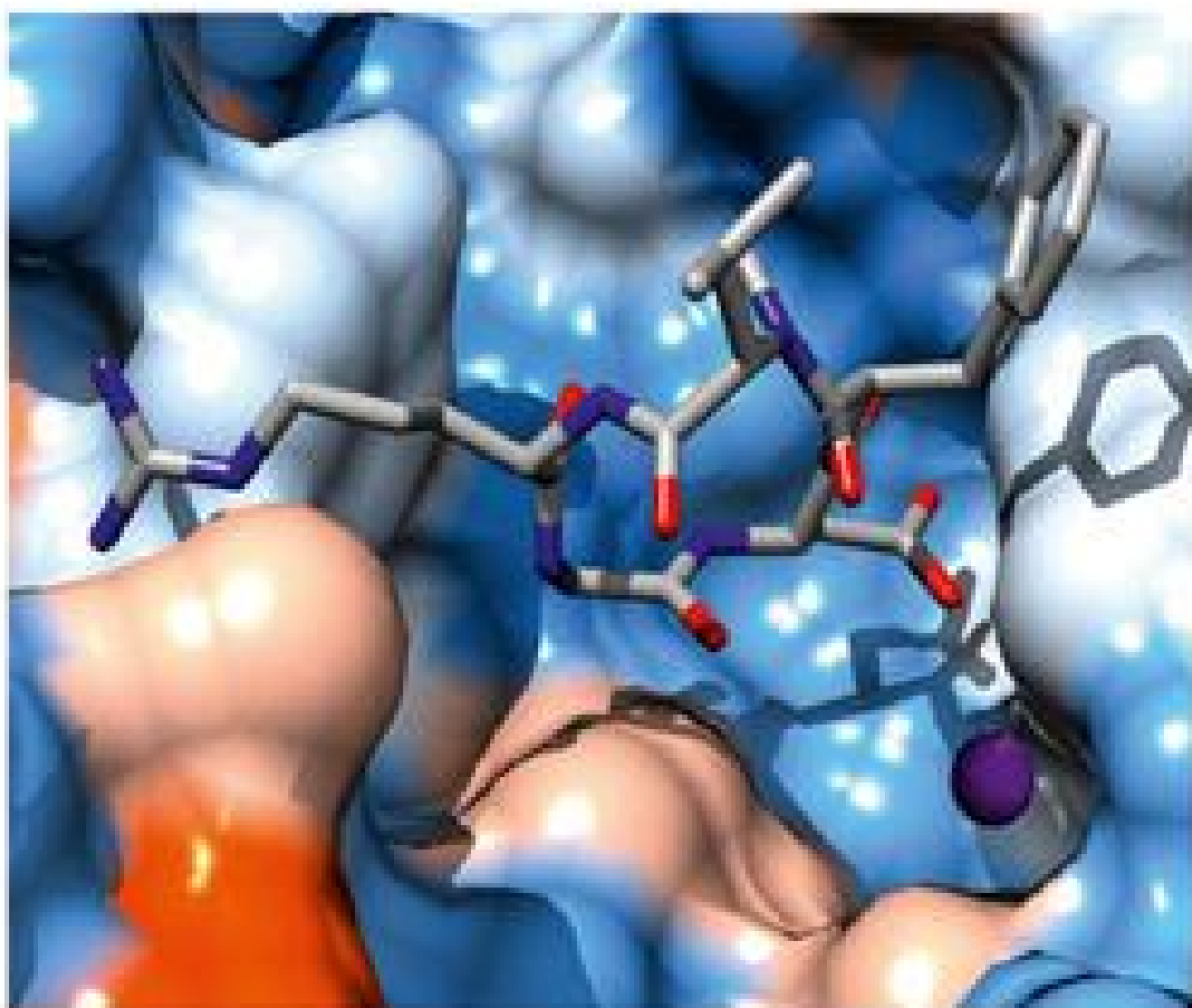


Norbert Sewald and
Hans-Dieter Jakubke

WILEY-VCH

Peptides: Chemistry and Biology

Second, Revised and Updated Edition



*Norbert Sewald and
Hans-Dieter Jakubke*

Peptides: Chemistry and Biology

Related Titles

Jakubke, H.-D., Sewald, N.

Peptides from A - Z

A Concise Encyclopedia

2008

ISBN: 978-3-527-31722-6

Groner, B. (Ed.)

Peptides as Drugs

Discovery and Development

2009

ISBN: 978-3-527-32205-3

Krauss, G.

Biochemistry of Signal Transduction and Regulation

Fourth, Enlarged and Improved Edition

2008

ISBN: 978-3-527-31397-6

Lindhorst, T. K.

Essentials of Carbohydrate Chemistry and Biochemistry

Third, Completely Revised and Enlarged Edition

2007

ISBN: 978-3-527-31528-4

Breslow, R. (ed.)

Artificial Enzymes

2005

ISBN: 978-3-527-31165-1

Whitford, D.

Proteins

Structure and Function

2005

ISBN: 978-0-471-49894-0

Schmuck, C., Wennemers, H. (eds.)

Highlights in Bioorganic Chemistry

Methods and Applications

2004

ISBN: 978-3-527-30656-5

Norbert Sewald and Hans-Dieter Jakubke

Peptides: Chemistry and Biology

Second, Revised and Updated Edition



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

The Authors

Prof. Dr. Norbert Sewald

Bielefeld University
Department of Chemistry
PO Box 100131
33501 Bielefeld
Germany
e-mail: norbert.sewald@uni-bielefeld.de

Prof. em. Dr. Hans-Dieter Jakubke

Leipzig University
Department of Biochemistry
Private address:
Höntzschstr. 1a
01465 Langebrück
Germany
e-mail:
Hans-Dieter.Jakubke@mailbox.tu-dresden.de

First Edition 2002

Second, Revised and Updated Edition 2009

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

Die Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Typesetting Thomson Digital, Noida, India

Printing Strauss GmbH, Mörlenbach

Binding Litges & Dopf GmbH, Heppenheim

Printed in the Federal Republic of Germany
Printed on acid-free paper

ISBN: 978-3-527-31867-4

Contents

Preface to the second edition XIII

Preface to the first edition XV

1	Introduction and Background	1
	References	4
2	Fundamental Chemical and Structural Principles	5
2.1	Definitions and Main Conformational Features of the Peptide Bond	5
2.2	Building Blocks, Classification, and Nomenclature	7
2.3	Analysis of the Covalent Structure of Peptides and Proteins	11
2.3.1	Separation and Purification	13
2.3.1.1	Separation Principles	13
2.3.1.2	Purification Techniques	17
2.3.1.3	Stability Problems	19
2.3.1.4	Evaluation of Homogeneity	20
2.3.2	Primary Structure Determination	20
2.3.2.1	End Group Analysis	21
2.3.2.2	Cleavage of Disulfide Bonds	24
2.3.2.3	Analysis of Amino Acid Composition	24
2.3.2.4	Selective Methods of Cleaving Peptide Bonds	26
2.3.2.5	N-Terminal Sequence Analysis (Edman Degradation)	28
2.3.2.6	C-Terminal Sequence Analysis	30
2.3.2.7	Mass Spectrometry	31
2.3.2.8	Peptide Ladder Sequencing	32
2.3.2.9	Assignment of Disulfide Bonds and Peptide Fragment Ordering	33
2.3.2.10	Location of Post-Translational Modifications and Bound Cofactors	34
2.4	Three-Dimensional Structure	36
2.4.1	Secondary Structure	36
2.4.1.1	Helices	37
2.4.1.2	β -Sheets	39
2.4.1.3	Turns	40
2.4.1.4	Amphiphilic Structures	42

2.4.2	Tertiary Structure	44
2.4.2.1	Structure Prediction	48
2.5	Methods of Structural Analysis	49
2.5.1	Circular Dichroism	49
2.5.2	Infrared Spectroscopy	51
2.5.3	NMR Spectroscopy	52
2.5.4	X-Ray Crystallography	54
2.5.5	UV Fluorescence Spectroscopy	55
2.6	Review Questions	56
	References	57
3	Biology of Peptides	63
3.1	Historical Aspects and Biological Functions	63
3.2	Biosynthesis	75
3.2.1	Ribosomal Peptide Synthesis	75
3.2.2	Post-Translational Modification	79
3.2.2.1	Enzymatic Cleavage of Peptide Bonds	79
3.2.2.2	Hydroxylation	80
3.2.2.3	Carboxylation	81
3.2.2.4	Glycosylation	81
3.2.2.5	Amidation	86
3.2.2.6	Phosphorylation	87
3.2.2.7	Lipidation	88
3.2.2.8	Pyroglutamyl Formation	90
3.2.2.9	Sulfation	91
3.2.2.10	Further Post-Translational Modifications	92
3.2.3	Nonribosomal Peptide Synthesis	94
3.3	Selected Biologically Active Peptides	96
3.3.1	Gastroenteropancreatic Peptide Families	96
3.3.1.1	The Gastrin Family	97
3.3.1.2	Secretin Family	98
3.3.1.3	The Insulin Superfamily	101
3.3.1.4	The Somatostatin Family	104
3.3.1.5	The Tachykinin Family	105
3.3.1.6	The Neuropeptide Y family	106
3.3.1.7	The Ghrelin Family	108
3.3.1.8	The EGF Family	109
3.3.2	Hypothalamic Liberins and Statins	110
3.3.2.1	Thyroliberin	112
3.3.2.2	Gonadoliberin	113
3.3.2.3	Corticoliberin	113
3.3.2.4	Growth Hormone-Releasing Hormone	114
3.3.3	Pituitary Hormones	115
3.3.3.1	Growth Hormone	115
3.3.3.2	Corticotropin	115

3.3.3.3	Melanotropin	117
3.3.4	Neurohypophyseal Hormones	118
3.3.4.1	Oxytocin	118
3.3.4.2	Vasopressin	119
3.3.5	Parathyroid Hormone and Calcitonin/Calcitonin Gene-Related Peptide Family	120
3.3.5.1	Parathyroid Hormone	120
3.3.5.2	Parathyroid Hormone-Related Peptides	121
3.3.5.3	The Calcitonin/Calcitonin Gene-Related Peptide Family	121
3.3.6	The Blood Pressure Regulating Peptide Families	123
3.3.6.1	Angiotensin–Kinin System	123
3.3.6.2	Endothelins and Endothelin-Like Peptides	125
3.3.6.3	Cardiac Peptide Hormones	127
3.3.7	Neuropeptides	128
3.3.7.1	Endorphins	131
3.3.7.2	Dynorphins	136
3.3.7.3	Hypocretins (Orexins)	136
3.3.7.4	Dermorphins	137
3.3.7.5	Deltorphins	138
3.3.7.6	Nociceptin/Orphanin and Nocistatin	138
3.3.7.7	Exorphins	139
3.3.7.8	The Adipokinetic Hormone/Red Pigment-Concentrating Hormone Family	140
3.3.7.9	Endomorphins	141
3.3.7.10	The Allatostatin Families	141
3.3.7.11	Neuromedins	142
3.3.7.12	Additional Neuroactive Peptides	143
3.3.8	Peptide Antibiotics	146
3.3.8.1	Nonribosomally Synthesized Peptide Antibiotics	147
3.3.8.2	Ribosomally Synthesized Peptide Antibiotics	152
3.3.9	Peptide Toxins	156
3.4	Review Questions	162
	References	163
4	Peptide Synthesis	175
4.1	Principles and Objectives	175
4.1.1	Main Targets of Peptide Synthesis	175
4.1.2	Basic Principles of Peptide Bond Formation	178
4.2	Protection of Functional Groups	181
4.2.1	N ^α -Amino Protection	182
4.2.1.1	Alkoxy carbonyl-Type (Urethane-Type) Protecting Groups	183
4.2.1.2	Carboxamide-Type Protecting Groups	192
4.2.1.3	Sulfonamide and Sulfenamide-Type Protecting Groups	192
4.2.1.4	Alkyl-Type Protecting Groups	192
4.2.2	C ^α -Carboxy Protection	193

4.2.2.1	Esters	194
4.2.2.2	Amides and Hydrazides	199
4.2.3	C-Terminal and Backbone N ^α -carboxamide Protection	199
4.2.4	Side-Chain Protection	201
4.2.4.1	Guanidino Protection	202
4.2.4.2	ω-Amino Protection	204
4.2.4.3	ω-Carboxy Protection	205
4.2.4.4	Thiol Protection	208
4.2.4.5	Imidazole Protection	211
4.2.4.6	Hydroxy Protection	214
4.2.4.7	Thioether Protection	216
4.2.4.8	Indole Protection	217
4.2.4.9	ω-Amide Protection	218
4.2.5	Enzyme-Labile Protecting Groups	220
4.2.5.1	Enzyme-labile N ^α -amino Protection	221
4.2.5.2	Enzyme-labile C ^α -carboxy Protection and Enzyme-labile Linker Moieties	223
4.2.6	Protecting Group Compatibility	223
4.3	Peptide Bond Formation	224
4.3.1	Acyl Azides	225
4.3.2	Anhydrides	226
4.3.2.1	Mixed Anhydrides	227
4.3.2.2	Symmetrical Anhydrides	229
4.3.2.3	N-Carboxy Anhydrides	229
4.3.3	Carbodiimides	231
4.3.4	Active Esters	235
4.3.5	Acyl Halides	237
4.3.6	Phosphonium Reagents	239
4.3.7	Guanidinium/Uronium Reagents	240
4.3.8	Immonium Type Coupling Reagents	242
4.3.9	Further Special Methods for Peptide Synthesis	243
4.4	Racemization During Synthesis	246
4.4.1	Direct Enolization	246
4.4.2	5(4H)-Oxazolone Mechanism	247
4.4.3	Racemization Tests: Stereochemical Product Analysis	249
4.5	Solid-Phase Peptide Synthesis (SPPS)	251
4.5.1	Solid Supports and Linker Systems	253
4.5.2	Safety-Catch Linkers	262
4.5.3	Protection Schemes	265
4.5.3.1	Boc/Bzl-protecting Groups Scheme (Merrifield Tactics)	265
4.5.3.2	Fmoc/tBu-Protecting Groups Scheme (Sheppard Tactics)	267
4.5.3.3	Three- and More-Dimensional Orthogonality	268
4.5.4	Chain Elongation	269
4.5.4.1	Coupling Methods	269
4.5.4.2	Undesired Problems During Elongation	269

4.5.4.3	Difficult Sequences	271
4.5.4.4	Chemical Strategies for SPPS Methodological Improvements	273
4.5.4.5	On-Resin Monitoring	273
4.5.5	Automation of the Process	274
4.5.6	Peptide Cleavage from the Resin	275
4.5.6.1	Acidolytic Methods	275
4.5.6.2	Side Reactions	276
4.5.6.3	Advantages and Disadvantages of the Boc/Bzl and Fmoc/tBu Schemes	277
4.5.7	Examples of Syntheses by Linear SPPS	277
4.5.8	Special Methods of Polymer-supported Synthesis	278
4.5.9	Microwave-Enhanced Peptide Synthesis	280
4.6	Biochemical Synthesis	281
4.6.1	Recombinant DNA Techniques	281
4.6.1.1	Principles of DNA Technology	282
4.6.1.2	Examples of Synthesis by Genetic Engineering	285
4.6.1.3	Cell-free Translation Systems	288
4.6.1.4	Proteins Containing Non-Proteinogenic Amino Acids – The Expansion of the Genetic Code	290
4.6.2	Enzymatic Peptide Synthesis	291
4.6.2.1	Approaches to Enzymatic Synthesis	291
4.6.2.2	Manipulations to Suppress Competitive Reactions	294
4.6.2.3	Substrate Mimetic Approach	295
4.6.3	Further Selected Biochemical Methods	297
4.6.3.1	Non-ribosomal Peptide Synthesis	297
4.6.3.2	Peptide Bond Formation by LF-Transferase	297
4.6.3.3	Antibody-catalyzed Peptide Bond Formation	297
4.7	Review Questions	300
	References	301
5	Synthesis Concepts for Peptides and Proteins	317
5.1	Strategy and Tactics	317
5.1.1	Linear or Stepwise Synthesis	317
5.1.2	Convergent Synthesis	320
5.1.3	Tactical Considerations	321
5.1.3.1	Selected Protecting Group Schemes	321
5.1.3.2	Preferred Coupling Procedures	324
5.2	Solution Phase Synthesis (SPS)	325
5.2.1	Convergent Synthesis Using Maximally Protected Segments	325
5.2.2	Convergent Synthesis Using Minimally Protected Segments	327
5.2.2.1	Chemical Approaches	327
5.2.2.2	Enzymatic Approaches	329
5.3	Solution Phase/Solid Phase-Hybrid Approaches	332
5.3.1	Solid Phase Synthesis of Protected Segments	332
5.3.2	SPS/SPPS-Hybrid Condensation of Lipophilic Segments	333

5.3.3	Phase Change Synthesis	335
5.3.4	SPS/SPPS-Hybrid Approach to Protein and Large Scale Peptide Synthesis	335
5.4	Optimized Strategies on a Polymeric Support	337
5.4.1	Standard SPPS	337
5.4.2	Convergent Solid-Phase Peptide Synthesis	339
5.4.3	Handle Approaches	341
5.4.3.1	Positively Charged Handles	341
5.4.3.2	Liquid Phase Method	342
5.4.3.3	Excluded Protecting Group Method	343
5.5	Chemical Ligation Strategies	343
5.5.1	Native Chemical Ligation	344
5.5.1.1	Facile Peptide Thioester Synthesis	346
5.5.1.2	Extended Native Chemical Ligation	346
5.5.1.3	Kinetically Controlled Ligation	348
5.5.1.4	Solid Phase Chemical Ligation	350
5.5.1.5	Alternative Approaches to Native Chemical Ligation	350
5.5.2	Expressed Protein Ligation (Intein-mediated Protein Ligation)	351
5.5.3	Prior Capture-mediated Ligation	353
5.5.3.1	Template-mediated Ligation	353
5.5.4	Non-native Chemoselective Ligation	355
5.5.4.1	Thioester- and Thioether-forming Ligations	355
5.5.4.2	Hydrazone- and Oxime-forming Ligations	356
5.5.5	Alternative Ligation Approaches	356
5.5.5.1	Staudinger Ligation	357
5.5.5.2	Ketoacid-hydroxylamine Amide Ligations	357
5.5.5.3	Expressed enzymatic ligation	358
5.5.5.4	Sortase-mediated Ligation	359
5.6	Review Questions	359
	References	360
6	Synthesis of Special Peptides and Peptide Conjugates	365
6.1	Cyclopeptides	365
6.1.1	Backbone Cyclization (Head-to-Tail Cyclization)	371
6.1.2	Side Chain-to-Head and Tail-to-Side Chain Cyclizations	380
6.1.3	Side Chain-to-Side Chain Cyclizations	380
6.2	Cystine Peptides	381
6.3	Glycopeptides	386
6.4	Phosphopeptides	395
6.5	Lipopeptides	398
6.6	Sulfated Peptides	402
6.7	Review Questions	403
	References	403
7	Peptide and Protein Design, Pseudopeptides, and Peptidomimetics	411
7.1	Peptide Design	413

7.2	Modified Peptides	418
7.2.1	Side-Chain Modification	418
7.2.2	Backbone Modification	421
7.2.3	Combined Modification (Global Restriction) Approaches	423
7.2.4	Modification by Secondary Structure Mimetics	425
7.2.5	Transition State Inhibitors	427
7.3	Peptidomimetics	428
7.4	Pseudobiopolymers	431
7.4.1	Peptoids	432
7.4.2	Peptide Nucleic Acids (PNA)	434
7.4.3	β -Peptides, Hydrazino Peptides, Aminoxy Peptides, and Oligosulfonamides	435
7.4.4	Oligocarbamates	437
7.4.5	Oligopyrrolinones	438
7.5	Macropeptides and <i>de novo</i> Design of Peptides and Proteins	439
7.5.1	Protein Design	439
7.5.2	Peptide Dendrimers	444
7.5.3	Peptide Polymers	447
7.6	Review Questions	447
	References	448
8	Combinatorial Peptide Synthesis	457
8.1	Parallel Synthesis	460
8.1.1	Synthesis in Teabags	461
8.1.2	Synthesis on Polyethylene Pins (Multipin Synthesis)	462
8.1.3	Parallel Synthesis of Single Compounds on Cellulose or Polymer Strips	464
8.1.4	Light-Directed, Spatially Addressable Parallel Synthesis	465
8.1.5	Liquid-Phase Synthesis using Soluble Polymeric Support	466
8.2	Synthesis of Mixtures	467
8.2.1	Reagent Mixture Method	468
8.2.2	Split and Combine Method	468
8.2.3	Encoding Methods	470
8.2.4	Peptide Library Deconvolution	474
8.2.5	Dynamic Combinatorial Libraries	476
8.2.6	Biological Methods for the Synthesis of Peptide Libraries	477
8.3	Review Questions	478
	References	479
9	Application of Peptides and Proteins	483
9.1	General Production Strategies	483
9.2	Improvement of the Therapeutic Potential	486
9.2.1	Peptide and Protein Drug Modifications	486
9.2.2	Peptide Drug Delivery Systems	488
9.3	Protein Pharmaceuticals	492
9.3.1	Importance and Sources	492

9.3.2	Endogenous Pharmaceutical Proteins	493
9.3.3	Engineered Protein Pharmaceuticals	493
9.3.3.1	Selected Recombinant Proteins	493
9.3.3.2	Peptide-Based Vaccines	497
9.3.3.3	Monoclonal Antibodies	498
9.3.3.4	Future Perspectives	500
9.4	Peptide Pharmaceuticals	502
9.4.1	Large-Scale Peptide Synthesis	502
9.4.2	Peptide Drugs and Drug Candidates	507
9.4.3	Peptides as Tools in Drug Discovery	515
9.4.4	Peptides Targeted to Functional Sites of Proteins	517
9.4.5	Peptides Used in Target Validation	517
9.4.6	High-throughput Screening (HTS) Using Peptides as Surrogate Ligands	518
9.4.7	Artificial Peptide Analogs in Drug Discovery	521
9.5	Review Questions	522
	References	523
10	Peptides in Proteomics	529
10.1	Genome and Proteome	529
10.2	Separation Methods	530
10.2.1	Depletion Strategies	530
10.2.2	Two-Dimensional Polyacrylamide Gel-Electrophoresis	530
10.2.3	Gel-Free Methods – Two-Dimensional Liquid Chromatography (2D-LC, MudPIT)	531
10.3	Peptide and Protein Analysis in Proteomics	532
10.3.1	Mass Spectrometry	532
10.3.2	Quantitative Proteomics	533
10.3.2.1	Metabolic Stable-Isotope Labelling	533
10.3.2.2	Tagging Methods	533
10.3.2.3	Enzymatic Stable-Isotope Labeling	535
10.4	Activity-Based Proteomics	535
10.4.1	Irreversibly Binding Affinity-Based Probes	536
10.4.2	Reversibly Binding Affinity-Based Probes	539
10.4.2.1	Inhibitor Affinity Chromatography (IAC)	540
10.4.2.2	Labelling Strategies with Reversibly Binding Protein Ligands	541
10.5	Review Questions	543
	References	543
	Glossary	547
	Index	559

Preface to the second edition

Peptides are continuously gaining increasing attention for application as lead compounds for the development of drugs, as drug molecules, as molecular tools for diagnostic purposes in biochemistry, medicinal sciences as well as proteome research. Since the first edition of this book has been published, a storming development in the peptide field took place. While formerly the development of synthetic methodology prevailed, currently peptide application in a biochemical, medicinal or analytical context predominates. Many different peptide conjugates have been tailored according to specific scientific questions.

The second thoroughly revised edition of this book especially takes into account these recent developments. Many sections have been updated carefully and some of them have been re-organized and extended significantly. This is especially true for Chapter 3, dealing with the biology of peptides, Chapter 5 that focuses on synthesis concepts for peptides and proteins, as well as Chapter 6 on special peptides and peptide conjugates. The recent developments of peptide drugs required an in-depth revision of Chapter 9 which is now an up to date reference on currently available peptide and protein therapeutics. The emerging role of peptides in an analytical context called for the addition of a new Chapter 10 providing details on the role of peptides on proteomics.

The second edition of our book also very much relies on advice, help and contributions by several colleagues. While it would lead too far to mention all of them, the most significant contributions have to be acknowledged in this preface. Chapter 2 was critically revised by Hans-Jörg Hofmann, University of Leipzig. The Section 6.2 was re-written by Luis Moroder, Max-Planck-Institute of Biochemistry, Martinsried. Dirk Ullmann and Volker Schellenberger usefully contributed to updating Chapter 9 on the application of peptides and proteins. Dr. Katherina Sewald prepared excellent graphical material and molecular formulae. Jens Conradi designed the new cover picture. Dr. Frank Weinreich, Lesley Belfit, Maike Peterson, Dr. Rosemary Whitelock, and Claudia Zschernitz, who were responsible for editing and production at Wiley-VCH, are acknowledged for carefully dealing with the manuscript and converting it into a book of high quality.

We very much hope that this second edition will meet the same positive response and enthusiastic comments by our colleagues as the first edition did.

Finally we would like to dedicate this book on Chemistry and Biology of Peptides to the memory of two pioneers in the field who favorably commented on the first edition, but passed away in the meantime: Robert B. Merrifield and Murray Goodman.

Bielefeld
and
Dresden-Langebrück
August 2008

Norbert Sewald

Hans-Dieter Jakubke

Preface to the first edition

The past decades have witnessed an enormous development in peptide chemistry with regard not only to the isolation, synthesis, structure identification, and elucidation of the mode of action of peptides, but also to their application as tools within the life sciences. Peptides have proved to be of interest not only in biochemistry, but also in chemistry, biology, pharmacology, medicinal chemistry, biotechnology, and gene technology.

These important natural products span a broad range with respect to their complexity. As the different amino acids are connected via peptide bonds to produce a peptide or a protein then many different sequences are possible – depending on the number of different building blocks and on the length of the peptide. As all peptides display a high degree of conformational diversity, it follows that many diverse and highly specific structures can be observed.

Whilst many previously published monographs have dealt exclusively with the synthetic aspects of peptide chemistry, this new book also covers its biological aspects, as well as related areas of peptidomimetics and combinatorial chemistry. The book is based on a monograph which was produced in the German language by Hans-Dieter Jakubke: *Peptide, Chemie und Biologie* (Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford), and first published in 1996. In this new publication, much of the material has been completely reorganized and many very recently investigated aspects and topics have been added. We have made every effort to produce a practically new book, in a modern format, in order to provide the reader with profound and detailed knowledge of this field of research. The glossary, which takes the form of a concise encyclopedia, contains data on more than 500 physiologically active peptides and proteins, and comprises about 20% of the book's content.

Our book covers many different issues of peptide chemistry and biology, and is devoted to those students and scientists from many different disciplines who might seek quick reference to an essential point. In this way it provides the reader with concise, up-to-date information, as well as including many new references for those who wish to obtain a deeper insight into any particular issue. In this book, the “virtual barrier” between peptides and proteins has been eliminated because, from

the viewpoint of the synthesis or biological function of these compounds, such a barrier does not exist.

This monograph represents a personal view of the authors on peptide chemistry and biology. We are aware however that, despite all our efforts, it is impossible to include all aspects of peptide research in one volume. We are not under the illusion that the text, although carefully prepared, is completely free of errors. Indeed, some colleagues and readers might feel that the choice of priorities, the treatment of different aspects of peptide research, or the depth of presentation may not always be as expected. In any case, comments, criticisms and suggestions are appreciated and highly welcome for further editions.

Several people have contributed considerably to the manuscript. All the graphical material was prepared by Dr. Katherina Stembera, who also typed large sections of the manuscript, provided valuable comments, and carried out all the formatting. We appreciate the kindness of Professor Robert Bruce Merrifield, Dr. Bernhard Streb and Dr. Rainer Obermeier for providing photographic material for our book. Margot Müller and Helga Niermann typed parts of the text. Dr. Frank Schumann and Dr. Jörg Schröder contributed Figures 2.19 and 2.25, respectively. We also thank Dirk Bächle, Kai Jenssen, Micha Jost, Dr. Jörg Schröder, and Ulf Strijowski for comments and proofreading parts of the manuscript.

Dr. Gudrun Walter, Maike Petersen, Dr. Bill Down, and Hans-Jörg Maier took care that the manuscript was converted into this book in a rather short period of time, without complications.

Bielefeld
and
Dresden-Langebrück
April 2002

Norbert Sewald

Hans-Dieter Jakubke

1

Introduction and Background

Peptide research has experienced considerable development during the past few decades. The progress in this important discipline of natural product chemistry is reflected in a flood of scientific data. The number of scientific publications per year increased from about 10 000 in the year 1980 to presently more than 20 000 papers. The introduction of new international scientific journals in this research area reflects this remarkable development.

A very useful bibliography on peptide research was published by John H. Jones [1]. The Houben-Weyl sampler volume E 22 "Synthesis of Peptides and Peptidomimetics" edited by Murray Goodman (Editor-in-Chief), Arthur Felix, Luis Moroder and Claudio Toniolo [2] represents the most comprehensive general treatise in this field. This work is a tribute to the 100th anniversary of Emil Fischer's first synthesis of peptides and is the successor of two Houben-Weyl volumes, in German, edited by Erich Wünsch in 1974 [3].

A number of very important physiological and biochemical functions of life are influenced by peptides. Peptides are involved as neurotransmitters, neuromodulators, and hormones in receptor-mediated signal transduction. More than 100 peptides with functions in the central and peripheral nervous systems, in immunological processes, in the cardiovascular system, and in the intestine are known. Peptides influence cell-cell communication upon interaction with receptors, and are involved in a number of biochemical processes, for example metabolism, pain, reproduction, and immune response.

The increasing knowledge of the manifold modes of action of bioactive peptides led to an increased interest of pharmacology and medical sciences in this class of compounds. The isolation and targeted application of these endogenous substances as potential intrinsic drugs are gaining importance for the treatment of pathologic processes. New therapeutic methods based on peptides for a series of diseases give rise to the hope that diseases, where peptides play a functional role, can be amenable to therapy.

Peptide chemistry contributes considerably to research in the life science area. Synthetic peptides serve as antigens to raise antibodies, as enzyme substrates to map the active site requirements of an enzyme under investigation, or as enzyme inhibitors to influence signaling pathways in biochemical research or pathologic processes in medical research. Peptide ligands immobilized to a solid matrix may

facilitate specific protein purification. Protein–protein interaction can be manipulated by small synthetic peptides. The “peptide dissection approach” uses relatively short peptide fragments that are part of a protein sequence. The synthetic peptides are investigated for their ability to fold independently, with the aim to improve the knowledge on protein folding.

The isolation of peptides from natural sources, however, is often problematic. In many cases, the concentration of peptide mediators ranges from 10^{-15} to 10^{-12} mol per mg fresh weight of tissue. Therefore, only highly sensitive assay methods, such as immunohistochemical techniques, render cellular localization possible. Although not all relevant bioactive peptides occur in such low concentrations, isolation methods generally suffer from disadvantages, such as the limited availability of human tissue sources. Complicated logistics during collection or storage of the corresponding organs, e.g., porcine or bovine pancreas for insulin production, additionally impose difficulties on the utilization of natural sources. Possible contamination of tissue used for the isolation of therapeutic peptides and proteins with pathogenic viruses is an enormous health hazard. Factor VIII preparations for treatment of hemophilia patients, isolated from natural sources, have been contaminated with human immunodeficiency virus (HIV), while impure growth hormone preparations isolated from human hypophyses after autopsy have led to the transmission of central nerve system diseases (Creutzfeld–Jacob disease). Nowadays, many therapeutic peptides and proteins are produced by recombinant techniques. Immunological incompatibilities of peptide drugs obtained from animal sources have also been observed. Consequently, the development of processes for the synthesis of peptide drugs must be pursued with high priority.

Chemical peptide synthesis is the classical method which has been mainly developed during the past four decades, although the foundations were laid in the early 20th century by Theodor Curtius and Emil Fischer. Synthesis has often been the final structural proof of many peptides isolated only in minute amounts from natural sources.

The production of polypeptides and proteins by recombinant techniques has also contributed important progress in terms of methodology. Genetically engineered pharmaproteins verify the concept of therapy with endogenous protein drugs (endopharmaceuticals). Cardiovascular diseases, tumors, auto-immune diseases and infectious diseases are the most important indications. Classical peptide synthesis has, however, not been questioned by the emergence of these techniques. Small peptides, like the artificial sweetener aspartame (which has an annual production of more than 5000 tons), and peptides of medium size remain the objectives of classical synthesis, not to mention derivatives with nonproteinogenic amino acids or selectively labeled (^{13}C , ^{15}N) amino acid residues for structural investigations using nuclear magnetic resonance (NMR).

The demand for synthetic peptides in biological applications is steadily increasing. At present, following the sequencing of the human genome, peptides have become a focus of biotechnological research. Their importance is growing as they represent the essential bioactive molecules inside the biological systems. The research fields of genomics and proteomics generate a huge number of new peptide targets which can

help combat diseases and strengthen the immune resistance. The new targets do not allow for an isolated position of peptide chemistry exclusively oriented toward synthesis. Modern interdisciplinary science and research require synthesis, analysis, isolation, structure determination, conformational analysis and molecular modeling as integrated components of a cooperation between biologists, biochemists, pharmacologists, medical scientists, biophysicists, and bioinformaticians. Studies on structure–activity relationships (SAR) involve a large number of synthetic peptide analogues with sequence variation and the introduction of nonproteinogenic building blocks. The ingenious concept of solid-phase peptide synthesis has exerted considerable impact on the life sciences, whilst methods of combinatorial peptide synthesis allow the simultaneous creation of peptide libraries which contain at least several hundred different peptides. The high yields and purities enable both *in vitro* and *in vivo* screening of biological activity to be carried out. Special techniques enable the creation of peptide libraries that contain several hundred thousand peptides; these techniques offer an interesting approach in the screening of new lead structures in pharmaceutical developments.

Peptide drugs, however, can be applied therapeutically only to a limited extent because of their chemical and enzymatic labilities. Many peptides are inactive when applied orally, and even parenteral application (intravenous or subcutaneous injection) is often not efficient because proteolytic degradation occurs on the locus of the application. However, sophisticated formulation techniques are very encouraging as they are capable of significantly enhancing the oral bioavailability of discovered active peptides. Application via mucous membranes (e.g., nasal) is promising. Despite the utilization of special depot formulations and new applications systems (computer-programmed minipump implants, iontophoretic methods, etc.) a major strategy in peptide chemistry is directed towards chemical modification in order to increase its chemical and enzymatic stability, to prolong the time of action, and to increase activity and selectivity towards the receptor.

The synthesis of analogues of bioactive peptides with unusual amino acid building blocks, linker or spacer molecules and modified peptide bonds is directed towards the development of potent agonists and antagonists of endogenous peptides. Once the amino acids of a protein that are essential for the specific biological mode of action have been revealed, these pharmacophoric groups may be incorporated into a small peptide. The development of orally active drugs is an important target. Rational drug design has contributed extensively in the development of protease-resistant structural variants of endogenous peptides, and in this context the incorporation of D-amino acids, the modification of covalent bonds, and the formation of ring structures (cyclopeptides) must be mentioned.

Peptidomimetics imitate bioactive peptides. The original peptide structure can hardly be recognized in these molecules, which induce a physiological effect by specific interaction with the corresponding receptor. Hence, a peptide structure may be transformed into a nonpeptide drug. This task is another timely challenge for peptide chemists, because only sufficient knowledge of the biologically active conformation of a peptide drug and of the interaction with the specific receptor enable the rational design of such peptide mimetics.

Today nearly 300 new peptide-based drugs for broad indications like cardiovascular diseases, bone metabolism disorders, type II diabetes and viral infections are at different stages of development. A breakthrough for peptides as important pharmaceuticals was the discovery of the 36-peptide enfuvirtide (T20; Fuzeon™) which is capable of docking on the surface of the AIDS virus and therefore blocking the virus from entering into human blood cells. The discovery of T20 has initiated research on viral entry inhibitors that form a new class of antiviral drugs. T20 is the first peptide-based drug to be produced on a metric ton scale by solid-phase peptide synthesis.

Interestingly, nanotubular structures based on cyclic tri- β -peptides have been synthesized that could be used in molecular electronic devices. Functional material from peptides promotes the development of nanotubular systems with interior and exterior functionalization which will widen the scope of these nanotubular structures to medical chemistry.

The variety of the tasks described herein renders peptide research an important and attractive discipline of modern life sciences. Despite the development of gene technology, peptide chemistry will have excellent future prospects because gene technology and peptide chemistry are complementary approaches.

References

- 1 J. H. Jones, *J. Pept. Sci.* **2000**, 6, 201.
- 2 M. Goodman, A. Felix, L. Moroder, C. Toniolo, Synthesis of Peptides and Peptidomimetics in *Houben-Weyl-Methoden der organischen Chemie*, Vol. E22, K. H. Büchel (Ed.), Thieme, Stuttgart, **2002**.
- 3 E. Wünsch, Synthese von Peptiden, in *Houben-Weyl Methoden der organischen Chemie*, Vol. 15, 1/2, E. Müller (Ed.), Thieme, Stuttgart, **1974**.

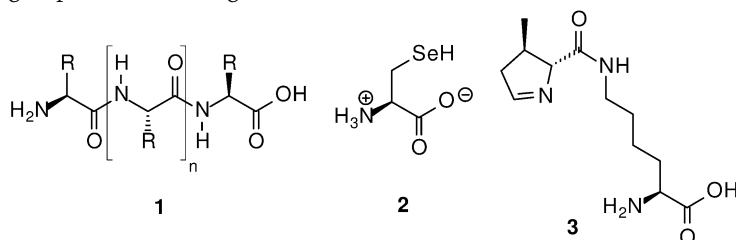
2

Fundamental Chemical and Structural Principles

2.1

Definitions and Main Conformational Features of the Peptide Bond

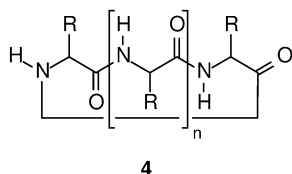
Peptides **1** are, formally, oligomers or polymers of amino acids, connected by amide bonds (peptide bonds) between the carboxy group of one amino acid and the amino group of the following amino acid.



Natural peptides and proteins encoded by DNA usually contain up to 22 different α -amino acids (including the secondary amino acid proline and the rare amino acids selenocysteine **2** and pyrrolysine **3**). The different side chains R of the amino acids contribute fundamentally to their biochemical mode of action. A collection of the names, structures, three-letter code, and one-letter code abbreviations of these proteinogenic amino acids is given on the inside front cover of this book.

Selenocysteine (Sec, U) **2**, which is found both in prokaryotes and eukaryotes, is encoded by a special tRNA with the anticodon UCA recognizing UGA triplets on mRNA. It is incorporated into proteins by ribosomal synthesis. Normally, the UGA codon serves as a stop codon.

Pyrrolysine (Pyl, O) **3** occurs in some methanogenic Archaea and was first isolated from *Methanosarcina barkeri* [1]. Pyl appears to be limited to the *Methanosarcinaceae* and the Gram-positive *Desulfotobacterium hafniense*.



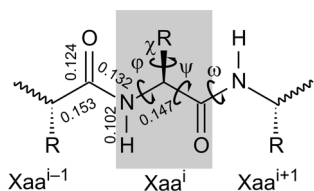


Figure 2.1 Torsion angles ϕ , ψ , ω , and χ^1 and bond lengths of the amino acid Xaa^i in a peptide.

Besides the great variety of linear peptides, there are cyclic peptides **4** of different ring sizes. Formally, homodetic cyclic peptides are formed by macrocyclization upon formation of a peptide bond between the amino and carboxy termini of a linear peptide.

In 1951, Pauling and Corey proved by X-ray crystallography of amino acids, amino acid amides, and simple linear peptides that the C–N bond length in a peptide bond is shorter than typical C–N single bonds and the CO double bond is longer than in aldehydes and ketones. Delocalization within the peptide confers partial double bond character onto the C–N bond. The conformation of the peptide backbone is characterized by the three torsion angles ϕ [$\text{C}(=\text{O})\text{--N--C}^\alpha\text{--C}(=\text{O})$], ψ [$\text{N--C}^\alpha\text{--C}(=\text{O})\text{--N}$], and ω [$\text{C}^\alpha\text{--C}(=\text{O})\text{--N--C}^\alpha$], as depicted in Figure 2.1.

The free rotation around the C–N amide bond is drastically restricted, because of the partial double bond character, with a rotational barrier of $65\text{--}90\text{ kJ mol}^{-1}$ dependent on the environment. Consequently, two rotamers of the peptide bond exist (Figure 2.2): the *trans*-configured peptide bond ($\omega = 180^\circ$) and the *cis*-configured peptide bond ($\omega = 0^\circ$). The former is energetically favored by approximately 8 kJ mol^{-1} and is preferred in most peptides. In cases where the amino group of the secondary amino acid proline is involved in a peptide bond, the energy of the *trans*-configured Xaa–Pro bond is increased. Consequently, the energy difference between the *cis* and *trans* isomers decreases.

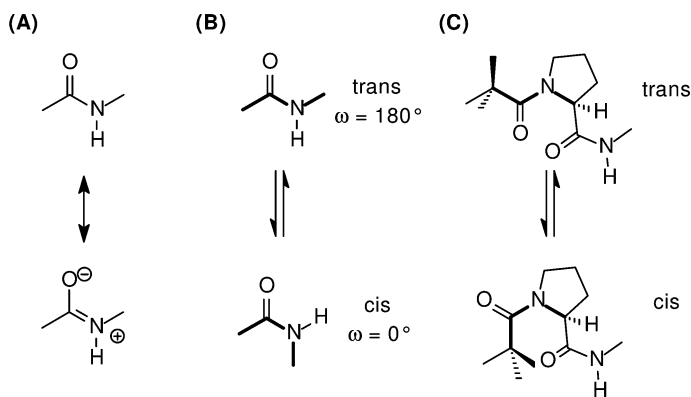
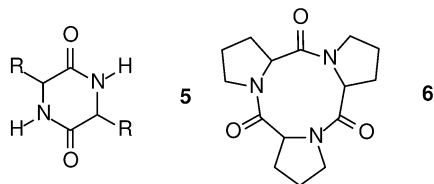


Figure 2.2 (A) Resonance stabilization of the peptide bond, (B) *cis/trans* isomerization and (C) *cis/trans* isomers of a Xaa–Pro bond.

The percentage of *cis*-configured Xaa-Pro bonds (6.5%) is approximately two orders of magnitude higher compared to *cis* peptide bonds between all other amino acids (0.05% *cis*). Several examples are known, where a peptide bond configuration in proteins has been assigned erroneously to be *trans* in X-ray crystallographic studies. The *cis/trans* isomerization of peptide bonds involving the secondary amino group of proline usually takes place in many proteins and plays an important role in folding and regulation phenomena. It has a half-life between 10 and 1000 s. *Peptidyl prolyl-cis/trans isomerases* (PPIases, EC 5.2.1.8) catalyze the isomerization of peptide bonds preceding the proline residue (prolyl bond) [2–4]. They are ubiquitous and abundant enzymes [3, 5, 6]. The human genome encodes certainly more than 40 different PPIases. Subfamilies of this enzyme class comprise cyclophilins (Cyp), FK506-binding proteins (FKBPs), parvulins, and protein phosphatase 2A phosphatase activator (PTPA) proteins. Cytosolic cyclophilins and FKBP are high-affinity receptors for the immunosuppressants cyclosporin A and FK506, respectively. In contrast to the classical mechanism of molecular chaperone action, PPIases catalyze a distinct chemical reaction, namely the rotation about prolyl peptide bonds. Besides their putative function in accelerating slow folding steps, compartmentalized PPIases are likely to have pleiotropic effects within cells.

Cis peptide bonds are present in the diketopiperazines **5**, which can be considered as cyclic dipeptides. Cyclic tripeptides with three *cis* peptide bonds are stable. As proline does not stabilize *trans*-configured peptide bonds, cyclo-(-Pro-Pro-Pro-) **6** and cyclo-(-Pro-Pro-Sar-) can be synthesized.

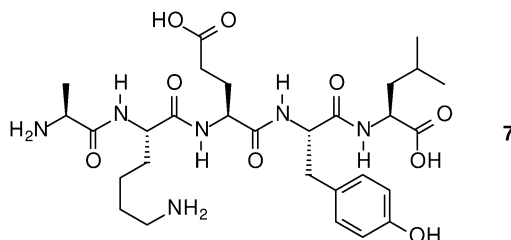


2.2

Building Blocks, Classification, and Nomenclature

Peptides are classified with Greek prefixes as di-, tri-, tetra-, penta-, . . . , octa-, nona-, decapeptides, etc., according to the number of amino acid residues incorporated. In longer peptides, the Greek prefix may be replaced by Arabic figures; for example, a decapeptide may be called 10-peptide, while a dodecapeptide is called 12-peptide. Formerly, peptides containing less than 10 amino acid residues were classified as “oligopeptides” (Greek *oligos* = few) and peptides with 10–100 amino acid residues were called “polypeptides”. The expression “protein” was used for derivatives containing more than 100 amino acids. This classification is absolutely formal and no longer strictly employed. Nowadays, one can find the notation “protein” also for polypeptides consisting of 50–100 amino acids.

The nomenclature formally considers peptides as N-acyl amino acids. Only the amino acid residue at the carboxy terminus of the peptide chain keeps the original



Alanyl-lysyl-glutamyl-tyrosyl-leucine

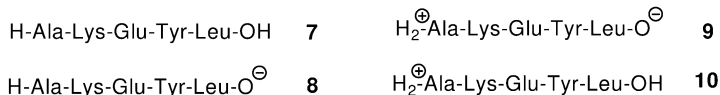


Figure 2.3 Structural formula, nomenclature, and three-letter code of the pentapeptide **7** in different ionic states (**8–10**).

name without suffix, all others are used with the original name and the suffix -yl (Figure 2.3). Consequently, peptide **7** is called alanyl-lysyl-glutamyl-tyrosyl-leucine.

A further simplification of a peptide formula is achieved by the three-letter code for amino acids (see inside cover). Linear peptide sequences are usually written horizontally, starting with the amino terminus on the left side and the carboxy terminus on the right side. When nothing is shown attached to either side of the three-letter symbol it should be understood that the amino group (always on the left) and carboxy group, respectively, are unmodified. This can be emphasized, e.g., Ala-Ala is equal to H-Ala-Ala-OH. Indicating free termini by presenting the terminal group is wrong. $\text{H}_2\text{N-Ala-Ala-COOH}$ implies a hydrazino group at one end and an α -keto acid derivative at the other. Representation of a free terminal carboxy group by writing H on the right is also wrong, because that implies a C-terminal aldehyde. Side chains are understood to be unsubstituted if nothing is shown, but a substituent can be indicated by use of brackets or attachment by a vertical bond up or down.

In peptide **7** alanine is the N-terminal amino acid and leucine the C-terminal amino acid. The three-letter code Ala-Lys-Glu-Tyr-Leu represents the pentapeptide **7** independent of the ionization state. If discrete ionic states of peptides should be emphasized, formulae **8**, **9** or **10** may be used for the anion, the zwitterion, and the cation, respectively.

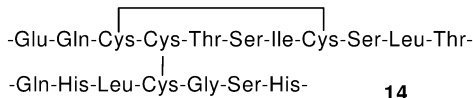
The three-letter code usually precludes that trifunctional amino acids with additional amino or carboxy functions located in the side chains (Lys, Glu, Asp) are connected by α -peptide bonds. This means, that the peptide bond is regularly formed between the C-1 (CO) of one amino acid and N-2 (N^{α}) of another amino acid. Special attention must be paid to abbreviations of isopeptide bonds in the three-letter code. The biochemically important peptide glutathione **11** (Figure 2.4) comprises, besides an α -peptide bond, also a γ -peptide bond. α -Glutamyl-lysine, N^{ϵ} - α -glutamyl-lysine, and N^{ϵ} - γ -glutamyl-lysine are constitution isomers of glutathione.

already included in the three-letter code representation and, therefore, not noted separately. An abbreviation Lys(NHBoc) would imply a Boc-protected ϵ -hydrazino group, and, likewise, the abbreviation Tyr(OtBu) would imply a *tert*-butylperoxo substituent.

The number and sequence of amino acids that are connected to give a peptide or a protein are called the “primary structure”. If the sequence of a peptide is completely known, the three-letter code symbols are listed sequentially, divided by a hyphen “-”, which symbolizes the peptide bond. Notably, if the full peptide sequence is given, for instance Ala-Lys-Glu-Tyr-Leu, no hyphens are written at the termini. However, if a partial sequence within a peptide chain, e.g., -Ala-Lys-Glu-Tyr-Leu- is written, additional hyphens are added at both termini. If partial sequences of a peptide have not yet been elucidated or if the presence of one amino acid in a special position has not been identified unambiguously, the tentative amino acid present in this position is given in brackets, separated by commas, as shown for **13**.

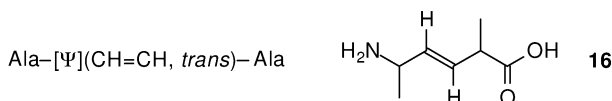
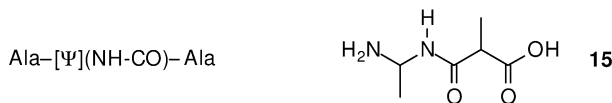


Covalent bonds between side-chain functional groups are also possible for the amino acid cysteine. A disulfide bond between two thiol groups of cysteine is formed upon oxidation to give a cystine residue. Different types of disulfide bridges can be distinguished (cf. Section 6.2). Intramolecular (intrachain) disulfide bridges are formed between two cysteine residues in one peptide chain, while intermolecular (interchain) disulfide bonds are formed between two different peptide chains, as shown for **14**.



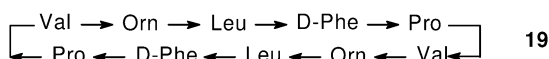
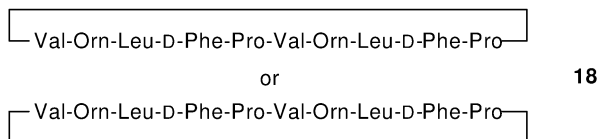
As nonproteinogenic building blocks such as hydroxycarboxylic acids, D-amino acids, and N-alkyl-amino acids may occur in peptides, an extension of the original definition for a peptide is required. Primarily, homomeric peptides and heteromeric peptides must be distinguished. While the former are composed exclusively of proteinogenic amino acids, the latter also contain nonproteinogenic building blocks.

Analogues of peptides (e.g. **15** and **16**), in which the peptide bond that connects two consecutive amino acid residues is replaced by another moiety, are named by placing the Greek letter psi (Ψ), followed by the replacing group in parentheses, between the residue symbols where the change occurs.



Further differentiation is made, according to the nature of the chemical bond, into homodetic peptides that contain exclusively peptide bonds (N^α -amide bonds) and heterodetic peptides that may also contain ester, disulfide, or thioester bonds. The sequence of a cyclic homodetic homomeric peptide can be written in two different variants: following the prefix “cyclo-”, the sequence is annotated in the three-letter code, set in brackets, and backbone cyclization is symbolized by a hyphen “-” before the first and after the last sequence position given. Gramicidin S can be written as shown in 17.

c(-Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-) 17



Alternatively, the sequence of the same peptide may be given as in 18, where cyclization is symbolized by a line above or below the sequence. In the case of a peptide cyclization that occurs across the backbone, this should be symbolized by horizontal lines starting from the N- and C-termini. A third possibility to display a cyclic peptide is shown in 19, where the direction of a peptide chain ($N \rightarrow C$) has to be symbolized by an arrow ($\rightarrow$) that points towards the amino acid located in the direction of the C-terminus. Cyclic heterodetic homomeric peptides are treated similarly in the three-letter annotation.

Since analogues of biologically active peptides are very often synthesized in order to study the structure–activity relationship, a brief introduction to the rules for the nomenclature of synthetic analogues of natural peptides should be given here in accordance with the suggestions by the IUPAC-IUB Joint Commission on Biochemical Nomenclature [7]. The most important rules are detailed in Table 2.1 for the example of a hypothetical peptide with the trivial name iupaciubin and the sequence Ala-Lys-Glu-Tyr-Leu.

2.3

Analysis of the Covalent Structure of Peptides and Proteins

According to Linderstrom-Lang, the structural description of proteins can be considered at four levels of organization (Figure 2.5). This description can be applied to polypeptides and is, to a limited extent, also valid for smaller peptides.

Primary structure, which is the subject of this chapter, comprises the number and sequence of amino acids connected consecutively by peptide bonds within the peptide chain. Secondary structure describes the local three-dimensional arrangement of the peptide backbone. Tertiary structure describes the three-dimensional

Table 2.1 Important nomenclature rules for peptide analogues.

Name/Three letter code	Description
Ala ¹ -Lys ² -Glu ³ -Phe ⁴ -Leu ⁵ [4-phenylalanine]iupaciubin [Phe ⁴]iupaciubin	The exchange of amino acid residues in a peptide is symbolized by the trivial name of the corresponding peptide preceded by the full name of the amino acid replacement and its position given in square brackets. Alternatively, the three letter code may be used instead of the full name of the amino acid. Multiple exchange is treated analogously.
(A) Arg-Ala ¹ -Lys-Glu-Tyr-Leu ⁵ Arginyl-iupaciubin (B) Ala ¹ -Lys-Glu-Tyr-Leu ⁵ -Met Iupaciubyl-methionin	The extension of a peptide may occur N-terminally (A) as well as C-terminally (B). The modified name is generated according to the previously discussed rules.
Ala-Lys ² -Thr ^{2a} -Glu ³ -Tyr-Leu endo-2a-threonine-iupaciubin endo-Thr ^{2a} -iupaciubin	An insertion of an additional amino acid residue is indicated by the prefix “endo” in combination with the number of the sequence position.
Ala-Lys ² -Tyr ⁴ -Leu des-3-glutamic acid-iupaciubin des-Glu ³ -iupaciubin	The omission of an amino acid residue is symbolized by the prefix “des” and the position.
(A) Val-  Ala-Lys-Glu-Tyr-Leu N ^{ε2} -Valyl-iupaciubin N ^{ε2} -Val-iupaciubin (B)  -Val Ala-Lys-Glu-Tyr-Leu N-(Iupaciubin-C ^{δ3} -yl)-valine Iupaciubin-C ^{δ3} -yl-Val	Substitutions on side-chain amino groups (A) or side-chain carboxy groups (B) are symbolized considering the general nomenclature rules.
Lys ² -Glu ³ -Tyr ⁴ Iupaciubin-(2-4)-peptide	The nomenclature for partial sequences that are derived from peptides with a trivial name uses the trivial name, followed by the numbers of sequence positions of the first and last amino acid within the partial sequence in brackets.

structure or overall shape of a single peptide chain resulting from the intramolecular interactions between amino acid side chains and secondary structure elements. The term quaternary structure (not shown in Figure 2.5) refers to the spatial arrangement of two or more polypeptide chains associated by noncovalent interactions, or in special cases linked by disulfide bonds, forming definite oligomer complexes. The term “domain” is applied to describe globular clusters within a protein molecule with more than 200 amino acid residues.

A pure, homogeneous compound is the precondition for structural or biochemical studies involving peptides. Pharmacologically active peptides for therapeutic application have to meet even more strict requirements. Before embarking on a discussion on structural analysis, it is useful to describe a few general methods that are specific for the separation and purification of peptides and proteins.

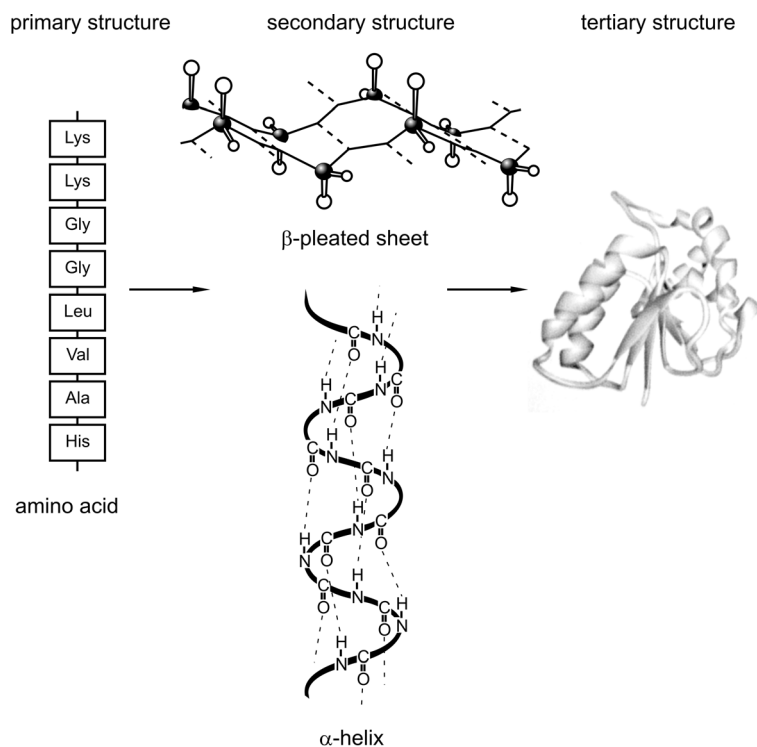


Figure 2.5 Important structural levels of proteins: Primary, secondary, and tertiary structure.

2.3.1

Separation and Purification

2.3.1.1 Separation Principles

Analysis and purification of naturally occurring peptides and of synthetic peptides relies on a series of separation techniques. In general, separation procedures are either applied at the preparative level, in order to isolate one or more individual components from a mixture for further investigations, or at the analytical level, with the goal of identifying and determining the relative amounts of some or all of the mixture components. In carrying out any preparative separation, studies at the analytical level are the initial steps to optimize separation conditions. Preparative separation methods for peptide mixtures should provide samples of high purity, whereas analytical methods include not only the evaluation of the final product of peptide synthesis, but also the monitoring of intermediates with respect to chemical and stereochemical purity. Nearly all practical procedures in the peptide or protein field are based on separation of solute components. Partition chromatography, which belongs to the liquid–liquid chromatography techniques, is the most often exploited technique. In this case, the components of a mixture are distributed between a liquid stationary and a liquid mobile phase because of different interactions. Due to the

Table 2.2 Selected methods for separation and purification of peptides and proteins.

Method	Remarks
Reversed phase HPLC (RP-HPLC)	Most popular HPLC variant used for separation of peptides and proteins; suitable for the assessment of the level of heterogeneity.
Ion-exchange chromatography (IEC)	Most commonly practised method for protein purification.
Size-exclusion chromatography (SEC)	Separation is solely based on molecular size; larger molecules elute more rapidly than smaller ones.
Affinity chromatography (AC)	A bioselective ligand chemically bound to an inert matrix retains the target component with selective affinity to the ligand.
Capillary electrophoresis (CE)	Separation of peptides and proteins is based on their differential migration in an electric field.
Multidimensional separations	2D gel electrophoresis and multidimensional chromatography approaches are capable to more accurately quantify the analyte, and are better compatible with mass spectrometry.
Ultrafiltration (UF)	Method for rapid concentration of protein solutions; lack of selectivity has severely restricted the use of UF for protein fractionation.
Two phase systems for protein separation and purification [11, 12]	Hydrophobic partitioning of proteins in aqueous two phase systems containing poly(ethylene glycol) and hydrophobically modified dextrans.

existence of charged species in aqueous solution, ion-exchange and electrophoresis are also common separation procedures, together with separation on the basis of molecular size. Additionally, adsorption chromatography, especially salt-promoted adsorption chromatography, has not lost its importance.

A comprehensive review by Larive *et al.* [8] focuses on selected applications of the separation and analysis of peptides and proteins published during 1997–1998. The review highlights the state of the art in this field. Selected separation and purification methods [9, 10] are compiled in Table 2.2, and methods of structural analysis will be described in Section 2.5.

Thin-layer chromatography, the simplest technique for peptide analysis, is performed in various solvent systems and uses different detection systems, often followed by electrophoresis. In particular, free peptides can be examined by paper electrophoresis or by thin-layer electrophoresis, either in an acidic solvent (dilute acetic acid) or under basic conditions. In the case of more than trace amounts of impurities, the product must be further purified prior to final evaluation. Chromatographic methods [13] and countercurrent distribution [14, 15] are the most commonly used methods for the purification of free peptides and of blocked intermediates. The classical crystallization procedure remains the simplest and most effective approach, but suffers from the low tendency of peptides to crystallize.

Normal-phase liquid chromatography [16] is used preferentially for the separation of hydrophilic peptides as they are very often not retained sufficiently on standard

reversed-phase-high performance liquid chromatography (RP-HPLC) packing material. The introduction of HPLC, both on an analytical and a preparative scale, opened a new area in separation and analysis of peptides and proteins and continues to be the method of choice [17–19]. Insoluble, hydrophilic support materials are used as the stationary phase in normal-phase HPLC. The major difference between “high-performance” or “high-pressure” chromatography and “low-pressure” bench-top chromatography is that HPLC employs columns and pumps that allow the application of very high pressures, with the advantage that particles with 3–10 μm mean diameter can be used as packing material for the columns. These conditions allow superior resolution within relatively short times for HPLC, compared to hours or even days for low-pressure chromatography systems. In RP-HPLC, the solid stationary phase is derivatized with nonpolar hydrophobic groups so that the elution conditions are the reverse of normal-phase liquid chromatography. Alkylsilanes with between 4 and 18 carbon atoms (referred to as C-4 to C-18 columns) are used for derivatization of the stationary phases. Retention of the solute occurs via hydrophobic interactions with the column support, and elution is accomplished by decreasing the ionic nature or increasing the hydrophobicity of the eluent. Commercially available reversed-phase columns allow rapid separation and detection of the components present in a mixture.

Ion-exchange chromatography (IEC) is easy to use for protein purification due to its high scale-up potential [20]. The separation principle of IEC is based on interaction of the protein’s net charge with the charged groups on the surface of the packing materials. Polystyrene and cellulose, as well as acrylamide and dextran polymers, serve as the preferred support materials for the ion exchanger. They are functionalized by quaternary amines, diethylaminoethyl (DEAE) or polyethylenimine for anion exchange, and sulfonate or carboxylate groups for cation exchange.

Hydrophobic interaction chromatography (HIC) is, besides thiophilic adsorption chromatography and electron donor–acceptor chromatography, the dominant method of salt-promoted adsorption chromatography [21].

Size-exclusion chromatography (SEC) [22, 23] continues to be an efficient separation method for proteins, though in peptide purification its resolving power is somewhat limited. Although low-molecular weight impurities can be separated without problems from a crude mixture of peptides, the separation of a target peptide from a closely related peptide mixture is practically not possible. In aqueous separation systems, SEC is also named gel filtration chromatography (GFC), whereas the alternative term gel permeation chromatography (GPC) is related to the application in nonaqueous separation systems.

Affinity chromatography was pioneered by Cuatrecasas [24]. In this technique [25], a low-molecular weight biospecific ligand is linked via a spacer to an inert, porous matrix, such as agarose gel, glass beads, polyacrylamide, or cross-linked dextrans. Monospecific ligands (analyte given in brackets) include hormones (receptors), antibodies (antigens), enzyme inhibitors (enzymes), and proteins (recombinant fusion proteins). Group-specific ligands (binding partner given in brackets) include, for example, lectins (glycoproteins), protein A and protein G (immunoglobulins G), calmodulin (Ca^{2+} -binding proteins), and dyes (enzymes). In general, affinity chro-

matography is useful if a high degree of specificity is required, for example in the isolation of a target protein present in a low concentration in a biological fluid or a cellular extract. Affinity methods have become popular for biological recognition using peptide combinatorial libraries (see Chapter 8) to optimize affinity-based separations [26]. Antibodies are widely used to isolate and purify the corresponding antigen. In contrast, it is also possible to isolate specific antibodies using an immobilized antigen on a column. Affinity chromatography often makes use of specific tags that have become indispensable tools. Besides facilitating the detection and purification of recombinant proteins, tags can improve yield, solubility and even folding of the fusion proteins [27].

Capillary electrophoresis (CE) or capillary zone electrophoresis (CZE), pioneered by Jorgenson and Lukacs [28], has been widely developed for the separation of peptides and proteins, including recombinant proteins [29, 30]. CE separates peptides and proteins based on their differential migration in solution in an electric field [31]. In accordance with ion-exchange HPLC, the separation is a function of the charge properties of the compound to be separated. However, the physical basis of separation is different in that it is typically performed in a capillary column of 50 μm i.d. and 30–100 cm length with a volume of 0.6–2 μL , and the injection volume is limited to <20 nL. A common set-up for peptide separation uses a buffer at pH 2.0. All charged peptides are cationic and migrate to the cathode. Under these conditions the effect of electroendosmosis is minimized. The mentioned limited sample injection volume in CE limits the detection of minor components in complex mixtures, and suitable preconcentration methods are often needed. Capillary isoelectric focusing (CIEF) provides high-efficiency separations and improved detection sensitivity for complex peptide mixtures [32].

The most widely used technique for protein separation on an analytical scale is two-dimensional polyacrylamide gel electrophoresis (2-DE). It aims at quantification and identification of proteins from complex mixtures and is nowadays widely employed in combination with mass spectrometry (MS) for proteome analysis. 2-DE usually combines protein separation according to pI by isoelectric focussing (IEF) with separation according to molecular mass by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). It allows the simultaneous resolution of more than 5000 proteins, depending on gel size and pH gradient [33].

Ultrafiltration (UF) is a pressure-driven separation process on the basis of molecular size using a membrane. Suspended solids and solutes of high molecular mass are retained. UF is used for the concentration of protein solutions and for the separation of molecules with significant mass differences. The nature of the membrane determines whether compounds permeate or are retained. UF membranes are classified according to the molecular mass limit ("molecular weight cut-off", MWCO), up to which molecules are allowed to permeate. Instead of applying pressure, the necessary force for ultrafiltration may be exerted by centrifugation (centrifugal ultrafiltration) [34].

Countercurrent chromatography refers to a form of support-free liquid–liquid partition chromatography [14, 15].

2.3.1.2 Purification Techniques

The application of a peptide or protein for either scientific investigations or commercial purposes usually requires isolation in a homogeneous state. Purification is defined as a process where the target molecule is separated from other compounds [35]. Separation and purification are, formally, different categories, but in practice they overlap to a great extent. Therefore, methods described in the preceding section are included in purification schemes. The purity is an essential prerequisite both for structural analysis and for reproducible results in biological investigations. Homogeneity and correct covalent structure must be the main goals as peptides and proteins have become an important class of potent therapeutic drugs (see Chapter 9).

Purification procedures can take considerable amounts of time and effort. Unfortunately, evidence is provided in the literature that insufficient attention is often paid to the correct evaluation of the final product. Insufficiently pure compounds may lead to totally erroneous conclusions. Residues such as Trp, Met, N-terminal Gln, and Cys are known to cause considerable problems from unwanted reactions during synthesis, work-up, handling, or storage. Cys, Met, and Trp are susceptible to oxidation. Whereas the oxidation of Cys and Met usually is reversible, this is not generally true for Trp. Gln in the N-terminal position of a peptide sequence is capable of forming a lactam structure, also known as pyroglutamyl residue. In several publications, peptides have been claimed to be either crude or pure, but often different categories of purity are used, such as homogeneous, very pure, and chromatographically pure. Obviously, different evaluation criteria are applied to each of these categories. In practice, the required degree of purity depends on the application. Employment as an antigen in the production of polyclonal antisera requires lower peptide purity, compared to a product used for conformational studies or special investigations on biological activity.

An accurate and specific assay for the component of interest is an essential prerequisite for peptide and protein purification. The quantitative result obtained from an assay is defined as the “activity”. This term is normally used for enzyme assays, but it can also be applied to the results of other assays. Enzyme assays are generally based on a specific reaction catalyzed by the enzyme, whilst assays for other proteins rely on their physical properties. For example, the purification of proteins bearing highly chromogenic cofactors can be easily monitored using spectrophotometric methods. A highly specific assay uses antibodies against the peptide or protein. The specific activity is a quantitative measure for purity, and is defined as the ratio of the amount of activity in a protein solution to the amount of total protein in the solution. The specific activity increases during purification by removing contaminating proteins that lack assay activity. A protein is enriched during purification, and will be completely pure after attaining the highest specific activity for that protein.

Most protein purification protocols involve multiple steps employing various techniques [36]. To obtain a cell-free solution of an intracellular protein, the cell must first be broken in order to release its contents. Insoluble particles, including

membranes, can then be removed by centrifugation or other bulk methods. Soluble proteins are separated and purified from the resulting crude solution. The release of membrane-bound proteins from the membrane can sometimes be performed with the use of detergents.

The purification of bacterial proteins starts with the physical separation of the protein from other bacterial nonprotein cell constituents. Generally, proteins secreted into the extracellular medium can be separated and purified more easily. In the case of a protein solution, most purification protocols start with a precipitation step by means of alcohols, heat, or salt. Ammonium sulfate precipitations are often used, since this salt causes precipitation of many proteins without affecting their biological activity.

Most purification methods subsequently involve various forms of chromatography. The most selective variants employ fusion proteins with different tags that can be purified on specific matrices by affinity chromatography [27]. Immobilized metal affinity chromatography (IMAC), pioneered by Porath *et al.* [37], continues to be very popular for the large-scale purification of proteins. Iminodiacetic acid or nitrilotriacetic acid (NTA), to which Ni^{2+} is typically chelated, are the most widely used support materials. IMAC is especially appropriate for recombinant proteins where the codons of (usually) six histidine residues (His-tag) are appended to the cDNA [38, 39]. Small-scale purification of proteins can easily be performed by the attachment of NTA to magnetic beads [40].

Tag-free recombinant proteins may, for example, be obtained by intein-mediated protein purification. Fusion proteins of the target sequence and a chitin-binding domain are produced recombinantly, purified on a chitin column and the formation of the tag-free target sequence is then brought about by intein-mediated protein splicing (see also Chapter 5) [41].

Numerous applications of preparative chromatography are cited in reviews for the purification of both recombinant proteins [42] and unstable proteins [43]. Countercurrent chromatography has also been used for protein purification [14, 15]. Further information on protein purification has been summarized in several excellent monographs [8–10, 35, 36]. A variety of methods has found application in the purification of potential peptide drugs, such as crystallization, countercurrent distribution, partition chromatography, GPC, low-pressure HIC, IEC and RP-HPLC.

As mentioned above, crystallization should be one of the most powerful purification methods, but its use is mainly limited to small oligopeptides up to pentapeptides. Countercurrent distribution and partition chromatography have now been largely substituted by more powerful purification procedures, such as preparative RP-HPLC. At present, most peptides with 50 residues or less can be purified using this technique. In practice, purification schemes which consist of two complementary techniques are very often used. For example, a crude synthetic peptide should first be separated from by-products, resulting from final deprotection steps, which are mostly uncharged, low-molecular weight compounds. For this purpose, IEC, GPC, or HIC are the methods of choice. For a necessary final purification step, RP-HPLC should be applied as a complementary technique. Such a two-step purification scheme seems to be superior to the application of various RP-HPLC steps using different mobile phases.

2.3.1.3 Stability Problems

Peptides and proteins differ from most other chemical compounds. Their chemical [44], physical [45], and enzymatic [46] instability presents a challenge, for example in the development of more stable peptide and protein drugs. In most cases the pure compounds are obtained as amorphous solids, preferentially by lyophilization (freeze-drying). Generally, peptides and proteins tend to retain variable amounts of water and acids (as counterions or residual acid) resulting from the final stage of purification or isolation. In addition, peptides are preferably stored as their corresponding acetates or hydrochlorides. Special stability programs for bulk peptide drugs are performed by testing their storage behavior at various temperatures for different periods of time [47]. Decomposition or deterioration may occur as a result of oxidation, absorption or release of moisture, exposure to heat or light, or interactions with surfaces.

Solid-state formulations are generally more stable than the corresponding aqueous formulations. In the latter case, the nature of the solvent, concentration, pH, and temperature have a great influence on stability. Adsorption onto the container, inactivation, racemization, oxidation, deamidation, chain cleavage, diketopiperazine formation, and rearrangements are the most well-known effects causing instability of peptides in solution. Surprisingly, several cases are known where protein stability in the solid state is inferior or comparable to that observed in solution [48, 49]. In the solid state, chemical instability is caused similarly to that in solution (bond cleavage, bond formation, rearrangement or substitution). Typical reactions are deamidation of Asn and Gln [50], oxidation at sulfur atoms of Cys and Met, disulfide exchange at Cys, peptide bond cleavage, β -elimination, and dimerization/aggregation [51]. Peptides and proteins in food, as well as in pharmaceutical formulations, may undergo reactions with nonprotein compounds. Loss of Lys in food proteins is primarily attributed to the Maillard reaction (Figure 2.6), but other amino acid residues such as Arg, Asn, are also capable of reacting with reducing sugars.

In the first step, a condensation reaction between the carbonyl group of a reducing sugar and an amino group forms an N-substituted glycosylamine, which is followed by conversion to a Schiff base. This may also be a problem in drug formulations. Subsequent cyclization and Amadori rearrangement yields the typical derivatives which cause discoloration of the formulation [52]. The reasons why these and other chemical degradation reactions can occur in “dry” formulations are still unclear. Moisture, temperature, and formulation excipients (e.g., polymers) are the main factors influencing peptide and protein chemical instability in the solid state. It has

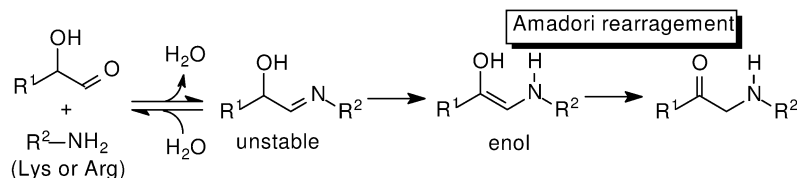


Figure 2.6 The Amadori rearrangement as the initial step of the Maillard reaction.

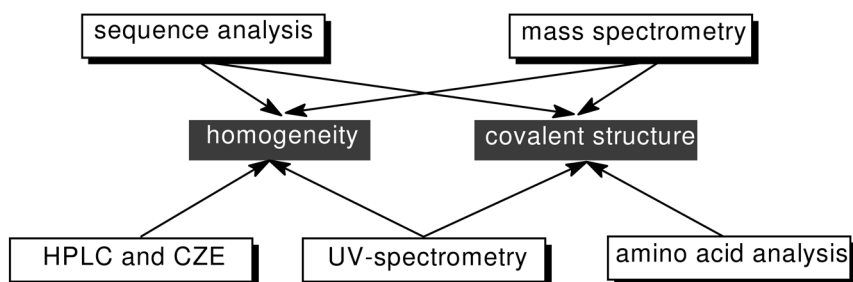


Figure 2.7 Analytical techniques used for the determination of homogeneity and covalent structure of peptides.

been suggested that proton activity in the solid state may have a similar influence on stability as does pH in the solution state. Details on the solid-state stability of proteins and peptides have been reviewed by Lai and Topp [53].

2.3.1.4 Evaluation of Homogeneity

After final purification, proof of homogeneity and structural characterization (see Section 2.5) are necessary to ascertain that the desired product, and not structural modifications of it, has been isolated. Numerous analytical methods are available to evaluate homogeneity and the covalent structure.

Figure 2.7 shows schematically the relative contribution of analytical methods to the goals of assessing homogeneity and structural integrity. Amino acid analysis (see Section 2.3.2.3) is necessary for the determination of the covalent structure, whereas sequence analysis (see Section 2.3.2.5) contributes to the evaluation of both homogeneity and covalent structure. Methods of structural analysis will be discussed in Section 2.5. Mass spectrometry with the dominant ionization methods matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), and the coupling of ESI to separation techniques, is an important tool in the analysis of peptides and proteins [54]. Nuclear magnetic resonance (NMR) spectroscopy (Section 2.5.3) is one of the premier analytical tools for structure elucidation. In particular, the use of multidimensional NMR methods for the structure determination of peptides and small proteins in solution has become routine. Ultraviolet (UV) spectroscopy is used to monitor the integrity of aromatic amino acids (particularly Trp), and is a routine tool for monitoring peptide purification. However, the information on the covalent structure of a peptide is very limited.

2.3.2

Primary Structure Determination

Amino acid sequence analysis is an essential component of protein structure determination [55–57]. The elucidation of the primary structure of insulin by Frederick Sanger in 1955 was a breakthrough of enormous biochemical significance and earned him the first of two Nobel prizes. Since that period, the amino acid

sequences of several thousand proteins have been elucidated. Refinement and automation of the procedures nowadays allow for sequence analysis of proteins below the 10 pmol level, whereas the experimental work involved in determining the insulin structure by Sanger and his coworkers took more than a decade and required about 100 g insulin.

In 1980, Sanger earned his second Nobel prize for the development of the technology for rapid DNA sequencing. Nucleic acid sequencing initially lagged far behind the corresponding techniques for peptides and proteins. Sanger's chain termination method (or dideoxy method) [58] and the chemical cleavage method according to Maxam and Gilbert [59] proved to be efficient methods for DNA sequencing. Nowadays, DNA sequencing is easier and faster than protein sequencing; furthermore, this new technique allows even the sequence determination of proteins that have never been isolated. Despite this progress, direct peptide and protein sequencing remains an indispensable tool for several reasons. Without doubt, the determination of amino acid sequences by degradation methods is a necessity to confirm conclusively the structure of isolated and synthetic peptides. Furthermore, determination of the primary structure from the DNA sequence does not provide information on post-translational modifications of proteins and the location of disulfide bonds. It must also be mentioned that a common error in DNA sequencing is the inadvertent insertion or deletion of a single nucleotide which requires the need to identify the open reading frame in the DNA structure. In addition, it is often not easy to identify and isolate the nucleic acid that encodes the protein of interest. Determining the amino acid sequence of a portion of the protein allows the chemical synthesis of the corresponding DNA fragment that can be used to identify and isolate the whole gene.

2.3.2.1 End Group Analysis

Before starting sequence analysis it is necessary to determine the number of different peptide chains in a protein by the analysis of N-terminal and/or C-terminal residues. Of course, this problem may be complicated if the amino group at the N-terminus and the carboxy group at the C-terminus are chemically blocked, or the peptide to be analyzed is cyclic.

N-terminal analysis can be performed using several chemical and enzymatic methods. Chemical methods are mostly based on the transformation or blocking of the N-terminal α -amino function, followed by hydrolysis, separation, and characterization of the terminal amino acid derivative (Figure 2.8).

The reagent 2,4-dinitrofluorobenzene was first used by Sanger [60], and this dinitrophenyl (DNP) method (Figure 2.8A) has found wide application. After labeling the N-terminus of the peptide chains, complete hydrolysis liberates the DNP amino acids that can be identified by chromatography.

The dansyl (1-dimethylaminonaphthalene-5-sulfonyl) method [61] (Figure 2.8B) is more sensitive; dansyl chloride provides dansyl amino acids which are strongly fluorescent and can be identified chromatographically at the 100 pmol level. In addition, a number of less important chemical methods for blocking N-terminal amino acids, such as arylsulfonation, carbamoylation or carboxymethylation, and

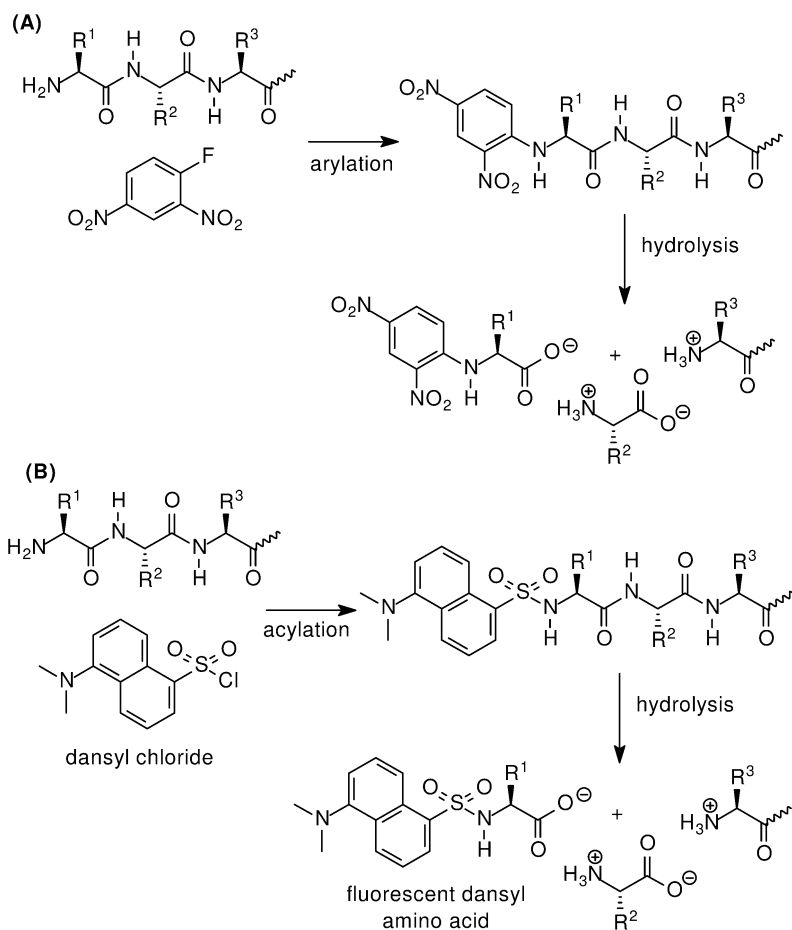


Figure 2.8 Identification of the N-terminus using (A) 2,4-dinitrophenylfluorobenzene and (B) dansyl chloride.

aminopeptidases as an enzymatic alternative can be used for N-terminal analysis. Enzymatic release of the N-terminal amino acid requires the presence of an unblocked α -amino function. Although porcine kidney leucyl aminopeptidase (LAP) hydrolyzes N-terminal L-amino acids (including glycine), peptide chains that bear L-leucyl residues are the preferred substrates [62]. Unfortunately, this enzyme does not release N-terminal arginine and lysine or any amino acid that is followed by proline. Furthermore, three aminopeptidases from *Bacillus stearothermophilus* [63] are suitable for the cleavage of N-terminal amino acid residues. Aminopeptidase I (API) shows a broad specificity; it releases not only neutral (preferentially aliphatic and aromatic), but also acidic and basic amino acids, including proline, from the N-terminus. Aminopeptidase N, which in 1980 was renamed to aminopeptidase M, shows a preference for neutral amino acids [64].

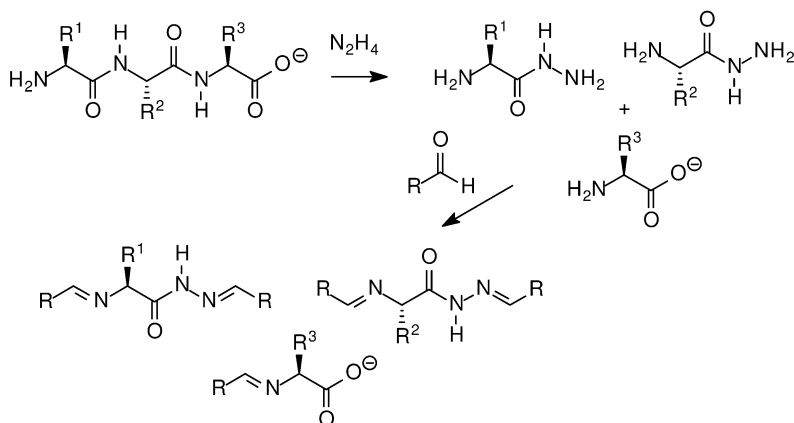


Figure 2.9 C-terminal group analysis according to Akabori.

Without doubt, Edman degradation (see Section 2.3.2.5) is one of the most powerful procedures for N-terminal residue identification.

C-terminal analysis can be performed both chemically and enzymatically, but has not gained as much importance as N-terminal procedures. Using the hydrazine method [65] according to Akabori, all amino acids (apart from the C-terminal ones) are converted into hydrazides on treatment with anhydrous hydrazine for 20–100 h at 90 °C in the presence of an acidic ion-exchange resin. Only the C-terminal amino acid residue is released as the free amino acid, and this can be identified chromatographically (Figure 2.9). C-terminal Cys, Gln, Asn, and Trp cannot be identified by this method, and Arg will be partly converted to Orn.

Another procedure is based on the treatment of the peptide with ammonium thiocyanate and acetic anhydride to give 1-acyl-2-thiohydantoin by cyclization (cf. Section 2.3.2.6). The 2-thiohydantoin derived from the C-terminal amino acid is formed after mild alkaline hydrolysis in addition to the N-acylpeptide containing one residue less.

The carboxypeptidase (CP) approach is more efficient (Figure 2.10), with CP A and CP B from bovine pancreas showing different, but complementary, specificities. CP A preferentially liberates C-terminal residues with aromatic side chains, but not Pro, Arg, Lys, and His, whereas CP B releases only C-terminal Arg, Lys, and His. The serine peptidase CP Y is much less specific and removes all residues, albeit Gly and Pro only very slowly.

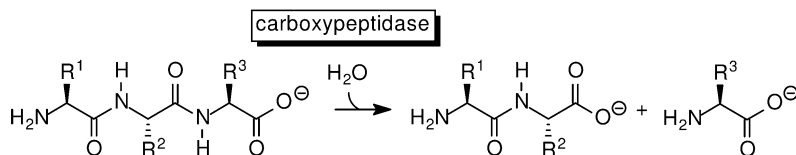


Figure 2.10 C-terminal group analysis by carboxypeptidase.

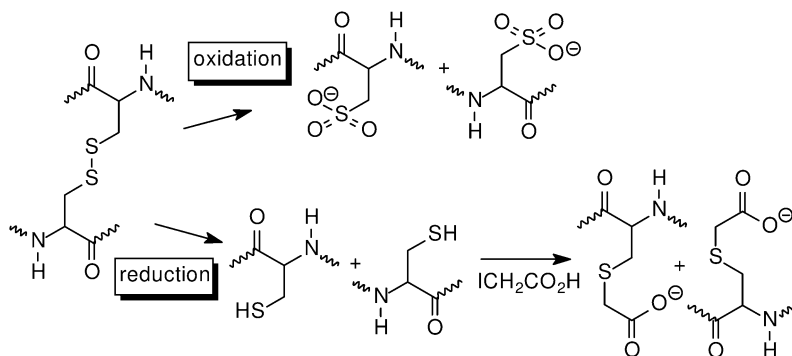


Figure 2.11 Oxidative and reductive cleavage of disulfide bonds.

2.3.2.2 Cleavage of Disulfide Bonds

Cleavage of disulfide bonds is necessary to separate disulfide-linked peptide chains and also to destroy the native peptide or protein conformation when it is stabilized by those bonds. The cleavage procedure involves hydrolysis of the protein under conditions that minimize the risk of disulfide bond exchange. Intra- or inter-chain disulfide bonds can be cleaved by reduction or oxidation (Figure 2.11).

Oxidation using performic acid, also pioneered by Sanger, converts all Cys residues to cysteic acid residues. Since cysteic acid is stable under both acidic and basic conditions, it is possible to determine the total Cys content as cysteic acid. Unwanted oxidation of Met residues to methionine sulfoxide/sulfone, as well as partial degradation of the Trp side chain, are serious disadvantages. The reductive cleavage of disulfide bonds is mostly achieved by treatment with a large excess of a suitable thiol, e.g., 2-mercaptoethanol 1,4-dithiothreitol/1,4-dithioerythritol (Cleland's reagent). The resulting free thiol groups should be blocked by alkylation with iodoacetic acid in order to prevent reoxidation in air. The location of the disulfide bonds is carried out in the final step of sequence analysis.

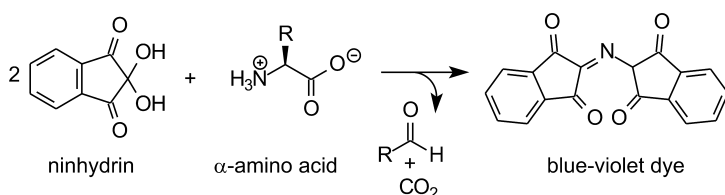
2.3.2.3 Analysis of Amino Acid Composition

Before starting the sequence analysis of a peptide chain it is necessary to know its amino acid composition. Amino acid analysis [66] can be performed by complete hydrolysis, followed by quantitative analysis of the liberated amino acids. For this purpose numerous chemical and enzymatic protocols are known, but none of these procedures alone is fully satisfactory. Besides hydrolysis with 6 M hydrochloric acid at 120 °C for 12 h, or with dilute alkali (2–4 M NaOH) at 100 °C for 4–8 h, mixtures of peptidases can also be used for complete peptide hydrolysis. Under acidic hydrolysis conditions, the polypeptide is dissolved in 6 M HCl and sealed in an evacuated tube to minimize the destruction of particular amino acids. Tryptophan and cysteine/cystine are especially sensitive towards oxygen. For the complete release of aliphatic amino acids, long hydrolysis times up to 100 h are sometimes required. Unfortunately, under such harsh conditions hydroxy-substituted amino acids (Ser, Thr, Tyr) are partially degraded, and Trp will be largely destroyed. Furthermore, Gln and Asn are

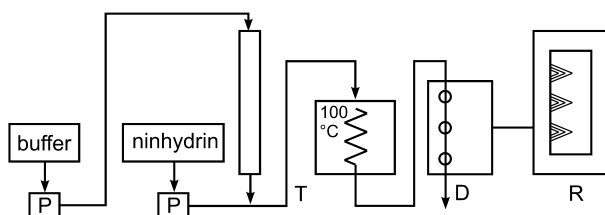
converted into Glu and Asp plus NH_4^+ , with the consequence that only the amounts of Asx (= Asn + Asp), Glx (= Glu + Gln), and NH_4^+ (= Asn + Gln) can be independently determined. Although Trp largely survives alkaline hydrolysis and, therefore, allows the content of this amino acid to be determined, this procedure causes decomposition, particularly of Ser and Trp, and Arg and Cys might also be damaged. Pronase consisting of a mixture of relatively unspecific peptidases from *Streptomyces griseus* is often used for enzymatic hydrolysis. However, the amount of peptidase used should not be more than 1% by weight of the polypeptide to be hydrolyzed, otherwise, self-degrading by-products might contaminate the final digest. The enzymatic procedure is mostly used for the determination of Trp, Asn and Gln because of the reasons mentioned above.

Based on the pioneering work of Stein and Moore, amino acid analysis has been automated. Such automated equipment separates amino acids by IEC, and the Stein and Moore protocol employs post-column derivatization with ninhydrin to give the blue color of Ruhemann's violet (Figure 2.12).

(A)



(B)



(C)

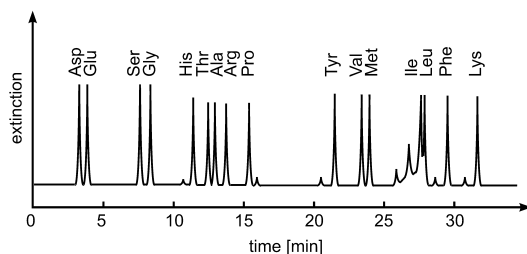


Figure 2.12 Principle of automatic amino acid analysis.
(A) Ninhydrin reaction (simplified). (B) Analyzer scheme.
P = pump, T = Teflon tubing, D = detector, R = recorder.
(C) Cation exchange HPLC amino acid analysis.

The intensity of the color can be measured spectrophotometrically, thus providing quantitative detection of each of the separated components. Nowadays, instruments which are based on partition chromatography, such as HPLC and gas-liquid chromatography (GLC) are in use for quantitative amino acid analysis. The derivatization of amino acids is necessary to introduce fluorophoric groups which conveniently permit quantitation of UV/visible absorption or fluorescence characteristics for HPLC, or convert amino acids into volatile derivatives for GLC. Various derivatization protocols for HPLC utilize fluorescent and UV-absorbing tags, e.g., dansyl, Fmoc, fluoresceamine, *iso*-indolyl [*o*-phthaldialdehyde (OPA) + 2-mercaptoethanol] condensation products, and Edman's reagent (see Section 2.3.2.5). The complete analysis of an amino acid mixture can be performed using a modern automated amino acid analyzer in a short time (minutes), with a sensitivity as low as 1 pmol of each amino acid.

2.3.2.4 Selective Methods of Cleaving Peptide Bonds

Efficient sequence analysis requires fragments of the protein to be analyzed, since stepwise degradation is limited to 40–80 residues (see Section 2.3.2.5). Therefore, longer polypeptide chains must be cleaved into fragments suitable for stepwise analysis by either enzymatic or chemical methods.

Several highly specific endopeptidases which allow complete and specific fragmentation are used for enzymatic cleavages. Trypsin cleaves at the carboxy side of lysyl and arginyl peptide bonds, thereby providing fragments with C-terminal Lys or Arg residues. In principle, it is possible to restrict cleavage to arginyl bonds by blocking the Lys ϵ -amino group. In principle, selective blocking of the ϵ -amino function can be performed by any acylating agent, as trypsin recognizes its specific substrate by a positively charged side-chain function. Using citraconic anhydride or ethyl trifluoroacetate, the appropriate protecting group can be deblocked by morpholine or very mild acid treatment, and this allows exposure of the Lys side-chain function for a second tryptic hydrolysis. Interestingly, Cys may be aminoalkylated by a haloalkylamine, e.g., 2-bromoethylamine, yielding a positively charged residue that enables tryptic cleavage. In contrast to trypsin, thrombin is more specific and only capable of cleaving a limited number of Arg bonds. Unfortunately, hydrolysis by this enzyme sometimes proceeds slowly and results in incomplete degradation of the substrate. Clostripain from the anaerobic bacterium *Clostridium histolyticum* is well known for its selective hydrolysis of arginyl bonds, but lysyl bonds are cleaved at a lower rate. The V8 protease from *Staphylococcus aureus* is highly specific for the cleavage of glutamyl bonds and is, therefore, widely used in sequence analysis. The less specific chymotrypsin provides an alternative set of peptide fragments bearing aromatic or hydrophobic aliphatic amino acids at the C-terminus. In general, too many small peptides will be formed, and this does not facilitate determination of the primary structure of a protein. The same is also true for other less specific endopeptidases, e.g., papain, subtilisin or pepsin, but these enzymes gained importance for isolating small peptide fragments containing disulfide bonds, phosphoserine or glycosyl moieties in their side chains.

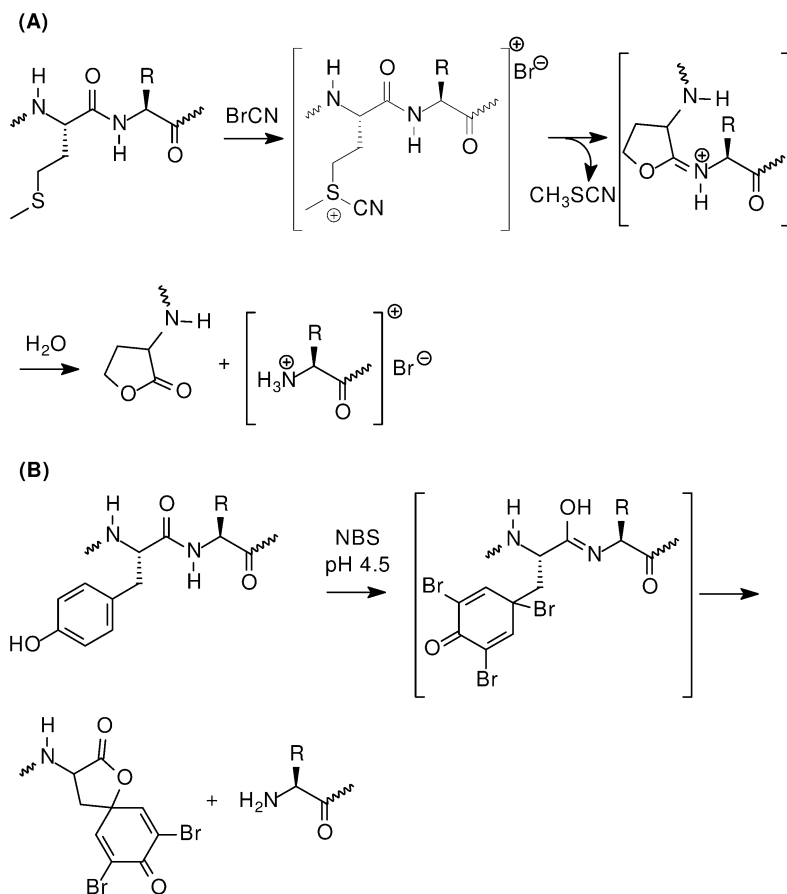


Figure 2.13 Chemical sequence-specific cleavage of peptide bonds with (A) cyanogen bromide (BrCN) and (B) *N*-bromosuccinimide (NBS).

For selective chemical cleavages, cyanogen bromide (BrCN) and *N*-bromosuccinimide are the favorite reagents in common use (Figure 2.13). Cleavage with BrCN [67] under acidic conditions (0.1 M HCl or 70% formic acid) denatures the protein and causes the formation of a peptidyl homoserine lactone from Met residues with release of the aminoacyl peptide. It should be taken into consideration that the C-terminal homoserine lactone as a reactive species may be used to anchor the peptide to an insoluble support for solid-phase sequencing (Section 2.3.2.5).

Trp is, like Met, another amino acid that occurs only rarely in proteins. Cleavage of tryptophyl bonds will, therefore, also form larger fragments. *N*-Bromosuccinimide cleaves not only tryptophyl bonds but also tyrosyl bonds [68]. The use of either 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine, which is generated *in situ*

from 2-(nitrophenylsulfenyl)-3-methyl-indole, *N*-bromosuccinimide [69], or 2-iodosobenzoic acid [70] allows a more selective cleavage of tryptophyl bonds.

2.3.2.5 N-Terminal Sequence Analysis (Edman Degradation)

The most efficient method for stepwise degradation of a peptide chain starting from the N-terminus was developed by Pehr Edman in 1949 [71]. This cyclic process consists of three steps: coupling, degradation, and conversion. Each cycle involves the reaction of the α -amino group at the N-terminus of a peptide chain with phenyl isothiocyanate (PITC) under slightly basic conditions, thereby forming the phenylthiocarbamoyl (PTC) adduct (Figure 2.14).

Extractive removal of excess PITC is followed by treatment with an anhydrous strong acid such as trifluoroacetic acid to cleave the N-terminal amino acid to give its 2-anilino-5(4*H*)-thiazolone derivative, without attacking the rest of the peptide bonds within the chain. The 5(4*H*)-thiazolone derivative is extracted and undergoes rearrangement in hot trifluoroacetic acid to give the more stable 3-phenyl-2-thiohydantoin (PTH). The PTH amino acids can be identified by TLC on silica gel,

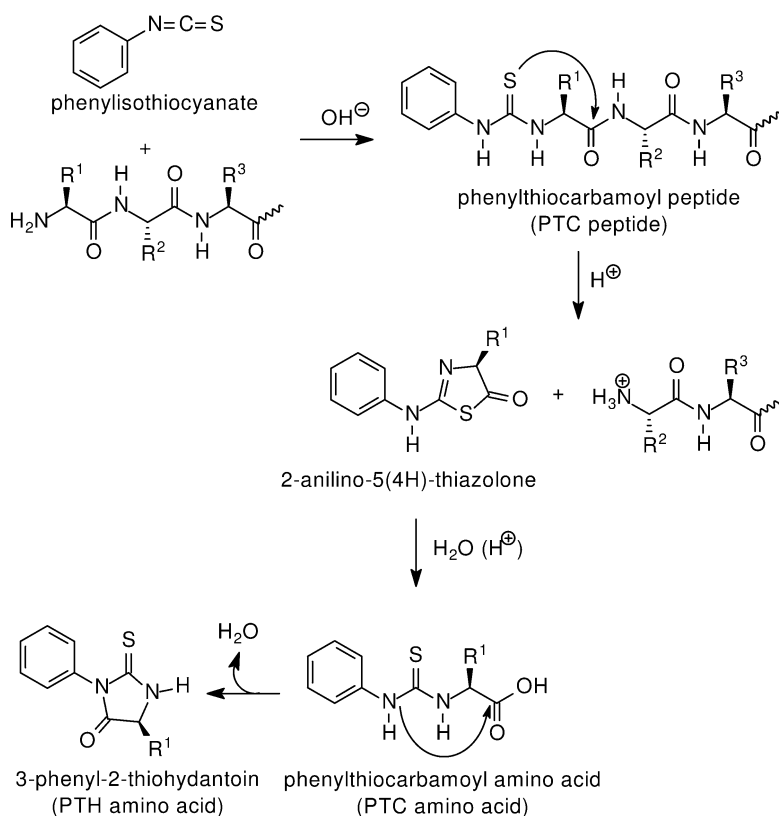


Figure 2.14 Edman method for N-terminal stepwise peptide degradation.

by RP-HPLC on a column with UV detection (2-anilino-5(4*H*)-thiazolone can be identified and quantified similarly), or by GLC, also in combination with MS. In order to obtain good results, precisely standardized conditions are required. The repetitive yield is defined as the total yield of one cycle. Unfortunately, the yield of PTH cannot achieve 100% at every step because of incomplete reactions and losses during the manipulations. Consequently, small amounts of PTHs corresponding to preceding sequence positions will be formed. At a stage when the yield of PTH from the newly exposed N-terminus is very low and the number of PTHs resulting from incompletely degraded chains is very high, interpretation of the results is invalid.

The Edman degradation has been improved by the development of the spinning-cup or liquid-phase sequencer [72]. Using this equipment, the analysis material is spread by centrifugal forces as a thin film on the inner wall of a spinning cup. The reagents and solvents are fed automatically to the bottom of the glass cup through a valve assembly and a special feeding line. The extracting solvents reach a groove in the cup, from where they are removed and leave the cup through an effluent line. All operations such as dissolution, concentration, drying under vacuum and extraction are carried out in the same system. The truncated peptide remains in the reaction vessel. Automated Edman degradation achieves repetitive yields of >90%. "Normal" repetitive yields of approximately 95% allow 30–40 cycles to be performed, with reliable results. The amount of material required for analysis can be significantly reduced by using a polymeric quaternary ammonium salt, called polybrene. This adheres strongly both to the analyte and to glass, and results in a form of immobilization of the peptide material.

Another breakthrough in sequence analysis was the development of the gas phase sequencer by Hewick *et al.* [73]. Using this simple principle, the amount of material used for analysis could be reduced even further. A chemically inert disk made of glass fiber, sometimes coated with polybrene, is used for the application of the analysis sample. Exact quantities of basic and acidic reagents, respectively, are delivered as a vapor in a stream of argon or nitrogen, and then added to the reaction cell at programmed times. Under these conditions the peptide loss can be significantly minimized, and such an instrument is capable of processing up to one residue per hour. The thiazolinone derivative is automatically removed and converted to the PTH derivative. The pulsed-liquid sequencer is a variant of the gas-phase sequencer. The acid is delivered as a liquid for very rapid degradation, which requires an accurately measured quantity sufficient to moisten the protein sample and to prevent it from being washed out. Such a procedure shortens the cycle by up to 30 min. Modern sequencers need, on average, as little as 10 pmol peptide or protein, though this must be free of salts and detergents.

The solid-phase sequencer, described by Laursen *et al.* [74], is based on a type of "reversed Merrifield technique" (see Section 4.5). The peptide is covalently immobilized on a suitable support, and this allows simple separation of excess phenyl isothiocyanate and 2-anilino-5(4*H*)-thiazolone. Initially, aminopolystyrene was applied as the matrix, but neither this material nor polyacrylamide could fulfil the requirements for optimum solid supports. Controlled-pore glass treated with 3-aminopropyltriethoxysilane displays improved properties for covalent attachment

of peptides to the free amino groups of the support by different coupling methods. Despite new attempts to use improved coupling methods for covalent fixation of the peptide to membranes, the solid-phase sequencing principle has not yet achieved the value of the gas-phase sequencer.

2.3.2.6 C-Terminal Sequence Analysis

The identification of C-terminal sequences is a useful complement to the Edman degradation procedure, especially in the investigation of N-terminally blocked peptides and proteins. Furthermore, C-terminal sequence information is of general interest for the control of recombinant protein products, for identification of post-translational truncations, for the confirmation of DNA sequence data, and for assisting the design of oligonucleotide probes for molecular cloning studies.

According to Schlack and Kumpf [75], the treatment of an *N*-acyl peptide with ammonium thiocyanate and acetic anhydride, or alternatively with the new derivatizing reagent triphenylgermyl isothiocyanate (Ph_3GeSCN , TPG-ITC) [76], leads to cyclization at the C-terminus, forming 1-acyl-2-thiohydantoin. Mild alkaline hydrolysis yields the 2-thiohydantoin corresponding to the C-terminal amino acid residue and the *N*-acyl peptide containing one amino acid residue less (Figure 2.15).

Further developed instruments [77] with improved chemistry for the activation and cleavage steps [78, 79] are now in use for C-terminal sequencing. Compared to Edman chemistry, C-terminal degradation techniques are less efficient and give rise to various side reactions. In particular, the yields are significantly lower than those obtained by Edman degradation. Proline normally stops C-terminal degradation and requires special derivatization. Combined application of Edman degradation with a C-terminal sequencer has resulted in improved sensitivity, length of degradation, and Pro passage. After initial Edman degradation, the sample is moved to the C-terminal instrument for continued sequencing [80]. C-Terminal sequence analysis is now also possible even at the 10 pmol level. Efficient combination of N- and C-terminal sequencing using the same analysis sample and degradation of more than 10

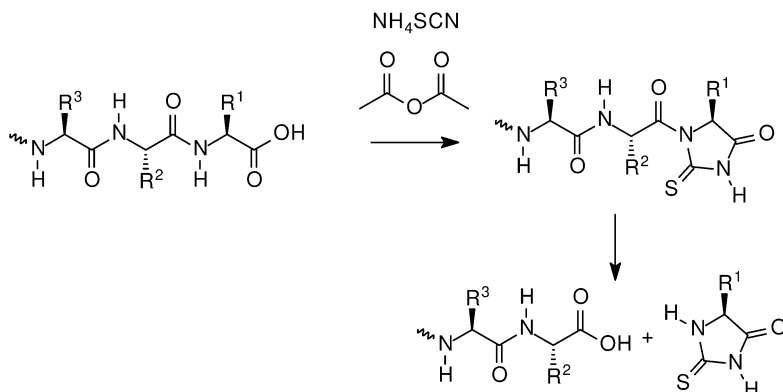


Figure 2.15 Schlack–Kumpf method for C-terminal stepwise peptide degradation.

residues is feasible. The present standard should allow routine analysis of most proteins, and hence C-terminal sequence analysis has become a useful complement to N-terminal degradation and MS.

Enzymatic degradation by carboxypeptidase is an attractive method for C-terminal sequencing, both by determination of the released amino acids and by identification of the truncated peptides with mass spectrometry.

2.3.2.7 Mass Spectrometry

Mass spectrometry has proved to be a very useful tool in the analysis of peptides and proteins [81], as well as in the rapidly developing area of proteome analysis [82, 83]. MALDI [84] and ESI [85] currently are the dominant methods for ionization of biomacromolecules. The latter technique may also be coupled to separation techniques such as HPLC and capillary electrophoresis (CE). The different techniques of biomolecular MS enable an exact determination of the molecular mass of a peptide or protein with high sensitivity and resolution.

Nanospray techniques have subsequently pushed the detection limits for peptide sequencing into the attomolar range [86]. The advances of the genome-sequencing projects and bioinformatics methods have also changed the position of MS, which is now becoming a deeply integrated tool in proteomics. Proteins separated on two-dimensional polyacrylamide gels can be identified at the sub-picomolar level [87]. In this approach, the protein present in the gel is digested by treatment with a protease (e.g., trypsin) and the resulting peptide fragments (peptide fingerprint, peptide mass map) are identified by searching against databases. MALDI peptide mapping operates efficiently when a statistically significant number of peptide peaks is detected and assigned unambiguously in the digest. Alternatively, nanoelectrospray MS can be applied if MALDI peptide mapping fails to identify the protein.

In particular, MALDI-MS and ESI-MS offer an alternative to the classical Edman peptide sequencing. The soft desorption of these techniques allows the transfer of high-molecular mass polypeptide ions into the gas phase without fragmentation [88, 89]. MALDI is often combined with a very sensitive time-of flight (ToF) detector. As a rule, short peptides can be directly sequenced, for example by MALDI-ToF techniques via post-source decay (PSD), while preceding cleavage to give suitable fragments is an imperative prerequisite for proteins and longer polypeptides. As indicated above (Section 2.3.2.6), either the released amino acid or the truncated peptides obtained from C-terminal sequencing using carboxypeptidase can be identified by MS coupled with various instrumental variations such as fast-atom bombardment [90], plasma desorption [91], ESI [92], or MALDI-MS.

Tandem mass spectrometry (MS/MS, or MSⁿ) was originally developed for the analysis of the molecular structure of single ions. These ions are selected by mass and further fragmented by employing collision-induced dissociation (CID) and subsequent analysis of the resulting fragments. MS/MS is routinely employed to determine the site and nature of a modification (post-translational or chemical). Targeted chemical modification of proteins may be detected by MS, and has been utilized as a probe of protein secondary structure or for the characterization of active sites.

Coupling to high-performance separation techniques allows quantitation. MS/MS is a suitable method for the sequence analysis of a peptide, and is proving indispensable for proteome analysis. Peptides are prone to fragment at amide bonds after low-energy collisions, resulting in a predictable fragmentation pattern. Consequently, sequence information can be obtained by this technique because most amino acids have unique masses. Only the pairs of Leu/Ile and Lys/Gln cannot be distinguished. Chemically modified amino acids have, in most cases, a molecular mass that is not identical with one of the naturally occurring amino acids and hence can be readily identified in the fragmentation pattern. Furthermore, disulfide bond assignment in proteins is possible using MALDI/MS.

The application of an ion-trap instrument for MS³ experiments and a computer algorithm for automated data analysis have led to a novel concept of two-dimensional fragment correlation MS and its application in peptide sequencing [93]. Even though the daughter ion (MS²) spectrum of a peptide contains the sequence information of the peptide, it is very difficult to decipher the MS² spectrum due to the difficulty in distinguishing the N-terminal fragments from the C-terminal fragments. This problem can be solved by taking a grand-daughter ion (MS³) spectrum of a particular daughter ion, since all fragment ions of the opposite terminus are eliminated in this spectrum. The sequence of the peptide can be derived from a two-dimensional plot of the MS² spectrum versus the intersection spectra (2-D fragment correlation mass spectrum). Using this technique, about 78% of the sequence of a tryptic digest of cytochrome c could be determined. Interestingly, this MS *de-novo* sequencing approach works with complex mixtures, does not require any additional wet chemistry step, and should be fully automated in the near future.

2.3.2.8 Peptide Ladder Sequencing

Peptide ladder sequencing combines ladder-generating chemistry and MS. The principle of peptide ladder sequencing based on Edman degradation with MALDI-MS was first developed by Chait *et al.* [94] in 1993.

N-terminal ladder sequencing requires a modified Edman procedure in which in every step the peptide is incompletely degraded to yield continuously a mixture of one amino acid-shortened peptides. Such a peptide ladder is formed when Edman degradation with PITC (cf. Figure 2.14) is performed in the presence of 5% phenyl isocyanate (PIC) as terminating reagent.

The phenylthiocarbamoyl peptide is cleaved in the presence of a strong acid to give the 5(4*H*)-thiazolone, while the phenylcarbamoyl peptide, formed from PIC, is stable under these conditions. During the next cycles the content of phenylcarbamoyl peptides (PC peptides) increases to a statistical mixture, thereby forming the peptide ladder. Analysis of the mixture using MALDI-MS allows direct sequence determination from the successive mass differences of the peptide ladder (Figure 2.16).

Since, contrary to the classical Edman procedure, quantitative derivatization is not necessary, one degradation cycle can be performed within approximately 5 min. Furthermore, application of the volatile trifluoroethyl isothiocyanate resulted in a significant optimization of this procedure and allowed peptide sequencing at the femtomole level [95]. The equal masses of Leu and Ile, and the small mass differences

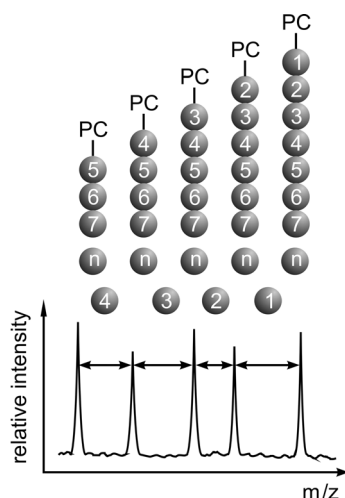


Figure 2.16 Derivation of the amino acid sequence from the peptide ladder from mass differences analyzed by MALDI-MS. PC = phenylcarbamoyl.

between Glu/Gln and Asp/Asn, represent serious problems. The latter requires sufficient mass resolution, provided for example by modern MALDI-ToF spectrometers in a range up to 5000 Da.

C-terminal ladder sequencing is based on the same principle, and initially was combined with C-terminal sequence analysis after carboxypeptidase treatment (Section 2.3.2.6). This procedure was applied both by time-dependent [96, 97] and concentration-dependent [98] CP digestion. Another C-terminal ladder sequencing approach uses Schlack–Kumpf chemistry (Section 2.3.2.6) coupled with MALDI-MS analysis of the truncated peptides [99]. This chemical one-pot degradation procedure is applied without purification steps, and no repetitions must be performed to obtain ladders of C-terminal fragments – so-called ragged ends. It could be demonstrated that the 20 common proteinogenic amino acid residues are compatible with this technique, but only up to eight residues of C-terminal sequences can be determined. MALDI instruments with delayed extraction [100, 101] allow the discrimination of all amino acids except Leu and Ile. Lys and Gln, having the same mass, can be distinguished after chemical acetylation with the formation of *N*^ε-acetyllysine.

2.3.2.9 Assignment of Disulfide Bonds and Peptide Fragment Ordering

In order to determine disulfide bond positions present in a peptide or protein, the latter is hydrolyzed under conditions such that the risk of disulfide exchange is minimized. The classical approach involves enzymatic cleavage using peptidases of low specificity, e.g., thermolysin or pepsin. The pairs of proteolytic fragments linked by disulfide bonds are then separated by diagonal electrophoresis. This technique includes electrophoretic separation in two dimensions using identical conditions. After electrophoresis in the first dimension, the electropherogram is exposed to performic acid vapor in order to oxidize cystine residues to cysteic acid (see

Section 2.3.2.2). Electrophoresis in the second dimension is then carried out under the same conditions. Those peptide fragments not modified by performic acid are located on the diagonal of the electropherogram, because their migration is similar in both directions. In contrast, cystine-containing peptide fragments will be oxidatively cleaved to give two new peptides that occur off-diagonally. Normally, fragments with an intrachain disulfide give only one new product, whereas fragments connected by interchain disulfide bonds are transformed to two new peptide products. After isolation of the parent disulfide-linked peptide fragment, the disulfide bond is cleaved, followed by alkylation and fragment sequence analysis. Alternative procedures are based on RP-HPLC and MS.

The strategy for sequence determination of a peptide or protein subsequent to enzymatic cleavage requires the correct ordering of the peptide fragments to be sequenced individually. This can be performed by comparing the sequences of two sets of overlapping peptide fragments that result from polypeptide cleavage with different sequence specificities (Section 2.3.2.4). The principle is shown in Table 2.3.

Amino acid analysis in the case of Table 2.3 gives the result that the polypeptide consists of 30 amino acids. The peptide has Phe at its N-terminus, as determined by Sanger's end-group analysis. Since the absence of Met excludes selective cleavage with BrCN (Section 2.3.2.4), enzymatic cleavage can be performed with chymotrypsin and trypsin, respectively. The pattern of five overlapping fragments leads to the conclusion that the original sequence corresponds to the B chain of human insulin.

2.3.2.10 Location of Post-Translational Modifications and Bound Cofactors

Post-translationally modified proteins (see Section 3.2.2) and those with permanently associated prosthetic groups fulfil important functions in biochemical pathways. In order to determine the location of modified amino acids, the protein must be degraded under conditions that are similar to those described for assignment of disulfide bonds (Section 2.3.2.9). Especially mild conditions are required to detect, for example, γ -carboxyglutamic acid (Gla) in various proteins as it is decarboxylated very easily under acidic conditions to give glutamic acid. Many proteins (and especially enzymes) contain covalently bound nonpeptide cofactors, e.g., N^ε-lipoyllysine in dihydrolipoyl transacetylase, 8 α -histidylflavin in succinate dehydrogenase, and biotin linked to the enzyme via the ϵ -amino group of a lysine residue. Subsequent to specific peptide chain cleavage reactions (as mentioned above), the resulting fragments containing the modified building block may be efficiently analyzed using various MS-based techniques (Section 2.3.2.7).

Nonenzymatic post-translational modifications cause protein degradation both *in vivo* and *in vitro*. The formation of 3-nitrotyrosine (3-NT), protein carbonyls, advanced glycation end-products (AGE), oxidation of methionine to methionine sulfoxide and tyrosine to dityrosine are selected examples that require exact quantification and characterization. The presence of 3-NT-containing proteins in tissue results from the reaction with reactive nitrogen species that are formed *in vivo* [102]. Besides reduction of 3-NT to 3-aminotyrosine, protein tyrosine nitration can be characterized directly by MS of a purified full-length or proteolytically digested peptide or protein [103, 104]. Protein carbonyls may be formed under oxidative stress

Table 2.3 Determination of the primary structure of a polypeptide by comparing the sequences of two sets of overlapping peptide fragments.

Analytical steps								Conclusion
Amino acid analysis								
Ala	1	Gln	1	Leu	4	Ser	1	The polypeptide consists of 30 amino acids.
Arg	1	Glu	2	Lys	1	Thr	2	
Asn	1	Gly	3	Phe	3	Tyr	2	
Cys	2	His	2	Pro	1	Val	3	
Determination of terminal groups according to Sanger								
Phe								Phe is the N-terminal amino acid residue.
Enzymatic cleavage								
Cleavage with chymotrypsin								Fragment C forms the C-terminus of the original peptide, because it does not contain either an aromatic amino acid (chymotrypsin cleavage) or a Lys/Arg residue (trypsin cleavage) at the C-terminus.
A Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr								
B Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe								
C Thr-Pro-Lys-Thr								
D 2 Phe, 1 Tyr								
Cleavage with trypsin								
E Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg								
F Gly-Phe-Phe-Tyr-Thr-Pro-Lys								
G Thr								
Sequence analysis and elucidation								
A B C FVNQHLGSHLVEALYLCGERGFFYTPKT E F								The polypeptide has the sequence of human insulin B chain.

by direct conversion of an amino acid side chain into a carbonyl, or by covalent attachment of carbonyl-carrying molecules like 4-hydroxynonenal. The detection of a protein carbonyl can be performed after derivatization with 2,4-dinitrophenylhydrazine (2,4-DNP) and either spectrophotometric analysis or staining with an anti-DNP antibody [105]. AGE formation is associated, for example, with aging and diabetes. Post-translational modifications of proteins (glycosylation, oxidation, phosphorylation) may be detected by MS. Electrospray and MALDI-ToF mass spectrometry have been applied to the direct analysis of, for example, glycosylated proteins. Methylglyoxal-dependent modifications to N^ε-carboxymethyllysine could be detected in human lens proteins as a result of age-dependent reactions [106]. Higher contents of N^ε-carboxymethyllysine-protein adducts have also been found in the peripheral

nerves of human diabetics [107]. General strategies for the separation and identification of glyated proteins have been reviewed [108]. Further oxidative modifications result from the oxidation of methionine to methionine sulfoxide, or tyrosine to dityrosine [109]. Hydroxyl radical oxidation leads to formation of 3-hydroxyvaline and 5-hydroxyvaline, as well as 3,4-dihydroxyphenylalanine (DOPA), *o*- and *m*-tyrosine, and dityrosine [110]. The oxidation of Met to methionine disulfoxide and of Tyr to dityrosine, as well as the hydroxylation of aromatic amino acid side chains in proteins, are often correlated with pathological phenomena and can be detected by MS/MS [111]. Deamidation [112] and diketopiperazine formation [113] complete the nonenzymatic post-translational modifications.

2.4

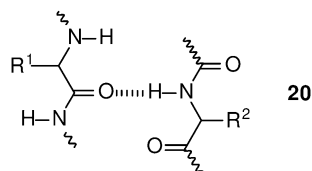
Three-Dimensional Structure

As already mentioned (Section 2.3), the structural features of peptides and proteins are described in terms of primary, secondary, and tertiary structure. While the primary structure classifies the number and sequence of the amino acid residues in a peptide or protein chain, the secondary structure describes ordered conformations of the peptide backbone, which are called secondary structure elements. The three-dimensional arrangement of secondary structure elements, which is essentially determined by the amino acid side chains, is called the tertiary structure.

2.4.1

Secondary Structure

The peptide chain conformation is characterized by the backbone torsion angles ϕ and ψ of the amino acid constituents and the dihedral angle ω of the peptide bond. (see Section 2.1, Figure 2.1). In comparison to alkyl chains, the conformational space is rather restricted. Due to the partial double bond character of the peptide bond leading to a relatively high rotational barrier, only torsion angle values around 0° (*cis*) or 180° (*trans*) are accepted for ω . An overview of the accessible torsion angle regions of ϕ and ψ can be obtained when considering the steric interactions during rotation around the corresponding bonds. These regions are often displayed in Ramachandran plots (Figure 2.17) [114, 115]. The actual conformation of a peptide chain under physiological conditions is essentially determined by hydrogen bonds between peptide bonds, specific side-chain interactions, hydrophobic effects and solvent influences.



Most typical secondary structure elements in peptides and proteins are characterized by specific hydrogen bond patterns. A hydrogen bond **20** basically is formed

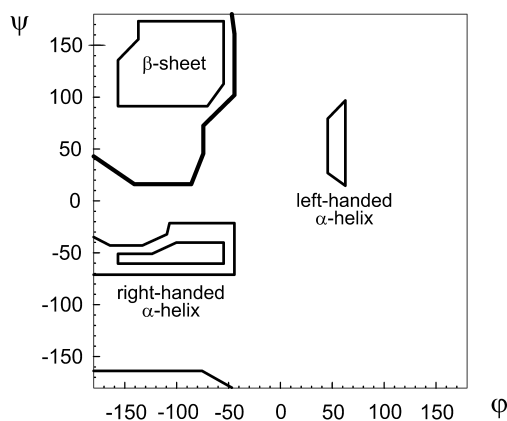


Figure 2.17 Ramachandran plot.

between the NH group (hydrogen bond donor) of one peptide bond and the carbonyl oxygen atom (hydrogen bond acceptor) of another peptide bond. The distance between the oxygen and the nitrogen atom in a hydrogen bond is about 280 pm.

The stabilization energy of a single hydrogen bond is relatively small (20 kJ mol^{-1}) compared to a covalent bond ($200\text{--}400 \text{ kJ mol}^{-1}$). Moreover, it has to be considered that hydrogen bond donor and acceptor groups in peptides or proteins are often part of hydrogen bonds to the solvent water. In most secondary structure elements that are stabilized by hydrogen bonding, multiple hydrogen bonds are formed, and these multiple interactions of such a cooperative system result in considerable stabilization. In the following sections, the most important secondary structure elements are described. If the torsion angles ϕ and ψ , respectively, of all amino acid constituents in such ordered structures have the same values, these secondary structure elements are called periodic. Helices and sheet structures belong to this group of secondary structure elements. If the torsion angle values of the amino acid constituents are different in the ordered structures, the secondary structures are non-periodic. This group is dominated by turn structures.

2.4.1.1 Helices

Helices are widely occurring secondary structure elements comprising screw-like arrangements of the peptide backbone stabilized by intramolecular hydrogen bonds aligned in parallel to the helix axis. Helices are characterized by a well-defined number of amino acid residues per turn (n), the helix pitch (h , repeat distance), and the number of skeleton atoms incorporated into the “ring” formed by the intramolecular hydrogen bond. Helices are chiral objects and the direction of the helical turn (handedness) is given by the letters “P” (plus) for clockwise and “M” (minus) for anti-clockwise helices. If the thumb of the right hand points along the helix axis in the direction of the peptide sequence (from the N- to the C-terminus) and the helix winds in the direction of the finger bowing, the helix is right-handed (P-helix). The opposite convention is valid for a left-handed helix (M-helix).

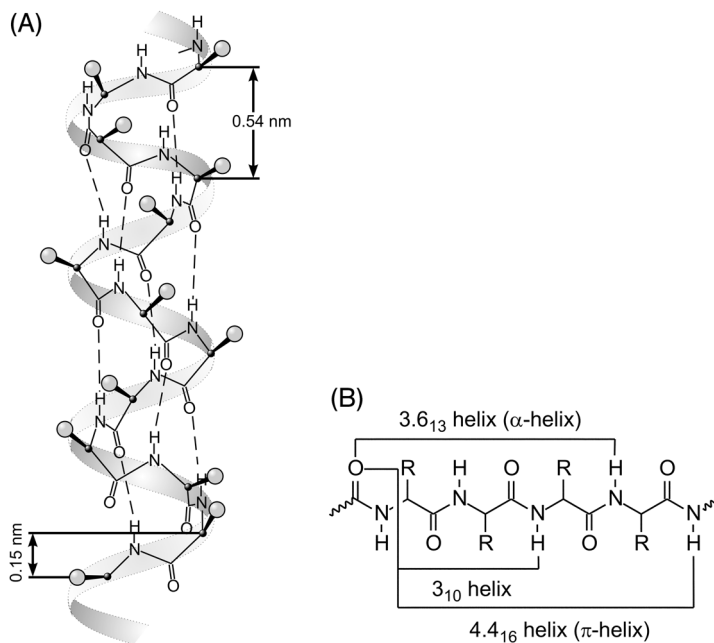


Figure 2.18 Hydrogen bond pattern in α -helical peptides (A) and schematic view of the hydrogen bond patterns in different helices (B).

The most common helix is the α -helix (Figure 2.18), which was originally proposed by Pauling and Corey based on theoretical investigations regarding the X-ray diffraction patterns of α -keratins. The α -helix of L-amino acid residues comprises a right-handed spiral arrangement of the peptide backbone with 3.6 amino acid residues per turn ($n = 3.6$), a helix pitch (h) of 540 pm, and the torsion angles $\phi = -57^\circ$, $\psi = -47^\circ$. It is stabilized by hydrogen bonds directed backwards from a C-terminal NH to an N-terminal CO ($\text{NH}^{i+4} \rightarrow \text{CO}^i$) forming a 13-membered “ring” (Figure 2.18). Consequently, the full nomenclature of such an α -helix is 3.6_{13} -P-helix. The amino acid side chains are oriented perpendicularly to the helix axis in order to reduce steric strain. A less prominent helix type is the 3_{10} -helix (3_{10} -P-helix, $\phi = 60^\circ$, $\psi = -30^\circ$), which sometimes caps α -helices in native peptides. In the 3_{10} -helix, the hydrogen bonds are also formed in the backward direction of the sequence ($\text{NH}^{i+3} \rightarrow \text{CO}^i$) [116]. The existence and role of the π -helix (4.4_{16} -P-helix, $\phi = -57^\circ$, $\psi = -70^\circ$) are under permanent debate [117, 118].

The nature of the amino acid side chains is of crucial importance for helix stability. Helix compatibility of a series of amino acids has been examined [119]. The amino acids proline and hydroxyproline as well as other (nonproteinogenic) N-alkylated amino acids are not able to act as hydrogen bond donors, and display high helix-breaking properties. However, they occur in the collagen triple helix. Glycine also does not have any conformational bias towards helix formation, whereas many other

amino acids (Ala, Val, Leu, Phe, Trp, Met, His, Gln) are highly compatible with helical structures. The general criteria for helix stabilization by amino acid residues are: (i) the steric requirements of the amino acid side chain; (ii) electrostatic interactions between charged amino acid side-chain functionalities; (iii) interactions between distant amino acid side chains ($i \leftrightarrow i + 3$ or $i \leftrightarrow i + 4$); (iv) the presence of proline residues; and (v) interactions between the amino acid residues at the helix termini and the electrostatic dipole moment of the helix.

α -Helices can only be formed by peptide chains of homochiral building blocks, which contain exclusively D- or exclusively L-amino acids. The right-handed α -helix (from L-amino acids) is the preferred conformation, for energetic and stereochemical reasons, but a minimum number of amino acids is required for the formation of this helix type.

It should be mentioned that helices can also be formed in sequences with L- and D-amino acids arranged in alternate order. Interestingly, such helices show a periodicity of dimer units, that is, the torsion angles of every second amino acid constituent have the same values. In such helices, hydrogen bonds are formed alternately in the forward and backward directions of the sequence, resulting in rings of different size. The hydrogen bonding pattern has some similarity to that of a parallel sheet structure (see Section 2.4.1.2). Therefore, such helices are called β -helices or mixed helices. In particular, in apolar environments, these helices are very stable [120]. The most prominent example is the channel-forming peptide gramicidin A exhibiting a helix with 20- and 22-membered hydrogen-bonded rings in alternate order [121].

2.4.1.2 β -Sheets

The hydrogen bond pattern of β -sheets differs fundamentally from that of helical structures, with hydrogen bonds being formed between two neighboring polypeptide chains. Two major variants of β -sheet structures may be distinguished.

- The parallel β -sheet, where the chains are aligned in a parallel manner (Figure 2.19A).
- The antiparallel β -sheet, where two neighboring peptide chains connected by hydrogen bonds are aligned in an antiparallel manner (Figure 2.19B).

An ideal β -sheet structure of a peptide chain is characterized by ϕ and ψ angles of $\pm 180^\circ$.

A hypothetical fully extended oligoglycine chain is characterized by the angles $\phi = -180^\circ$ and $\psi = 180^\circ$. This periodic conformation, known as the extended or “zig-zag” conformation, however, cannot be accommodated without distortion when side chains are present. In this case, an antiparallel β -pleated sheet displays torsion angles $\phi = -139^\circ$ and $\psi = 135^\circ$ (Figure 2.19). β -Pleated sheets are found in silk fibroin and other β -keratins, as well as in several domains of globular proteins. The side chains of amino acids involved in β -sheet formation are aligned in an alternating manner towards both sides of the β -sheet. β -Sheet structure is much more complex than a simple ribbon diagram might imply. Several variants of fundamental conformational states of polypeptides in the β -region are known, in which the backbone is extended near to its maximal length, and to more complex architectures in which extended

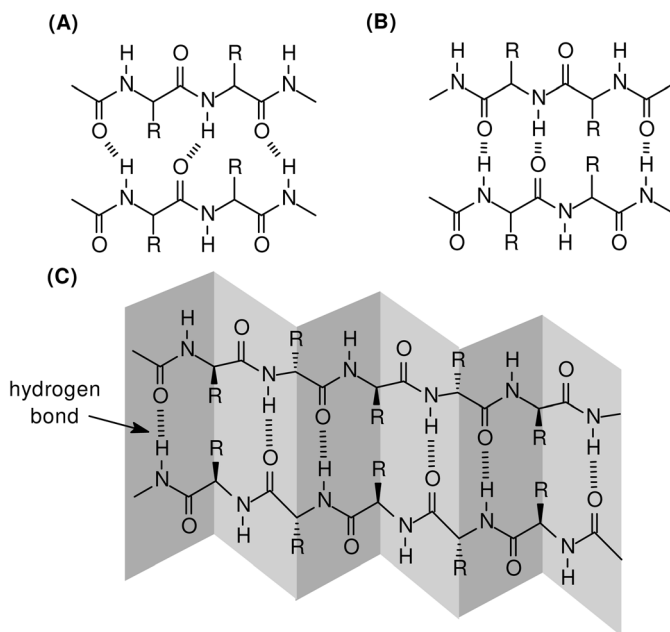


Figure 2.19 Hydrogen bond pattern in parallel (A) and antiparallel (B, C) β -pleated sheets.

segments are linked by turns and loops [122]. β -Sheets may occur in a twisted, curled, or backfolded form. They usually exhibit a right-handed twist, favored by nonbonded intrastrand interactions and interstrand geometric constraints. With respect to tertiary structure, layers of β -sheets are usually oriented relative to each other either at a small angle ($\sim 30^\circ$) in aligned β -sheet packing, or close to 90° in orthogonal β -sheet packing [123]. Statistical studies of proteins of known structure revealed that β -branched and aromatic amino acids most frequently occur in β -sheets, while Gly and Pro tend to be poor β -sheet-forming residues. β -Sheets often occur in the hydrophobic core of proteins. Consequently, the β -sheet-forming propensities of amino acids may reflect the hydrophobic requirement more than a real β -sheet-forming propensity. The β -pleated sheet has been postulated as a structure into which any amino acid could substitute [122].

2.4.1.3 Turns

Helices and sheet structures are uni-directional. Therefore, loops are required to reverse the direction of a polypeptide chain. The characteristic secondary structure element of a loop is the reverse turn. A polypeptide chain cannot fold into compact globular structure without involving tight turns that usually occur on the exposed surface of proteins. Hence, turns play a role in molecular recognition and provide defined template structures for the design of new molecules such as drugs, pesticides, and antigens. Turns are also found in small peptides. Often, but not necessarily, turns are stabilized by a hydrogen bond between an amino group located

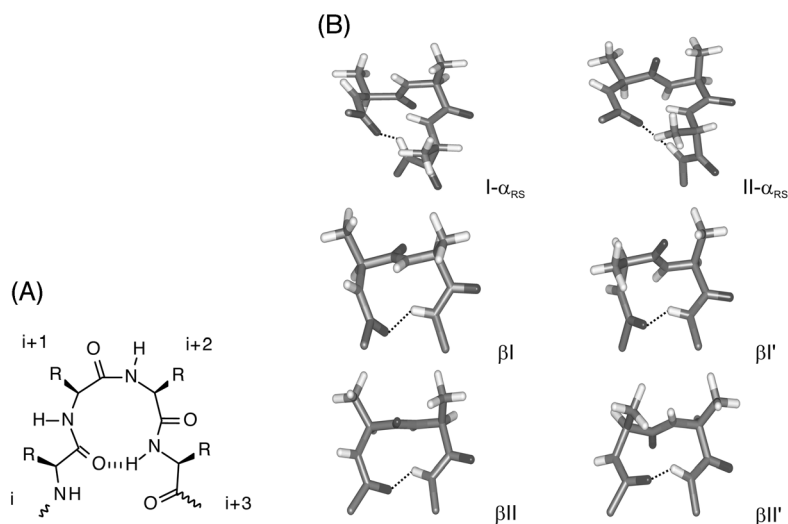


Figure 2.20 Schematic view of a β -turn (A) and selected α - and β -turns (B).

C-terminally and a carboxy group located N-terminally. Turns are classified according to the number of amino acid residues involved as γ -turns (three amino acids), β -turns (four amino acids), α -turns (five amino acids), or π -turns (six amino acids) (Figure 2.20). The conformation of α -turns is determined by the dihedral angles of the three central amino acids ($i + 1$, $i + 2$, $i + 3$). Nine α -turn types have been classified by Pavone *et al.* [124] (Table 2.4).

A general criterion for the existence of a β -turn is that the distance of the atoms $C^\alpha(i)$ and $C^\alpha(i + 3)$ is smaller than 7 Å. β -Turns can be further classified according to the characteristic dihedral angles φ and ψ (Figure 2.20; Table 2.4) of the second ($i + 1$) and the third ($i + 2$ amino acid). Originally β -turns were classified into types I, II, and III and the pseudo-mirror images (I' , II' , III'). Subsequently, the definition was broadened (Table 2.4) [125] and the number of β -turns types was increased. However, since type III β -turns are the basic structural elements of the 3_{10} -helix, type III β -turns have been eliminated from this classification. These loop structures differ, therefore, in the spatial orientation not only of the NH and CO functions of the amino acids in positions $i + 1$ and $i + 2$, but also of the side chains [122, 126]. The three-dimensional side-chain orientation is given by the vectors $C^\alpha \rightarrow C^\beta$. If functional groups are present in the side chain, they may be crucial for peptide–receptor interactions.

Some types of β -turns are stabilized intrinsically by certain amino acids. Proline has the highest tendency to occur in a reverse turn. The pyrrolidine ring in L-proline restricts the dihedral angle φ to -60° : Therefore, proline with a *trans*-configured peptide bond is found preferentially in position $i + 1$ of βI - or βII -turns. Proline with a *cis*-peptide bond occurs in position $i + 2$ of a βVIa -turn, which is also named a proline-turn. Mainly D-proline, but also D-amino acids, in general have a high preference to occur in position $i + 1$ of a $\beta II'$ -turn. Glycine is often considered as

Table 2.4 Characteristic torsion angles (in degrees) in the most important secondary structures.

Type	φ_{i+1}	ψ_{i+1}	φ_{i+2}	ψ_{i+2}	φ_{i+3}	ψ_{i+3}
γ -turn	75	-64				
γ^i -turn	-79	69				
β I turn	-60	-30	-90	0		
β I' turn	60	30	90	0		
β II turn	-60	120	80	0		
β II' turn	60	-120	-80	0		
β IV turn	61	10	-53	17		
β VIa turn (1)	-60	120	-90	0		
β VIa turn (2)	-120	120	60	0		
β VIIb turn	-135	135	-75	160		
β VIII turn	-60	-30	-120	120		
I- α_{RS} turn	-60	-29	-72	-29	-96	-20
I- α_{LS} turn	48	42	67	33	70	32
II- α_{RS} turn	-59	129	88	-16	-91	-32
II- α_{LS} turn	53	-137	-95	81	57	38
I- α_{RU} turn	59	-157	-67	-29	-68	-39
I- α_{LU} turn	-61	158	64	37	62	39
II- α_{RU} turn	54	39	67	-5	-125	-34
II- α_{LU} turn	-65	-20	-90	16	86	37
I- α_C turn	-103	143	-85	2	-54	-39
3_10 -helix	-60	-30				
α -Helix	-57	-47				
π -Helix	-57	-70				
Polyproline I helix ($\omega = 0^\circ$)	-75	160				
Polyproline II helix ($\omega = 180^\circ$)	-75	145				
antiparallel β -pleated sheet	-139	135				
parallel β -pleated sheet	-119	113				

“proteinogenic D-amino acid”, because it is, like D-amino acids, often found in positions $i + 1$ or $i + 2$, respectively, of a turn [126]. β -Turns are the most economical peptide geometry that can link extended segments of the peptide chain with direction reversal that enables the formation of a so-called β -hairpin. The latter can serve as a module in more extended antiparallel β -sheet structures.

2.4.1.4 Amphiphilic Structures

Amphiphilic (amphipathic) compounds are at the same time both hydrophilic and hydrophobic (Figure 2.21). In amphiphilic helices, one side of the helix mainly presents hydrophobic residues, while the other side mainly contains hydrophilic residues. Interactions between hydrophobic residues in an aqueous environment mediate molecular self-assembly of amphiphilic peptides and stabilize the aggregate.

Correct placement of hydrophobic and hydrophilic residues within an amino acid sequence promotes the formation of amphiphilic helices (Figure 2.22). Helices may associate as α -helical hairpins (helix–turn–helix), coiled coils [127], or helix bundles. The association may be either inter- or intramolecular. In a hydrophilic environment,

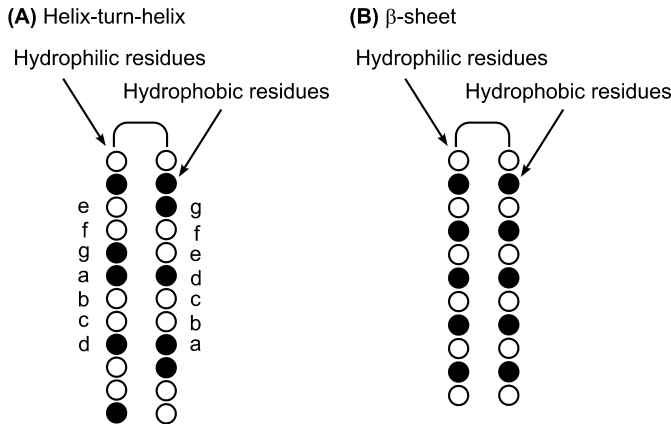


Figure 2.21 Positioning of hydrophobic and hydrophilic residues within an amino acid sequence that fold as helix-turn-helix (α -helical hairpin, A) or β -sheet (β -hairpin, B) motif.

the hydrophobic residues assist the formation of a secondary structure (α -helix, β -sheet; Figure 2.21) and form a stabilizing core between the single secondary structure elements. The hydrophobic surface areas are not exposed to the aqueous (hydrophilic) environment because of the association of two or more helices via a hydrophobic interhelix interface.

Amphiphilic β -sheets are usually composed of alternating polar and nonpolar residues within the β -strand sequence. Consequently, association of the β -strands gives an amphiphilic β -hairpin or β -sheet, where the hydrophobic faces may associate in a sandwich-like fashion. Amphiphilic helices also interact with lipid membranes. The 26-peptide melittin is present in a monomeric form at low concentrations in aqueous media of low ionic strength, where it is largely in a random coil conformation according to circular dichroism (CD) spectroscopy. However, at the lipid bilayer interface, melittin adopts a highly helical conformation. Peptides such as melittin

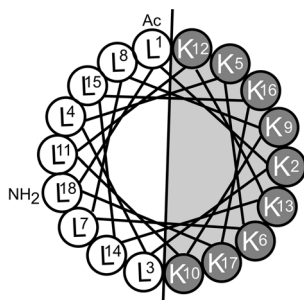


Figure 2.22 Sequence and Schiffer-Edmundson's helical wheel representation of the amphipathic α -helical 18-peptide LKLLKKLLKK¹⁰LKKLLKKL [128].

belong to the defense system of a variety of species, and enhance the permeability of the lipid membrane by directly disturbing the lipid matrix. They usually disrupt the transmembrane electrochemical gradient. Magainins and cecropins have been shown to adopt α -helical secondary structure in membrane environments [129]. Many amphiphilic peptides are thought to form transmembrane helical bundles, though whether amphiphilic peptides align along the lipid bilayer plane or whether they form transmembrane helical structures remains a controversial issue. A lipid bilayer is ~ 30 Å thick, which corresponds to the length of a 20-peptide α -helical structure. Hydrophobic or amphiphilic peptide segments of about 20 amino acids length are often found in natural ion channel proteins. These peptides are regarded as being able to penetrate membranes perpendicularly [130]. The detergent-like characteristics of amphiphilic helical peptides might provide an alternative explanation for their cytotoxic activity [131].

2.4.2

Tertiary Structure

In helices or β -sheets, the conformation of a polypeptide chain is determined not only by hydrogen bonds but also by additional interactions and bonds that stabilize the chain's three-dimensional structure (Figure 2.23). Metal chelation by different groups within a protein fold (e.g., zinc finger) is a further stabilizing factor.

The disulfide bond is the second type of covalent bond besides the amide bond in a polypeptide chain. It contributes sequence-specifically to formation of the so-called tertiary structure, the conformation of the full peptide chain. A disulfide bond is formed by oxidation of the SH groups of two cysteine residues. Intramolecular disulfide bonds are formed within a single polypeptide chain, while intermolecular disulfide bonds occur between different peptide chains. A torsion angle of $\sim 110^\circ$ is observed at the disulfide bridge. The disulfide bonds are not cleaved thermally because of the high bond energy of 200 kJ mol^{-1} (cf. C–C: 330 kJ mol^{-1} , C–N: 260 kJ mol^{-1} , C–H: 410 kJ mol^{-1}). However, they can be cleaved reductively by an excess of reducing agents such as thiols (e.g., dithiothreitol, DTT).

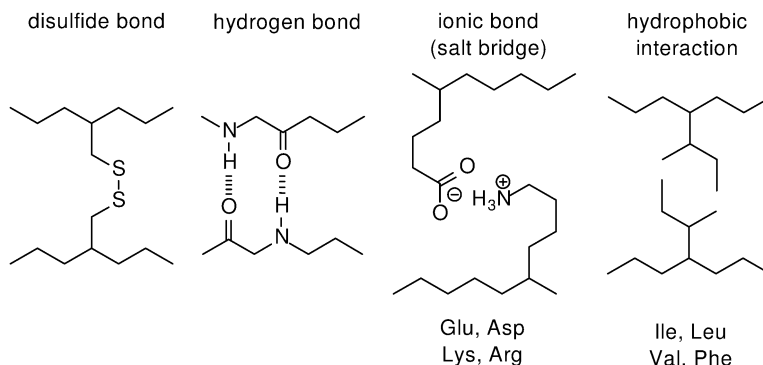


Figure 2.23 Stabilizing interchain interactions between amino acid side chains.

Hydrogen bonds may be formed not only between structural elements of the peptide bonds (as shown for secondary structures), but also between suitable side-chain functionalities of trifunctional amino acids.

The functional groups in the side chains of acidic or basic amino acids are completely or partially ionized at physiological pH. Therefore, electrostatic interactions are observed between acidic and basic groups. These ionic bonds (salt bridges) between carboxylate groups (aspartyl, glutamyl or the C-terminus) and N-protonated residues (arginine, lysine, histidine or the N-terminus) with a bond energy of $\sim 40\text{--}85 \text{ kJ mol}^{-1}$ influence peptide conformation. Electrostatic interactions are also extended to the hydrate shell. Moreover, ion–dipole and dipole–dipole interactions are observed; these occur because of electrostatic interactions between polarizable groups (SH groups, OH groups) with relatively small binding contributions. Generally, electrostatic interactions are highly dependent on the pH, salt concentration, and dielectric constant of the medium.

Hydrophobic interactions, which describe the aggregation phenomena of apolar molecules or structure elements in an aqueous environment, are eminently important in the stabilization of peptide chain conformations due to the occurrence of amino acids with apolar side chains. Such hydrophobic effects cannot be explained by the relatively weak van der Waals–London forces between the apolar groups. On the contrary, interactions between an apolar solute and the solvent water could be even more favorable than between the apolar solutes themselves, when only focussing on the energy or enthalpy contributions. The hydrophobic effect is determined by entropy effects. Hydrophobic regions are covered in aqueous solution by a molecular layer of higher ordered water molecules connected with an entropy decrease in the solvent. Aggregation of apolar groups or molecules increases entropy again by reducing the contact surface area between apolar solutes and solvent allowing some part of the water molecules to take part in the aqueous hydrogen bonding network. Thus, the hydrophobic effect is solvent-driven. Comparable effects can be observed in all structured solvents.

Figure 2.24 illustrates these aspects, which can be quantitatively understood on the basis of the Gibbs–Helmholtz equation $\Delta G = \Delta H - T\Delta S$. State 1 in Figure 2.24 is thermodynamically disfavored because of the higher degree of order of the water molecules around the apolar side chains. Although both hydrophobic residues in state 2 are ordered to a higher degree now, the degree of order of the water molecules has decreased dramatically ($\Delta S > 0$). The contribution of $T\Delta S$ is so considerable that the free enthalpy of the system decreases for the transition from state 1 to state 2 ($\Delta G < 0$). The influence of the enthalpy is quite small in this case.

In general, the three-dimensional arrangement of a peptide chain in a globular polypeptide or protein is characterized by a relatively small content of periodic structural elements (α -helix, β -sheet) and shows an unsymmetrical and irregular structure. The cooperativity between hydrogen bonds and hydrophobic interactions and other noncovalent interactions is basically the reason for the formation of stable three-dimensional structures. Under physiological conditions, thermodynamically stable conformations of a biologically active peptide with a minimum of free enthalpy occur.

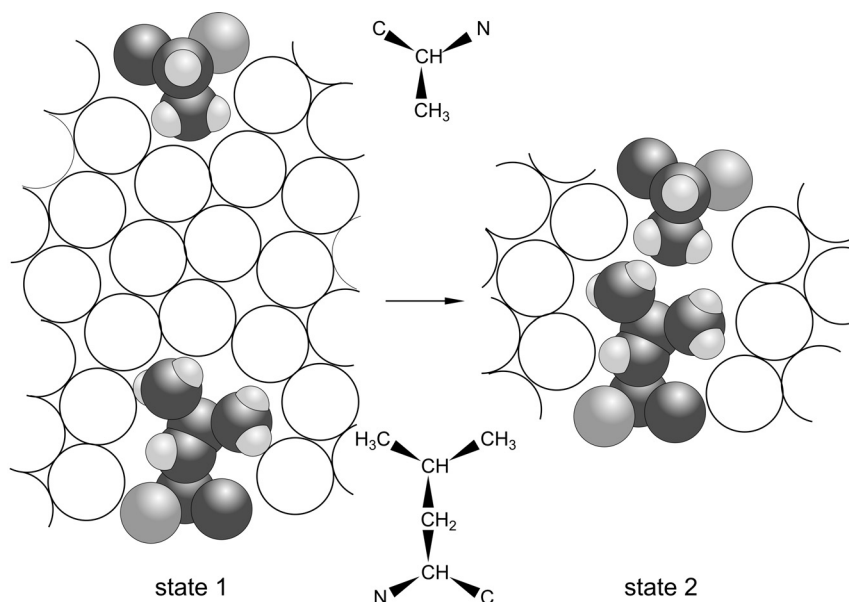


Figure 2.24 Hydrophobic interactions between Ala and Leu side chains in an aqueous medium.

Tertiary structure formation is based on supersecondary structure elements, these being formed by the association of several secondary structures. The helix–turn–helix motif ($\alpha\alpha$, Figure 2.25A), the β -hairpin motif ($\beta\beta$, Figure 2.25B), the Greek key motif (β_4), and the $\beta\alpha\beta$ motif belong to the most frequent supersecondary structure elements. The increasing amount of protein structural data led to the classification of tertiary protein folds. To date, more than 500 distinct protein tertiary folds have been characterized, and these represent one-third of all existing globular folds. The rigid framework of secondary structure elements is the best defined part of a protein structure. The special organization of secondary structure elements (topology) may be divided into several classes. Tertiary structure is formed by packing secondary structure elements into one or several compact globular units (domains). Tertiary fold family classification [132] is used in different databases (SCOP [133] and CATH [134]). The overall agreement between these databases suggests the existence of a natural logic in structural classification. In the CATH database, structures are grouped into fold families depending on both the overall shape and the connectivity of the secondary structure. In general, three classes of domains (tertiary structure elements) can be distinguished (Figure 2.25):

1. Structures containing only α -helices (e.g., 7-helix bundles, Figure 2.25C; α -superhelix, Figure 2.25F)
2. Structures containing exclusively antiparallel β -pleated sheets (e.g., β -propellers, Figure 2.25D)
3. Structures containing α -helices and β -sheets (e.g., TIM barrel, Figure 2.25E; α,β -superhelix, Figure 2.25G)

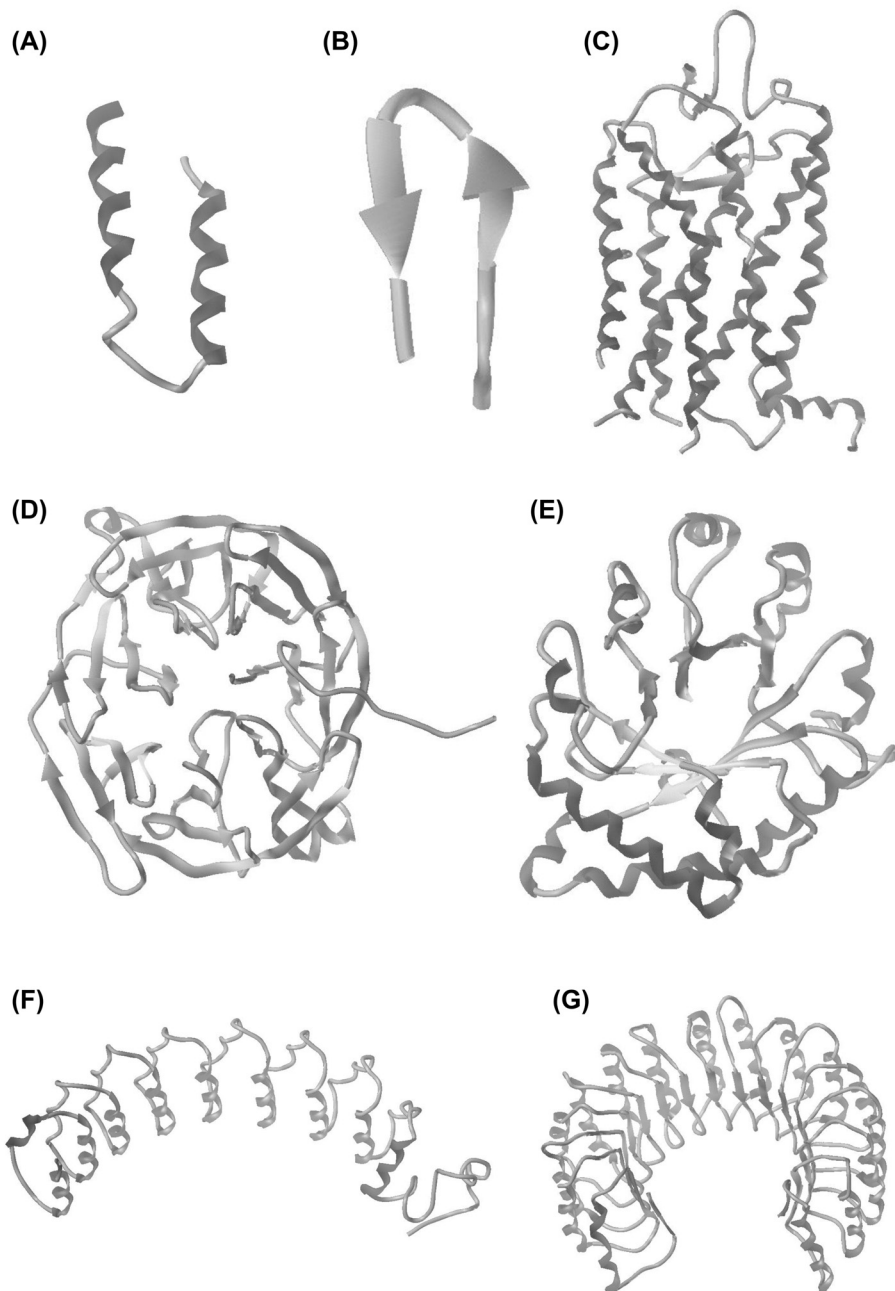


Figure 2.25 Examples of tertiary folds.

Many efforts have been made to predict protein structural classes. In contrast to α -helices and β -sheets, very few methods have been reported for predicting tight turns [135].

2.4.2.1 Structure Prediction

Approaches to protein structure prediction are based on the thermodynamic hypothesis which postulates the native state of a protein as the state of lowest free energy under physiological conditions [136]. Studies of protein folding (structure prediction, fold recognition, homology modeling, and homology design) generally make use of some form of effective energy function.

Considering the fact that a protein consisting of 100 amino acids may adopt three possible conformations per amino acid residue, and that the time for a single conformational transition is 10^{-13} s, the average time a protein would need to find the optimum conformation for all 100 residues by searching the full conformational space is 10^{27} years. In fact, protein folding takes place on a time scale of seconds to minutes. This contradiction has been called the Levinthal paradox, which states that there is insufficient time to search randomly the entire conformational space available to a polypeptide chain in the unfolded form [137]. Levinthal concluded that protein folding follows multiply branched pathways where local secondary structure is formed in the first stage, which is governed by the interaction of neighboring amino acid side chains. Hence, secondary structure formation is considered to be an early event in protein folding. Alternatively, it has been proposed that initially a hydrophobic collapse takes place to form a partially folded structure from which secondary structure subsequently forms. Protein structure prediction remains the “holy grail” of protein chemistry. Until recently, with few exceptions, the prediction of protein structure has been more conceptually than practically important. Predictions were rarely accurate enough to deduce biological function or to facilitate the structure-based design of new pharmaceuticals.

Protein structure prediction basically relies on two strategies: (i) *ab initio* prediction [138]; and (ii) homology modeling [139, 140]. Homology models are based primarily on the database information. Threading, or fold-recognition methods lie between these two extremes, and involve the identification of a structural template that most closely resembles the structure in question. In cases where homologous (>30% homology) or weakly homologous sequences of known structure are not available, the most successful methods for structure prediction rely on the prediction of secondary structure and local structure motifs. Secondary structure prediction is gaining increasing importance for the prediction of protein structure and function [141]. Secondary structures of peptides and proteins are partly predictable from local sequence information based on knowledge of the intrinsic propensities of amino acids to form a helix or a β -sheet. The prediction of ambivalent propensities can assist in the definition of regions that are prone to undergo conformational transitions. Although primary protein sequence information may provide an educated guess about function (especially when well-characterized homologous sequences exist), incomplete annotation, false inheritance and multiple structural and functional domains may disturb the interpretation of database searches based on primary sequence alone.

2.5

Methods of Structural Analysis

The three-dimensional structure of a peptide or a protein is the crucial determinant of its biological activity. As the various genome projects are steadily approaching their final goal, attention now returns to the functions of the proteins encoded by the genes. This will increasingly transform structural biology into structural genomics [142]. Especially in drug design and molecular medicine, the mere information of a gene sequence is not sufficient to obtain information about a corresponding protein at the molecular level. A large number of methods have been established on the way to the 3-D structure of peptides and proteins [143]. Preliminary structural information can be obtained by biochemical or biophysical methods, e.g. electrophoresis, gel filtration or dynamic light scattering. X-ray small angle scattering (SAXS) on protein solution can provide information of the molecular shape, whereas electron microscopy (EM) allows studies of large macromolecules and bimolecular complexes. Like SAXS, EM provides information on the size and shape of macromolecules up to a resolution of 10 Å requiring only small amounts of the sample. NMR (cf. Section 2.5.3) is well suited for solution studies of macromolecules investigating molecular interactions and molecular flexibility, whereas macromolecular X-ray crystallography (cf. Section 2.5.4) provides most of the 3D protein structure information.

Besides the structural analysis methodology discussed in this chapter, techniques for biomolecular interaction analysis such as surface plasmon resonance [144], fluorescence correlation spectroscopy [145, 146], and microcalorimetry [147] are steadily gaining importance.

2.5.1

Circular Dichroism [148, 149]

Linear polarized light consists of two circular polarized components of opposite helicity but identical frequency, speed, and intensity. When linear polarized light passes through an optically active medium, for instance a solution containing one enantiomer of an optically active compound, the speed of light in matter is different for the left and right circular polarized components (different refractive indices). In such a case a net rotation of the plane of polarization is observed for the linear polarized light. Consequently, enantiomerically pure or enriched optically active compounds can be characterized by the optical rotation index and optical rotatory dispersion.

However, it is not only the speed of the two circular polarized components but also the extinction by chiral chromophores that may be different. If this is the case, elliptically polarized light is observed. CD spectroscopy detects the wavelength dependence of this ellipticity, and positive or negative CD is observed when either the right- or the left-circular polarized component is absorbed more strongly.

CD is a method of choice for the quick determination of protein and peptide secondary structure. Proteins are often composed of the two classical secondary

structure elements, α -helix and β -sheet, in complex combinations. Besides these ordered regions, other parts of the protein or peptide may exist in a random coil state. CD spectroscopy is a highly sensitive method to distinguish between α -helical, β -sheet, and random coil conformations. However, it is only able to estimate the proportions of these secondary structure elements and not their positions (*vide infra*). Therefore, information obtained by CD spectroscopy is limited compared to that obtained by NMR or X-ray diffraction. CD data can be valuable as a preliminary guide to peptide and protein conformational transitions under a wide range of conditions. Peptides and proteins that lack non-amino acid chromophores (e.g., prosthetic groups) do not exhibit absorption or CD bands at wavelengths above 300 nm. The amide group is the most prominent chromophore of peptides and proteins to be observed by CD spectroscopy. Two electronic transitions of the amide chromophore have been characterized. The $n\text{--}\pi^*$ transition is usually quite weak and occurs as a negative band around 220 nm. The energy (wavelength) of the amide $n\text{--}\pi^*$ transition is sensitive to hydrogen bond formation. The $\pi\text{--}\pi^*$ transition is usually stronger, and is registered as a positive band around 192 nm and a negative band around 210 nm.

As already mentioned, the proportions of α -helical secondary structure, β -sheet conformation, and random coil can be determined by CD spectroscopy. An α -helical conformation is usually characterized by a negative band at 222 nm ($n\text{--}\pi^*$), a negative band at 208 nm, and a positive band at 192 nm. Short peptides do not usually form stable helices in solution; however, it has been shown that the addition of 2,2,2-trifluoroethanol (TFE) leads to an increase in the helix content of most peptides [150]. As discussed in Section 2.4.1.2, β -sheets in proteins are less well-defined, compared to the α -helix, and can be formed either in a parallel or antiparallel manner. β -Sheets display a characteristic negative band at 216 nm and a positive band of comparable size close to 195 nm. Random coil conformations (unordered conformations) are usually characterized by a strong negative CD band just below 200 nm. Several algorithms are available that allow computational secondary structure analysis of peptides and proteins by fitting the observed spectrum with the combination of the characteristic absorption of the three secondary structures mentioned above. One interesting approach is to deconstruct a protein into a series of synthetic peptides that are then analyzed by CD [151]. In special cases, aromatic side chains of amino acids and disulfide bridges may also serve as chromophores that can give rise to CD bands.

However, UV absorption spectra and electronic CD spectroscopy (ECD) contain limited information because of their inherently low resolution, high signal overlap, and conformational insensitivity of the few accessible electronic transitions. Vibrational circular dichroism (VCD) is a new method that provides alternative views on the conformation of proteins and peptides. The vibrational region (IR range) of the spectrum resolves characteristic transitions of the amide functionality. VCD sensitively discriminates β -sheet and helices as well as disordered structure. For peptides, VCD analysis is often combined with DFT calculations of the expected spectra for comparison. Much recent work has focused on empirical and theoretical VCD analyses of peptides, with detailed prediction of helix, sheet and hairpin spectra [152].

2.5.2

Infrared Spectroscopy [153]

Fourier transform infrared spectroscopy (FT-IR) is nowadays commonly applied to peptides and proteins, and is mainly used to estimate the content of secondary structure elements [154]. A common approach towards these studies has been to prepare peptides that correspond to protein epitopes, and to examine the structure under various conditions (solvent systems, biomembrane models). Additionally, turn-forming model peptides [155–157] and helix models [158, 159] have been investigated. Furthermore, protein hydration and the structural integrity of lyophilized proteins are often examined by FT-IR. The samples may be investigated in solution, in membrane-like environments, and in the solid state, for example after adsorption onto a solid surface. In the latter cases the technique of attenuated total reflectance (ATR) is often involved [160]; this is especially valuable when the analyte displays low solubility or tends to associate in higher concentrations. The sample is applied in the solid state, but may be hydratized as in its natural environment.

ATR-FT-IR is one of the most powerful methods for recording IR spectra of biological materials in general, and for biological membranes in particular [161]. It can also be applied to proteins that cannot be studied by X-ray crystallography or NMR. ATR-FT-IR requires only very small amounts of material (1–100 μg) and provides spectra within minutes. It is especially suited for studies concerning structure, orientation, and conformational transitions in peptides and membrane proteins. Furthermore, temperature, pressure, and pH may be varied in this type of investigation and, additionally, specific ligands may be added. No external chromophores are necessary.

The amide N–H stretching band (νNH), the amide I band around 1615 cm^{-1} ($\nu\text{C=O}$), and the amide II band around 1550 cm^{-1} (δNH) are the characteristic features that are usually examined by FT-IR spectroscopy of peptides and proteins. Recently, side-chain carboxy groups of proteins (Asp, Glu) and the OH group of the tyrosine side chain have become targets of FT-IR investigations. Signal deconvolution techniques are usually applied in order to identify the distinct absorption frequencies. The formation of hydrogen bonds characteristically influences the free amide vibrations mentioned above, and allows distinction to be made between α -helical, β -sheet, and random coil conformation.

IR spectroscopy allows monitoring of the exchange rate of amide protons, and hence provides collective data for all amino acid residues present in a protein or peptide. Polarized IR spectroscopy provides information on the orientation of parts of a protein molecule present in an ordered environment.

The replacement of an amide hydrogen by deuterium is highly sensitive to changes in the environment. Moreover, the exchange kinetics may be used to detect conformational changes in the protein structure. Protons exposed on the protein surface will undergo hydrogen/deuterium exchange much more rapidly than those present in the protein core. Amide protons present in flexible regions buried in the protein or involved in secondary structure formation are characterized by medium exchange rates. It has been shown that careful analysis of the amide I band during the

deuteration process provides information that helps to assign the exchanging protons to a secondary structure type [162].

2.5.3

NMR Spectroscopy [162–165]

NMR spectroscopy is one of the most widely used analytical techniques for structure elucidation [166]. Nowadays, more-dimensional NMR methods are used routinely for the resonance assignment and structure determination of peptides and small proteins. While ^1H is the nucleus to be detected in unlabeled peptides and proteins, proteins uniformly labeled with ^{13}C and ^{15}N provide further information by the application of heteronuclear NMR techniques. NMR studies directed towards the elucidation of the three-dimensional protein structure rely, especially for proteins, on isotope labeling with ^{13}C and ^{15}N in connection with three- and more-dimensional NMR methods [167]. These labeled proteins are usually provided by the application of overexpression systems utilizing isotope-enriched culture media.

For some time, the limit for the elucidation of the three-dimensional protein structure by NMR with respect to the molecular mass was considered to be 30 kDa. However, recent developments and the construction of ultra-high-field superconducting magnets have extended the molecular mass range of molecules to be investigated with NMR well beyond 100 kDa. NMR spectra of such large proteins usually suffer from considerable line broadening, but this has been overcome by the development of transversal relaxation-optimized spectroscopy (TROSY) developed by Wüthrich's group [168]. At molecular masses above ~ 20 kDa, spin diffusion becomes a limiting problem because of the longer correlation time of the protein. Consequently, the transversal relaxation time becomes short, which leads to line broadening. These limitations may be overcome by random partial deuteration of proteins [169] and by using the TROSY technique. The molecular mass is not the only determinant for successful NMR structure determination. Before NMR studies may be conducted, investigations must be carried out in order to establish whether the peptide or protein forms aggregates, and whether the folded state of the protein is stable under the experimental conditions.

NMR studies can be applied either in the solution phase or in the solid phase. The peptide or protein is dissolved in aqueous or nonaqueous solvents that may also contain detergents (for the analysis of proteins in a membrane-like environment). In recent years, solid-state NMR has been increasingly used in the area of membrane protein structure elucidation [8]. Solid-state NMR spectroscopy is an attractive method to investigate peptides that are immobilized on solid surfaces, or peptide aggregates such as amyloid fibrils [132].

Crystallization of the proteins, which remains the major obstacle in X-ray structural analysis, is not necessary in NMR studies. The solution conditions (pH, temperature, buffers) can be varied over a wide range. Nowadays, even NMR investigations on the folding pathway of a protein are amenable. In such cases, partially folded proteins are often considered as models for transient species formed during kinetic refolding [170]. Moreover, NMR studies provide a dynamic picture of the protein under investigation.

The chemical shift value is one of the classical NMR parameters used in structural analysis. Insufficient signal dispersion of larger molecules, however, requires the application of additional parameters. Typically, scalar couplings (through-bond connection) and dipolar couplings (through-space connection, nuclear Overhauser effect, NOE) are used for the assignment of the nuclei. In particular, NOE information provides valuable data on the spatial relationship between two nuclei. These internuclear distances (e.g., interproton) are usually indispensable for elucidation of the three-dimensional structure and are used together with other geometric constraints (covalent bond distances and angles) for the computation of three-dimensional protein or peptide structures.

To determine the three-dimensional structure of a peptide, initially all signals in the NMR spectra are assigned to the amino acid residues present. If the peptide sequence is known – which is usually the case for compounds obtained by synthesis or overexpression – the primary structure can be verified by inter-residue NOE signals. If the sequence is unknown, it may be established by analysis of NOESY spectra.

The crosspeak volume integrals of NOESY spectra also provide valuable distance information for the corresponding nuclei. Calibration of these crosspeak volumes with those of known internuclear distance (e.g., geminal protons) leads to a direct conversion of the crosspeak volume values to interproton distances. The sequential distances $d(\text{H}^\alpha, \text{H}^N)$, $d(\text{H}^N, \text{H}^N)$, and $d(\text{H}^\beta, \text{H}^N)$ depend on the torsion angles of the bonds involved. Consequently, the coupling constants between two protons provide information on the torsion angle between these two protons according to the Karplus equation. Regular secondary structures are also characterized by a variety of medium-range or long-range backbone interproton distances that are sufficiently short to be observed by NOE. Helices and turns usually display short sequential and medium-range $^1\text{H}^1\text{H}$ distances, while β -sheets usually display short sequential and long-range backbone $^1\text{H}^1\text{H}$ distances. The characteristic internuclear distances in helices, β -sheets and β -turns are compiled in the classical monograph by Wüthrich [171].

NMR studies are especially helpful in determining the three-dimensional peptide structure when applied to conformationally constrained peptides. This is the case for either cyclic peptides or peptides containing sterically hindered amino acids. Linear unconstrained peptides are usually too flexible to adopt a preferred conformation in solution.

The amide N–H exchange rates in D_2O solutions and the temperature coefficients of the NH proton chemical shift are further valuable data for conformational analysis of peptides and proteins. An NH proton that is exposed to the solvent will undergo much faster exchange by deuterons in D_2O . Plots of the amide chemical shift versus temperature are usually linear, and their slope is referred to as the temperature coefficient. In general, the temperature coefficients of amide proton resonances for extended peptide chain structures are of the order of -6 to -10 ppb K^{-1} , while greater values of the temperature coefficient (>-4 ppb K^{-1}) are correlated with solvent-inaccessible environments or hydrogen bonds. Consequently, higher temperature coefficients should correlate with slowly exchanging amide protons, and a combination of these data can be used in order to identify regular secondary structures, especially in the case of small linear or cyclic peptides.

Automation programs and tools for the recognition of spin systems have been designed on the basis of pattern recognition techniques [172] in order to assist sequence-specific assignments. Once an ensemble of three-dimensional structures has been calculated from the NMR data, it has to be refined with respect to geometry and constraint violations. In addition to the determination of three-dimensional protein structures in solution, NMR provides valuable information on local structure, conformational dynamics and on the interaction of the protein with small molecules. Consequently, NMR is nowadays a highly versatile tool, for example, in industrial drug research, because it can detect very weak ligand–protein interactions that are characterized by only millimolar binding constants [173, 174]. In this context, transferred NOE experiments should be mentioned especially, as they may reveal the protein-bound conformation of a small-molecular weight ligand, provided that an equilibrium between the bound and unbound state exists [175]. Furthermore, the SAR by NMR technique (cf. Section 7.1) as described by Fesik *et al.* has proved to be a valuable tool in drug discovery.

2.5.4

X-Ray Crystallography

X-ray crystallography has become the predominant technique for 3D structure determination and allows visualization of the macromolecular structure down to very small building blocks, the individual atoms and, sometimes, even electrons. Some proteins have been characterized by this technique at a resolution <1 Å, while the standard resolution of newly determined protein structures is in the range of 2 Å. Meanwhile, more than 41 000 X-ray structures of protein and protein complexes have been deposited in the protein structure database, Protein Data Bank (PDB), [176, 177]. X-ray crystallography relies on the diffraction of an X-ray beam at the single crystal lattice. The wavelength range of X-rays corresponds to the size of the diffracting structures (atomic radii and lattice constants), and the observed diffraction pattern results from a superposition of the diffracted beams. The rotating-crystal method uses monochromatic X-rays at a constant wavelength with variation of the angle of incidence, while the Laue methods apply polychromatic X-rays of different wavelengths at a constant angle of incidence. The latter method, when used in combination with a cyclotron beam as the radiation source, provides sufficient data for structural analysis within a short period of time. Ultimately, this will enable time-resolved X-ray analysis of proteins in order to study catalytic mechanisms, for example. Nowadays, intense, highly focused X-rays from integrated cyclotron radiation beamlines facilitate the traditionally tedious and time-consuming process of structural analysis [178].

The protein to be characterized must first be purified and crystallized. Sequence analysis prior to X-ray diffraction experiments is helpful to support the assignment. Protein crystallization can be seen as a multi-parameter optimization process where the optimum concentration and purity of the protein, the concentration of the precipitation agent, ionic strength, pH value, buffer, and temperature may be varied. The presence of detergents is necessary for the crystallization of some proteins, and consequently protein crystallization is the crucial bottleneck in structural analysis by

X-ray diffraction. Additionally, novel methods for automatic crystallization and automatic crystallographic data analysis [179] facilitate this method and will eventually provide methodology for high-throughput determination of three-dimensional protein structures.

The reflex pattern obtained upon X-ray diffraction provides information on the crystal lattice constants, the symmetry of the crystal, and its space group. The number of observed reflexes should be as high as possible. After data scaling, the phase problem must be solved. The phase determination uses, for example, heavy metal ions incorporated into the crystal (soaking), but in other cases a known protein structure with sequence homology >35% to the protein under investigation may be used as a starting structure for the refinement. Once a suitable molecular model has been obtained, it is further refined, and then finally validated to provide a three-dimensional structure that is related as closely as possible to the electron density map obtained from X-ray diffraction.

In contrast to proteins that are highly hydrated even in the crystalline state, peptide crystals usually display limited hydration. Smaller peptide molecules usually provide very good crystals, and a resolution of <1 Å is observed in X-ray analysis. On occasion, multiple conformations may be encountered in peptide crystals due to the flexibility of peptides and to similar conformer energies. These conformers may be present in the same crystal, or in separate crystals [180].

Nowadays, X-ray crystallography is combined with complementary methods to get structural information at different levels to highlight the link between the biology of the cellular processes caused by the 3D structure of the specific protein and the underlying chemistry of protein function. Examples are the placement of crystal structures into SAXS envelopes of large multiprotein complexes, the application of EM envelopes for crystallographic structure solution by molecular replacement and the combination of X-ray structure analysis with molecular dynamics simulations or quantum chemical calculations.

2.5.5

UV Fluorescence Spectroscopy

Fluorescence spectroscopy [181] is widely used in peptide and protein chemistry, either observing intrinsic fluorophors (Trp, Tyr), fluorescent cofactors, or extrinsic fluorophors that are used to label the protein. As fluorescence spectroscopy involves electronic transitions, it can be applied to study very fast processes. Interactions of proteins with other proteins [182], nucleic acids [182], small ligands, and membranes can be monitored, as well as protein folding [183] and conformational transitions [184]. The association of peptides can be monitored by fluorescence quenching [185]. Usually, fluorescence intensity, anisotropy, and emission wavelength may be observed. Intracellular processes may be observed by fluorescence spectroscopy when the protein to be examined is expressed by genetic engineering as a fusion protein with green fluorescent protein (GFP) [186].

Fluorescence resonance energy transfer (FRET) is one technique involving fluorescence spectroscopy that is being applied increasingly [187]. The fluorescence of

one chromophore present in a molecule can be quenched by interaction with other chromophores. This is not a process that involves collision or complex formation that is usually required for electronic coupling of the fluorophore and the quencher molecule. In FRET, energy is transferred across even bigger distances of up to 10 nm. Efficient energy transfer according to the FRET mechanism is only possible when the absorption range of the acceptor chromophore and the fluorescence range of the donor chromophore overlap. The donor must be a fluorescent moiety with a sufficiently long fluorescence lifetime. Moreover, the donor and acceptor must be oriented correctly with respect to each other. It is noteworthy that FRET is a radiationless process and proportional to R^{-6} , where R is the distance between the donor and the acceptor. Therefore, the distance R between acceptor and donor can be determined in FRET experiments. FRET can be used to monitor interactions between biomolecules. In addition, the cleavage of synthetic substrates by proteases may be easily monitored, even in a highly parallel manner by labeling the substrate with the fluorescence donor on one end and the fluorescence acceptor on the other end [188].

Fluorescence correlation spectroscopy (FCS) [146] is a method used to monitor biomolecular interactions. Single molecules can be detected because a confocal volume element in the femtoliter range is observed. Concentrations between 10^{-9} and 10^{-15} M can be detected, and the diffusion times of fluorescently labeled molecules will change upon binding to a potential binding partner. The diffusion time observed is subsequently correlated with the molecular size. This method allows the determination of kinetic parameters [189].

2.6

Review Questions

- Q2.1. How many amino acids are genetically encoded?
- Q2.2. Name two rare genetically encoded amino acids and draw their structure.
- Q2.3. Explain the difference between a *cis*- and a *trans*-configured peptide bond
- Q2.4. Give the three-letter-code abbreviation of the dipeptide alanyl-alanine.
- Q2.5. Give two methods for peptide separation on a preparative scale.
- Q2.6. Explain the mechanism of N-terminal sequencing by Edman degradation.
- Q2.7. How does cyanogen bromide react with peptides?
- Q2.8. Discuss C-terminal sequencing according to Schlack–Kumpf.
- Q2.9. Name the characteristic torsion angles used to describe peptide conformations.
- Q2.10. What is a Ramachandran plot?
- Q2.11. Name three basic elements of secondary structure found in peptides and proteins and discuss their properties.

Q2.12. What is a tertiary fold?

Q2.13. What are the requirements for NMR-based structure analysis of larger peptides and proteins < 30 kDa?

References

- 1 G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, 296, 1459.
- 2 G. Fischer, H. Bang, C. Mech, *Biomed. Biochim. Acta* **1984**, 43, 1101.
- 3 G. Fischer, A. Aumüller, *Rev. Physiol. Biochem. Pharmacol.* **2004**, 148, 105.
- 4 L. Min, D. B. Fulton, A. H. Andreotti, *Front. Biosci.* **2005**, 10, 385.
- 5 G. Fischer, H. Bang, C. Mech, *Biomed. Biochim. Acta* **1984**, 43, 1101.
- 6 L. Min, D. B. Fulton, A. H. Andreotti, *Front. Biosci.* **2005**, 10, 385.
- 7 IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) , Nomenclature and Symbolism for Amino Acids and Peptides, Recommendations, 1983, *Eur. J. Biochem.* **1994**, 138, 9.
- 8 C. K. Larive, S. M. Lunte, M. Zhong, M. D. Perkins, G. S. Wilson, G. Gokulrangan, T. Williams, F. Afroz, C. Schöneich, T. S. Derrick *et al.* *Anal. Chem.* **1999**, 71, 389.
- 9 P. Cutler (Ed.), *Protein Purification Protocols*, Humana Press, Totowa, NJ. **2004**.
- 10 L. R. Bonner, *Protein Purification*, Taylor & Francis, New York, **2007**.
- 11 F. Hachem, B. A. Andrews, J. A. Asenjo, *Enzyme Microb. Technol.* **1996**, 19, 507.
- 12 M. Lu, F. Tjerneld, *J. Chromatogr. A* **1997**, 766, 99.
- 13 S. Gedela, N. R. Medicherla, *Chromatographia* **2007**, 65, 511.
- 14 Y. Ito, Y. Ma, *J. Chromatogr. A* **1996**, 753, 1.
- 15 Y. Ma, Y. Ito, *Anal. Chim. Acta* **1997**, 352, 411.
- 16 T. Yoshida, *J. Chromatogr.* **1998**, 88, 105.
- 17 W. S. Hancock, *Handbook of HPLC for the Separation of Amino Acids, Peptides, and Proteins*, Volumes 1 and 2, CRC Press, Boca Raton, **1984**.
- 18 C. T. Mant, R. S. Hodges, *High Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis, and Conformation*, CRC Press, Boca Raton, **1991**.
- 19 J. K. Swadesh, *HPLC: Practical and Industrial Applications*, CRC Press, Boca Raton, **2001**.
- 20 P. R. Levison, C. Mumford, M. Streater, A. B. Nielsen, N. D. Pathirana, S. E. Badger, *J. Chromatogr. A* **1997**, 760, 151.
- 21 A. Vailaya, C. Horvath, *Ind. Eng. Chem. Res.* **1996**, 35, 2964.
- 22 J. Porath, *J. Protein Chem.* **1997**, 16, 463.
- 23 H. G. Barth, B. E. Boyers, C. Jackson, *Anal. Chem.* **1998**, 70, 251.
- 24 P. Cuatrecasas, *Adv. Enzymol.* **1972**, 36, 29.
- 25 P. Bailon, G. K. Ehrlich, W.-J. Fung, W. Berthold, (Eds.), *Affinity Chromatography: Methods and Protocols*, Humana Press, **2007**.
- 26 J. A. Buettner, C. A. Dadd, B. L. Masecar, D. J. Hammond, *Int. J. Peptide Protein Res.* **1996**, 47, 70.
- 27 D. S. Waugh, *Trends Biotechnol.* **2005**, 23, 316.
- 28 J. W. Jorgenson, K. D. Lukacs, *Anal. Chem.* **1981**, 53, 1298.
- 29 K. M. Hutterer, J. W. Jorgenson, *Anal. Chem.* **1999**, 71, 1293.
- 30 G. Teshima, S. L. Wu, *Methods Enzymol.* **1996**, 27.
- 31 M. Hadley, M. Gilges, J. Senior, A. Shab, P. Camilleri, *J. Chromatogr. B* **2000**, 745, 177.
- 32 Y. Shen, J. Berger, G. A. Anderson, R. D. Smith, *Anal. Chem.* **2000**, 72, 2154.

- 33 A. Görg, W. Weiss, M. J. Dunn, *Proteomics* **2004**, *4*, 3665.
- 34 R. Ghosh, *Protein Bioseparation Using Ultrafiltration*, Imperial College Press, London, **2003**.
- 35 T. Vorherr, *Speciality Chemicals Magazine*, May **2007**, *27* (4), 40.
- 36 D. M. Bollag, M. D. Rozycki, S. J. Edelstein, *Protein Methods*, Wiley, New York, **1996**.
- 37 J. Porath, J. Carlsson, I. Olsson, G. Belfrage, *Nature* **1975**, *258*, 598.
- 38 N. K. Shin, D.-Y. Kim, C.-S. Shin, M.-S. Hong, J. Lee, H.-C. Shin, *J. Biotechnol.* **1998**, *62*, 143.
- 39 H. Rogl, K. Kosemund, W. Kühlbrandt, I. Collinson, *FEBS Lett.* **1998**, *432*, 21.
- 40 Z. Ji, D. I. Pinon, L. J. Miller, *Anal. Biochem.* **1996**, *240*, 197.
- 41 C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem. Int. Ed.* **2000**, *39*, 450.
- 42 J. E. Dyr, J. Suttmar, *J. Chromatogr. B* **1997**, *699*, 383.
- 43 M. Kaufmann, *J. Chromatogr. B* **1997**, *699*, 347.
- 44 J. W. van Nispen in: *Topics in Pharmaceutical Sciences*, D. D. Breimer, P. Speiser (Eds.), Elsevier, Amsterdam, **1987**.
- 45 H. R. Costantino, R. Langer, A. M. Klibanov, *J. Pharm. Sci.* **1994**, *83*, 1662.
- 46 J. W. van Nispen, R. Pinder, *Annu. Rep. Med. Chem.* **1986**, *21*, 51.
- 47 L. Chafetz, *Pharm. Technol.* **1984**, *24*.
- 48 R. Pearlman, T. J. Nguyen, *J. Pharm. Pharmacol.* **1992**, *44*, 178.
- 49 R. G. Strickley, G. C. Visor, L. Lin, L. Gu, *Pharm. Res.* **1989**, *6*, 971.
- 50 N. Bhatt, K. Patel, R. T. Borchardt, *Pharm. Res.* **1990**, *7*, 593.
- 51 G. M. Jordan, S. Yoshioka, T. Terao, *J. Pharm. Pharmacol.* **1994**, *46*, 182.
- 52 M. J. Hageman, *Drug Dev. Ind. Pharm.* **1988**, *14*, 2047.
- 53 M. C. Lai, E. M. Topp, *J. Pharm. Sci.* **1999**, *88*, 489.
- 54 A. L. Burlingame, R. K. Boyd, S. J. Gaskell, *Anal. Chem.* **1996**, *68*, 559.
- 55 F. Sanger, *Annu. Rev. Biochem.* **1988**, *57*, 1.
- 56 M. Walker, *The Protein Protocols Handbook*, Humana Press, Totowa, NJ, **2002**.
- 57 B. J. Smith, *Protein Sequencing Protocols*, Humana Press, Totowa, NJ, **2003**.
- 58 F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463.
- 59 A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560.
- 60 F. Sanger, *Biochem. J.* **1945**, *39*, 507.
- 61 W. R. Gray, B. S. Hartley, *Biochem. J.* **1963**, *89*, 59.
- 62 H. Hanson, D. Glaser, M. Ludewig, H. G. Mannsfeldt, M. John, *Hoppe-Seyler's Z. Physiol. Chem.* **1967**, *348*, 689.
- 63 H. Zuber, G. Roncari, *Angew. Chem. Int. Ed. Engl.* **1967**, *6*, 880.
- 64 H. Feracci, S. Maroux, *Biochim. Biophys. Acta* **1980**, *599*, 448.
- 65 S. Akabori, K. Ohno, T. Ikenaka, Y. Okada, H. Hanafusa, I. Haruna, A. Tsugita, K. Sugae, T. Matsushima, *Bull. Chem. Soc. Jpn.* **1956**, *29*, 507.
- 66 C. Cooper, N. Packer, K. Williams (Eds.), *Amino Acid Analysis Protocols*, Humana Press, Totowa, NJ, **2000**.
- 67 E. Gross, B. Witkop, *J. Am. Chem. Soc.* **1961**, *83*, 1510.
- 68 A. Patchornik, W. B. Lawson, B. Witkop, *J. Am. Chem. Soc.* **1958**, *80*, 4747.
- 69 A. Fontana, W. E. Savie, M. Zombonin in: *Methods in Peptide and Protein Analysis*, C. Birr (Ed.), Elsevier/North Holland Biomedical Press, Amsterdam, **1980**.
- 70 W. C. Mahoney, P. K. Smith, M. A. Hermodson, *Biochemistry* **1981**, *20*, 443.
- 71 P. Edman, *Arch. Biochem.* **1949**, *22*, 475.
- 72 P. Edman, G. Begg, *Eur. J. Biochem.* **1967**, *1*, 80.
- 73 R. W. Hewick, M. W. Hundapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **1981**, *256*, 7990.
- 74 R. A. Laursen, A. G. Bonner, *Chem. Eng. News* **1970**, *48*, 52.
- 75 P. Schlack, W. Kumpf, *Hoppe-Seyler's Z. Physiol. Chem.* **1926**, *154*, 125.
- 76 J. Li, S. Liang, *Anal. Biochem.* **2002**, *302*, 108.

- 77 G. R. Stark, *Biochemistry* **1968**, 7, 1796.
- 78 V. L. Boyd, M. Bozzini, G. Zon, R. L. Noble, R. J. Mattaliano, *Anal. Biochem.* **1992**, 206, 344.
- 79 J. M. Bailey, N. R. Shenoy, M. Ronk, J. E. Shively, *Protein Sci.* **1992**, 1, 68.
- 80 T. Bergmann, E. Cederlund, H. Jörnvall, *Anal. Biochem.* **2001**, 290, 74.
- 81 G. Siuzdak, *Mass Spectrometry for Biotechnology*, Academic Press, San Diego, **1996**.
- 82 T. Rabilloud, *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*, Springer, Berlin, **2000**.
- 83 R. M. Kamp, J. J. Calvete, T. Choli-Papadopoulou (Eds.), *Methods in Proteome and Protein Analysis*, Springer, **2004**.
- 84 M. Karas, F. Hillenkamp, *Anal. Chem.* **1988**, 60, 2299.
- 85 J. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **1985**, 246, 64.
- 86 M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, *Nature* **1996**, 379, 466.
- 87 A. Shevchenko, A. Loboda, A. Shevchenko, W. Ens, K. G. Standing, *Anal. Chem.* **2000**, 72, 2132.
- 88 B. T. Chait, S. B. H. Kent, *Science* **1992**, 257, 78.
- 89 R. W. Nelson, M. A. McLean, T. W. Hutchens, *Anal. Chem.* **1994**, 66, 1408.
- 90 C. V. Bradley, D. H. Williams, M. R. Hanley, *Biochem. Biophys. Res. Commun.* **1982**, 104, 1223.
- 91 K. Klarskov, K. Breddam, P. Roepstorff, *Anal. Biochem.* **1989**, 180, 28.
- 92 C. E. Smith, K. L. Duffin in: *Techniques in Protein Chemistry*, R. H. Angeletti (Ed.), Academic Press, San Diego, **1993**.
- 93 Z. Zhang, J. S. McElvain, *Anal. Chem.* **2000**, 72, 2337.
- 94 B. T. Chait, R. Wang, R. C. Beavis, S. B. H. Kent, *Science* **1993**, 262, 89.
- 95 M. Bartlett-Jones, W. A. Jeffery, H. F. Hansen, D. J. Pappin, *Rapid. Commun. Mass Spect.* **1994**, 8, 737.
- 96 B. Thiede, B. Wittmann-Liebold, M. Bienert, E. Krause, *FEBS Lett.* **1995**, 357, 65.
- 97 H. A. Woods, A. Y. C. Huang, R. J. Cotter, G. R. Pasternack, D. M. Pardoli, E. M. Jaffee, *Anal. Biochem.* **1995**, 226, 15.
- 98 D. H. Patterson, G. E. Tarr, F. E. Regnier, S. A. Martin, *Anal. Chem.* **1995**, 67, 3971.
- 99 B. Thiede, J. Salnikov, B. Wittmann-Liebold, *Eur. J. Biochem.* **1997**, 244, 750.
- 100 R. S. Brown, J. J. Lennon, *Anal. Chem.* **1995**, 67, 1998.
- 101 R. M. Whittall, L. Li, *Anal. Chem.* **1995**, 67, 1950.
- 102 B. Halliwell, *FEBS Lett.* **1997**, 411, 157.
- 103 D. Yi, G. A. Smythe, B. C. Blount, M. W. Duncan, *Arch. Biochem. Biophys.* **1997**, 344, 253.
- 104 J. P. Crow, Y. Z. Ye, M. Strong, M. Kirk, S. Barnes, J. S. Beckmann, *J. Neurochem.* **1997**, 69, 1945.
- 105 R. L. Levine, J. A. Williams, E. R. Stadtman, E. Shacter, *Methods Enzymol.* **1994**, 233, 346.
- 106 M. Ahmed, E. Brinkmann Frye, T. P. Degenhardt, S. R. Thorpe, J. W. Baynes, *Biochem. J.* **1997**, 324, 565.
- 107 K. Sugimoto, Y. Nishizawa, S. Horiuchi, S. Yagihashi, *Diabetologia* **1997**, 40, 1380.
- 108 I. Miksik, Z. Deyl, *J. Chromatogr. B* **1997**, 699, 311.
- 109 C.-C. Chao, Y.-S. Ma, E. R. Stadtman, *Proc. Natl. Acad. Sci. USA* **1997**, 94, 2969.
- 110 S. Fu, R. T. Dean, *Biochem. J.* **1997**, 324, 41.
- 111 K. L. Schey, L. Finley, *Acc. Chem. Res.* **2000**, 33, 299.
- 112 R. Zand, M. X. Li, X. Jin, D. Lubman, *Biochemistry* **1998**, 37, 2441.
- 113 S. Capasso, L. Mozzarella, *Peptides* **1998**, 19, 389.
- 114 C. Ramakrishnan, G. N. Ramachandran, *Biophys. J.* **1965**, 5, 909.
- 115 G. N. Ramachandran, V. Sasisekharan, *Adv. Protein. Chem.* **1968**, 23, 283.
- 116 M. Crisma, F. Formaggio, A. Moretto, C. Toniolo, *Biopolymers* **2006**, 84, 3.
- 117 T. M. Weaver, *Protein Sci.* **2000**, 9, 201.
- 118 M. N. Fodje, S. Al-Karadaghi, *Protein Eng.* **2002**, 15, 353.

- 119 D. Obrecht, M. Altorfer, U. Bohdal, J. Daly, W. Huber, A. Labhardt, C. Lehmann, K. Müller, R. Ruffieux, P. Schönholzer, C. Spiegler, C. Zumburn, *Biopolymers* **1997**, 42, 575.
- 120 C. Baldauf, R. Günther, H.-J. Hofmann, *Biopolymers* **2005**, 80, 675.
- 121 D. W. Urry, M. C. Goddall, J. D. Glickson, D. F. Mayers, *Proc. Natl. Acad. Sci. USA* **1971**, 68, 1907.
- 122 K. S. Rotondi, L. M. Gierasch, *Biopolymers (Pept. Sci.)* **2006**, 84, 13.
- 123 C. K. Smith, L. Regan, *Acc. Chem. Res.* **1997**, 30, 153.
- 124 V. Pavone, G. Gaeta, A. Lombardi, F. Nastro, O. Maglio, *Biopolymers* **1996**, 38, 705.
- 125 E. G. Hutchinson, J. M. Thornton, *Protein Sci.* **1994**, 3, 2207.
- 126 K. Möhle, M. Gufmann, H.-J. Hofmann, *J. Comput. Chem.* **1997**, 18, 1415.
- 127 P. Burkhard, J. Stetefeld, S. V. Strelkov, *Trends Cell Biol.* **2001**, 11, 82.
- 128 S. Castano, I. Cornut, K. Büttner, J. L. Dasseux, J. Dufourcq, *Biochim. Biophys. Acta* **1999**, 1416, 161.
- 129 B. Bechinger, *J. Membr. Biol.* **1997**, 156, 197.
- 130 S. Futaki, *Biopolymers* **1998**, 47, 75.
- 131 B. Bechinger, R. Kinder, M. Helmle, T. C. B. Vogt, U. Harzer, S. Schinzel, *Biopolymers* **1999**, 51, 174.
- 132 C. Zhang, C. DeLisi, *Cell. Mol. Life Sci.* **2001**, 58, 72.
- 133 <http://scop.mrc-lmb.cam.ac.uk/scop>.
- 134 <http://www.cathdb.info>.
- 135 K.-C. Chou, *Anal. Biochem.* **2000**, 286, 1.
- 136 T. Lazaridis, M. Karplus, *Curr. Opin. Struct. Biol.* **2000**, 10, 139.
- 137 C. Levinthal, *J. Chim. Phys.* **1968**, 65, 44.
- 138 R. Bonneau, D. Baker, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, 30, 173.
- 139 B. Al-Lazikani, J. Jung, Z. Xiang, B. Honig, *Curr. Opin. Chem. Biol.* **2001**, 5, 51.
- 140 P. Koehl, *Curr. Opin. Struct. Biol.* **2001**, 11, 348.
- 141 B. Rost, *J. Struct. Biol.* **2001**, 134, 204.
- 142 R. A. Meyers, *Proteins: From Analytics to Structural Genomics*, Wiley-VCH, Weinheim, **2006**.
- 143 A. Schmidt, V. S. Lamzin, *Cell. Mol. Life Sci.* **2007**, 64, 1959.
- 144 R. L. Rich, D. G. Myszka, *Curr. Opin. Biotechnol.* **2000**, 11, 54.
- 145 R. Rigler, E. Elson, *Fluorescence Correlation Spectroscopy: Theory and Applications*, Springer, Berlin, **2001**.
- 146 E. Haustein, P. Schwill, *Ann. Rev. Biophys. Biomol. Struct.* **2007**, 36, 151.
- 147 P. L. Privalov, A. I. Dragan, *Biophys. Chem.* **2007**, 126, 16.
- 148 N. Berova, K. Nakanishi, R. W. Woody, *Circular Dichroism: Principles and Applications*, Wiley-VCH, Weinheim, **2000**.
- 149 G. D. Fasman (Ed.) *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York, **1996**.
- 150 J. W. Nelson, N. R. Kallenbach, *Proteins Struct. Funct. Genet.* **1986**, 1, 211.
- 151 P. Wittung-Stafshed, *Biochim. Biophys. Acta* **1998**, 1382, 324.
- 152 T. A. Keiderling, *Curr. Opin. Chem. Biol.* **2002**, 6, 682.
- 153 H. Günzler, H.-U. Greulich, *IR Spectroscopy: An Introduction*, Wiley-VCH, Weinheim, **2002**.
- 154 E. Vass, M. Hollósi, F. Besson, R. Buchet, *Chem: Rev.* **2003**, 103, 1917.
- 155 E. Vass, M. Kurz, R. K. Konat, M. Hollosi, *Spectrochim. Acta A* **1998**, 54, 773.
- 156 E. Vass, E. Lang, J. Samu, Z. Majer, M. Kajtar-Peredy, M. Mak, L. Radics, M. Hollosi, *J. Mol. Struct.* **1998**, 440, 59.
- 157 E. Vass, J. Samu, Z. Majer, M. Hollosi, *J. Mol. Struct.* **1998**, 408, 207.
- 158 G. Yoder, P. Pancoska, T. A. Keiderling, *Biochemistry* **1997**, 36, 15123.
- 159 D. K. Graff, B. Pastrana-Rios, S. Y. Venyaminov, F. G. Prendegast, *J. Am. Chem. Soc.* **1997**, 119, 11282.
- 160 K. Oberg, A. L. Fink, *Anal. Biochem.* **1998**, 256, 92.
- 161 C. Vigano, L. Manciu, F. Buyse, E. Goormaghtigh, J.-M. Ruysschaert, *Biopolymers* **2000**, 55, 373.

- 162 J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, New York, 2007.
- 163 H. Friebohn, *Basic One- Two-Dimensional NMR Spectroscopy*, Wiley-VCH, Weinheim, 2005.
- 164 G. A. Webb, *Nuclear Magnetic Resonance*, A Special Periodical Report, Vol. 34, Royal Society of Chemistry, 2005.
- 165 K. Wüthrich, *J. Biol. NMR* 2003, 27, 13.
- 166 G. T. Montelione, D. Zheng, Y. J. Huang, K. C. Gunsalus, T. Szyperski, *Nature Struct. Biol., Struct. Genomics Suppl.* 2000, 982.
- 167 K. H. Gardner, L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.* 1998, 27, 357.
- 168 K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* 1997, 94, 12386.
- 169 S. J. Archer, D. J. Domaille, E. Laue, *Annu. Rep. Med. Chem.* 1996, 31, 299.
- 170 J. Balbach, V. Forge, N. A. van Nuland, S. L. Winder, P. J. Hore, C. M. Dobson, *Nature Struct. Biol.* 1995, 2, 865.
- 171 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
- 172 D. Croft, J. Kemmink, K.-P. Nedig, H. Oschkinat, *J. Biomol. NMR* 1997, 10, 207.
- 173 T. Diercks, M. Coles, H. Kessler, *Curr. Opin. Chem. Biol.* 2001, 5, 285.
- 174 J. M. Moore, *Biopolymers* 1999, 51, 221.
- 175 C. B. Post, *Curr. Opin. Struct. Biol.* 2003, 13, 581.
- 176 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* 2000, 28, 235.
- 177 <http://www.rcsb.org>.
- 178 E. Abola, P. Kuhn, T. Earnest, R. C. Stevens, *Nature Struct. Biol., Struct. Genomics Suppl.* 2000, 973.
- 179 V. S. Lamzin, A. Perrakis, *Nature Struct. Biol., Struct. Genomics Suppl.* 2000, 978.
- 180 I. L. Karle, *Acc. Chem. Res.* 1999, 32, 693.
- 181 M. P. Brown, C. Royer, *Curr. Opin. Biotechnol.* 1997, 8, 45.
- 182 J. J. Hill, C. A. Royer, *Methods Enzymol.* 1997, 278, 390.
- 183 M. R. Eftink, M. C. R. Shastry, *Methods Enzymol.* 1997, 278, 258.
- 184 M. R. Eftink, *Biochemistry* 1998, 63, 276.
- 185 A. Karlström, A. Unden, *Biopolymers* 1997, 41, 1.
- 186 S.-H. Park, R. T. Raines, *Protein Sci.* 1997, 6, 2344.
- 187 D. L. Andrews, A. A. Demidov, *Resonance Energy Transfer*, Wiley, Chichester, 1999.
- 188 H. R. Stennicke, M. Renatus, M. Meldal, G. S. Salvesen, *Biochem. J.* 2000, 350, 563.
- 189 T. Winkler, U. Kettling, A. Koltermann, M. Eigen, *Proc. Natl. Acad. Sci. USA* 1999, 96, 1375.

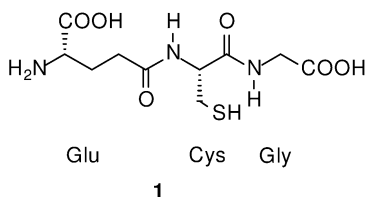
3

Biology of Peptides

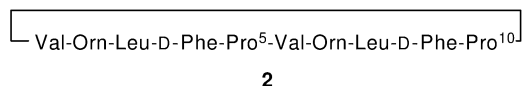
3.1

Historical Aspects and Biological Functions [1–4]

Although Emil Fischer had already predicted at the beginning of the last century that organic chemistry and biology would, in the near future, turn to peptide and protein research, interest in this field was rather poor until the early 1940s. At that time, the number of known biologically attractive peptides was very limited.



In addition to glutathione **1** and carnosine, only gramicidin S and a few other mold-derived compounds (e.g., enniatins) had been discovered. As early as 1888, glutathione (GSH) was observed as a reducing agent in yeast by Rey-Pailhade. It was isolated from liver, yeast and muscle by Hopkins in 1921 and synthesized by Harington and Mead in 1935 [5]. GSH occurs in virtually all cells of animals, most plants and bacteria and is involved in metabolism, transport and cellular protection. It acts as a biological redox agent, as a coenzyme and cofactor, and as a substrate in certain reactions catalyzed by glutathione S-transferase. GSH scavenges free radicals and reduces peroxides, and is important in the lens and in certain parasites lacking catalase to remove H_2O_2 [6]. The synthesis of GSH, and also the synthesis of carnosine by Sifferd and du Vigneaud [7] in the same year, had profited from the invention some years earlier of the benzyloxycarbonyl group as an amino-protecting group. The amino acid sequence of *gramicidin S* **2** was elucidated as early as 1947 as being a cyclic decapeptide with a repeated sequence of two pentapeptides [8].



Gramicidin S primarily acts on Gram-positive bacteria. An understanding of its biosynthesis, which is performed by the gramicidin S-synthetase, an enzyme

complex of two subunits, was achieved at a much later date, however [9]. Interestingly, the well-known naturally occurring dipeptide carnosine, H- β -Ala-His-OH, is involved in a variety of activities related to detoxification of the body from free radical species and the byproducts of membrane lipid peroxidations. Possible anti-aging activities have been described, and appropriate preparations have been sold as dietary or anti-aging cosmetic products – the so-called “elixir of youth”, “miraculous anti-aging supplement”, and anti-aging amino acid supplement”. N $^{\alpha}$ -acetylcarnosine (NAC) is known as the “best-selling” derivative of carnosine. It is stable against the human serum carnosinase which cleaves carnosine into its constituents of β -alanine and histidine. NAC has been developed for ophthalmic application as lubricating eye drops [10, 11].

The importance and broad functional role of peptides in life processes became apparent only in the 1950s and early 1960s, when the continuous development of increasingly sensitive analytical methods and techniques for isolation and purification signaled the start of a new era in this research field. Size-exclusion chromatography [12], chromatography on cellulose-based ion-exchangers [13], countercurrent distribution [14] and other methods developed in various areas of biochemistry complemented the techniques for peptide isolation that had been developed previously. Several important peptide hormones, e.g., insulin [15] and glucagon [16] from the pancreas, corticotropin [17] from the adenohypophysis, α -melanotropin [18] and β -melanotropin [19] from the pars intermedia of the hypophysis or species-dependent (e.g. chicken, whale, porpoise) from the neurohypophysis, oxytocin [20, 21] and vasopressin [22, 23] from the neurohypophysis, and angiotensin [24, 25] from blood plasma were elucidated during the 1950s. The isolation of peptides from natural sources has been – and will remain in the future – a necessity of peptide research, as was stated by the unforgotten Josef Rudinger at the Third American Peptide Symposium in 1972:

“In the work on the smaller biologically active peptides, isolation is to my mind still the most difficult and critical phase and its practitioners have my sincere respect and admiration.”

The deduction of amino acid sequences from the experimentally determined cDNA sequences of peptides that are synthesized ribosomally has not changed the importance of peptide isolation. Unexpected post-translational modifications and uncertainties concerning the cleavage sites of the precursors are, besides other problems, important points to maintain and to improve the methods of peptide isolation from natural sources.

Until now, innumerable peptides with a variety of biological and physiological effects have been detected, isolated, characterized and mostly synthesized. Without doubt, the sequence determination and, especially, the chemical synthesis of oxytocin, performed by the Nobel prize winner Vincent du Vigneaud [26], were a historical milestone in peptide research, since this synthetic hormone had a biological activity indistinguishable from that of the natural hormone isolated from the neurohypophysis.

Peptides and their higher correlates, the proteins, fulfil crucial functions in almost all processes of the living cell. The distinction between what constitutes a peptide and what constitutes a protein should increasingly be only of academic interest, and becomes increasingly unclear, both from the synthetic and structural points of view, as peptides increase in length. It holds true to state that proteins are large peptides. With increasing length, peptides have a greater tendency to exhibit elements of secondary structure, forming preferred solution conformations. Furthermore, the problems connected with the chemical synthesis of proteins in the range of 100 to 150 amino acid residues are basically the same as in the synthesis of peptides or, in a general sense, of peptide chemistry.

Many of these naturally occurring bioactive molecules act as *enzymes* [27–30]. These biocatalysts mediate very precisely and effectively all chemical reactions within and outside the cells. Enzymes differ from ordinary chemical catalysts by having higher reaction rates, greater reaction specificity, milder reaction conditions and, last but not least, the capacity of regulation. The capability of binding their substrates specifically through geometrically and physically complementary interactions permits enzymes to act absolutely stereospecifically in substrate binding and catalysis.

Living organisms require a highly complex chemical signaling system in order to coordinate activities at any level of their organization. Intercellular *biochemical communication* is mainly based on the action of *hormones* [31, 32] and neurotransmitters used by all living organisms to coordinate their activities at any level of their organization. The chemically heterogeneous group of hormones contains a vast number of peptides, proteins, and amino acids. Peptide and protein hormones exert their action via binding to specific receptors. A *receptor* [33] is a protein on the cell membrane or within the cytoplasm or cell nucleus capable of binding its ligand (L) according to the law of mass action ($R + L = [RL]$). The binding parameter of a radiolabeled ligand to a receptor can be determined from a plot of bound ligand (B)/free ligand (F) versus bound ligand ($B/F = (B_{\max} - B)/K_L$). This is known as a Scatchard plot, named after its originator, George Scatchard. K_L is operationally defined as the ligand concentration at which the receptor is half-maximally occupied by ligand, and can be determined from the tangential slope $-1/K_L$. The hormone–receptor interactions stimulate the synthesis and activation of specific enzymes via signaling cascades involving second messengers such as cAMP, cGMP, diacylglycerol (DAG), inositol triphosphate (IP_3), and Ca^{2+} . Many peptide and protein hormone receptors mediate adenylate cyclase activation and inhibition, respectively, via *G protein-coupled receptors* which are known as seven transmembrane receptors (7TM receptors). On the other hand, *receptor tyrosine kinases* detect ligands and propagate signals via tyrosine kinases of intracellular domains. The release of prostaglandins, steroid hormones, thyroid hormones, peptides, and glycoproteins can also be stimulated by peptide and protein hormones. Sometimes, internalization of a receptor is induced by an excess of a peptide hormone which leads to disappearance of the receptor – described as receptor down-regulation. This results in temporary stimulation followed by inhibition of the target cells. The term peptide agonist is used for analogues of native peptide hormones that trigger the hormone

signal in the same manner. Peptide antagonists are analogues that act as competitive inhibitors, occupy the appropriate receptor, displace the agonist from the receptor, but do not transmit the hormone signal.

In the early days, it was essential when working with peptide hormones to avoid degradation by proteolytic enzymes that might be co-extracted from the appropriate tissues. Without doubt, the extraction technique using ethanol containing hydrochloric acid that was first employed in the extraction of insulin from pancreatic tissue was a major breakthrough. Alternatively, in the case of relatively thermostable peptides the proteolytic tissue enzymes could be inactivated by brief boiling and subsequent extraction with cold 0.2 M HCl. The simultaneous isolation of gastrin by Gregory and Tracy in Liverpool and secretin by Jorpes and Mutt in Stockholm was described in the early 1960s. Human secretin was sequenced after its isolation from 181 g of intestinal tissue obtained after surgery. The isolation of a species variant of a known peptide hormone is much easier to perform when guided by radioimmunoassay – a highly sensitive technique introduced by Yalow and Berson [34].

Peptides and proteins are responsible for the regulation of biochemical processes in complex organisms. Figure 3.1 outlines such regulation by peptides. Autocrine hormones (A) act on the same cell that released them, as demonstrated, for example, by the T-cell proliferation stimulating interleukin 2. Paracrine hormones (B), which are also named local mediators, are directed to surrounding cells by diffusion. Endocrine hormones (C) are synthesized and transported to distant

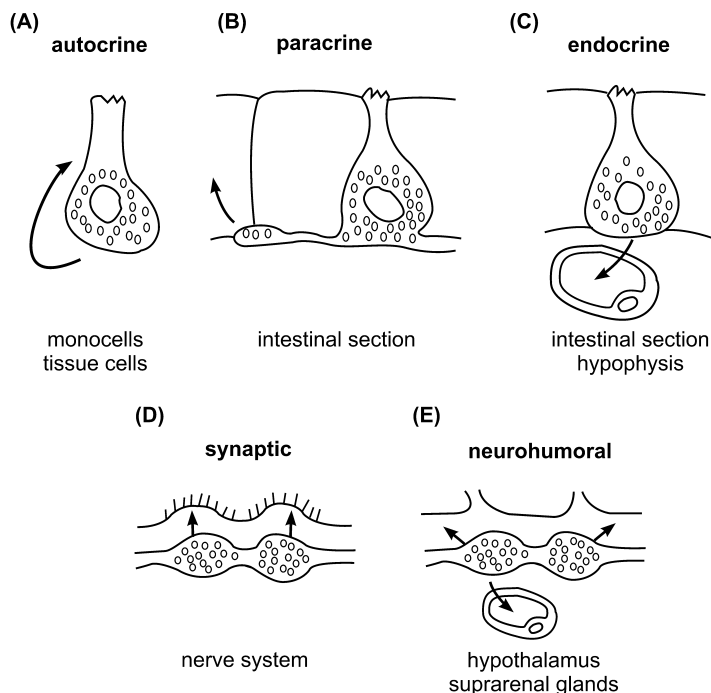


Figure 3.1 Biochemical communication exerted by peptides.

cells via the bloodstream; examples include insulin and glucagon. Furthermore, neurotransmitters (D) – many of which are also peptides (e.g., β -endorphin, enkephalins, somatostatin) – elicit chemically transmitted nerve impulses across most synapses. Neurohormones (E) act as mediators on nerve cells. The brain contains many neuroactive peptides with complicated relationships, though to date only a fraction of these have been discovered [35]. Various peptides originally considered to be brain peptides exerting multiple effects in the central nervous system (CNS) have also been found in the gut, and in other locations. In other words, peptides occurring in one place exert effects in other places. An interactive blood-brain barrier (BBB) helps to regulate the passage of numerous peptides from the periphery to the CNS and vice versa [36]. Based on their lipophilicity and other physico-chemical properties, some peptides are capable of crossing the BBB by simple diffusion, whereas other peptides cross by saturable transport systems. Without doubt, the BBB provides a regulatory system involved in controlling the communication between the CNS and the rest of the body.

Biological membranes [37–39] contain a large number of proteins, such as integral proteins with nonpolar surface regions which allow hydrophobic association with the bilayer core, and peripheral proteins which bind to integral proteins on the membrane surface by polar interactions. For example, membrane proteins are involved in controlling the permeability of the membrane and supporting the transport of solutes across it against thermodynamic gradients. Due to the presence of these nonpolar cores, biological membranes are highly impermeable to most ionic and polar molecules; hence, the transport of these substances requires the action of specific transport proteins. For the transport of various ions (e.g., Na^+ , K^+ , Ca^{2+} , Cl^- , and metabolites, for example amino acids, pyruvate, nucleotides, and sugars), specific transmembrane transport proteins are required. Channel-forming ionophores, for example gramicidin A, For-Val-Gly-Ala-D-Leu-Ala-D-Val-Val-D-Val-Trp-D-Leu¹⁰-Trp-D-Leu-Trp-D-Leu-Trp-NH-CH₂-CH₂-OH, permit the passage of protons and alkali cations, but this channel is blocked by Ca^{2+} . The transmembrane channel is formed by dimerization of gramicidin A in a head-to-head fashion (Figure 3.2).

Porins, originally discovered in the outer membrane of Gram-positive bacteria, are transmembrane proteins consisting of trimers of identical subunits, and each subunit

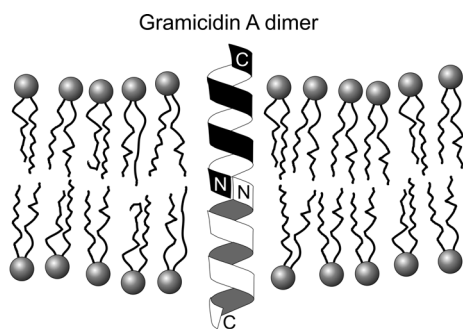


Figure 3.2 Schematic representation of a transmembrane channel formed by a gramicidin A dimer.

forms a 16-stranded antiparallel β -barrel forming a solvent-accessible channel along its barrel axis. They form aqueous channels used for the passive diffusion of water, ions and small molecules. Porins occur in bacterial cell walls, and also in fungal, plants, mammalian and other vertebrate cell membranes [40]. Special carrier proteins transport smaller metabolic intermediates from one location to another. Electrons generated (e.g., by oxidation of $\text{NADH} + \text{H}^+$ and FADH_2 in the citric acid cycle) pass through four protein complexes, termed the electron-transport chain. The free energy released by the electron transport is conserved by the generation of an electrochemical proton gradient across the inner mitochondrial membrane. The energy stored in the proton gradient is utilized by the proton-pumping F_1F_0 -ATPase, an oligomeric protein, in the synthesis of ATP. The tetrameric protein hemoglobin is a very important oxygen delivery system which transports the required amount of oxygen from the lungs, gills, or skin to the tissue for use in respiration [41].

Very specialized proteins are involved in the *immune system* used by vertebrates to defend themselves against other organisms [42–45]. Cellular immunity which targets parasites, fungi, virally infected cells, and foreign tissue is mediated by T lymphocytes or T cells, whereas humoral immunity directed against bacterial infections, and the extracellular phases of viral infections is mediated by a huge and diverse collection of antibodies or immunoglobulins. The latter are produced by B lymphocytes or B cells. Antibodies are glycoproteins consisting of two identical light (L) chains and two identical heavy (H) chains. There are two types of light chain ($L = \kappa$ or λ), each of which can be associated with any of five types of heavy chains ($\alpha, \delta, \epsilon, \gamma, \mu$). Secreted human immunoglobulins (Ig) consist of the following classes, in which the Greek letter designates the type of heavy chain corresponding to the class of Ig [IgA: $(L_2\alpha_2)_{n=1-3}$, IgD: $L_2\delta_2$, IgE: $L_2\epsilon_2$, IgM: $(L_2\mu_2)_5$, IgG: $L_2\gamma_2$]. Four IgG subclasses (IgG₁₋₄) exist, differing in their γ chains. Each heavy chain contains an *N*-linked oligosaccharide. The four subunits of the Ig associate by disulfide bonds as well as by noncovalent interactions, thereby forming a Y-shaped symmetric dimer, as confirmed by electron microscopy and X-ray diffraction.

As shown schematically in Figure 3.3, each light (L) chain and heavy (H) chain consists of a variable (V) region (V_L and V_H) and the remaining constant region. The two identical F_{ab} segments form, with their terminal variable regions, the two antigen-binding sites of the antibody. Papain cleaves the IgG molecule in its hinge region, yielding two F_{ab} fragments and one F_c fragment. A virtually unlimited variety of antigen-binding sites is generated from the immune system by somatic recombination and somatic mutation. This generation of antibody diversity provides a shield against almost any antigen attacking the organism.

An antigenic determinant that is commonly (but not very precisely) called an epitope characterizes the portion of an antigen that makes contact with a particular antibody or T-cell receptor. There is a fundamental difference between epitopes recognized by T cells and those recognized by antibodies or B cells. T cells recognize protein antigens in a fragmented form on the surface of antigen-presenting cells. The cellular immune response begins when a macrophage engulfs and partially digests a foreign antigen and then displays the resulting antigenic fragments on its surface.

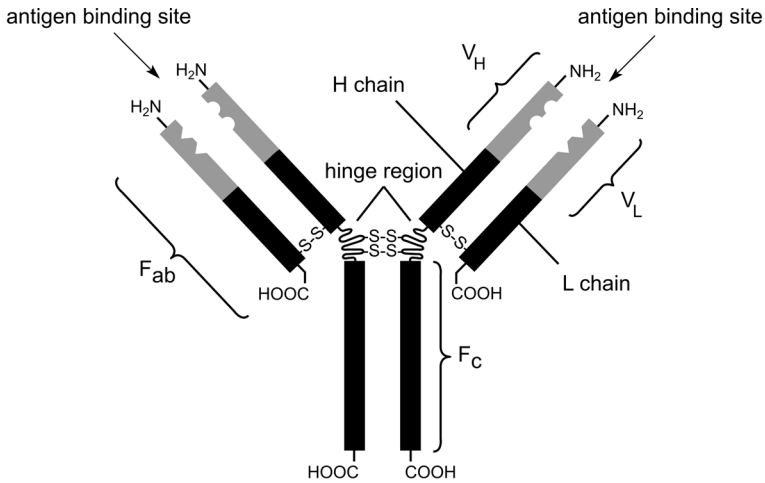


Figure 3.3 Schematic diagram of IgA antibody molecule.

MHC proteins (MHC molecules) (Figure 3.4) are surface glycoproteins acting as markers of individuality; they are the antigen-presenting markers that allow the immune system to distinguish body cells from invading antigens and cells of the immune system from other cells, respectively [46–48]. They are membrane-bound

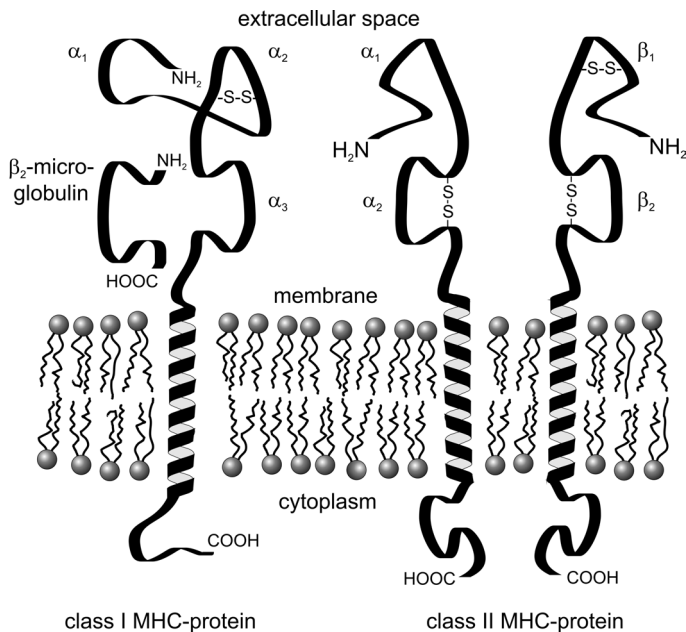


Figure 3.4 Major histocompatibility complex (MHC) proteins.

proteins encoded by the major histocompatibility complex (MHC; *histo* refers to tissue). With the Class I MHC proteins and the Class II MHC proteins, two principal classes of MHC molecules were recognized long before the native function was understood. Class I MHC molecules (~ 350 aa; $M_r \sim 44$ kDa) are folded into five domains: a C-terminal cytoplasmic domain (~ 30 aa), a transmembrane domain (~ 40 aa), and three external domains (~ 90 aa each, designated α_3 , α_2 , α_1). The Class I MHC proteins are also invariably associated (1:1) with an extracellular, non-glycosylated small protein, termed β_2 -microglobulin (β_2m), which does not span the membrane, and is separately encoded by a gene of a different chromosome. The α_3 domain and β_2m , the closest located to the membrane, are both homologous to an immunoglobulin domain. Class II MHC proteins are heterodimeric transmembrane glycoproteins and consist of an α chain ($M_r \sim 33$ kDa) and a β chain ($M_r \sim 28$ kDa). Each (~ 230 aa) of the two chains forms two conserved immunoglobulin-like domains (α_1 , α_2 and β_1 , β_2 , respectively). Class I MHC proteins occur on most cells, whereas Class II MHC molecules are found on macrophages and B lymphocytes. The peptide fragments are presented to the T-cell receptor either via a Class I or Class II MHC molecule. As a rule, Class I MHC molecules present peptides to CD8 (predominantly cytotoxic) T cells, and Class II MHC molecules to CD4⁺ (mainly helper) T cells.

The *complement system* serves as an essential biological defense system. It is directed against foreign invaders by eliminating foreign cells via complement fixation – that means killing foreign cells by binding and lyzing their cell membranes, by inducing the phagocytosis of foreign particles (opsonization), and triggering local acute inflammatory processes. The complement system consists of approximately 20 plasma proteins and is characterized by two related activation pathways: the antibody-dependent classical pathway and the antibody-independent alternative pathway. Similar to the blood clotting system, both pathways mainly comprise the sequential activation of a series of serine proteases. Each activated protease acts only on the next component of the cascade. Most of the complement protein names bear the capital letter “C” followed by a component number, and active proteases are indicated by a bar over the component number [49]. The classical pathway consists of the recognition unit (C1q, C1r, C1s) assembling on cell surface-bound antibody–antigen complexes, the activation unit (C2, C3, C4) amplifying the recognition event via a proteolytic cascade, and the membrane attack unit (C5–C9) that punctures the antibody-marked plasma membrane of the cell; this results in cell lysis and death. The alternative pathway is thought to provide the initial response to the invasion of microorganisms since it is activated in the absence of antibodies. It uses a couple of the same components as the classical pathway. The alternative pathway leads to the assembly of the membrane attack unit, but it proceeds via a series of reactions which are activated by whole bacteria, polymers of bacterial origin, and virus-infected host cells.

Blood clotting (blood coagulation) involves a double cascade of proteolytic reactions resulting in the formation of a gel (blood clot, thrombus) which is sufficiently dense to prevent bleeding from a wound [49]. In this bifurcated cascade, nearly 20 different blood coagulation factors are involved. Most of these substances are plasma glycoproteins which are synthesized in the liver. The human blood coagulation factors are,

historically, designated by Roman numerals, and the activated form is indicated by the subscript “a”, e.g., XII_a for the activated form of factor XII. With the exception of fibrin, the active clotting factors are serine proteases. The human blood coagulation factors are listed in Table 3.1. The primary stages of the human blood clotting cascade are formally divided into the so-called intrinsic and extrinsic pathways. In the intrinsic pathway, the protein components are contained in the blood, whereas in the extrinsic pathway one of its important factors (VII) occurs in the tissues. The Stuart factor (X) may be activated by factor IX_a (a product of the intrinsic pathway) or by factor VII_a (a product of the extrinsic pathway). The initiation of blood clotting in the intrinsic pathway is performed by the contact system in which XII (in the presence of high-molecular weight kininogen, HMK) is activated by adsorption to a negatively charged surface such as glass or kaolin *in vitro*. Collagen and platelet membranes cause the same effect *in vivo*. In the last two steps, which require a phospholipid membrane surface and Ca²⁺, factor XI_a activates factor IX, and the latter activates factor X in the presence of factor VIII_a. The start of the extrinsic pathway is characterized by proconvertin (VII) activation either by factor XII_a or thrombin. The VII_a formed promotes the activation of X in the presence of Ca²⁺, phospholipid membrane, and tissue factor III.

Fibrous proteins are responsible for the mechanical properties of bone, hair, skin, horn, tendon, muscle, feather, tooth, and nails. Keratin consists of a group of fibrous proteins occurring in wool, hair, hooves, claws, horns and feathers, and also in skin, connective tissues and intermediate filaments [50–52]. α -Keratins occur in mammals, having around 30 various variants, whereas β -keratins are found in birds, reptiles, and in the silk of insects and arachnids. α -Keratin forms closely associated pairs of α -helices in which each pair is composed of a type I and a type II keratin chain twisted in parallel into a left-handed coil. α -Keratins are subdivided into “hard” or “soft” keratins, depending on whether they have high or low Cys contents. Hard keratins occur, for example, in hair, horn and nail, whereas soft keratins are found in skin and callus. The elasticity of wool fibers and hair is based on the coiled coil’s tendency to untwist after stretching, and to recover its original conformation after relaxing the external force. *Silk fibroin* ($M_r \sim 365$ kDa) is an important β -keratin, in which the sequence $-(\text{Gly-Ser-Gly-Ala-Gly-Ala})_n-$ is repeated many times [53]. The polypeptide chains form antiparallel β -pleated sheets in which the peptide chains extend parallel to the fiber axis. Skin in higher animals contains an extensive network of *intermediate filaments* (IF) made of keratin-forming protein fibers 100–150 Å in diameter.

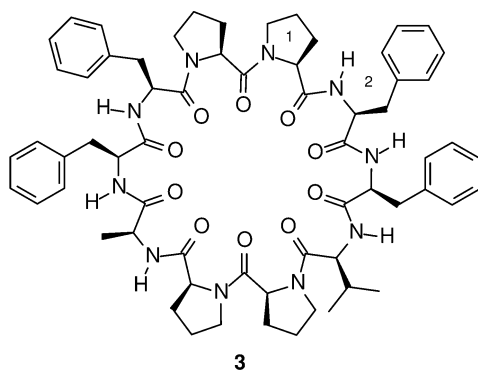
Collagen is the major component of the natural extracellular matrix, playing a central role in its organization and mechanical properties. It occurs as the important stress-bearing component of connective tissues such as bone, cartilage, teeth, ligament, tendon, and the fibrous matrices of skin and blood vessels [54]. Collagen is composed of three polypeptide chains. Within the types of collagen, the most abundant is *type I collagen* ($M_r \sim 285$ kDa; width: ~ 14 Å; length: ~ 3000 Å) characterized by the chain composition $[\alpha 1(I)]_2\alpha 2(I)$, and mainly occurring in bone, tendon, skin, blood, cornea, and blood vessels. Further important members are *type II collagen* $[\alpha 1(II)]_3$, which is found in cartilage and intervertebral disk, and *type III collagen*

Table 3.1 Blood coagulation factors.

No.	Common name	M_r (kDa)	Properties and functions
I	Fibrinogen	~340	(A α) ₂ (B β) ₂ γ ₂ ; A and B represent the N-terminal 16- and 14-residue fibrinopeptides which are cleaved by thrombin forming the fibrin monomer $\alpha_2\beta_2\gamma_2$
II	Prothrombin (II _a : Thrombin)	~72	Zymogen of thrombin; glycoprotein with 579 aa, several disulfide bonds; three domain structure (N-terminal 40-residue Gla-domain with 10 Gla residues, and two ~115 residue kringle domains); Factor X _a cleaves between Arg ²⁷¹ -Thr ²⁷² and Arg ³²⁰ -Ile ³²¹ releasing the N-terminal propeptide 1–271 and thrombin connected with the A and B peptides linked by one of the disulfide bonds
III	Tissue factor or thromboplastin	~30	Membrane glycoprotein occurring in many tissues; involved in the intrinsic pathway
IV	Calcium ions		Promote the binding of the factors IX, X, VII and II to acidic phospholipids of cell membranes where activation is performed. Stabilizing functions for I, V, and other factors during activation, and for the subunit dissociation of XIII
V	Proaccelerin (V _a : accelerin)	~350	V _a promotes the binding of X and II to platelets, where the activation of X is performed, II is converted to II _a
VII	Proconvertin	~50	Involved in the extrinsic pathway; VII _a mediates the activation of X significantly in the presence of Ca ²⁺ , phospholipid membrane and III; VII _a activates also IX
VIII	Antihemophilic factor	~265	Activated VII acts as an accessory factor during the activation of X by activated Christmas factor (IX _a); VII forms a complex with the von Willebrandt factor during circulation in blood
IX	Christmas factor	~56	Single chain Gla-containing glycoprotein; IX _a activated by XI _a and VII _a , respectively, activates factor X proteolytically, and X _a cleaves II in the preceeding step of the clotting cascade
X	Stuart factor	~55	X may be activated via either the intrinsic pathway (IX _a + VIII _a + Ca ²⁺) or the extrinsic pathway (VII _a + III + Ca ²⁺)
XI	Plasma thromboplastin antecedent (PTA)	~136	Two chain glycoprotein joined by disulfide bond(s); XI belongs to the four glycoproteins of the so-called contact system
XII	Hageman factor	~75	Single chain glycoprotein acting as the first factor in the intrinsic pathway; XII is activated by plasmin, kallikrein, and the high molecular weight kininogen (HMK); XII belongs to proteins of the contact system
XIII	Fibrin-stabilizing factor (FSF)	~300/ ~160	FSF is a transamidase present both in platelets and plasma; platelet FSF consists of two a chains (~75 kD), whereas plasma FSF has two additional b chains (~88 kD); XIII _a cross-links the rather fragile soft clots to the more stable hard clots
	Prekallikrein	~69	Zymogen of the serine protease kallikrein which activates the Hageman factor (XII)
	High molecular weight kininogen (HMK)	~70	Activation yields a kinin that is involved in the activation of XII; HMK belongs to the proteins of the contact system

$[\alpha 1(\text{III})]_3$, which occurs in blood vessels and fetal skin. Mutations of type I collagen usually result in *osteogenesis imperfecta* (brittle bone disease). *Elastin* [55], a structural protein with rubber-like elastic properties, is the main component of the elastic yellow connective tissue occurring, e.g., in the lungs and aorta. Elastin forms a three-dimensional network of fiber cross-linked by allysine aldol, lysinonorleucine, desmosine, and isodesmosine. It has an unanticipated regulatory function during arterial development, controlling the proliferation of smooth muscle and stabilizing arterial structure [56].

Toxic peptides and proteins are used by various species for defense purposes, or are employed in struggles for limited resources. Several venoms (e.g., of bees, wasps, and snakes) contain well-characterized peptides and proteins. Both in Europe and in the United States, the notorious toadstool *Amanita phalloides* is responsible for 95% of the casualties occurring after ingestion of poisonous fungus. After World War II, Theodor Wieland and coworkers were engaged in the elucidation of the constituents of the poisonous *Amanita* mushrooms. At the very end of these research activities, the formulae and three-dimensional structures of the toxic constituents of the green death cap *Amanita phalloides* and of the destroying angels, *Amanita virosa* and their relatives, could be characterized as peptides and were named phallotoxins and amatoxins [57]. Interestingly, *antamanide* **3**, which is a cyclic decapeptide of *A. phalloides*, protects experimental animals from intoxication by *Amanita*-derived phalloidin.



Frog skin has been used for medical purposes for centuries, and indeed is still used today in some South American countries. Amphibian skin is a rich source of various bioactive compounds, including various peptides which are produced by holocrine-type serous glands in the integument. The peptides are stored as granules in the lumen of the cells and are released upon specific stimulation. They are involved in the regulation of physiological functions of the skin, or in defense against predators or microorganisms. A systematic study of the occurrence of bioactive peptides in more than 100 amphibian species was performed in the early 1970s by Vittorio Ersparmer [58], and led to the discovery of a huge number of peptides with a range of pharmacological activities. Normally, peptides originating from ribosome-mediated biosynthesis contain only L-amino acids, whereas peptides with D-amino acids are very rare. The first example of a D-amino acid-containing peptide discovered in an

animal was *dermorphin* 4, which was isolated from the skin of the South American frog *Phyllomedusa sauvagei* by Erspamer and his colleagues in 1981. Dermorphin is released from a biosynthetic precursor containing several copies of dermorphin with the L-isomer only. The conversion of L-Ala into D-Ala occurs post-translationally. Dermorphin binds preferably to the μ -receptors. With [Lys⁷]dermorphin, [Trp⁴,Asn⁷]dermorphin, and [Trp⁴,Asn⁷](dermorphin-(1–5)), three additional dermorphin analogues were found in the skin of *Phyllomedusa bicolor*. The dermorphins are potent analgesics in rodents and primates, including humans [59]. Further examples of neuroactive peptides are *achatin-I* (H-Gly-D-Phe-Ala-Asp-OH) from the ganglia of the African giant snail *Achatina fulica*, and *fulicin* (H-Phe-D-Asn-Glu-Phe-Val-NH₂) which was isolated from the ganglia of the of *A. fulica*. The sequences of nine *fulicin gene-related peptides* (FGRP-1 to -9) were predicted from the cDNA encoding the ganglia fulicin precursor, and the transcripts were detectable in the heart [60].

H-Tyr-D-Ala-Phe-Gly-Tyr⁵-Pro-Ser-NH₂

4

During the past decades, over 500 peptides have been detected in microorganisms, with *Bacillus* sp. and *Actinomyces* sp. the main sources. In contrast, only a few plants and animals are yet known as sources of these types of peptides.

Peptide antibiotics [61–63] consist of a chemically heterogeneous group of peptides produced by microorganisms (bacteria and fungi) that kill or inhibit the growth of other microorganisms. Originally, they were classified, related to their chemical structure, into homomeric and heteromeric peptide antibiotics as well as into linear and cyclic peptides. Furthermore, according to their mode of biosynthesis, the hundreds of known peptide antibiotics can be classified into non-ribosomally synthesized peptide antibiotics (e.g., bacitracins, polymyxins, gramicidins) and ribosomally synthesized compounds (e.g., lantibiotics [64] and related peptides). The first totally chemical synthesis of a natural lanthionine peptide – nisin – was described by Tetsuo Shiba and coworkers in 1987 [65].

Nonribosomally synthesized peptides are very often drastically modified, and are largely produced by bacteria. Ribosomally synthesized peptide antibiotics act as a major component of the natural host defense molecules of the producing species. In 1962, Kiss and Michl [66] noted for the first time the occurrence of antimicrobial and hemolytic peptides in the skin secretions of *Bombina variegata*, and this led to the discovery of the antimicrobial peptide bombinin [67]. A database of antimicrobial peptides from amphibian skin [68] is available at: <http://www.bbcm.univ.trieste.it/~tossi/search.htm>.

During the late 1980s, the *magainins* were isolated from the African clawed frog *Xenopus laevis* by Michael Zasloff [69]. These antimicrobial peptides, which are obtained from various amphibian species, represent the best-studied class of peptide antibiotics [70]. They are considered as effector molecules of innate immunity acting as the first line of defense against bacterial infections [71, 72]. Some of the most studied mammalian peptides are the *defensins* [73]. The α -defensins (also known as “classical defensins”) and β -defensins are predominantly β -sheet structures

stabilized by three disulfide bonds, and contain a high arginine content [74]. The human α -defensins, HNP-1 (30 aa), HNP-2 (29 aa), and HNP-3 (30 aa) are constituents of the microbicidal granulae of neutrophils [75]. *Thionins* were the first antimicrobial peptides to be isolated from plants [76] and were found to be toxic against both Gram-positive and Gram-negative bacteria, yeast, fungi and various mammalian cell types [77]. Plant defensins consisting of 27 to 84 amino acid residues have eight disulfide-linked cysteines comprising a triple-stranded antiparallel β -sheet structure with only one α -helix. They have a high antifungal activity [78]. *Cecropins* [79] are insect-derived linear peptides found in the hemolymph of the giant silk moth (*Hyalopora cecropia*), and form α -helices in solution. These positively charged peptides form voltage-dependent ion channels in planar lipid membranes, but are not lethal to mammalian cells at microbicidal levels.

Prebiotic peptides should be mentioned since, during the course of chemical evolution, they were produced ahead of all other oligomer precursors of biomolecules [80, 81]. The formation of peptides under primordial Earth conditions have been simulated experimentally to investigate the real chances for the formation of precursor molecules under such conditions. It is well known that the formation of peptides from amino acids in aqueous solution is a thermodynamically and kinetically unfavorable process. Various condensation reactions have been proposed which can be categorized into melt processes, heterogeneous systems involving mineral catalysis, and condensation reactions induced by various reagents in the homogeneous phase, respectively. In particular, “salt-induced peptide formation” (SIPF) seems to be a very simple way of obtaining peptides in aqueous solution, and the possible combination of SIPF with clay mineral catalysis might provide insight into the chemical evolution of peptides on the primitive Earth. Other work has provided evidence in favor of an important role of α -amino acid *N*-carboxyanhydrides as activated peptide monomers and likely as early free energy carriers.

3.2

Biosynthesis

3.2.1

Ribosomal Peptide Synthesis

Although the formation of a peptide bond is known to be a relatively simple chemical reaction, the biosynthesis of polypeptides is a very complex process. Ribosomal peptide synthesis involves deoxyribonucleic acid (DNA) encoding genetic information, and two different types of ribonucleic acids which convey (as messenger RNA, mRNA) the genetic information from the nucleus to the ribosome, and carry (as transfer RNA, tRNA) the specific amino acids appended enzymatically to the site of peptide bond formation of the ribosome [82]. The sequence of amino acid residues for each polypeptide is like a blueprint stored in the encoding DNA of the genes in the chromosomes [83]. Each of the 22 proteinogenic amino acids in a polypeptide is encoded by a triad of codon bases in the gene [84, 85]. The 21st proteinogenic amino

acid *selenocysteine* (Sec, U) [86] is ribosomally incorporated into proteins by a cognate aminoacyl-transfer RNA pair, Sec-tRNA^{Sec}, via read-through of an internal, in-frame UGA codon, which is normally the *opal* stop codon.

While in eukaryotes most tRNA are directly loaded with the corresponding amino acid, in bacteria and archaea some of them (Asn-tRNA, Gln-tRNA, and Cys-tRNA) are first loaded with an amino acid precursor, which is then converted to the cognate amino acid. Asn-tRNA and Gln-tRNA are formed in most prokaryotes from Asp-tRNA or Glu-tRNA, respectively. Such an indirect pathway prevails for Sec-tRNA in all organisms. Cys-tRNA synthesis in methanogens and Sec-tRNA biosynthesis proceeds via Ser-tRNA and O-phosphoseryl-tRNA [87]. Unlike selenocysteine, the 22nd genetically encoded amino acid, *pyrrolysine* (Pyl, O) is charged directly onto a dedicated tRNA by a cognate aminoacyl-tRNA synthetase [88, 89].

In the first step of ribosomal synthesis, named transcription, mRNA is synthesized enzymatically under the direction of the DNA template by copying the encoded sequence of deoxyribonucleotides onto a molecule of mRNA with the complementary sequence of ribonucleotides. RNA polymerase initiates transcription on the antisense strand of a gene at a position designed by its promoter. Although prokaryotic mRNA transcripts do not require additional processing, eukaryotic mRNAs bear an enzymatically appended 5' cap and very often an enzymatically generated poly(A) tail. In addition, the introns of primary eukaryotic mRNA transcripts are removed by splicing mechanisms.

Amino acid activation is the first step in the aminoacylation process of tRNA, catalyzed by aminoacyl-tRNA synthetases (aatRS). An amino acid reacts with adenosine triphosphate (ATP) under elimination of pyrophosphate to yield a mixed anhydride, the aminoacyl adenylate, which normally remains tightly bound to the enzyme (Figure 3.5A). In the second step, the highly activated aminoacyl moiety is transferred to the appropriate tRNA, thereby forming the aminoacyl-tRNA and liberating adenosine monophosphate (AMP) (Figure 3.5B). The reaction is driven to completion by the pyrophosphatase-catalyzed hydrolysis of inorganic pyrophosphate generated in the first step. At least one specific tRNA and one aminoacyl-tRNA synthetase exist for every amino acid. Two classes of these enzymes are known, each containing ten members. The accurate translation process requires two exact recognition steps. The first step is the choice of the correct amino acid for the covalent attachment to a tRNA, this being catalyzed by the appropriate aatRS. The second step is recognition of the amino acid-loaded tRNA, as specified by the mRNA sequence. A proofreading or editing step by the appropriate aatRS enhances the fidelity of tRNA loading with its cognate amino acid at the expense of ATP hydrolysis. The tRNA vary in length from 60 to 95 ribonucleotides ($M_r \sim 18\text{--}28$ kDa) and contain up to 25% of modified bases. In all tRNA the acceptor group for the amino acid is terminated by the sequence CCA with a free 3'-OH group. In aminoacyl-tRNA the amino acid is esterified to the 2'- or 3'-OH group of their 3'-terminal ribose moiety.

In the translation process, the appropriate tRNA is selected only through codon-anticodon interactions, without any participation of the aminoacyl group. The tRNA can recognize several degenerate codons through wobble base pairing, according to

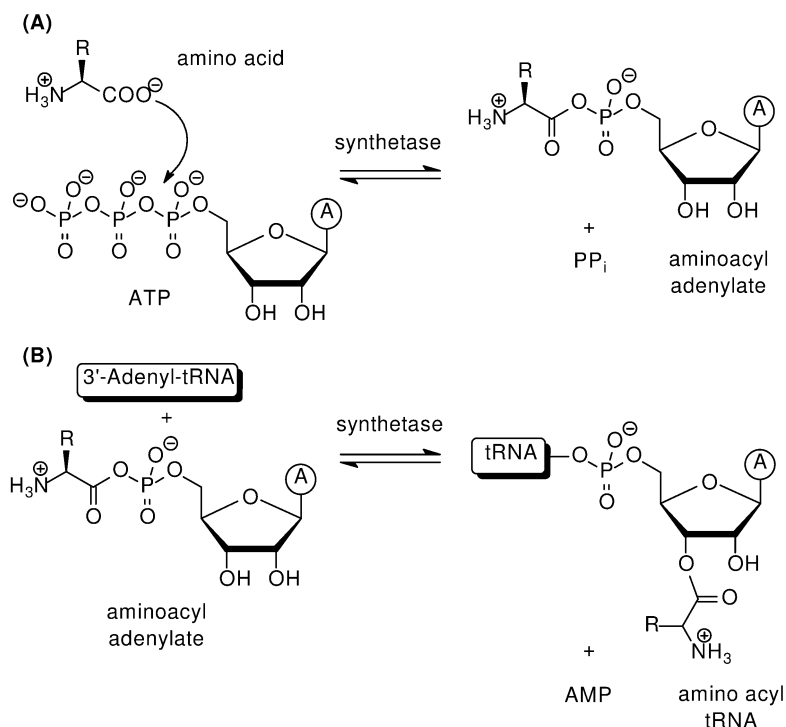


Figure 3.5 Two-step process of tRNA aminoacylation catalyzed by aminoacyl-tRNA synthetases (aatRS).

the wobble hypothesis suggested by Crick. He proposed that the first two codon-anticodon pairs show normal Watson-Crick geometry, whereas the third codon-anticodon pairing allows limited conformation adjustments in its pairing geometry based on a small amount of flexibility or “wobble” in the third codon position.

Ribosomal peptide synthesis proceeds stepwise from the N- to the C-terminus by reading the mRNAs in the 5' → 3' direction. The ribosome consists of a small and a large subunit, and contains (e.g., in the simple *E. coli* version) three rRNA molecules and 52 protein molecules. Eukaryotic ribosomes are, however, larger and more complex than prokaryotic ribosomes. The ribosomes that are active in protein synthesis are tandemly arranged on the mRNA like beads on a string, named polyribosomes or polysomes.

In prokaryotic peptide synthesis, N^α-formylmethionine (fMet) is the N-terminal residue in the chain initiation process. The appropriate tRNA_f^{Met} differs from tRNA_m^{Met} which bears internal methionine residues. Initiation is a complex three-stage process requiring the two ribosomal subunits, fMet – tRNA_f^{Met}, and the initiation factors IF-1, IF-2, and IF-3. The ribosome has various tRNA binding sites. The P site, which normally binds the peptidyl-tRNA, is, in the initiation step, occupied with fMet – tRNA_f^{Met} whereas the A site binds the incoming aminoacyl-tRNA.

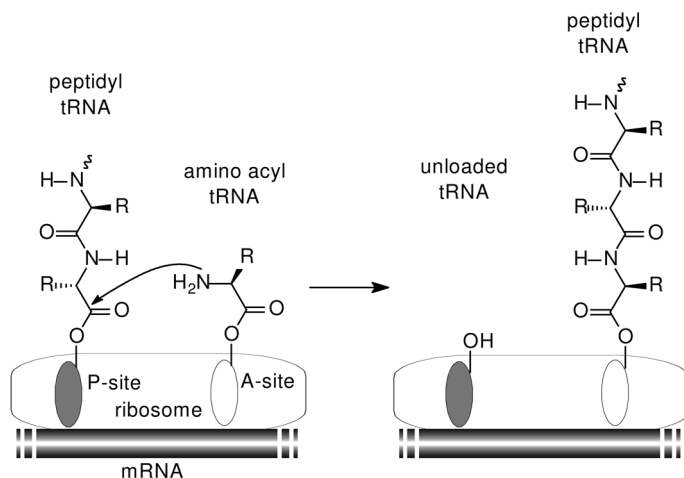


Figure 3.6 Schematic view of the ribosomal peptide bond formation catalyzed by the peptidyltransferase center.

Furthermore, the ribosome contains a third t-RNA-binding site, the E site (derived from exit), which temporarily binds the outgoing tRNA.

Chain elongation is a three-stage process adding amino acid residues step by step, starting from the carboxy group of fMet, at a rate of up to 40 residues per second. Three elongation factors (EF-Tu, EF-Ts, and EF-G) and GTP are involved in the elongation cycle. Peptide bond formation occurs in the second stage of this cycle by nucleophilic attack on the ester moiety of the peptidyl-tRNA in the P site by the amino group of the aminoacyl-tRNA in the A site (Figure 3.6).

This process does not require ATP, and is catalyzed by the *peptidyl transferase center* (PTC) [90, 91]. PTC is a ribozyme catalyzing peptide bond formation. It resides in the large ribosomal subunit on the 23S rRNA and catalyzes both peptide bond formation and peptide release. The ribosome-catalyzed reaction occurs with a speed of $\sim 15\text{--}50$ peptide bonds per second. Peptide bond formation on the ribosome proceeds via the substrate-assisted proton shuttle mechanism, in which the proton shuttle occurs via the 2'-hydroxy of A76 of the P-site tRNA, and is assisted by neighbouring ribosomal RNA and water molecules. In this mechanism no chemical catalysis by proteins is involved, instead it involves conformational and environmental changes of the active site [92–95].

The resulting unloaded P site tRNA is then transferred to the E site (not shown in Figure 3.6) and, simultaneously, in the so-called translocation process the peptidyl-tRNA from the A site is transferred together with its bound mRNA to the P site, thus preparing the ribosome for the next elongation cycle.

Chain termination in *E. coli* starts when the termination codons UAA, UGA, or UAG are recognized by the release factors RF-1 (UAA and UAG) and RF-2 (UAA and UGA). In addition, RF-3, which is a GTP-binding protein, stimulates, together with GTP, the binding process of the other two releasing factors. As a result of the binding process the peptidyl transferase center hydrolyzes the ester bond of the

peptidyl-tRNA, and the synthesized polypeptide is released. The resulting unloaded tRNA and the release factors dissociate from the ribosome with concomitant hydrolysis of the RF-3-bound GTP to GDP and P_i . The translation process in eukaryotes resembles that in prokaryotes, but differs from it in certain details.

3.2.2

Post-Translational Modification

This approach is an essential prerequisite for mature polypeptides and proteins, and the most important and fundamental process of proteomic complexity [96]. Proteins can be modified post-translationally by covalent attachment of one or more of several classes of molecules, by the formation of intra- or intermolecular linkages, by proteolytic processing of a newly synthesized polypeptide chain, or by any combination of these events. Post-translational covalent modifications occurring in Nature include glycosylation, phosphorylation, lipidation, acylation, methylation, hydroxylation, sulfation, iodination, carboxylation, nucleotidylation, ADP-ribosylation, pyroglutamyl formation, amino acid modification, and numerous other types.

Normally, the translation product cannot be considered as a functional protein. After the assembly of the complete sequence of a protein, some of the amino acid building blocks may be involved in post-translational modifications. Most modifications are performed after release of the polypeptide from the ribosome, but modifications such as disulfide bridge formation or N-terminal acetylation very often occur in the nascent polypeptide chain. Enzymes catalyzing processing reactions are mainly located in the endoplasmic reticulum (ER) and Golgi apparatus; among these are enzymes which catalyze disulfide bridge formation, iodination, and glycosylation. Peptide chain folding is stabilized mainly by noncovalent interactions, and in many cases the association of subunits forming oligomeric proteins is an important event after translation. In general, many proteins are modified by limited proteolysis and/or by derivatization of specific amino acid residues.

3.2.2.1 Enzymatic Cleavage of Peptide Bonds

Limited proteolytic cleavage by specific peptidases is a common process in post-translational modification. Although it seems to be a waste of cellular resources, proteolytic reactions belong to the most important types in the maturation of proteins.

Proteolytic removal of the N-terminal fMet or Met (in eukaryotic proteins) residues most likely occurs in all synthesized polypeptides shortly after their release from the ribosome. Only a few mature proteins contain the formyl or even the N-terminal methionine residue.

Eukaryotic membrane-bound ribosomes synthesize transmembrane proteins and proteins that are destined for secretion. The question arises as to how these proteins are differentiated from soluble and mitochondrial proteins assembled by free ribosomes. According to the *signal hypothesis*, established by the Nobel laureate Günter Blobel, all secreted proteins are synthesized with N-terminal signal peptide sequences of 13 to 36 predominantly hydrophobic amino acid residues. Immediately after the signal peptide sequence of the preprotein enters the lumen of the rough

endoplasmic reticulum (RER), it is specifically cleaved by a membrane-bound signal peptidase, thereby releasing the signal peptide (pre-sequence). Likewise, many proteins are synthesized as inactive precursors – termed proproteins – that must be activated by limited proteolysis that cleaves the so-called propeptide sequence. Normally, the initial biosynthesis product is termed a prepro-protein. The activation of proenzymes (zymogens) of the complement system and proteolysis in the cascade system of blood coagulation should be mentioned in this context. Some polypeptides are synthesized as segments of so-called polyproteins containing the sequences of two or more polypeptides. This applies to polypeptide hormones, ubiquitin, and virus proteins (such as those causing AIDS or poliomyelitis):

- Ubiquitin is synthesized as several tandem repeats, termed polyubiquitin.
- Proopiomelanocortin (POMC) is a prototype of a polyprotein with a cleavage pattern that varies among different tissues, yielding a different set of peptide hormones.
- The hormone insulin is derived from an inactive single-chain, 84-residue precursor termed proinsulin. Correct formation of the disulfide bonds is performed efficiently *in vivo* by assembly of the longer chain of proinsulin containing the A and B chains, together with an internal segment known as the C chain. The active hormone is formed after proteolytic excision of the C chain from proinsulin.

Another example is the biosynthesis of collagen, a fibrous triple-helix protein that forms the major extracellular component of connective tissue. N- and C-terminal sequences, each containing about 100 residues, are constituents of the polypeptide chains of procollagen. These propeptide sequences direct the formation of the collagen triple helix, and they are removed by amino- and carboxyprocollagen peptidases after folding of the triple helix. In addition, collagen assembly requires chemical modification of Pro and Lys residues, and this will be discussed below.

3.2.2.2 Hydroxylation

The assembly of collagen in the course of collagen biosynthesis is a very good example of protein maturation by post-translational modification. When the nascent procollagen moves into the RER of the fibroblasts, the sequence-specific enzymes prolyl hydroxylase and lysyl hydroxylase catalyze the regioselective hydroxylation of proline and lysine residues to give 4-hydroxy- or 3-hydroxyproline (Hyp) and 5-hydroxylysine (Hyl), respectively. Furthermore, glycosyl transferases catalyze the attachment of sugar residues to Hyl residues. These modifications occur prior to triple helix formation, because the hydroxylases and glycosyl transferases involved do not act on the helical structure.

Besides the examples mentioned above, polypeptides and proteins of biological interest often include structural elements exceeding the 22 genetically encoded amino acid residues. Both the side-chain functionalities and the terminal amino and carboxy groups of proteins are covalently modified. More than 150 different types are known. Thirteen of the encoded amino acids contain functional groups at the side chain as possible modification sites.

3.2.2.3 Carboxylation

Carboxylation of glutamate building blocks in certain proteins to give γ -carboxyglutamate (Gla) residues is catalyzed by vitamin K-dependent carboxylases. Some proenzymes in the blood-clotting cascade undergo this type of modification. Human prothrombin contains in its Gla domain sequence 6–32 no fewer than 10 Gla residues. The synthesis of Gla from Glu, known as the liver reaction cycle [97], requires vitamin K hydroquinone as an essential cofactor that is transformed into vitamin K-2,3-epoxide during Gla synthesis. This epoxide is converted to vitamin K in two sequential reactions. Proteins containing Gla residues involved in Ca^{2+} binding have also been detected in other tissues following the discovery of Gla residues in clotting factors.

3.2.2.4 Glycosylation

Glycosylation of proteins is more abundant than all other types of post-translational modifications taken together. Co- and posttranslational modification has been estimated to occur with more than half of all natural proteins. Glycoproteins contain covalently linked carbohydrate moieties in proportions between <1 and >70 mass%, and can be found in both soluble and membrane-bound forms in all cells, and also in the extracellular matrix and in extracellular liquids. The carbohydrate modification of glycoproteins can exert a wide range of biological functions, e.g., as hormone activities, embryogenesis, neuronal development, proliferation of cells and their organization into specific tissues, inflammatory reactions, and immune surveillance. Dramatic changes in cell-surface carbohydrates are often connected with tumor progression.

Glycoproteins are characterized by microheterogeneity with respect to the carbohydrate portion. They occur naturally in a variety of forms (glycoforms [98]) that share the same sequence, but differ in both the nature and site of glycosylation. These microheterogeneous mixtures complicate the determination of the exact function in structure–activity relationships.

Especially in eukaryotic cells, many secreted and membrane-associated proteins, as well as those located inside membranous organelles, are glycosylated. Protein glycosylation plays an important role in biological processes [99–103], for example in protein folding, cellular differentiation, cell–cell communication, and the sliminess of mucosa. The hydrophilic carbohydrate moieties exert considerable influence on the structure, polarity, and solubility of the proteins. The addition of large glycan structures to the protein backbone may dramatically alter its conformation and consequently influence the protein functions. Furthermore, it may substantially affect the thermal stability and targeting/clearance properties of the glycosylated protein. Co- or post-translational modification of proteins with oligosaccharide moieties may also change the folding pathway of the polypeptide chain by interactions between the protein and the oligosaccharide part.

The higher steric requirements, shielding of potential cleavage sites, and charged oligosaccharide groups confer higher stability towards proteolytic degradation of the protein. While native coerulein has a half-life of 54 h, the enzymatic degradation of the terminal N-acetylneuraminic acid with neuraminidase reduces the half-life to 5 min [104]. A galactose-recognizing receptor system (asialoglycoprotein receptor) is responsible for rapid degradation in the liver. This receptor system

recognizes the terminal galactose residue exposed by cleavage of N-acetylneuraminic acid and initiates lysosomal degradation. By contrast, ribosomal enzymes in the Golgi complex are addressed for transport into the lysosomes by the attachment of mannose-6-phosphate units to the terminus of the oligosaccharide chains [105].

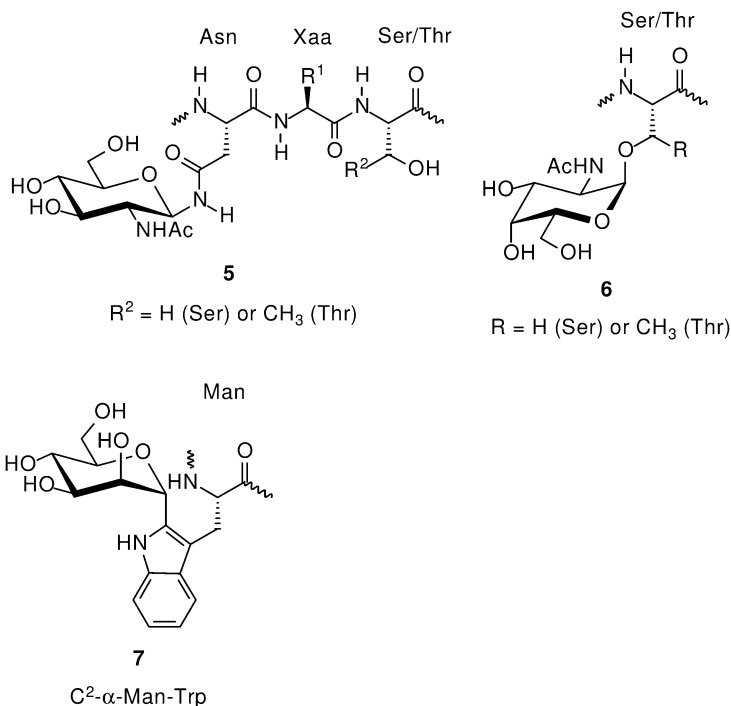
The carbohydrate proportion of glycoproteins in tumor cells is often changed significantly in comparison with regular cells. The investigation of so-called tumor-associated antigens provides promising starting points for diagnostics and immune therapy [106].

Carbohydrate moieties on cell surfaces can be considered as molecular codes. Although often only seven or eight monosaccharide units occur in a single glycosyl residue of a mammalian cell [107], the multifunctionality of these monomers results in an unusually high number of possible complex structures. Considering the possibility of branched derivatives, the stereochemistry of the glycosidic bond, together with modifications of hydroxylic and amino groups of the protein, millions of different structures are possible even for tetrasaccharides. These many different ways of encoding information may all lead to different biological functions.

On the cell surface many carbohydrate groups of glycoconjugates are involved in different types of biochemical recognition processes responsible for growth, development, infection, immune response, cell adhesion, formation of metastases, and different events of signal transduction. Viral invasion also may involve membrane-bound glycoproteins. Erroneous glycosylation is thought to be associated with autoimmune diseases, infectious diseases, or cancer. Such compounds often occur in minute concentrations and are very difficult to isolate, to characterize, and to synthesize. The carbohydrate portions of glycoproteins are of major importance in the realm of cell-cell and cell-matrix recognition. Cells utilize glycoproteins for encoding information on the protein-folding pathway, the localization (intracellular organelles, cell surface, or protein export), or on the recognition of other proteins. Carbohydrate-mediated cell adhesion is also of major biological importance and may be induced either by tissue damage or infections [108]. The interaction of the glycoprotein E-selectin (endothelial leukocyte adhesion molecule, ELAM-1) and an oligosaccharide on the surface of neutrophilic cells represents such an adhesion process. E-selectin is expressed on the surface of endothelial cells and recognizes the tetrasaccharide sialyl-Lewis^X (Sle^X) that forms the terminus of a glycolipid on the surface of neutrophilic cells [109, 110]. Cell adhesion is initiated by cytokines or other inflammatory compounds (e.g., leukotrienes, toxins, lipopolysaccharides) that induce the formation of E-selectin. The interaction between E-selectin on the endothelial cell and the complementary receptor on leukocytes present in the bloodstream leads to retardation of the leukocyte in the so-called rolling process on the surface of the endothelium. In addition, adhesion between the leukocytes and the endothelial cells is mediated by interaction of integrins present on the leukocyte and proteins such as intercellular adhesion molecule 1 (ICAM-1) containing an RGD sequence (-Arg-Gly-Asp-), or the protein VCAM-1 containing the recognition sequences LDV (-Leu-Asp-Val-) or IDSP (-Ile-Asp-Ser-Pro-). As a consequence of these interactions, the leukocyte is attached to the endothelium, undergoes extravasation, and resolves the inflammatory process.

The tetrasaccharide sialyl-Lewis^X is also found on the surface of several tumor cells. Cancer cells clearly utilize this type of adhesion for metastasizing processes, and spread with the bloodstream throughout the body. Synthetic sialyl-Lewis^X derivatives compete in solution with leukocytes for binding of E-selectin. As a consequence, these derivatives may act as antiadhesive agents and are regarded as potential antitumor or anti-inflammatory agents.

In eukaryotes, the protein portion of glycoproteins is synthesized ribosomally, and the attachment of the carbohydrate moieties occurs either co- or post-translationally. The majority of all naturally occurring glycoproteins can be classified either as N-glycosides or as O-glycosides. In the first case N-acetylglucosamine is linked via a β -glycosidic bond to the amide side chain of an asparagine residue **5** or, in O-glycosides, the saccharide residues are linked to the hydroxy groups of serine, threonine (**6**) or tyrosine.



Some unusual glycoproteins have recently been described such as O-glycosides of hydroxylysine and hydroxyproline and even the C-glycoside of tryptophan **7** [111].

At least three different types of N-glycosides can be distinguished: complex, high-mannose, and hybrid (Figure 3.7).

These types are all characterized by a Man₃GlcNAc₂ core sequence attached to the protein via an N-glycosidic bond to an Asn residue within the sequence Asn-Xaa-Ser/Thr, where Xaa is any amino acid residue except Pro (and sometimes Asp) [112]. On occasion, C⁶ of the first GlcNAc residue is glycosylated with α -L-fucose.

The biosynthesis of N-linked oligosaccharides [113] occurs in the ER and Golgi complex and begins in the ER with the multistep formation of a lipid-linked

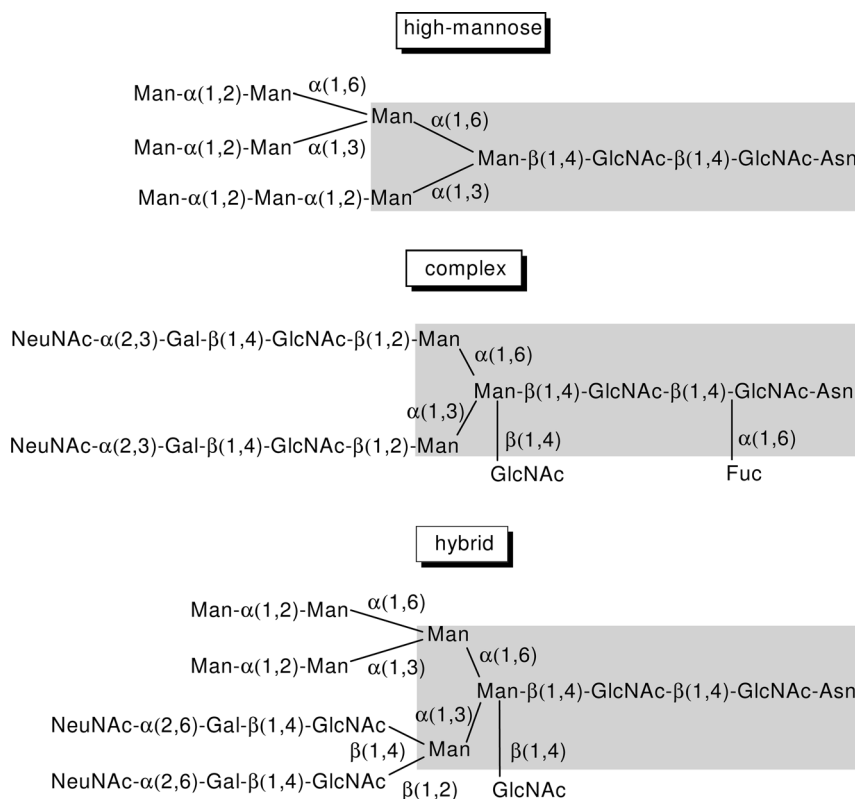


Figure 3.7 Structural motifs of N-glycopeptides.

precursor that consists of dolichol pyrophosphate linked to a common 14-residue core oligosaccharide of the composition (N-acetylglucosamine)₂(mannose)₉(glucose)₃. Each of the 14 monosaccharide units is added stepwise to the dolichol pyrophosphate carrier, catalyzed by a unique glycosyltransferase. The saccharide moiety is then transferred to an Asn residue of the growing peptide chain. Processing begins in the ER by enzymatic removal (“trimming”) of three glucose residues and one mannose residue. The glycoprotein is then transported in membranous vesicles to the Golgi apparatus in order for processing to be continued. Trimming of further mannose residues followed by attachment of various other monosaccharide units finally provides the typical primary structures of N-linked oligosaccharides, as shown in Figure 3.7.

The most common form of O-linked glycoproteins results from the modification of the side chain of serine and threonine by N-acetyl- α -D-galactosamine [114]. The synthesis of these glycans occurs in the Golgi complex where the GalNAc moiety of UDP-GalNAc is transferred to the hydroxy function of serine or threonine under catalysis of the polypeptide GalNAc transferase. O-linked glycans are synthesized in a

highly diverse manner by the O-glycosylation process. Each monosaccharide has three or four attachment sites for linkage of other sugar residues, linked by α - or β -glycosidic bonds, and allows the formation of branched glycan structures. Glycoproteins that are modified by dense clusters of O-glycosylation are termed mucins. Seven different types of O-linked glycans have been found in humans [115]:

1. The mucin-type O-glycans with the motif GalNAc α 1-Ser/Thr present in eight different core structures occur in secreted proteins and proteins of the plasma membrane.
2. Glycosaminoglycans (GAG) with the motif GlcA β (1,3)Gal β (1,3)Gal β (1,4)Xyl β 1-Ser are present in proteoglycans.
3. β -O-linked N-acetyl glucosamine (GlcNAc β 1-Ser/Thr) is observed in nuclear and cytoplasmic proteins and appears to regulate protein stability, sub-cellular localization, and protein-protein interactions.
4. β -O-linked galactose occurs attached to hydroxylysine (Hyl) in collagens.
5. α -O-linked mannose is found as a less common modification in glycoproteins in the brain, nerves, and skeletal muscle. The oligosaccharide NeuAc α (2,3)Gal β (1,4)GlcNAc β (1,2)Man α 1-Ser/Thr is found in α -dystroglycan, a skeletal muscle extracellular matrix protein.
6. β -O-linked glucosyl residues (Glc β 1-Ser), often with α -connected Xyl α (1,3)Xyl residues in position 3, occur in EGF protein domains.
7. α -O-linked fucosyl saccharides (Fuc α 1-Ser/Thr) often with β -connected NeuAc α (2,6)Gal β (1,4)GlcNAc residues in position 3 have also been found in EGF protein domains.

All O-glycans are classified according to the first sugar attached to a Ser, Thr, or hLys residue of the protein component. The most common form in humans is the mucin-type O-glycan bearing GalNAc at the reducing end. The eight mucin-type core structures are compiled in Figure 3.8 [111].

The carbohydrate moiety of mucins can also function as a ligand for cell adhesion events. A well-studied example is the homing of leukocytes to the site of inflammation, an event dependent on the interaction of mucin oligosaccharide moieties with selectins [116]. Mucins of tumor cells display striking alterations in glycosylation. Aberrations of glycosylation are often accompanied with tumor progression and prognosis [117].

As well as GlcNAc and GalNAc, proteins can also be modified by fucosides, mannosides, xylosides, and glucosides. For example, fucosides linked to Ser and Thr have been found in various proteins, e.g., tissue-type plasminogen activator, and several clotting factors. A novel structural class of glycoproteins has been found in human RNase 2, where an α -D-mannopyranose is connected to the C-2 of a Trp residue through a C-glycosidic linkage like in 7 [118]. Furthermore, glycoproteins modified with an oligosaccharide linked to Ser or Thr via a phosphodiester linkage comprise a new type of protein modification which is termed protein phosphoglycosylation [119].

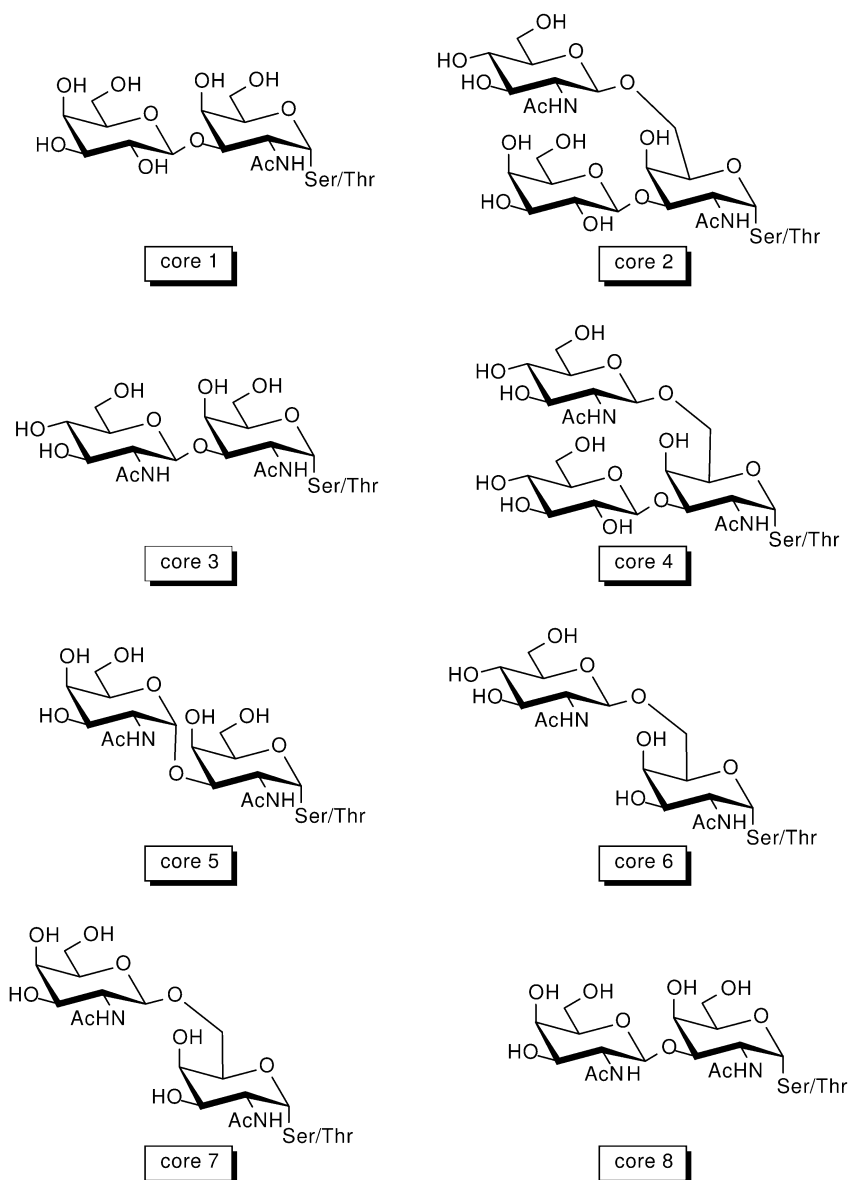


Figure 3.8 Structural core motifs of mucin-type O-linked glycans.

The urgent need for alternative sources of homogeneous glycoproteins is a challenge for chemical glycoprotein and glycopeptide synthesis (see Section 6.3).

3.2.2.5 Amidation

Amidation of bioactive peptides is performed using the bifunctional enzyme peptidylglycine α -amidating monooxygenase (PAM, EC 1.14.17.3) by N-oxidative

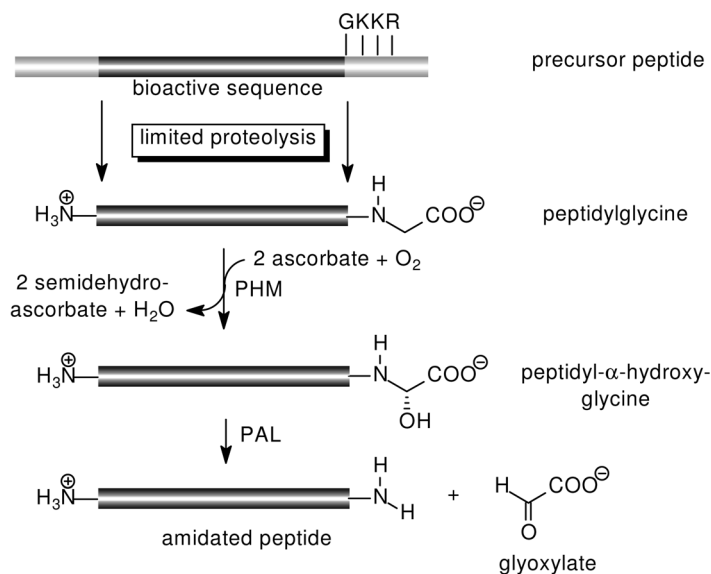


Figure 3.9 Amidation reaction catalyzed by the bifunctional enzyme peptidylglycine α -hydroxylating monooxygenase (PHM); PAL = peptidyl- α -hydroxyglycine α -amidating lyase; PHM = peptidylglycine α -hydroxylating monooxygenase.

cleavage of a glycine-extended precursor [120]. The N-terminal domain of PAM is the ascorbate-dependent peptidylglycine α -hydroxylating monooxygenase (PHM) that catalyzes the stereospecific hydroxylation of the glycine C $^\alpha$ in peptidylglycine substrates, whereas the peptidyl- α -hydroxyglycine α -amidating lyase (PAL) is responsible for generation of the α -amidated peptide product and glyoxylate (Figure 3.9). The isolation of peptide amides with all 20 amino acid amides underlines the broad substrate specificity of PAM.

3.2.2.6 Phosphorylation

Nature uses the phosphorylation and dephosphorylation of serine, threonine, and tyrosine residues in polypeptides and proteins as a universal mechanism in the regulation of many cellular processes, for example in the activation or deactivation of enzymes in signal transduction and protein trafficking. Protein phosphorylation in prokaryotic systems mediates chemosensing of extracellular signals which eventually leads to chemotaxis.

Protein kinases and protein phosphatases, respectively, effect the phosphorylation and dephosphorylation of serine, threonine, and tyrosine residues present in proteins. All organisms utilize this phenomenon for the regulation of a variety of intracellular processes. Protein-tyrosine kinases (PTK) convert an extracellular signal (binding of a ligand to a membrane-bound receptor) into an intracellular signal (phosphorylation of tyrosine residues situated on specific proteins) [121–123].

Covalent protein phosphorylation is important for many purposes, including control of the cell cycle, regulation of cell growth and proliferation, regulation of transcription, and signal transduction. Furthermore, the biological functions of many receptors for hormones, neurotransmitters, peptide-derived drugs, and growth factors are known to be regulated by reversible phosphorylation. The phosphorylation of a single Ser residue is sometimes sufficient to regulate the enzyme activity, a classical example being the control of glycogen synthase and glycogen phosphorylase. These enzymes are regulated by phosphorylation/dephosphorylation integrated in amplifying cascades controlled by the hormones insulin, glucagon, and epinephrine, and also Ca^{2+} . Other phosphoproteins include casein, ovalbumin, phosphovitin, and vitellin.

Phosphorylated tyrosine residues (pTyr) serve, for example, as recognition sites for binding by other signaling proteins. These are often assembled in a modular fashion like the Src homology 2 (SH2) domain or phosphotyrosine-binding domains (PTB). Both of these display high affinity towards ligands containing pTyr. SH2 domains are clusters of approximately 100 amino acids binding to peptides, phosphorylated on tyrosine residues, with an affinity in the range of 10–100 nM [124]. In principle, two categories of SH2 domains can be distinguished according to the recognition motif:

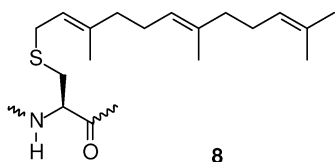
1. Class 1 SH2 domains recognize sequences of the type pTyr-Xaa-Xaa-Ile/Pro (Xaa: hydrophilic amino acids).
2. Class 2 SH2 domains bind with high affinity to the motif pTyr-Yaa-Zaa-Yaa (Yaa: hydrophobic amino acid, Zaa: any amino acid).

pTyr phosphate groups are removed by protein tyrosine phosphatases (PTPase). Consequently, physiological or pathological events in this type of signaling cascade may involve either phosphorylation (protein tyrosine kinases), recognition and binding of phosphorylated proteins (SH2 and PTB domains), or dephosphorylation (protein tyrosine phosphatases). All three phenomena may be addressed in order to influence possible pathological events [125]. While tyrosine phosphorylation only represents a minor subset of protein phosphorylations of a cell, it profoundly influences signal transduction and, hence, cellular responses such as differentiation, mitogenesis, migration, and survival [126]. Inappropriate tyrosine phosphorylation is closely connected with pathological phenomena, including oncogenesis. While only 0.03% of protein-bound phosphate is found on tyrosine residues in normal cells, this percentage is significantly (10- to 20-fold) increased in some malignant tumors.

3.2.2.7 Lipidation

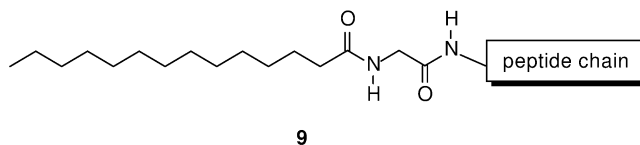
Lipidation of proteins is performed by the enzyme-catalyzed covalent attachment of different types of lipid groups by, for example, prenylation, acylation with fatty acids, and glycosylphosphatidylinositol groups. Lipid-modified proteins are very often attached to cell membranes, with the lipid moiety anchoring the protein to membranes and mediating protein–protein interactions. In many cases these lipid-linked proteins are involved in the transduction of extracellular signals across the plasma membrane and into the nucleus.

Prenylation of polypeptides and proteins with polyisoprenoids belongs to the functionally most important post-translational modification [127–129]. S-Farnesylation and S-geranylgeranylation – the two different types of prenylation – are, in principle, similar in their behavior, but they differ in specificity with regard to the C-terminal sequence of proteins. Prenylation can be performed via a covalent attachment of either a C₁₅ farnesyl or C₂₀ geranylgeranyl moiety to the cysteine residue of the CAAX motif present in the protein to be modified (A: aliphatic amino acid, X: variable residue). Farnesylation, catalyzed by the farnesyltransferase (FTase), is preferred for proteins with the CAAX motif where X is either M, S, Q or A, whereas proteins with X = L are geranylgeranylated under the catalysis of the geranylgeranyltransferase (GGTase-I). FTase catalyzes transfer of the C₁₅ farnesyl moiety from farnesyl pyrophosphate (FPP) to a protein containing the C-terminal CVLS motif yielding the farnesylated protein **8**.



When the prenyl moiety has been attached to the Cys residue of the protein via a thioether linkage, the VLS tripeptide is proteolytically cleaved and a methyl ester is formed by a protein methyltransferase at the new C-terminal Cys residue. This modification pattern was identified in Ras proteins [130]. Ras proteins are plasma membrane-bound GTP-binding proteins that influence numerous signal transduction processes and play a crucial role as a molecular switch involved in tumor formation [131].

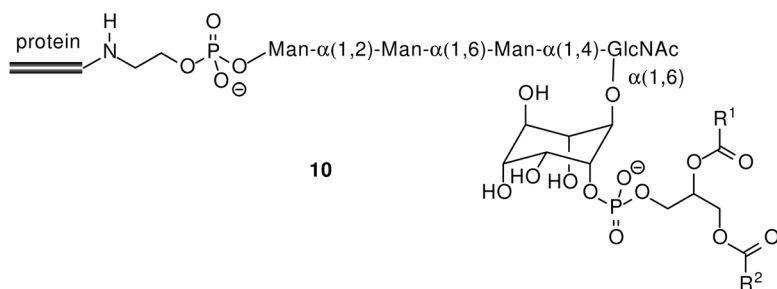
Acylation with fatty acids is performed either co- or post-translationally in various protein families that cover important functional properties [132]. The acyl moieties are linked either via thioesters or via amide bonds. N-Acylation has been found either in the form of myristoyl (C₁₄) groups linked to an N-terminal glycine of a protein **9** [133] or, on occasion, as palmitoyl moieties attached to lysine ϵ -amino groups.



N-Terminal acetylation via common biochemical acetylating agents leads to N-acetylated peptides such as α -melanotropin (α -MSH), or some proteins. In histone H4, the N-terminal Ser is N-acetylated and may be O-phosphorylated. Some other Lys residues in positions 5, 8 and 12 of this protein are N $^{\epsilon}$ -acetylated. Interestingly, the side-chain amino function Lys²⁰ is either mono- or dimethylated. N $^{\epsilon}$ -Palmitoyl-lysine has been found in adenylate cyclase from *Bordetella pertussis*, where it is formed

enzymatically by a specific enzyme [134]. Palmitoyl (C_{16}) groups are most commonly linked to specific cysteine residues in protein S-acylation [135], this modification occurring post-translationally in the cytosol. Palmitoylated proteins are occasionally also prenylated. Thioester-linked palmitic acid may undergo acylation/deacylation cycles due to the labile nature of the thioester bond. In addition, only a few O-acylated peptides or proteins have been elucidated [136].

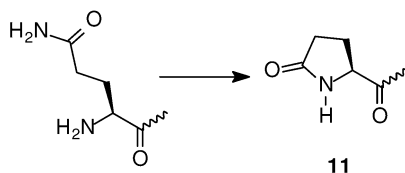
In GPI-linked proteins **10** the glycosylphosphatidylinositol moieties anchor the protein to the exterior surface of the eukaryotic plasma membrane.



GPI-linked proteins are associated with, for example, enzymes, recognition antigens, receptors, and immune system proteins. The glycan is a core tetrasaccharide that varies with the identity of the protein. In addition, diversity in the fatty acid residues (R^1 and R^2) must be stated.

3.2.2.8 Pyroglutamyl Formation

Pyroglutamic acid (<Glu, pGlu, Pyr or Glp) formation via intramolecular cyclization of N-terminal Gln residues is another example of post-translational modification. The aminolytic cyclization of an N-terminal glutamine ($R = NH_2$) in peptides irreversibly gives terminated pyroglutamyl peptides **11**. The acyltransferase *glutaminyl cyclase* (QC, EC 2.3.2.5) is responsible for N-terminal pyroglutamate formation from glutaminyl precursors in peptides and proteins. The first QC was isolated from the latex of *Carica papaya* in 1963 [137]. Later, it was established that glutaminyl cyclases occur in both animal and plant sources. They are abundant in mammalian neuroendocrine tissues, such as hypothalamus and pituitary, and are highly conserved from yeast to human. From the crystal structure of human glutaminyl cyclase it follows that a single zinc ion in the active site is coordinated to three conserved residues and one water molecule [138]. Naturally occurring pyroglutamyl peptides include eisenin (<Glu-Gln-Ala-OH), pelvetin (<Glu-Gln-Gln-OH) from algae, or the liberins thyroliberin and gonadoliberin as well as other peptides including physalaemin and neurotensin.



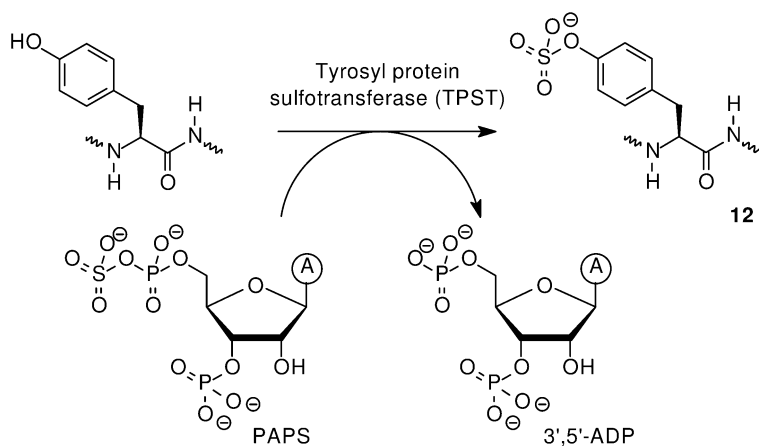


Figure 3.10 Tyrosyl protein sulfotransferase (TPST) catalyzed reaction.

3.2.2.9 Sulfation

Sulfation of tyrosine residues **12** (Figure 3.10) is performed in cholecystokinin, gastrin, and related peptide hormones.

Protein tyrosine sulfation is emerging as a widespread post-translational modification in multicellular eukaryotes. It is one of the most common modifications in secreted and transmembrane proteins and a key modulator of extracellular protein–protein interactions. Evidence suggests up to 1% of all tyrosine residues of the total protein content in an organism can be sulfated [139]. Proteins that contain tyrosine residues sulfated at the phenolic hydroxy function participate in protein–protein interactions, mediated by recognition of the sulfate group. Today, proteins of different types and different modes of action are known to be sulfated at tyrosine residues, among them α -choriogonadotropin, heparin cofactor II, and gastrin II. Secretory proteins are the most abundant, but not exclusive, *in vivo* substrates for tyrosine sulfation [140]. An O-sulfated tyrosine of factor VIII is required in blood coagulation for efficient association with the von Willebrand factor. Tyrosine sulfation of the P-selectin glycoprotein ligand-1 (PSGL-1) mediates high-affinity binding to P-selectin, which is crucial for leukocyte adhesion, while tyrosine sulfation of chemokine receptors is required for optimal interaction with chemokine ligands in leukocyte chemotaxis. Hence, tyrosine sulfation is clearly one of the key modulators mediating inflammatory response. Tyrosine sulfation does not seem to be a comparable system of dynamic regulation like tyrosine phosphorylation. While tyrosine phosphorylation regulates the activity of cytosolic and nuclear proteins, sulfation modulates interactions between extracellular proteins, e.g. between secreted proteins and transmembrane proteins [139–142]. However, the functional significance of tyrosine sulfation remains the subject of many ongoing investigations. Moreover, tyrosine sulfation of the HIV-1 co-receptor CCR5 is a precondition for viral entry into host cells and tyrosine sulfation plays a central role in erythrocyte invasion by the malaria parasite *Plasmodium vivax* [141, 142].

While tyrosine phosphorylation is supposed to be reversible and dynamic because of the interplay of kinases and phosphatases, tyrosine sulfation is still not well understood. The modification is introduced by tyrosyl protein sulfotransferases (TPST) in the Golgi apparatus using 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as the sulfate donor (Figure 3.10) [139, 141]. Different sulfatases have been shown to cleave sulfate esters, e.g. in sulfated carbohydrates, steroids and proteins. They contain a C^α -formylglycine residue (see Section 3.2.2.10) in the active site, which is crucial for the enzyme activity. Two alternative cleavage mechanisms have been suggested, both involve the ability of a carbonyl group to form a *gem*-diol (Figure 3.11) [143–145].

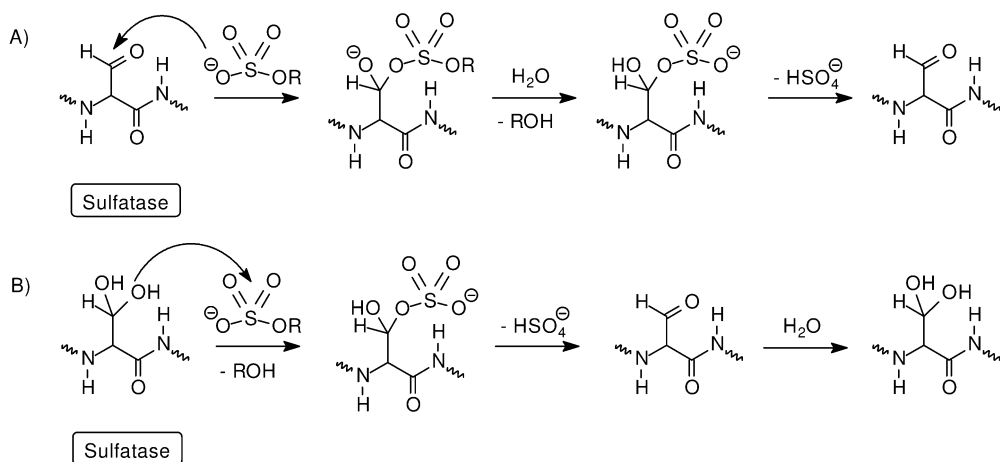


Figure 3.11 Mechanism of sulfatase action by (A) addition-hydrolysis or (B) transesterification-elimination.

3.2.2.10 Further Post-Translational Modifications

As mentioned above, a cysteine residue, strictly conserved in the active site of sulfatases in prokaryotic and eukaryotic sulfatases, is post-translationally modified into a C^α -formylglycine. The oxidation occurs in the endoplasmic reticulum before the nascent polypeptide is folded. The aldehyde is part of the catalytic site and is likely to act as an aldehyde hydrate (Figure 3.11) [143, 144, 146].

Arginine methylation in proteins is a post-translational modification involved in various cellular processes like signaling, RNA processing, transcription, and subcellular transport. There are three main forms of methylated arginine identified in eukaryotes: N^ω -monomethylarginine, asymmetric $N^{\omega,\omega'}$ -dimethylarginine, and symmetric $N^{\omega,\omega'}$ -dimethylarginine [147, 148].

Lysine side chain N-acetylation is a post-translational modification common to histone proteins. It neutralizes the electrostatic interaction between unmodified lysine in histones and the negatively charged DNA phosphate backbone, thus weakening the histone–DNA association in chromatin and activating transcription

[149]. Lysine side chain N-acetylation has also been proposed as a molecular switch in the tumor suppressor protein p53, a transcription factor that regulates the cell cycle. Acetylation of p53 promotes transcriptional activation of target genes [150].

N^{ϵ} -(Carboxymethyl)lysine belongs to the advanced glycation end products (AGE). AGE-modified proteins are formed upon reaction between lysine and the aldehyde group of glucose (Amadori rearrangement). They originate from a Maillard reaction followed by oxidative cleavage (Figure 2.6).

Specific reaction of the thiol groups of the cysteine side chain with the tripeptide glutathione (Section 3.1) under formation of a disulfide-bridged conjugate is termed S-glutathionylation. This post-translational modification may be reverted by small redox proteins such as sulfiredoxin and glutaredoxin. Altered levels of S-glutathionylation have been associated with pathological settings [151].

Oxidation of methionine residues results in *R*- or *S*-stereoisomers of methionine sulfoxide (MetO), while most cells contain stereospecific methionine sulfoxide reductases that catalyze reduction of MetO residues back to methionine residues. The MetO content of proteins has been shown to increase with age [152].

Formation of isoaspartyl residues (with a peptide bond involving the side chain carboxylate) in proteins occurs via aspartimide formation from asparagine or aspartate (Sections 4.2.4.3, 4.2.4.9) and represents a major source of spontaneous protein damage under physiological conditions. The enzyme *L*-isoaspartyl O-methyltransferase O-methylates the free α -carboxy group of the isoaspartyl residue, followed by aspartimide formation and subsequent hydrolysis to regenerate the aspartyl residue [153].

When the thiol group of a cysteine residue in a protein reacts with NO under oxidative conditions, an S-nitrosylated protein with a cysteine S–NO bond is formed. Thiol S-nitrosylation and NO transfer reactions (transnitrosation reactions) play a role in various events of cell signalling (e.g. regulation of ion channels, G-protein coupled reactions, receptor stimulation, activation of nuclear regulatory proteins) [154].

Tyrosine residues in proteins may be nitrated to give 3-nitrotyrosine, which dramatically alters both the structure and the function of the protein due to a shift in the pK_a of the tyrosine hydroxy group from 10.1 to 7.2. The modification may also affect tyrosine O-phosphorylation with consequences e.g. for signalling processes [155].

ADP-ribosylation is a post-translational modification of proteins that influences cell signaling and the control of cellular processes like DNA repair and apoptosis. Mono-ADP-ribosylation comprises the transfer of ADP and ribose from NAD^+ to acceptor proteins, catalyzed by cellular ADP-ribosyltransferases or by certain bacterial toxins. A glycosidic bond is formed between the C^1 of ribose and the side chains of Arg, Asn, or Cys [156].

Ubiquitinylation is a well-characterized process where the 76 amino acid protein ubiquitin is C-terminally activated as a thioester and then attached to the lysine side chain of a protein substrate. Cellular proteins marked by ubiquitinylation are “labelled” for proteasome-mediated degradation [157]. Sumoylation [158] is biochemically related to ubiquitinylation, but is functionally distinct from it.

Sumoylation also involves covalent attachment of a small protein moiety, SUMO, to substrate proteins. However, conjugation of SUMO does not typically lead to degradation of the substrate but rather appears to regulate protein–protein interactions and intracellular localization.

3.2.3

Nonribosomal Peptide Synthesis

Nonribosomal peptide synthesis (NRPS) is an approach to peptide synthesis performed in prokaryotes which allows the synthesis of chemically diverse peptides characterized by high D-amino acid content, N-methylation, unusual heteroaromatic moieties, cyclic structures, and both N- and C-terminally modified building blocks. NRPS occurs without predetermination of the amino acid sequence by nucleic acids, and is mechanistically performed according to the template-directed peptide synthesis. The existence of a poly- or multi-enzymatic pathway to peptides was first suggested by Fritz Lipmann in the mid-1950s [159], and this idea was later defined as being template-directed peptide synthesis catalyzed by nonribosomal peptide synthetases (NRPS). Some of these enzymes are multimeric complexes, while others are single proteins. It has been reported that each module is composed of about 1000–1200 aa, giving a mass of the complete enzyme of $M_r \sim 2$ MDa. The peptide synthetases show a unique modular arrangement of functional units composed of individual domains catalyzing the successive condensation of their substrates by adenylation, thiolation, modification, and transpeptidation (for selected reviews, see [160–162]). The nonribosomal multienzyme thiotemplate mechanism is shown in simplified form in Figure 3.12.

According to this mechanism, peptide bond formation is performed on multi-enzyme complexes in a repeated two-step process that includes ATP-dependent aminoacyl adenylation with similarities to the amino acid activation procedure catalyzed by aminoacyl-tRNA synthetases, and aminoacyl thioesterification at specifically reactive thiol groups of the multienzyme. The initially formed unstable aminoacyl adenylate is subsequently transferred to a thiotemplate site, where it is bound as a thioester to the cysteamine moiety of an enzyme-bound 4'-phosphopantetheine (4'-PP), which is covalently bound to each thiolation domain. It has been demonstrated that peptide synthetases require post-translational modifications in order to become catalytically active. The active holoforms are formed from the inactive apoproteins by post-translational transfer of the 4'-PP residue of coenzyme A to the hydroxy group of a highly conserved serine residue. This is located at the C-terminal region of each substrate activating unit, and is defined as the acylation or thiolation domain. In this binding state, the thiol-activated amino acid derivative can undergo modifications such as epimerization (as shown in Figure 3.12) or N-methylation. Transfer of the thioester-activated carboxy group of one residue to the adjacent amino group of the next amino acid results in a stepwise elongation in a series of transpeptidation reactions. This biosynthesis system can be described as a template-driven assembly. A series of very large multifunctional peptide synthetases characterized by a modular arrangement perform the synthesis in an ordered fashion. Only the protein template is genetically encoded. A single peptide synthetase

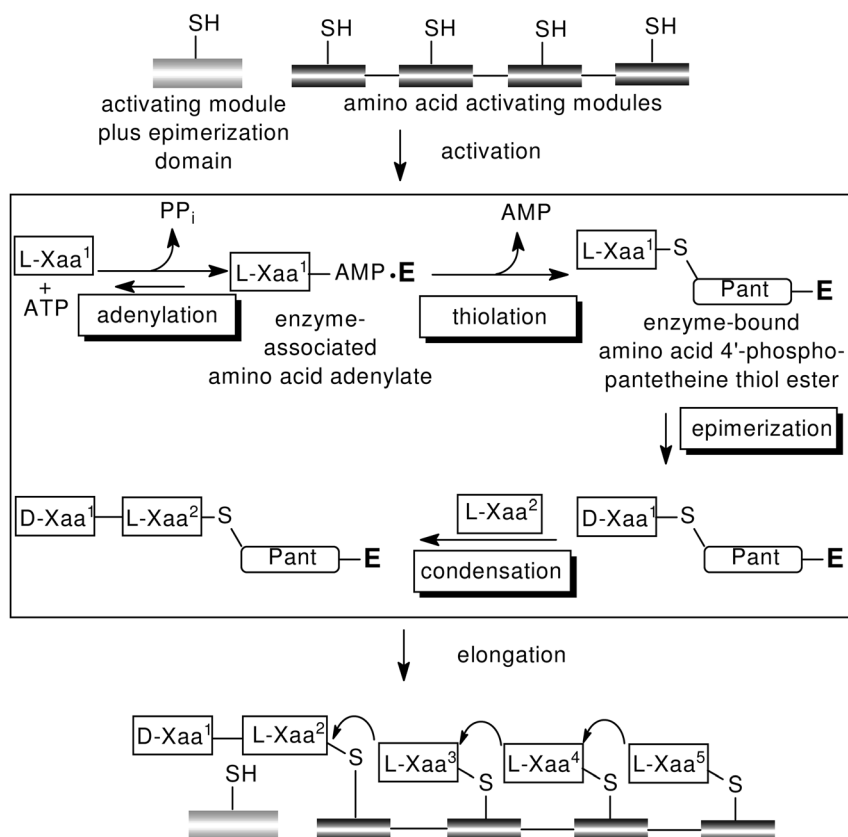


Figure 3.12 Simple scheme of nonribosomal peptide synthesis.

gene contains four to six modules, allowing the addition of four to six building blocks. Each module is capable of recognizing a residue, followed by its activation (if necessary), its modification, and its addition to the growing peptide chain. Such a minimal module can activate both amino acid and hydroxy acid residues. This model, which involves multiple carriers at multifunctional templates, was derived from investigations of the gramicidin S synthetase multienzyme complex. Based on this mechanism, various peptide products containing either encoded amino acids, or hydroxyamino acids, D-configured amino acids, or other unusual derivatives, or amino acids with further modifications such as acylation, glycosylation, N-methylation, heterocyclic ring formation, can be formed. There is evidence that more than 300 different building blocks are found in nonribosomally synthesized peptide structures. Nonribosomally synthesized peptides are produced exclusively by microorganisms, and the structural diversity is immense. Gram-positive bacteria of the species *Bacillus*, *Actinomyces* and *Streptomyces* belong to the most important producers.

Gramicidin S and the tyrocidin peptides produced by *Bacillus brevis* strains represent the best-investigated examples related to the biochemistry of nonribosomal peptide synthesis. In gramicidin S 2, two identical pentapeptides are linked via head-to-tail linkage. In contrast, the tyrocidins contain the sequence Val-Orn-Leu-Phe-Pro only as one copy, while the second pentapeptide of tyrocidins A, B, C, and D is assembled from a different set of amino acid building blocks with variation in specific positions. Gramicidin S and tyrocidin underline the relationship between peptide structure and multienzyme assembly of cyclic peptides in *Bacillus brevis* strains. Both peptides share a pentapeptide sequence that is encoded by two and three multi-enzymes, respectively.

The cyclosporin synthetase from the fungus *Beauveria nivea* is one of the largest known integrated enzyme structures, with a molecular weight of ~1400 kDa. This synthetase catalyzes at least 40 reaction steps during the final assembly of the undecapeptide chain of cyclosporin A and its cyclization [163]. In summary, there are several advantages of the application of NRPS in biotechnology, including peptide bond formation in water, without any protection of the amino acid building blocks. Epimerization domains enable selective incorporation of D-amino acids into the growing peptide chain, and finally cyclic peptides and other modified peptides can be obtained.

3.3

Selected Biologically Active Peptides

Living organisms require a highly complex chemical signaling system in order to coordinate activities at any level of their organisation. Intracellular biochemical communication is mainly based on the action of hormones and neurotransmitters. The chemically heterogeneous group of hormones contains a huge number of peptides, proteins, and even amino acids. In the following sections the most important peptide hormone families will be discussed. In addition, there will follow information on a number of individual bioactive peptides including peptide antibiotics and selected peptide toxins.

3.3.1

Gastroenteropancreatic Peptide Families

In 1902 Bayliss and Starling discovered blood-borne regulation by specific messenger molecules. Starting from the observation that acidification in the duodenum stimulated pancreatic secretion, a substance was extracted from the duodenal mucosa, which, injected into the blood stream, released bicarbonate from the denervated pancreas. Three years later, Starling then coined the word *hormone* as a general designation for blood-borne messengers and, in the same year, Edkins discovered an acid stimulatory hormone in extract of the antrum. Hence, the first two hormones, secretin and gastrin, were peptide hormones, both of gastrointestinal origin.

Gastrointestinal hormones are released to the circulation from endocrine cells, and also from neurons in the gastrointestinal tract. At present, more than 30 peptide hormone genes are known to be expressed throughout the digestive tract. Consequently, the gut is the largest endocrine hormone-producing organ in the body, both in terms of the number of endocrine cells and the number of hormones [164–166]. According to the classical concept of gastrointestinal hormones, which prevailed until the 1970s with only three known hormones secretin, gastrin, and cholecystokinin (CCK), a gut hormone was defined as a substance produced by one type of endocrine cell dispersed in a relatively well-characterized region of the gastrointestinal tract. From here, it is released to the bloodstream upon an appropriate stimulus, and reaches its target organ in order to elicit an acute response. However, modern developments in cellular and molecular biology demand a new definition of gastrointestinal hormones [167]. In fact, the well-known widespread expression of gastrointestinal hormone genes outside the gastrointestinal tract makes these peptide hormones multifunctional regulators of general physiological and biochemical interest, whilst acting simultaneously as acute metabolic hormones, local growth factors, neurotransmitters, and fertility factors. The resulting complexity has been further increased by the fact that individual genes for gut peptides encode various peptides, which release, in a tissue- and cell-specific manner, a very large number of different, biologically active peptides. Originating from one common ancestral gene, the gastrointestinal hormones form structural homology groups.

3.3.1.1 The Gastrin Family

This family (Table 3.2) comprises the mammalian hormones gastrin and cholecystokinin, the protochordate neuropeptide cionin, and the frog skin peptide cerulein. The homology of this family is decisively concentrated in and around the well-defined active site, and the common C-terminal 4-peptide sequence (-Trp-Met-Asp-Phe-NH₂). In the literature this family is also often termed the cholecystokinin-gastrin family.

Gastrin (GT) is produced in the G cells of the antrum and duodenum as preprogastrin (101 aa) which yields, after co- and post-translational processing, mainly

Table 3.2 Primary structures of gastrin family peptides.

Name	Abbr.	Sequence
Gastrin-34	hG-34	QLGPGGPPHLVADPSKKQGPWLEEEEEAY ^S GWMDFa
Gastrin-17	hG-17	QGPWLEEEEEAY ^S GWMDFa
Cholecystokinin-39	pCCK-39	YIQQARKAPSGRVSMIKNLQSLDPSHRISDRDY ^S MGWMDFa
Cholecystokinin-22	CCK-22	NLQSLDPSHRISDRDY ^S MGWMDFa
Cholecystokinin-8	CCK-8	DY ^S MGWMDFa
Cerulein ^a	CRL	<EQDY ^S TGWMDFa
Cionin ^a		NY ^S Y ^S GWMDFa

^aNot present in mammals; Y^S = Tyr(SO₃⁻).

the bioactive peptides G-71, G-34, G-17, and G-6 [166, 168]. The formation of prepro-gastrin is stimulated by partially digested protein, amino acids, and by the vagus nerve in response to stomach distension. Gastrin stimulates gastric acid and pepsinogen secretion. The cDNA-deduced amino acid sequence of human pro-gastrin comprises 80 residues. After co-translational cleavage of the 21-residue signal peptide from prepro-gastrin by signalase, pro-gastrin is transported to the Golgi apparatus, where O-sulfation of Tyr⁶⁶ by tyrosyl-protein sulfotransferase is performed. Human *gastrin-71* has the following sequence: SWKPRSQQPD¹⁰APLGT GANRD²⁰LPLPWLEQQG³⁰PASHHRR↓QLG⁴⁰PQGGPHLVAD⁵⁰PSKK↓QGPWLE⁶⁰EEEEAY⁵GWMD⁷⁰Fa Y⁵: Tyr(SO₃⁻). The arrows ↓ indicate the dibasic processing sites providing *gastrin-17* and *gastrin-34*, respectively, with N-terminal pyroglutamic acid residues formed by subsequent glutaminy cyclizations. GT is structurally related to cholecystokinin and they share some biological and pharmacological effects, exerting their biological functions via membrane GPCRs, termed CCK receptors, located on multiple cellular targets and in the CNS and peripheral organs.

Cholecystokinin (CCK) is involved in the control of multiple functions, both in the gastrointestinal tract regulating digestive functions, and in the CNS acting as neurotransmitters. It was formerly named cholecystokinin/pancreozymin (CCK-PZ), and originally isolated from the porcine duodenum as a 33-peptide (CCK-33) with sequence similarities to gastrin [169]. CCK-8 was found later in the rat brain. The CCK gene which is expressed in gastrointestinal tissues encodes prepro-cholecystokinin, which by differentiated endoproteolytic cleavages is processed to six CCK peptides varying in length from 58 to 5 amino acids: CCK-5, CCK-8, CCK-22, CCK-33, CCK-39, and CCK-58. CCK-33 and CCK-39 are the most abundant CCK peptides and have a pivotal function in regulating the exocrine pancreas. The six CCK peptides have the same C-terminal octapeptide sequence as shown for CCK-8, in which the tyrosine residue is O-sulfated. Two types of CCK receptor (type A, “alimentary” and type B, “brain”) have been identified; these were later renamed CCK1 receptor (CCK1R) and CCK2 receptor (CCK2R) [168, 170].

Cerulein (caerulein, CRL) is a member of the gastrin family of nonmammalian origin, which was firstly isolated from the skin of the Australian frog *Litoria caerulea* and from the skin of other amphibians. Synthetic CRL, named *ceruletide*, is applied in X-ray diagnostics and in the diagnosis of pancreatic function. Activation of the receptor causes, for example, breakdown of phosphoinositides, and mobilization of cellular calcium. In general, it acts in a similar manner to cholecystokinin. Due to the stimulating effect of the small intestine, ceruletide is used therapeutically in postoperative intestinal atonia and paralytic ileus, and can also be employed for the expulsion of gallstones and in biliary colic.

Cionin is a disulfotyrosyl hybrid of cholecystokinin and gastrin. It was isolated from the neural ganglion of the protochordate *Ciona intestinalis*.

3.3.1.2 Secretin Family

This family (Table 3.3) comprises secretin, glucagon, glucagon-like peptides (GLP-1 and GLP-2 both encoded by one gene), vasoactive intestinal polypeptide (VIP) and peptide histidine isoleucine amide (both encoded by one gene), growth hormone-

Table 3.3 Primary structures of secretin family peptides.

Name/Abbr.	Sequence
Secretin	HSDGTFTSEL ¹⁰ SRLREGARLQ ²⁰ RLLQGIVa
Glucagon	HSQGTFTSDY ¹⁰ SKYLSRRAQ ²⁰ DFVQWLMNT
GLP-1	HAEGTFTSDV ¹⁰ SSYL EGQAAK ²⁰ EFIAWLVKGR ³⁰ a
GLP-2	HADGRFSDEM ¹⁰ NTILDNLAAR ²⁰ DFINWLIQTK ³⁰ ITDRa
VIP	HSDAVFTDNY ¹⁰ TRLRKQMAVK ²⁰ KYLNSILNa
PHI	HAGGVFTSDF ¹⁰ SRLGQLSAK ²⁰ KYLESIIa
GHRH	YADAIFTNSY ¹⁰ RKVLGQLSAR ²⁰ KLLQDIMSRQ ³⁰ GESNQERGAR ⁴⁰ RARLa
PACAP-38	HSDGIFTD SY ¹⁰ SRYRKQMAVK ²⁰ KYLA AVLGR ³⁰ YKQRVKNKa
GIP	YAEGTFISNY ¹⁰ SIAMDKIHQQ ²⁰ DFVNWLLAQK ³⁰ GKKNDWKHNI ⁴⁰ TQ

releasing hormone, glucose-dependent insulinotropic polypeptide (GIP), pituitary adenylate cyclase-activating polypeptide (PACAP). The extendins, as well as helospectine I and II and helodermin, are from the toxins of *Heloderma horridum* and *Heloderma suspectum*, respectively, and are not present in mammals.

Secretin is released by gastric acid from S cells in the duodenum. It was originally discovered in the duodenal mucosa as a hormone enhancing the secretion of bicarbonate, enzymes, and K^+ from the exocrine pancreas. The glycine-extended S-28, and S-30, extended by Lys-Arg, were identified in porcine gut extracts in the mid-1980s as additional secretins with full biological activity. Although the existence of S-28 and S-30 was not surprising, the discovery of the bioactive S-71 was unexpected. The latter is produced after RNA splicing. S-71 contains the sequence of non-amidated S-27, followed by a Gly-Lys-Arg extension and a further C-terminal extension of 41 aa. The activation of the receptor stimulates the adenylate cyclase/protein kinase A cascade and enhances Ca^{2+} concentration in various cellular systems [171, 172].

Glucagon is formed in the α -cells of the islets of Langerhans located in the endocrine part of the pancreas [172, 173]. An increased secretion of glucagon occurs in response to a decrease in blood glucose concentration. Glucagon stimulates glucose release through both glycogenolysis and lipolysis. Glucagon and its receptor are expressed in numerous other tissues. The human glucagon gene encodes *preproglucagon* (180 aa). Post-translational proteolytic cleavage of the precursor protein gives rise to the following main cleavage fragments: genuine pancreatic glucagon (sequence 33–61), glucagon-like peptide-1 (GLP-1) (sequence 78–108) and GLP-2 (sequence 126–159). In the α -cells of the pancreatic islets the precursor protein is processed to release the pancreatic glucagon, whereas GLP processing remains incomplete. On the other hand, the L cells of the gut process the precursor in a different way to release GLPs, but glucagon remains as a part of glicentin. This prohormone fragment can be further processed to glucagon-related pancreatic peptide (GRPP) and oxyntomodulin. Glucagon and GLP-1 show an extensive sequence homology, but they exert their action through different specific receptors. Glucagon is used in the treatment of hypoglycemia; for this it can be applied parenterally, by using a portable pump, or in a depot form, nasally or as eye drops.

Glucagon-like peptides (GLP) are glucagon-like because of their ~50% sequence homology with glucagon. *GLP-1*, *GLP-1-(7–36)-amide* (proglucagon 78–107-NH₂), was originally predicted to consist of 37 residues (proglucagon 72–108; N-terminally extended sequence: His-Asp-Glu-Phe-Glu-Arg-), and, therefore, the secreted hormone is frequently alluded to as *GLP-1-(7–36)-amide*. The insulinotropic peptide amide is secreted from endocrine cells in the gut mucosa in response to meal ingestion, and belongs to the integrins. *GLP-1* suppresses glucagon secretion, promotes satiety, delays gastric emptying, and stimulates peripheral glucose uptake. The insulinotropic effect is preserved in patients with type 2 diabetes mellitus, in whom the secretion of glucagon is also inhibited. After *GLP-1* infusion, blood glucose may be completely normalized. *GLP-1* appears to regulate plasma glucose levels through various and independent mechanisms. It is an excellent candidate option for the treatment of patients with type 2 diabetes mellitus. The structure of *GLP-2* (proglucagon 126–159), was confirmed by synthesis. *GLP-2* is secreted from gut endocrine cells and promotes nutrient absorption. *GLP-2* regulates gastric motility, gastric acid secretion, intestinal hexose transport, and increases the barrier function of the gut epithelium [174, 175].

Vasoactive intestinal polypeptide (VIP) was first isolated from porcine intestine and is a potent vasodilator, a major growth stimulator, a bronchodilator, and a neuronal survival-promoting agent. VIP is widely distributed in the peripheral and central nervous systems, and acts both as neurotransmitter and as a hormone. For this reason, it is not only an intestinal peptide but, in accordance with its early given abbreviation, it is also a very important peptide. VIP is co-synthesized with peptide histidine isoleucine amide on the same peptide precursor [176, 177].

Peptide histidine isoleucine amide (PHI) is a 27-peptide amide isolated from porcine intestine. It shares the same peptide precursor with the vasoactive intestinal peptide (VIP), and occurs together with VIP, for example, in the neurons of the central nervous system, in the digestive tract, urogenital tract, lungs, and wall of the gallbladder. The biological activities are similar to those of VIP and secretin [178].

Growth hormone-releasing hormone (GHRH), *somatoliberin*, is released from neurosecretory terminals in the median eminence and stimulates the synthesis and release of growth hormone. GHRH is expressed in the arcuate nucleus of the hypothalamus and also in other tissues, such as gonads, intestine, immune tissues and the placenta. The application of GHRH for the treatment of children with growth hormone deficiency is very important [179].

Glucose-dependent insulinotropic polypeptide (GIP), also referred to as *gastric inhibitory polypeptide*, YAEGTFISNY¹⁰SIAMDKIHQQ²⁰DFVNWLLAQK³⁰GKKNDWKHNI⁴⁰TQ, is a 42-peptide involved in the regulation of fat and glucose metabolism [180]. GIP was initially discovered from porcine intestinal extracts in 1970. The early assumed inhibitory effect on gastric acid secretion could not be confirmed. More importantly, GIP acts as an incretin hormone. *Incretins* are insulinotropic hormones of the gut released by nutrients. They stimulate insulin secretion in physiological concentrations in the presence of elevated blood glucose levels. The incretin effect has been defined as the ratio between the integrated insulin response to an oral

glucose load and an isoglycemic intravenous glucose infusion. This effect is estimated as being responsible for 50–70% of the insulin response to glucose, and is mainly caused by the two intestinal insulin-stimulating hormones GLP-1 and GIP that only fulfill the incretin definition of the known gut hormones. GIP is released from K cells in the gut after meal ingestion, acting in concert with glucagon-like peptide 1 to augment glucose-stimulated insulin secretion. Furthermore, GIP cells are also found in the entire small intestinal mucosa.

Pituitary adenylate cyclase-activating polypeptide (PACAP), a neuropeptide occurring in two variants: a full-length 38-peptide, PACAP-38, and a truncated 27-peptide, PACAP-27, that is equivalent to the N-terminal of PACAP-38. PACAP was originally isolated from ovine hypothalamus and has been found in virtually every tissue in the body. PACAP is a multifunctional peptide hormone influencing diverse biological functions, e.g., smooth muscle and cardiac muscle relaxation, regulation of cell cycle, bone metabolism, and endocrine/paracrine function [181].

3.3.1.3 The Insulin Superfamily

The insulin superfamily [182, 183] (Figure 3.13) is an ancient group of small, structurally related proteins, such as insulin, insulin-like growth factors (IGF), and



Figure 3.13 Selected members of the insulin family.

the relaxin-like peptide family. Insulin-related peptides have also been described in both vertebrates and invertebrates, including nematodes, mollusks, and insects. Many of these hormones initiate an evolutionary conserved signal transduction mechanism by acting via a heterotetrameric, membrane-spanning receptor tyrosine kinases.

Insulin [184] is a 51-polypeptide hormone consisting of the A chain (acidic chain; 21 aa) and the B chain (basic chain, 30 aa) linked by two disulfide bridges (Cys⁷_A–Cys⁷_B/Cys²⁰_A–Cys¹⁹_B). One more intrachain disulfide bond is located between Cys⁶–Cys¹¹ within the A chain. The name insulin is derived from Latin *insula*, “island”, as it is secreted from the β cells of the islets of Langerhans in response to high blood glucose levels. Although most species have only one type of insulin, three rodents (laboratory rat, mouse, spiny mouse) and two fish (toadfish, tuna) have two distinct insulin types. In biosynthesis, the mature mRNA transcript encodes prepro-insulin bearing an N-terminal hydrophobic, 16 aa signal sequence. The latter is removed by a signal peptidase when prepro-insulin traverses the membrane into the lumen of the ER. The resulting 84 aa proinsulin containing the disulfide bonds is transported to the Golgi complex, where insulin is formed after enzymatic excision of the connecting peptide (C-peptide) between the C-terminal A chain and the N-terminal B chain. Insulin secretion as a response to elevated blood glucose levels is triggered by an increase in the concentration of free cytosolic Ca²⁺. Inositol triphosphate as the messenger releases Ca²⁺ from internal stores. As a rule, the biosynthesis of insulin is stimulated at glucose concentrations >2–4 mM, whereas glucose concentrations 4–6 mM are required for the stimulation of insulin secretion. The normal concentration of circulating insulin in human blood, which can be determined radioimmunologically, is 1 ng mL⁻¹. The liver is the major site of insulin degradation, but most peripheral tissues also contain insulin-degrading enzymes. Insulin acts together with its counterpart glucagon (which has largely opposing effects) to maintain the correct concentration of blood glucose. An increased blood glucose level (hyperglycemia) is the physiological signal for the synthesis and secretion of this hormone. Insulin increases cell permeability for glucose and other monosaccharides, as well as fatty acids and amino acids. Furthermore, it stimulates liver, muscle, and adipose cells to store these metabolites for further use in the form of glycogen, protein, and fat. In diabetes mellitus, insulin is either not secreted in sufficient concentrations or does not efficiently stimulate its target cells. An abnormally high concentration of ketone bodies (ketosis) is one of the most dangerous effects. The insulin receptor is a transmembrane glycoprotein ($M_r \sim 300$ kDa) with tyrosine-specific protein kinase activity converting the extracellular insulin signal into the cell. The receptor is a tetramer ($\alpha_2\beta_2$), and consists of two identical subunits joined together by disulfide bonds. The catalytic domain of the tyrosine-specific protein kinase is exposed on the cytoplasmic side of the plasma membrane. When activated by insulin binding, it phosphorylates itself, thereby enhancing the activity of the kinase, and transfers the terminal phosphate group from ATP to the hydroxy group on a tyrosine residue of selected proteins in the target cell. In 1922, Banting and Best utilized a classical endocrinological procedure in their discovery of insulin. They pulverized pancreas,

filtered the crude tissue extract, and observed that it produced hypoglycemia in an experimental dog.

Insulin-like growth factors (IGF) are polypeptides with structural and functional resemblance to insulin and relaxin [185–187]. The IGF family of ligands, binding proteins and receptors is an important growth factor system involved in both the development of the organism and the maintenance of the normal function of many cells of the body. It also shows powerful anti-apoptotic effects. IGF occur in the blood of vertebrates and play an important role in the regulation of growth and differentiation of an ever-increasing number of tissues. IGF are known to associate with high affinity to a family of circulating binding proteins (IGF-BP) that influence IGF availability and activity. IGF I is a single-chain 70-polypeptide ($M_r \sim 7.6$ kDa) containing three intrachain disulfide bridges. IGF I mediates, in interaction with other hormones, the effects of growth hormone on bone growth. Similar to IGF II, IGF I possesses growth-promoting activity on chick embryo fibroblasts at a concentration of 10^{-9} M. *IGF II* is a single-chain 67-polypeptide ($M_r \sim 7.5$ kDa) containing three intrachain disulfide bridges. It participates in the fetal and embryonal development of the nervous system and the bones. Interest in IGF and their effects on carcinogenesis have increased because high serum concentrations of IGF I are associated with an increased risk of breast, prostate, colorectal, and lung cancers.

The *relaxin-like peptide family* [188–190] comprises seven peptides with high structural but low sequence similarity. To this family belong three *relaxin peptides* (relaxin-1, relaxin-2, relaxin-3) and four *insulin-like peptides* (INSL-3, INSL-4, INSL-5, INSL-5). In most other mammals, there are only two *RLN* genes that encode relaxin and relaxin-3. It has been shown that the recently discovered *RLN3* gene is the ancestral gene for the relaxin peptide family. However, the function of relaxin-3 remains to be determined. Its highest expression is in the brain, which suggests a neuropeptide role.

Relaxin shows structural similarity to insulin [191–193]. It is first synthesized as a prohormone that is comprised of a signal sequence and a B-C-A domain configuration. The C-peptide is removed during processing. Rat relaxin consists of an A chain (24 aa) and a B chain (35 aa) linked by two interchain disulfide bonds ($\text{Cys}_A^{11}-\text{Cys}_B^{14}/\text{Cys}_A^{24}-\text{Cys}_B^{26}$). In addition, the A chain contains an interchain disulfide bond ($\text{Cys}^{10}-\text{Cys}^{15}$). Relaxin is produced in the reproductive tract of many mammals during pregnancy, and acts to maintain pregnant uterine tissues and relax the pubic ligaments in preparation for birth. Furthermore, it promotes development of the mammary apparatus, thus enabling normal lactational performance. In humans, relaxin is produced and secreted in small amounts during pregnancy, but until now no physiological function for circulating relaxin has been identified. Interestingly, diverse therapeutic actions of relaxin on non-reproductive tissues have been described that have clinical implications. For example, relaxin has been shown to promote wound healing and to reduce fibrosis in the heart, kidney, liver, and lung. In humans, three genes have been established: *RLN1* (encoding human relaxin-1, H1), *RLN2* (human relaxin-2, H2), and *RLN3* (human relaxin-3, H3). The A chains of H1–H3 consist of 24 aa differing in their sequences, whereas differences also exist in the length of the B chains (H1 and H3: 28 aa; H2: 29 aa). In most other mammals,

there exist only two *RLN* genes, encoding relaxin and relaxin-3. The relaxin receptor is a leucine-rich repeat-containing G protein-coupled receptor 7 (LGR7).

Insulin-like peptides (INSLP) comprise the four insulin-like peptides 3 to insulin-like peptide 6 which are encoded in human by the genes *INSL3–INSL6* [188, 194–196]. They are structurally related to insulin. The INSLP each contain individual A and B chains that are connected by two disulfide bonds and formed from a prohormone after proteolytic cleavage of the intervening C chain. INSLP-3, also known as relaxin-like factor (RLF) and Leydig cell insulin-like peptide, is an A (26 aa)–B (31 aa) heteromer covalently linked by two interchain disulfide bonds ($\text{Cys}_A^{11}\text{–Cys}_B^{10}/\text{Cys}_A^{24}\text{–Cys}_B^{22}$). Until now, the functions of the insulin-like peptide 4 (INSLP-4, placentin, early placenta insulin-like peptide, EPIL), insulin-like peptide 5 (INSLP-5, relaxin/insulin like factor-2, RIF2), and insulin-like peptide 6 (INSLP-6, relaxin/insulin like factor-1, RIF1) have not been elucidated; neither have their appropriate receptors been identified.

3.3.1.4 The Somatostatin Family

The somatostatin family consists of somatostatin and corticostatin (Figure 3.14). *Somatostatin* (SST) is a heterodetic, cyclic 14-peptide released by the hypothalamus. It is a potent inhibitor of growth hormone (GH) release, but not of growth hormone synthesis. SST also inhibits the secretion of several other hormones, such as glucagon, thyrotropin, insulin, cholecystokinin, and gastrin. Furthermore, SST is a regulator of many other biological processes. Besides the hypothalamus, SST-14 and SST-like peptides (including SST-28) have been found in the stomach, pancreas, and also in the central and peripheral nervous systems, thymus, gastrointestinal tract, ovaries, and even in plants (tobacco, spinach). SST infusions have been applied to the treatment of bleeding peptic ulcers and gastrointestinal lesions, for the preventive treatment of stress ulcers, and in the healing of fistulae of the small intestine and gallbladder. Many analogues have been synthesized and tested to identify some with higher selectivity and a longer half-life. Some analogues are used in the treatment of growth hormone- and thyrotropin-secreting pituitary tumors, carcinoid tumors. Especially, lanreotide, H-D-βNal-Cys-Tyr-D-Trp-Lys-Val-Cys-Thr-ol (disulfide bond: $\text{C}^2\text{–C}^7$), and vapreotide, H-D-Phe-Cys-Tyr-D-Trp-Lys-Val-Cys-Trp-OH (disulfide bond:

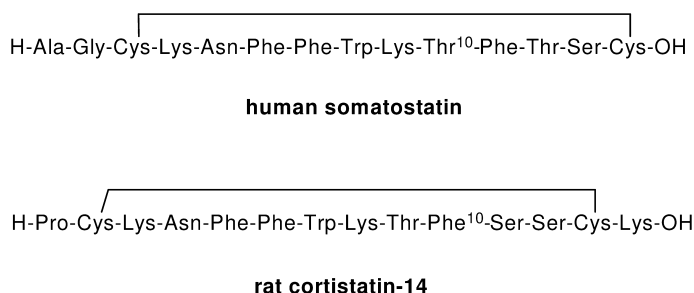


Figure 3.14 Members of the somatostatin family.

C²–C⁷) have become available for clinical use. Furthermore, radiolabeled somatostatin analogues such as ⁹⁰Y-DOTA-Tyr³-octreotide (⁹⁰Y – DOTATOC) with the β-emitter ⁹⁰Y have been developed for radiotherapy [197, 198].

Cortistatin (CST) is a cyclic neuropeptide and is named after its predominantly cortical expression and ability to depress cortical activity. The rat precursor protein is predicted to have 112 aa with a striking sequence homology to somatostatin at its distal C-terminal end. After removal of the 27 aa signal sequence, procortistatin could be further cleaved to rCST-29, rCST-14, and additional products by further cleavage. In addition, CST has also been cloned from humans, and mouse sources. CST is capable of binding to all five cloned somatostatin receptors, and shares many pharmacological and functional effects with SST, including the depression of neuronal activity and inhibition of cell proliferation. CST acts also as an anti-inflammatory peptide with therapeutic action on both lethal endotoxemia and inflammatory bowel disease (IBD) [199, 200].

3.3.1.5 The Tachykinin Family

Tachykinins comprise a large peptide family which is phylogenetically ancient and has been well conserved during evolution [201, 202]. Structurally related tachykinin peptides have been isolated from mammals, reptiles, birds, amphibia and fish. Their activities are involved in the central and peripheral nervous systems as well as in the cardiovascular and immune systems of both lower and higher life forms. The mammalian tachykinins, selected members are listed in Table 3.4, act as neurotransmitters, paracrine or endocrine factors and neuroimmunomodulators having functions in the nervous system, gastrointestinal tract and cardiovascular system [203, 204]. They are encoded on three different genes, termed *TAC1*, *TAC3* and *TAC4*. Multiple receptors, termed NK₁, NK₂, and NK₃, of the mammalian peptides belong to the superfamily of G protein-coupled receptors that have distinct pharmacological features. The tachykinin peptides are characterized by a conserved C-terminal sequence –Phe-Xaa-Gly-Leu-Met-NH₂, where Xaa represents an aromatic (Xaa = Phe

Table 3.4 Primary structures of mammalian tachykinins.

Name	Origin	Sequence
Substance P	bovine	RPKPQQFFGLMa
Neurokinin A	porcine	HKDSVFGLMa
Neurokinin B	porcine	DMHDDFFVGLMa
Neurokinin K	porcine	DADSSIEKQVALLKALYGHGQISHKRHGQISHKRKDSVFGLMa
Neurokinin γ	rabbit	DAGHGQISHKRKDSVFGLMa
Endokinin A	human	DGGEQTLSTEAEETWVIVALEEGAGPSIQLQLQEVKTGKASQFFGLMa
Endokinin B	human	DGGEQTLSTEAEETWEGAGPSIQLQLQEVKTGKASQFFGL Ma
Hemokinin-1	mouse	RSRTRQFYGLMa
Endokinin C	human	KKAYQLEHTFQGLLa
Endokinin D	human	VGAYQLEHTFQGLLa

or Tyr) or hydrophobic (Xaa = Val or Ile) building block. For a long time, the tachykinin family was only represented by substance P (SP), neurokinin A (NKA), neurokinin B (NKB) and two elongated analogs of NKA, neuropeptide γ (NP γ) and neuropeptide K (NPK). Nowadays this family is expanded by new members, hemokinin-1 (HK-1) and endokinins A-D. Substance P [205, 206] exerts a wide spectrum of biological effects such as stimulation of the smooth muscle, salivary secretion, and lowering of the blood pressure. The activity is mediated via the G protein-coupled receptors NK₁, NK₂ or NK₃. SP occurs both in the gastrointestinal tract and in the brain, and acts a neurotransmitter in various brain regions. The *neurokinins* [207–209] are widely distributed in the central and peripheral nervous systems, and the two neurokinins (NKA and NKB) act as neurotransmitters or neuromodulators. Together with substance P, the NK play an important role in pain transmission, neurogenic inflammation, smooth muscle contraction, secretion, vasodilation, and activation of the immune system. The *endokinins* (EK) [210, 211] with cardiovascular activity are encoded on the human *TAC4 gene*, and are mainly distributed in the adrenal glands and in the placenta. The occurrence in the placenta suggests that EK may be involved in controlling blood flow during pregnancy. EK A/B 10-mers show equivalent affinity for the three tachykinin receptors as substance P. Whilst EK C and EK D, which possess a previously uncharacterized tachykinin motif (FQGLLa), display only low potency. The novel mammalian tachykinin *hemokinin-1* (HK-1) was discovered in lymphoid B hematopoietic cells of mouse bone marrow. HK-1 is uniquely expressed outside the neural tissue in immune tissues and is a full agonist at the NK receptors, with highest selectivity for NK₁ [211, 212].

The amino termini of the tachykinins are quite variable, but classes of vertebrate tachykinins with structural similarities have been termed as SP-like, NKA-like, and NP γ -like tachykinins.

Nonmammalian tachykinins [58, 201] that occur in cold-blooded animals also show a wide spectrum of pharmacological activities in certain mammalian tissue preparations. Selected members are listed in Table 3.5. In 1952, Erspamer isolated a substance, subsequently named eledoisin, from extracts of salivary glands of the cuttlefish (*Eledone moschata*). Ten years later its structure was elucidated, starting from 10 000 pairs of posterior salivary glands of 1450 kg of the eight-armed cephalopod. Eledoisin causes contraction of smooth muscle preparations and marked peripheral vasodilatation, resulting in a blood pressure decrease. Physalaemin acts similarly, but is three to four times more potent. Five tachykinins have been isolated from the skin of the Australian frog *Pseudophryne güntneri*. Whereas three of these are similar to kassinin, the two other peptides are similar to SP. Locustatachykinin I (GPSGFYGV_R) and locustatachykinin II (APLSGFYGV_R) from the insect *Locusta migratoria* are characterized by significant changes in the C-terminal sequence part compared with the other tachykinins.

3.3.1.6 The Neuropeptide Y family

The neuropeptide Y family (Table 3.6) (neuropeptide tyrosine, NPY family) comprises neuropeptide Y, pancreatic polypeptide and peptide YY. An alternative desig-

Table 3.5 Primary structures of selected non-mammalian tachykinins.

Name	Source	Sequence
Eledoisin	<i>Eledone moschata</i> (musky octopus)	<EPSKDAFIGL ^a Ma
Uperolein	<i>Uperoleia marmorata</i> (marbled toadlet)	<EPDPNAFYGL ^a Ma
Uperin	<i>Uperoleia inundata</i> (flood plain toadlet)	<EADPNAFYGL ^a Ma
Physalaemin	<i>Physalaemus fuscumaculatus</i> (brown-spotted dwarf frog)	<EADPNKFYGL ^a Ma
Phyllomedusin	<i>Phyllomedusa bicolor</i> (giant leaf frog)	<EPNPNRFIGL ^a Ma
Kassinin	<i>Kassina senegalensis</i> (Senegal running frog)	DVPKSDQFVGL ^a Ma
Hylambatin	<i>Kassina (Hylambates) maculata</i> (spotted running frog)	DPPDPNRFYGL ^a Ma
Ranamargarin	<i>Rana margaratae</i> (Margareta's odorous frog)	DDASDRAKKFYGL ^a Ma
Ranakinin	<i>Rana ridibunda</i> (Central Asian marsh frog, brain)	KPNPERFYGL ^a Ma
Bufokinin	<i>Bufo marinus</i> (marine toad, gut)	KRPDPQFYGL ^a Ma
Ranatachykinin A	<i>Rana catesbeiana</i> (bull frog, brain)	KPSPDRFYGL ^a Ma
Carassin	<i>Carassius auratus</i> (goldfish, brain)	SPANAQITRKRHKINSFVGL ^a Ma
Locustatachykinin I	<i>Locusta migratoria</i> (migratory locust)	GPSGFYGV ^a Ra
Locustatachykinin II	<i>Locusta migratoria</i> (migratory locust)	APLSGFYGV ^a Ra
Scylliorhinin I	<i>Scylliorhinus caniculus</i> (dogfish, gut)	AKFDKFYGL ^a Ma
Scylliorhinin II	<i>Scylliorhinus caniculus</i> (dogfish, brain)	^a SPSNSKCPDGPDCFVGL ^a Ma

^aDisulfide bond: C⁷-C¹³

nation PP-fold family was proposed by Rehfeld in 1998. The members of this family are 36-peptide amides exerting various biological functions, such as vasoconstriction, intestinal functions, stimulation of food intake, regulation of circadian rhythms, and release of pituitary sex hormones. The five receptors of the NPY family – Y1, Y2, Y4, Y5, and y6 – belong to the rhodopsin-like superfamily (class 1) of G protein-coupled receptors.

Neuropeptide Y (NPY) [213–215] occurs widely distributed in the central and peripheral nervous system in the brains of vertebrates with the highest concentration

Table 3.6 Primary structures of neuropeptide Y family peptides.

Name	Origin	Sequence
Neuropeptide Y (NPY)	porcine	YPSKPDNPGE ¹⁰ DAPAEDLARY ²⁰ YSALRHYINL ³⁰ ITRQRYa
Pancreatic polypeptide (PP)	human	APLEPVYPGD ¹⁰ NATPEQMAQY ²⁰ AADLRRYINM ³⁰ ITRPRY ^a
Peptide YY (PYY)	porcine	YPAKPEAPGE ¹⁰ DASPEELSR ²⁰ YASLRHYLNL ³⁰ VTQRQRYa

in the hypothalamus. Most of the vascular effects of NPY, and many of its physiological actions, are mediated through the Y1 receptor. The Y5 receptor subtype seems to play a role in NPY-induced feeding. NPY inhibits the formation of cAMP which is stimulated by, e.g., forskolin or isoproterenol, and the calmodulin-stimulated phosphodiesterase. It increases intracellular Ca^{2+} concentration in vascular smooth muscle cells. However, more biological and other effects of NPY have been described.

Pancreatic polypeptide (PP) [216, 217] is a 36-peptide amide formed in the PP cells of the pancreas. The PP sequences among different higher mammals differ only by 1–4 amino acids. The three-dimensional structures of different PP are known as the PP-fold. The plasma level of PP is increased by electric stimulation of the vagus, food intake and several other stimulating factors. PP inhibits exocrine secretion from the pancreas, which is stimulated, e.g., by secretin, cerulein, and cholecystokinin.

Peptide YY (PYY) [218, 219] is released in the gut postprandially by the L cells of the lower intestine, and acts as an inhibitor of gastric acid and motility via neural pathways. Beside other peptides, such as PP, GLP-1, and CCK, PYY and its endogenous form PYY_{3–36} are known as short-term regulators of food-intake, in contrast to the long-term satiety regulators leptin and insulin. PYY is involved in the regulation of energy homeostasis in obese children.

3.3.1.7 The Ghrelin Family

Members of the ghrelin family are compiled in Table 3.7. *Ghrelin* is a 28-peptide with an essential n-octanoic moiety esterified with the hydroxy group of Ser³ originally discovered as the endogenous ligand of the growth hormone secretagogue receptor (GHS-R). The name “ghrelin” is related to “ghre”, a word root in Proto-Indo-European languages for “grow”. Almost simultaneously another peptide with the same amino acid sequence as rat ghrelin was identified in mouse. It was named *motilin-related peptide* (MTLRP) because its precursor was similar to that of motilin. In addition, *motilin homolog* with identical partial sequence 1–14 of human ghrelin was already submitted as part of a US patent application. However, both amino acid sequences were derived from the nucleotide sequence. Since it became true that the unique post-translational modification is essential for the biological activity of this peptide hormone it seems therefore indicated to use only the the name ghrelin coined

Table 3.7 Primary structures of ghrelin family peptides.

Name	Origin	Sequence
Ghrelin	human	GSS ^X FLSPEHQ ¹⁰ RVQQRKESKK ²⁰ PPAKLQPR
MTLRP	mouse	GSS FLSPHEQ ¹⁰ KVQQRKES
Motilin homolog		GSS FLSPENQ ¹⁰ RVQQ
Ghrelin	rat/murine	GSS ^X FLSPEHQ ¹⁰ KAQQRKESKK ²⁰ PPAKLQPR
Des-Gln ¹⁴ -ghrelin	rat, splice variant	GSS ^X FLSPENQ ¹⁰ RVQQRKESKKP ²⁰ PAKLQPR
Motilin (MTL)	human	FVPIFTYGEI ¹⁰ QRMQEKERNK ²⁰ GQ

X = CH₃–(CH₂)₆–CO–

by the true discoverers Kojima *et al.* [220, 221]. Ghrelin is secreted primarily from the stomach, and secondarily from the small intestine and colon. Furthermore, it may also be expressed in the hypothalamus, the pancreatic islets, the pituitary, and several tissues in the periphery. In addition to its predictable effect on growth hormone secretion, ghrelin has an important function in the short-term regulation of appetite and the long-term regulation of energy balance and glucose homeostasis. Recent reports have implicated ghrelin in the regulation of gastrointestinal, cardiovascular, immune function, and bone physiology. Obestatin, derived from the same gene as ghrelin, shows opposite actions on energy homeostasis and gastrointestinal function. Ghrelin is highly conserved among mammals, and has even been discovered in fish, chickens, dog, goose, emu, turkey, and bullfrogs. *Des-Gln¹⁴-ghrelin*, a splice variant with n-octanoyl modification at the hydroxy group of Ser³, like ghrelin, stimulates the release of growth hormone after injection into rats. The discovery of ghrelin was a milestone in the understanding of the interplay between the stomach and the brain, and has brought new light into the neuroendocrine networks that regulate food intake, energy balance, gastrointestinal function, and growth [222–224]. *Motilin* (MTL) [225, 226] was originally isolated from porcine gut, and is mainly expressed and secreted by enterochromaffin cells of the small intestine. MTL is released from pro-motilin, besides motilin-associated peptides (MAP). The sequence of motilin is highly conserved among species. In the past MTL was not a member of related peptides, but the discovery of ghrelin has changed this. It stimulates the contraction of smooth muscle in the gastrointestinal tract. Ghrelin and MTL, and GHS-R and the motilin receptor, are structurally related. In principle, motilin can activate GHS-R-expressing cells, but its activity is very weak. It has been suggested that ghrelin and motilin might have evolved from a common ancestral peptide. Interestingly, *des-Gln¹⁴-ghrelin* shows an even better alignment with motilin than ghrelin itself.

3.3.1.8 The EGF Family

The *epidermal growth factor (EGF) family* [227, 228] comprises epidermal growth factor (EGF), transforming growth factor- α (TGF- α), amphiregulin (AR), heparin-binding epidermal growth factor (HB-EGF), betacellulin (BTC), epiregulin (ER), and the neuroregulins 1–3. Furthermore, cripto, heregulin and pox virus growth factors, occurring in vaccinia, Shope and myxoma viruses, belong to this group. The members of the EGF family bind to the EGF receptor (EGFR). However, additional receptors, similar in structure, have been identified, including the human EGF receptors 2 (HER-2) and 3 (HER-3). EGF and TGF- α are the prototype members of the EGF family. Within the gastrointestinal tract, EGF is synthesized in the submaxillary glands and Brunner's glands, and smaller amounts are produced within the exocrine pancreas. EGF is also found in gastric juice and within the intestinal lumen as a result of secretion from salivary glands. It is assumed that EGF exerts trophic and biological properties within the gastrointestinal tract via interaction at the mucosal surface. TGF- α is produced by many cell types, including most epithelial cells. TGF- α mRNA and the bioactive peptide has been found directly in human and rodent stomach, colonic and small epithelium. TGF- α appears to act via the same receptor as EGF. Furthermore, it

has been proposed that TGF- α may play an important role in the development of intestinal malignancies. The members of this family share six conserved Cys residues, have one or more EGF domains, and a transmembrane domain.

3.3.2

Hypothalamic Liberins and Statins

A group of peptide hormones synthesized in the neurons of various distinct nuclei of the hypothalamus were initially named as releasing factors and release-inhibiting factors. Nowadays, these molecules are called releasing hormones or, according to an IUPAC-IUB recommendation, liberins. Moreover, the release-inhibiting factors are better designated as release-inhibiting hormones or statins (Table 3.8).

The hypothalamus is the lowest part of the midbrain and, in certain nuclear areas, nervous excitation is transformed into hormonal signals, for example releasing hormones and neurohypophyseal hormones. The releasing hormones, which are stored in the eminence of the hypothalamus in nanogram quantities, have a half-life of the order of a few minutes. The classical idea of hypothalamic regulation of anterior pituitary (AP) hormone secretion holds that release of each AP hormone

Table 3.8 Hypothalamic releasing hormones (liberins)^a and release-inhibiting hormones (statins)^b.

Name ^b	Synonym	Abbr.	Major Effects
Thyroliberin	Thyrotropin-releasing hormone	TRH	Stimulates TSH release
Gonadoliberin	Gonadotropin-releasing hormone	GnRH	Stimulates both lutropin and follitropin release
Luliberin	Lutinizng hormone-releasing	LH-RH	
Folliliberin	hormone Follicle hormone-releasing hormone	FSH-RH	
Corticoliberin	Corticotropin-releasing hormone	CRH	Stimulates corticotropin (ACTH) release
Prolactoliberin	Prolactin-releasing hormone	PRH	Stimulates prolactin release
Prolactostatin	Prolactin-release inhibiting hormone	PIH	Inhibits prolactin release
Somatoliberin	Somatotropin-releasing hormone	SRH	Stimulates somatotropin release
	Growth hormone-releasing hormone	GRH	Stimulates growth hormone release
Somatostatin	Somatotropin-release inhibiting hormone	SIH	Inhibits somatotropin (growth hormone) release
Melanotropin	Melanotropin-releasing hormone	MRH	Stimulates melanotropin release
Melanostatin	Melanotropin-release inhibiting hormone	MIH	Inhibits melanotropin release

^aInstead of releasing hormone and release inhibiting hormone the terms releasing factor (RF) and release-inhibiting factor (RIF) are also in use.

^bAccording to IUPAC-IUB recommendation.

is controlled specifically by a corresponding hypothalamic-releasing hormone. The binding of a given hypothalamic-releasing hormone to specific receptors in its target cell increases the concentration of cytosolic Ca^{2+} , thereby selectively stimulating the release of the appropriate hormone. Generally, the hypothalamic hormones are delivered via a direct circulatory connection to the anterior pituitary, where they stimulate or inhibit the release of the appropriate trophic hormone into the bloodstream. In most cases the trophic hormones stimulate the appropriate endocrine glands to secrete the corresponding endocrine hormones. The classical relationship between hypothalamus, hypophysis, and target glands or target tissues in humans is shown schematically in Figure 3.15.

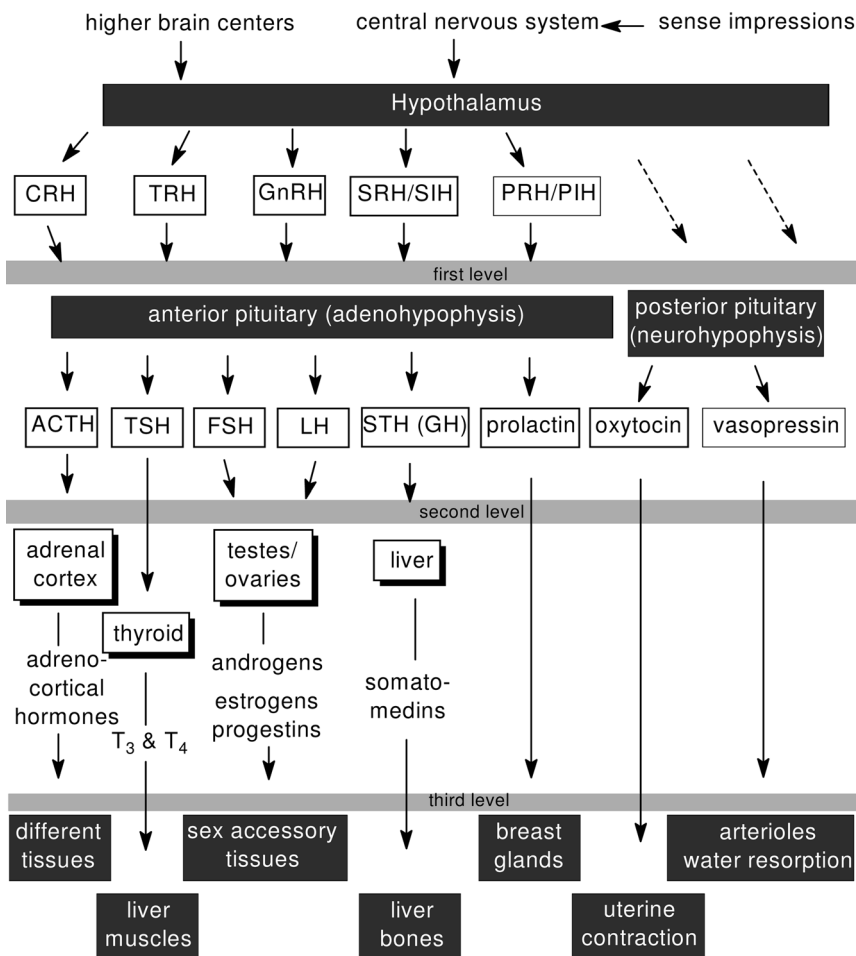
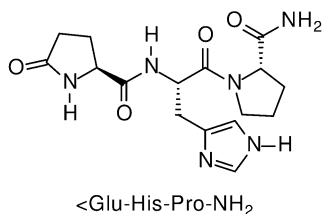


Figure 3.15 Relationship between hypothalamus, hypophysis and target tissues including action of the hormones of the second and third level.

3.3.2.1 Thyroliberin

Thyroliberin (TRH), <Glu-His-Pro-NH₂, was initially discovered in the nervous system (brain), pancreas, intestine, and the prostate gland. It is formed as prepro-TRH which, in the rat, consists of 225 residues and contains five copies of the TRH sequence. After release of the TRH sequences by limited proteolysis, the N-terminal glutamine is cyclized to pyroglutamic acid, whilst the C-terminal glycine is converted to the amide function under catalysis by peptidylglycine-amidating monooxygenase giving the primary structure of TRH **13**. TRH stimulates the anterior pituitary to produce and secrete the trophic hormone thyrotropin, which in turn stimulates the thyroid gland to synthesize and secrete thyroxine (T₄) and triiodothyronine (T₃). The latter two hormones cause a negative feedback on the secretion of TRH and thyrotropin. Further biological effects of TRH are the secretion of prolactin, the inhibition of glucagon in the pancreas, contraction of the urinary bladder, vasodilatation in the cardiovascular system, and action as a neurotransmitter in the CNS, with various specific effects. TRH has found application as a diagnostic for thyroid and hypophyseal functions. Some TRH preparations for oral, parenteral, and intranasal application are available commercially to treat spinal injury, schizophrenia, epilepsy, motor neuron disease, depression, and circulatory shock. Some analogues are more effective than native TRH; for example, <Glu-3MeHis-Pro-NH₂ and the corresponding *N*-amylamide possess eight- and ten-fold the potency of the native hormone, respectively. In contrast, a higher number of analogues which have central activity but which do not release TSH have been described. Pyroglutamyl-peptidase II (thyroliberinase), which cleaves specifically the pGlu-His bond, is an important enzyme in the breakdown of TRH. Metabolites of TRH, such as cyclo-(-His-Pro-) and pGlu-His-Gly-OH show a central appetite-inhibiting effect. The latter compound, named *anorexigenic peptide*, was found in the urine of women suffering from anorexia nervosa. Furthermore, a correlation between increased cyclo-(-His-Pro-) plasma levels with weight loss in bulimia and anorexia nervosa has been reported. TRH was the first releasing hormone to be isolated, and its structure was elucidated independently in 1969 by the groups of Andrew Schally [229] and Roger Guillemin [230] using extracts of hypothalami from approximately 3 million animals (sheep and pigs). TRH exerts its biological effects via the stimulation of cell-surface receptors, TRH-R, that belong to the G protein-coupled receptors [231, 232]. Interestingly, in the cryptic region of the prepro-TRH the sequence of the 22-peptide corticotropin release-inhibiting factor has been elucidated.



13

3.3.2.2 Gonadoliberin

Gonadoliberin (GnRH) **14** stimulates collectively the synthesis and secretion of the gonadotropins, e.g., luteinizing hormone (LH), and follicle-stimulating hormone (FSH) in the adenohypophysis, that in turn regulate ovulation/spermatogenesis. GnRH is identical with luteinizing hormone-releasing hormone (LH-RH) and follicle-stimulating hormone-releasing hormone (FSH-RH).

<Glu-His-Trp-Ser-Tyr⁵-Gly-Leu-Arg-Pro-Gly¹⁰-NH₂

14

The sequence of gonadoliberin is formed initially as a part of the precursor protein (prepro-GnRH) consisting of 92 aa. The release of GnRH from the precursor requires C-terminal amidation by peptidylglycine α -amidating monooxygenase. GnRH occurs not only in the hypothalamus but also in the brain, liver, heart, pancreas, kidneys and adrenal glands, gonads, and small intestine. Many thousands of analogues have been synthesized and tested biologically. Several superagonists (e.g., leuprolide, nafarelin, buserelin) have found clinical application for disorders such as prostate cancer or endometriosis, and many antagonists are undergoing evaluation as both male and female contraceptive agents. Antagonists of the “third generation” are currently in clinical trials for induced hormone suppression, e.g., in sex steroid-dependent benign and malignant diseases, and for premature LH surges in assisted reproduction. Members include *cetrorelix*, Ac-[D-Nal¹, D-p-Cl-Phe², D-Pal³, D-Cit⁶, D-Ala¹⁰]GnRH (SB-75) and *detirelix*, Ac-[D-Nal¹, D-p-Cl-Phe², D-Trp³, D-Harg(Et₂)⁶, D-Ala¹⁰]GnRH. GnRH was first isolated by Andrew Schally and coworkers [233–235].

3.3.2.3 Corticoliberin

Corticoliberin (CRH) **15** is formed in the hypothalamus, released into the capillaries of the portal vessels and transported to the anterior pituitary where it stimulates the production of POMC, a polypeptide containing the sequences of ACTH and several other hormones. Figure 3.16 shows the hormone control circuit for the synthesis of adrenocorticotropin (ACTH), including the feedback regulation.

H-Ser¹-Glu-Glu-Pro-Pro⁵-Ile-Ser-Leu-Asp-Leu¹⁰-Thr-Phe-His-Leu-Leu¹⁵-Arg-Glu-Val-Leu-Glu²⁰-Met-Ala-Arg-Ala-Glu²⁵-Gln-Leu-Ala-Gln-Gln³⁰-Ala-His-Ser-Asn-Arg³⁵-Lys-Leu-Met-Glu-Ile⁴⁰-Ile-NH₂

15

The post-translationally released ACTH is secreted into the plasma, interacts with receptors at the surface of the adrenal lobe, and stimulates especially the synthesis of glucocorticoids. ACTH inhibits the release of CRH, and the glucocorticoids inhibit the release of both ACTH and CRH. Furthermore, dopaminergic neurons originating in the hypothalamus have major regulatory function on the synthesis and release of POMC. CRH occurs in nerve fibers throughout the brain, but it has been also found in the adrenal medulla, pancreas, placenta, and testes. As shown above, it stimulates the synthesis and release of POMC in the adenohypophysis and placenta, which is subsequently degraded to form, for example, corticotropin (ACTH), β -endorphin, and melanotropins. CRH also functions as thyrotropin-releasing factor in non-

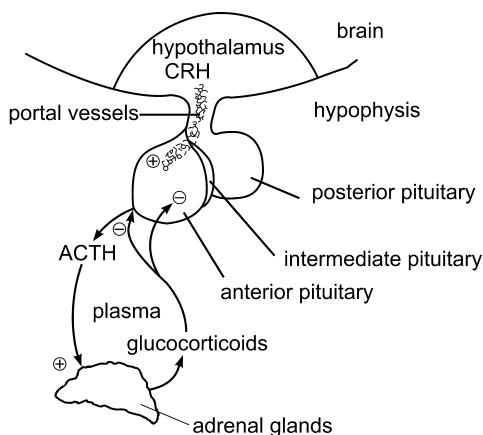


Figure 3.16 Representative example for the regulatory relationship between the hypothalamus, anterior pituitary (AP), the target tissue (adrenal cortex) demonstrated for the axis corticotropin-releasing hormone (CRH), adrenocorticotropin (ACTH), and adrenocortical hormones. + = stimulation, – = inhibition.

mammalian species. In vertebrates exist four paralogous genes that encode CRH, urocortin/urotensin 1, urocortin 2, or urocortin 3, comprising the CRH family [236]. CRH is the central trigger of the hypothalamic-pituitary-adrenal axis and, together with the related peptides urocortin 1, 2, and 3, it is involved in the regulation of behavioral, endocrine, autonomic, cardiovascular, reproductive, metabolic, gastrointestinal and immune systemic activities. The 41-poly peptide amide shows structural similarities to sauvagine from the skin of *Phyllomedusa sauvagii*, and urotensin I from the urophysis of the white sucker (*Catostomus commersoni*), respectively.

3.3.2.4 Growth Hormone-Releasing Hormone

Growth hormone-releasing hormone (GHRH) **16**, according to IUPAC-IUB recommendation also named somatoliberin or somatotropin-releasing hormone (SRH), is released from neurosecretory terminals in the median eminence and stimulates the synthesis and release of growth hormone (somatotropin) [179]. Human GHRH is synthesized as part of the precursor prepro-hGHRH. The GHRH receptor belongs to the family of seven transmembrane receptors coupled to a heterotrimeric GTP binding protein. GHRH is very important for the treatment of children with growth hormone deficiency. Structure–activity studies led to the conclusion that GHRH can be shortened at the C-terminus. In children, hGHRH-(1–29)-amide still has 50% of the activity of the native hormone. Somatostatin (cf. Section 3.3.1.4) is the corresponding release-inhibiting hormone.

H-Tyr¹-Ala-Asp-Ala-Ile⁵-Phe-Thr-Asn-Ser-Tyr¹⁰-Arg-Lys-Val-Leu-Gly¹⁵-Gln-Leu-Ser-Ala-Arg²⁰-Lys-Leu-Leu-Gln-Asp²⁵-Ile-Met-Ser-Arg-Gln³⁰-Gln-Gly-Glu-Ser-Asn³⁵-Gln-Glu-Arg-Gly-Ala⁴⁰-Arg-Ala-Arg-Leu-NH₂

3.3.3

Pituitary Hormones

The secretion of pituitary hormones, also termed trophic hormones, is controlled by the hypothalamus (cf. Figure 3.15). The pituitary hormones (Table 3.9) are released into the bloodstream and stimulate their corresponding glands or tissues (adrenal cortex, thyroid, testes/ovaries, liver, and other special tissues) to secrete the appropriate endocrine hormones. Alternatively, they may exert their biological effect directly, as shown by growth hormone.

3.3.3.1 **Growth Hormone**

Growth hormone (GH), also named somatotropin, somatotrophic hormone (STH), is a single-chain proteohormone ($M_r \sim 21.5$ kDa) produced by somatotrophs of the adenohypophysis [237–239]. Until recently, the consensus was that its synthesis is regulated by the hypothalamic growth hormone-releasing hormone (GHRH), and the appropriate release-inhibiting hormone somatostatin, some being also produced peripherally. Additionally, secretion of GH is stimulated by ghrelin acting via the growth hormone secretagogue (GHS) receptor (GHS-R). Furthermore, in some species, GH secretion is influenced by other peptides/proteins synthesized in the hypothalamus and in the periphery, such as insulin-like growth factor 1 (IGF-1), leptin, pituitary adenylate cyclase-activating polypeptide (PACAP) and thyroliberin (TRH). GH is an anabolic hormone that regulates the metabolism of proteins, sugars, fats and minerals in mammals. Human GH consists mainly of a four-helix bundle (191 aa; two disulfide bridges). The binding of GH activates its receptor (620 aa) to stimulate growth and metabolism in muscle, bone, and cartilage cells. The receptor consists of an N-terminal extracellular ligand binding domain (hGHbp, 238 aa), a single (probably helical) transmembrane segment, and a C-terminal cytoplasmic domain. Human GH causes its receptor to dimerize. This ligand-induced dimerization has important implications for the mechanism of signal transduction. The ligand-induced receptor dimerization is important for the specific mechanism of signal transduction, which might also be verified for other growth factors. GH promotes the longitudinal growth and regeneration of bone, and a mainly tumor-promoted overproduction of GH causes excessive growth (gigantism). In adults GH influences only the growth of soft tissues, and this results in enlarged hands and feet, as well as thickened facial features (acromegaly). Insufficient growth (dwarfism) based on GH deficiency in the period before skeletal maturity can only be treated by application of human GH, as animal GH is without effect in humans. Recombinant hGH is used in the therapy of hypophyseal dwarfism, in the treatment of Turner's syndrome, in the healing of wounds, and for increasing the mass and density of bones.

3.3.3.2 **Corticotropin**

The 39-peptide hormone corticotropin (adrenocorticotropin, adrenocorticotrophic hormone, ACTH) 17, is synthesized in response to stimulation by corticoliberin, and stress. An opposite action in the sense of a corticostatin is exerted by the

Table 3.9 Hypophyseal hormones.^a

Name ^b	Synonym	Abbr.	Major Effects	Structure
Thyrotropin	Thyroid stimulating hormone	TSH	Synthesis and secretion of the thyroid hormones	Glycoprotein ^c α chain: 96 aa β chain: 112 aa
Lutropin	Luteinizing hormone	LH	Acts together with FSH in stimulation growth of the gonads and the synthesis of the sex hormones	Glycoprotein ^c α chain: 96 aa β chain: 121 aa
Follitropin	Follicle-stimulating hormone	FSH	Responsible for development and function of male and female gonads; promotes spermatogenesis in male testes, and controls maturation of female follicle	Glycoprotein ^c α chain: 96 aa β chain: 111 aa
Prolactin	Lactotropin	PRL	Increases in female mammals the milk production, and initiates maternal behaviour	Single chain protein ^c (198 aa)
	lactogenic hormone luteotropic hormone luteotropin	LTH		
Somatotropin	Somatotropic hormone	STH	Promotes species specific the longitudinal growth and bone regeneration; stimulates synthesis of somatomedins	Single chain protein ^c (191 aa) 2 disulfide bonds
	Growth hormone	GH		
Lipotropin	Lipotropic hormone	LPH	Stimulates lipid metabolism, especially, fatty acid mobilisation from lipid depots	Single chain protein β-LPH: 91 aa
Corticotropin	Adrenocorticotropic hormone	ACTH	Stimulates in the adrenal cortex the synthesis of glucocorticoids and mineralocorticoids	Single peptide chain: 39 aa
Melanotropin	Melanocyte stimulating hormone	MSH	Causes dispersion of melanin in the melanophores of the skin of cold-blooded vertebrates; α-MSH is present extensively in th CNS of humans and other animals	Mammalian α-MSH ^d : 13 aa β-MSH: 18–22 aa

^aNot included: oxytocin and vasopressin which are formed in hypothalamus and stored in the posterior pituitary.

^bAccording to IUPAC–IUB recommendation.

^cHuman proteins.

^dOccurs primarily in the hypophyseal pars intermedia and in the hypothalamus.

corticotropin release-inhibiting factor. The latter is a 22-peptide (FIDPELQRSW¹⁰EEKEGEGVLM²⁰PE) from the cryptic region of the prepro-thyrotropin releasing hormone [prepro-TRH-(178–199)] assumed to fulfill the criterion of the endogenous CRIF [240, 241]. ACTH is formed as a proteolytic cleavage product of proopiomelanocortin (POMC) in the adenohypophysis (cf. Figure 3.16). POMC contains seven different peptide hormones, among them corticotropin and melanotropin. The tissue-specific post-translational processing of POMC in the anterior and intermediate lobes of the pituitary provides the signal sequence, the N-terminal fragment, corticotropin, and β -lipoprotein (β -LPH), whereas in the intermediate lobe only further proteolytic cleavage occurs, yielding γ -MSH (from the N-terminal fragment), α -MSH and corticotropin-like intermediate lobe peptide (CLIP) from corticotropin, as well as β -endorphin together with γ -lipotropin (γ -MSH) from β -LPH. ACTH is the key mediator of pituitary-dependent regulation of adrenal steroidogenesis. It stimulates a Ca^{2+} -dependent process of the synthesis of glucocorticoids and mineralocorticoids. Native ACTH and active analogues are used as a diagnostic aid for adrenal cortex function, and also as therapeutic agents in the treatment of several disorders, such as insufficient function of the adrenal cortex, multiple sclerosis, collagen diseases, inflammatory rheumatic diseases, radicular pain syndrome, etc. Recently, it has been reported that ACTH-(1–24) inhibits the proliferation of adrenocortical tumors *in vivo* [242, 243].

H-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly¹⁰-Lys-Pro-Val-Gly-Lys¹⁵-Lys-Arg-Arg-Pro-Val²⁰-Lys-Val-Tyr-Pro-Asn²⁵-Gly-Ala-Glu-Asp-Glu³⁰-Ser-Ala-Glu-Ala-Phe³⁵-Pro-Leu-Glu-Phe-OH

17

3.3.3.3 Melanotropin

Melanotropin, also named melanocyte-stimulating hormone (MSH), or melanocortin, is a peptide hormone produced in the pars intermedia of the hypophysis under the control of melanoliberin and melanostatin. Species lacking the pars intermedia (e.g., chicken, whale) produce it in the neurohypophysis. In addition, MSH is formed by neurons of the CNS. Three different MSH sequences are parts of the biosynthetic precursor protein proopiomelanocortin (POMC), and are released during the processing of POMC in the intermediate lobe (Table 3.10). All three peptides have been highly conserved during evolution, but their exact biological function in mammals is

Table 3.10 Primary structures of melanocortin peptides.

Name	Origin	Sequence
α -MSH	mammalian	Ac-SYSMEHFRWG ¹⁰ KPVa
β -MSH	human	DEGPYRMEHFRWGSPPKD
γ_1 -MSH	rat	YVMGHFRWDR ¹⁰ Fa
γ_2 -MSH	rat	TVMGHFRWER ¹⁰ FG
γ_3 -MSH	rat	YVMGHFRWDR ¹⁰ FGPRNSSSAG ²⁰ GSAQ

All derived melanocortin peptides contain the common amino acid motif HFRW.

still largely obscure. α -MSH is the most potent naturally occurring melanocortin peptide. The melanocortins exert their effects by activating seven-transmembrane domain G protein-coupled receptors MC1R to MC5R. They have a wide and varied distribution and occur in the brain, adrenal glands, circulating leukocytes and resident peritoneal macrophages. The receptors are coupled to adenylyl cyclase forming cAMP after activation. α -MSH exerts pleiotrophic functions, including the modulation of a wide range of inflammatory stimuli; this would be consistent with a cytoprotective role for this hormone in protecting skin cells from exogenous stress, as would occur following UV exposure or exposure to agents inducing inflammation or oxidative stress. α -MSH also modulates both cutaneous and uveal melanoma cell behavior. Superpotent analogues of α -MSH have been described, for example, *melanotan I* (MTI), Ac-Ser-Tyr-Ser-Nle-Glu-His-D-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH₂, and *melanotan II* (MTII), Ac-Nle-c-[Asp-His-D-Phe-Arg-Trp-Lys]-NH₂, which have been tested clinically for investigations on tanning of the skin (MTI) and for the diagnosis and treatment of male erectile dysfunction (MTII). Human β -MSH shows the same biological activity as α -MSH in the Anolis skin test, but β -MSH from other species has lower activity. Three different species of γ -MSH are generally described. In dependence on the target tissue, melanocortin receptor activation leads to various physico-pharmacological effects, such as inhibition of food intake and activation of metabolic rate mediated by MC4R, and skin pigmentation mediated by MC1R [244–246]. In summary, the *melanocortin system* (MCS) consists of the melanocortin peptides, a family of five melanocortin receptors, corticotropin, the endogeneous melanocortin antagonists agouti protein and agouti-related protein. In addition, with mahagony and syndecan-3 two ancillary proteins have been discovered which modulate the activity of the melanocortin peptides [247, 248].

3.3.4

Neurohypophyseal Hormones

The neurohypophysis is the posterior lobe of the pituitary, and is anatomically distinct from the adenohypophysis. It secretes oxytocin and vasopressin (Figure 3.17), two homologous heterodetic cyclic peptides that are primarily synthesized in the hypothalamus.

3.3.4.1 Oxytocin

Oxytocin (OT) is synthesized in hypothalamic cells which project either to the neurohypophysis or to sites within the CNS. Synthesis occurs in the hypothalamus

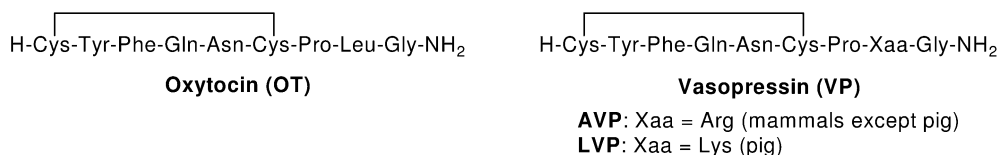


Figure 3.17 Neurohypophyseal hormones.

together with the precursor protein neurophysin I. The precursor protein is transported via the tractus paraventriculo-hypophyseus to the posterior lobe. It is released proteolytically from the precursor in this storage site in response to an appropriate biological stimulation, and secreted into the bloodstream. The neurohypophyseal OT release has long been associated with uterine smooth muscle contraction during labor and milk ejection during lactation. Cholesterol and Mg^{2+} probably function as allosteric modulators. OT also plays an important role in many other reproduction-related functions, e.g., control of the estrous cycle length, ovarian steroidogenesis, and follicle luteinization in the ovary. Furthermore, in the male, OT is a stimulator of spontaneous erections in rats and is involved in ejaculation. However, the function of intracerebral OT remains unclear. It has been suggested that brain OT influences the formation of social bonds and bonding behaviors, and is involved in the modulation of the neuroendocrine reflexes. The human OT receptor (388 aa) is a typical class I G protein-coupled receptor primarily coupled via G_q proteins to phospholipase C- β . Based on the structural similarities to vasopressin, OT shows some vasopressin activity. A huge number of analogues with prolonged or dissociated biological effects, as well as those with antagonistic properties, have been synthesized. Because OT is likely to be involved causally in preterm labor, the design of synthetic peptide and non-peptide OT antagonists as potential tocolytic agents for the prevention of preterm births is an area of intensive research. Of the many OT antagonists described to date, only one, the combined vasopressin V_{1a} and OT receptor antagonist atosiban has been approved. However, atosiban is far from being an ideal OT antagonist. In 1953, OT was structurally elucidated and one year later chemically synthesized as the first peptide hormone by the Nobel laureate Vincent du Vigneaud [249–251].

3.3.4.2 Vasopressin

Vasopressin (VP) is synthesized by the magnocellular cells, mainly in the nucleus supraopticus, and also in the paraventricular nuclei of the hypothalamus as part of neurophysin II, a precursor protein bearing the VP sequence following the N-terminal signal sequence. In this form VP is transported through the supraoptico-hypophyseal tract to the posterior pituitary, where it is stored and then released proteolytically into the blood after an adequate stimulus. VP shows antidiuretic activity, and in higher doses it increases blood pressure. The structural similarity to oxytocins results in a slight oxytocin effect. VP elicits a variety of responses, both centrally and peripherally. For example, it induces water conservation by the kidney, contributing to regulation of osmotic and cardiovascular homeostasis. VP deficiency causes *diabetes insipidus*. The biological actions of AVP are mediated by three G protein-coupled receptor subtypes: the V_{1a} (vascular), V_{1b} (V_3) (pituitary), V_2 (renal) receptors, and the oxytocin (OT) receptors. AVP V_{1a} receptors occur in many tissues, including the CNS, and mediate the vascular effects of AVP by causing vasoconstriction of the vascular smooth muscle cells. V_{1b} receptors present in the pituitary, pancreas, adrenals and CNS mediate the release of corticotropin by AVP. The commercial drug [1-desamino-D-Arg⁸]VP (DDAVP) has 400 times the effect of native AVP on the kidneys, with a practically negligible effect on blood pressure.

[1-Desamino-penicillamine,2-*O*-methyltyrosine]AVP belongs to the widely used antagonists. Many useful analogues have been synthesized, e.g., those with increased V_1 -antagonistic potency and reduced V_2 agonism (antidiuretic effects), as well as analogues with CNS activity, such as [desGly⁹-NH₂]VP. Sequence determination and the first chemical synthesis was performed by Vincent du Vigneaud in 1953/54, together with those of oxytocin [252–254].

3.3.5

Parathyroid Hormone and Calcitonin/Calcitonin Gene-Related Peptide Family

Hormones are involved in maintaining Ca^{2+} homeostasis, the normal extracellular calcium concentration being ~ 1.2 mM. Ca^{2+} is an essential ion that is involved in many biological processes, for example as a second messenger mediating hormone signals, as a necessary cofactor for a number of extracellular enzymes, as a structure-stabilizing factor for proteins and lipids in cell membranes, cytoplasm, organelles and chromosomes, and as a constituent of bone as hydroxyapatite, $Ca_5(PO_4)_3OH$. Indeed, bones are the main Ca^{2+} reservoir of the body. Besides vitamin D, peptide hormones are also implicated in controlling Ca^{2+} metabolism.

3.3.5.1 Parathyroid Hormone

Parathyroid hormone (PTH), also named parathormone or parathyrin, is an endocrine 84 aa polypeptide hormone which is secreted by the parathyroid gland [255–257]. PTH regulates the metabolism of calcium and phosphate in the body. It is expressed as a 115-peptide precursor (prepro-PTH) and secreted as an 84-peptide in response to subnormal serum calcium levels, but the major functions are associated with the N-terminal 34-residue sequence part (hPTH-1-34: SVSEIQLMHN¹⁰LGKHLNSMER²⁰VEWLRKKLQD³⁰VHNF). PTH increases the Ca^{2+} serum level by promoting its resorption from bone and kidney, as well as increasing its dietary absorption from the intestine. PTH stimulates bone resorption by osteoclasts, and also inhibits collagen synthesis by osteoblasts. Furthermore, it promotes the production of the active form of vitamin D in the kidney that, in turn, increases the export of intestinal Ca^{2+} to the blood. PTH displays essentially the opposite effect of calcitonin. The actions of PTH are mediated via a single G protein-coupled receptor, now referred to as the common PTH/PTHrP receptor, or PTHR1. Nowadays, beside the recombinant native hormone (rhPTH-(1–84) [Preos[®]], two very interesting fragments/analogues (rhPTH-1–34) [Forteo[®]] and [Leu²⁷]cyclo-(Glu²²-Lys²⁶)hPTH-(1–31)NH₂ [Ostabolin-C[®]] are available. The PTH analogues treat osteoporosis by stimulating bone formation and strengthening bone micro-architecture in humans. Forteo[®], from Eli Lilly, has been available commercially for some years, while Preos[®] from NPS Pharmaceutical has recently completed Phase III clinical trials and should be available shortly. The analogue Ostabolin-C[®], from Zelos Therapeutics, is currently in Phase II clinical trials. These products are versatile peptides. Studies have been started using these compounds in cancer patients as a novel tool to treat bone marrow depletion caused by chemotherapeutic drugs and ionizing radiation.

3.3.5.2 Parathyroid Hormone-Related Peptides

Parathyroid hormone-related peptides (PTHrP) are a family of peptide hormones produced by almost all tissues in the body. PTHrP acts as a paracrine regulator in several tissues, and plays a central role in the physiological regulation of bone formation by promoting the recruitment and survival of osteoblasts. The human gene is expressed to three PTHrH isoforms of 139, 141, and 173 aa. In analogy to the parathyroid hormone most of the known biological functions are exerted by the N-terminal PTHrP-(1–34), AVSEHQLLHD¹⁰KGKSIQDKRR²⁰RFFLHHLIAE³⁰IHTA, with 60% sequence similarity to PTH-(1–34). The PTH/PTHrP receptor (PTHR1) binds both PTHrP and parathyroid hormone (PTH), which activates by coupling to multiple G proteins both adenylate cyclase/protein kinase A- and phospholipase C/protein kinase C-dependent signaling cascades. PTHrP analogues which strongly inhibit PTHrP adenylate cyclase-stimulation seem to be useful for the treatment of malignancy-associated hypercalcemia in animal trials, but failed in human tests [258, 259].

3.3.5.3 The Calcitonin/Calcitonin Gene-Related Peptide Family

The calcitonin/calcitonin gene-related peptide family (Figure 3.18) comprises calcitonin (CT), calcitonin gene-related peptides (CGRP1 and 2), amylin, adrenomedullin (ADM), and intermedin/adrenomedullin-2 (IMD). The biological actions of the members of this peptide family are mediated through binding to the two closely related type II G protein-coupled receptors (GPCRs), the calcitonin receptor, and the calcitonin receptor-like receptor (CRLR). The peptides of the calcitonin/calcitonin gene-related family show overlapping biological effects owing to their structure and cross-reactivity between receptors.

Calcitonin is a 32 aa peptide hormone with important functions in calcium homeostasis and bone remodeling. CT from various species contains an N-terminal Cys¹–Cys⁷ disulfide bridge and a C-terminal proline amide. Amino acid residues conserved across species are Cys¹, Leu⁴, Ser⁵, Thr⁶, Cys⁷, Gly²⁸, and, furthermore, both Leu⁹ and Leu¹⁶ are common to 11 of 12 known sequences. CT is widely used therapeutically in the treatment of bone disorders. The physiological function of CT is to maintain skeletal mass during periods of calcium stress, such as during growth, pregnancy, and lactation. Furthermore, it plays a central role in controlling calcium homeostasis and maintaining serum calcium without significant fluctuations. Extensive studies on structure–activity relationships have been carried out to understand the structural basis for the activity of this hormone. Interestingly, elcatonin, [Asu^{1–7}]eel calcitonin, is an analogue of eel calcitonin in which the disulfide bridge is replaced by an ethylene bridge. It shows full biological activity in comparison to human CT. This analogue and salmon CT are widely used clinically because of their superior potency. CT and special analogues have been found safe for the treatment of a number of bone disorders, such as Paget's disease, Sudeck's atrophy, high-turnover osteoporosis, and hypercalcemia caused by malignancy. Salmon calcitonin is a well-tolerated peptide drug with wide therapeutic importance. Beside its administration parenterally for long-term treatments, suitable oral formulation should facilitate the application of sCT in

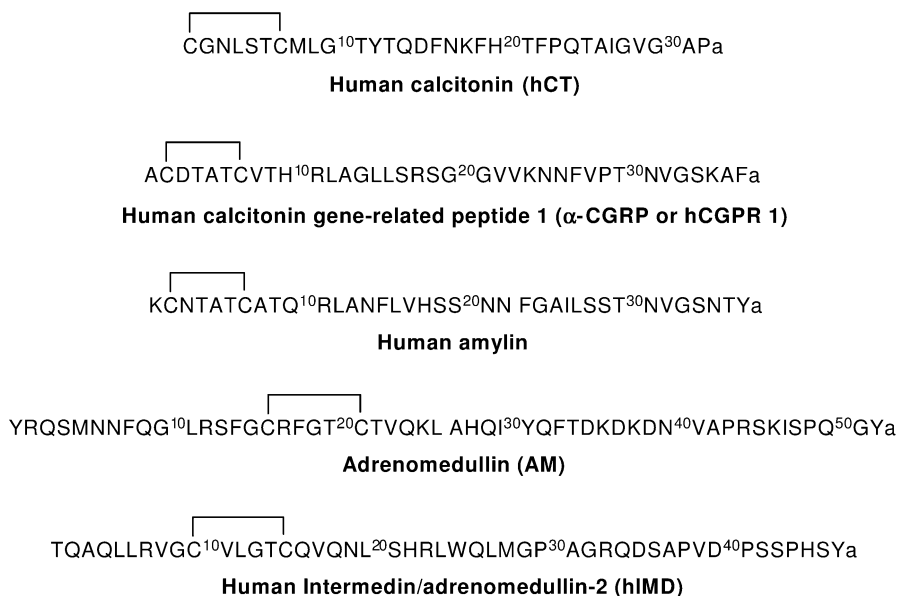


Figure 3.18 Calcitonin/calcitonin gene-related peptide family.

the treatment of osteoporosis and other bone diseases. Studies along this line are in progress. Especially, second-generation analogues of CT with reduced side effects and new dosage forms (nasal and, potentially, oral) will enhance the usefulness of calcitonin therapy [260, 261].

Calcitonin gene-related peptide (CGRP) is a 37-peptide and was discovered by alternative RNA processing of the calcitonin (CT) gene in 1982 [262]. It is a product from alternative splicing of the primary RNA transcript of the CT/CGRP gene. h- β -CGRP or hCGRP2 differs from the homologous peptide h- α -CGRP (cf. Figure 3.18) by three residues (Asn³/Met²²/Ser²⁵). In the CNS, splicing of the CT/ α -CGRP gene produces α -CGRP, whereas in the C cells of the thyroid gland CT is formed. CGRP is widely distributed in the nervous system and in the cardiovascular system. Besides these areas, the CGRP receptor is also present in the adrenal and pituitary glands, kidney, bone, and exocrine pancreas. CGRPs have been reported to be involved in the regulation of responses to sensory stimuli and to play a role in the regulation of the dilation of blood vessels and heart contractility [263, 264].

Amylin, also named islet amyloid polypeptide (IAPP), diabetes-associated peptide (DAP), is a 37-peptide amide first isolated from an insulinoma and from pancreatic amyloid deposits of patients with non-insulin-dependent (type II) diabetes mellitus (NIDDM). Amylin shares 46% sequence homology with the two calcitonin gene-related peptides and 20% with human calcitonin. It may be involved in the pathogenesis of type II diabetes by deposition as amyloid within the pancreas, which

leads to β -cell destruction. Amylin has a vasodilatory effect like other members of the calcitonin/calcitonin gene-related family. It also acts as an agonist at the calcitonin receptor, but efficient signaling of amylin requires the formation of a receptor complex of the calcitonin receptor with receptor activity-modifying proteins (RAMPs) [265].

Adrenomedullin (AM) is a vasoactive 52-peptide amide and shares 24% sequence homology with CGRP. AM was first discovered in human pheochromocytoma tissue in 1993 [266], and later found in the normal adrenal medullae, kidneys, lungs, and blood vessels. The human gene contains four exons separated by three introns and codes for a longer preprohormone of 185 aa, which is processed post-translationally, originating AM and proadrenomedullin N-terminal 20-peptide (PAMP). Both peptides participate in many physiological functions, including vasodilatation, bronchodilatation, neurotransmission, regulation of hormone secretion, brain functions, renal homeostasis, and antimicrobial activities. Recently, it has been reported that AM and PAMP might be potent inducers of angiogenesis which is required for the maintained growth of solid tumors [267, 268].

Intermedin/adrenomedullin-2 (IMD) was initially isolated from the puffer fish, *Fugu rubripes* [269]. The sequence is conserved across species including human, rat and mouse [270]. IMD is expressed in the gastrointestinal tract, pituitary, kidney, lung, thymus, submaxillary glands, and brain. It has been reported that IMD functions as a pituitary paracrine factor regulating prolactin release. From *in vivo* studies it has been concluded that IMD treatment led to blood pressure reduction in both normal and spontaneously hypertensive rats, mediated via CRLR/RAMP receptor complexes. An inhibitory effect on food intake and gastric emptying caused by IMD has also been described.

3.3.6

The Blood Pressure Regulating Peptide Families

Blood pressure-regulating peptides possess a wide variety of physiological actions. A huge number of peptides, including the angiotensin–kinin system, endothelins, angiotensin II, and [Arg]vasopressin stimulate vasoconstriction and thus increase blood pressure. They are involved in homeostasis together with vasodilators such as plasma kinins, substance P, and atrial natriuretic peptide. In this chapter only some main families which are involved in blood pressure regulation are discussed briefly.

Other peptides, such as tachykinins, vasopressin, oxytocin, neuropeptide Y, which also influence blood pressure are found in other chapters.

3.3.6.1 Angiotensin–Kinin System

The angiotensin–kinin system plays an important role in the regulation of blood pressure. The tissue hormones belonging to the classes of the angiotensins and kinins are released into biological fluids from precursor proteins by the action of proteases.

Table 3.11 Primary structures of plasma kinins.

Name	Sequence	
Bradykinin (BK)	Kinin-9	H-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-OH
Kallidin	Kinin-10	H-Lys-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-OH
Methionyl-lysyl-bradykinin	Kinin-11	H-Met-Lys-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-OH

Kinins are peptide hormones implicated in many physiological and pathological processes, including reduction of blood pressure and regulation of sodium homeostasis, inflammation and the cardioprotective effects of preconditioning. The plasma kinins (Table 3.11) comprise bradykinin, kallidin, and methionyl-lysyl-bradykinin. The kallikrein-kinin system (KKS) of the plasma generates bradykinin, whereas the tissue KKS is responsible for the formation of kallidin. The kinin action is mediated via two types of kinin receptor, the type 1 (B_1) and type 2 (B_2). It has been reported that endogenous kinin peptides play a role in the regulation of coronary vascular tone, and in mediating the hypotensive effect of ACE inhibition [271].

Bradykinin causes dilation of blood vessels, resulting in a decrease in blood pressure, and promotes contraction of the smooth muscles of the bronchia, intestine and uterus. The B_1 receptor, mainly involved in inflammatory processes, binds [des-Arg⁹]bradykinin more strongly than bradykinin, whereas the B_2 receptor shows a greater affinity to the native hormone. The degradation of bradykinin occurs primarily in the lungs by kininase II (identical to ACE), but also by other endopeptidases and exopeptidases. Enzymatically stable agonists for the B_2 receptor are potential therapeutics for infarct treatment, whereas antagonists of bradykinin are potential agents for the treatment of inflammation. Stabilized antagonists have also been developed and tested for the treatment of small cell lung carcinoma and prostate carcinoma [272, 273]. Bradykinin and various analogues such as [Thr⁶]bradykinin and C-terminally shortened or lengthened bradykinins have been isolated from the skin of amphibia. *Bradykinin-potentiating peptides* (BPP) are peptides of different structure and from different natural sources, which are able to potentiate the bradykinin actions in various systems [274]. First, the inhibition of angiotensin-converting enzyme (ACE) was assumed to be the main mechanism of BPP potentiating actions. However, besides ACE inhibition, inhibition of other bradykinin-degrading enzymes, the influence on the bradykinin receptor, on the signal pathways and crosstalk with pathways of other hormones were also observed.

Angiotensins (AT) (Table 3.12) are tissue peptide hormones occurring both in the periphery and in the brain, with influence on blood pressure [275, 276]. The source of AT is angiotensinogen, a plasma protein ($M_r \sim 60$ kDa) of the α_2 -globulin fraction which is initially cleaved by the kidney aspartyl protease renin yielding the inactive AT I. AT II is formed by proteolytic cleavage of AT I catalyzed by angiotensin-converting enzyme (ACE). ACE is a zinc metallopeptidase with great importance in the regulation of blood pressure as well as fluid and salt balance in mammals. The dipeptidylcarboxypeptidase ACE catalyzes the conversion of the inactive AT I into

Table 3.12 Primary structures of angiotensins.

Name	Abbr.	Sequence
Angiotensin I	AT I	H-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu ¹⁰ -OH
Angiotensin II	AT II	H-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-OH
Angiotensin III	AT III	H-Arg-Val-Tyr-Ile-His-Pro-Phe-OH
Angiotensin IV	AT IV	H-Val-Tyr-Ile-His-Pro-Phe-OH

the potent vasoconstrictor AT II. Circulating AT II induces vasoconstriction, aldosterone release, sodium and water retention, increases fluid intake, and plays a key role in the regulation of blood pressure and fluid homeostasis. Two AT II receptors, AT₁ and AT₂, with similar binding affinity for AT II, are known. They belong to the superfamily of seven membrane-spanning G protein-coupled receptors. AT II effects, including vasoconstriction, increased aldosterone secretion and sympathetic tone and cardiac and vascular hypertrophy, are predominantly mediated via the AT₁ receptor. However, the functions of the AT₂ receptor have not been clearly characterized. AT II is formed in many tissues, including most peripheral organs and the brain. One-fifth of the adult population suffers from chronic hypertension. The application of AT II inhibition, either by ACE inhibitors or by AT₁ receptor blockade with specific non-peptidic antagonists, for treatment of hypertension, ischemic heart disease, and heart failure, was started during the early 1970s and remains important to the present day. AT II is inactivated in the blood by angiotensinase. AT III is formed by cleavage of the N-terminal Asp of AT II under catalysis by aminopeptidase A. AT IV results from the cleavage of AT III with aminopeptidase N. AT III acts as a central regulator of vasopressin release and blood pressure. It has been reported that angiotensin-converting enzyme 2, acting as a carboxypeptidase, catalyzes the formation of both AT I-(1-9) and AT II-(1-7). The latter is involved in the activation of peripheral vasodilator mechanisms, and shows antitrophic effects mediated by the inhibition of protein synthesis. Furthermore, AT II-(1-7) amplifies the vasodilator actions of bradykinin. Structure–activity relationship studies have led to a vast number of agonists and antagonists being identified, though only two are worthy of mention at this point. The agonist [Asn¹,Val⁵]AT II (Hypertensin[®], CIBA) is administered in collapse or shock situations in order to restore normal blood pressure as quickly as possible. In contrast, the antagonist Saralasin[®] (Sarenin), H-Sar-Arg-Val-Tyr-Val-His-Pro-Ala-OH, has found application in the diagnosis of AT II-dependent forms of hypertonia; it is also used to treat donor kidneys before transplantation in order to minimize loss of function of the transplanted organ as a result of ischemia.

3.3.6.2 Endothelins and Endothelin-Like Peptides

Mammalian endothelins (ET) form with snake venom endothelin-like peptides, also called sarafotoxins (SRTX), a homogenous family of potent vasoconstrictor peptides [277]. Endothelins and sarafotoxins (Table 3.13) act on the vascular system

Table 3.13 Primary structures of endothelins and endothelin-like peptides (sarafotoxins).

Name	Abbr.	Source	Sequence ^b
Endothelin-1	ET-1	<i>Homo sapiens</i> (Man)	CSCSSSLMDKE ¹⁰ CVYFCHLDII ²⁰ W
Endothelin-2	ET-2	<i>Homo sapiens</i> (Man)	CSCSSWLDKE ¹⁰ CVYFCHLDII ²⁰ W
Endothelin-3	ET-3	<i>Homo sapiens</i> (Man)	CTCFTYKDKE ¹⁰ CVYYCHLDII ²⁰ W
Sarafotoxin-a	SRTX-a	<i>Atractaspis engaddensis</i> (Burrowing Asp)	CSCCKDMDKE ¹⁰ CLNFCHQDVI ²⁰ W
Sarafotoxin-b	SRTX-b	<i>Atractaspis engaddensis</i> (Burrowing Asp)	CSCCKDMDKE ¹⁰ CLYFCHQDVI ²⁰ W
Sarafotoxin-c	SRTX-c	<i>Atractaspis engaddensis</i> (Burrowing Asp)	CTCNDMDKE ¹⁰ CLNFCHQDVI ²⁰ W
Sarafotoxin	Btx	<i>Atractaspis bibronii</i> (Burrowing Asp)	CSCADMDKE ¹⁰ CLYFCHQDVI ²⁰ W
Sarafotoxin-e	SRTX-e	<i>Atractaspis engaddensis</i> (Burrowing Asp)	CTCKDMDKE ¹⁰ CLYFCHQGII ²⁰ W
Endothelin-β	VIC ^a	<i>Mus Sp.</i> (Mouse)	CSCNSWLDKE ¹⁰ CVYFCHLDII ²⁰ W

^aVIC = vasoactive intestinal contractor peptide;^bDisulfide bonds: C¹–C¹⁵/C³–C¹¹.

via identical receptors and modulate the contraction of cardiac and smooth muscle in different tissues in vertebrates. This similarity is surprising, since SRTX are strongly toxic components isolated from the venom of snakes of the genus *Atractaspis*, while ET are endogenous hormones of the mammalian vascular system. The latter are produced in picomole amounts by endothelial cells of the cardiovascular systems of vertebrates. ET-1 is formed only in endothelial cells, and is the most abundant member of the family. It acts in an autocrine or paracrine manner as a vasoconstrictor on the vasculature. It is involved in many physiological and pathological processes, including vascular changes associated with pulmonary hypertension and sepsis. ET-2 is released in the epithelial cells of the kidney and intestine, while ET-3 is found in the brain. The receptors selective for ET-1 were designated ET-RA, with agonist potency ET-1 > ET-2 ≫ ET-3, whereas the non-selective receptor (ET-1 = ET-2 = ET-3) was named ET-RB. ET can also relax the muscles, in some cases either by direct receptor activation or indirectly by releasing endothelium-derived relaxing factors, such as NO or eicosanoids. ETs modulate chronotropy, ionotropy, bronchoconstriction, and neurotransmission. Furthermore, they act as regulators for other hormones and neurotransmitters. Intravenous application results in the regulation of blood pressure (both pressor and depressor effects), local blood flow, and kidney function. ETs also play a role in chronic diseases, such as hypertension. For this reason, the development of suitable ET antagonists for medical applications is of major importance. A cyclic pentapeptide isolated from the cultured broth of *Streptomyces misakiensis*, cyclo-(D-Trp-D-Glu-Ala-D-Val-Leu-), has been found to be a potent endothelin antagonist. Starting from this natural lead compound, a synthetic analogue (BQ-123) has been obtained where the inhibitory potency is improved by more than two orders of magnitude [278, 279]. The designation sarafotoxin is derived from the

Hebrew spelling of the species *Atractaspis engaddensis* – *Saraf Ein Gedi* – from which the SRTX were first isolated and characterized [280]. The SRTX belong to the most toxic snake toxins yet described. They show 60% homology with ET. All SRTX contain two conserved disulfide bridges (C^1-C^{15}/C^3-C^{11}) analogously to the endothelins. The SRTX cause strong vasoconstriction of coronary arteries. Death after intoxication with SRTX peptides is the result of cardiac ischemia or infarction [281].

3.3.6.3 Cardiac Peptide Hormones

Natriuretic peptides (NP), more precisely cardiac hormones, comprise a group of six peptides (Figure 3.19) produced by atrial cardiocytes in the heart of mammals exerting natriuretic-diuretic, vasorelaxant and other actions lowering blood pressure and controlling electrolyte homeostasis [282–286]. Beside *atrial natriuretic peptide* (ANP), *brain natriuretic peptide* (BNP) and *C-type natriuretic peptide* (CNP) three additional members of this peptide family are known. The members of the cardiac hormone family are synthesized by three different genes in the heart and then stored as three different prohormones (pro-BNP: 108 aa; pro-CNP: 103 aa, and pro-ANP: 126 aa). Pro-ANP (also termed: atriopeptigen, cardionatrin IV) is secreted from the atria and then processed proteolytically by corin, a type II transmembrane serine protease, to the human circulating 28-peptide ANP, and the N-terminal pro-AFP-(1-98). The latter contains three peptide hormones sequences: *long-acting*

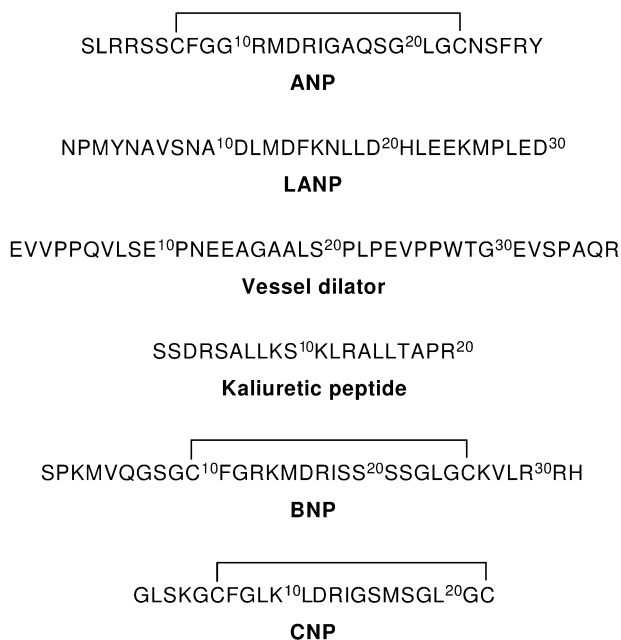


Figure 3.19 Primary structure of human natriuretic peptides.

natriuretic peptide, LANP [pro- AFP-(1-30)], *vessel dilator* [pro-AFP-(31-67)], and *kaliuretic peptide* [pro-AFP-(79-98)] [287, 288]. These three hormones vasodilate blood vessels, and cause a natriuresis and diuresis in animals and humans. BNP, discovered in 1988, is now considered to have been misnamed because it is 10-fold higher concentrated in the heart than in the brain. ANP and BNP are synthesized primarily in the heart muscle cells of the atria (atrial cardiocytes), and also in comparatively smaller amounts in the heart ventricles and in non-cardiac areas, including the CNS, kidneys, and gonads. The sixth member of the cardiac hormones, named C-type natriuretic peptide, was also detected first in the brain in 1990. CNP is synthesized in the vascular endothelium. The main known biological properties of these peptide hormones are blood pressure regulation and the maintenance of plasma volume in animals and humans. The biological effects are mediated via specific natriuretic peptide receptors (NPR) on the surface of the target cells. The receptor NPR-A, a guanylyl cyclase-coupled receptor widely distributed within the body, including vascular smooth muscle, kidneys, adrenal glands, brain and heart, mediates the biological actions of both ANP and BNP. NPR-B, a second guanylyl cyclase-coupled receptor, is responsible for CNP signaling. Cardiac hormones have been shown not only to be powerful prognostic markers in the setting of heart failure, but also to play an important role as a target for therapy. BNP (Nesiritide/Natrecor®) is commercially available and has been widely used in the U.S. for the treatment of acute congestive heart failure. Due to a side effect of Nesiritide, vessel dilator seems to have better conditions for the treatment of chronic heart failure. The measurement of plasma NP levels is an important tool in the diagnosis of chronic heart failure (CHF) and acute coronary syndromes. Furthermore, the therapeutic use of ANP and BNP has opened an active area of inquiry, either by the application of the cardiac hormones themselves or by preventing their degradation. Especially, long-acting analogues of ANP and BNP should have great potential in the treatment of chronic CHF and essential hypertension.

3.3.7

Neuropeptides

In the early 1970s, de Wied [289] coined the term *neuropeptide* for the first time to describe an endogenous substance synthesized in nerve cells and involved in nervous-system functions. Later this definition has evolved from this origin-oriented one to a functional one: neuropeptides are present throughout the central nervous system as well as in peripheral organs such as the pancreas, the adrenal glands, and the cells of the immune system. Neuropeptides are peptides in the range between two and about 50 amino acid residues which fulfil a pivotal function in human life because they are involved in almost all physiological processes [290–292]. Post-translational processing of polypeptide precursors into active neuropeptides by a common mechanism is a usual pathway for their formation. However, some neuropeptides can be produced from a single polypeptide. Additionally, neurons with the same gene encoding a protein are capable of releasing

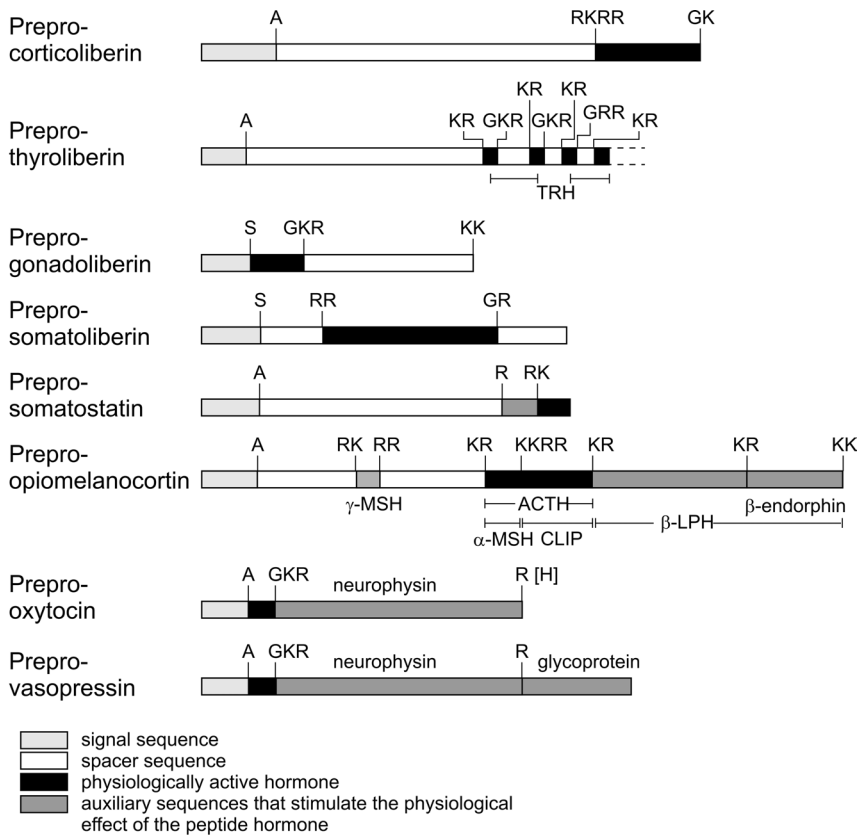


Figure 3.20 Polypeptides as precursors of bioactive peptides.

different neuropeptides due to differences in the way that each neuron processes the polypeptide. Neuropeptides that are synthesized in the form of large protein precursors (Figure 3.20) undergo proteolytic processing to yield the bioactive peptides. The signal peptide sequence is required for vectorial transport across the membranes of the ER. The conformation of the signal peptide, typically a β -turn, is important for the recognition of the exact cleavage site. The eukaryotic signal peptidase is an integral membrane protein with strong sequence similarity to the related enzyme isolated from bacteria. Precursor proteins of opioid peptides, tachykinins, and some releasing hormones have been characterized. Prior to the release of the bioactive peptide, propeptides may be modified by amidation, acylation, or glycosylation. The cleavage of propeptides at pairs of basic amino acids or single basic residues is catalyzed by trypsin-like endopeptidases (prohormone converting enzymes). Post-translational processing can generate a number of active peptides from a single precursor protein [293] which has for example been

demonstrated for POMC [294], this being the common precursor for ACTH, β -endorphin, MSH, corticotropin-like intermediate lobe peptide (CLIP), and related peptides.

Neuropeptides are stored in intracellular vesicles and act as neurohormones upon release into the bloodstream. Neuropeptides released into the synaptic cleft are neuromodulators that inhibit the action of excitatory neurotransmitters, or neuromediators which prolong the action of neurotransmitters. In the extracellular space, neuropeptides are involved both in pharmacodynamic and pharmacokinetic cascades. Neuropeptides interact with selective receptors on the target cells. The homeostasis of neuropeptides is characterized by the equilibrium between active regulation of cellular function via receptor interactions and their own catabolism. With very few exceptions, almost all receptors are coupled to G proteins characterized by seven membrane-spanning α -helices [295]. Nowadays, molecular biology procedures have been used to discover new neuropeptides. For example, calcitonin gene-related peptide was shown to exist as a result of alternative RNA splicing of the calcitonin gene, and the subtractive hybridization approach led to the discovery of the orexins. Furthermore, orphan G protein-coupled receptors have been used as targets for the identification and isolation of their endogenous ligands, e.g., nociceptin for the opioid N/OFQ receptor.

Neuropeptides are capable of exerting multiple effects in the CNS and elsewhere. An enormous number of neuropeptide actions mediated via the different receptor types has been elucidated since 1975. Examples of CNS effects caused by peptides (selected members in brackets) in humans and experimental animals are involved, for example, in:

- *feeding* [β -endorphin, motilin, corticotropin-releasing hormone, NPY, CCK, galanin, GLP-1, neuropeptide FF, bombesin, amylin, somatostatin, galanin, β -casomorphin, angiotensin II, enterostatin, oxytocin, gastrin-releasing peptide, pituitary adenylate cyclase-activating peptide]
- *behavior* [α -MSH, calcitonin gene-related peptide, NPY, melanin-concentrating hormone, vasopressin, neurotensin, CCK-8, oxytocin, ACTH]
- *stress* [atrial natriuretic factor, opioid peptides, NPY, substance P]
- *thermogenesis* [NPY]
- *thermoregulation* [dermorphin, NPY, delta sleep-inducing peptide, vasopressin, α -MSH, neurotensin, cyclo-(His-Pro)]
- *sleep* [delta sleep-inducing peptide, vasopressin, corticotropin-like intermediate lobe peptide]
- *learning* [substance P]
- *memory* [vasopressin, galanin, CCK]
- *aggression* [α -MSH]
- *alcohol uptake* [tachykinin]

- *anorectic effects* [amylin, calcitonin gene-related peptide]
- *anxiety* [α -MSH, motilin, melanin-concentrating hormone, NPY]
- *drinking* [angiotensin II, tachykinins, opioid peptides]
- *pain* [neurotensin, vasopressin, CCK]
- *sexual behavior* [α -MSH, melanin-concentrating hormone]
- *locomotion* [neurotensin, NPY, amylin, vasopressin, melanin-concentrating hormone]

The important effects of peptides in the CNS have been underlined by increasing numbers of publications that have been reviewed for the years 1980–1985 [296], 1986–1993 [297], 1994–1999 [298, 299], 2000–2007 [300–306]. Recently, the 30th issue of the annual review of research concerning the endogenous opioid system that summarizes published papers during 2007 has been published [306]. This excellent volume continues the tradition initiated by Abba Kastin, Gayle Olson, Richard Olson, David Coy, Anthony Vaccarino, and continued by Richard J. Bodnar and colleagues in the reviews spanning from 1978 through 2007.

3.3.7.1 Endorphins

The designation endorphins is the generic term for endogenous opioid peptides. They are opiate-like endogenous peptides with morphine-like activity. The name is related to *endogenous morphine*. Endorphins exert their effects through opioid receptors, which have been discovered throughout the mammalian CNS and peripheral nerve tissues such as guinea pig ileum, the mouse *vas deferens* and human gastrointestinal tract. There are at least three multiple receptor subtypes: μ , δ and κ ; the existence of other receptors, such as ϵ and σ , has been suggested, but has not yet been confirmed. Based on their differing pharmacology, the following subclasses of receptors – δ_1 , δ_2 , κ_1 , κ_2 and κ_3 – have been proposed and, due to their differing location, also the subclasses μ_1 and μ_2 .

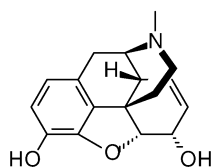
Opiates such as morphine **18** have been widely used by clinicians both for the blockade of severe pain and for anesthesia. Although at present synthetic or semisynthetic morphine analogues such as the phenylpiperidine derivative fentanyl **19** are used, opium extracts have been the natural source for morphine for several thousand years. Indeed, it was surprising to find that nature provided morphine not only as a plant product but also (as structurally related opiates) in mammals, albeit in very small amounts and (presumably) during certain metabolic periods [307].

Collier first postulated the existence of “endogenous morphine” in 1972 at the International Congress of Pharmacology in San Francisco:

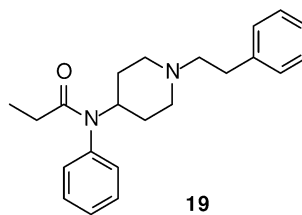
“The receptor for a foreign drug is really the receptor for a humoral substance with which the foreign molecule also interacts.... We do not yet know the natural function of the macromolecule(s) with which morphine interacts.”

The binding of endogenous opioid receptor ligands was independently demonstrated one year later by three groups [308–310], and in 1975 with the enkephalins the first endogenous peptide ligands were discovered by Kosterlitz and coworkers [311], followed by the characterization of four major types of opioid receptors some years later [312, 313].

Enkephalins (EK) were the first isolated endogenous opioid peptides from porcine brain (enkephalos = brain). Methionine enkephalin (Met-enkephalin) **20** and leucine enkephalin (Leu-enkephalin) **21** show a strong antinociceptive effect *in vivo*, comparable to that of the most potent opioids used in medicine such as morphine **18**, fentanyl **19**, and sufentanyl. The enkephalins were found to occur naturally and bind as physiological agonists to the opioid receptors in the brain. They act as the endogenous ligands of the μ , δ and κ opiate receptors in the brain.

**18**

H-Tyr-Gly-Gly-Phe-Met-OH

20**19**

H-Tyr-Gly-Gly-Phe-Leu-OH

21

The enkephalins act both as neurotransmitters and neuromodulators. The application of enkephalins as analgesics has been retarded by their poor stability *in vivo* and by their inability to effectively penetrate the blood-brain barrier (BBB). An effective BBB transport of glycosylated enkephalins has been described by several groups. Many analogues have been synthesized in an attempt to develop new anodynes that would be more potent and less addictive but, until now no definitive alternatives to the application of classical analgesics for pain treatment have been identified [314, 315].

Enkephalins and other so-called “typical” opioid peptides originate from three large precursor proteins, as shown in Figure 3.21. These typical opioid peptides share the N-terminal sequence H-Tyr-Gly-Gly-Phe-, and most of the members bind to more than one type of opioid receptor. The so-called “atypical” opioid peptides originate from a variety of precursor proteins with different N-terminal amino acid sequences and a conserved N-terminal tyrosine residue [316].

The proenkephalin sequence contains four copies of Met-enkephalin, one of Leu-enkephalin, and two extended forms of Met-enkephalin. Prodynorphin (proenkephalin B) is the precursor for α -neoendorphin and the dynorphins, and contains three copies of Leu-enkephalin. Prepro-opimelanocortin, which is transformed into POMC by cleavage of the signal sequence, is the source of ACTH and β -lipotropin. These hormones are cleaved to give smaller bioactive peptides in a tissue-specific manner (Figure 3.22).

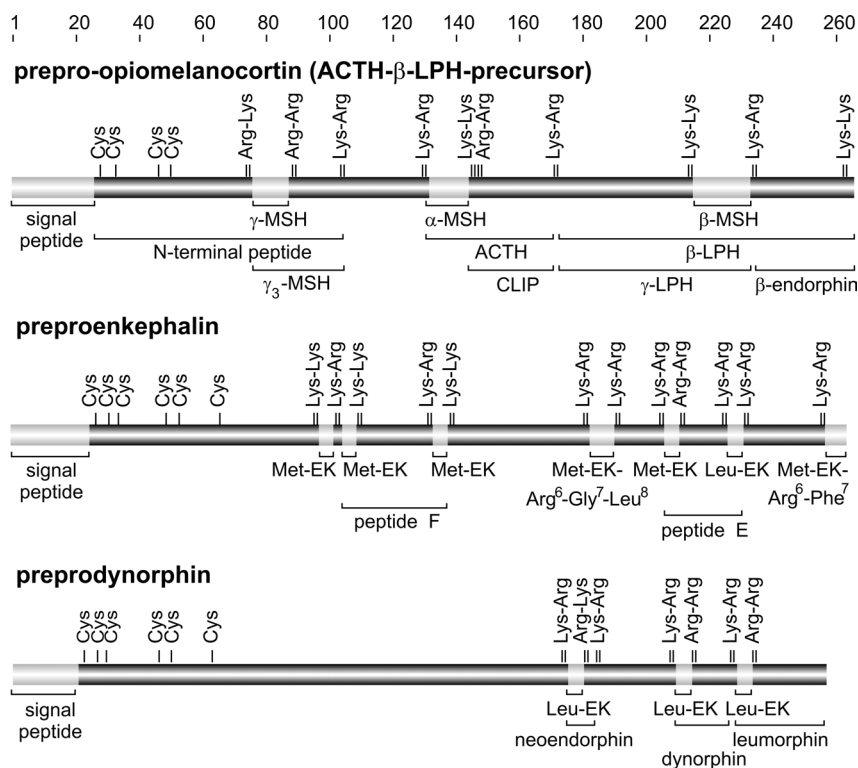


Figure 3.21 Schematic representation of the sequences of opioid peptide precursors (EK = enkephalin).

POMC is proteolytically cleaved in the anterior lobe, yielding the N-terminal fragment, ACTH, and β-lipotropin (β-LPH). However, the latter polypeptide hormones are further cleaved to yield γ-MSH, α-MSH, CLIP, γ-LPH, and β-endorphin in the intermediate lobe only. The resulting range of hormones (Table 3.14) shows different activities, underlining the fact that the peptides formed both in the anterior and intermediate lobes and the corresponding cleavage products of the proenkephalins are physiologically distinct.

The N-terminal tyrosine of the opioid peptides is indispensable for binding to opioid receptors. However, the activation of the peptide receptor complex requires the N-terminal dipeptide sequence which is directly related to the tyramine moiety of morphine **18** or other opiate alkaloids. Besides this dipeptide “message part”, the opioid peptides contain – according to the proposal of Schwyzner – the “address” sequence which is composed of Gly-Phe or Phe that is responsible for their high affinity to opioid receptors. The C-terminal sequence part differentiates the affinity of opioid peptides toward the receptor types. The primary structures of β-endorphin (sequence 61–91 of β-lipotropin) and related sequences of the other endorphin peptides are compiled in Table 3.15.

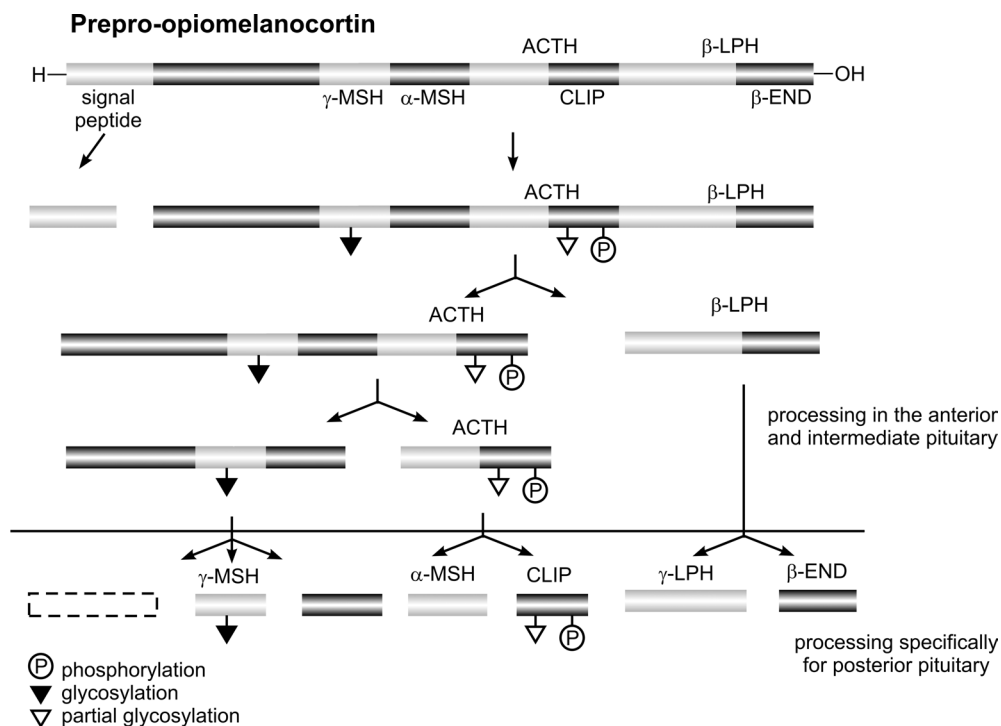


Figure 3.22 Tissue-specific processing of proopiomelanocortin (POMC) in the anterior pituitary and intermediate pituitary.

Morphine **18** shows relatively high affinity toward μ -receptors, where two subtypes (μ_1 , μ_2) have been proposed. Fentanyl **19** similarly binds with high selectivity to the μ -receptor. Naloxone **22**, naltrexone **23**, and nalorphine **24** are the most commonly used opioid antagonists with selectivity toward the μ -receptor.

Table 3.14 Different sets of opioid peptides released from polyprotein precursors.

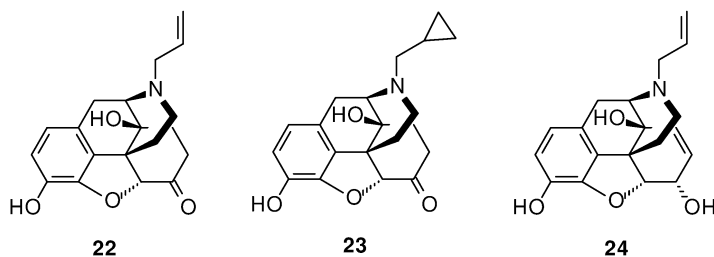
Proopiomelanocortin	Proenkephalin	Prodynorphin
ACTH	Met-enkephalin	Dynorphin A (1–17)
β -Lipotropin ^a	Leu-enkephalin	Dynorphin A (1–8)
γ -MSH	Met-enkephalin-Arg ⁶ -Phe ⁷	α -Neoendorphin ^b
α -MSH	Met-enkephalin-Arg ⁶ -Gly ⁷ -Leu ⁸	Dynorphin B (Rimorphin)
Corticotropin like		Dynorphin B29 (Leumorphin)
intermediate lobe peptide (CLIP)		Leu-enkephalin

^a β -Lipotropin contains β -MSH and the endorphins (α , β , γ , δ).

^b α -Neoendorphin contains β -neoendorphin.

Table 3.15 Primary structures of human endorphins.

Name	Sequence
β -endorphin	YGGFMTSEKS ¹⁰ QTPLVTLFKN ²⁰ AIKNAYKKG ³⁰ E
C-fragment	YGGFMTSEKS ¹⁰ QTPLVTLFKN ²⁰ AIKNAY
δ -endorphin	YGGFMTSEKS ¹⁰ QTPLVTLFK
γ -endorphin	YGGFMTSEKS ¹⁰ QTPLVTL
α -endorphin	YGGFMTSEKS ¹⁰
Met-enkephalin	YGGFM



However, they also show affinity toward δ - and κ -receptors, reversing the effects of agonists on these receptors. δ -Receptor subtypes have been classified into δ_1 and δ_2 or into δ_{cx} and δ_{ncx} , whereas for the κ -receptor the subtypes κ_1 , κ_2 , κ_3 , κ_{1a} , κ_{1b} , κ_{2a} , and κ_{2b} have been described [317]. It has been shown above that three opioid receptors (μ , κ , σ) were postulated even before the discovery of the first endogenous opioids, and the isolation of the first opioid peptides led to the characterization of the δ -receptor. Furthermore, other types of novel, less well-characterized opioid receptors (ϵ , λ , ι , ξ) have also been postulated. However, the σ -receptor is no longer regarded to be an opioid receptor. With very few exceptions, all cloned opioid receptors share the general structure of G protein-coupled receptors (GPCR), with an extracellular N-terminal region, seven transmembrane domains and intracellular C-terminal tail structure. At present, molecular biology procedures have been used to discover new neuropeptides. For example, the subtractive hybridization approach led to the discovery of the orexins. Furthermore, orphan G protein-coupled receptors have been used as targets for the identification and isolation of their endogenous ligands, e.g., nociceptin for the opioid N/OFQ receptor. An enormous number of opioid actions mediated via the different opioid receptor types have been elucidated. They are involved in analgesia, learning and memory, eating and drinking, pregnancy, development and endocrinology, mental illness and neurologic disorders as well as locomotion, respiration and thermoregulation [318–320]. The term endorphin is used in the narrower sense not for all peptides with opiate-like effects, but rather more preferentially to fragments of β -lipotropin. Therefore, dynorphins, and α -neoendorphin would not fit this narrower definition. In principle, the same is also true for the enkephalins. However, from the historical point of view the enkephalins were included in this chapter.

Table 3.16 Primary structures of dynorphins.

Name	Sequence ^a
Dynorphin A (Dyn A)	YGGFLRRIRP ¹⁰ KLKWDNQ
Dynorphin B (Dyn B)	YGGFLRRDFK ¹⁰ VVT
Big dynorphin (Big Dyn)	YGGFLRRIRP ¹⁰ KLKWDNQKRY ²⁰ GGFLRRDFKV ³⁰ VT
α -Neoendorphin	YGGFLRKYPK ¹⁰
β -Neoendorphin	YGGFLRKYP
Rimorphin	YGGFLRRQFK ¹⁰ VVT

^aPorcine.

3.3.7.2 Dynorphins

Dynorphins (Dyn) comprise a family of opioid peptides that are released from the precursor preprodynorphin, as shown schematically in Figure 3.21. The peptides of this family (Table 3.16) are putative endogenous ligands for the κ -opioid receptor (KOR). The sequence of Big Dyn [human pro-dynorphin-(207–238)] contains Dyn A [Big Dyn-(1–17)], and Dyn B [Big Dyn-(20–32)] (Table 3.15). The dynorphins are synthesized in several brain areas, including the striatum, hippocampus, basal ganglia, cerebral cortex, and spinal cord. The dynorphins interact preferentially with KOR, and are critical for regulation of nociceptive transmission, stress-induced responses, and substance dependence. In humans, the dynorphins may be involved in the development of various neuropsychiatric disorders. It has been shown that pro-dynorphin expression is altered in the brain of drug abusers and psychiatric patients. Furthermore, from KOR-mediated effects of salvinorin A it has been suggested that dynorphins may contribute to perceptual distortions in dementia, schizophrenia, and bipolar disorders [321–323]. The longer dynorphins were first detected in the hypothalamus, hypophysis, and duodenum of the pig. Later, they were isolated from adrenal medulla, guinea pig heart, and rat duodenum.

3.3.7.3 Hypocretins (Orexins)

The hypocretins (Hcrt), also known as orexins, are two neuropeptides that are produced in the lateral hypothalamic area [324, 325]. The neurons producing hypocretins are located in the lateral hypothalamic area but project broadly to various parts of the brain, and they have been implicated in the control of energy homeostasis and arousal maintenance. Human (bovine, rat, mouse) hypocretin-1 (orexin-A), and human hypocretin-2 (orexin-B) are derived from the 131-residue preprohypocretin, which shows sequence similarities with members of the secretin family (Figure 3.23). The name hypocretin reflects their hypothalamic origin and the similarity to secretin, but additionally they have also been termed orexins because of the stimulated acute food intake when administered to rats during daytime. In principle, the two terms are interchangeable and are both used extensively in the literature. The hypocretins stimulate appetite and stereotypic behavior associated with feeding. They also play

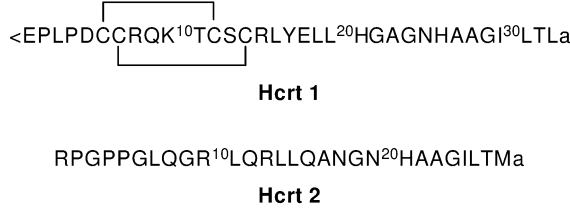


Figure 3.23 Primary structure of the hypocretins.

roles in regulating drinking behavior, neuroendocrine function, and the sleep–wake cycle. The Hcrt were discovered by three groups during the late 1990s. They have been initially identified as endogenous ligands for an orphan G protein-coupled receptor which is broadly expressed in the central nervous system. The initial orphan G protein-coupled receptor, Hcrt1 (also referred to as OX1R) bound Hcrt-1 with high affinity, but Hcrt-2 with 100- to 200-fold lower affinity. A related receptor Hcrt2 (OX2R), sharing 64% identity with Hcrt1, identified by searching database entries with Hcrt-1, showed a high affinity for the two hypocretins. Both receptors are highly conserved (95%) across species. The Hcrt1 (OX1R) receptor shows structural similarities to certain neuropeptide receptors, especially to the Y₂ neuropeptide Y receptor.

3.3.7.4 Dermorphins

The dermorphins [326, 327] (Table 3.17) are potent analgesics in rodents and primates, including humans. Dermorphin (DM) was first isolated from the skin of the frog *Phyllomedusa sauvagei*. Besides the deltorphins, DM was the first example of D-amino acid-containing peptides occurring in animals. Dermorphin is released from a biosynthetic precursor containing several copies of dermorphin with the L-isomer in position 2 only. The conversion of L-Ala into D-Ala is a post-translationally enzyme-catalyzed process. Dermorphin binds preferably to the μ-receptors. In the guinea pig ileum test, DM is 57 times as active as Met-enkephalin and, when given intravenously, it shows a significantly higher analgesic activity than morphine.

Table 3.17 Primary structures of dermorphins.

Name	Source	Sequence
Dermorphin	<i>Phyllomedusa sauvagei</i>	H-Tyr-D-Ala-Phe-Gly-Tyr-Pro-Ser-NH ₂
[Hyp ⁶]dermorphin	<i>P. rhodai</i> , <i>P. burmeisteri</i>	H-Tyr-D-Ala-Phe-Gly-Tyr-Hyp-Ser-NH ₂
[Lys ⁷]dermorphin	<i>P. bicolor</i>	H-Tyr-D-Ala-Phe-Gly-Tyr-Pro-Lys-NH ₂
[Trp ⁴ ,Asn ⁷]dermorphin	<i>P. bicolor</i>	H-Tyr-D-Ala-Phe-Trp-Tyr-Pro-Asn-NH ₂
[Trp ⁴ ,Asn ⁵ (dermorphin-(1–5)]	<i>P. bicolor</i>	H-Tyr-D-Ala-Phe-Trp-Asn-NH ₂

Table 3.18 Primary structures of deltorphins.

Name	Sequence
Deltorphin (Dermenkephalin)	H-Tyr-D-Met-Phe-His-Leu-Met-Asp-NH ₂
[D-Ala ²]deltorphin I	H-Tyr-D-Ala-Phe-Asp-Val-Val-Gly-NH ₂
[D-Ala ²]deltorphin II	H-Tyr-D-Ala-Phe-Glu-Val-Val-Gly-NH ₂

3.3.7.5 Deltorphins

The deltorphins (DT) [328, 329] (Table 3.18) are members of a small family of highly selective δ -opioid 7-peptide amides isolated from the skin of the South American frog *Phyllomedusa bicolor* and *P. sauvagei*. They contain D-Met or D-Ala, respectively, in position 2, and together with the dermorphins they are the first naturally occurring regulatory peptides that contain D-amino acids. Deltorphin is released from a biosynthetic precursor additionally containing four copies of dermorphin. A propeptide from *Phyllomedusa bicolor* contains three copies of [D-Ala²]deltorphin I, and one copy of [D-Ala²]deltorphin II. In contrast to dermorphin, DT binds to δ -receptors. The varying selectivity for the opioid receptors can be explained by charge effects and differences in the hydrophobicity of the C-terminal part of the peptides.

3.3.7.6 Nociceptin/Orphanin and Nocistatin

Nociceptin/orphanin (N/OFQ) is a 17-neuropeptide (Figure 3.24) capable of inducing a variety of biological actions by activating a specific “orphan opioid receptor” (ORL-1, now designated NOP) [330–332]. NOP shares significant homology with classical opioid receptors, but does not bind classical opioid ligands. N/OFQ is involved in the control of several biological activities, not limited to nociception, but also including learning and memory, motivation, stress and anxiety, and the regulation of cardiovascular, hormonal, renal, and intestinal functions. N/OFQ and NOP occur not only in brain and spinal cord, but also in peripheral tissues of several species, such as humans, rat and mouse, where they regulate important functions, e.g., anxiety, analgesia/hyperalgesia, depressive state, and food intake. N/OFQ shows a clear resemblance to dynorphin A. However, dynorphin A is inactive at the NOP receptor, but is a potent agonist of the κ -opioid receptor. The broad pharmacological profile of N/OFQ has initiated the development of novel non-peptidic NOP receptor

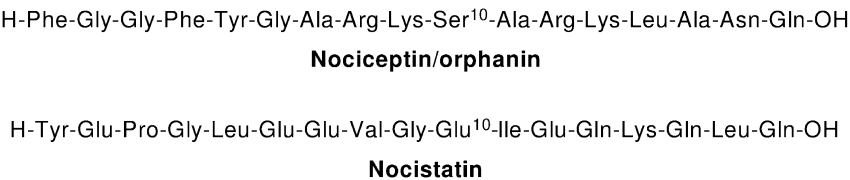


Figure 3.24 Primary structures of nociceptin/orphanin und nocistatin.

ligands, antagonists and agonists which are protease-resistant and bioavailable. N/OFQ is synthesized as part of a larger precursor polypeptide (176 aa) together with the 17-peptide *nocistatin* (Figure 3.24) which play opposing roles in the CNS [333, 334]. Simultaneous administration of nocistatin blocks allodynia and hyperalgesia induced by nociceptin/orphanin. Nocistatin is widely present in the spinal cord and brain, and may play an important function in the CNS, including involvement in nociception, learning, and memory.

3.3.7.7 Exorphins

Exorphins (Table 3.19) are protein-derived opioid peptides of exogenous origin that are mainly formed from food proteins during digestion [335–339]. Preferentially, milk proteins, such as α -casein, β -casein, κ -casein, α -lactoglobulin, β -lactoglobulin and lactoferrin contain partial fragments that behave like opioid receptor ligands under *in vitro* and *in vivo* conditions. The most well-known members of the exorphins are doubtlessly the β -casomorphins (β -CM), also named β -casorphins. These β -casein-derived opioid peptide receptor ligands were the first milk protein-derived opioid peptides acting as receptor ligands. β -CM with opioid activity have been found in human, bovine, ovine, and water buffalo β -casein. They can be released from β -casein both in the adult and neonate organism, where they might exert opioid activities. β -Casomorphin-7 (β -CM-7) was the first example of a peptide with morphine-like activity. In contrast to the moderate potencies and receptor selectivities of the natural β -CM, some of the synthetic analogues display high agonist potencies and remarkable μ -receptor selectivities, e.g. Morphiceptin is the N-terminal tetrapeptide amide of β -CM-7 and had been synthesized prior to isolation from commercial enzymatic digest of casein. It is regarded as a standard μ -selective opioid receptor ligand. α -Casein exorphins, also termed *casoxins*, are derived from bovine α -casein. α -Casein exorphin (1–7) corresponds to the partial sequence 105–111 of bovine α_s -casein, and a fragment shortened by the C-terminal Glu show moderate typical opioid properties

Table 3.19 Primary structures of selected exorphins.

Name	Source	Sequence
β -Casomorphin-7	Casein	H-Tyr-Pro-Phe-Pro-Gly-Pro-Ile
Morphiceptin	Synthetic	H-Tyr-Pro-Phe-Pro-NH ₂
α -Casein exorphin (1–7)	bovine- α_s -Casein	H-Arg-Tyr-Leu-Gly-Tyr-Leu-Glu-OH
Casoxin D (1–7)	human-Casein	H-Tyr-Val-Pro-Phe-Pro-Pro-Phe-OH
Bovine casoxin (1–6)	Synthetic	H-Ser-Arg-Tyr-Pro-Ser-Tyr-OMe
α -Lactorphin	α -Lactalbumin	H-Tyr-Gly-Leu-Phe-NH ₂
β -Lactorphin	β -Lactalbumin	H-Tyr-Leu-Leu-Phe-NH ₂
Gluten-exorphin A5	Wheat gluten	H-Gly-Tyr-Tyr-Pro-Thr-OH
LVV-hemorphin-7	Hemoglobin	H-Leu-Val-Val-Tyr-Pro-Trp-Thr-OH
VV-hemorphin-7	Hemoglobin	H-Val-Val-Tyr-Pro-Trp-Thr-Gln-OH
Hemopressin	Hemoglobin	H-Pro-Val-Asn-Phe-Lys-Phe-Leu-Ser-His-OH

in vitro. Casoxin D (1–7) corresponds to human α _s-casein [158–164]. It is released upon peptic-chymotryptic digestion of a human casein, and was reported to act as an opioid antagonist. From tryptic and peptic digests of bovine κ -casein preparations, two peptides have been isolated showing low opioid antagonist properties. The bovine casoxin (1–6) could be regarded as a low-affinity μ - and κ -selective opioid receptor antagonist. Furthermore, α -lactorphin and β -lactorphin are specific fragments of α -lactalbumin and β -lactoglobulin, respectively, bearing an N-terminal Tyr residue. They have been synthesized and act as μ -opioid receptor ligands with low potency. *Gluten-exorphins* are opioid peptides released from wheat gluten pepsin–elastase digest. The sequence of gluten-exorphin A5 occurs 15 times in the primary structure of wheat gluten. The A5 peptide is selective for δ -receptors. *Hemorphins* [340, 341] are peptides with opioid-like activity found after enzymatic hydrolysis of hemoglobin. The small peptides between seven and ten amino acid residues correspond to fragments 31–35, 31–37, 31–38, 31–40 (each starting with H-Leu-Val-Val), as well as 32–38 and 32–41 (starting with H-Val-Val-) of the bovine hemoglobin β -chain. Interestingly, it has been reported that the generation of VV-hemorphin-7 from globin can be performed by peritoneal macrophages. A putative physiological role of hemorphins has been discussed. The bioactive 9-peptide *hemopressin* [342] is derived from the α ₁-chain of hemoglobin and was initially isolated from rat brain extract. Rat hemopressin has been reported to possess systemic vasopressor activity in rats, to decrease systemic arterial pressure in rabbit and mouse, and to inhibit peripheral hyperalgesia via a naloxone-independent mechanism.

3.3.7.8 The Adipokinetic Hormone/Red Pigment-Concentrating Hormone Family

The adipokinetic hormone/red pigment-concentrating hormone family (AKH/RPCH peptide hormone family) (Table 3.20) comprises peptide hormones produced in the neurosecretory organs of crustaceans and insects, and named after the first fully characterized members and their most prominent functions [343]. These include the aggregation of pigment in the epidermal cells of crustaceans by the *red pigment-concentrating hormone* (RPCH, today denoted as Panbo-RPCH), whereas the *adipokinetic hormones* (AKH) in insects regulate the levels of circulating metabolites such as lipids, carbohydrates and proline by activating phosphor-ylases or lipases in the fat body cell.). The members of this peptide hormone family consist of 8 to 10 amino acid residues bearing both a blocked N-terminus

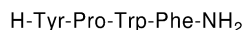
Table 3.20 Primary structures of AKH/RPCH family peptides.

Name (Abbr.)	Sequence
Locmi-AKH-I	pGlu-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly-Thr-NH ₂
Locmi-AKH-II	pGlu-Leu-Asn-Phe-Ser-Ala-Gly-Trp-NN ₂
Locmi-AKH-III,	pGlu-Leu-Asn-Phe-Thr-Pro-Trp-Trp-NH ₂
Panbo-RPCH	pGlu-Leu-Asn-Phe-Ser-Pro-Gly-Trp-NH ₂

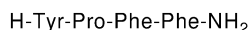
(pyroglutamate residue) and C-terminus (amide), respectively. They are characterized by aromatic amino acids at position four (Phe or Tyr) and eight (Trp). Besides these well-known members of this family, some AKH/RPCH peptides are produced in the brain and not in the retrocerebral corpora cardiaca (CC), for example in the migratory locust and the African malarial mosquito. The complete sequence of the locust's AKH, today denoted as *Locmi-AKH-I*, was elucidated in 1976. Insects contain up to three AKH peptides (isoforms), as demonstrated by the other two peptides produced by the African migratory locust: *Locmi-AKH-II* and *Locmi-AKH-III*. The mobilization of substrates for high-energetic phases is the major function of AKH in insects. However, AKH peptides also play a multifunctional role and exert pleiotropic actions [344]. Beside the members listed above, a huge number of AKHs such as Phymo-AKH (from *Phymateus morbillosus*), Empepe-AKH (from *Empusa pennata*), Manto-CC (*Mantophasmatodea*), and others (Psein-AKH, Grybi-AKH) are known. *Panbo-RPCH* is a crustacean color-change hormone discovered in the eye stalks of the prawn *Pandalus borealis* in 1972 [345, 346]. It was described as the first neuropeptide from invertebrates. The highly conserved RPCH is the only member of the AKH/RPCH family occurring in crustaceans. However, some insects contain the crustacean form Panbo-RPCH. The major biological function of Panbo-RPCH is the aggregation of pigment granules in the epithelial chromatophores.

3.3.7.9 Endomorphins

Endomorphins [347–349] are 4-peptides isolated from the brain showing the highest affinity and selectivity for the μ -opioid receptor of all known endogenous opioids. From the neuroanatomical distribution it follows that the endomorphins reflect their potential endogenous function in many major physiological processes including perception of pain, responses related to stress, and complex functions such as reward, arousal, and vigilance, as well as autonomic, cognitive, neuroendocrine, and limbic homeostasis. *Endomorphin-1* (EM-1) **25** shows 4000- and 15 000-fold preference over the δ - and κ -receptors, respectively, and may be a natural ligand for the μ -receptor. *Endomorphin-2* (EM-2) **26** shows potent μ -selective activity and is well-positioned to serve as an endogenous modulator of pain in its earliest stages of perception. Contrary to EM-2, which is more present in the spinal cord and lower brainstem, EM-1 is more widely and densely prevalent throughout the brain compared to EM-2.



25



26

3.3.7.10 The Allatostatin Families

The allatostatins (AST) [350–354] comprise families of pleiotropic neuropeptides first isolated from the brain of the cockroach *Diploptera punctata*. AST inhibit the synthesis of the juvenile hormone in the *corpora allata*, which regulates insect metamorphosis. Mandibular organs are the crustacean homologs of insect corpora

allata and produce precursors of juvenile hormone with putatively similar functions. Three families of allatostatins in insects have been isolated: (i) the *cockroach type* (> 70 peptides, consensus sequence FGLa), comprising e.g. Dippu-AST 1 (LYDFGLa), Dippu-AST 2 (AYSIVSEYKR¹⁰LPVYNFGLa), Dippu-AST 5); (ii) the *cricket type* (consensus sequence WX₆Wa), with Grybi-AST 1 (H-Gly-Trp-Gln-Asp-Leu-Asn-Gly-Gly-Trp-NH₂) and Grybi-AST 5 (H-Ala-Trp-Asp-Gln-Leu-Arg-Pro-Gly-Trp-NH₂); and (iii) the *Manduca type* (consensus sequence PISCF). Members of the latter family are highly homologous (<EXRZRQCYFN¹⁰PISCF with X = V,I and Z = F,Y) between the species *Manduca sexta*, *Drosophila melanogaster*, and *Anopheles gambiae*. The FGLamides, W(X)₆Wamides, and PISCFs act rapidly and reversibly. However, although these three family types occur in all groups of insects studied, they act as inhibitors of juvenile hormone production in only some groups. Only the FGLamide-type AST have been isolated in crustaceans, in which they may function to stimulate production of hormone by the mandibular glands, as occurs in early cockroach embryos. It has also been proposed that the inhibition of juvenile hormone synthesis appeared secondarily, and that the ancestral function was the modulation of myotropic activity. Further effects of the different AST include endocrine and interneuronal functions, neuromodulatory effects, and direct action on biosynthetic pathways; mostly being species- or order-specific. Picomolar concentrations of the *Drosophila melanogaster* Drome-AST 3 (H-Ser-Arg-Pro-Tyr-Ser-Phe-Gly-Leu-NH₂) from the head of *Drosophila* activate a fruitfly G protein-coupled receptor that shows striking sequence similarities to mammalian galanin and somatostatin/opioid receptors.

3.3.7.11 Neuromedins

Neuromedins (NM) consist of a heterogeneous group of neuropeptides preferentially occurring in mammals (Table 3.21). *Neuromedin B* (NMB) [355] was originally

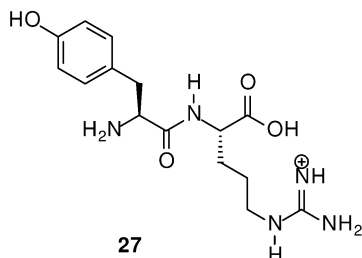
Table 3.21 Primary structures of neuromedins (NM).

Name	Source	Sequence
NMB	<i>Sus</i> Sp. (Pig)	GNLWATGHFMa
NMC	<i>Sus</i> Sp. (Pig)	GNHWAVGHLMa
NMS-36	<i>Rattus</i> Sp. (Rat)	LPRLHTDSRMATIDFPKKDPTTSLGRPFFLFRPRNa
NMS-33	<i>Bombina</i> Sp. (Firebelly Toads)	FLFQFSRAKDPSLKIGDSSGIVGRPFFLFRPRNa
NMS-17	<i>Bombina</i> Sp. (Firebelly Toads)	DSSGIVGRPFFLFRPRNa
NMU-8	<i>Sus</i> Sp. (Pig)	YFLFRPRNa
NMU-23	<i>Rattus</i> Sp. (Rat)	YKVNEYQGPVAPSGGFFLFRPRNa
NMU-25	<i>Homo sapiens</i> (Man)	FRVDEEFQSPFGSRSGYFLFRPRNa
NMU-25	<i>Sus</i> Sp. (Pig)	FKVDEEFQGPVQSNRRYFLFRPRNa
NMN	<i>Sus</i> Sp. (Pig)	KIPYIL

isolated from porcine spinal cord. NMB occurs both in CNS and the gastrointestinal tract. It exerts smooth muscle contraction, exocrine and endocrine secretions of gastrointestinal tissues, pancreas and pituitary, and various central effects and *in vitro* effects. N-terminally extended forms (NMB-30 and NMB-32) show similar activities to NMB. *Neuromedin C* (NMC) is a bombesin-like peptide isolated from porcine spinal cord exerting stimulating effects on smooth muscle of rat uterus [356]. NMC shows sequence identity with the C-terminal part 18–27 of the gastric-inhibitory polypeptide and sequence homology with the C-terminus of bombesin. The name NMC was coined since it is closely related to NMB. *Neuromedin S* (NMS) is a neuropeptide of different chain length with potent anorexigenic and circadian rhythm-modulating properties [357, 358]. It was identified and isolated from human, rat, and mouse brain tissues, and shows structural similarities to neuromedin U. Rat NMS-36 has been shown to be an endogenous ligand for neuromedin U receptor type 1 and type 2, respectively. Intracerebroventricular (i.c.v.) administration of NMS decreases food intake in rats more potently and persistently than is observed with the same dose of neuromedin U. The NMS analogues NMS-17, and NMS-33 were isolated from the dermal venoms of Eurasian bombinid toads. *Neuromedin U* (NMU), U is derived from uterus, is a neuropeptide of different chain length first isolated from the spinal cord of the pig. NMU was detected in various vertebrates, e.g., pig, human, rat, dog, chicken, rabbit and frog, showing a classical brain–gut distribution. Additionally, NMU possesses peripheral actions, e.g., adrenocortical function, modification of intestinal ion transport, and splanchnic blood flow. A significant reduction in food intake (anorexigenic effect) was found after intracerebroventricular (i.c.v.) administration to rats. Furthermore, elevations in body temperature and heart rate, as well as augmentation of stress responses, complete the spectrum of actions [359, 360]. The C-terminal 5-peptide amide sequence FRPRNa is completely conserved among some of the neuromedins. *Neuromedin N* (NMN) is a small neuropeptide with a similar biological profile to neurotensin (NT) isolated from porcine spinal cord. It exhibited a contractile activity on guinea pig ileum. NMN is synthesized as part of a larger precursor which also contains neurotensin (NT) and neurotensin-like peptide. In the brain, processing of pro-NT/NMN gives rise to NMN and NT, whereas in the gut processing leads mainly to the formation of NT and large NMN, a big peptide ending with the NMN sequence at its C-terminus [361, 362]. The designations neuromedin L (NML) for neurokinin A and neuromedin K (NMK) for neurokinin B are antiquated and should not longer be used.

3.3.7.12 Additional Neuroactive Peptides

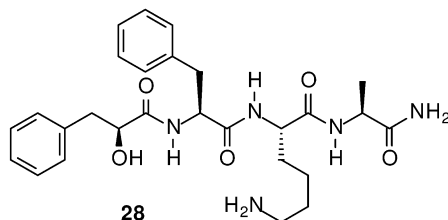
The neuroactive dipeptide *kyotorphin* 27 was first isolated from rat hypothalamus [363]. The sequence is part of neokytorphin (H-Thr-Ser-Lys-Tyr-Arg-OH) which has also been found in bovine brain and seems to be the propeptide of kyotorphin. It has been shown that angiotensin-converting enzyme (ACE) is capable of cleaving neokytorphin, yielding kyotorphin. Kyotorphin promotes the release of [Met] enkephalin.



Achatin is a neuroexcitatory peptide 4-peptide (H-Gly-D-Phe-Ala-Asp-OH) isolated from the ganglia of the African giant land snail *Achatina fulica* [364]. Achatin-I contains a D-amino acid in position 2, whereas achatin-II with the L-isomer in the same position shows neither physiological nor pharmacological activities. Because of Na^+ , achatin I induced a voltage-dependent inward current on the giant neuron.

Proctolin, H-Arg-Tyr-Leu-Pro-Thr-OH, acts as an excitatory neurotransmitter and was isolated from the proctodeal muscles of 125 000 cockroaches (~125 kg) *Periplaneta americana* in 1975 [365]. It initiates strong contractions at the proctodaeum at a concentration of only $10^{-9} \text{ mol L}^{-1}$.

Antho-Kamide 28, *antho-Rlamide I* (L-3-phenyllactyl-Tyr-Arg-Ile-NH₂), and *antho-RNamide I* (L-3-phenyllactyl-Leu-Arg-Asn-NH₂) [366] are neuropeptides isolated from the simple sea anemone species *Anthopleura elegantissima* characterized by an N-terminally acylated phenyllactyl moiety. The coelenterates contain a multitude of neuropeptides.



Delta sleep-inducing peptide (DSIP), H-Trp-Ala-Gly-Gly-Asp-Ala-Ser-Gly-Glu-OH, is a 9-peptide first isolated from cerebral venous blood of sleeping rabbits in 1977 [367]. It is widely distributed in the brain and other tissues. The concentration in the brain increases during hibernation. DSIP and appropriate analogues mediate sleep-like states (δ -slow-wave sleep) after intraventricular application into the rabbit brain. From various studies it can be concluded that DSIP passes the blood-brain barrier. DSIP was initially regarded as a candidate sleep-promoting factor. However, in part due to the lack of isolation of the DSIP gene, protein and related receptor, the link between DSIP and sleep has never been further characterized [368].

Diazepam-binding inhibitor peptide (DBIP) [369], also called anxiety peptide, octadecaneuropeptide (ODN), QATVGDVNTD¹⁰RPGLDLK, is an 18-peptide isolated from human and rat brain extracts. DBIP is a bioactive fragment of the diazepam-

binding inhibitor (DBI), and shows a number of behavioral and neurophysiological activities. It interacts with the benzodiazepine receptor and appears to increase anxiety [370].

Neurotensin (NT), <ELYENKPRRP¹⁰YIL, which occurs in the brain and gastrointestinal tract, was originally isolated from calf hypothalamus in 1971 [371]. NT acts as a neuromodulator of dopamine transmission and on anterior pituitary hormone secretion. NT exerts potent hypothermic and analgesic effects in the brain. Furthermore, it is a paracrine and endocrine modulator of the digestive tract and the cardiovascular system of mammals. NT is synthesized together with neuromedin N (NN) and the *neurotensin-like peptide* (H-Lys-Leu-Pro-Leu-Val-Leu-OH) as a larger precursor. Processing of pro-NT/NN in the brain gives rise to NN and NT whereas, in the gut, processing mainly leads to the formation of NT and large NN. Four neurotensin receptors have been identified, two of which are G protein-coupled receptors (NTR-1, NTR-2). NTR-3 and NTR-4 are structurally different and represent entirely new types of neuropeptide receptor, as they occur primarily intracellularly and are related to yeast sorting receptors (sortilin, SorLA). NT receptor antagonists contributed to an understanding of the role of NT in antipsychotic drug actions, psychostimulant sensitization and the modulation of pain. It is assumed that NT plays a key role in stress-induced analgesia [372].

Neuropeptide FF (NPFF), H-Phe-Leu-Phe-Gln-Pro-Gln-Arg-Phe-NH₂, was initially detected in bovine brain using antisera directed against the molluscan peptide FMRFamide [373]. NPFF is a peptide with pleiotropic functions in the mammalian CNS, including feeding processes, cardiovascular regulation, opiate interactions, pain modulation, insulin release, and electrolyte imbalance. With NPFF₁ and NPFF₂, two G protein-coupled receptors have been characterized as specific NPFF receptors [374, 375].

Head activator (HA), <EPPGGSKVIL¹⁰F, is an 11-neuropeptide conserved from hydra to humans. It is involved in the development of neuronal cells, and acts in hydra as an important factor in head regeneration. This morphogenic peptide was first isolated and sequenced from the coelenterates *Anthopleura elegantissima* and *Hydra vulgaris* in 1981 [376]. Later, HA was also isolated from tissues of higher organisms, e.g., bovine or human hypothalamus and rat intestine. In *Hydra*, HA initiates head-specific growth and differentiation processes, while in mammals it acts as a growth factor in neuronal development [377].

Neuropeptide 26RFa, VGTALGSLAE¹⁰ELNGYNRKKG²⁰GFSRFa, is a 26-peptide amide first isolated from a brain extract of the European green frog *Rana esculenta*. It possesses a C-terminal FRFa motif, which suggests it to be a member of a new subfamily of vertebrate RFamide peptides. Neuropeptide 26RFa has been reported to be a potent stimulator of appetite in mammals. Other data suggest that this neuropeptide exerts various neuroendocrine regulatory functions at the pituitary and adrenal levels [378].

Neuropeptide F (NPF), PDKDFIVNPS¹⁰DLVLDNKAAL²⁰RDYLRQINEY³⁰FAIIGRPRFa, is a 39-peptide amide first isolated from the flatworm *Moniezia expansa*

[379]. It shows sequence homology with members of the vertebrate neuropeptide Y family.

Neuropeptide S (NPS), SFRNGVGTGM¹⁰KKTSFQRAKS²⁰, acts as a transmitter. The name was coined since the N-terminal residue of NPS in all species is always serine (S). The NPS receptor (NPSR) shares moderate homology with other members of the G protein-coupled receptor (GPCR) supergene family. NPS is involved in sleep–wakefulness regulation. Its primary structure is highly conserved among vertebrates. NPS promotes arousal and reduces anxiety-like behavior in rats and mice [380, 381].

Neuropeptide W (NPW), WYKHVASPRY¹⁰HTVGRAAGLL²⁰MGLRRSPYLW³⁰ (hNPW), was first isolated from the porcine hypothalamus as an endogenous ligand for the G protein-coupled receptors GPR7 and GPR8 [382]. NPW is named after the tryptophan residues (W) at the N- and C-termini of the peptide. Rat NPW differs from the human peptide by only one amino acid in position 17 (Ser instead of Ala). NPW is present in antral G cells of rat, mouse, and human stomach [383].

Prolactin-releasing peptide (PrRP), hPrRP31: SRTHRHSMEI¹⁰RTPDIQPAWY²⁰ASR GIRPVGR³⁰Fa, is a 31-neuropeptide amide that, together with the N-terminally truncated PrRP20, was originally reported to function in the anterior lobe of the pituitary gland to stimulate prolactin release [384]. Later it was suggested that these peptides have various functions not only in the pituitary gland but also in other tissues. For example, they exert inhibition of food intake and body weight, the stimulation of sympathetic tone, and the activation of stress hormone secretion in the CNS. Recently, it was reported that PrRP acts as an endogenous regulator of cell growth [385].

3.3.8

Peptide Antibiotics

Antibiotics, derived from the Greek *antibios* which means “against life”, are a chemically heterogeneous group of substances produced by microorganisms (bacteria and fungi) that kill or inhibit the growth of other microorganisms. Peptide antibiotics [61–63] are both linear and cyclic peptides that can be further subdivided into homomeric peptide antibiotics exclusively composed of amino acids, and heteromeric ones additionally containing non-amino acid-derived building blocks. Furthermore, homodetic and heterodetic peptide antibiotics can be distinguished based on the character of the covalent bonds. For example, tyrocidins, gramicidin S, bacitracins, cyclosporins, viomycin, and capreomycin belong to the homomeric homodetic cyclic peptide antibiotics. Linear peptides, such as gramicidins A–C, bleomycin, and peplomycin are found in the family of heteromeric peptide antibiotics. The representatives of cyclic peptides can be either homodetic (polymyxins, colistines, etc.) or heterodetic (vancomycin, dactinomycin, ristocetin A, etc.). In the following sections the hundreds of known peptide antibiotics will be classified according to the mode of their biosynthesis (cf. Sections 3.2.1 and 3.2.3, respectively) into nonribosomally synthesized peptide antibiotics and ribosomally synthesized peptides.

3.3.8.1 Nonribosomally Synthesized Peptide Antibiotics

Nonribosomally synthesized peptide antibiotics are largely produced by bacteria, though a few representatives are produced by *Streptomyces* or lower fungi. They display some unusual structural features which may in part account for their antibiotic properties. Peptide antibiotics often contain nonproteinogenic amino acids. In general, building blocks with D-configuration, N-methylation or nonstandard structural features are incompatible with ribosomal synthesis. Many peptide antibiotics are cyclic and branched-cyclic peptidolactones and depsipeptides. Such peptide structures are not found in animal cells. Because of the cyclic structures and the high content of nonproteinogenic constituents, these antibiotics are resistant to proteolysis.

The antibiotic action targets various metabolic areas, including nucleic acid and protein biosynthesis, energy metabolism, cell-wall biosynthesis, and nutrient uptake. The streptogranins, for example, act as protein synthesis inhibitors, while the Gram-positive bacteria-specific bacitracin inhibits the transfer of peptidoglycan (cell wall) precursors to bactoprenol pyrophosphate. Although traditionally gramicidin S has been considered to be selective against Gram-positive bacteria, it has been shown also to have high activity against Gram-negative bacteria and the fungus *Candida albicans*. Most peptide antibiotics are relatively toxic, and so few are used clinically:

- Bacitracin and polymyxin B have been used in antibacterial chemotherapy.
- Thiopeptin, thiostrepton, and enduracidin are used as stock feed additives and veterinary drugs.
- The cationic lipopeptide colistin (polymyxin E) was modified in order to reduce its systemic toxicity; the resultant product, the methosulfate derivative colimycin, has been used as an aerosol formulation against *Pseudomonas aeruginosa* lung infections.

The modification of existing peptide antibiotics represents one approach to increase the potential of therapeutic applications. For example, the semisynthetically modified water-soluble compounds dalfopristin and quinupristin are derived from the potent, but rather insoluble, streptogranins – which were first discovered in the 1950s. They have been clinically tested in combination as a parenteral agent against resistant Gram-positive bacteria.

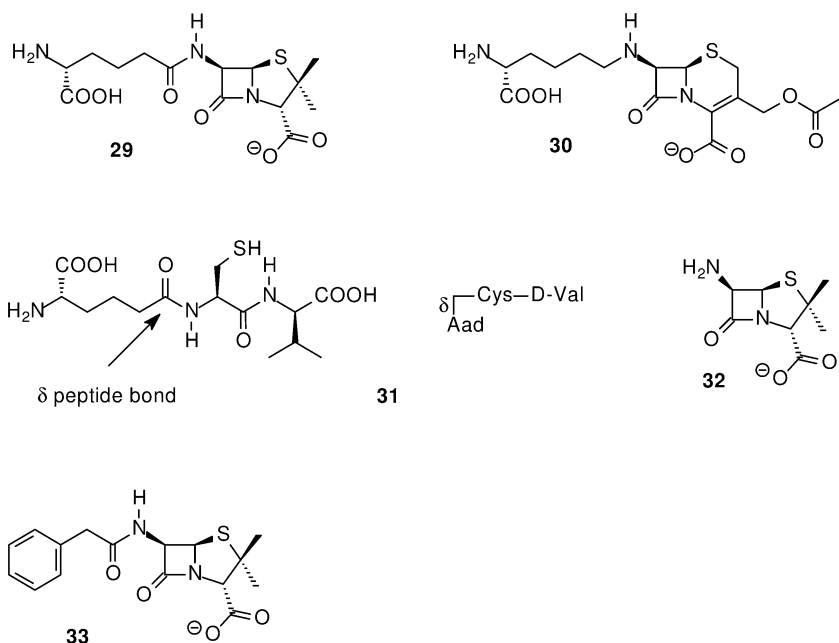
A second interesting approach comprises a combination of peptide synthesis modules to yield novel structures. These can, for example, be used as templates for chemical synthesis and diversity, and analogues of gramicidin S with variations in amino acid sequence, ring size, charge, etc., have been designed and synthesized in this way. Such compounds display greater selectivity towards bacteria than towards mammalian cells.

Although clearly not a peptide antibiotic, penicillin should be mentioned in this context, as the biosynthesis of penicillin N **29** and cephalosporin C **30** in *Cephalosporium acremonium* starts from cysteine and valine, together with α -aminoadipic acid, catalyzed by the ACV synthetase. The enzyme involved not

only catalyzes the formation of the two peptide bonds but also epimerizes the valine residue, yielding the tripeptide δ -(L- α -aminoadipoyl)-L-Cys-D-Val (ACV) **31** as an intermediate. The next step in the biosynthesis of **29** involves oxidative cyclization of **31** by the enzyme isopenicillin N synthase (IPNS) in the presence of Fe^{2+} ions, ascorbate, and oxygen [386].

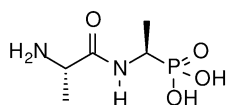
The growth-inhibitory properties of penicillins were first observed in 1928 by Alexander Fleming, in a *Staphylococcus* culture, and the application of penicillin against bacterial infections was first tested in humans in 1941.

All penicillins are characterized by the bicyclic β -lactam-thiazolidine structure, while the acyl moiety of the 6-aminopenicillanic acid (6-APA) **32** is variable. The latter has no antibiotic activity, and can be isolated as a fermentation product of *Penicillium chrysogenum*, though it is also available via enzymatic hydrolysis of benzylpenicillin (penicillin G) **33** using a bacterial acylase.



Thousands of new penicillin derivatives have been prepared in a semisynthetic approach using the enzymatic conversion of **29** or **33** into 6-aminopenicillanic acid **32** followed by acylation of the free amino group.

The structural diversity within the nonribosomally synthesized peptide antibiotics is extremely high, and does not permit strict definition of the families of related peptide antibiotics. These compounds vary in size, ranging from modified dipeptides (e.g., bacilysin or the phosphopeptide alaphosphin **34**) to peptides containing 16 to 17 amino acid residues (e.g., thiostrepton and enduracin).

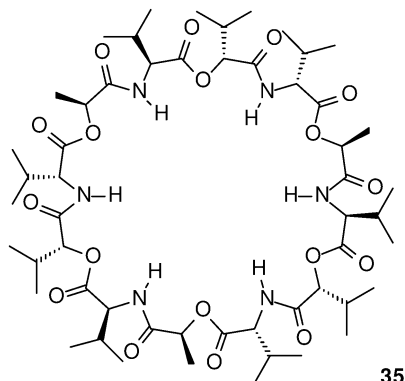


34

For this reason, only some well-investigated examples of peptide antibiotics have been selected for description in this chapter.

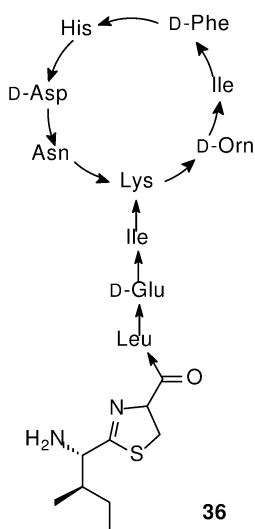
- *Gramicidins A–C* are linear membrane active 15-peptide antibiotics with alternating D- and L-amino acid residues, a formyl group at the N-terminus, and a characteristic C-terminal ethanolamine moiety. Depending on the N-terminal amino acid (Val or Ile), a differentiation may be made between [Val]Gramicidin A–C and [Ile]Gramicidin A–C. [Val]Gramicidin A, $\text{HCO-Val-Gly-Ala-D-Leu-Ala-D-Val-Val-D-Val-Trp-D-Leu}^{10}\text{-Trp-D-Leu-Trp-D-Leu-Trp-NH-CH}_2\text{-CH}_2\text{-OH}$, is capable of transporting monovalent cations, e.g., alkali ions, NH_4^+ , and H^+ , through membranes formed by a dimer (left-handed antiparallel double-stranded helix, cf. Figure 3.2). Gramicidin B has a Phe residue in position 11, whereas Gramicidin C has Tyr in this position. Gramicidins are broad-spectrum antibiotics effective against Gram-positive, Gram-negative bacteria and fungi. However, they are toxic to human red blood cells. Analogues with decreasing toxicity have been synthesized by modifying the ring size and amphipathicity.
- *Alamethicin* is produced by the fungus *Trichoderma viride* and is one of the most extensively investigated members of the long peptaibol antibiotics. Alamethicin consists of a natural microheterogeneous peptaibol mixture of which 23 members have been sequenced up to 2004. All alamethicins are 19-peptides blocked at the N-terminus by an acetyl moiety, and at the C-terminus by the 1,2-aminoalcohol L-phenylalaninol (Pheol). Alamethicin contains several aminoisobutyric acid (Aib) residues and L-phenylalaninol (Pheol) at the C-terminus. The alamethicins are classified into two groups. The major group is called F50 and is composed of neutral peptides (Gln^{18}), whereas the minor group, termed F30, consists of acidic peptides (Glu^{18}). Only the F30/6 analogue bears two acidic amino acids (Glu^7 , Glu^{18}), while in all other alamethicins Gln^7 is conserved. The alamethicins are rich (7–10 aa) in the strongly helicogenic, non-coded α -aminoisobutyric acid (Aib), and contain two well-spaced proline residues in positions 2 and 14. A major component of the natural alamethicin mixture is the F50/5 analogue with the following sequence: $\text{Ac-Aib-Pro-Aib-Ala-Aib-Ala-Gln-Aib-Val-Aib}^{10}\text{-Gly-Leu-Aib-Pro-Val-Aib-Aib-Glu-Gln-Pheol}$, the total synthesis of which in solution by an easily tunable segment condensation approach was described in 2004 [387].
- *Valinomycin 35*, $\text{cyclo-(D-Val-Lac-Val-D-Hiv)}_3$, is a symmetric cyclodepsipeptide isolated from *Streptomyces fulvissimus* [388]. It is formed by three repeating units of a sequence of D-Val, L-lactic acid (Lac), Val and D-hydroxyisovaleric acid (D-Hiv). Valinomycin acts as an ion carrier which selectively transports potassium ions

across membranes and shows the best K^+ / Na^+ selectivity of all K^+ -ionophores known to date. It is especially active against *Mycobacterium tuberculosis*.

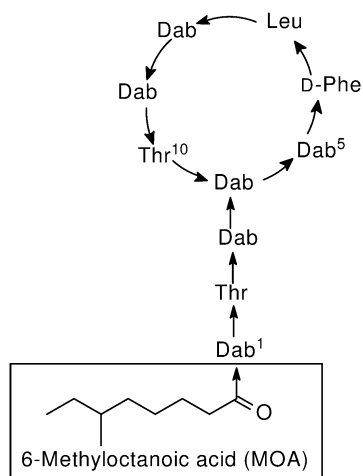


35

- *Bacitracin A* **36** is the most well-known member of the metallopeptide antibiotics produced by *Bacillus subtilis* and *Bacillus licheniformis*. They show a potent bactericidal activity directed primarily against Gram-positive bacteria, and are useful particularly in ophthalmology. The antibiotic activity of bacitracin A is apparently due to the inhibition of bacterial cell wall biosynthesis. It has been reported that bacitracin forms a strong complex with long-chain polyisoprenyl pyrophosphates that requires the presence of a bivalent metal cation such as Zn^{2+} or Ni^{2+} . Bacitracin interferes with sterol biosynthesis in mammalian cells by binding to pyrophosphate intermediates, thus accounting for its human toxicity [389].

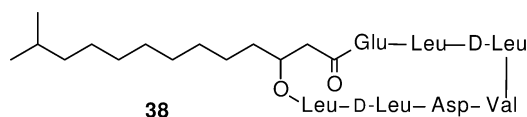


36



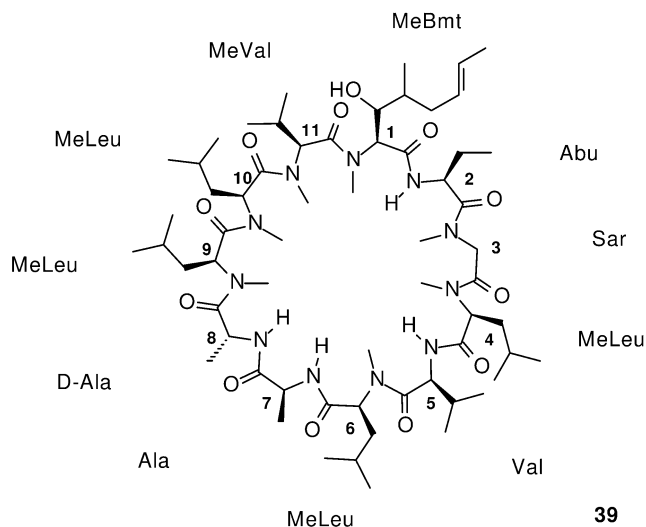
37

- *Polymyxins* are fatty acid-containing, branched cyclic peptides produced by *Bacillus polymyxa* possessing antibiotic activity against Gram-negative bacteria [390]. They consist of ten amino acids, six of them being L- α,γ -diaminobutyric acid (Dab). The structure of *polymyxin B₁* **37** is shown as a member of this peptide group. The N-terminal Dab residue is always acylated with 6-methyloctanoic acid or iso-octanoic acid, which is responsible for the strong amphiphilic, detergent-like properties of the peptide. The activity of polymyxins is based on the high affinity of this rather basic peptide to the negatively charged core moiety of the lipopolysaccharide (LPS) in the outer leaflet of the Gram-negative bacterial outer membrane.
- *Surfactin* **38** is a branched cyclic lipopeptide produced from *Bacillus subtilis* [391]. 3-Hydroxy-12-methyltridecanoic acid is linked via an amide bond to the N-terminal Glu of the peptide part, and the cycle is formed by an ester bridge between the carboxy group of the terminal Leu residue and the β -hydroxy group of the tridecanoic acid. Surfactin is the most powerful bio-detergent so far detected, based on its pronounced amphiphilicity. It also shows promising properties as an antiviral and anti-mycoplasma agent.



- *Cyclosporin A* (CsA) **39**, cyclo-(-MeBmt¹-Abu-Sar-MeLeu-Val⁵-MeLeu-Ala-D-Ala-MeLeu-MeLeu¹⁰-MeVal-), is the representative member of the cyclosporins comprising a group of homodetic cyclic peptide antibiotics that are produced by the fungus *Beauveria nivea* (previously named *Tolypocladium inflatum*). CsA exhibits a remarkable spectrum of various biological activities, including antifungal, antiviral, antiparasitic, anti-inflammatory, as well as immunosuppressive activities. It is a highly effective agent for the treatment of autoimmune disorders and for preventing organ-transplant rejection. Until now, CsA has been the main component of 25 naturally occurring cyclosporins that differ in the basic primary structure of CsA by amino acid substitutions in different positions. The first total chemical synthesis was described by Wenger [392] in 1984, but since then several hundred CsA analogues have been chemically synthesized. The biosynthesis in *Beauveria nivea* is accomplished by the multienzyme cyclosporin synthetase, which is the largest integrated enzyme structure so far known, catalyzing about 40 reaction steps including the final assembly of the undecapeptide chain of CsA and its cyclization. By *in vitro* synthesis, many new analogues of CsA became available which exerted remarkable immunosuppressive activity *in vitro*. CsA inhibits the activity of the peptidyl prolyl *cis/trans* isomerase cyclophilin which is a cytosolic CsA-binding protein. The immunosuppressive property of CsA is based on the ability of its complex with cyclophilin to prevent the expression of genes involved in the

activation of T lymphocytes by interfering with the appropriate intracellular signaling pathways [393, 394].



3.3.8.2 Ribosomally Synthesized Peptide Antibiotics

This class of peptide antibiotics plays a major role in natural host defenses. In general, these peptides do not exceed 50 amino acid residues, they are rich in positively charged amino acids, and are distinctly amphiphilic. β -Structures stabilized by disulfide bonds, amphipathic α -helices, extended structures and loops are characteristic structural features. A short survey of defense peptides was provided in Section 3.1. Some well-studied examples have been selected for a further short description.

Bacteriocins [395] comprise a group of antimicrobial peptides produced by different bacteria that kill or inhibit the growth of other bacteria. Bacteriocins are characterized by the presence of lanthionine and 3-methyllanthionine. Their importance results from their potential application as biopreservatives in food to inhibit the growth of spoilage or pathogenic bacteria. Bacteriocins can be simply divided into those produced by Gram-positive bacteria and others produced by Gram-negative bacteria. Probably colicins were the first bacteriocins and are the prototypes of the first group. Furthermore, bacteriocins of Gram-positive bacteria are divided into five subtypes. From which the lantiobiotics form the first subtype.

- *Colicins* are produced by various *E. coli* strains and are cytotoxic towards related bacteria. They parasitize outer membrane receptors, the physiological purpose of which is the transport of metabolites, metals, vitamins, and sugars for entry into susceptible cells; the lethal effects are either pore formation or intracellular nuclease activity.

- *Lantibiotics* [64, 396–399] comprise a class of post-translationally modified peptides of 19 to 38 residues. They are complex polycyclic molecules formed by the dehydration of selected Ser and Thr residues yielding in the first step 2,3-didehydroalanine (Dha) and (Z)-2,3-didehydrobutyrine (Dhb) to form lanthionine (Lan) or methyllanthionine (MeLan), respectively. The name lantibiotics is derived from “lanthionine-containing antibiotic peptides”. The unusual amino acid lanthionine is composed of two alanine residues connected across a thioether linkage that connects their β -carbons similarly to the way the amino acid cystine connects two alanine residues across a disulfide bridge. Although the existence of the lantibiotic nisin has been known since 1927, and subtilin (another group member) was discovered in 1944, it was not until the early 1970s that the complete structure of these two lantibiotics was elucidated in the pioneering studies of Erhard Gross. The post-translational modifications common to all lantibiotics are shown schematically in Figure 3.25 for nisin.

The gene organization for the biosynthetic machinery is performed in clusters, including information for the antibiotic prepeptide, the modifying enzymes, and accessory functions (e.g., special proteases and ABC transporter, immunity factors and regulatory peptides). According to Jung, the lantibiotics were subdivided into type A lantibiotics (e.g., nisin and subtilin), and type B lantibiotics (e.g., cinnamycin, ancovenin, and duramycins). In a recent review a new classification scheme has been proposed by Willey and van der Donk [399] that places all lantibiotics in one of three classes (Class I, II, or III) based on the pathway by which maturation of the peptides occurs and the presence or absence of antibiotic activity. Class I lantibiotics such as nisin A, subtilin, epidermin or Pep5 are modified by separate dehydratases (LanB) and cyclase (LanC). Class II lantibiotics with the selected representatives cinnamycin, mersacidin and lacticins are dehydrated and cyclized by a single LanM-type enzyme. Class III lantibiotics, e.g. Sap B secreted from *S. coelicolor*, SapT from *S. tendae*, and AmfS produced by *S. griseus*, have little or no antibiotic activity but instead perform alternative physiological function for the producing cell.

Many lantibiotics are bacteriocidal against a variety of Gram-positive bacteria at nanomolar levels and some are active against methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant enterococci and oxacillin-resistant Gram-positives. They exert their antibiotic activity by either forming pores in the target cell membrane or inhibiting cell wall biosynthesis. Nisin has been used for food preservation for more than 40 years. Importantly, there is little evidence of stable resistance mutants arising in food products that have been treated with nisin. There is interest reported in expanding the use of nisin to food packaging, as well as introducing it in clinical applications. Interestingly, mutacin 1140 from *Streptococcus mutans* is the major causative agent of dental caries. Cinnamycin and duramycins are potent inhibitors of phospholipase A2. The morphogenetic peptides SapB and SapT are both hydrophobic and surface active and may function as biosurfactants. Generally, since protein engineering is readily performed with ribosomally synthesized peptides, the lantibiotics seem to be especially amenable to rational drug design.

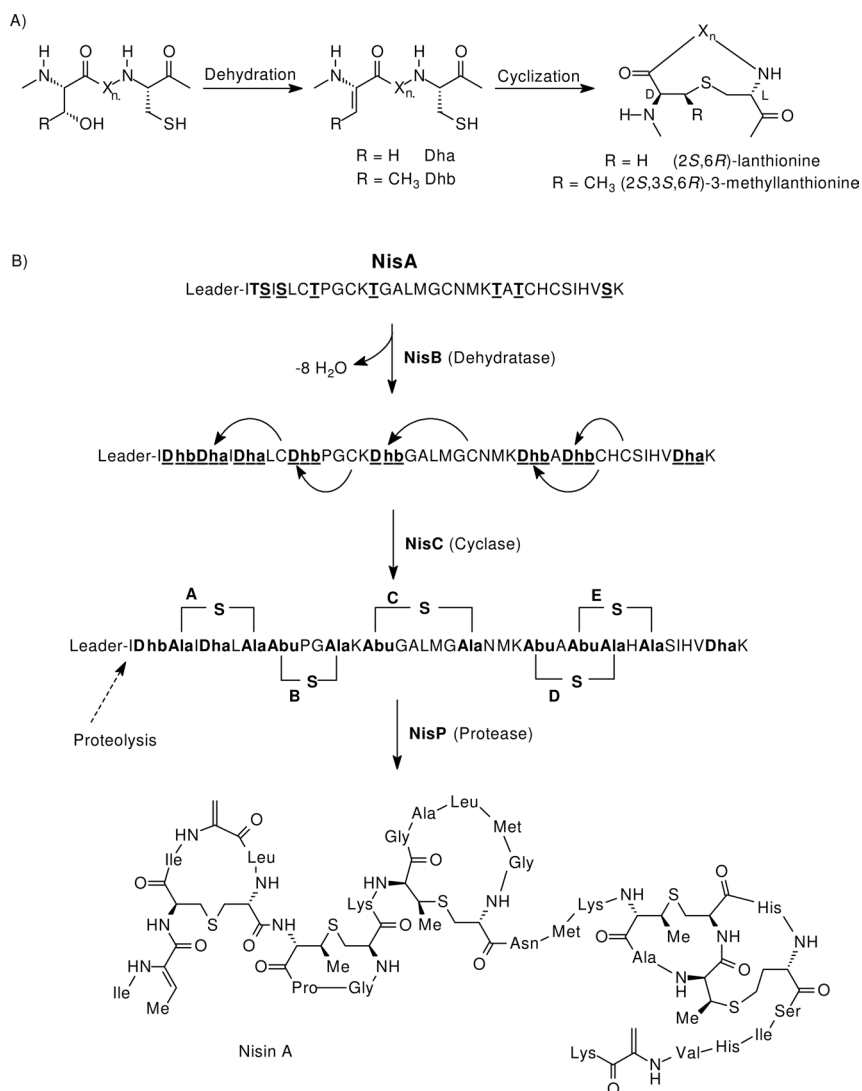


Figure 3.25 Schematic representation of lantibiotic biosynthesis demonstrated for nisin (slightly changed according to Willey and van den Donk [399]. (A) Formation of lanthionine (Lan) and methylanthionine (MeLan) in lantibiotic propeptides. (B) Formation of nisin.

According to a classification suggested by Boman [79] the structurally heterogeneous defense peptides can be subdivided into five groups:

Linear, preferentially helical peptides lacking cysteine form the first group with the cecropins and the magainins as prototypes.

- *Cecropins* [79, 400] form a group of insect-derived antimicrobial peptides first isolated from *Drosophila* and from the pupae of the giant silk moth *Hyalophora*

cecropia, from which the name is derived. *Cecropin A*, KWKLFKKIEK¹⁰VGQ NIRDGII²⁰KAGPAVAVVG³⁰QATQIAKa, and *cecropin B* with 35 aa, are positively charged linear peptides from *H. cecropia*, forming time-variant and voltage-dependent ion channels in planar lipid membranes. Cecropins are not lethal for mammalian cells at microbiocidal levels, and have been administered safely to animals.

- *Magainins* [69–71, 401] are synthesized in the skin glands of frogs and other amphibians. They are active *in vitro* against a variety of bacteria, fungi, and protozoa. These amphibian-derived peptides are characterized by a helical, amphiphilic structure resulting in high membrane affinity. Magainins with 20 to 25 residues are shorter than cecropins but share the overall features of charge distribution, helicity and amphiphilic properties. Magainin 1, GIGKFLH-SAG¹⁰KFGKAFVGEI²⁰MKS, and magainin 2, [Lys¹⁰, Asn²²]magainin 1, from *Xenopus laevis* and other magainins are α -helical ionophores which dissipate ion gradients in membranes, causing lysis. MSI-78, GIGKFLK KAK¹⁰KFGKAFV-KIL²⁰KKa, a synthetic analogue based on the magainin consensus sequence, has been developed as a drug for the topical treatment of Gram-positive infections.

Linear proline- and arginine-rich peptides form the second group which is extremely heterogeneous. Members occur, for example, in bovine neutrophils, pig intestine, honeybee, *Drosophila*, and other insects.

- *Bac5*, RFRPPIRRPP¹⁰LRPPFYPPFR²⁰PPIRPPIFPP³⁰IRPPFRPPLR⁴⁰FP, was isolated from ox, and may interact and disorganize bacterial membranes. *Indolicidin*, ILPWKWPWWP¹⁰WRRa, was first isolated from neutrophils of *Bos taurus*. Its antimicrobial action involves the bacterial cytoplasmic membrane. *Apidaecin Ia*, GNNRPVYIPQ¹⁰PRPPHPRIa, was isolated from honeybee and is a representative of the apidaecins. All these peptides are equally active against Gram-negative and Gram-positive bacteria.

The third group is characterized by one disulfide bond and is represented by the magainin-related brevenins and batenectin.

- *Brevenin-1*, FLPVLGIAA¹⁰KVVPALFCKI²⁰TKC (disulfide bond: C¹⁸-C²³), isolated from the Japanese frog *Rana brevipoda porsa*, and *batenectin*, RLCRIV-VIRV¹⁰CR (disulfide bond: C³-C¹¹), from bovine neutrophils are members of this group. These peptides show highly specific activity against bacteria, but are also toxic towards eukaryotic cells.

Peptides with more than one disulfide bond and β -sheet structures comprise the fourth group, including the best-studied families of animal and insect defensins.

- *Defensins* [73–75, 402] are potent antimicrobial and proinflammatory peptides. The mammalian defensins are cationic peptides ($M_r \sim 3.5\text{--}4.5$ kDa), rich in lysine and arginine containing six cysteine residues which form three characteristic disulfide bridges. They can be divided into three groups: α -defensins, β -defensins, and θ -defensins. Antimicrobial peptides of plants or insects have

also been termed defensins, though they have different structural features in comparison to vertebrate defensins. α -Defensins consist of 29–35 aa and contain three disulfide bonds that stabilize the triple-stranded β -sheet. Until now, six different human α -defensins (HD) have been described. Plant α -defensins do not form ion-permeable pores in artificial membranes, but may act through a receptor-mediated mechanism. β -Defensins consist of 29–35 aa, and have been isolated from different sources.

The fifth group in the Boman classification refers to fragments derived from factors already known but associated with other functions.

- *GIP(7-42)* is a proteolytical fragment of gastric inhibitory polypeptide (GIP) lacking the six N-terminal amino acid residues of GIP, whereas *DBI(32-86)* is a fragment of the diazepam-binding inhibitor (DBI). Both peptides show antibacterial activity against the Gram-positive bacterium *B. megaterium*.

In general, eukaryotic defense peptides that are produced constitutively are targeted to either secretory glands or to storage in granulae. This mainly applies to frog skin peptides, tachyplesins, and mammalian defensins. In contrast, the biosynthesis of many insect peptides (e.g., cecropins) occurs in response to an acute challenge by living bacteria or macromolecules (peptidoglycan or lipopolysaccharides) derived from bacteria. In both cases, the defense peptides are synthesized as inactive prepro-peptides like the genetically encoded bacterial peptides. There is generally only one processing step in bacteria to remove the propeptide part or leader peptide sequence, whereas the usual two-step maturation process is observed for eukaryotic defense peptides.

3.3.9

Peptide Toxins

During evolution venomous animals evolved a vast array of peptide and protein toxins [403, 404], either for defence against predators or for attack in aggressive competition for limited nutrient resources. Peptide toxins are usually found in animal venom associated with specialized glands secreting peptides injected via hollow teeth, stings, harpoons and nematocyst tubules. This envenomation apparatus allows their delivery into the soft tissue of animals via subcutaneous, intramuscular or intravenous routes. Most venoms comprise a highly complex mixture of peptides, often showing diverse and selective pharmacologies. Since these peptides are directed against a wide variety of pharmacological targets, they form an invaluable source of ligands for investigating the properties of these targets in various experimental paradigms. A couple of these peptide toxins have been used *in vivo* for proof-of-concept studies, with several having undergone preclinical or clinical development, e.g., for the treatment of pain, multiple sclerosis, and cardiovascular diseases.

Peptide and protein toxins are mostly low-molecular weight, single-chain compounds produced by snakes and invertebrates, as well as by virulent strains of bacteria and some plants. The isolation of toxic peptides from venoms of various snakes was

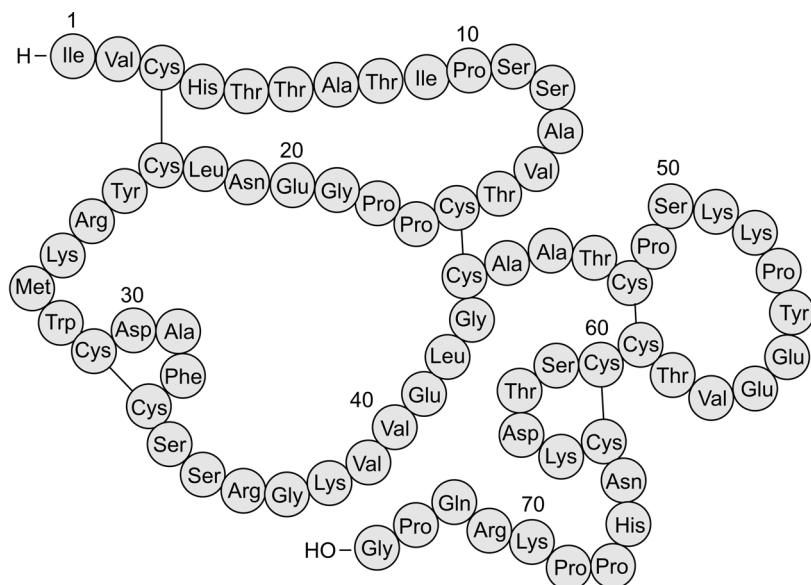


Figure 3.26 Primary structure of α -bungarotoxin.

first attempted in the late 1960s. After elucidation of the amino acid sequence of a neurotoxin from the cobra *Naja nigricollis*, the primary structure of α -bungarotoxin (Figure 3.26) was described in 1971 [405]. The latter toxin was isolated from the venom of the Chinese krait *Bungarus multicinctus*. α -Bungarotoxin strongly binds to the nicotinic acetylcholine receptor, and has been a useful tool in the characterization of this receptor. The neurotoxic 71-peptide α -cobratoxin [406] from the venom of the Thai cobra *Naja naja siamensis* shows similar activity, and binds specifically to the acetylcholine receptor, thereby inhibiting its opening.

Conotoxins [407–410] comprise the peptide neurotoxins of the marine predatory cone snails (genus *Conus*) from which over 500 species are known. The conotoxins contain on average 8 to 30 aa residues and a content of Cys between 22 and 50%. Although small in size, the conotoxins contain many of the structural elements found in larger proteins, including α -helices, β -sheets and β -turns, hence they are often referred to as mini-proteins. It has been reported that there are potentially 50 000 different conotoxins present in the venoms of living species in the genus *Conus*. Conotoxins have been divided into superfamilies (A, M, O, P, S, I, T), according to their disulfide bond framework, and further subdivided into classes according to their mode of action. Selected families of conotoxins are, for example, α -, α A-, μ -, μ O-, χ -, ρ -, δ -, κ -, κ A-, κ I-, δ -, σ -, τ -, ψ - and ω -conotoxins, contryphans, conulakins, conantokins and conopressins. α -Conotoxins are perhaps the most widespread members of this peptide group so far discovered, and the venom of most *Conus* species is likely to contain several of these. α -Conotoxin (GIA) (Figure 3.27), and shortened, slightly sequence-altered forms (GI, GII and MI) block postsynaptic cholinergic receptors in neuromuscular connections. μ -Conotoxins belonging to the

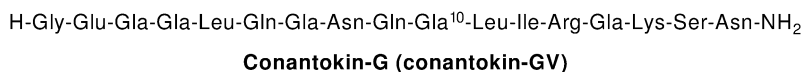
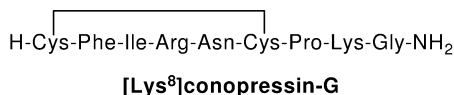
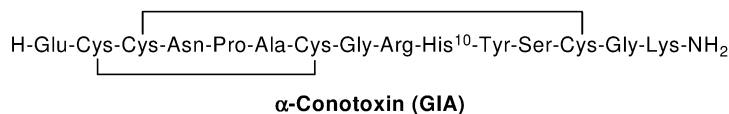


Figure 3.27 Primary structures of selected conotoxins.

M-superfamily are Cys-rich 23-peptide amides (GVIIIA, GVIIIB, GVIIIC) and were the first known peptide inhibitors that block muscular Na^+ channels causing paralysis and death. ω -Conotoxins belong to the O-superfamily. After injection into mice, they cause shaking and are therefore named also shaker-peptides. They block voltage-sensitive Ca^{2+} channels at cholinergic nerve terminals, thus inhibiting the release of acetylcholine. The first δ -conotoxin to be characterized, TxVIA (originally termed as ‘King Kong Peptide’), was isolated from the venom of the molluscivorous *C. textile*. The κ -conotoxin class belongs to the O-superfamily. The members of this class contain three disulfide bonds and were the first conotoxins to block voltage-gated potassium channels. Conopressins are basic 9-peptide amides with structural similarities to vasopressin. Smooth muscle effects result after injection into mice; therefore, they may be involved in the distribution of paralyzing toxins in the guest organism. $[\text{Lys}^8]\text{conopressin-G}$ and $[\text{Arg}^8]\text{conopressin-S}$ are typical members of this family. The conantokins (also named “sleepier peptides”) lacking cysteine are further peptides with neurotoxic effects. The name is derived from the Philippine “*antokin*”, which means “sleepily”. Conantokin-G and the conantokins-T and -GS have been identified. It is assumed that the Gla residues in conantokin-G promote the binding of the receptor on cell membranes. The conantokins-G and -T are antagonists of the NMDA receptor.

Scorpion peptide toxins have been produced from more than 700 scorpion species from which about 40% are capable of damaging mammals, including human beings. The primary neurotoxic components of scorpion venoms are basic peptides ($M_r \sim 8 \text{ kDa}$). Independent of chain length, all the toxins contain four intrachain disulfide bridges. The first peptide to be tested *in vivo* was *margatoxin* (MgTX), $\text{TIINVKCTSP}^{10}\text{KQCLPPCKAQ}^{20}\text{FGQSAGAKCM}^{30}\text{NGKCKCYPH}$, from the scorpion *Centruroides margaritatus*, which is a potent blocker of voltage-gated potassium channels $\text{K}_v 1.1$, $\text{K}_v 1.2$, and $\text{K}_v 1.3$. MgTX also inhibits the delayed-type hypersensitivity reaction to tuberculin in the mini-pigs, as determined by both the size of induration and the extent of T-cell infiltration. The effect of MgTX was similar in magnitude to that of the immunosuppressant agent FK 506, although lower concentrations of peptide were needed to achieve similar efficacy. Toxin V from *Leiurus quinquestriatus quinquestriatus* (LppV, Middle East), and toxin II from *Androctonus*

australi Hector (AaH, North Africa) are both 64-peptides that cause membrane potential-dependent slackening of Na^+ channel activation. This effect results from a voltage-dependent high-affinity binding of the scorpion toxin on the Na^+ channel. In contrast to the α -toxins, β -toxins bind voltage independently on the Na^+ channel. Chlorotoxin, MCMPCFTTDH¹⁰QMARKDDCC²⁰GGKGRGKCYG³⁰PQCLCR, isolated from the scorpion *Leiurus quinquestriatus* blocks specifically Ca^{2+} -activated chloride channels. The radioactive ¹³¹I-chlorotoxin analogue has cytolytic activity, and therefore the potential to selectively affect tumors and gliomas of neuroectodermal origin.

Spider peptide toxins paralyze insects by blocking the neuromuscular transmission mediated via glutamate receptors. Argiotoxin from *Argiope lobata* was the first spider venom to be structurally elucidated. The curtatocins from the venom of *Hololena curta* are cysteine-rich peptide amides with 36–38 amino acid residues. The toxic effect of these peptides results from an irreversible presynaptic neuromuscular blockade.

Sea anemone toxins are divided into type I and type II toxins that cause neurotoxicity by binding on neural and muscular Na^+ channels, respectively. Sh-K, RSCID-TIPKS¹⁰RCTTAFQCKHS²⁰MKYRLSFCBK³⁰TCGTC, is a K_v 1.1, K_v 1.2, and K_v 1.3 channel blocker. The 49-peptide anthopleurin A (Ax-I) containing three disulfide bridges was isolated from the sea anemone species *Anthopleura*. Ax-I is a type I toxin, whereas the 48-peptide toxin Sh-I from *Stichodactyla helianthus* is a type II toxin.

Bee venom contains various neurotoxic and cytolytic peptides (Table 3.22) in addition to the enzymes hyaluronidase and phospholipase A2 [411].

Mellitin corresponds to the sequence 44–69 of prepro-mellitin of the queen-bee. It occurs in the venom in 50-fold molar excess over other constituents. The hemolytic and surface tension-decreasing activity of mellitin is based on the distribution of hydrophobic amino acid residues in the N-terminal sequence part, and hydrophilic building blocks in the C-terminal part. The resulting tenside character is probably a prerequisite for the pharmacological and biochemical action.

Table 3.22 Primary structures of bee, wasp and hornet venom peptides.

Name	Source	Sequence
Mellitin	<i>Apis</i> Sp. (honeybee)	GIGAVLKVIT ¹⁰ TGLPALISWI ²⁰ KRKRQQ
Mastoparan	Wasp, hornet, yellow jacket	INLKALAAIA ¹⁰ KKILa
Mastoparan C	<i>Vespa crabro</i> (European hornet)	LNLKALLAVA ¹⁰ KKILa
Eumenine	<i>Anterhynchium flavomarginatum</i>	INLLKIAKGI ¹⁰ IKSLa
mastoparan-AF (EMP-AF)	<i>micado</i> (mason wasp)	
Mast cell-granulating peptide (MCDP)	<i>Apis</i> Sp. (bee)	IKCNCKRHVI ¹⁰ KPHICRKICG ²⁰ KNa
Bombolitin I	<i>Megabombus pennsylvanicus</i> (American bumblebee)	IKITTMLAKL ¹⁰ GKVLAHVa
Apamin	<i>Apis</i> Sp. (bee)	CNCKAPETAL ¹⁰ CARRCQHa
Crabolin	<i>Vespa crabro</i> (European hornet)	FLPLILRKIV ¹⁰ TALa

Mast cell-degranulating peptide (MCDP) is a 22-peptide amide containing two disulfide bridges. According to the name, it causes degranulation of mast cells and the release of high histamine concentrations into the tissue. Additionally, MCDP shows anti-inflammatory activity that is 100-fold that of hydrocortisone. MCDP causes mast cell degranulation and histamine release at low concentrations, and has anti-inflammatory activity at higher concentrations. Like apamin, MCDP blocks Ca^{2+} -dependent K^{+} channels in neurons.

Bombolitins [412] are another class of mast cell-degranulating peptides. The bombolitins I to V were first isolated from the venom of the bumblebee *Megabombus pennsylvanicus*. They are structurally similar. The bombolitins lyse erythrocytes and liposomes, release histamine from rat peritoneal mast cells, and stimulate phospholipase A 2 from different sources.

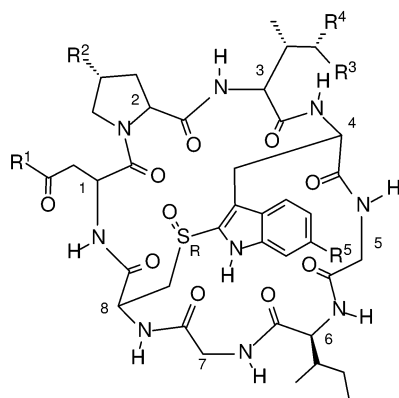
Mastoparan is a 14-peptide amide found in the venom of wasps. It degranulates mast cells and induces the release of catecholamines and serotonin from adrenal chromaffin cells or platelets. Furthermore, mastoparan shows hemolytic activity, and forms artificial ion channels in guest cell membranes. *Mastoparan C*, [Leu^{1,7},Val⁹]*mastoparan* occurs, like *crabolin*, in the venom of the European hornet. *Eumenine mastoparan-AF*, EMP-AF, was isolated from the venom of the solitary wasp *Anterhynchium flavomarginatum micado*, the most common eumenine wasp found in Japan. Additionally to its degranulation and hemolytic activity, EMP-AF also affects neuromuscular transmission in the lobster walking leg preparation [413].

Apamin is an 18-peptide amide of the bee venom (1–3% of the venom) and causes neurotoxic effects. The two arginine residues are of essential importance for the biological activity. Similar to MCDP, apamin blocks selectively Ca^{2+} -dependent K^{+} channels in neurons; this results in serious disturbances of CNS function.

In contrast to the peptide toxins described above, the poisonous constituents of the notorious toadstool *Amanita phalloides* are complex cyclic peptides. Undoubtedly, most fatal intoxications by mushrooms occur after ingestion of the *Amanita* species. The structural elucidation, synthesis, structure–activity relationship studies and biochemical studies were mainly carried out by Theodor Wieland and coworkers [57]. This team was able to explore the molecular events by which the amatoxins and phallotoxins exert their biological and toxic activities.

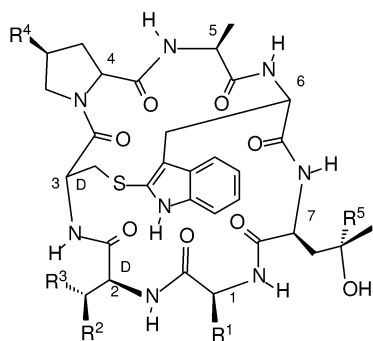
Amatoxins are bicyclic 8-peptides – from which up to nine individuals have been isolated so far – and solely responsible for such effects. The naturally occurring amatoxins (α -amanitin, β -amanitin, γ -amanitin, ϵ -amanitin, amanin, amanin amide, amanullin, amanullinic acid, proamanullin), are derived from one parent molecule (Figure 3.28), and differ only by the number of hydroxy groups and by an amide carboxy exchange. As an example, α -amanitin is termed systematically cyclo-(L-asparaginyl-trans-4-hydroxy-L-prolyl-(R)-4,5-dihydroxy-L-isoleucyl-6-hydroxy-2-mercapto-L-tryptophyl-glycyl-L-cysteinyl)-cyclo-(4-8)-sulfide-(R)-S-oxide. The lethal dose of amatoxins in humans could be estimated from accidental intoxications to be about 0.1 mg kg^{-1} body weight, or even lower. Wieland *et al.* [414] have developed a simple test for the discrimination between poisonous *Amanita* (except *A. visosa* specimens) and edible fungi as follows:

(A)



	R ¹	R ²	R ³	R ⁴	R ⁵
α-Amanitin	NH ₂	OH	CH ₂ OH	OH	OH
β-Amanitin	OH	OH	CH ₂ OH	OH	OH
γ-Amanitin	NH ₂	OH	CH ₃	OH	OH
ε-Amanitin	OH	OH	CH ₃	OH	OH
Amanin	OH	OH	CH ₂ OH	OH	H

(B)



	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶
Phalloin	CH ₃	CH ₃	OH	OH	CH ₃	CH ₃
Phalloidin	CH ₃	CH ₃	OH	OH	CH ₂ OH	CH ₃
Phallisin	CH ₃	CH ₃	OH	OH	CH ₂ OH	CH ₂ OH
Prophalloin	CH ₃	CH ₃	OH	H	CH ₃	CH ₃
Phallacin	CH(CH ₃) ₂	OH	CO ₂ H	OH	CH ₃	CH ₃

Figure 3.28 Selected naturally occurring toxic peptides of *Amanita* mushrooms. (A) amanitoxins, (B) phallotoxins.

- Press a piece of the mushroom (cap or stalk) onto crude paper (news-print).
- Mark the spot with a pencil, and after drying moisten the spot with a drop of concentrated hydrochloric acid.
- A greenish-blue color developing within 5–10 min indicates the presence of almost all amanitoxins.

Phallotoxins form a second group of less-toxic peptides occurring beside the amanitoxins in all toxic *Amanita* species. The lethal doses are generally higher compared to those of the amanitoxins. The basic structural formula of the phallotoxins is also shown in Figure 3.28. All seven naturally occurring members (phalloidin, phalloin, phallisin, prophalloin, phallacin, phallacidin, phallisacin) of this group are derived from the same cyclodeptide backbone consisting of seven amino acids and cross-linked from residue 3 to residue 6 by tryptathionine. For example, the systematic name of phalloidin (R¹, R², R⁶=CH₃; R², R³=OH; R⁵=CH₂OH) is

cyclo(-L-alanyl-D-threonyl-L-cysteinyl-*cis*-4-hydroxy-L-prolyl-L-alanyl-2-mercapto-L-tryptophyl-4,5-dihydroxy-leucyl)-cyclo(3-6)-sulfide. As mentioned above, the cyclic 10-peptide antamanide **3**, which is also found in *Amanita phalloides*, protects experimental animals from intoxication by phalloidin. Based on their strongly specific potential of inhibiting eukaryotic RNA polymerase II (B), the amatoxins have importance as diagnostic tools for recognizing whether a cellular process is dependent on the function of those enzymes. The binding of phallotoxins to F-actin stabilizes filamentous structures of the cytoskeleton; therefore, by conjugation with fluorescent molecules the phallotoxins can be used to visualize such cellular structures.

3.4

Review Questions

- Q3.1. Give a definition for autocrine, paracrine, and endocrine hormones.
- Q3.2. What are MHC proteins and where are they important?
- Q3.3. Describe the structure and composition of an antibody.
- Q3.4. Name proteins involved in the blood clotting cascade.
- Q3.5. Give a brief overview of the processes in ribosomal protein synthesis.
- Q3.6. What is the codon for selenocysteine? How is selenocysteine loaded to its tRNA?
- Q3.7. In which direction does ribosomal peptide synthesis proceed – from the N- to the C-terminus or from the C- to the N-terminus? In which direction is the mRNA read?
- Q3.8. Name at least five post-translational modifications of proteins.
- Q3.9. What types of N-glycans do you know?
- Q3.10. How are N-glycans formed in biosynthesis?
- Q3.11. How are O-glycans formed in biosynthesis?
- Q3.12. Describe some protein–carbohydrate binding motifs of O-glycans.
- Q3.13. Discuss the physiological roles of O-phosphorylation.
- Q3.14. Which lipid residues are usually found in lipidated proteins?
- Q3.15. What are pyroglutamyl peptides and where do they occur?
- Q3.16. Which amino acid residue can be sulfated?
- Q3.17. How does nonribosomal peptide synthesis proceed?
- Q3.18. Name some members of the insulin family and discuss their physiological role.
- Q3.19. Where do tachykinins occur and what is their physiological role?
- Q3.20. What are liberins and statins? Give some examples and describe their roles.
- Q3.21. Name the sequence of thyroliberin (TRH) and describe its physiological role.
- Q3.22. What is the physiological role of oxytocin?
- Q3.23. Name some classes of neuropeptides.
- Q3.24. Give examples of ribosomally synthesized peptide antibiotics.
- Q3.25. Give examples of nonribosomally synthesized peptide antibiotics.
- Q3.26. What are conotoxins?

References

- 1 G. Schmidt, *Top. Curr. Chem.* **1986**, 136, 109.
- 2 D. T. Krieger, *Science* **1983**, 222, 975.
- 3 J. H. Schwarz in: *Principles of Neural Science*, E. R. Kandel, J. H. Schwarz, T. M. Jessel (Eds.), Elsevier, Amsterdam, **1991**.
- 4 V. Mutt in: *Peptides 1992*, C. H. Schneider, A. N. Eberle (Eds.), Escom, Leiden, **1993**.
- 5 C. R. Harington, T. H. Mead, *Biochem. J.* **1935**, 29, 1602.
- 6 M. E. Anderson, *Methods Enzymol.* **1985**, 113, 548.
- 7 R. H. Siffert, V. du Vigneaud, *J. Biol. Chem.* **1935**, 108, 753.
- 8 R. Connsden, A. H. Gordon, A. J. P. Martin, R. L. M. Synge, *Biochem. J.* **1947**, 41, 596.
- 9 W. Gevers, H. Kleinkauf, F. Lipmann, *Proc. Natl. Acad. Sci. USA* **1969**, 63, 1335.
- 10 V. P. Skulachev, *Biochemistry (Moscow)* **2000**, 65, 749.
- 11 A. Guiotto, A. Calderan, P. Ruzza, G. Borin, *Curr. Med. Chem.* **2005**, 12, 2293.
- 12 J. Porath, P. Flodin, *Nature* **1959**, 183, 1657.
- 13 E. A. Peterson, H. A. Sober, *J. Am. Chem. Soc.* **1956**, 78, 751.
- 14 L. C. Craig, T. P. King, *Fed. Proc.* **1958**, 17, 1126.
- 15 A. P. Ryle, F. Sanger, L. F. Smith, R. Kitai, *Biochem. J.* **1955**, 60, 541.
- 16 W. W. Bromer, L. G. Sinn, A. Staub, O. K. Behrens, *J. Am. Chem. Soc.* **1956**, 78, 3858.
- 17 C. H. Li, I. I. Geschwind, R. D. Cole, I. D. Raacke, J. I. Harris, J. S. Dixon, *Nature* **1955**, 176, 687.
- 18 J. I. Harris, P. Roos, *Nature* **1956**, 178, 90.
- 19 J. I. Harris, A. B. Lerner, *Nature* **1957**, 179, 1346.
- 20 V. du Vigneaud, C. Ressler, S. Trippett, *J. Biol. Chem.* **1953**, 205, 949.
- 21 H. Tuppy, *Biochim. Biophys. Acta* **1953**, 11, 449.
- 22 V. du Vigneaud, H. C. Lawler, E. A. Popenoe, *J. Am. Chem. Soc.* **1953**, 75, 4880.
- 23 R. Acher, J. Chauvet, *Biochim. Biophys. Acta* **1953**, 12, 487.
- 24 D. F. Elliot, W. S. Peart, *Nature* **1956**, 177, 527.
- 25 L. T. Skeggs, K. E. Lentz, J. R. Kahn, N. P. Shumway, *J. Exp. Med.* **1956**, 104, 193.
- 26 V. du Vigneaud, C. Ressler, J. M. Swan, C. W. Roberts, P. G. Katsoyannis, S. Gordon, *J. Am. Chem. Soc.* **1953**, 75, 4879.
- 27 W. P. Jencks, *Catalysis in Chemistry and Enzymology*, Dover Publications, New York, **1987**.
- 28 C. Walsh, *Enzymatic Reaction Mechanism*, Freeman, New York, **1979**.
- 29 M. L. Bender, R. J. Bergeron, M. Komiyama, *The Bioorganic Chemistry of Enzymatic Catalysis*, Wiley, New York, **1984**.
- 30 T. C. Bruice, S. J. Benkovic, *Biochemistry* **2000**, 39, 6267.
- 31 L. Carpo, *Hormones*, Freeman, New York, **1985**.
- 32 W. König, *Peptide and Protein Hormones: Structure, Regulation, Activity*, VCH, Weinheim, **1993**.
- 33 C. Stanford, R. Horton, *Receptors: Structure and Function*, Oxford University Press, **2001**.
- 34 R. S. Yalow, S. A. Berson, *In vitro Procedures with Radioisotopes in Medicine*, IAEA, Wien, **1970**.
- 35 W. Pan, A. J. Kastin, W. A. Banks, J. E. Zadina, *Peptides* **1999**, 20, 1127.
- 36 A. J. Kastin, W. Pan, L. M. Maness, W. A. Banks, *Brain Res.* **1999**, 848, 96.
- 37 J. B. Finean, R. Coleman, R. H. Mitchell, *Membranes and Their Cellular Functions*, Blackwell Scientific Publications, Oxford, **1984**.
- 38 M. Bretscher, *Sci. Am.* **1985**, 253, 100.
- 39 M. K. Jain, *Introduction to Biological Membranes*, Wiley, New York, **1988**.
- 40 S. W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Paupit, J. N. Jansonius, J. R. Rosenbusch, *Nature* **1992**, 358, 727.

- 41 M. F. Perutz, *Sci. Am.* **1978**, 239, 92.
- 42 F. W. Alt, T. K. Blackwell, G. D. Yancopoulos, *Science* **1987**, 238, 1079.
- 43 G. Kroemer, A. C. Martinez, *Immunol. Today* **1992**, 13, 401.
- 44 J. Kuby, *Immunology*, Freeman, New York, **1994**.
- 45 T. Hunkapiller, L. Hood, *Adv. Immunol.* **1989**, 44, 1.
- 46 L. D. Barber, P. Parham, *Annu. Rev. Cell Biol.* **1993**, 9, 163.
- 47 P. Malik, P. Strominger, *J. Immunol. Methods* **2000**, 234, 83.
- 48 I. Stefanova, J. R. Dorfman, R. N. Germain, *Nature* **2002**, 420, 429.
- 49 T. Halkier, *Mechanisms in Blood Coagulation, Fibrinolysis, and the Complement System*, Cambridge University Press, Cambridge, **1991**.
- 50 P. M. Steinert, D. A. D. Parry, *Annu. Rev. Cell. Biol.* **1985**, 1, 41.
- 51 E. Fuchs, *Annu. Rev. Cell Dev. Biol.* **1995**, 11, 123.
- 52 L. Kreplak, J. Doucet, P. Dumas, F. Briki, *Biophys. J.* **2004**, 87, 640.
- 53 D. Kaplan, W. W. Adams, B. Farmer, C. Viney, *Silk Proteins*, American Chemical Society, New York, **1994**.
- 54 M. E. Nimni, *Collagen*, CRC Press, Boca Raton, **1988**.
- 55 L. Roberts, W. Hornebeck (Eds.), *Elastin and Elastases*, CRC Press, Boca Raton, **1989**.
- 56 D. Y. Li, B. Brooke, E. C. Davis, R. P. Mecham, L. K. Sorensen, B. B. Boak, E. Eichwald, M. T. Keating, *Nature* **1998**, 393, 276.
- 57 T. Wieland, *Peptides of Poisonous Amanita Mushrooms*, Springer, New York, **1986**.
- 58 V. Ersparmer in: *Amphibian Biology*, Vol. 1, H. Heatwole (Ed.), Surray Beatty & Sons, Chipping Norton, **1994**.
- 59 M. Broccardo, V. Ersparmer, G. Falconieri Ersparmer, G. Importa, G. Linari, P. Melchiorri, P. C. Montecucchi, *Br. J. Pharmacol.* **1981**, 73, 625.
- 60 Y. Yasuda-Kamatani, M. Kobayashi, A. Yasuda, F. Fujita, H. Minakato, K. Nomoto, M. Nakamura, F. Sakiyama, *Peptides* **1997**, 18, 347.
- 61 R. E. W. Hancock, *Lancet* **1997**, 349, 418.
- 62 H. Kleinkauf, H. von Döhren, *Biochemistry of Peptide Antibiotics*, de Gruyter, Berlin, **1990**.
- 63 R. E. W. Hancock, D. S. Chapple, *Antimicrob. Agents Chemother.* **1999**, 43, 1317.
- 64 R. W. Jack, G. Bierbaum, H.-G. Sahl, *Lantibiotics and Related Peptides*, Springer, New York, **1998**.
- 65 K. Fukase, M. Kitazawa, A. Sano, K. Shimbo, H. Fujita, S. Horimoto, T. Wakamiya, T. Shiba, *Tetrahedron Lett.* **1988**, 29, 795.
- 66 G. Kiss, H. Michl, *Toxicon* **1962**, 1, 33.
- 67 A. Csordas, H. Michl, *Monatsh. Chem.* **1970**, 101, 182.
- 68 M. Simmaco, G. Mignogna, D. Barra, *Biopolymers* **1998**, 47, 435.
- 69 M. Zasloff, *Proc. Natl. Acad. Sci. USA* **1987**, 84, 5449.
- 70 D. Barra, M. Simmaco, *Trends Biotechnol.* **1995**, 13, 205.
- 71 M. Zasloff, *Curr. Opin. Immunol.* **1992**, 4, 3.
- 72 D. Barra, M. Simmaco, H. Boman, *FEBS Lett.* **1998**, 430, 130.
- 73 T. Ganz, M. E. Selsted, R. I. Lehrer, *Eur. J. Haematol.* **1990**, 44, 1.
- 74 S. H. White, W. C. Wimley, M. E. Selsted, *Curr. Opin. Struct. Biol.* **1995**, 5, 521.
- 75 R. I. Lehrer, A. K. Lichtenstein, T. Ganz, *Annu. Rev. Immunol.* **1993**, 11, 105.
- 76 R. Fernandez de Caley, B. Gonzalez-Pascual, F. Garcia-Olmedo, P. Carbonero, *Appl. Microbiol.* **1972**, 23, 998.
- 77 W. F. Broekaert, B. P. A. Cammue, M. F. C. Debolle, K. Thevissen, G. W. Desamblanx, R. W. Osborne, *Crit. Rev. Plant Sci.* **1997**, 16, 297.
- 78 H. G. Boman, D. Hultmark, *Annu. Rev. Microbiol.* **1987**, 41, 103.
- 79 H. G. Boman, *Annu. Rev. Immunol.* **1995**, 13, 61.
- 80 B. M. Rode, *Peptides* **1999**, 20, 773.

- 81 R. Pascal, L. Boiteau, A. Commeyras, *Top. Curr. Chem.* **2005**, 259, 69.
- 82 R. L. P. Adams, J. P. Knowler, D. P. Leader, *The Biochemistry of the Nucleic Acids*, Chapman & Hall, London, **1992**.
- 83 S. Benzer, *Sci. Am.* **1962**, 206, 70.
- 84 M. W. Nierenberg, *Sci. Am.* **1963**, 208, 80.
- 85 F. H. C. Crick, *Sci. Am.* **1962**, 207, 66.
- 86 T. L. Hendrickson, *Nature Struct. Mol. Biol.* **2007**, 24, 100.
- 87 K. Sheppard, J. Yuan, M. J. Hohn, B. Jester, K. M. Devine, D. Söll, *Nucleic Acids Res.* **2008**, 36, 1813.
- 88 G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, 296, 1459.
- 89 J. A. Krzycki, *Curr. Opin. Microbiol.* **2005**, 8, 706.
- 90 N. Polacek, A. S. Mankin, *Crit. Rev. Biochem. Mol.* **2005**, 40, 285.
- 91 M. D. Erlacher, K. Lang, B. Wotzel, R. Rieder, R. Micura, N. Polacek, *J. Am. Chem. Soc.* **2006**, 128, 4453.
- 92 M. Beringer, M. V. Rodnina, *Mol. Cell* **2007**, 26, 311.
- 93 M. V. Rodnina, M. Beringer, W. Wintermeyer, *Trends Biochem. Sci.* **2007**, 32, 20.
- 94 T. M. Schmeing, K. S. Huang, D. E. Kitchen, S. A. Strobel, *Mol. Cell* **2005**, 20, 437.
- 95 S. Trobo, J. Aqvist, *Biochemistry* **2006**, 45, 7049.
- 96 C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Englewood, Polo, Roberts and Co. Publ., **2006**.
- 97 J. W. Suttie, *Annu. Rev. Biochem.* **1985**, 54, 472.
- 98 T. W. Rademacher, R. B. Parekh, R. A. Dwek, *Annu. Rev. Biochem.* **1988**, 57, 785.
- 99 A. Varki, *Glycobiology* **1993**, 3, 97.
- 100 B. Imperiali, *Acc. Chem. Res.* **1997**, 30, 452.
- 101 B. Imperiali, S. E. O'Connor, *Curr. Opin. Chem. Biol.* **1999**, 3, 643.
- 102 P. Sears, C.-H. Wong, *Cell. Mol. Life Sci.* **1998**, 54, 223.
- 103 C. M. Taylor, *Tetrahedron* **1998**, 54, 11317.
- 104 G. Ashwell, A. G. Morell, *Adv. Enzymol. Relat. Areas Mol.* **1974**, 41, 99.
- 105 K. V. Figura, A. Hasilik, *Trends Biochem. Sci.* **1984**, 9, 29.
- 106 G. F. Springer, *Science* **1984**, 224, 1198.
- 107 A. T. Beyer, J. E. Sadler, J. I. Rearick, J. C. Paulson, R. L. Hill, *Adv. Enzymol.* **1981**, 52, 24.
- 108 L. A. Lasky, *Science* **1992**, 258, 964.
- 109 M. L. Phillips, E. Nudelman, R. C. A. Gaeta, M. Perez, A. K. Singhal, S. Hakomori, J. C. Paulson, *Science* **1990**, 250, 1130.
- 110 G. Waltz, A. Arufo, W. Kalanus, M. Bevilacqua, B. Seed, *Science* **1990**, 250, 1132.
- 111 T. Buskas, S. Ingale, G.-J. Boons, *Glycobiology* **2006**, 16, 113R.
- 112 G. W. Hart, K. Brew, G. A. Grant, R. A. Bradshaw, W. Lennarz, *J. Biol. Chem.* **1979**, 254, 9747.
- 113 I. Ellgaard, A. Helenius, *Curr. Opin. Cell Biol.* **2001**, 13, 431.
- 114 H. C. Hang, C. R. Bertozzi, *Bioorg. Med. Chem.* **2005**, 13, 5021.
- 115 S. Wopereis, D. J. Lefeber, Éva Morava, R. A. Wevers, *Clin. Chem.* **2006**, 52, 574.
- 116 K. Ley, G. S. Kansas, *Nat. Rev. Immunol.* **2004**, 4, 325.
- 117 J. C. Byrd, R. S. Bresalier, *Cancer Metastasis Rev.* **2004**, 23, 77.
- 118 A. Gonzales de Peredo, D. Klein, B. Macek, D. Hess, J. Peter-Katalinic, J. Hofsteenge, *Mol. Cell Proteomics* **2002**, 1, 11.
- 119 T. Ilg, *Parasitol. Today* **2000**, 16, 489.
- 120 S. T. Prigge, A. S. Kolhekar, B. A. Eiper, R. E. Mains, L. M. Amzel, *Science* **1997**, 278, 1300.
- 121 F. Marks, *Protein Phosphorylation*, VCH, Weinheim, **1996**.
- 122 M. J. Hubbard, P. Cohen, *Trends Biochem. Sci.* **1993**, 18, 172.
- 123 P. Cohen, *Trends Biochem. Sci.* **1992**, 17, 408.
- 124 B. Schaffhausen, *Biochim. Biophys. Acta* **1995**, 1242, 61.

- 125 T. R. Bruke, Z.-J. Yao, Jr., D.-G. Liu, J. Voigt, Y. Gao, *Biopolymers* **2001**, 60, 32.
- 126 D. J. Dunican, P. Doherty, *Biopolymers* **2001**, 60, 45.
- 127 J. M. Troutman, M. J. Roberts, D. A. Andres, H. P. Spielmann, *Bioconj. Chem.* **2005**, 16, 1209.
- 128 S. L. Wolda, J. A. Glomset, *J. Biol. Chem.* **1988**, 263, 5997.
- 129 D. M. Leonard, *J. Med. Chem.* **1997**, 40, 2971.
- 130 D. R. Lowy, B. M. Willumsen, *Annu. Rev. Biochem.* **1993**, 62, 851.
- 131 A. Wittinghofer, H. Waldmann, *Angew. Chem. Int. Ed.* **2000**, 39, 4192.
- 132 P. J. Casey, X. Y. Buss, *Methods Enzymol.* **1995**, 250, 435.
- 133 M. D. Resh, *Biochim. Biophys. Acta* **1999**, 1451, 1.
- 134 M. Hackett, L. Guo, J. Shabanowitz, D. Hunt, E. L. Hewlett, *Science* **1994**, 266, 433.
- 135 J. T. Dunphy, M. E. Linder, *Biochim. Biophys. Acta* **1998**, 1436, 245.
- 136 W. D. Branton, C. G. Fields, V. L. VanDrise, G. B. Fields, *Tetrahedron Lett.* **1993**, 34, 4885.
- 137 M. Messer, *Nature* **1963**, 197, 1299.
- 138 K.-F. Huang, Y.-L. Liu, W.-J. Cheng, T.-P. Ko, A. H.-J. Wang, *Proc. Natl. Acad. Sci. USA* **2005**, 102, 13117.
- 139 F. Monigatti, B. Hekking, H. Steen, *Biochim. Biophys. Acta* **2006**, 1764, 1904.
- 140 A. Hille, T. Bräulke, K. von Figura, W. B. Huttner, *Eur. J. Biochem.* **1990**, 188, 577.
- 141 E. Chapman, M. D. Best, S. R. Hanson, C.-H. Wong, *Angew. Chem. Int. Ed.* **2004**, 43, 3526.
- 142 C. Seibert, T. P. Sakmar, *Biopolymers* **2008**, 90, 459.
- 143 T. Dierks, M. R. Lecca, P. Schlotterhose, B. Schmidt, K. von Figura, *EMBO J.* **1999**, 18, 2084.
- 144 S. R. Hanson, M. D. Best, C.-H. Wong, *Angew. Chem. Int. Ed.* **2004**, 43, 5736.
- 145 D. Ghosh, *Cell. Mol. Life Sci.* **2007**, 64, 2013.
- 146 K. von Figura, B. Schmidt, T. Selmer, T. Dierks, *Bioessays* **1998**, 20, 505.
- 147 M. T. Bedford, S. Richard, *Mol. Cell* **2005**, 18, 263.
- 148 S. Pahlich, R. P. Zakaryan, H. Gehring, *Biochim. Biophys. Acta* **2006**, 1764, 1890.
- 149 L. Hong, G. P. Schroth, H. Matthews, P. Yau, E. M. Bradbury, *J. Biol. Chem.* **1993**, 268, 305.
- 150 A. G. Li, L. G. Piluso, X. Cai, B. J. Gadd, A. G. Ladurner, X. Liu, *Mol. Cell* **2007**, 28, 408.
- 151 D. M. Townsend, *Mol. Interv.* **2007**, 7, 313.
- 152 E. R. Stadtman, H. Van Remmen, A. Richardson, N. B. Wehra, R. L. Levine, *Biochim. Biophys. Acta* **2005**, 1703, 135.
- 153 K. J. Reissner, D. W. Aswad, *Cell. Mol. Life Sci.* **2003**, 60, 1281.
- 154 B. M. Gaston, J. Carver, A. Doctor, L. A. Palmer, *Mol. Interv.* **2003**, 3, 253.
- 155 W.-S. Yeo, S. J. Lee, J. R. Lee, K. P. Kim, *BMB Reports* **2008**, 41, 194.
- 156 D. Corda, M. Di Girolamo, *EMBO J.* **2003**, 22, 1953.
- 157 D. Kornitzer, A. Ciechanover, *J. Cell. Physiol.* **2000**, 182, 1.
- 158 V. G. Wilson, D. Rangasamy, *Exp. Cell Res.* **2001**, 271, 57.
- 159 F. Lipmann in: *The Mechanism of Enzyme Action*, W. D. McElroy, B. Glass (Eds.), Hopkins, Baltimore, **1954**.
- 160 H. Kleinkauf, H. von Döhren, *Eur. J. Biochem.* **1990**, 192, 1.
- 161 M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, 97, 2651.
- 162 S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, 105, 715.
- 163 A. Lawen, J. Dittmann, B. Schmidt, D. Riesner, H. Kleinkauf, *Biochimie* **1992**, 74, 511.
- 164 S. G. Schultz, G. M. Makblout, *Handbook of Physiology: The Gastrointestinal System II*, Bethesda, MD: American Physiological Society, **1989**, p. 87.
- 165 J. H. Walsh, G. J. Dockray, *Gut Peptides*, Raven, New York, **1994**.
- 166 J. R. Rehfeld, *Horm. Metabol. Res.* **2004**, 36, 735.

- 167 J. F. Rehfeld, *Physiol. Rev.* **1998**, 78, 1087.
- 168 M. Dufresne, C. Seva, D. Fourmy, *Physiol. Rev.* **2006**, 86, 805.
- 169 V. Mutt, J. E. Jorpes, *Eur. J. Biochem.* **1968**, 6, 156.
- 170 M. Ruiz-Gayo, M. C. Gonzalez, S. Fernandez-Alfonso, *Regul. Pept.* **2006**, 137, 179.
- 171 I. P. Lam, F. K. Siu, J. Y. Chu, B. K. Chow, *Int. Rev. Cytol.* **2008**, 265, 159.
- 172 G. G. Nussdorfer, M. Bahcelioglu, G. Neri, L. K. Malendowicz, *Peptides* **2000**, 21, 309.
- 173 J. J. Holst, M. Bersani, A. H. Johnsen, H. Kofod, B. Hartmann, C. Orskov, *J. Biol. Chem.* **1994**, 269, 18827.
- 174 D. D. De Leon, M. F. Crutchlow, J.-Y. Nina Hann, D. A. Stoffers, *Int. J. Biochem. Cell Biol.* **2006**, 38, 845.
- 175 P. B. Jeppesen, *Gastroenterology* **2006**, 130, S127.
- 176 S. I. Said, V. Mutt, *Eur. J. Biochem.* **1972**, 28, 199.
- 177 T. W. Moody, I. Gozes, *Curr. Pharm. Des.* **2007**, 13, 1099.
- 178 K. Tatemoto, V. Mutt, *Proc. Natl. Acad. Sci. USA* **1981**, 78, 6603.
- 179 L. A. Frohman, R. D. Kineman, in: *Handbook of Physiology*, J. L. Kostyo H. M. Goodman (Eds.), Section 7, *The Endocrine System*, Volume V, *Hormonal Control of Growth*, Oxford University Press, New York, **1999**, p. 189.
- 180 J. J. Meier, M. A. Nauck, *Horm. Metab. Res.* **2004**, 36, 859.
- 181 H. Vaudry A. Arimura (Eds.), *Pituitary Adenylate Cyclase Activating Polypeptides*, Kluwer, Dordrecht (NL), **2002**.
- 182 S. J. Chan, Q. P. Cao, D. Steiner, *Proc. Natl. Acad. Sci. USA* **1990**, 87, 9319.
- 183 I. Claeys, G. Simonet, J. Poels, T. van Loy, L. Vercammen, A. de Loof, J. Vanden Broeck, *Peptides* **2002**, 23, 807.
- 184 J. P. Mayer, F. Zhang, R. D. DiMarchi, *Biopolymers (Pept. Sci.)* **2007**, 88, 687.
- 185 P. de Meyts, J. Whittaker, *Nat. Rev. Drug Discovery* **2002**, 1, 769.
- 186 E. M. Spencer (Ed.), *Insulin-like Growth Factors, Somatomedins*, de Gruyter, Berlin, **1983**.
- 187 G. Fürstenberger, H. J. Senn, *Lancet Oncol.* **2002**, 3, 298.
- 188 T. N. Wilkinson, T. P. Speed, G. W. Tregear, R. A. D. Bathgate, *BMC Evol. Biol.* **2005**, 5, 14.
- 189 S. Y. Hsu, *Trends Endocrinol. Metab.* **2003**, 14, 303.
- 190 R. A. D. Bathgate *et al.*, in: *Physiology of Reproduction*, E. Knobil, J. D. Neill, (Eds.), Elsevier, San Diego, 3rd Edn., **2006**, p. 679.
- 191 R. A. D. Bathgate, C. S. Samuel, T. C. D. Burazin, A. L. Gundlach, G. W. Tregear, *Trends Endocrinol. Metabol.* **2003**, 14, 207.
- 192 T. N. Wilkinson, R. A. Bathgate, *Adv. Exp. Med. Biol.* **2007**, 612, 1.
- 193 O. D. Sherwood, *Endocr. Rev.* **2004**, 25, 205.
- 194 R. Ivell, R. A. D. Bathgate, *Biol. Reprod.* **2002**, 67, 699.
- 195 D. J. Scott, P. Fu, P.-J. Shen, A. Gundlach, S. Layfield, A. Riesewijk, H. Tomiyana, J. M. Hutson, G. W. Tregear, R. A. D. Bathgate, *Ann. Acad. Sci. N. Y.* **2005**, 1041, 16.
- 196 K. J. Rosengren, S. Zhang, F. Lin, N. L. Daly, D. J. Scott, R. A. Hughes, R. A. D. Bathgate, D. J. Craik, J. D. Wade, *J. Biol. Chem.* **2006**, 281, 28287.
- 197 W. W. de Herder, S. W. J. Lamberts, *Curr. Opin. Oncol.* **2002**, 14, 53.
- 198 A. M. Comaru-Schally, A. V. Schally, *Int. J. Oncol.* **2005**, 26, 301.
- 199 A. D. Spier, L. de Lecea, *Brain Res. Rev.* **2000**, 33, 228.
- 200 H. Rubinfeld, I. Shimon, *Pediatr. Endocrinol. Rev.* **2006**, 4, 106.
- 201 C. Severini, G. Improta, G. Falconieri-Erspamer, S. Salvadori, V. Erspamer, The Tachykinin Peptide Family, *Pharmacol Rev.* **2002**, 54, 285.
- 202 J. M. Conlon, The Tachykinin Peptide Family with Particular Emphasis on Mammalian Tachykinins and Tachykinin Receptor Agonists, In: P. Holzer (Ed.), *Handbook of Experimental Pharmacology*, Vol. 64, Springer, Heidelberg, **2004**, p. 25.

- 203 L. Liu, E. Burcher, *Peptides* **2005**, 26, 1369.
- 204 N. M. Page, *Peptides* **2005**, 26, 1356.
- 205 S. E. Leeman, S. L. Ferguson, *Neuropeptides* **2000**, 34, 249.
- 206 S. Harrison, P. Geppetti, *Int. J. Biochem. Cell. Biol.* **2001**, 33, 555.
- 207 E. Munekata, *Comp. Biochem. Physiol.* **1991**, 98c 171.
- 208 Z. Gao, N. P. Peet, *Curr. Medicinal Chem.* **1999**, 6, 374.
- 209 C. A. Maggi, *Trends Biochem. Sci.* **2000**, 21, 173.
- 210 M. M. Kurtz, R. Wang, M. K. Clements, M. A. Cascieri, C. P. Austin, B. R. Cunningham, N. M. Page, *Cell Mol. Life Sci.* **2004**, 61, 1652.
- 211 M. N. Page, N. J. Bell, S. M. Gardiner, I. T. Manyonda, K. J. Brayley, P. G. Strange, P. J. Lowry, *Proc. Natl. Acad. Sci. USA* **2003**, 100, 6245.
- 212 Y. Zhang, L. Lu, C. Furlonger, G. E. Wu, C. J. Paige, *Nat. Immunol.* **2000**, 1, 392.
- 213 A. Bettio, A. G. Beck-Sickinger, *Biopolymers (Pept. Sci.)* **2001**, 60, 420.
- 214 R. Bader, O. Zerbe, *ChemBioChem* **2005**, 6, 1520.
- 215 L. Grundemar, S. Bloon (Eds.), *Neuropeptide Y and Drug Developments*, Academic Press, New York, **1997**.
- 216 J. R. Kimmel, L. J. Hayden, H. G. Pollock, *J. Biol. Chem.* **1975**, 250, 9369.
- 217 R. L. Batterham, C. W. Le Roux, M. A. Cohen, A. J. Park, S. M. Ellis, M. Patterson, G. S. Frost, M. A. Ghatei, S. R. Bloom, *J. Clin. Endocrinol. Metab.* **2003**, 88, 3989.
- 218 R. L. Batterham, M. A. Cowley, C. J. Small, H. Herzog, M. A. Cohen, C. L. Dakin, A. M. Wren, A. E. Brynes, M. J. Low, M. A. Ghatei, R. D. Cone, S. R. Bloom, *Nature* **2002**, 418, 650.
- 219 C. L. Roth, P. J. Enriori, K. Harz, J. Woelfle, M. A. Cowley, T. Reinehr, *J. Clin. Endocrinol. Metab.* **2005**, 90, 6386.
- 220 M. Kojima, H. Hosoda, Y. Date, M. Nakazato, H. Matsuo, K. Kangawa, *Nature* **1999**, 402, 656.
- 221 M. Kojima, K. Kangawa, *Physiol. Rev.* **2005**, 85, 495.
- 222 T. L. Peeters, *J. Physiol. Pharmacol.* **2003**, Suppl. 4, 95.
- 223 N. A. Tritos, E. G. Kokkotou, *Mayo Clin. Proc.* **2006**, 81, 653.
- 224 A. F. Leite-Moreira, J. B. Soares, *Drug Discov. Today* **2007**, 12, 276.
- 225 J. C. Brown, V. Mutt, J. R. Dryburgh, *Can. J. Physiol. Pharmacol.* **1971**, 49, 399.
- 226 Z. Itoh, *Peptides* **1997**, 18, 593.
- 227 G. Carpenter, M. I. Wahl, The Epidermal Growth Factor Family, in: *Peptides, Growth Factors and their Receptors I*, M. B. Sporn, A. B. Roberts (Eds.), Springer, New York, **1991**, 69.
- 228 D. Zhang, M. X. Slivkowski, M. Mark, G. Frantz, R. Akita, Y. Sun, K. Hillan, C. Crowley, J. Brush, P. J. Godowski, *Proc. Natl. Acad. Sci. USA* **1997**, 94, 9562.
- 229 R. M. G. Nair, J. F. Barrett, C. Y. Bowers, A. V. Schally, *Biochemistry* **1970**, 9, 1103.
- 230 R. Burgus, T. F. Dunn, D. Desiderio, D. N. Ward, W. Vale, R. Guillemin, *Nature* **1970**, 226, 321.
- 231 G. A. Scalabrino, N. Hogan, K. M. O'Boyle, G. R. Slatore, D. J. Ghregg, C. M. Fitchett, S. M. Draper, G. W. Bennett, P. M. Hinkle, K. Bauer, C. H. Williams, K. F. Tipton, J. A. Kelly, *Neuropharmacology* **2007**, 52, 1472.
- 232 S. Engel, M. C. Gershengorn, *Pharmacol. Ther.* **2007**, 113, 410.
- 233 A. Nagy, A. V. Schally, *Biol. Reprod.* **2005**, 73, 851.
- 234 A. Padula, *Animal Reprod. Sci.* **2005**, 88, 115.
- 235 J. P. Moreau, P. Delavault, J. Blumberg, *Clin. Ther.* **2006**, 28, 1485.
- 236 G. C. Boorse, R. J. Denver, *Gen. Comp. Endocrinol.* **2006**, 146, 9.
- 237 C. H. Li, J. D. Yamashiro, *J. Am. Chem. Soc.* **1970**, 92, 7608.
- 238 E. E. Muller, V. Locatelli, D. Cocchi, *Physiol. Rev.* **1999**, 79, 511.
- 239 L. L. Anderson, S. Jeftinija, C. G. Scannes, *Exp. Biol. Med.* **2004**, 229, 291.

- 240 E. Redei, P. A. Rittenhouse, S. Revskoy, R. F. McGivern, F. Aird, *Ann. N. Y. Acad. Sci.* **1998**, 840, 456.
- 241 D. Engler, E. Redei, I. Kola, *Endocr. Rev.* **1999**, 20, 460.
- 242 N. Gallo-Payet, M. D. Payet, *Microsc. Res. Tech.* **2003**, 61, 275.
- 243 O. Zwermann, D. M. Schulte, M. Reincke, F. Beuschlein, *Eur. J. Endocrinol.* **2005**, 153, 435.
- 244 A. Eberle, *Melanotropins: Chemistry, Physiology and Mechanisms*, Karger, Basel, **1988**.
- 245 P. C. Eves, S. MacNeil, J. W. Haycock, *Peptides* **2006**, 27, 444.
- 246 M. E. Hadley, R. T. Dorr, *Peptides* **2006**, 27, 921.
- 247 J. E. Wikberg, R. Muceniece, I. Mandrika, *et al.*, *Pharmacol. Res.* **2000**, 42, 393.
- 248 I. Gantz, T. M. Fong, *Am. J. Physiol. Metab.* **2003**, 284, E468.
- 249 V. du Vigneaud, C. Ressler, C. J. M. Swan, C. W. Roberts, *J. Am. Chem. Soc.* **1953**, 75, 4879.
- 250 F. Fahrenholz, *Physiol. Rev.* **2001**, 81, 629.
- 251 M. Manning, L. L. Cheng, S. Stoev, N. C. Wo, W. Y. Chan, H. H. Szeto, T. Durroux, B. Mouillac, C. Barberis, *J. Peptide Sci.* **2005**, 11, 593.
- 252 V. du Vigneaud, D. T. Gish, P. G. Katsoyannis, *J. Am. Chem. Soc.* **1954**, 76, 4751.
- 253 L. L. Cheng, S. Stoev, M. Manning, S. Derick, A. Pena, M. B. Mimoun, G. Guillon, *J. Med. Chem.* **2004**, 47, 2375.
- 254 G. Griebel, J. Stemmelin, C. S. Gal, P. Soubrie, *Curr. Pharm. Des.* **2005**, 11, 1549.
- 255 J. F. Whitfield, *Exp. Opin. Invest. Drugs* **2005**, 14, 251.
- 256 J. P. Potts, *J. Endocrinol.* **2005**, 187, 311.
- 257 P. M. Guerreiro, J. L. Renfro, D. M. Power, A. V. M. Canario, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2007**, 292, R679.
- 258 G. J. Strewler, *N. Engl. J. Med.* **2000**, 342, 177.
- 259 T. J. Martin, *J. Clin. Invest.* **2005**, 115, 2322.
- 260 P. M. Sexton, D. M. Findlay, T. J. Martin, *Curr. Med. Chem.* **1999**, 6, 1067.
- 261 C. L. Huang, L. Sun, B. S. Moonga, M. Zaidi, *Cell. Mol. Biol. (Noisy-le-grand)* **2006**, 52, 33.
- 262 S. G. Amara, V. Jonas, M. G. Rosenfeld, E. S. Ong, R. M. Evans, *Nature* **1982**, 298, 240.
- 263 C. Shih, G. W. Bernard, *Clin. Orthop. Relat. Res.* **1997**, 334, 335.
- 264 P. J. Goadsby, *Drugs* **2006**, 65, 1.
- 265 T. J. Rink, K. Beaumont, J. Koda, A. Young, *Trends Pharmacol. Sci.* **1993**, 14, 113.
- 266 K. Kitamura, K. Kangawa, M. Kawamoto, Y. Ichiki, S. Nakamura, H. Matsuo, T. Eto, *Biochem. Biophys. Res. Commun.* **1993**, 192, 553.
- 267 J. P. Hinson, S. Kapas, D. M. Smith, *Endocrine Rev.* **2000**, 21, 138.
- 268 A. Martinez, M. J. Miller, E. J. Unsworth, J. M. Siegfried, F. Cuttitta, *Endocrinol.* **1995**, 136, 4099.
- 269 Y. Takai, S. Hyodo, T. Katafuchi, N. Minamino, *Peptides* **2004**, 25, 1643.
- 270 D. Bell, B. J. McDermott, *Brit. J. Pharmacol.* **2008**, 153, S247.
- 271 A.-M. Duncan, A. Kladis, G. L. Jennings, A. M. Dart, M. Esler, D. J. Campbell, *Am. J. Physiol.* **2000**, 278, R897.
- 272 J. M. Stewart, L. Gera, E. J. York, D. C. Chan, E. J. Whalleya, P. A., Bunn, Jr., R. J. Vavrek, *Biol. Chem.* **2001**, 382, 37.
- 273 L. M. F. Leeb-Lundberg, D. S. Kang, M. E. Lamb, D. B. Fathy, *Pharmacol. Rev.* **2005**, 57, 27.
- 274 S. Mueller, I. Paegelow, S. Reißmann, *Signal Transduction* **2006**, 6, 5.
- 275 J. M. Saavedra, *Cell Mol. Neurobiol.* **2005**, 25, 485.
- 276 C. M. Ferrario, *Hypertens.* **2006**, 47, 515.
- 277 F. Ducancel, *CMLS, Cell Mol. Life Sci.* **2005**, 62, 5286.
- 278 S. A. Douglas, M. Gellai, M. Ezekiel, E. H. Ohlstein, *J. Hypertens.* **1994**, 12, 561.

- 279 F. L. Marasciulo, M. Montagnani, M. A. Potenza, *Curr. Med. Chem.* **2006**, *13*, 1655.
- 280 Y. Kloog, I. Ambar, M. Sokolovsky, E. Kochta, Z. Wolberg, A. Bdolah, *Science* **1988**, *242*, 268.
- 281 F. Ducancel, *Toxicon* **2002**, *40*, 1541.
- 282 A. J. De Bold, H. B. Borenstein, A. T. Veress, W. A. Sonnenberg, *Life Sci.* **1981**, *28*, 89.
- 283 N. C. Trippodo, A. A. MacPhee, F. E. Cole, H. L. Blakesley, *Proc. Soc. Exp. Biol. Med.* **1982**, *170*, 502.
- 284 A. J. de Bold, B. G. Bruneau, Natriuretic Peptides, in: *Handbook of Physiology, Section 7: The Endocrine System*, Vol. III: Endocrine Regulation of Water and Electrolyte Balance, J. C. S. Fray, M. H. Goodman (Eds.), Oxford University Press, New York, **2000**, p. 377.
- 285 C. Moro, M. Berlan, *Fund. Clin. Pharmacol.* **2006**, *20*, 41.
- 286 D. L. Vesely, *Clin. Exp. Pharmacol. Physiol.* **2006**, *33*, 169.
- 287 D. L. Vesely, J. S. Norris, J. M. Walters, R. R. Jespersen, D. A. Baeyens, *Biochem. Biophys. Res. Commun.* **1987**, *148*, 1540.
- 288 D. L. Vesely, *Cardiovasc. Haematolog. Disorders-Drug Targ.* **2007**, *7*, 47.
- 289 D. de Wied, *Excerpta Med. Int. Congr. Ser.* **1974**, *359*, 653.
- 290 E. R. Kandel, J. H. Schwartz, T. M. Jessel (Eds.), *Principles of Neural Sciences*, McGraw-Hill Medical, New York, **2000**.
- 291 T. Hökfelt, T. Bartfei, F. Bloom, *Lancet Neurol.* **2003**, *2*, 463.
- 292 T. Hökfelt, C. Broberger, Z.-Q. D. Xu, V. Sergeev, R. Ubink, M. Diez, *Neuropharmacol.* **2000**, *39*, 1337.
- 293 H. Gainer, J. T. Russel, Y. P. Loh, *Neuroendocrinol.* **1985**, *40*, 171.
- 294 R. E. Mains, B. A. Eipper, *J. Biol. Chem.* **1979**, *254*, 7885.
- 295 J. E. Zadina, W. A. Banks, A. J. Kastin, *Peptides* **1986**, *7*, 497.
- 296 J. E. Zadina, W. A. Banks, A. J. Kastin, *Peptides* **1986**, *7*, 497.
- 297 B. Ahmed, A. J. Kastin, W. A. Banks, J. E. Zadina, *Peptides* **1994**, *15*, 1105.
- 298 W. Pan, A. J. Kastin, W. A. Banks, J. E. Zadina, *Peptides* **1999**, *20*, 1127.
- 299 A. L. Vaccarino, G. A. Olson, R. D. Olson, A. J. Kastin, *Peptides* **1999**, *20*, 1527.
- 300 R. J. Bodnar, M. M. Hadjimarkou, *Peptides* **2002**, *23*, 2307.
- 301 R. J. Bodnar, M. M. Hadjimarkou, *Peptides* **2003**, *24*, 1241.
- 302 R. J. Bodnar, G. E. Klein, *Peptides* **2004**, *25*, 2205.
- 303 R. J. Bodnar, G. E. Klein, *Peptides* **2005**, *26*, 2629.
- 304 R. J. Bodnar, G. E. Klein, *Peptides* **2006**, *27*, 3391.
- 305 R. J. Bodnar, *Peptides* **2007**, *28*, 2435.
- 306 R. J. Bodnar, *Peptides*, **2008**, *29*, 2292.
- 307 S. Spector, J. Donnerer in: *Handbook of Experimental Pharmacology*, A. Herz, H. Akil, E. J. Simon (Eds.), Springer, Berlin, **1993**.
- 308 C. B. Pert, S. H. Snyder, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 2243.
- 309 E. J. Simon, J. M. Hiller, I. Edelman, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 1947.
- 310 L. Terenius, *Acta Pharmacol. Toxicol.* **1973**, *33*, 377.
- 311 J. Hughes, T. W. Smith, H. W. Kosterlitz, L. A. Morgan, H. R. Morris, *Nature* **1975**, *258*, 577.
- 312 G. W. Pasternak, *The Opiate Receptors*, Humana Press, Totowa, N.Y., **1988**.
- 313 R. J. Knapp, K. N. Hawkins, G. K. Lui, J. E. Shook, J. S. Heyman, F. Porreca, V. J. Hruby, T. F. Burks, H. I. Yamamura, *Adv. Pain Res. Ther.* **1990**, *14*, 45.
- 314 G. Cardillo, L. Gentilucci, A. R. Qasem, F. Sgarzi, S. Spampinato, *J. Med. Chem.* **2002**, *10*, 2755.
- 315 R. Polt, M. Dhanasekaran, C. M. Keyari, *Med. Res. Rev.* **2005**, *25*, 557.
- 316 H. Teschemacher in: *Handbook of Experimental Pharmacology*, A. Herz, H. Akil, E. J. Simon (Eds.), Springer, Berlin, **1993**.
- 317 C. J. Fowler, G. L. Fraser, *Neurochem. Int.* **1994**, *24*, 401.

- 318 R. J. Rodgers, S. J. Cooper, *Endorphins, Opiates, and Behavioural Processes*, Wiley, 1988.
- 319 T. Hökfelt, T. Bartfai, F. Bloom, *Lancet Neurol.* **2003**, 2, 463.
- 320 R. Ubink, L. Calza, T. Hökfelt, *Trends Neurosci.* **2003**, 26, 604.
- 321 Y. L. Hurd, *Mol. Psychiat.* **2002**, 7, 75.
- 322 D. J. Sheffler, B. L. Roth, *Trends Pharmacol. Sci.* **2003**, 24, 107.
- 323 F. Merg, D. Filliol, I. Usynin, I. Bazov, N. Bark, Y. L. Hurd, T. Yakovleva, B. L. Kieffer, G. Bakalkin, *J. Neurochem.* **2006**, 97, 292.
- 324 L. de Lecea, J. G. Sutcliffe (Eds.), *Hypocretins*, Springer, 2005.
- 325 L. de Lecea, J. G. Sutcliffe, *FEBS J.* **2005**, 5675.
- 326 P. Melchiorri, L. Negri, *Gen. Pharmacol.* **1996**, 27, 1099.
- 327 A. B. Usenko, T. G. Emel'yanova, N. F. Myasoedov, *Biol. Bull.* **2002**, 29, 2002.
- 328 V. Erspamer, P. Melchiorri, G. Falconieri-Erspamer, L. Negri, R. Corsi, C. Severini, D. Barra, M. Simmaco, G. Kreil, *Proc. Natl. Acad. Sci. USA* **1989**, 86, 5188.
- 329 L. H. Lazarus, S. D. Bryant, P. S. Cooper, S. Salvadori, *Prog. Neurobiol.* **1999**, 57, 377.
- 330 R. K. Reinscheid, H. Nothacker, O. Civelli, *Peptides* **2000**, 21, 901.
- 331 M. H. Heinricher, *Life Sci.* **2005**, 77, 3127.
- 332 M. Broccardo, G. Linari, R. Guerrini, S. Agostini, C. Petrella, G. Improta, *Peptides* **2005**, 26, 1590.
- 333 G. M. Scoto, N. Santangelo, C. Parenti, *Neurosci. Lett.* **2005**, 387, 126.
- 334 E. Okuda-Ashitaka, S. Ito, *Peptides* **2000**, 21, 1101.
- 335 H. Teschemacher, G. Koch, V. Brantl, *Biopolymers* **1997**, 43, 99.
- 336 K. Neubert, B. Hartrodt, I. Born, A. Barth, H.-L. Reuthrich, G. Grecksch, U. Schrader, C. Liebmann in: *β-Casomorphins and Related Peptides*, F. Nyberg, V. Brantl (Eds.), Fyris-Tryk, Uppsala, 1990, 15.
- 337 H. Meisel, R. J. FitzGerald, *Br. J. Nutr.* **2000**, 84, 27.
- 338 A. Trompette, J. Claustre, F. Caillon, G. Jourdan, J. A. Chayvialle, P. Plaisancié, *J. Nutr.* **2003**, 133, 3499.
- 339 H. Teschemacher, *Curr. Pharm. Design* **2003**, 9, 1331.
- 340 A. A. Karelin, M. M. Philippova, E. V. Karelina, V. T. Ivanov, *Biochem. Biophys. Res. Commun.* **1994**, 202, 410.
- 341 Q. Zhao, I. Garreau, F. Sannier, J. M. Piot, *Biopolymers* **1997**, 43, 75.
- 342 H. Lipperton, B. Lin, B. Gumusel, N. Witriol, A. Wasserman, M. Knight, *Peptides* **2006**, 27, 2284.
- 343 G. Gäde, *Biochem. J.* **1991**, 275, 671.
- 344 G. Gäde, H. G. Marco, in: *Studies in Natural Product Chemistry (Bioactive Natural Products)*, Atta-ur-Rahman (Ed.), Vol. 33, Elsevier, Amsterdam, **2005**, p. 69.
- 345 P. Fernlund, L. Josefsson, *Science* **1972**, 177, 173.
- 346 G. Gäde, *Annu. Rev. Entomol.* **2004**, 49, 93.
- 347 J. E. Zadina, L. Hackler, L.-J. Ge, A. J. Kastin, *Nature* **1997**, 386, 499.
- 348 J. E. Zadina, S. Martin-Schild, A. A. Gerall, A. J. Kastin, L. Hackler, L.-J. Ge, X. Zhang, *Ann. N. Y. Acad. Sci.* **1999**, 897, 136.
- 349 J. Fichna, A. Janecka, J. Costentin, J.-C. Do Rego, *Pharmacol. Rev.* **2007**, 59, 88.
- 350 A. P. Woodhead, B. Stay, S. L. Seidel, M. A. Khan, S. S. Tobe, *Proc. Natl. Acad. Sci. USA* **1989**, 86, 5997.
- 351 W. G. Bendena, B. C. Donly, S. S. Tobe, *Ann. N. Y. Acad. Sci.* **1989**, 897, 311.
- 352 J. G. Yoon, B. Stay, *J. Comp. Neurol.* **1995**, 363, 475.
- 353 N. Birgül, C. Weise, H.-J. Kreienkamp, D. Richter, *EMBO J.* **1999**, 18, 5892.
- 354 B. Stay, S. S. Tobe, *Annu. Rev. Entomol.* **2007**, 52, 277.
- 355 H. Ohki-Hamazaki, *Prog. Neurobiol.* **2000**, 62, 297.
- 356 G. Gasmi, A. Singer, J. Forman-Kay, B. Sarkar, *J. Peptide Res.* **1997**, 49, 500.
- 357 T. Ida, K. Mori, M. Miyazato, Y. Egi, S. Abe, K. Nakahara, *Endocrinol.* **2005**, 146, 4217.

- 358 K. Mori, M. Miyazato, T. Ida, N. Murakami, R. Serino, Y. Ueta, M. Kojima, K. Kangawa, *EMBO J.* **2005**, 24, 325.
- 359 P. H. Jethwa, C. J. Small, K. L. Smith, A. Seth, S. J. Darch, C. R. Abbot, K. G. Murphy, J. F. Todd, M. A. Ghatei, S. R. Bloom, *Am. J. Physiol. Endocrinol. Metab.* **2005**, 289, E301.
- 360 T. Kawai, A. Shibata, K. Kurosowa, Y. Sato, S. Kato, K. Ohki, T. Hashimoto, N. Sakura, *Chem. Pharm. Bull.* **2006**, 54, 659.
- 361 P. R. Dobner, D. L. Barber, L. Villa-Komaroff, C. McKiernan, *Proc. Natl. Acad. Sci. USA* **1987**, 84, 3516.
- 362 J.-P. Vincent, J. Mazella, P. Kitabki, *Trends Pharmacol. Sci.* **1999**, 20, 302.
- 363 H. Takagi, H. Shiomi, H. Ueda, H. Amano, *Nature* **1979**, 282, 410.
- 364 Y. Kamatani, H. Minakata, P. T. Kenny, T. Iwashita, K. Watanabe, K. Funase, X. P. Sun, A. Yongsiri, K. H. Kim, P. Novalis-Li, *Biochem. Biophys. Res. Commun.* **1989**, 160, 1015.
- 365 A. N. Starratt, B. E. Brown, *Life Sci.* **1975**, 17, 1253.
- 366 C. J. P. Grimmlikhuijzen, K. L. Rinehart, E. Jacob, D. Graff, R. K. Reinscheid, H.-P. Nothacker, A. L. Staley, *Proc. Natl. Acad. Sci. USA* **1990**, 87, 5410.
- 367 G. A. Schoenenberger, *Eur. Neurol.* **1984**, 23, 321.
- 368 V. M. Kovalzon, T. V. Strekalova, *Neurochemistry* **2006**, 97, 303.
- 369 P. Ferrero, A. Guidotti, B. Conti-Tronconi, E. Costa, *Neuropharmacol.* **1984**, 23, 1359.
- 370 J. Leprince, D. Cosquer, G. Bellemere, D. Chatenet, H. Tollemere, S. Jegou, M.-C. Tonon, H. Vaudry, *Peptides* **2006**, 27, 1561.
- 371 R. E. Carraway, S. E. Leeman, *Biochem. J.* **1973**, 248, 6854.
- 372 P. R. Dobner, *Cell Mol. Life Sci.* **2005**, 62, 1946.
- 373 H.-Y. T. Yang *et al.*, *Proc. Natl. Acad. Sci. USA* **1985**, 82, 7757.
- 374 N. Vyas, C. Mollereau, G. Cheve, C. R. McCurdy, *Peptides* **2006**, 27, 990.
- 375 J. M. Nystedt, A. Brandt, F. S. Vilim, E. B. Ziff, P. Panula, *Peptides* **2006**, 27, 1020.
- 376 H. C. Schaller, H. Bodenmüller, *Proc. Natl. Acad. Sci. USA* **1981**, 78, 7000.
- 377 I. Franke, F. Buck, W. Hampe, *Eur. J. Biochem.* **1997**, 244, 940.
- 378 N. Chartrel, F. Bruzzzone, J. Leprince, H. Tollemere, Y. Anouar, J.-C. Do-Rego, I. Segalas-Milazzo, L. Guilhaudis, P. Cosette, T. Jouenne, G. Simonnet, M. Vallarino, J.-C. Beauvillain, J. Costentin, H. Vaudry, *Peptides* **2006**, 27, 1110.
- 379 A. G. Maule, C. Shaw, W. Halton, L. Thiem, C. F. Johnston, I. Fairweather, H. D. Buchanan, *Parasitology* **1991**, 102, 309.
- 380 Y. L. Xu, S. Reinscheid, S. Huitron-Resendiz, S. Clark, Z. Wang, S. Lin, F. Burcher, J. Zeng, N. Ly, S. Henriksen, *Neuron* **2004**, 43, 487.
- 381 R. K. Reinscheid, Y. L. Xu, *FEBS J.* **2005**, 272, 5689.
- 382 Y. Shimomura, M. Harada, M. Goto, T. Sugo, Y. Matsumoto, M. Abe, T. Watanabe, T. Asami, C. Kitada, M. Mori, H. Onda, M. Fujino, *J. Biol. Chem.* **2002**, 277, 35826.
- 383 M. S. Mondal, H. Yamaguchi, Y. Date, K. Toshinai, T. Kawagoe, T. Tsurata, H. Kageyama, S. Shioda, Y. Shimomura, M. Mori, N. Nakazato, *J. Endocrinol.* **2006**, 188, 49.
- 384 S. Hinuma, Y. Habata, R. Fujii, Y. Kawamata, M. Hosoya, *Nature* **1998**, 393, 272.
- 385 W. K. Samson, M. M. Taylor, *Peptides* **2006**, 27, 1099.
- 386 P. L. Roach, I. J. Clifton, C. M. H. Hensgens, N. Shibata, C. J. Schofield, J. Hajdu, J. E. Baldwin, *Nature* **1997**, 387, 827.
- 387 C. Peggion, I. Coin, C. Toniolo, *Biopolymers (Pept. Sci.)* **2004**, 76, 485.
- 388 J. S. Davies, *J. Peptide Sci.* **2003**, 9, 471.
- 389 L.-J. Ming, J. D. Epperson, *J. Inorg. Biochem.* **2002**, 91, 46.
- 390 A. L. Meyer, *Curr. Opin. Pharmacol.* **2005**, 5, 490.

- 391 F. Peypoux, J. M. Bonmatin, J. Wallach, *Appl. Microbiol. Biotechnol.* **1999**, 51, 553.
- 392 R. M. Wenger, *Helv. Chim. Acta* **1984**, 67, 502.
- 393 H. Cho, Y. Chung, *Arch. Pharm. Res.* **2004**, 27, 662.
- 394 A. R. Hamel, F. Hubler, A. Carrupt, R. M. Wenger, M. Mutter, *J. Peptide Res.* **2004**, 63, 147.
- 395 R. James, C. Lazdunski, F. Pattus (Eds.), *Bacteriocins, Microcins and Lantibiotics*, Springer, Berlin, **1992**.
- 396 G. Jung, H.-G. Sahl (Eds.), *Nisin and Novel Lantibiotics*, Escom, Leiden, **1991**.
- 397 H. G. Sahl, G. Bierbaum, *Annu. Rev. Microbiol.* **1998**, 52, 41.
- 398 C. Chatterjee, M. Paul, L. Xie, W. A. van der Donk, *Chem. Rev.* **2005**, 105, 633.
- 399 J. M. Willey, W. A. van den Donk, *Annu. Rev. Microbiol.* **2007**, 61, 477.
- 400 A. Pillai, S. Ueno, H. Zhang, J. M. Lee, Y. Kato, *Biochem. J.* **2005**, 390, 207.
- 401 A. J. De Lucca, T. J. Jacks, K. A. Brodgen, *Ann. Agric. Environ. Med.* **1996**, 3, 37.
- 402 C. Beisswenger, R. Bals, *Curr. Protein Pept. Sci.* **2005**, 6, 255.
- 403 A. L. Harvey (Ed.) Snake toxins, in: *International Encyclopedia of Pharmacology of Therapeutics*, Pergamon Press, New York, **1991**.
- 404 J. E. Alouf, J. H. Freer (Eds.) *Sourcebook of Bacterial Protein Toxins*, Academic Press, London, **1991**.
- 405 D. Mebs, K. Narita, C. Y. Lee, *Biochem. Biophys. Res. Commun.* **1971**, 44, 711.
- 406 E. Karlson, H. Arnberg, D. Eaker, *Eur. J. Biochem.* **1971**, 21, 1.
- 407 B. M. Olivera, L. J. Cruz, *Toxicon* **2001**, 39, 7.
- 408 C.-Z. Wang, C.-W. Chi, *Acta Biochim. Biophys. Sinica* **2004**, 36, 713.
- 409 C. J. Armishaw, P. F. Alewood, *Curr. Protein Pept. Sci.* **2005**, 6, 221.
- 410 R. S. Norton, B. M. Olivera, *Toxicon* **2006**, 48, 780.
- 411 R. C. Hider, *Endeavour, New Series* **1988**, 12, 60.
- 412 A. Argiolas, J. J. Pisano, *J. Biol. Chem.* **1985**, 260, 1437.
- 413 K. Konno, M. Hisada, H. Naoko, Y. Itagaki, N. Kawai, A. Miwa, T. Yasuhara, Y. Morimoto, Y. Nakata, *Toxicon* **2000**, 38, 1505.
- 414 T. Wieland, L. Wirth, E. Fischer, *Liebigs Ann. Chem.* **1949**, 564, 152.

4

Peptide Synthesis

4.1

Principles and Objectives

As the biosynthesis even of complicated proteins occurs *in vivo* within seconds or minutes, the relatively tedious classical chemical synthesis of peptides and proteins in the laboratory appears to be a somewhat hopeless enterprise, especially with regard to speed. The first chemical synthesis of insulin during the early 1960s took over two years to complete. Today, peptide and protein synthesis have become much more efficient, not only by the use of recombinant DNA techniques but also with the advent of automated solid-phase peptide synthesis and multiple peptide synthesis.

Peptides and proteins can be considered as encoded biomolecules. The translation products are determined genetically and synthesized for highly specific functions. Following the sequencing of the human genome, specific peptide sequences are under intensive investigation by all facets of the genomics and proteomics research field. The above-mentioned seemingly hopeless competition between a peptide chemist and nature takes on a completely new aspect since a huge number of new peptide targets are the focus of synthesis research to find new active compounds which can help combat diseases in humans. Recent approvals for peptide drugs demonstrate the utility of peptides for broad indications, e.g. cardiovascular diseases, bone metabolism disorders, viral infections and type II diabetes. Furthermore, new technologies have been developed for faster and cost-effective peptide synthesis.

4.1.1

Main Targets of Peptide Synthesis

The *confirmation of suggested primary structures* by chemical total synthesis is the most convincing structural proof. Despite application of the latest techniques, misinterpretations in the sequence determination of peptides have in the past led to incorrect suggestions of primary structures, most notably for corticotropin, human growth hormone, and motilin. Erroneous sequences have been identified upon comparison of products obtained by total syntheses and by isolation from natural sources.

Moreover, in some cases the final structural proof has been obtained only by chemical syntheses.

The *design of bioactive peptide drugs* is of great importance since the value of peptides as drugs has been recognized for a long time. Numerous analogues of peptides have been synthesized in order to identify the structural parameters for biological activity. In the case of the classical peptide hormone oxytocin, early efforts succeeded in obtaining an analogue by the exchange of Gln⁴ with Thr, the analogue subsequently displaying much higher biological activity than the native hormone. [Thr⁴]oxytocin is thought to undergo a much more efficient interaction with the receptor compared with the native hormone, and hence this phenomenon has been considered a “chemical mutation” which would have succeeded in evolution, but probably only in the distant future. A systematic amino acid exchange is hampered by the variety of possible combinations of proteinogenic amino acids in bioactive peptides; however, combinatorial chemistry has today opened a series of completely new perspectives. In addition to the exchange of amino acids by other proteinogenic building blocks, the modification by isofunctional or isosteric amino acids is also a viable approach. Extensive investigations on the influence of the side chain on biological activity have been performed. Furthermore, modifications of the C- and N-terminal amino and carboxy groups, of functional groups in side-chain positions, and of the configuration of the peptide backbone are of central interest in structure–activity relationships (SAR). All such chemical modifications of a bioactive peptide drug have a major influence on the conformation, so that conformational studies of native drugs in comparison with their modified analogues are central targets of timely research. The conformational analysis of linear peptides and analogues with modified structure is not only very difficult but is also often ambiguous, because linear peptides are rather flexible unless stable secondary structures such as helices, β -sheets, turns, and other folds are present. As a result, conformational studies are usually performed with cyclic oligopeptides (often using NMR spectroscopy) because these compounds display a greater degree of conformational homogeneity. The main disadvantages of using a peptide drug are the low bioavailability and short half-life, especially after oral administration. Possibilities exist by which peptide degradation by proteases can be reduced, and uptake enhanced. For example, chemical modification (e.g., cyclization, D-amino acid substitution, or the incorporation of peptide mimetics) leads to the production of compounds with higher stability *in vivo*. Other strategies aimed at stabilizing peptides *in vivo* in order to increase their half-life include the coadministration of protease inhibitors, or the use of a specific formulation such as emulsions, liposomes, or nanoparticles. Another alternative is to use a pro-drug to alter the peptide’s physico-chemical properties. Peptides may be used as lead structures in order to create nonpeptidic bioactive compounds, the process often being supported by conformational studies and molecular modeling. In the course of this structure-based design, an amino acid sequence of a bioactive peptide is transformed into a nonpeptide lead compound. Here, functional and structural information is combined and the peptide backbone topography with respect to the pharmacophoric groups is reproduced in the small nonpeptide lead compound. These substances must display pharmacological activity, especially after oral administration, and must

imitate the peptide in terms of its receptor interaction (agonist activity); alternatively, it may block this interaction (antagonist activity). Last but not least, new sophisticated screening tools including library-array-technologies have been consolidated, and thus represent a fast lead finding process for modern drug discovery.

The *preparation of pharmacologically active peptides and proteins* is of great importance, both for the production of peptides on the mg scale for research or multi-ton quantities for drugs in the future. Although today more than 30 peptides are on the market and approximately 200 new peptide-based drugs are under different stages of development, it has been suggested that the market will double in the next few years. Based on economic considerations, classical pharmacologically active peptides, such as oxytocin, ACTH, and secretin, have been synthesized chemically. Furthermore, peptides like these are characterized by a higher purity than those isolated from natural sources. Synthetic secretin costs about 10% less than the natural product isolated from porcine intestine, and a similar situation applies to many other peptide drugs. In addition to economic reasons, the better accessibility via a synthetic approach is important because many peptide drugs occur naturally only in nanogram quantities. Moreover, chemical synthesis is often preferred in the case of peptides and proteins that occur specifically in humans, as isolation following autopsy is generally not a viable route. With the development of recombinant techniques, the industrial production of pharmacologically active proteins using biotechnological methods has become a viable route on a preparative scale. In particular, immune modulators and tissue proteins represent the focal point of promising new concepts in tumor therapy. In addition, diagnostics are being directed increasingly towards enzymes and monoclonal antibodies, while vaccines based on specialized proteins and peptides are also under development. Today, pharmacological research and medicine stand on the brink of a series of completely new developments, with the most important therapeutic areas likely to be those of cardiovascular conditions, tumors, and autoimmune or infectious diseases.

The *synthesis of model peptides* comprises an important task in peptide research. Tailor-made model peptides serve to study conformational behavior, using a variety of physico-chemical methods. Peptide substrates may also be used in studies of enzymology, while investigations into the antigenicity of larger proteins can be made using specialized multiple peptide synthesis techniques (see Chapter 8). Peptide epitope mapping especially is directed towards diagnostic targets and vaccine development. Protein–protein interactions often rely on the specific recognition of surface-exposed amino acid residues, which may be presented in either of two ways (Figure 4.1):

- as a continuous epitope, where the amino acids involved occur consecutively in the primary structure;
- as a discontinuous epitope, where the crucial amino acids are located in different positions within the protein sequence.

An overlapping series of short peptide sequences (6–15 amino acids) is synthesized on the basis of the protein primary structure. Binding studies then reveal the sequences necessary for protein–protein recognition. In future, model peptides

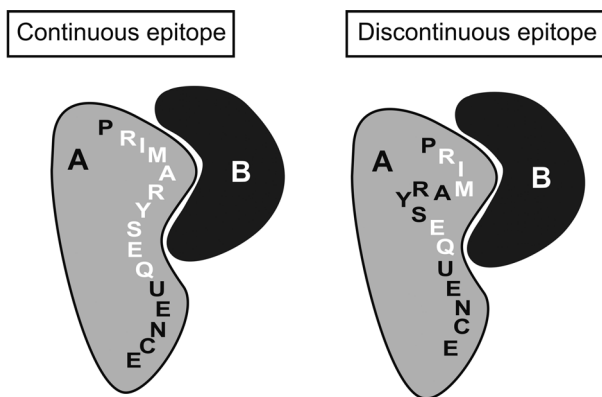


Figure 4.1 Schematic view of the interaction between two proteins A and B via continuous or discontinuous epitopes.

will serve as valuable tools in investigations of the tertiary structure of polypeptide chains. The question of whether the general rules of protein folding derived from the primary structure (amino acid sequence) can be maintained remains a controversial issue. Synthetic model peptides imitating protein surfaces can also be synthesized, and the flexibility of the peptide backbone can be reduced by conformational restrictions.

4.1.2

Basic Principles of Peptide Bond Formation

The formation of a peptide bond – resulting in a dipeptide – is seemingly a very simple chemical process. The two component amino acids are connected by a peptide (amide) bond with the elimination of water (Figure 4.2).

The synthesis of a peptide bond under mild reaction conditions can only be achieved after activation of the carboxy component (A) of one amino acid. The second amino acid (as the amino component, B) attacks the activated carboxy component in a nucleophilic attack, with formation of the dipeptide (A–B, Figure 4.3). If the amino function of the carboxy component (A) is unprotected, then formation of the peptide

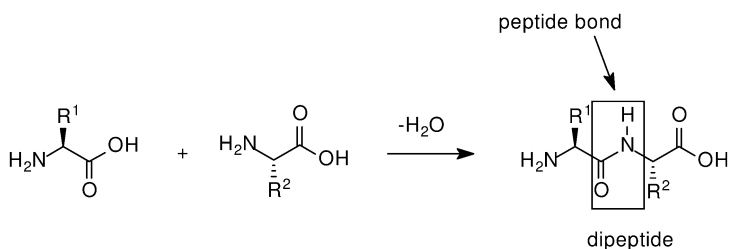


Figure 4.2 Simplified scheme of peptide bond formation.

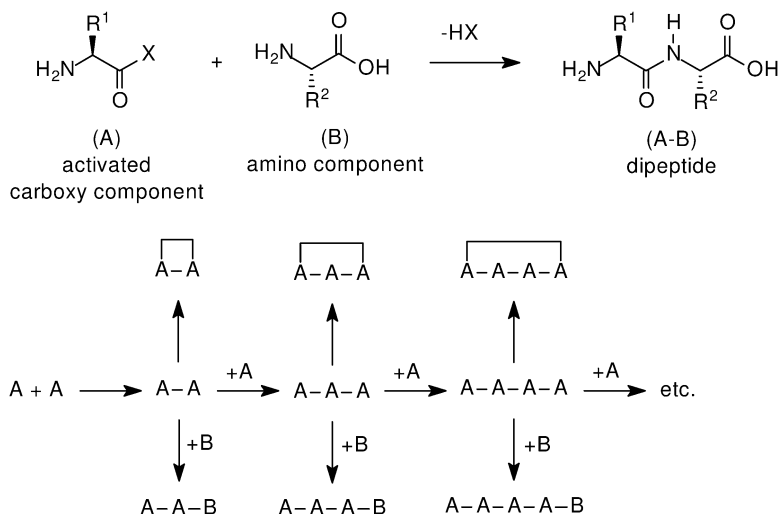


Figure 4.3 Schematic view of possible side reactions of a peptide synthesis involving an activated N^α -unprotected carboxy component (A). Linear and cyclic oligomers are formed besides the desired dipeptide A-B. X = activating group; R^1 , R^2 = amino acid side chains.

bond occurs in an uncontrolled manner (Figure 4.3, lower part). Linear and cyclic peptides are formed as undesired by-products, together with the target compound A-B. Consequently, during the course of peptide synthesis all functional groups not involved in peptide bond formation must be blocked, both temporarily and reversibly.

Peptide synthesis – the formation of each peptide bond – is therefore a three-step procedure (Figure 4.4).

1. The preparation of a partially protected amino acid is required. After this protection, the zwitterionic structure of the amino acids is no longer present.
2. The formation of the peptide bond occurs in two partial steps. The N-protected amino acid must be activated at the carboxy function in order to be converted into a reactive intermediate. Subsequently, the formation of the peptide bond occurs. This coupling reaction can be performed either as a one-pot reaction or in two separate, consecutive steps.
3. Selective or total cleavage of the protecting groups is carried out. Although total deprotection is required only when the peptide chain has been fully assembled, selective cleavage of the protecting groups is usually necessary in order to continue the peptide synthesis.

Peptide synthesis becomes further complicated by the fact that ten of the proteinogenic amino acids (Ser, Thr, Tyr, Asp, Glu, Lys, Arg, His, Sec, and Cys) have functional groups in the side chain which need to be selectively protected.

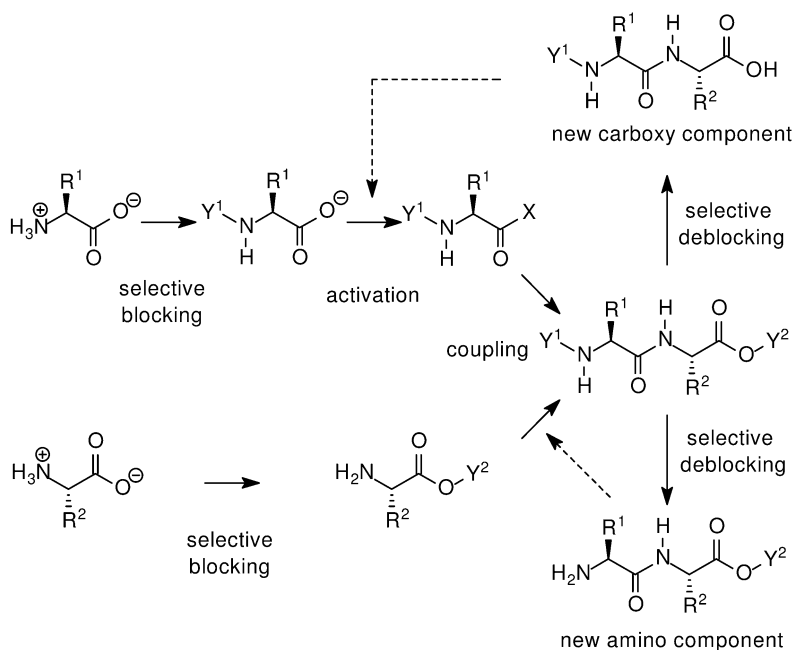


Figure 4.4 The multi-step process of peptide synthesis.

Y^1 = amino-protecting group; Y^2 = carboxy protecting group; R^1 , R^2 = amino acid side chains.

Because of the different requirements with respect to selectivity, a distinction must be made between the intermediary (temporary, or transient) and semipermanent protecting groups. The intermediary groups are used for temporary protection of the amino or carboxy function involved in subsequent bond formations. These groups must be cleaved selectively under conditions that do not interfere with the stability of peptide bonds already present, or that of semipermanent protecting groups at amino acid side chains. The semipermanent protecting groups are usually cleaved only at the end of the peptide synthesis or, occasionally, also at an intermediate stage. The different types of protecting groups will be described in Section 4.2.

Activation of the carboxy component and subsequent bond formation (the coupling reaction) should, under ideal conditions, occur at a high reaction rate and without racemization or formation of by-products. High yields are required upon application of equimolar amounts of both components, but unfortunately no such chemical coupling method yet exists which meets these requirements. The number of appropriate methods applicable to practical synthesis is rather small compared to the high number of coupling methods described. Further details will be provided in Section 4.3.

During the course of peptide synthesis, a number of reactions occur which involve functional groups that are usually connected to a chiral center (glycine is the only exception), and consequently there is a potential risk of racemization (see Section 4.4).

Cleavage of the protecting groups is the last step in the peptide synthesis cycle. Except for the synthesis of dipeptides, where total deblocking is required, the selective cleavage of protecting groups in order to extend the peptide chain is of major importance. The synthetic strategy requires thoughtful planning, which is described as the “strategy and tactics of peptide synthesis” (see Chapter 5). Depending on the strategy chosen, either the protecting group of the N^α -amino function is cleaved selectively, or the protecting group of the carboxy function is cleaved. The term “strategy” refers to the sequence of condensation reactions between single amino acids to produce a peptide. Usually, a distinction is made between stepwise synthesis and segment condensation (fragment condensation). When peptide synthesis is performed in solution (also referred to as “conventional synthesis”), for example with difficult sequences, the stepwise peptide chain elongation is limited in most cases to the assembly of shorter fragments. When the synthesis of a longer peptide is planned, the target molecule must be divided into appropriate segments. Suitable fragments have to be identified, perhaps where only minimal C-terminal epimerization occurs during fragment condensation. Following stepwise assembly of the single segments, the latter are connected to produce the target compound. The tactics of peptide synthesis also include selection of the most suitable combination of protecting groups, together with application of the optimal coupling methods for each segment condensation.

Solid-phase peptide synthesis (SPPS) is, in its original version, a variant of the stepwise synthesis of peptides and proteins. It is a synthetic concept where the growing peptide chain is attached to an insoluble polymeric support, and was first described by Robert Bruce Merrifield in 1963. Today, this method is referred to as the Merrifield synthesis in honor of the 1984 Nobel laureate (for details, see Section 4.5). Nowadays, segment condensations may also be performed on a polymeric support (see Chapter 5, Section 4.3).

4.2

Protection of Functional Groups

Temporary and semipermanent protecting groups are distinguished with respect to cleavage selectivity. A temporary protecting group must fulfil the following requirements:

1. Introduction of the protecting group leads to an amino acid derivative that is no longer present in a zwitterionic structure.
2. Cleavage must occur without hampering the stability of semipermanent protecting groups or peptide bonds.
3. Racemization must not occur during all necessary operations.
4. Stability and characterization of protected intermediates must be favorable.
5. Solubility of the protected amino acid components must be favorable.

According to the tactical requirements, semipermanent protecting groups are usually cleaved at the end of the peptide synthesis. In this reaction, only those

deblocking reagents that do not damage the final product can be used. The requirements with respect to cleavage selectivity are less stringent because temporary protecting groups may also be cleaved in the final deblocking step. Criteria (3), (4), and (5) must also be applied to the semipermanent protecting groups.

The selection of protecting groups must be carried out in accordance with the strategy envisaged for the synthesis of a particular peptide (see Chapter 5). Protecting group orthogonality is a criterion that is especially important for the success of a peptide synthesis.

Protection schemes rely on either quantitative rate differences of deprotection reactions or on mechanistically different deprotection reactions in the case of orthogonal protecting groups [1]. Orthogonality means that subsets of protecting groups present in a molecule can be cleaved selectively under certain reaction conditions, while all other protecting groups remain intact. Often, orthogonality refers to a mechanism-based distinction, for example between semipermanent and temporary protecting groups. It generally provides the opportunity of substantially milder overall reaction conditions.

In the following section, the most important protecting groups are discussed in the context of the functional group to be blocked.

4.2.1

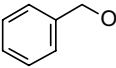
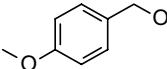
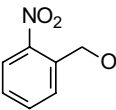
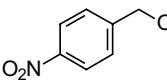
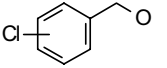
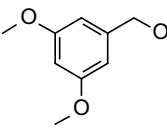
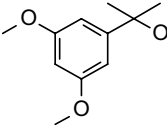
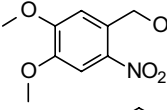
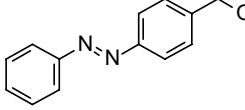
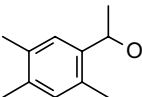
N^α-Amino Protection

Protecting groups of this type are used for the N^α-amino group of all proteinogenic amino acids (including the imino group of Pro), for N^ω-amino groups (e.g., Lys, Orn) and also, if necessary, for the temporary blocking of N-acyl amino acid hydrazides and other functional groups. In principle, an amino group can be blocked reversibly by acylation, alkylation, and alkyl-acylation. Protecting groups based on sulfur and phosphorus derivatives have also been developed. Salt formation at the amino group does not offer a genuine protection during the operations of peptide synthesis. Several hundred different amino-protecting groups have been developed during the past decades – which confirms the non-existence of a universal, ideal amino-protecting group. A classification can be made based on the structure of the protecting group, or on the cleavage conditions: acidolysis, base cleavage, reduction/oxidation, nucleophilic substitution, and photolysis. Because most peptides are sufficiently stable under moderately acidic conditions, amino-protecting groups with different labilities towards acidic deblocking agents are preferred. Alternately, the base-labile compound 9-fluorenylmethoxycarbonyl group (Fmoc) has found widespread application. It has been estimated that, in practical terms, over 80% of all syntheses use these two deblocking methods. Some protecting groups may be cleaved by hydrogenolysis. Protection of the amino group can also be achieved easily by acylation and by alkylation. The disadvantage of structurally and chemically simple acyl residues (e.g., acetyl-, benzoyl-, monochloroacetyl-) is that the carboxamide group of the protecting function is chemically very similar to the peptide bond, which renders selective cleavage very difficult or even impossible, and this eventually led to the development of urethane-type amino-protecting groups.

4.2.1.1 Alkoxycarbonyl-Type (Urethane-Type) Protecting Groups

A variety of urethane-type amino-protecting groups, which are cleavable under different reaction conditions, is available. Selected representatives are listed in Table 4.1; the most prominent groups are discussed in the following section.

Table 4.1 Selected amino-protecting groups (PG) of the urethane type $R'-O-CO-NH-R$ (H_2N-R = amino acid).

Name	Abbr.	$R'-O$	Cleavage	Ref.
Benzyloxycarbonyl	Z		H_2/Pd ; $HBr/AcOH$; $Na/liq. NH_3$	[2]
4-Methoxybenzyloxy carbonyl	Z(OMe)		$TFA/H_2/Pd$; $Na/liq. NH_3$	[3, 4]
2-Nitrobenzyloxy carbonyl	Z(2-NO ₂)		like Z, but more labile); photolysis	[5]
4-Nitrobenzyloxy carbonyl	Z(NO ₂)		like Z, but more stable to $HBr/HOAc$	[6, 7]
Chlorobenzyloxy carbonyl	Z(Cl)		like Z, but more stable to H_2/Pd or $HBr/AcOH$	[8–12]
3,5-Dimethoxybenzyl oxycarbonyl	Z(3,5-OMe)		photolysis	[13]
α,α -Dimethyl-3,5-dimethoxy benzyloxycarbonyl	Ddz		photolysis; 5% TFA in CH_2Cl_2	[14]
6-Nitroveratryloxycarbonyl (4,5-Dimethoxy-2-nitrobenzyloxycarbonyl)	Nvoc		like Z; photolysis	[5]
4-Phenylazobenzyloxy-carbonyl	Pz		H_2/Pd , $HBr/AcOH$	[15]
α -Methyl-2,4,5-trimethylbenzyl oxycarbonyl	Tmz		3% TFA in $CHCl_3$,	[16]

(Continued)

Table 4.1 (Continued)

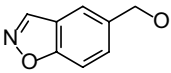
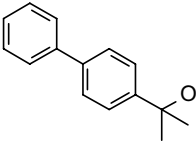
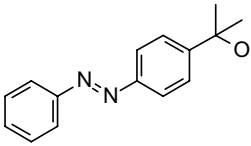
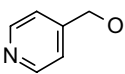
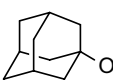
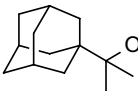
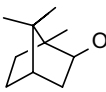
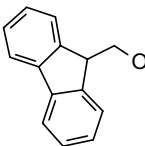
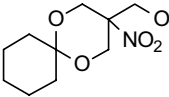
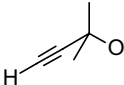
Name	Abbr.	R'-O	Cleavage	Ref.
Benzisoxazol-5- yloxycarbonyl	Bic		triethylamine in DMF and solvolysis at pH 7	[17]
2-(Biphenyl-4-yl)-2- propoxycarbonyl	Bpoc		80% AcOH, MgClO ₄	[18, 19]
(4-Phenylazophenyl)- isopropoxycarbonyl	Azoc		like Bpoc	[20]
Isonicotinyloxycarbonyl	iNoc		acid stable Zn/AcOH H ₂ /Pd;	[21]
<i>tert</i> -Butoxycarbonyl	Boc		TFA; TFA/ CH ₂ Cl ₂ ; HCl/or- ganic solvent	[3, 22]
2-Cyano- <i>tert</i> - butoxycarbonyl	Cyoc		weak base (e.g. aq. K ₂ CO ₃ , triethylamine)	[23]
2,2,2-Trichloro- <i>tert</i> - butoxycarbonyl	Tcboc		Co(II)-phthalocya- nine; Zn/AcOH	[24]
Adamantyl-1- oxycarbonyl	Adoc		TFA	[25]
1-(1-Adamantyl)-1- methylethoxy-carbonyl	Adpoc		3% TFA in CH ₂ Cl ₂ ; labile to H ₂ /Pd	[26]
Isobornyloxycarbonyl	Iboc		TFA; stable to H ₂ /Pd and bases	[27]
Fluorenyl-9- methoxycarbonyl	Fmoc		piperidine; DBU; 2-aminoethanol; morpholine; liq. NH ₃	[28, 29]

Table 4.1 (Continued)

Name	Abbr.	R'-O	Cleavage	Ref.
(2-Nitrofluoren-9-yl)methoxycarbonyl	Fmoc (NO ₂)		photolysis	[30]
2-(4-Toluenesulfonyl)ethoxycarbonyl	Tsoc		NaOC ₂ H ₅ / ethanol	[31]
Methylsulfonyl ethoxycarbonyl	Msc		base catalysed β-elimination	[32]
2-(4-Nitrophenylsulfonyl)ethoxycarbonyl	Nsc		like Fmoc	[33–35]
2-(tert-Butylsulfonyl)-2-propenyloxycarbonyl	Bspoc		nucleophile	[36]
1,1-Dioxo-benzo[b]thien-2-ylmethoxycarbonyl	Bsmoc		nucleophile, e.g. 2% tris(2-amino-ethyl) amine	[37, 38]
2-(Methylsulfonyl)-3-phenyl-2-propenyloxycarbonyl	Mspoc		nucleophile	[39]
Allyloxycarbonyl	Alloc		Pd ⁰ /nucleophile	[40–42]
2-(Trimethylsilyl)ethoxycarbonyl	Teoc		F ⁻ , e.g. TBAF	[43]
Triisopropylsilyl ethoxycarbonyl	Tipseoc		F ⁻ , e.g. TBAF	[44]
Piperidinylloxycarbonyl	Pipoc		electrolysis; H ₂ /Pd	[45]

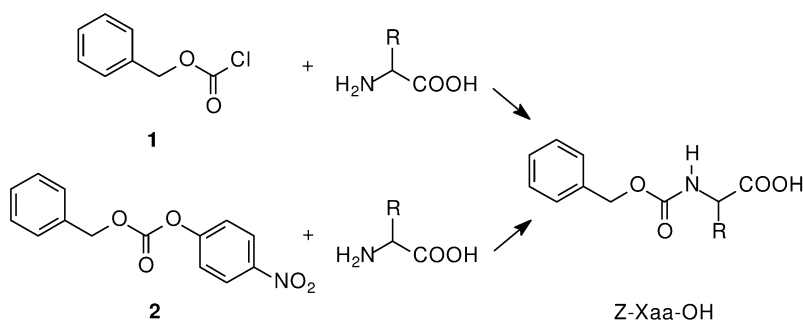
(Continued)

Table 4.1 (Continued)

Name	Abbr.	R'-O	Cleavage	Ref.
Cyclopentyloxycarbonyl	Poc		HBr/AcOH; stable to H ₂ /Pd	[3]
3-Nitro-1,5-dioxaspiro [5.5]undec-3- ylmethoxycarbonyl	PTnm		"relay deprotection": H ₃ O ⁺ , then pH 8.5	[46]
2-Ethynyl-2-propyl oxycarbonyl	Epoc		H ₂ , Lindlar-cat.	[47]

Benzyloxycarbonyl group The introduction of the benzyloxycarbonyl group into peptide synthesis provided a major breakthrough in the development of modern peptide chemistry. Today, this amino-protecting group is among those used most frequently, and is abbreviated as Z, in honor of its inventor, Leonidas Zervas. In older references, the three-letter abbreviation Cbz is also referred to as the carbobenzyloxy (Cbo) group. Introduction of the Z group in amino acids is usually performed by reaction with benzylchloroformate **1** (the correct names benzyloxycarbonyl chloride or benzylchlorocarbonate are seldom used) under Schotten–Baumann conditions in the presence of NaOH, NaHCO₃ or MgO, mostly in aqueous organic solvents under vigorous stirring (Scheme 4.1).

Amino acid esters can be converted into the Z-protected derivatives in chloroform. Selective protection in the case of amino acids with functional groups in the side chain often requires special reaction conditions. Arginine may be selectively N^α-monoprotected using benzyl chloroformate in a buffered NaHCO₃–NaOH solution at pH 9–11 [48]. Usually, Z-amino acids are obtained in high yields in crystalline form, but products obtained as an oil can be crystallized as salts with dicyclohexylamine. Benzyl-4-nitrophenylcarbonate **2** and similarly activated benzyl derivatives can also be used to introduce this group. Cleavage of the Z group is



Scheme 4.1

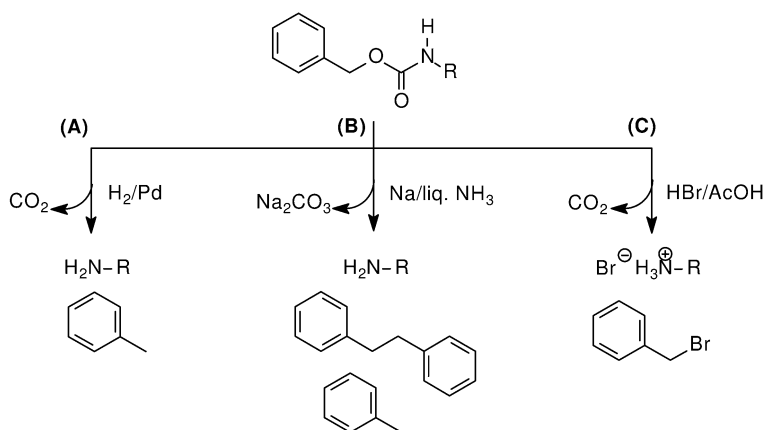


Figure 4.5 Cleavage conditions of the benzyloxycarbonyl group. $\text{H}_2\text{N-R}$ = amino acid.

achieved by acidolysis using hydrogen bromide in acetic acid, liquid HBr, boiling trifluoroacetic acid (TFA), HBr/TFA, liquid hydrogen fluoride (HF), or by catalytic hydrogenation using different catalysts under a broad scope of reaction conditions (Figure 4.5).

The catalytic hydrogenation proceeds smoothly in organic solvents (e.g., acetic acid, alcohols, dimethylformamide (DMF)) or in aqueous organic solution with palladium black, palladium on charcoal, or palladium on barium sulfate. Only toluene and carbon dioxide are formed in addition to the desired free peptide. Cessation of CO_2 evolution or H_2 consumption is used to monitor the cleavage reaction. Although hydrogenolytic cleavage of Z groups fails when cysteine/cystine is present in the peptide, the reaction is successful in the presence of $\text{BF}_3 \cdot \text{OEt}_2$. Acidolysis is performed preferably using 2N hydrogen bromide in acetic acid, although several other variants (e.g., hydrogen chloride, hydrogen iodide) and solvents (e.g., dioxan, nitromethane, TFA, tetrachloromethane) have been suggested. Side reactions can also occur for this standard method; for example, O-acetylations have been observed at threonine or serine, while S-transetherifications may occur in the presence of methionine. Tryptophan or nitroarginine building blocks may be destroyed, and benzyl esters or carboxamide groups are sometimes cleaved under these conditions. However, these side reactions may be minimized under certain reaction conditions. The Z group can also be cleaved with anhydrous liquid hydrogen fluoride (HF method).

More than 20 different benzyloxycarbonyl type groups with a variety of substituents have been described, mainly with regard to tuned sensitivity towards acidolytic cleavage reagents. The 4-methoxybenzyloxycarbonyl residue provides a methodological advantage because the protected amino acids are crystalline. It can be cleaved in the presence of the Z group with anhydrous TFA at temperatures below 0°C . Z groups with halogen substituents in the 3- or 4-position of the phenyl ring are also of interest because they lead to higher stability of the modified Z group towards acidic reagents. This can be used to better differentiate between protecting groups of the Z type and Boc type (see below). The (3,5-dimethoxybenzyl)oxycarbonyl,

(2-nitrobenzyl)oxycarbonyl, and (6-nitroveratryl)oxycarbonyl groups can be cleaved photochemically.

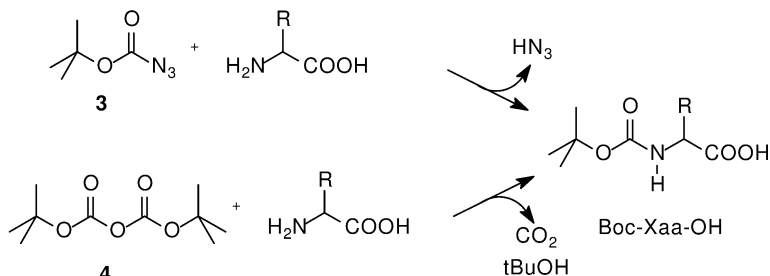
***tert*-Butoxycarbonyl group** The *tert*-butoxycarbonyl group (Boc) is, like the Z and the Fmoc group, one of the most important amino-protecting groups. It should also be mentioned that amino-protecting groups of the urethane type safeguard racemization-free peptide bond formations during the stepwise assembly of a peptide chain starting from the C-terminus.

The *tert*-butoxycarbonyl azide **3** is an excellent acylation reagent. It reacts with the salts of amino acids in water/dioxan mixtures in the presence of tertiary amines or magnesium oxide or under pH control with 2N to 4N NaOH to give the Boc-protected amino acids Boc-Xaa-OH, but also with amino acid esters in pyridine. pH-Stat-controlled reaction conditions are advantageous.

Although this acylation protocol has been proved to be very efficient, further methods for the introduction of the Boc group have been described, mainly because of the explosive nature of azides. The *tert*-butoxycarbonyl fluoride obtained from chlorocarbonyl fluoride and *tert*-butanol at -25°C has attracted special attention because it gives Boc-protected amino acid in high yields under pH-stat conditions. *tert*-Butyl-*S*-4,6-dimethylpyrimidyl-2-thiocarbonate, 2-*tert*-butoxycarbonyloximino-2-phenylacetonitrile and, preferably, di-*tert*-butyldicarbonate (Boc)₂O **4** (Scheme 4.2) are also used as reagents. Amino acid salts can be used in aqueous solution with solubilizing additives such as dioxan or tetrahydrofuran.

The Boc protecting group is compatible with most coupling methods for peptide synthesis. It can be cleaved under mild acidolytic conditions, and is resistant towards catalytic hydrogenation, alkaline hydrolysis and reduction with Na/liquid ammonia. Hydrogen chloride in acetic acid, dioxan, ether, and ethyl acetate are important cleavage reagents. A cleavage method frequently used is treatment with non-aqueous TFA at temperatures around 0°C , which smoothly cleaves Boc groups (Figure 4.6).

Undesired *tert*-butylations may occur during acidolysis at nucleophilic sites of the peptide chain. For instance, the indole system of tryptophan or the thioether group of methionine can be *tert*-butylated. The intermediate carbenium ions responsible for these side reactions are usually trapped using scavengers such as anisole or thioanisole. Further deblocking reagents are $\text{BF}_3 \cdot \text{OEt}_2$, 2-mercaptoethanesulfonic acid, aqueous TFA, 98% formic acid, an excess of sodium *tert*-butoxide, or ceric ammonium nitrate. A mild method using tetrabutylammonium fluoride (TBAF) has



Scheme 4.2

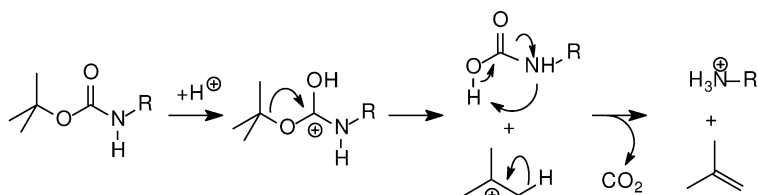


Figure 4.6 Mechanism of acidolytic cleavage of the *tert*-butoxycarbonyl group.
 $\text{H}_2\text{N-R}$ = amino acid or peptide.

also been reported [49]. However, a smooth reaction has not been observed in all cases. This variety of reagents can be used when other protected functional groups and the Z group are present. Modified analogues of the *tert*-butoxycarbonyl group have been described on many occasions, and some representatives are discussed in the following section. The (α,ω -dimethyl-3,5-dimethoxybenzyl)oxycarbonyl group (Ddz) can be cleaved both acidolytically with 5% TFA in methylene chloride, and also photolytically. Therefore, it provides an interesting variant for special applications. The 2-(4-biphenyl)-2-propoxycarbonyl residue (Bpoc) can be cleaved with 80% acetic acid, and is therefore often combined with protecting groups for ω -amino-, hydroxy-, and carboxy functionalities based on the *tert*-butyl type. The 2-ethynyl-2-propyloxycarbonyl group is cleaved by catalytic hydrogenation (with Lindlar catalyst), even in the presence of methionine-containing peptides. Alternately, it can be cleaved by acidic hydrolysis under the formation of a cyclic carbonate. A different variant is the 2-cyano-*tert*-butoxycarbonyl residue (Cyoc) which can be cleaved by β -elimination under weakly basic conditions.

9-Fluorenylmethoxycarbonyl group The 9-fluorenylmethoxycarbonyl group (Fmoc) is the only amino-protecting group of the urethane type of widespread practical importance that can be cleaved under mildly basic conditions [28].

The sensitivity of Fmoc-protected amino acids (Figure 4.7; 5), especially towards secondary amines, enables deblocking of the amino group with dilute solutions of piperidine or diethylamine in DMF under mild conditions. Usually, 20% piperidine in DMF is applied routinely when the cleavage of the Fmoc group occurs within seconds at room temperature. Other reagents, such as 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) or fluoride ion may be used as an alternative. The reaction proceeds according to an E1cB mechanism, with initial proton abstraction to give the stabilized dibenzocyclopentadienyl anion. The dibenzofulvene formed reacts with piperidine to give a stable adduct as a by-product of the cleavage reaction (Figure 4.7). In cases where no primary or secondary amine is used, the addition of 2% piperidine is recommended to trap the dibenzofulvene.

Under these conditions, neither Z nor Boc groups nor other modern amino-protecting groups usually applied are attacked. In order to prevent premature cleavage under peptide-coupling conditions, addition of the weakly acidic additive 1-hydroxybenzotriazole has been recommended. CaCl_2 addition has been reported to increase Fmoc lifetime, and hence may be used for the synthesis of Fmoc-protected peptide segments by saponification of C-terminal esters or linkers [50]. Interestingly, the Fmoc group is not totally inert towards standard hydrogenolytic cleavage conditions.

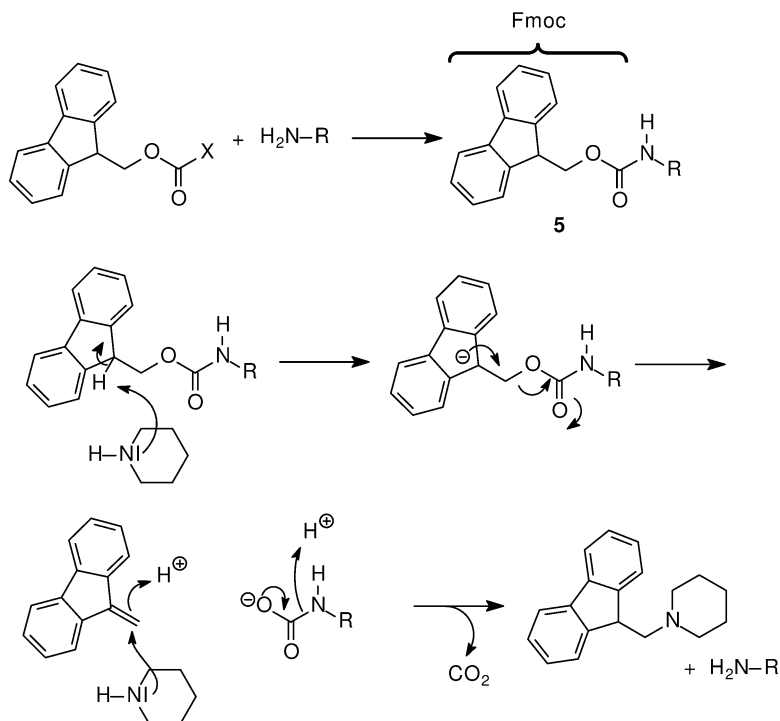


Figure 4.7 Mechanism of Fmoc cleavage. $\text{H}_2\text{N-R}$ = amino acid or peptide.

The reagents Fmoc-Cl and, alternately, Fmoc-OSu (9-fluorenylmethyl-*N*-succinimidyl carbonate), and Fmoc-2-mercaptobenzothiazole [51] have been recommended for the formation of Fmoc-protected amino acids. The latter Fmoc-2-MBT reagent is proposed for avoiding the formation of undesired side reactions. Although, the Fmoc group was introduced into peptide chemistry in 1970, its application in the Merrifield solid-phase peptide synthesis was reported much later. Nowadays, the Fmoc tactic is one of the most preferred synthetic concepts in solid-phase peptide synthesis. The acid stability, combined with the mildly basic cleavage conditions, allows the application of acid-labile semipermanent protecting groups of the *tert*-butyl type for side-chain functionalities in solid-phase peptide synthesis.

Monoisooctyl-Fmoc (Mio-Fmoc, 2-(2-ethylhexyl)-fluorene-9-methoxycarbonyl) and diisooctyl-Fmoc (Dio-Fmoc, 2,7-bis(2-ethylhexyl)-fluorene-9-methoxycarbonyl) are base-labile protecting groups derived from the Fmoc group with improved solubility in organic solvents [52].

New base-labile N^α -protecting groups The 2-(4-nitrophenylsulfonyl)ethoxycarbonyl group Nsc (see Table 4.1) has recently been considered as a valuable alternative to the Fmoc group, because it is more base-stable and, therefore, less sensitive towards weakly basic media [53–55]. Nsc is more hydrophilic than Fmoc, and less prone to aspartimide formation. Like Fmoc, Nsc forms in the deprotection reaction an olefin

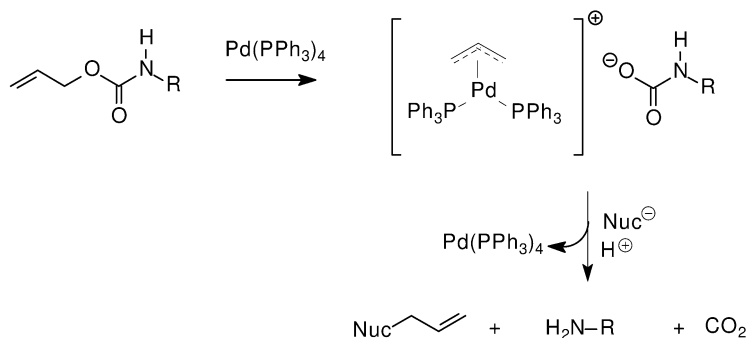


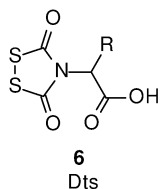
Figure 4.8 Mechanism of Alloc cleavage. $\text{H}_2\text{N-R}$ = amino acid or peptide.

that subsequently is trapped by the secondary base piperidine. In the Fmoc case, the formation of the dibenzofulvene piperidine adduct is reversible, while the corresponding adduct in the Nsc case is formed irreversibly.

Carpino *et al.* developed the new base-labile amino-protecting groups Bspoc, Bsmoc, and Mspoc (see Table 4.1) [36–39]. Cleavage relies on the conjugate addition of a nucleophile (e.g., secondary amine) with concomitant liberation of the carbamate. Bsmoc is stable towards acids and tertiary amines, but is cleaved readily with secondary amines. As the deprotection step is, simultaneously, the scavenging step, fewer side reactions should occur.

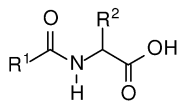
Protecting groups that are orthogonal to Boc and Fmoc The N^α -allyloxycarbonyl (Alloc) group [40–42] may be used for the synthesis of labile derivatives (see Sections 6.3–6.5). It is completely orthogonal to Boc and Fmoc and, therefore, finds application in the synthesis of cyclic peptides (see Chapter 6.1). The deprotection step involves palladium(0)-catalyzed allyl transfer (Figure 4.8) to various nucleophiles (e.g., amines, amine-borane complexes, carboxylates, organosilanes, organostannanes) [56]. Amine-borane complexes are especially suited as allyl scavengers under neutral conditions without undesired N-allylation of the deprotected product [56]. Tandem deprotection-acylation of N^α -Alloc derivatives has also been reported [57].

The dithiasuccinoyl group Dts **6** [58–60] may formally be regarded as a urethane thio analogue that can be cleaved by treatment with thiols. It is stable to acids and bases and, consequently, orthogonal to Boc or Fmoc. The 2-(4-trifluoromethylphenylsulfonyl)ethoxycarbonyl group (Tsc), which is cleaved with 0.1N LiOH, THF/ H_2O , has been described to be orthogonal to Fmoc [61].



4.2.1.2 Carboxamide-Type Protecting Groups

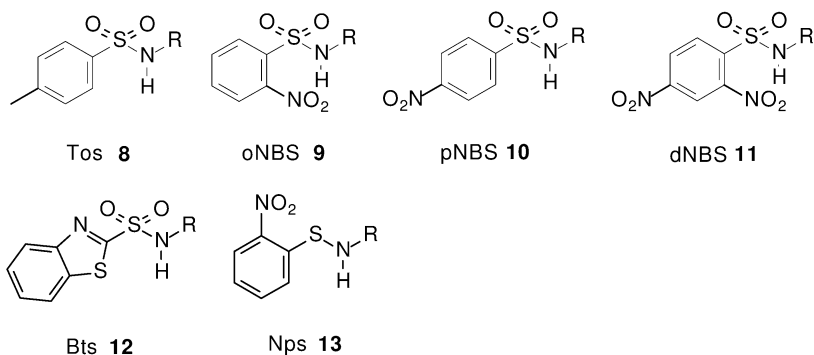
Representatives of the carboxamide-type protecting groups are the formyl group For **7** ($R^1 = H$) which can be cleaved by solvolysis, oxidation, or hydrazinolysis, and the trifluoroacetyl group Tfa **7** ($R^1 = CF_3$), which is labile towards alkaline conditions.



7 For: $R^1 = H$
Tfa: $R^1 = CF_3$

4.2.1.3 Sulfonamide and Sulfenamide-Type Protecting Groups

Arenesulfonyl groups feature some properties that render them potential amino-protecting groups. In the past, the 4-toluolsulfonyl (tosyl, Tos) group **8** was mainly used, though this suffers from the disadvantage that it is only cleavable by reduction with Na/NH_3 . In principle, N^α -sulfonamide protecting groups allow N-alkylation of amino acids [62, 63] and do not suffer from racemization via the oxazolone mechanism (cf. Section 4.4.2) [64]. N^α -Arenesulfonyl amino acid chlorides are highly reactive species that are well suited for peptide synthesis [65]. Consequently, N^α -arenesulfonamide-type protecting groups experienced a renaissance with the introduction of the *o*- and *p*-nitrobenzenesulfonyl groups (*o*Nbs **9**, *p*Nbs **10**) [62, 66] and of dNbs **11** [67] and Bts **12** [63]. These protecting groups can be removed by treatment with thiophenol or alkanethiols, and hence are orthogonal to *tert*-butyl and 9-fluorenylmethyl-type protecting groups.

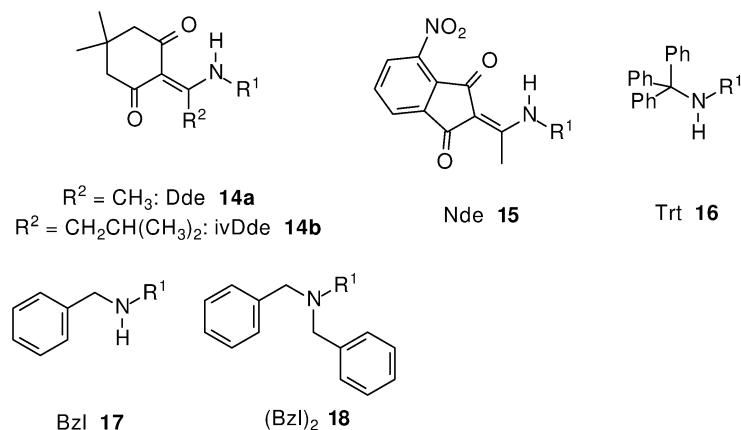


The 2-nitrophenylsulfenyl (Nps) group **13**, which has been used in solution-phase peptide synthesis, may be removed by HCl in inert solvents or with the use of thiol reagents [68–70].

4.2.1.4 Alkyl-Type Protecting Groups

The protecting groups 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl Dde **14a** ($R^2 = CH_3$) [71, 72], 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl ivDde

14b ($R^2 = iBu$) [73], and 1-(4-nitro-1,3-dioxindan-2-ylidene)ethyl Nde **15** [74] are alkyl-type protecting groups that are orthogonal to Boc and Fmoc, because they are cleaved by hydrazinolysis.



The most popular protecting group of the alkyl type is the triphenylmethyl (Trt) group **16** that can be cleaved acidolytically under mild conditions. The activation of Trt-protected amino acids is hampered because of the steric hindrance. Mono- or disubstituted N-benzyl protecting groups **17**, **18** have received very little attention in peptide synthesis, but they can be cleaved by catalytic hydrogenation in the presence of palladium black at elevated temperatures. This fact probably led to the finding that the exchange of an ethyl by a benzyl residue in the ethoxycarbonyl group, as described by Fischer, could result in an amino-protecting group being cleavable by hydrogenolysis. This paved the way for the protecting groups of the urethane type (as the protecting groups based on carbamates are referred to).

4.2.2

C^α-Carboxy Protection

There are few exceptions (e.g., [75–77], cf. Section 5.1.1) from the usual synthetic strategy where a peptide chain is assembled from the C-terminus to the N-terminus (C → N direction). Consequently, C-terminal protection has not been investigated as extensively as N-terminal protection. C^α-carboxy-protecting groups in solution-phase synthesis, and also as linker moieties in solid-phase synthesis, must be orthogonal to the temporary N^α-protecting group.

In principle, it is possible to perform peptide synthesis without complete protection of the carboxy group of amino components. Removal of the zwitterionic structure of the corresponding amino acid or the peptide is achieved by salt formation. Here, alkali or alkaline earth metal salts of amino acids can usually be reacted in water or aqueous dioxan with the corresponding activated carboxy components. The corresponding carboxylate salts of tertiary bases (*N*-methyl morpholine, *N*-ethyl piperidine, triethy-

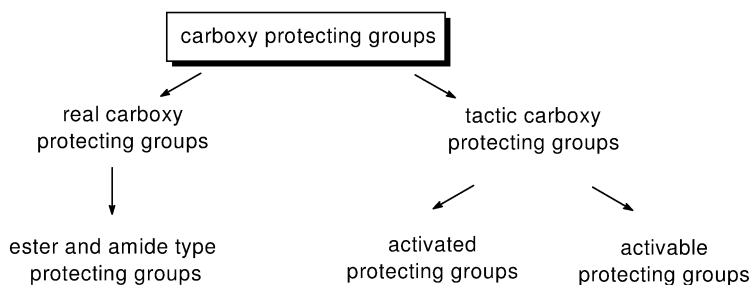


Figure 4.9 Classification of carboxy-protecting groups.

lamine, 1,1,3,3-tetramethylguanidine, etc.) have been recognized as important for peptide synthesis in organic solvents, especially in DMF. The mixed anhydride and azide methods are, besides active esters, best suited for the so-called salt coupling. A major advantage is that, after the coupling reaction, the carboxy group is liberated by acidification. This protocol is not universally applicable, as side reactions or solubility problems may occur without genuine protection of the carboxy group of the amino component used in the reaction.

Esters of the methyl, ethyl, benzyl, *tert*-butyl, or aryl type are appropriate for the reversible blocking of the carboxy group. Carboxy-protecting groups may be divided into two categories on the basis of strategic-tactic criteria (Figure 4.9):

- The first group contains protecting groups which can be deblocked normally after the end of the synthesis by regeneration of the free carboxy group.
- The second group contains derivatives that are activated towards aminolysis at the end of a synthesis, either immediately or after special chemical transformation.

Although these two different types have been classified as genuine and nongenuine carboxy-protecting groups, a better discrimination between them may be to refer to “real” and “tactic” carboxy-protecting groups. The expression “tactic” protecting groups emphasizes the application of these functionalities in special synthetic strategies where the mere role of a protecting group is exceeded.

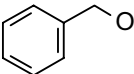
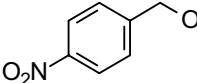
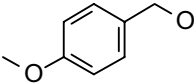
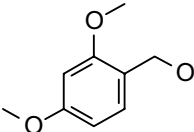
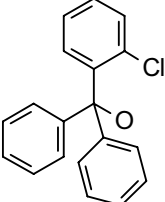
4.2.2.1 Esters

Activated carboxy-protecting groups can be used only under special conditions for peptide synthesis. In order to safeguard the desired reaction pathway, the activation method used for the coupling must have a higher acylation potential compared with the activated carboxy-protecting group of the amino component. This so-called “backing-off” strategy makes use preferentially of active esters as carboxy-protecting groups, and the mixed anhydride method for coupling. In this case, the N-protected carboxy component is transformed into an active ester. The N $^{\alpha}$ -protecting group is then cleaved and subsequently acylated with another N-protected amino acid to form the protected dipeptide with the active ester moiety located at the C-terminus. The mixed anhydride method is mainly applied in these cases. Successive repetition of these steps allows the formation of segments in a stepwise manner, starting from the C-terminus and without practical risk of racemization. These segments can then

be connected with other segments via the activated C-terminus to give larger peptides. Carboxy-protecting groups that, potentially, can be activated behave like real protecting groups during the coupling reaction, and the risk of undesired side reactions by uncontrolled self-condensation can be neglected in practical cases. One example of an ester that is transformed into the reactive form immediately before the envisaged coupling step is the 4-(methylthio)phenyl ester, which can be transformed oxidatively into the 4-(methylsulfonyl)phenyl ester.

Most of the currently known carboxylic protecting groups are esters derived from either primary, secondary, or tertiary alcohols (Table 4.2).

Table 4.2 Selected carboxy-protecting groups of the ester type RCO–OR' (RCO–OH = amino acid with a side chain carboxy group).

Name	Abbr.	R'–O	Cleavage	Ref.
Methyl	Me		alkaline or enzymatic hydrolysis	[78]
Ethyl	Et		alkaline or enzymatic hydrolysis	[79]
Benzyl	Bzl		H ₂ /Pd; HBr/AcOH; liq. HF; alkaline hydrolysis	
4-Nitrobenzyl	Nbz		H ₂ /Pd; alkaline hydrolysis; stable to HBr/AcOH	
4-Methoxybenzyl	Mob		TFA/anisole; H ₂ /Pd; liq. HF; HCl/nitromethane; alkaline hydrolysis	
2,4-Dimethoxybenzyl	2,4-Dmb		1% TFA/CH ₂ Cl ₂	
o-Chlorotrityl	Trt(2-Cl)		1% TFA/CH ₂ Cl ₂	

(Continued)

Table 4.2 (Continued)

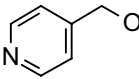
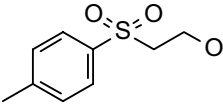
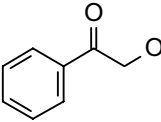
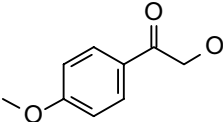
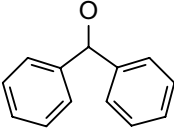
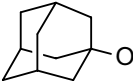
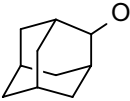
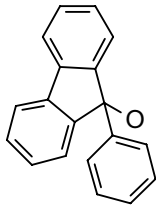
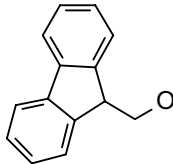
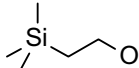
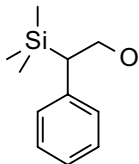
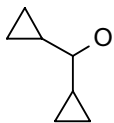
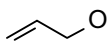
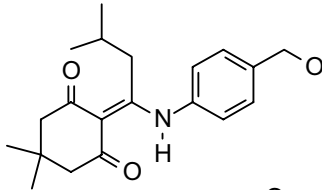
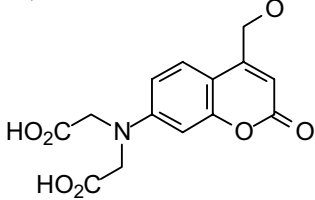
Name	Abbr.	R'-O	Cleavage	Ref.
4-Picolyl (pyridyl-4-methyl)	Pic		H ₂ /Pd; alkaline hydrolysis; electrolytic reduction	
2-(Toluene-4-sulfonyl)ethyl	Tse		β-elimination with Na ₂ CO ₃ in H ₂ O/dioxan	
Phenacyl	Pac		sodium thiophenolate; H ₂ /Pd; Zn/AcOH	
4-Methoxyphenacyl	Pac(OMe)		UV-photolysis at 20 °C	
Diphenylmethyl (Benzhydryl)	Dpm		H ₂ /Pd; TFA at 0 °C; BF ₃ ·OEt ₂ /AcOH at 25 °C; HCl/AcOH	
<i>tert</i> -Butyl	tBu		TFA; HCl/AcOH; HBr/AcOH; BF ₃ ·OEt ₂ /AcOH	
Cyclohexyl	Cy		strong acids	
1-Adamantyl	1-Ada		TFA	
2-Adamantyl	2-Ada		strong acids	

Table 4.2 (Continued)

Name	Abbr.	R'-O	Cleavage	Ref.
9-Phenylfluoren-9-yl	Pf		>1% TFA/ scavenger/ CH ₂ Cl ₂	[80]
9-Fluorenylmethyl	Fm		piperidine; TBAF	[81, 82]
2-Trimethylsilylethyl	TMSE		TBAF	[83, 84]
2-Phenyl-2-trimethylsilylethyl	PTMSE		TBAF	[85]
Dicyclopropylmethyl	Dcpm		1% TFA/ CH ₂ Cl ₂	[86]
Allyl	All		Pd(0)/ nucleophile	[41]
4-{N-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl]-amino}benzyl	Dmab		Hydrazine	[87]
{7-(bis(carboxymethyl)amino (coumarin-4-yl) methyl	BCMAMC		UV irradiation at 405 nm	[88]

For additional information see [89–91].

Methyl and ethyl esters have been used for peptide synthesis by Curtius and Fischer, and can be obtained easily on reaction of the amino acid with the corresponding alcohol and thionyl chloride at low temperature. Presumably, the alkyl chlorosulfinate (AlkO-SO-Cl) is formed as a reactive intermediate from SOCl_2 and the alcohol with cleavage of HCl . This intermediate reacts with the amino acid, liberating SO_2 and HCl to give the amino acid ester hydrochloride. Benzyl esters can be prepared in a similar manner. Using a different approach, cesium or tertiary ammonium salts of N^α -protected amino acids can be esterified with an alkyl halide. Methyl esters are also easily obtained with diazomethane.

Cleavage of these protecting groups on completion of a peptide synthesis proceeds by mild alkaline hydrolysis in mixtures of water and an organic solvent such as dioxan, methanol, ethanol, acetone, or DMF to solubilize the starting materials. The alkyl esters are usually applied only for the synthesis of short-chain peptides, as hydrolytic cleavage becomes very difficult with increasing chain length, and extreme saponification conditions (excess of alkali and long reaction times) are necessary. Consequently, racemization and other side reactions are observed. Enzyme-catalyzed deprotections often provide ingenious alternatives. Both methyl and ethyl ester resist hydrogenolysis and mild acidolysis. Transformation of these esters into the corresponding hydrazides (the starting materials for azide synthesis) using hydrazine is possible, as is the transformation into amides by aminolysis.

Benzyl esters of free amino acids are obtained by direct esterification with benzyl alcohol in the presence of acidic catalysts (e.g., 4-toluenesulfonic acid, hydrogen chloride, benzene sulfonic acid, polyphosphoric acid). The water formed during the course of the esterification is usually removed by azeotropic distillation (with benzene, toluene, or tetrachloromethane). Benzyl esters of N -protected amino acids are obtained after activation of the carboxylic group with $\text{N,N'$ -dicyclohexyl carbodiimide (DCC) or thionyl chloride and reaction with benzyl alcohol, and also by transesterification reactions. Benzyl ester derivatives can be cleaved by alkaline hydrolysis, catalytic hydrogenation, treatment with saturated HBr /acetic acid (12 h at room temperature or 1–2 h at 50–60 °C), by treatment with sodium in liquid ammonia, or with liquid hydrogen fluoride.

Free amino acids are transformed into the *tert*-butyl esters by addition of isobutene under acidic condition or by transesterification with *tert*-butyl acetate. The hydroxy groups of serine or threonine are transformed under these conditions into the *tert*-butyl ethers. N -protected amino acids such as Z -amino acids can be reacted with isobutene in the presence of catalytic amounts of sulfuric acid or with *tert*-butyl acetate in the presence of perchloric acid to give the corresponding *tert*-butyl ester. The *tert*-butyl ester group, which is easily cleavable by acidolysis (TFA , HCl/AcOH , $\text{BF}_3 \cdot \text{OEt}_2/\text{AcOH}$), is resistant towards hydrogenolysis and quite stable towards alkaline hydrolysis, hydrazinolysis, and ammonolysis. The dicyclopropylmethyl group (Dcpm) can be cleaved with 1% TFA in dichloromethane, even in the presence of side chain-protecting groups of the *tert*-butyl type [86].

The above-mentioned representatives have been used in many practical applications. Esters represent the prototype of anchoring groups in peptide synthesis on either a soluble or insoluble polymeric support. Consequently, this type of carboxy-protecting group is of crucial importance from a methodological point of view.

The carboxy-protecting groups mentioned can be used both as α - and as ω -carboxy-protecting groups in the case of amino dicarboxylic acids. Peptides with a C-terminal amino dicarboxylic acid can be assembled very easily, because in this case both carboxy groups can be blocked for solution synthesis with the same protecting group. Peptides with an internal or N-terminal amino dicarboxylic acid cannot be assembled without taking differential blocking measures [89–91].

In certain specialized cases, for example the head-to-tail cyclization of peptides (cf. Section 6.1), the use of special carboxy-protecting groups with a third dimension in orthogonality is required. This may rely on differences in cleavage kinetics, as are present in the combination of super acid-sensitive esters such as 2,4-dimethoxybenzyl ester (2,4-Dmb), 2-chlorotrityl [Trt(2-Cl)] or the corresponding linkers with Fmoc/tBu chemistry. Different cleavage mechanisms lead to complete orthogonality, as for example in the application of allyl esters or the Dmab group [87] in Fmoc/tBu chemistry. Recently, the {7-(bis(carboxymethyl)amino(coumarin-4-yl))methyl (BCMACM) group as a new photolabile carboxyl protecting group has been described [88]. The BCMACM group is compatible with the widely used Fmoc/tBu chemistry. The water-soluble protecting group can be removed under mild conditions by UV irradiation at 405 nm and has proved to be an efficient carboxyl protecting group for NCL (cf. Section 5.4.2.2) at Glu-Cys sites.

4.2.2.2 Amides and Hydrazides

Although hydrazides may also be used for C $^{\alpha}$ - and side-chain carboxy group protection, further azide couplings are usually obligatory under such circumstances. N'-substituted hydrazides are also of great importance in this context. The preparation is achieved by reaction of the N-protected amino acid with a partially blocked hydrazine derivative using the carbodiimide or mixed anhydride method. After cleavage of the amino-protecting group, the corresponding derivative is used for a segment synthesis and selective deblocking of the hydrazide protecting group is performed on completion of the synthesis. The resulting N-terminally protected, C-terminal hydrazide can then be transformed into the azide and used as an activated carboxy component (Figure 4.10).

Phenylhydrazides of protected amino acids or peptides are cleaved oxidatively by a tyrosinase isolated from mushrooms to provide the free acid [93].

Carboxamides are principally suited as carboxy-protected amino acid or peptide derivatives. Selective chemical methods for cleavage were lacking before highly specific peptide amidases became available. The carboxamide group is more easily hydrolyzed than a peptide bond, although because of the rather small differences selective cleavage is seldom successful.

4.2.3

C-Terminal and Backbone N $^{\alpha}$ -carboxamide Protection

Additional protection of backbone amide groups may be necessary to avoid prominent side reactions such as aspartimide formation (cf. Sections 4.2.4.3 and 4.2.4.9) or interchain aggregation of the resin-bound peptides containing “difficult sequences” (cf. Section 4.5.4.3). This modification appears to prevent or disrupt unwanted

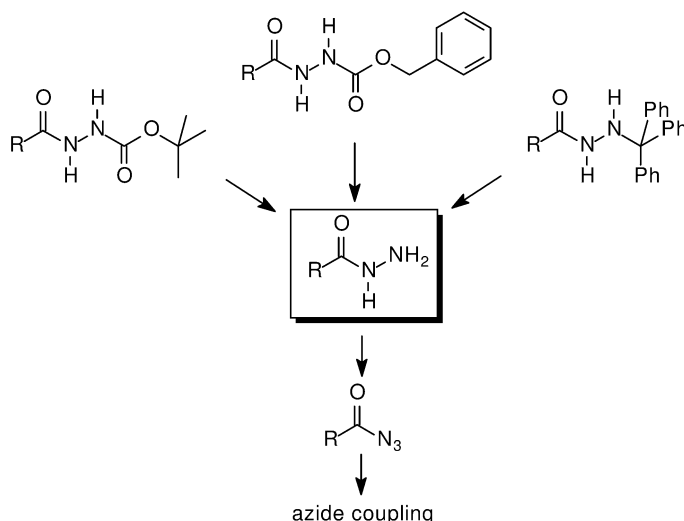
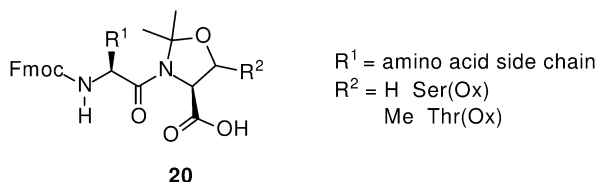


Figure 4.10 Differently protected hydrazide groups as potentially activatable carboxy-protecting groups. R-CO = peptidyl residue.

hydrogen-bonding networks, both by removal of the native amide hydrogen and by alteration of the backbone conformation based on the tertiary amide bond that is formed. Weygand *et al.* [94] used the 2,4-dimethoxybenzyl group for this purpose. Based on the reversible N^α -substitution effect studied by Weygand *et al.* [94], Narita *et al.* [95], and Bartl *et al.* [96], an elegant solution to this problem of difficult sequences was developed by Sheppard *et al.* These authors introduced the 2-hydroxy-4-methoxybenzyl group (Hmb) that has proven very useful for this purpose [97, 98]. It comprises an acid-labile N-benzyl-type protecting group with an additional 2-hydroxy substituent. Therefore, acyl capture of an activated amino acid by this substituent is followed by peptide bond-forming intramolecular O $\rightarrow$ N acyl transfer (Figure 4.11). Improved acyl transfer kinetics is observed in the case of the photolabile 2-hydroxy-6-nitrobenzyl group (Hnb) [99]. The incorporation of either Hmb, Hnb or SiMb (3-methylsulfinyl-4-methoxy-6-hydroxybenzyl)-protected derivatives [100] at every fifth to seventh sequence position during SPPS using preformed *N,O*-bis-(Fmoc)-*N*-(Hmb/Hnb) amino acid derivatives **19** seems to be sufficient in order to suppress aggregation.

Alternatively, pseudoproline dipeptide building blocks Fmoc-Xaa-Ser(Ox)-OH or Fmoc-Xaa-Thr(Ox)-OH **20**, as developed by Mutter *et al.* [101–103], may be employed. This approach has been reported to be superior to Hmb protection, provided that the sequence is compatible with the incorporation of the pseudoproline dipeptides analogues [104].



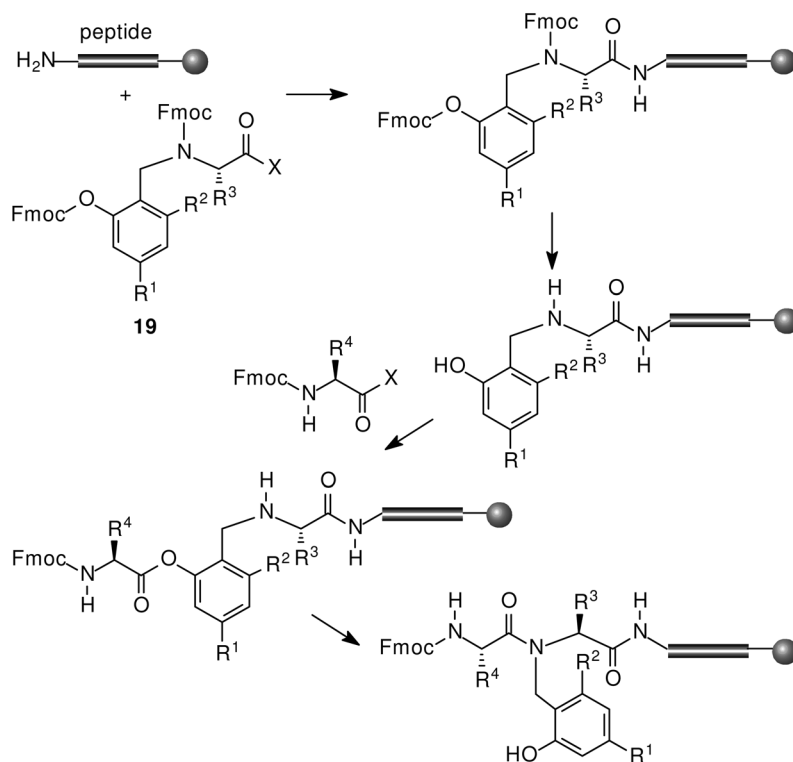


Figure 4.11 Acyl capture and O $\rightarrow$ N acyl transfer involving Hmb ($R^1 = \text{OMe}$; $R^2 = \text{H}$) or Hnb ($R^1 = \text{H}$; $R^2 = \text{NO}_2$). R^3 , R^4 = amino acid side chains; X = activating group.

4.2.4

Side-Chain Protection

As the side chain functionality of an amino acid exerts a crucial influence on the reactivity of the α -amino or ω -carboxy group, it is especially important that those side-chain functionalities which give rise to undesired side reactions be protected in any situation. In particular, α -amino or ω -carboxy groups of diamino carboxylic acids or amino dicarboxylic acids must be mentioned in this context. Selective blocking represents a crucial problem in peptide chemistry, and selection of the protecting group combination is a question that must be addressed by synthesis tactics. The ω -protecting groups of trifunctional amino acids are termed “semipermanent protecting groups” (see Section 4.1.2) because they are usually only cleaved on completion of peptide synthesis. The thiol group of cysteine, and usually also the guanidino group of arginine, also require semipermanent masking. In other cases, side reactions arising from side-chain functional groups can be suppressed or minimized to a certain extent by using special reaction conditions. Despite this possibility of minimal protection, maximum protection is preferred in a practical sense, and this applies especially to SPPS.

4.2.4.1 Guanidino Protection

Although the guanidino group of arginine is usually protected by protonation under normal reaction conditions due to its strongly basic character, the low solubility of the corresponding derivatives in organic solvents hampers peptide synthesis. Partial acylation of the guanidino group may occur under these conditions, and represents a crucial disadvantage.

Derivatives of arginine hydrobromide can be used, in principle, as amino components in peptide synthesis. If required, N^α -benzyloxycarbonyl-protected arginine hydrobromide can be obtained very easily by treatment of Z-Arg-OH with 1.4N HBr in methanol and subsequent precipitation with absolute ether.

Ideally, all three side-chain nitrogen atoms should be protected, but most strategies rely on N^ω - or $N^{\omega'}$, N^δ -protection, respectively. Despite the existence of a variety of blocking possibilities described for the strongly basic guanidino group, no ideal protecting group yet exists. The different protecting groups can be divided into four classes: nitro; urethane (acyl); arenesulfonyl; and trityl types. However, the latter derivatives are poorly soluble in organic solvents and hence are not widely used.

The ω -nitro group is stable toward TFA, HBr/AcOH, and alkali. Treatment with liquid HF and several reductive methods (Zn/AcOH, electrolysis) or catalytic hydrogenation with Pd or Raney-nickel are suitable methods for cleavage. ω -Nitroarginine is prone to other side reactions during acylation and cleavage. Lactam formation (Figure 4.12; **21** $\rightarrow$ **22**) and consecutive aminolysis have been observed upon activation of protected ω -nitro arginine derivatives ($R^2 = \text{NO}_2$).

Protecting groups of the urethane type have been successfully applied to achieve N^ω -mono-acylation. In this case, the synthesis of disubstituted N^α, N^ω -diacyl compounds can be performed most easily.

Z-Arg(Z)-OH can only be coupled with hydrochlorides of the amino component (additional protonation of the guanidino group) to give peptide derivatives because in other cases lactam formation by attack of the N^δ to the activated carboxy group has been observed (Figure 4.12, $R^2 = \text{Z}$).

Differently substituted N^α, N^ω -diacyl derivatives, such as Z-Arg[Z(4- NO_2)]-OH or Z-Arg(Boc)-OH [105], show very little tendency towards lactam formation. The differences in deblocking selectivity – the 4-nitrobenzyloxycarbonyl group is stable

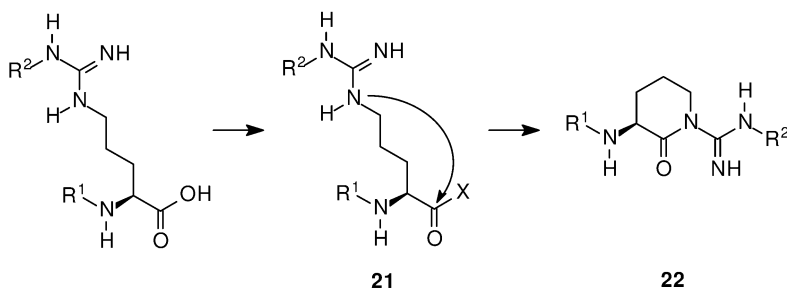


Figure 4.12 δ -Lactam formation in carboxy-activated arginine derivatives.

towards HBr/AcOH – permit a variety of peptide synthetic applications. Derivatives on the basis of Z-Arg(ω , δ -Z₂)-OH, and especially the corresponding 4-nitrophenyl- or N-hydroxysuccinimidyl esters, are excellent building blocks for arginine-containing peptides [89]. The N^ω,N^δ-di(1-adamantyloxycarbonyl) derivative Adoc is obtained from Z-Arg-OH and 1-adamantyl chloroformate [106]. Subsequently, different N^α-protecting groups may be introduced after hydrogenolytic cleavage of the Z group. However, the guanidino group is in this protection scheme and in the case of the N^ω-Boc derivative still sufficiently nucleophilic to be acylated. Consequently, substantial conversion of arginine residues to ornithine residues via acylation of the unprotected ω -nitrogen during coupling and subsequent intramolecular decomposition during deprotection has been observed [107]. N^ω,N^δ-Bis-Boc protection is not in all cases a viable alternative because of the rather poor coupling efficiency.

A single arene sulfonyl moiety offers complete protection of the guanidino group. Consequently, members of this class belong to the most commonly used guanidino protecting groups (Figure 4.13). Albeit undesired lactam formation upon activation of the carboxy group cannot be excluded completely by N^ω-monoprotection using arene sulfonyl groups, it may be suppressed sufficiently by the application of additives such as 1-hydroxybenzotriazole. The 4-toluenesulfonyl group (Tos) **23** can only be removed by acidolysis with liquid HF or by reduction with Na/liquid NH₃. Arene sulfonyl-type protecting groups for the guanidino function have been modified with respect to acid lability by the introduction of electron-donating substituents on the aryl residue. Groups such as 4-methoxybenzenesulfonyl **24** (Mbs) [108], 2,4,6-trimethylbenzenesulfonyl **25** (Mts) [109], 4-methoxy-2,3,6-trimethylbenzenesulfonyl

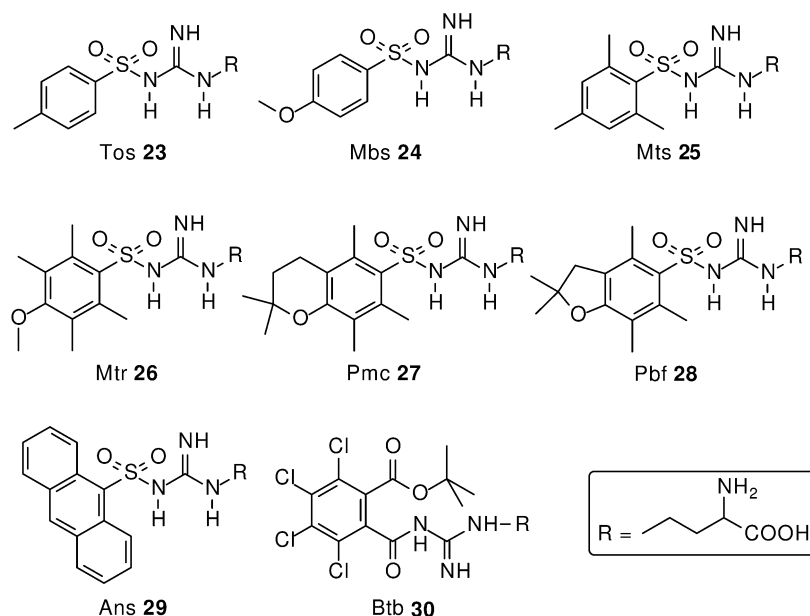


Figure 4.13 Selected protecting groups for the guanidino function of arginine.

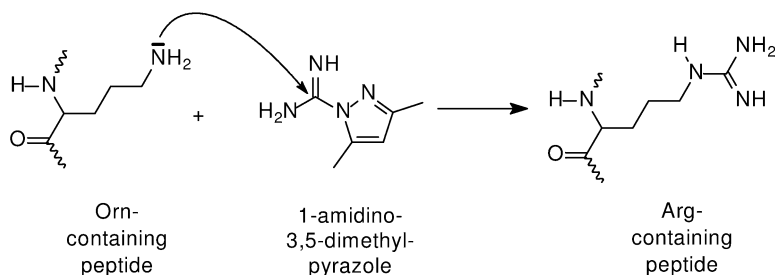


Figure 4.14 Selective guanylation of ornithine-containing peptides.

(Mtr) **26** [110], 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc) **27** [111], and 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) **28** [112] have become very popular in SPPS.

Acid lability increases in the series Tos < Mbs < Mts < Mtr < Pmc < Pbf, with Pbf displaying the best deprotection kinetics.

The Mts group is highly compatible with Boc as the temporary protecting group and can (unlike Tos) be removed with trifluoromethane sulfonic acid (TFMSA)/TFA/thioanisole. Mtr has often been used in the Fmoc tactics, but requires prolonged cleavage time with TFA/thioanisole. The acid lability of the Pmc group is similar to that of the Boc group; therefore, Boc (e.g., for Lys) and Pmc or Pbf (for Arg) may be advantageously combined as side chain-protecting groups. The risk of tryptophan, serine, or threonine sulfonation has been shown to be minimized by the addition of suitable scavengers [113].

The 9-anthracenesulfonyl (Ans) group **29**, which is cleaved by mild reducing agents [114], and the 1-*tert*-butoxycarbonyl-2,3,4,5-tetrachlorobenzoyl (Btb) group **34** [115] have also been successfully applied in both liquid-phase and solid-phase peptide synthesis (Figure 4.14).

Alternatively, the guanidino group can be introduced by selective guanylation of ornithine-containing peptides on completion of peptide synthesis. Besides other guanylation reagents, 1-amidino-3,5-dimethylpyrazole is well suited for this purpose.

4.2.4.2 ω -Amino Protection

Diaminocarboxylic acids such as lysine and ornithine often require orthogonal protection schemes for the α - and ω -amino groups, as the protection of the ϵ -amino group is mandatory. Appropriate protecting groups are presented in Table 4.1. The ϵ -amino group of lysine is more basic and more nucleophilic than the α -amino group, and can be converted into the ϵ -imino derivative with one equivalent of benzaldehyde in the presence of one equivalent of LiOH. Under basic reaction conditions the free α -amino group can then be selectively acylated. Subsequent cleavage of the imine derivative to form, e.g., Z-Lys(H)-OH allows orthogonal protection of the ϵ -amino group, for example with Boc₂O to give Z-Lys(Boc)-OH (Figure 4.15). Alternatively, the α -amino and α -carboxy groups may be simultaneously deactivated by the formation of a chelate complex with copper(II) ions. Subsequently, the ϵ -amino group can then be selectively acylated to give, for example, H-Lys(Boc)-OH.

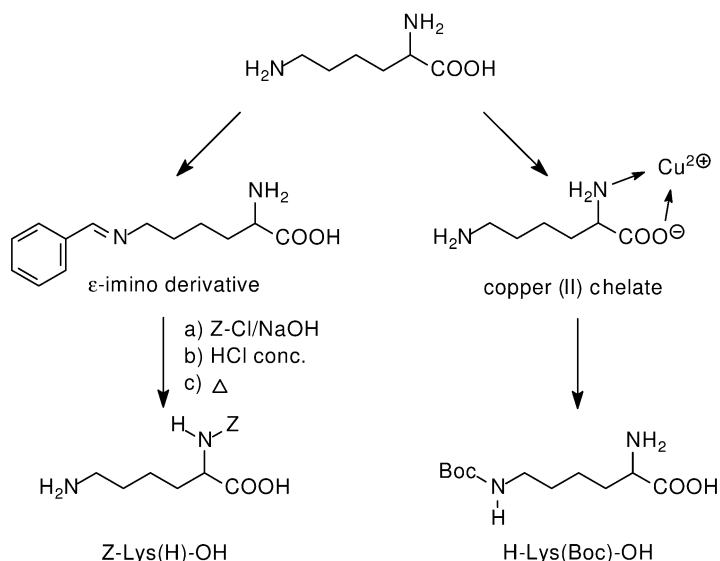


Figure 4.15 Alternative protection of the α - and ϵ -amino function of lysine.

In Boc tactics, the N^ϵ -amino group of lysine usually is protected with the 2-chlorobenzoyloxycarbonyl group [Z(2-Cl)], which displays 60-fold less acid lability compared to the Z group. This prevents premature deprotection during acidolytic Boc cleavage in the course of peptide chain assembly.

In the Fmoc approach, lysine side-chain protection as the Boc derivative can be regarded as the optimum combination.

On occasion, completely orthogonal protection of the N^ϵ -amino group is necessary, and in these special cases the trifluoroacetyl group (Tfa), the 3-nitro-2-pyridinesulfonyl group (Npys), the 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl (Dde) **14a** [71, 72], or the allyloxycarbonyl (Alloc) [40–42] groups are available. The Tfa and Npys groups are very compatible with Boc chemistry, as cleavage occurs in the first case with piperidine and in the latter case with triphenylphosphine or 2-pyridinethiol-1-oxide. Dde **14a** ($R^2 = \text{CH}_3$) and the related ivDde **14b** ($R^2 = \text{iBu}$) [73] groups are orthogonal to Boc and Fmoc, because they are removed with dilute hydrazine solutions. Cleavage results in the formation of a pyrazole derivative, and this can be monitored spectroscopically at 300 nm.

Alloc is an orthogonal protecting group to be applied both in Boc and Fmoc chemistry, as deprotection occurs upon treatment with a suitable nucleophile in the presence of a palladium catalyst.

A new relay deprotection strategy using the PTnm group (Table 4.1) for N^ϵ -amino protection of lysine has been developed by Wang *et al.* [46].

4.2.4.3 ω -Carboxy Protection

Orthogonal protection is required in most cases for the aminodicarboxylic acids aspartic acid and glutamic acid. Formation of isoaspartyl peptides via the correspond-

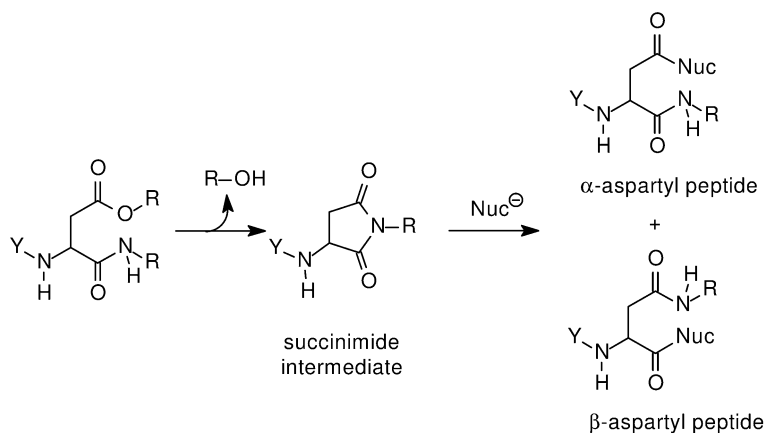


Figure 4.16 Formation of isoaspartyl peptides via a succinimide intermediate.

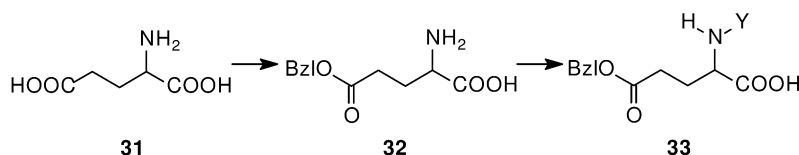
ing succinimide intermediate is the most prominent side reaction of Asp side-chain esters (Figure 4.16).

Here, the nitrogen atom of the amino acid located C-terminally with respect to the Asp residue attacks the Asp side-chain ester moiety. Subsequent ring opening with a nucleophile (e.g., water or amines such as piperidine) yields a mixture of α -aspartyl and β -aspartyl peptides [116], a phenomenon that is also observed upon aging of Asp-containing proteins in biological systems (cf. Section 3.2.2.10).

This problem may, in most cases, be circumvented in Fmoc chemistry by application of the β -*tert*-butyl ester of Asp. However, several publications have reported that, under certain circumstances, aspartimide formation and subsequent side reactions occur in Fmoc protocols, even with *tert*-butyl-protected derivatives. Aspartimide formation appears to be sequence-dependent [117, 118]. The main factors influencing this side reaction are the base employed, the β -carboxy protecting group, and the residue located C-terminally next to the Asp residue. Sequences containing Asp(OtBu)-Gly, Asp(OtBu)-Thr(tBu), Asp(OtBu)-Cys(Acm), Asp(OtBu)-Asn(Trt), and Asp(OtBu)-Asp(OtBu), respectively, seem to be most problematic [119]. The base deprotonates the Asp-Xaa amide group, which then attacks the side-chain ester. Aspartimides are prone to base-catalyzed epimerization and to nucleophilic ring opening (e.g., by piperidine), giving mixtures of α - and β -piperidines, aspartyl- and isoaspartyl peptides together with the corresponding diastereomers formed by epimerization. Addition of HOBt has been shown to minimize these side reactions, but if they cannot be completely suppressed then additional protection (e.g., with Hmb, Hnb) of the amide nitrogen becomes obligatory (cf. Section 4.2.3).

In Boc chemistry the β -cyclohexyl ester Asp(OCy) that is nowadays favored over the benzyl ester efficiently prevents aspartimide formation, even in sensitive sequences. Prolonged treatment with strong acids during protecting group cleavage may, however, be accompanied by cyclization to the aspartimide.

Glutamic acid is usually protected at the side chain with the same groups as used for aspartic acid. Glu derivatives are less prone to undergo this type of cyclization;



Scheme 4.3

however, during HF cleavage in Boc chemistry dehydration to the acylium ion and further side reactions are sometimes observed.

Regioselective acid-catalyzed esterification at the ω -carboxy group of glutamic acid 31 (e.g., with benzyl alcohol) can be achieved using H_2SO_4 as the acid, and yields approximately 75% of the desired side-chain ester derivative 32 of glutamic acid. This monoester can then be converted into the N^{α} -urethane-protected derivative 33 (Scheme 4.3). The protocol may also be applied to derivatives of aspartic acid.

Alternatively, as shown in Figure 4.17, Z-protected aspartic acid 34 is converted into the cyclic anhydride 35 upon treatment with acetic acid. After alcoholysis and separation of the two regioisomers 36 and 37, the α -ester Z-Asp(OH)-OR 36 is converted to the *tert*-butyl derivative Z-Asp(OtBu)-OR 38 and subsequently saponified by alkali to give Z-Asp(OtBu)-OH 39. The Z group may then be exchanged by other amino-protecting groups such as Fmoc [120].

The β -2,4-dimethyl-3-pentyl ester has been suggested as a novel side chain-protecting group for aspartic acid, and is supposed to prevent undesired side reactions – especially during solid-phase synthesis [121]. If orthogonal protection of the side chains of Asp and Glu is required, viable alternatives are available with the base-labile 9-fluorenylmethyl ester (OFm) [80], the allyl ester (OAll) [41], which is

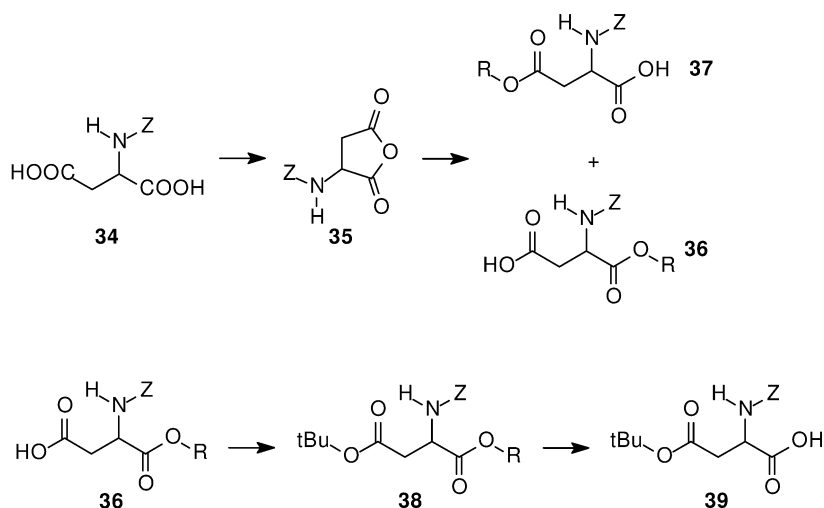


Figure 4.17 Protecting group conversion starting from Z-Asp-OH.

cleaved by palladium-catalyzed transfer to a suitable nucleophile, and Dmab, which is cleaved by hydrazinolysis [87].

4.2.4.4 Thiol Protection [122]

The structural requirements of cysteine-containing peptides further complicate peptide synthesis. In some cases, the Cys residue is needed in its free thiol form, whilst other peptide products require regioselective intra- or intermolecular disulfide bridge formation. Solutions to this problem will be addressed in Section 6.2.

The high nucleophilicity, the ease of oxidation, and the acidic character of the cysteine thiol group require selective, semipermanent blocking of this functional group during all synthetic operations (Table 4.3).

Table 4.3 Thiol protecting groups R–S–R' (RSH = amino acid with a thiol group).

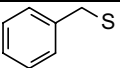
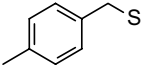
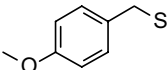
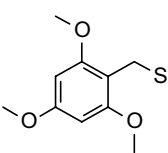
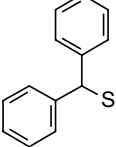
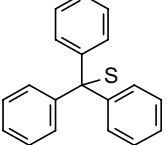
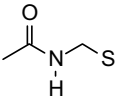
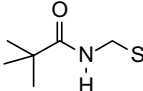
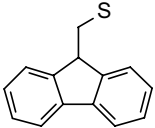
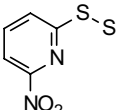
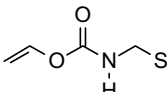
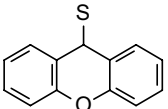
Name	Abbr.	R'–S	Cleavage	Ref.
Benzyl	Bzl		Na/liq. NH ₃	[123]
4-Methylbenzyl	Bzl(4-Me)		strong acids; Ti(TFA) ₃	[124]
4-Methoxybenzyl	Mob		HF; TFA; TFMSA Hg ^{II}	[125]
2,4,6-Trimethoxybenzyl	Tmb		TFA	[126]
Diphenylmethyl	Dpm		TFA; HBr/AcOH; HF; Nps-Cl followed by reduction (B/C) ^a	[127, 128]
Trityl	Trt		HF; TFA; I ₂ ; Hg ^{II} ; TFA/AcOH; AgNO ₃ ; Nps-Cl, then reduction (A/B/C) ^a	[129–132]
<i>tert</i> -Butyl	tBu		Nps-Cl	[133]
Acetamidomethyl	Acm		Hg ^{II} ; I ₂ ; Tl ^{III} ; Nps-Cl followed by reduction (A/B/C) ^a	[134, 135]

Table 4.3 (Continued)

Name	Abbr.	R'-S	Cleavage	Ref.
Trimethylacetamido-methyl	Tacm		Hg ^{II} ; I ₂ ; AgBF ₄	[136, 137]
9-Fluorenylmethyl	Fm		piperidine	[138]
<i>tert</i> -Butylsulfanyl	StBu		thiolysis; phosphines	[139–142]
3-Nitro-2-pyridyl-sulfanyl (2-Nitropyridine sulfenyl)	SNpys		thiolysis “low” HF	[143]
Allyloxycarbonyl aminomethyl	Allocam		Pd ⁰ ; Bu ₃ SnH	[144]
9H-Xanthen-9-yl	Xan		TFA/CH ₂ Cl ₂ /Et ₃ SiH; I ₂	[145]

^aA: iodolysis, B: rhodanolysis, C: Kamber method.

The *S*-benzyl group was introduced in 1930 by du Vigneaud in peptide chemistry for the synthesis of oxytocin, and proved to be a major breakthrough for further synthetic projects involving peptides containing either cysteine or cystine. As this protecting group can be cleaved only by reduction with sodium in liquid ammonia, significant damage to longer peptides may occur because of the rather drastic reaction conditions, as was observed in the first total synthesis of insulin.

S-Benzyl-protected cysteine residues are especially prone to racemization, however, epimerization has been observed for resin attachment of protected (Acm, *t*Bu, Trt, StBu, Tacm, Bzl(4-Me)) Cys residues and even for internal Cys residues during repetitive cleavage of the N^α-protecting group in Fmoc tactics [146–148].

Several substituted benzyl derivatives have been introduced into peptide chemistry in order to facilitate cleavage of *S*-benzyl-type protecting groups. The 4-methylbenzyl residue [Bzl(4-Me)] [124] is cleaved by strong acids and thallium trifluoroacetate, respectively. Higher acid lability is displayed by the 4-methoxybenzyl derivative Mob, which is stable toward TFA and iodine, respectively, but is cleaved by boiling TFA, HF, Na/liq. NH₃, mercury(II) salts, TFMSA, etc. The 4-methylbenzyl and the 4-methoxybenzyl derivatives require treatment with liquid HF for cleavage and,

therefore, are often employed in Boc chemistry. The 2,4,6-trimethylbenzyl (Tmb) group and the 2,4,6-trimethoxybenzyl group (Tmob) are cleaved by acidolysis with TFA and consequently are compatible with Fmoc chemistry [126]. Further potential representatives of this class of compounds have been developed as acid-labile thiol-protecting groups. However, many of them were found to be too resistant toward acids and so were not used in any practical application.

Thiol-protecting groups of the S-acyl type do not have any practical relevance, because acyl migrations from sulfur to nitrogen are energetically favored and represent prominent side reactions.

The triphenylmethyl group (trityl group, Trt) has received much attention, as it may be cleaved by silver salts, mercury(II) salts, iodine, or TFA in the presence of ethanethiol, alkyl sulfenyl chlorides, etc. Deprotection with TFA provides the free thiol, which is in particular useful in the synthesis of peptide antigens for conjugation to carrier proteins. Considerable base-catalyzed racemization has been observed upon *in situ* activation of Fmoc-Cys(Trt)-OH with uronium reagents such as TBTU [149].

The *tert*-butyl group has also been used for thiol protection [133, 150] as it is labile towards both TFMSA and mercury(II) salts.

A different type of thiol-protecting group is found in the N-acyl N,S-acetals, most of which are compatible with both Boc and Fmoc chemistry. The acetamidomethyl (Acm) group is completely stable towards acidolysis, and is removed by treatment with mercury(II) salts at pH 4, thallium(III) trifluoroacetate, or iodine. Oxidizing agents such as iodine simultaneously induce disulfide formation. Structural analogues of the Acm group are the chloroacetamidomethyl group, the isobutylamidomethyl group, and the trimethylacetamidomethyl group, Tacm [136, 137]. The combination of thiol-protecting groups with incongruous cleavage conditions enables differential disulfide formation in peptides containing more than two cysteine residues.

In particular, Trt, Acm, and Tacm represent thiol-protecting groups of eminent importance, as they allow for a simultaneous oxidative coupling to give cystine derivatives. Iodolysis (treatment with iodine, method A, Table 4.3 [151]), rhodanolysis (treatment with rhodanides, method B, Table 4.3 [152]), and the Kamber method, method C, Table 4.3 [153] are suitable for the simultaneous conversion of a thiol-protected peptide to the corresponding cystine (disulfide) derivative without separate deprotection. Oxidative cleavage by iodine, however, may be accompanied by methionine S-oxidation, iodination of Tyr residues [154], and the formation of tryptophan thioesters [155]. Additionally, iodine or thallium(III) oxidation may cause Acm migration from Cys to Ser, Thr, or Gln side chains [156].

Disulfide derivatives also serve as protecting groups for thiols. The ethylsulfanyl ($R' = S\text{-Et}$), the *tert*-butylsulfanyl ($R' = S\text{-tBu}$), and the 2-nitropyridinesulfenyl ($R' = S\text{-Npys}$) groups belong to this class of unsymmetrical disulfides, and can be selectively reduced by thiols such as 2-mercaptoethanol or by phosphines. In the former case of a thiol-disulfide exchange reaction, the excess of a free thiol acting as a deblocking reagent shifts the equilibrium in the desired direction. Removal of the S-alkylthio group occurs selectively, as many other protecting groups (including several thiol-protecting groups) are stable toward thiolysis.

The general protection scheme of thiol groups by an *S*-sulfo moiety (S-SO_3^-) should also be mentioned in this context [157], because the corresponding cysteine *S*-sulfo derivatives obtained from oxidative sulfolysis from *S*-alkylthio or *S*-arylthio cysteine derivatives possess sufficient stability for some peptide synthetic procedures. This temporary blocking is removed by reduction with thioglycolic acid at pH 5.

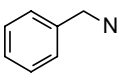
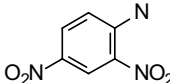
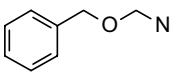
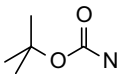
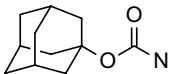
Protected cysteine residues are prone to several side reactions: β -elimination leading to the formation of dehydroalanine derivatives is favored under basic reaction conditions where subsequent conjugate addition of nucleophiles such as piperidine takes place. Oxidation and alkylation may also occur.

More than 70 variants for the protection of the cysteine thiol group have been described, and further details may be found in several monographs [89, 122]. An interesting thiol-protecting scheme for the formation of intrachain disulfide bridges has been demonstrated in an endothelin synthesis [158].

4.2.4.5 Imidazole Protection

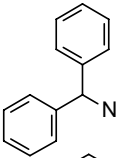
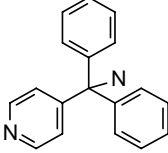
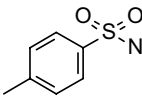
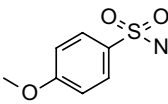
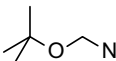
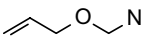
Histidine is one of the most problematic building blocks in peptide synthesis [159, 160]. Side reactions may occur when protection of the imidazole ring of histidine is omitted, because these derivatives are easily *N*-acylated. Some histidine derivatives are also relatively insoluble in the solvents normally used in peptide synthesis. Undesired side reactions such as racemization, formation of cyclic imidazolides, and *N*-guanylations of the imidazole ring have been observed, especially upon activation of C-terminal histidine residues. These problems may be avoided by reversible masking of the imidazole function (Table 4.4).

Table 4.4 Histidine side chain protecting groups $\text{R}'\text{-N}^{\text{im}}\text{-R}$ ($\text{N}^{\text{im}}\text{-R}$ = histidine moiety).

Name	Abbr.	$\text{R}'\text{-N}^{\text{im}}$	Cleavage	Ref.
Benzyl	Bzl			[161]
2,4-Dinitrophenyl	Dnp		2-mercapto-ethanol; thiophenol; stable to acids	[162, 163]
Benzyloxymethyl	Bom		strong acids	[164, 165]
<i>tert</i> -Butoxy-carbonyl	Boc		TFA; HBr/AcOH; HF; stable to HCl/dioxan	[166]
Adamantyl-1-oxycarbonyl	Adoc		TFA	[25, 167]
Triphenylmethyl	Trt		TFA	[168, 169]

(Continued)

Table 4.4 (Continued)

Name	Abbr.	R'-N ^{im}	Cleavage	Ref.
Diphenylmethyl	Dpm		6 N HBr/AcOH (3 h); HCOOH (10 min); TFA (1 h)	[170]
Pyridyldiphenylmethyl	Pdpm		catalytic or electrolytic reduction; Zn/AcOH	[171]
4-Toluenesulfonyl (Tosyl)	Tos		strong acids	[172, 173]
4-Methoxy-benzene-sulfonyl	Mbs		TFA/(CH ₃) ₂ S	[174]
tert-Butoxymethyl	Bum		TFA; stable to piperidine;	[175]
Allyl	All		Pd ⁰ /nucleophile;	[176]
Allyloxymethyl	Allom		Pd ⁰ /nucleophile	[177]

In the past, the imidazole group was protected mainly as the N-benzyl derivative – a method that suffers from a very limited scope of cleavage conditions. This problem was overcome by the development of other protecting groups such as the 2,4-dinitrophenyl group, which is acid stable and can be cleaved by thiolysis (2-mercaptoethanol at pH 8 for 1 h; thiophenol) or by hydrazinolysis.

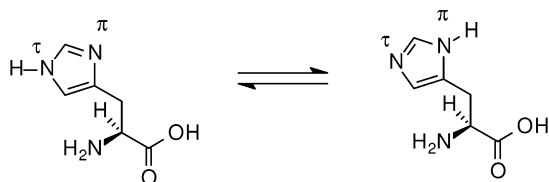
Imidazole-protecting groups of the urethane type (Z, Boc, Adoc) are stable towards HCl in organic solvents, but as N-acyl imidazolides they may give rise to undesired acyl transfer reactions. The pronounced lability towards aminolysis and hydrolysis leads to problems with these permanent protecting groups during other deblocking reactions.

The N^{im}-Boc group is stable toward piperidine and is, therefore, suited for the application in SPPS according to Fmoc tactics. It is cleaved by HBr/HOAc, TFA, or liquid HF, respectively. The structurally related Adoc group is removed by treatment with TFA.

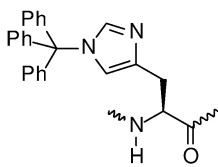
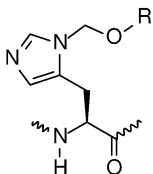
Arylsulfonyl groups have increasingly received attention as imidazole-protecting groups. The N^{im}-(4-toluenesulfonyl) group is often used in combination with Boc as the temporary protecting group. It is cleaved under the general deprotection

conditions with liquid HF or TFMSA, but it has been reported to be susceptible to cleavage by HOBT. Modification of the aryl residue with electron-donating groups increases the acid lability of the arylsulfonyl protecting group, as has already been discussed in connection with the arylsulfonyl-based guanidino protecting groups. The 4-methoxybenzenesulfonyl (Mbs) group and the 2,4,5-trimethylbenzenesulfonyl group, respectively, are labile toward TFA in the presence of dimethylsulfide.

Some confusion exists in the literature concerning protection of the two non-equivalent, but similarly reactive heterocyclic nitrogen atoms, designated π and τ , as shown in **40**. In the past the location of the N^{im} -protecting groups was not specifically considered or was often incorrectly indicated. However, the location of the protecting groups is important in connection with His racemization. Regiospecific protection of the π -nitrogen prevents racemization but is not easy to realize because of the comparable reactivity of the two heterocyclic N atoms.

**40**

While the two regioisomeric derivatives are formed during alkylation in a ratio of 3:1, steric hindrance of the alkylation reagent favors N^{τ} -alkylation. Trityl chloride reacts regioselectively at the N^{τ} position, and N^{τ} -trityl-protected histidine **41** belongs to the most suitable acid-labile protecting groups that are very compatible with Fmoc/tBu tactics. The trityl residue [168, 169] is stable to the conditions of coupling and deprotection during Fmoc synthesis. It can be cleaved with dilute acetic acid at slightly elevated temperatures or with TFA at room temperature, but not with 1N HCl at room temperature. Surprisingly, the lack of N^{π} -protection in the N^{τ} -trityl derivatives does not favor racemization at C^{α} of the corresponding carboxy-activated histidine derivatives, although usually only N^{π} -protection is regarded to be devoid of racemization. Although the N^{τ} -trityl group prevents racemization by its steric bulk to a certain extent [178], epimerization has nevertheless been observed, for example upon condensation of Fmoc-His(Trt)-OH to a 4-benzyloxybenzylalcohol linker [179].

**41****42**

In this respect, the N^π -benzyloxymethyl (Bom) group **42** ($R = \text{benzyl}$) is often used in combination with N^α -Boc protection [180]. This N^π protection of the imidazole significantly decreases the risk of racemization [181]. The Bom group is resistant toward TFA, but it can be easily cleaved with HF, TFMSA, or HBr/HOAc – which renders it especially suited for Boc/Bzl tactics. Formaldehyde is liberated during HF deprotection, and may damage residues such as N-terminally located Cys and Trp [182].

The N^π -*tert*-butoxymethyl (Bum) group **42** ($R = \text{tert-butyl}$) that may be applied in N^α -Fmoc tactics is characterized by higher acid lability, though the preparation of the protected derivative is less straightforward. Application of the N^π -allyl [176] and N^π -allyloxymethyl [177] groups has also been suggested.

4.2.4.6 Hydroxy Protection

In principle, protection of the primary and secondary hydroxy function of serine and threonine (and of the aromatic hydroxy group of tyrosine) is not necessary, especially when these residues are located in the amino component. However, a substantial excess of the acyl component, as is generally used for example in SPPS, may lead to partial O-acylation (especially of tyrosine residues) as the basic reaction conditions generate the highly nucleophilic phenolate anion. The azide method was regarded as the most useful variant for the activation of hydroxy-substituted amino acids and of peptides with C-terminal hydroxy-substituted amino acids. However, aromatic nitrosations in the case of tyrosine or side reactions based on Curtius rearrangements have been observed during acyl azide synthesis. Tyrosine is also prone to electrophilic aromatic substitution during deprotection procedures. An increased risk of racemization has been reported for carboxy components containing C-terminal tyrosine.

The application of active esters originally failed with these derivatives, as N-protected active esters with unprotected hydroxy side-chain functionalities could not be obtained in sufficient purity. However, nowadays the corresponding 2,4-dinitrophenylesters and other active ester derivatives (e.g., Z-Ser-OPcp, Boc-Thr-OSu) can be obtained satisfactorily. On occasion, the syntheses applying components with free hydroxy groups suffer from further undesired side reactions, such as $N \rightarrow O$ aminoacyl migrations, O-acetylations upon deprotection (e.g., with HBr/HOAc), and racemization under the basic conditions of hydrazinolysis. In any case, the advantages of a global protection scheme with the imminent risk of incomplete deprotection reactions should be compared individually to the above-mentioned strategy with partial protection.

The blocking of hydroxy groups with benzyl or *tert*-butyl type groups is of eminent practical importance (Table 4.5). O-Benzyl threonine is obtained from threonine by reaction with benzyl alcohol in the presence of 4-toluenesulfonic acid and subsequent alkaline saponification of the concomitantly formed benzyl ester [183]. O-Benzyl tyrosine is formed by direct alkylation of the tyrosine-copper complex with benzyl bromide [184]. Interestingly, cleavage of the O-benzyl group in Z-Tyr(Bzl)-OMe during hydrogenolysis can be efficiently suppressed by the addition of 2,2'-dipyridyl [185].

tert-Butyl ethers of hydroxy amino acids are very easily accessible by acid-catalyzed *tert*-butylation of amino acid benzyl esters with isobutene and subsequent hydrogenolysis [186].

Table 4.5 Hydroxy-protecting groups R'-O-R (HO-R = amino acid with a hydroxy group).

Name	Abbr.	R'-O	Cleavage	Ref.
Benzyl	Bzl		H ₂ /Pd; HF; Na/liq. NH ₃ ; HBr/dioxan	[187]
2,6-Dichlorobenzyl	Dcb		strong acids; stable to TFA	[188]
Diphenylmethyl (benzhydryl)	Dpm		H ₂ /Pd; TFA (reflux)	[189]
Cyclohexyl	Cy		TFMSA/thioanisole	[190]
2-Bromobenzyloxy carbonyl	Z(2-Br)		strong acids; stable to TFA	[191, 192]
<i>tert</i> -Butyl	tBu		TFA; HCl/TFA; conc. HCl (0 °C, 10 min)	[193]
1-Benzyloxycarbonyl amino-2,2,2-trifluoroethyl	Zte		H ₂ /Pd; acidolysis	[194]
Methylthiomethyl	Mtm		CH ₃ I in wet acetone (NaHCO ₃)	[195]
Allyl	All		Pd ⁰ , nucleophile;	[41]
Allyloxycarbonyl	Alloc		Pd ⁰ , nucleophile	[41, 196]

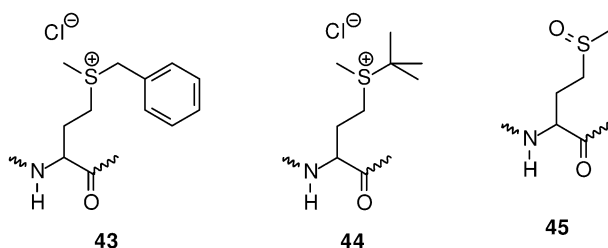
O-Benzyl groups can be cleaved by hydrogenolysis or with liquid HF, and hence are appropriate for Boc chemistry. During acidolysis, highly reactive benzyl cations are formed that often give rise to side reactions, but this can be suppressed by the addition of cation scavengers such as anisole or resorcinol. Furthermore, acidolytic deprotection reactions sometimes may be incomplete. Tyrosine residues to be used for Boc chemistry are normally protected using either the 2,6-dichlorobenzyl (Dcb) or the 2-bromobenzyloxycarbonyl [Z(2-Br)] groups. The 2-bromobenzyloxycarbonyl group [191, 192] is stable towards nucleophiles and TFA, and can be removed together with other side-chain-protecting groups under strongly acidolytic conditions on completion of the

synthesis. The 2,4-dinitrophenyl (Dnp) group can be used for side-chain protection of Tyr and is removable by thiolysis. A base-resistant protecting group for the phenol of Tyr is the 3-pentyl (Pen) group [197] which found application in the combined solid-phase and solution approach for the synthesis of large peptides [198].

The *tert*-butyl ether of Ser, Thr, or Tyr residues is very compatible with Fmoc chemistry. It is removed with TFA (1–2 h at room temperature), with HCl/TFA, and with concentrated HCl at 0 °C under an inert gas atmosphere within 10 min. Fmoc-Ser (tBu)-OH has been reported to undergo racemization under continuous-flow solid-phase synthesis involving preactivation. However, the correct choice of coupling reagent and the base employed (collidine) reduces racemization to an acceptable extent [199].

4.2.4.7 Thioether Protection

The methionine thioether group usually does not cause severe complications during peptide synthesis, and the introduction of amino- or carboxy-protecting groups generally succeeds without difficulties. Side reactions that occur during deprotection reactions with methionine-containing peptides are much more problematic. Partial *S*-demethylation producing homocysteine derivatives has been reported during deblocking reactions with sodium in liquid ammonia, though this method is rarely used nowadays. *S*-Alkylations may occur on acidolytic cleavage of benzyl- or *tert*-butyl-type protecting groups. The reaction of the benzyl or *tert*-butyl cations, respectively, with the thioether group gives rise to the formation of sulfonium salts, and this sometimes cannot be avoided completely, even by the addition of scavengers such as anisole or ethyl methylsulfide. An exchange of *S*-alkyl groups has also been reported in this context. The sulfonium salt **43** may react to give *S*-benzyl homocysteine. The sulfonium salt **44** has been unambiguously identified. *S*-Methylation of methionine to give the sulfonium salt has been recommended as a protecting measure [200].



The oxidation sensitivity of the methionine thioether group represents another problem. Oxidizing agents such as peroxides or oxygen (or even dimethylsulfoxide (DMSO), a solvent that is frequently used in peptide chemistry), may lead to methionine *S*-oxide **45** formation. There are reports that biologically active peptides become devoid of their activity upon oxidation of a methionine residue. The sulfoxide formation may be reduced or even completely avoided when oxidants are strictly excluded or special cleavage cocktails are used [201].

On the other hand, methionine sulfoxide formation has been suggested as a protecting measure for the thioether moiety. *S*-Oxidation leads to the creation of a

new stereogenic center, and diastereomeric peptides containing methionine S-oxides have been separated by crystallization of the picrates. The sulfoxide group can be reduced, for example with thioglycolic acid, *N*-methylmercaptoacetamide, or ammonium fluoride/2-mercaptoethanol (the latter method does not affect disulfide bonds).

Hydrogenolytic cleavage of protecting groups is possible for methionine-containing peptides despite the presence of the methionine sulfur, when a base or $\text{BF}_3 \cdot \text{OEt}_2$ is added or by hydrogen transfer reaction using formic acid as the hydrogen donor.

Methionine is normally used both in Boc and in Fmoc chemistry without further side-chain protection. Oxidation is avoided in Fmoc chemistry by adding ethyl methylsulfide or thioanisole during cleavage and deprotection.

4.2.4.8 Indole Protection

The indole ring of tryptophan usually does not require any protecting measures. A broad variety of synthetic conditions permit the incorporation of N^{in} -unprotected tryptophan, both as the amino or as the carboxy component. However, when the azide method is used for peptide coupling, N^{in} -nitrosation may occur during the synthesis of the acyl azide from the hydrazide, especially in the case of incorrect stoichiometry. The indole system is prone to oxidation by either oxygen or peroxides and, therefore, an inert gas atmosphere and absolutely peroxide-free solvents are required. Side reactions at the indole ring (e.g., electrophilic aromatic substitutions) have been observed during some deprotection reactions: *tert*-butylation may occur both at N^{in} and at different positions of the indole ring upon acidolysis of *tert*-butyl-type protecting groups. This also applies to other reactions involving carbenium ions [202]. In most cases, these side reactions are suppressed by the addition of scavengers.

There are few possibilities available for indole ring protection. N^{in} -Formylation increases the stability against alkylation and oxidation [203, 204] and can be removed with dilute aqueous piperidine solution, TFMSA, hydrazine, or hydroxylamine. Consequently, the formyl derivatives are applicable in Boc chemistry. Under basic reaction conditions the formyl group may migrate to free α -amino groups.

In principle, acylation of the indole nitrogen with the Z-group is also a viable alternative. This derivative is stable towards some acids, but is cleaved slowly by TFA and more rapidly by HF. Deprotection by hydrogenolysis and hydrolysis, respectively, is possible.

Although tryptophan may be used without side-chain protection in Fmoc protocols, problems may arise during final deprotection. While indole *tert*-butylation is usually suppressed by the addition of scavengers, sulfonation by the cleavage products of arenesulfonyl-type guanidino-protecting groups may be deleterious. In these cases, Fmoc-Trp(Boc)-OH provides a viable alternative for preventing Trp sulfonation, even in the case of peptides containing multiple arginine residues [205]. Indole Boc protection also diminishes the propensity of trialkylsilane scavengers to reduce the indole ring to an indoline ring. The N^{in} -cyclohexyloxycarbonyl (Hoc) group [206] is base-resistant. Further possibilities for indole protection are available [207–209].

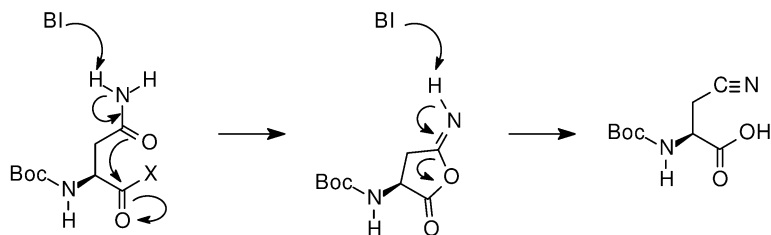


Figure 4.18 Mechanism for the dehydration of carboxy-activated asparagine. Glutamine dehydration follows an analogous mechanism. B = base; X = electron-withdrawing activating group.

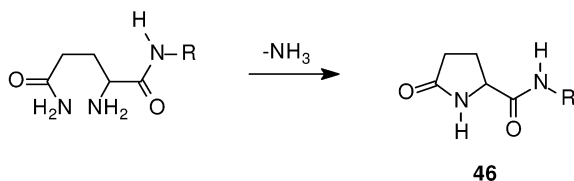
4.2.4.9 ω -Amide Protection

From a chemical point of view, the ω -carboxamide groups of glutamine and asparagine, as well as the C-terminal amides of amino acids and peptides, are rather unreactive functionalities that do not require further protection. While this is usually true for the C-terminal carboxamides, side reactions sometimes occur at the side-chain carboxamides of Asn and Gln during peptide synthesis. Especially under the conditions of carbodiimide couplings, undesired nitrile formation from the carboxamide is very common (Figure 4.18). Activating reagents such as BOP, PyBOP, and HBTU are also reported to favor this dehydration, which can be suppressed by the addition of HOBT or HOAt to the coupling reaction. As this side reaction occurs only during activation of Asn or Gln residues because of mechanistic considerations, it should not cause any problem once these residues have been incorporated into the peptide chain.

Hydrazinolytic conditions may convert the primary amides into the corresponding hydrazides, and the conversion of carboxamides into esters during solvolytic deprotection reactions has also been observed.

Cyclization reactions of asparagine in a peptide chain involving the β -carboxamide group with release of ammonia have also been reported. The tendency towards deamidation depends both on the primary and secondary structure, as well as on the solvent [210–212]. The resulting succinimide derivative (cf. Figure 4.16) may subsequently undergo hydrolysis to give a mixture of α -aspartyl and β -aspartyl peptides.

Peptides containing glutamine in the N-terminal position are easily converted into pyroglutamyl peptide derivatives **46** (Scheme 4.4). This side reaction is catalyzed by acids and is, therefore, especially problematic in Boc tactics. It may be avoided by the application of high TFA concentrations for Boc cleavage. Base-catalyzed pyroglutamate formation in Fmoc chemistry is less common.



Scheme 4.4

Most of the side reactions described above can be suppressed by reversible blocking of the side-chain carboxamide group, which also confers better solubility in organic solvents on the peptides – which is indispensable for normal peptide synthetic operations. Unprotected amide groups tend to form intra- or intermolecular hydrogen bonds, and this results in lower solubility in organic solvents and higher solubility in aqueous solvents. Substituted N-benzyl derivatives that can be cleaved under acidolytic conditions are well suited for this purpose. N-Methoxybenzyl residues or the N-diphenylmethyl protecting group can be introduced directly by the reaction of N^α-protected asparagine or glutamine, respectively, with the benzyl or diphenylmethyl alcohol (Figure 4.19A). Alternately, the corresponding N^α/C^α-protected aminodicarboxylic acids (aspartic acid or glutamic acid) are transformed into the side chain-protected carboxamides with benzyl amine or diphenylamine derivatives and DCC as coupling reagent (Figure 4.19B).

All amide-protecting groups are cleaved with liquid HF. Cleavage with TFA at ambient temperature is also possible for some derivatives, except for the 4-methoxybenzyl and the diphenylmethyl protecting groups. The addition of anisole as a scavenger favors clean cleavage reactions, as the intermediate benzyl-type cations are trapped.

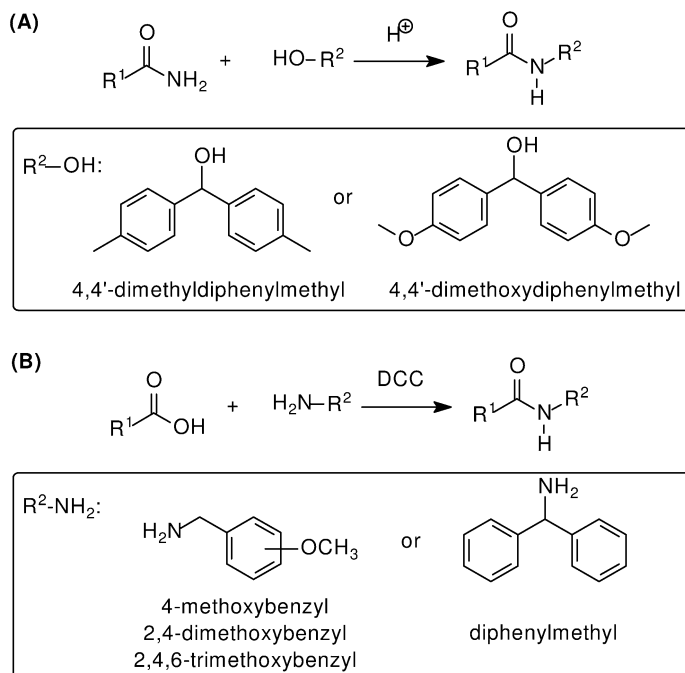
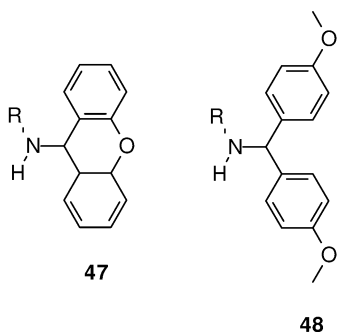


Figure 4.19 Introduction of ω-amide-protecting groups into amino acid derivatives via N-alkylation of a carboxamide (A) or amide formation (B).

The N-trimethoxybenzyl (Tmb) and the N-triphenylmethyl (Trt) groups are stable towards piperidine and can be removed by treatment with TFA [213], though Trp alkylation may be detrimental with Tmb [214]. The trityl cation usually does not alkylate Trp or other sensitive side chains. Hence, Trt and Tmb are very appropriate carboxamide-protecting groups for solid-phase synthesis in combination with Fmoc as the temporary protecting group.

The base-stable 9-xanthenyl (Xan) [215] **47** or the 4,4'-dimethoxydiphenylmethyl (Mbh) [216] **48** groups have found some application for amide protection in Boc/Bzl chemistry, as they can be removed with strong acids. However, they display somewhat limited acid stability, and this results in progressive cleavage during the consecutive operations of peptide synthesis according to the Boc protocol. On the other hand, they provide protection against nitrile formation when Asn or Gln are coupled.



4.2.5

Enzyme-Labile Protecting Groups

The synthesis of acid- and base-labile peptide derivatives and peptide conjugates requires the application of protecting groups that can be cleaved under very mild or even neutral reaction conditions. In particular, glyco-, phospho-, and lipopeptides, as well as sulfated peptides, (cf. Sections 6.3–6.6) belong to this class of sensitive compounds. The treatment of such derivatives with bases often leads to β -elimination reactions in the case of serine and threonine derivatives, or to other undesired side reactions. Treatment with acids in several cases causes degradation of the nonpeptide moiety of the conjugate. Therefore, these targets may require the application of enzyme-labile protecting groups.

The first peptide synthesis using the Bz-Phe residue as an enzyme-labile protecting group to be cleaved with chymotrypsin was already described in 1955 [217]. Subsequently, the development of enzyme-labile blocking groups has experienced a logical progression from the early specialized application to the most sophisticated synthetic strategies [218–220]. Enzymatic manipulation of protecting groups comprises more than simple deprotection, as regiospecificity and stereospecificity are favorable features of enzyme reactions. In addition, because of the mild reaction conditions, enzymes can compete successfully with chemical

reagents in several cases of protecting group manipulation. The specificity of the biocatalyst ensures that neither the peptide bonds nor other protecting groups present are attacked. Consequently, enzymatic protecting group techniques represent a particularly efficient methodology for this type of sensitive compound [221–225].

4.2.5.1 Enzyme-labile N^α-amino Protection

The first enzyme-labile urethane-type protecting groups were developed by Waldmann *et al.* [223]. The overall strategy utilized in many of these derivatives employs an enzyme-labile bond, where cleavage may additionally result in a fragmentation reaction of the urethane, thus liberating the free amino group (Figure 4.20).

The tetrabenzylglucosyloxycarbonyl group (BGloc) [226] comprises O-benzyl-protected glucose as the carbohydrate moiety which is attached to the urethane oxygen. After removal of the benzyl groups, a mixture of α -glucosidase from baker's yeast and β -glucosidase from almonds results in cleavage of the glycosidic bond, thus liberating the carbamic acid (Figure 4.21). Based on this concept, a whole set of enzyme-labile carbohydrate-based protecting groups can be generated that may be selectively removed using a glycosidase. Furthermore, tetra-O-acetyl- β -D-glucopyranosyl-oxycarbonyl (AGloc) and tetra-O-acetyl- β -D-galactopyranosyl-oxycarbonyl (AGaloc) have been developed as urethane-type protecting groups. These are regarded as second-generation enzyme-labile urethane-type protecting groups which are readily attached to the desired amino acid and, after further synthetic steps, may be cleaved in two separate or sequential enzymatic steps [227].

The acetoxybenzyloxycarbonyl group [Z(OAc)] [228, 229] and the chemically related phenylacetyloxybenzyloxycarbonyl group [Z(OAcPh)] [230] are cleaved by an esterase or lipase [Z(OAc), Figure 4.22] or penicillin G acylase [Z(OAcPh)]. These enzymes recognize the enzyme-labile ester bond and selectively cleave it. Consequently, a 4-hydroxybenzylurethane is formed that spontaneously undergoes fragmentation with release of the desired amino group. Depending on the acyl residue selected (acetate or phenylacetate), enzymes with different selectivity may be used.

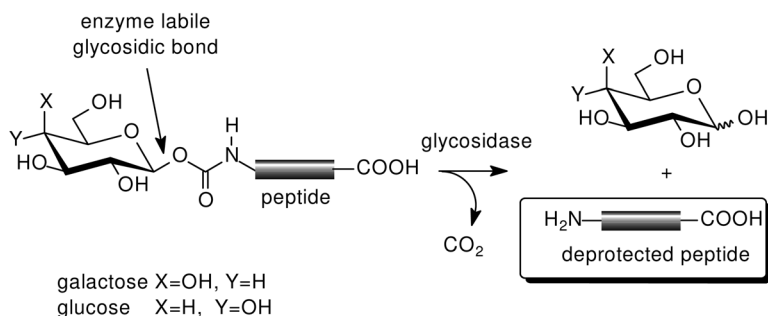


Figure 4.20 Enzyme-labile carbohydrate-derived protecting groups.

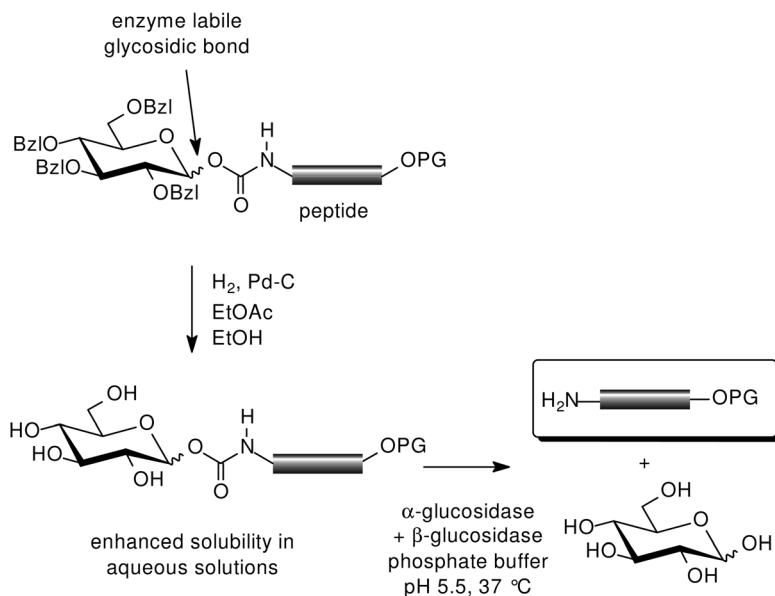


Figure 4.21 Enzymatic cleavage of the BGloc protecting group (PG: protecting group).

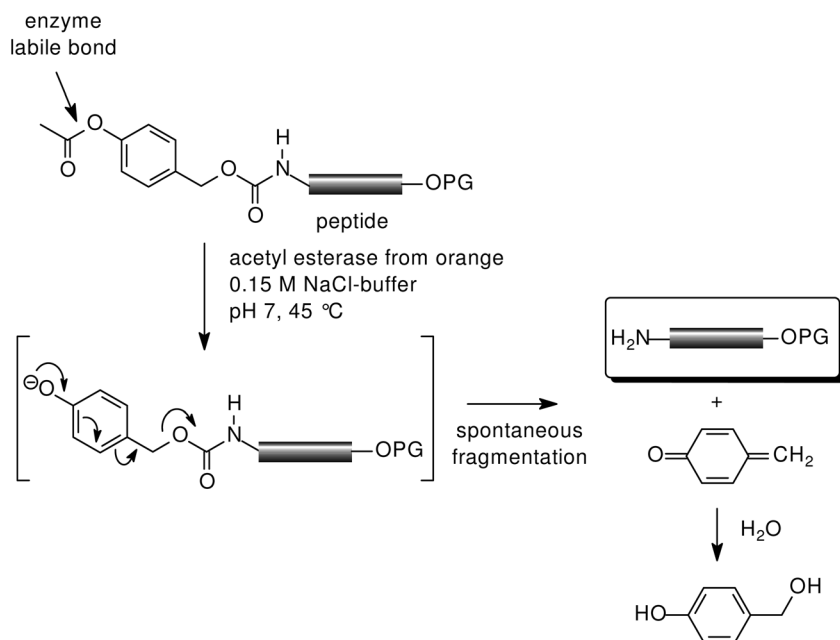


Figure 4.22 Mechanism of enzymatic Z(OAc) protecting group cleavage (PG: protecting group).

4.2.5.2 Enzyme-labile C $^{\alpha}$ -carboxy Protection and Enzyme-labile Linker Moieties

Esters form one class of carboxy-protecting groups that are labile towards cleavage by enzymes, and can be removed by lipases. The lipase from *Mucor javanicus* selectively deblocks heptyl esters of the C-terminus in sensitive amino acid derivatives, without any undesired side reactions [231, 232]. However, some problems are associated with the use of the very hydrophobic heptyl ester, these being mainly due to the low solubility of the substrate and, hence, the low turnover rate, especially in the presence of further hydrophobic amino acids. The more hydrophilic 2-(*N*-morpholino)-ethyl ester (MoEt) and the methoxyethoxyethyl ester (Mee) circumvent this problem. The Mee ester may be cleaved by lipase-catalyzed hydrolysis or even papain-catalyzed hydrolysis [233, 234].

Waldmann's group was also able to transfer the principle of deprotection by enzyme-induced fragmentation of enzyme-labile urethane-type protecting groups to an enzyme-labile ester group [235]. Cleavage of the 4-phenylacetoxybenzyl (PAOB) ester is mechanistically related to Z(OAcPh) cleavage (Figure 4.22), and is initiated by the action of penicillin G acylase at pH 7 and room temperature, with liberation of phenylacetate. The phenolate formed concomitantly undergoes fragmentation to give the desired C-terminally deblocked dipeptide and the quinone methide, which is efficiently trapped by water, forming 4-hydroxybenzyl alcohol.

Another strategy uses amino acid choline esters (Figure 4.23) that may be cleaved by the enzyme butyryl choline esterase. This commercially available biocatalyst recognizes positively charged choline esters and cleaves them under very mild conditions [236, 237]. Ester groups of sensitive peptide conjugates, such as phosphorylated and glycosylated peptides [238], lipidated peptides [239, 240] and nucleopeptides [241] can be hydrolyzed under very mild conditions with choline esterases, such as acetylcholine esterase (AChE) and butyrylcholine esterase (BChE).

4.2.6

Protecting Group Compatibility

Protecting group orthogonality [1] is required for the different types of protecting groups that are used for a particular synthesis. Orthogonality means that each class of

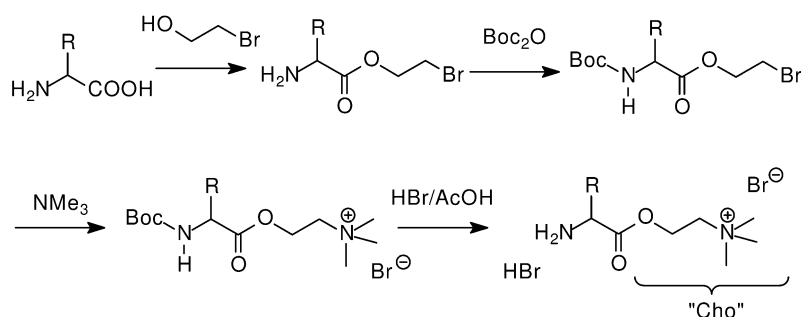


Figure 4.23 Synthesis of amino acid choline esters.

protecting group (temporary, semipermanent) is cleaved independently from the other. Therefore, the chemical cleavage mechanisms should be different for an optimum degree of orthogonality. However, in most cases strict orthogonality cannot be achieved easily and the protecting schemes rely on differences in cleavage kinetics.

The protecting group schemes most frequently used in solid-phase synthesis, and which often are also suitable for solution-phase syntheses, are detailed briefly in Section 4.5.3.

For example, the application of Boc as the temporary protecting group excludes the use of semipermanent protecting groups at the side chains and at the C-terminus with comparable acid-lability. Consequently, semipermanent protection in Boc tactics must employ functional groups that resist the conditions necessary for the iterative Boc cleavages, such as benzyl-type groups. These tactics are, therefore, called the Boc/Bzl tactics or Boc/Bzl protection scheme.

Fmoc as the temporary protecting group is cleaved after each coupling step by treatment with bases. In such cases, side-chain and C-terminal protection may employ *tert*-butyl-type groups. Hence, this is referred to as the Fmoc/*t*Bu tactics or Fmoc/*t*Bu protection scheme.

Special protection schemes may also be required, for example in the synthesis of fully side chain-protected peptides, or for subsequent cyclization reactions.

4.3

Peptide Bond Formation

Peptide bond formation is a nucleophilic substitution reaction of an amino group (nucleophile) at a carboxy group involving a tetrahedral intermediate. However, carboxylic acids react at room temperature with ammonia or amines to give an ammonium salt instead of a carboxamide. Therefore, the carboxy component ($X-R^2=OH$) must be activated prior to peptide bond formation (Figure 4.24).

Furthermore, the peptide coupling reaction must be performed under mild conditions, and preferably at room temperature. Activation of the carboxy component (an increase in the electrophilicity) is achieved by the introduction of electron-accepting moieties. Groups which exert either an inductive (-I) effect or mesomeric (-M) effect (or both) decrease the electron density at the $C=O$ group, thereby favoring the nucleophilic attack of the amino component. The latter attacks, with its nitrogen lonepair, the electrophilic position of the carboxy group to give the tetrahedral zwitterionic intermediate. Peptide bond formation is then completed by dissociation

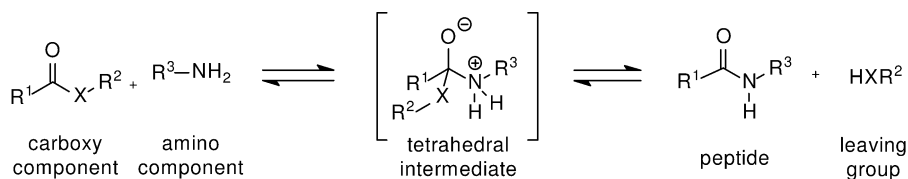


Figure 4.24 Peptide bond formation via carboxyl activation ($X=O$ or S).

of the leaving group (nucleofuge R^2X^-). The leaving group capacity (nucleofugicity) is another factor which influences the reaction rate.

The variation of XR^2 provides a broad spectrum of methods for peptide bond formation, and this methodological arsenal has been summarized in several monographs [89, 91, 242–248].

It is beyond the scope of this book to discuss all the available methods in detail. Rather, attention will be focused on the reagents that are generally accepted in practical applications. Among these are, e.g., the acyl azides, anhydrides, carbodiimides, active esters, and acyl halides, together with new developments in peptide coupling approaches [249] such as HOBt and HOAt-based phosphonium, guanidinium/uronium and immonium type coupling reagents.

4.3.1

Acyl Azides

The acyl azide method [250] was introduced into peptide chemistry as early as 1902 by Theodor Curtius (251), and hence is one of the oldest coupling methods. Besides free amino acids and peptides, which are coupled with an acyl azide in alkaline aqueous solution, amino acid esters can be used in organic solvents. In contrast to many other coupling methods, the addition of an auxiliary base or a second equivalent of the amino component in order to trap the hydrazoic acid is not necessary.

For a long time the azide method was considered to be the only racemization-free coupling method, and it experienced a tremendous renaissance following the introduction of selectively cleavable amino-protecting groups. The starting materials for this method are the crystalline amino acid or peptide hydrazides, respectively, that are easily accessible from the corresponding esters by hydrazinolysis.

These hydrazides are transformed into the azides by *N*-nitrosation with one equivalent of sodium nitrite in hydrochloric acid at -10°C and simultaneous water elimination. Alternately, mixtures of acetic acid, tetrahydrofuran, or DMF with hydrochloric acid have been used. The acyl azide is extracted from the aqueous layer, for example, with ethyl acetate, washed, dried, and reacted with the correspond-

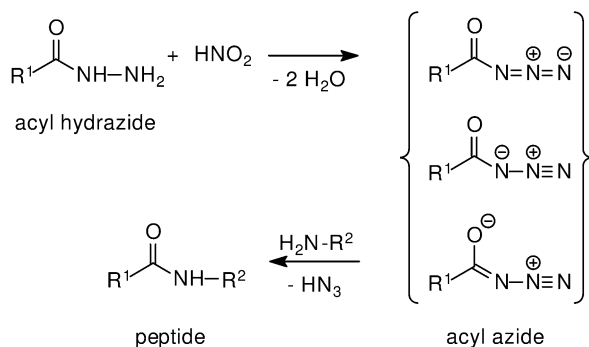


Figure 4.25 Classical multistep peptide synthesis via acyl azide preparation.

ing amino component. Some acyl azides can be precipitated upon dilution with iced water.

Diphenylphosphoryl azide (DPPA) may also be used for the synthesis of acyl azides [252]. The Honzl–Rudinger method uses *tert*-butyl nitrite for the N-nitrosation and allows the application of organic solvents in azide couplings [253]. The coupling reaction should be performed at low temperature because of the inherent thermal instability of acyl azides. At higher temperature, the Curtius rearrangement, i.e., conversion of the acyl azide to an isocyanate, becomes a prominent side reaction and eventually leads to the undesired formation of ureas. As a consequence of the low reaction temperature (e.g., 4 °C), the reaction rate is rather low and the peptide coupling reactions often take several days to reach completion.

Hydrazinolysis of ester groups in longer N-terminally protected peptides is often difficult and, consequently, the application of orthogonally N'-protected hydrazine derivatives is the method of choice. The peptide segments assembled according to this backing-off strategy can be used for azide couplings after selective deblocking of the hydrazide group.

Although, as mentioned previously, the azide method has been considered as the method with the lowest tendency towards racemization, an excess of base in the reaction is accompanied by considerable racemization. Hence, any contact with bases during the coupling reaction must be avoided; for example, ammonium salts of the amino component should be neutralized with *N,N*-diisopropylamine or *N*-alkylmorpholine instead of triethylamine.

Despite all the limitations mentioned, the azide method is still important, especially for segment condensations, because of its low tendency towards racemization, the applicability with O-unprotected serine or threonine components, and the highly variable use of N'-protected hydrazides.

4.3.2

Anhydrides

Basic considerations on the application of anhydrides in peptide synthesis date back to the early investigations of Theodor Curtius in 1881 in the synthesis of hippuric acid. Curtius obtained Bz-Gly_{*n*}-OH (*n* = 2–6) together with hippuric acid upon reaction of silver glycinate with benzoyl chloride. In those early days it was argued that N-benzoyl amino acids or N-benzoyl peptides formed unsymmetric anhydrides with benzoic acid as reactive intermediates when treated with benzoyl chloride.

About 70 years later, Theodor Wieland took advantage of these findings when his group made the mixed anhydride method available for modern peptide synthesis. Nowadays, besides this method, symmetrical anhydrides and N-carboxy anhydrides (NCA, Leuchs anhydrides) formed intramolecularly from the carboxy groups of the amino acid and a carbamic acid are used in peptide couplings. Last, but not least, it should be mentioned that unsymmetrical anhydrides are often involved in biochemical acylation reactions.

4.3.2.1 Mixed Anhydrides

Carboxylic acids and inorganic acids, respectively, are used for the formation of mixed anhydrides [254]. However, only a few representatives of the variety of synthetic variants and acids used have found widespread practical application and, in most cases, alkyl chloroformates (or correctly, alkyl chlorocarbonates) are used. Ethyl chloroformate (ethyl chlorocarbonate), which in the past was used very frequently, has nowadays mainly been replaced by isobutyl chloroformate (isobutyl chlorocarbonate). The mechanism of the mixed anhydride method is shown in Figure 4.26.

The regiochemistry of the aminolysis reaction of the mixed anhydride, which is initially formed from the carboxy component and the chloroformate, depends on the electrophilicity and/or the steric situation of the two competing carboxy groups. In the case of mixed anhydrides formed from an N-protected amino acid carboxylate (carboxy component) and an alkyl chlorocarbonate (activating component, e.g., from alkyl chloroformate), the amine nucleophile predominantly attacks the carboxy group of the amino acid component with formation of the desired peptide derivative and release of the activating component as free acid (Figure 4.26, reaction 2). On application of an alkyl chloroformate (R^1 = isobutyl, ethyl, etc.), the free monoalkyl carbonic acid formed is unstable and immediately decomposes into carbon dioxide

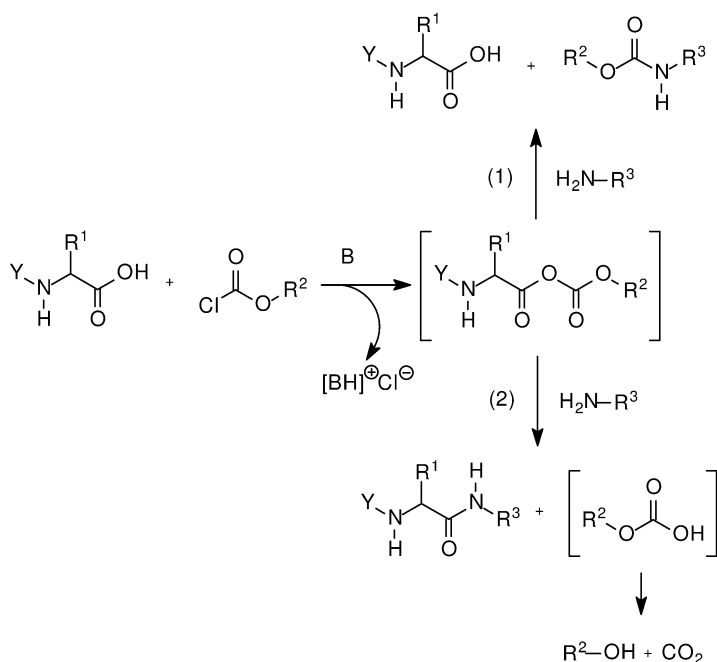


Figure 4.26 Mechanism of mixed anhydride formation and regiochemistry of aminolysis.

Y = amino-protecting group; R^1 = side chain of the carboxy component; R^2 = substituent of the chlorocarbonate; $\text{H}_2\text{N-R}^3$ = amino component; B = tertiary base.

and the corresponding alcohol. However, there are some reports of the opposite regiochemistry (Figure 4.26, reaction 1) of the nucleophilic attack, with formation of a urethane and regeneration of the N-protected amino acid component.

In order to form a mixed anhydride, N-protected amino acids or peptides are dissolved in dichloromethane, tetrahydrofuran, dioxan, acetonitrile, ethyl acetate, or DMF, respectively, and treated with one equivalent of a tertiary base (*N*-methylpiperidine, *N*-methylmorpholine, *N*-ethylmorpholine, etc.). The required alkyl chloroformate is then added at temperatures between -15 and -5 °C under vigorous stirring to form the unsymmetrical anhydride (activation). After a short activation period, the nucleophilic amino acid component is added. If it is used as an ammonium salt (which requires more base), then overstoichiometric application of base must be avoided. The mixed anhydride method can be performed very easily and is one of the most powerful coupling methods, provided that the reaction protocol described above is followed strictly.

Profound studies on the stability of mixed anhydrides, on the minimization of the undesired urethane formation and racemization eventually led to an improved mechanistic understanding and to an increase in efficiency of this coupling method, which now finds widespread application. An investigation into the reasons for excess urethane formation (Figure 4.26, reaction 1), especially after the activation of isoleucyl or valinyl residues, led to the conclusion that this prominent side reaction can be prevented by using dichloromethane as the solvent and *N*-methylpiperidine as the tertiary base. The relatively high stability of mixed anhydrides towards hydrolysis allows their purification by washing the organic phase containing the mixed anhydride with water [255–257]. The stability of mixed anhydrides obtained with alkyl chloroformates also depends on the alkyl group present. Mixed anhydrides from Boc-, Z-, and Fmoc-protected amino acids and isopropyl chloroformate can be purified and isolated. These are much more stable than the derivatives obtained with ethyl or isobutyl chloroformate [258, 259]. In the absence of a suitable nucleophile, the decomposition of a mixed anhydride in organic solvents starts with cyclization to produce 2-alkoxy-5(4*H*)-oxazolones, with liberation of carbon dioxide and the alcohol R^2 -OH [260, 261]. Symmetrical anhydrides and esters are formed as by-products.

Several interesting aspects with respect to the practical realization of a mixed anhydride coupling should be noted: Although aqueous DMF is a good solvent both for the formation of the mixed anhydride and for the subsequent coupling reaction, it promotes racemization to a much greater extent than either tetrahydrofuran or halogenated solvents, as has been shown in the reaction of Z-Gly-Xaa-OH (Xaa = Ala, Leu, Val, Phe) with H-Val-OEt [262]. Isopropyl chloroformates are superior to ethyl or isobutyl chloroformates. Interestingly, activation with ethyl chloroformate in DMF or *N*-methylpyrrolidone leads to less racemization than activation with isobutyl chloroformate. Nevertheless, mixed anhydride formation with ethyl chloroformate, and also the application of triethylamine as a tertiary base, are currently of very little practical importance. Side reactions during mixed anhydride couplings have been observed initially upon activation of N^α -tosyl-, N^α -trityl-, and N^α -trifluoroacetyl-protected amino acids, respectively.

Mixed anhydride formation with pivalic acid [263] as the activation reagent is sometimes recommended, especially for N^α -acyl-protected asparagine. These unsymmetrical anhydrides are prepared analogously from N^α -acyl amino acids and pivaloyl chloride, and react in high yields with amino nucleophiles. The strong + I effect of the *tert*-butyl group in pivalic acid reduces the electrophilicity at this carbonyl group and leads, together with the steric hindrance, to the desired regioselective attack of the nucleophile on the activated amino acid.

For mechanistic reasons, the mixed anhydride method can be extended to other activated species formed by reaction of the carboxylic acid with, e.g., propylphosphonic acid anhydride (PPA) [264], phosphoric acid-derived species, phosphoric acid derived species (cf. Section 4.3.6), carbodiimides (cf. Section 4.3.3), boronic acid derivatives [265] or ethoxyacetylene [266].

4.3.2.2 Symmetrical Anhydrides

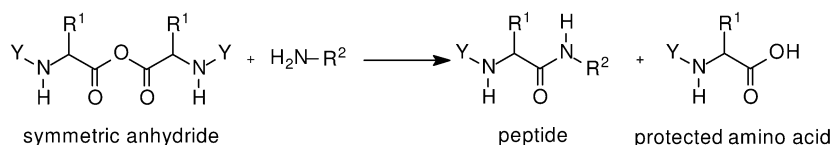
Symmetrical anhydrides of N^α -acyl amino acids (Scheme 4.5) are highly reactive intermediates for peptide bond formation. Ambiguous regioselectivity upon reaction with amine nucleophiles is not possible, in contrast to the mixed anhydride method. However, a maximum peptide yield of 50% (with reference to the carboxy component) can be obtained.

Although the free N^α -acyl amino acid formed on aminolysis of a symmetrical anhydride together with the desired peptide can be recovered by extraction with saturated sodium bicarbonate solution, the practical importance of this method initially was very small. Symmetrical anhydrides can be obtained from N^α -protected amino acids with phosgene or, more conveniently, with carbodiimides. The reaction of two equivalents of N^α -acylamino acid with one equivalent of carbodiimide favors formation of the corresponding symmetrical anhydride that may be isolated, but can also be used in the subsequent coupling reaction without further purification. Symmetrical anhydrides based on N^α -alkoxycarbonyl amino acids are stable towards hydrolysis. This permits purification analogously to that described above for the mixed anhydrides.

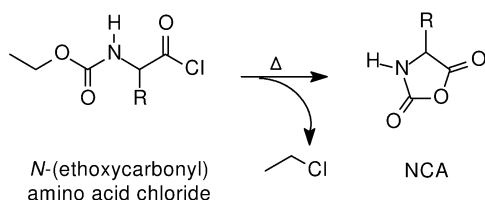
The general interest in symmetrical anhydrides for stepwise peptide chain elongation increased as Boc-protected amino acid derivatives became commercially available at reasonable prices. Although crystalline symmetrical anhydrides can be obtained commercially, *in situ* preparation remains the variant of choice.

4.3.2.3 N-Carboxy Anhydrides

Simultaneous activation of the carboxy group and acyl protection of an amino acid is found in the N-carboxy anhydrides (NCA), which were discovered in 1906 by Hermann Leuchs. In the German literature they are, therefore, also often called



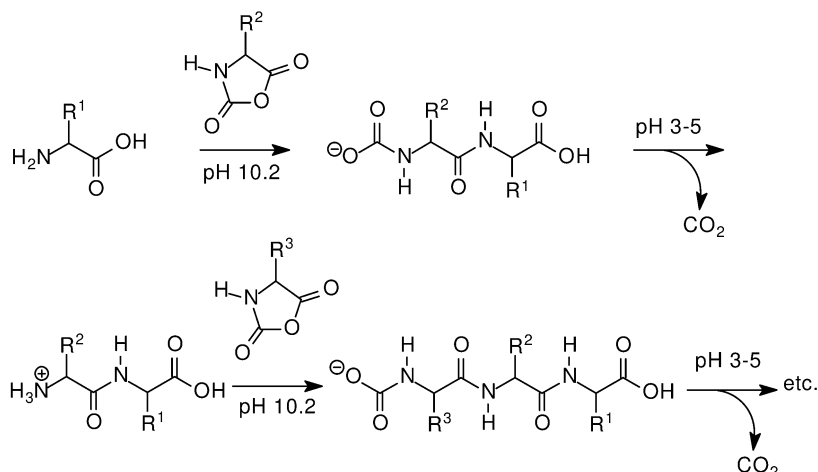
Scheme 4.5

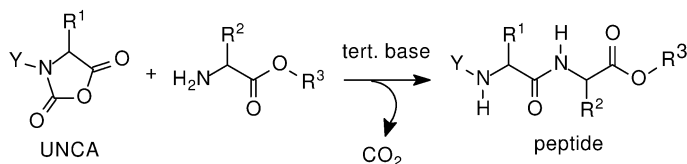
**Scheme 4.6**

Leuchs-anhydrides. In principle, these derivatives should possess ideal preconditions for application in peptide synthesis.

The first *N*-carboxy anhydride (1,3-oxazolidine-2,5-dione) was obtained by thermal elimination of chloroethane from *N*-(ethoxycarbonyl) amino acid chloride (Scheme 4.6).

An elegant procedure for the preparation of these derivatives is the reaction of free amino acids with phosgene via the corresponding carbamoyl chloride as an intermediate. However, even traces of water polymerize *N*-carboxy anhydrides, because the carbamic acid initially formed spontaneously decarboxylates to give a free amine that acts as a nucleophile in further ring-opening reactions. Therefore, the practical application of the NCA method in peptide synthesis was seriously limited until 1966, when correct reaction conditions were published which allowed for controlled peptide synthesis with *N*-carboxy anhydrides in an aqueous medium [267]. As shown in Figure 4.27, amino acids and peptides are rapidly acylated by *N*-carboxy anhydrides at low temperature and at pH 10.2. The intermediate peptide carbamates are decarboxylated by a decrease in the pH to values between 3 and 5. The next coupling cycle starts with an increase of the pH to 10.2, and proceeds by addition of the next *N*-carboxy anhydride. Vigorous stirring of the reaction mixture is necessary in order to minimize the risk of undesired carboxylate exchange between the intermediary

**Figure 4.27** Controlled peptide synthesis with *N*-carboxy anhydrides in an aqueous medium.



Scheme 4.7

peptide carbamate and the amino component. Exact control of the pH value (pH 10.2–10.5 for amino acids, and pH 10.2 for peptides) is another precondition, as hydantoins are formed as by-products at pH >10.5.

The sulfur analogues of the N-carboxy anhydrides, N-thiocarboxy anhydrides (NTA), can be favorably applied in peptide syntheses, because of the higher stability of the thiocarbamate salts. The acylations can be performed at pH values as low as 9–9.5, which prevents the possible hydrolytic conversion to the hydantoins. The NCA/NTA method is especially suited for segment syntheses without isolation of reaction intermediates. Trifunctional amino acids (except lysine and cysteine) do not require side-chain protection. This method has been used to assemble several segments of the ribonuclease S-protein that were then combined, using the azide method to produce the full-length S-protein [268].

In 1990, the NCA methodology again attracted significant interest, as *urethane-type protected N-carboxy anhydrides* (UNCA) were synthesized and applied to peptide syntheses [269]. NCA can be acylated at the ring nitrogen with suitable reagents for the introduction of a suitable protecting group (Y = Boc, Z, or Fmoc) in an aprotic solvent in the presence of a tertiary base to give the corresponding UNCA (Scheme 4.7). UNCA can be obtained for most amino acids as crystalline compounds, and are stable in the absence of water. They display high reactivity towards nucleophiles, and form the required peptide bonds at high rates in most anhydrous solvents (with the exception of alcohols) normally used in peptide synthesis, both in solution or on a polymeric support. Carbon dioxide is the only by-product, and there is no risk of either oligomerization or polymerization, because the amino terminus of the growing peptide chain is still urethane-protected after the coupling reaction.

Likewise, N-trityl- and N-phenylfluorenyl-protected NCA have been reported [270].

4.3.3

Carbodiimides

Carbodiimides can be used for the condensation of a carboxy group and an amino group [271, 272]. Initially *N,N'*-dicyclohexyl carbodiimide (DCC) was used and is now well established, because it is relatively cheap and soluble in the solvents used in peptides synthesis. During peptide bond formation the carbodiimide is converted into the corresponding urea derivative, which in the case of *N,N'*-dicyclohexyl urea precipitates from the reaction solution. Clearly, the different reaction rates of aminolysis and hydrolysis of the reactive intermediate after carbodiimide activation

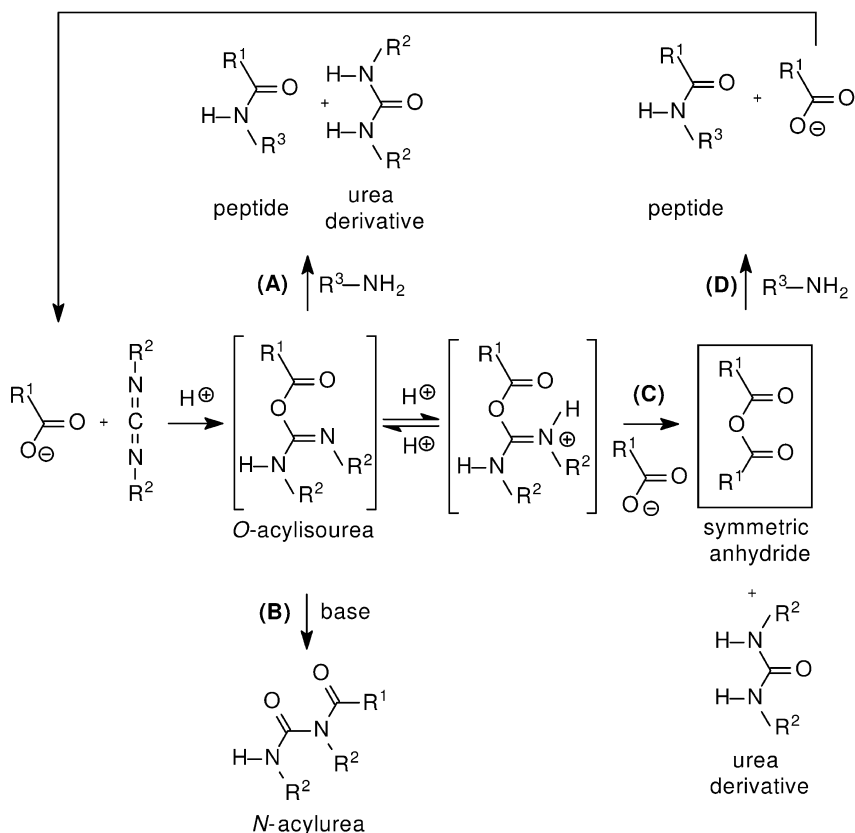
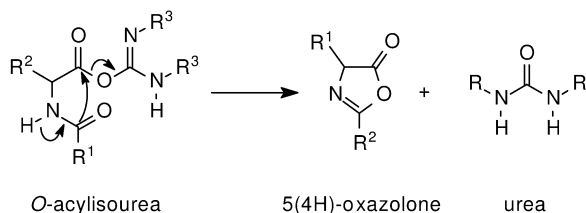


Figure 4.28 Mechanism of carbodiimide-mediated peptide couplings. $R^1\text{COOH}$ = carboxy component (amino acid or peptide); R^2 = e.g. cyclohexyl (DCC); $R^3\text{-NH}_2$ = amino component (amino acid or peptide).

enable peptide couplings in aqueous media. The reaction mechanism for carbodiimide-mediated peptide couplings has been established in extensive studies performed by several groups, and is shown in Figure 4.28.

The carboxylate anion adds to the protonated carbodiimide with formation of a highly reactive *O*-acylisourea; this has not yet been isolated, but its existence is postulated by close analogy to stable compounds of this class. The *O*-acylisourea reacts with the amino component to produce the protected peptide and the urea derivative (Figure 4.28, reaction A). Alternatively, the *O*-acylisourea – in equilibrium with its *N*-protonated form – is attacked by a second carboxylate as the nucleophile forming the symmetrical amino acid anhydride together with *N,N'*-disubstituted urea (Figure 4.28, reaction C). The former reacts with the amino acid component to give the peptide derivative and the free amino acid (Figure 4.28, reaction D). An undesired side reaction is the base-catalyzed acyl migration from the isourea oxygen to nitrogen; this results in an *N*-acylurea that does not undergo further aminolysis (Figure 4.28, reaction B). This $\text{O} \rightarrow \text{N}$ acyl migration is catalyzed not only by an



Scheme 4.8

excess of base, but also by the basic amino component or the carbodiimide. Polar solvents additionally favor this reaction pathway.

Furthermore, the *O*-acylisourea may undergo cyclization to give the 5(4*H*)-oxazolone (Scheme 4.8), again with liberation of the *N,N'*-dialkylurea as a good leaving group. This intramolecular side reaction explains the high propensity of the carbodiimide method towards racemization, when applied to *N*-acyl-protected amino acids and peptides. 5(4*H*)-Oxazolones are themselves acylating agents, but are prone to racemization. Derivatives with *N*-protecting groups of the urethane type are the only exception. Contrary to earlier speculation, 2-alkoxy-substituted 5(4*H*)-oxazolones are real intermediates that are formed during this reaction from the *O*-acylisourea (cf. Section 4.4.2).

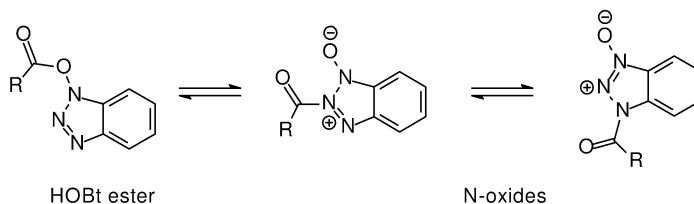
The DCC method is a reliable variant for the stepwise assembly of peptides using urethane-type protected amino acids (Z, Boc, Fmoc) in both solid-phase and solution-phase peptide synthesis, as well as for segment condensations. However, only peptide segments containing C-terminal glycine or proline residues should be used as acyl components, because of the risk of epimerization at the C-terminal position. The *N*-acyl urea formation mentioned above may be a prominent side reaction that can be suppressed by keeping the reaction at low temperatures, or by using nonpolar solvents in which the dicyclohexyl urea should, nevertheless, dissolve. Free dimethylamine, which is present in low concentrations even in highly pure DMF or *N,N*-dimethylacetamide also favors the $O \rightarrow N$ acyl migration to produce *N*-acyl urea. When activated with DCC, unprotected glutamine or asparagine side chains may undergo dehydration of the carboxamide group to yield the nitrile. Therefore, a global (maximum) protection strategy for the side-chain functionalities should be applied when DCC couplings are involved.

Besides *N*-acyl urea formation, a serious problem in peptide synthesis when using the carbodiimide method – and especially in the synthesis of difficult sequences – is imposed by incomplete separation of the *N,N'*-dialkyl urea, and especially of *N,N'*-dicyclohexyl urea. Several carbodiimide derivatives which have been substituted with different tertiary amino or quaternary ammonium groups have been introduced. The ureas formed during peptide synthesis with these modified reagents are easily separated from the reaction mixture because of their increased solubility in aqueous or acidic aqueous solutions. Reagents of this type, e.g., 1-ethyl-3-(3'-dimethylaminopropyl)-carbodiimide HCl salt (EDC, WSC·HCl) or 1-cyclohexyl-3-(3'-trimethylammoniopropyl) carbodiimide iodide, are appropriate for peptide couplings, even in alcohol [273] (or aqueous solution as required), for coupling reactions involving

proteins or peptide cyclizations. In addition, polymeric carbodiimides have been developed which allow straightforward separation of the urea formed during the reaction. Unsymmetrically substituted carbodiimides have been reported to be suitable for suppression of the $O \rightarrow N$ acyl migration forming the undesired *N*-acyl urea, as was shown for 1-benzyl-3-ethylcarbodiimide [274]. Simultaneously, a lower tendency towards racemization has been observed. At present, DCC, EDC and diisopropyl carbodiimide (DIC) are frequently used in peptide synthesis.

The *DCC-additive method* is based on the addition of a selected nucleophile that reacts faster than the competing acyl transfer and generates an intermediate still active enough to couple with the amino component and also suppress or diminish both *N*-acyl urea formation and racemization. The DCC-additive protocol was introduced into peptide chemistry in 1966 [275, 276], whereby one equivalent of DCC and two equivalents of *N*-hydroxysuccinimide (HOSu) were applied simultaneously (Wünsch–Weygand protocol). Without doubt, the König–Geiger protocol [277] belongs to the DCC-based coupling methods that have found widespread application. In this variant, 1-hydroxybenzotriazole (HOBt) is mainly preferred as the additive. Like all additives of this type, 1-hydroxybenzotriazole (HOBt) reacts rapidly with the *O*-acyl isourea initially formed to produce the corresponding benzotriazole active ester. The latter is less reactive than the *O*-acyl isourea, reducing racemization and avoiding the formation of other unreactive by-products. The N^α -protected amino acid HOBt ester (Scheme 4.9) is thought to exist in solution in an equilibrium with two *N*-oxide forms. The X-ray crystal structure of Trt-Met-OBt has been determined and corresponds to the *N*-oxide form on the right-hand side in Scheme 4.9 [278].

Benzotriazolyl esters undergo aminolysis up to three orders of magnitude faster than succinimidyl or related esters. Thus, rearrangements, cyclization, and the formation of symmetrical anhydrides are efficiently suppressed. Dehydration of asparagine or glutamine residues is also not observed. Likewise, 2-alkoxy-5(4*H*)-oxazolones react immediately with HOBt or HOSu, respectively, to produce the active esters that undergo peptide coupling with amino components present in the reaction mixture. Many other *N*-hydroxy compounds besides HOSu and HOBt have been suggested as potential additives for the DCC method, for example 1-hydroxy-5-chloro-benzotriazole (Cl-HOBt) [279] or the currently most efficient additive 1-hydroxy-7-azabenzotriazole (HOAt). The latter combines the features of HOBt and a tertiary base in the same molecule, and is reported to accelerate reaction rates as well as provide concomitant optimal coupling yields and minimal racemiza-



Scheme 4.9

tion [280–283]. Mechanistic details are given in Section 4.3.4. Further additives are, e.g., ethyl 1-hydroxy-1*H*-1,2,3-triazole-4-carboxylate (HOCT) [284] and other hydroxy-substituted triazole and tetrazole derivatives [285].

Cl-HOBt is more acidic than HOBt and the chloro substituent its structure and makes it a less hazardous reagent. Due to its efficiency and stability Cl-HOBt is more suitable for industrial use. For example, it was used together with HBTU to suppress racemization during fragment condensations in the industrial synthesis of T20 (enfuvirtide, Fuzeon™). Nowadays, onium (phosphonium and guanidinium/uro-nium) salts (cf. Sections 4.3.6 and 4.3.7) of hydroxybenzotriazole derivatives are also available as excellent coupling reagents.

4.3.4

Active Esters

Peptide bond formation by ester aminolysis ($X=O$, S) proceeds analogously to ester saponification according to a $B_{AC}2$ mechanism (Figure 4.29). Nucleophilic attack of the amino component on the activated carboxy moiety of the carboxy component leads, in the rate-determining step of this reaction, to the tetrahedral intermediate, the formation of which is favored by electron-withdrawing groups XR^3 .

The rate of peptide bond formation correlates with the leaving group capacity of XR^3 . For example, a good leaving group is the conjugate base of a relatively strong acid, and this is the case when the leaving group anion $^-XR^3$ is well-stabilized by inductive or mesomeric effects which facilitate cleavage of the $(C=O)-XR^3$ bond.

The initial application, e.g., of thiophenyl-, 4-nitrophenyl-, 2,4,5-trichlorophenyl- and pentafluorophenyl esters triggers further development of a broad variety of active esters [286].

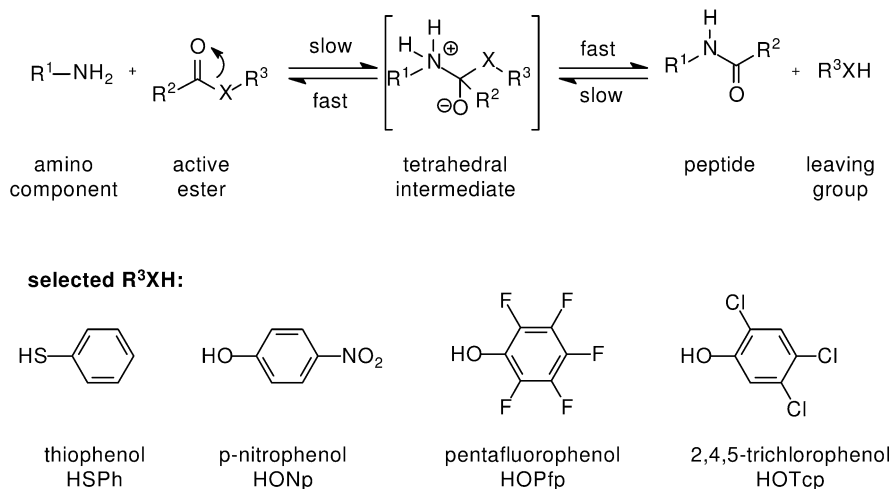


Figure 4.29 Mechanism of conventional ester aminolysis.

Active esters of protected amino acids can be synthesized by the mixed anhydride method, the carbodiimide method, the carbonate method, and other more specialized variants. Protected amino acid active esters can also be obtained starting from amino acid salts in a “one-pot” reaction, provided that the carbonate component has been carefully chosen. Treatment of amino acid salts with benzyl (4-nitrophenyl) carbonate or *tert*-butyl (4-nitrophenyl)carbonate, respectively, gives the corresponding Z- or Boc-protected derivatives that, upon acidification and further addition of a carbodiimide, yield the amino-protected 4-nitrophenyl esters. These are crystalline compounds that may be stored in the dark at room temperature for longer periods of time and which display high reactivity in aminolysis reactions. Aminolysis may even be further enhanced after the addition of catalysts such as azoles or N-hydroxy compounds. Halogen-substituted phenyl esters usually are crystalline compounds with high aminolysis activity. Separation of the halogenated phenols liberated during reaction of these active esters with amines is often less problematic than separation of the corresponding 4-nitrophenols. The “backing off” protocol is a suitable method for the synthesis of optically pure N-protected activated peptidyl esters.

For mechanistic reasons, the isoxazolium method [287] can be discussed in the context of this chapter, because the N-ethyl-5-phenylisoxazolium-3'-sulfonate (Woodward reagent K) forms an N-protected enol ester with N-protected amino acids and peptides under mildly basic conditions.

Anchimeric assisted active esters are characterized by a different mechanistic pathway compared to the $B_{AC}2$ mechanism of substituted alkyl esters, unsubstituted or substituted aryl esters, or their thio analogues. The significant mechanistic difference to the latter active esters is based on intramolecular base catalysis by neighboring group effects (anchimeric assistance). The first examples were N-hydroxypiperidinyl esters described by Young *et al.* [288] and 8-quinolyl esters described by Jakubke and Voigt [289]. The 8-quinolyl esters featured an additional proton-accepting group stabilizing a hydrogen-bonded transition state during aminolysis. This led to high aminolytic activity and extensively decreased the racemization-prone oxazolone formation. This general mechanistic principle is outlined in Figure 4.30.

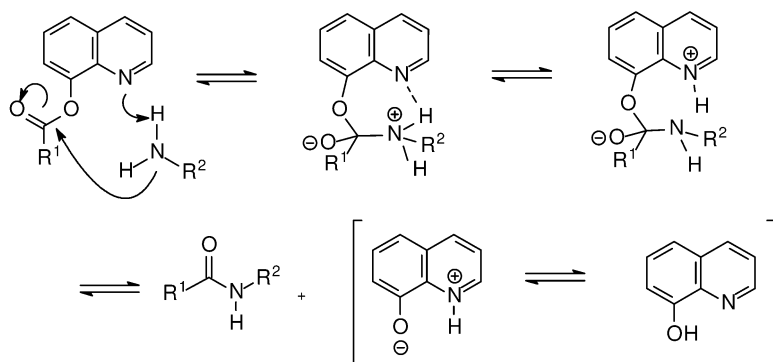
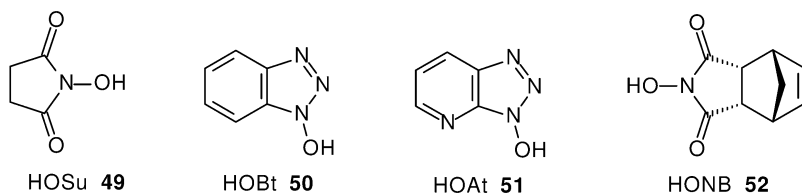


Figure 4.30 Intramolecular base catalysis during aminolysis of acylamino 8-quinolyl esters. R^1COOH = carboxy component (protected amino acid or peptide); R^2NH_2 = amino component (amino acid or peptide).

Comparison of the pK_a values of the isomeric 3-hydroxyquinoline (8.06), 6-hydroxyquinoline (8.88), and 8-hydroxyquinoline (9.89) leads to the conclusion that the leaving group ability of the corresponding bases should decrease in the same order. The opposite holds true; 8-quinolyl esters display significantly higher aminolysis reactivity because nucleophilic attack of the amino group on the carboxy group of the ester is favored by the anchimeric assistance of the quinoline nitrogen. This nitrogen atom acts as a basic catalyst and “delivers” the attacking amino group to the reactive center. As the reaction proceeds, the tetrahedral intermediate is further stabilized by an intramolecular hydrogen bond. Racemization of 8-quinolyl esters is disfavored, because the amide oxygen cannot attack the weakly electrophilic ester group intramolecularly to give an oxazolone. Hence, efficient discrimination between aminolysis and oxazolone-based racemization (cf. Section 4.4.2) is possible.

As demonstrated by Carpino *et al.* [290], this very mechanistical principle applies also to those active esters that have high practical importance, such as derivatives of *N*-hydroxysuccinimide (HOSu) **49**, 1-hydroxybenzotriazole (HOBT) **50**, the 7-aza analogue of HOBT, commonly referred to as 7-aza-1-hydroxybenzotriazole (HOAt) **51**, and *N*-hydroxy-5-norbornene-*endo*-2,3-dicarboximide (HONB) **52** [291]. Special structural features of HOBT esters have already been discussed in Section 4.3.3. In the DCC/additive one-pot procedure the appropriate active ester is formed *in situ* as an intermediate and subsequently reacts with the amino component. Most of these moieties have been recommended as highly efficient additives for DCC coupling reactions. Especially, 7-aza-1-hydroxybenzotriazole is most efficient and its active esters are characterized by very little racemization. HOAt as an additive is superior to HOBT for stepwise peptide couplings and segment condensations with respect to reaction rate and the degree of racemization. Furthermore, in one-pot coupling procedures using HOBT- and HOAt-based onium salts the appropriate active esters are also prepared *in situ* as will be discussed in Sections 4.3.6 and 4.3.7, respectively.



4.3.5

Acyl Halides [292]

By analogy to organic chemistry, an obvious method of activating a carboxy group for amide formation at room temperature or below would utilize the acid chloride. However, amino acid chlorides have until recently [65, 293] rarely been used as they are reputed to be “overactivated” and prone to numerous side reactions, including racemization.

Despite this, because the Fmoc group is highly stable under acidic conditions and does not readily undergo S_N1 or S_N2 displacement reactions at the 9-fluorenylmethyl residue, the Fmoc-protected amino acids have emerged as ideal substrates for conversion into the corresponding acid chlorides. Consequently, these derivatives have experienced a renaissance in peptide synthesis. All common amino acids lacking polar side chains, as well as a number of those bearing benzyl- or allyl-based side-chain protection, can be converted to stable, crystalline acid chlorides. For chemical reasons, amino acids bearing *tert*-butyl-protected side chains cannot generally be accommodated. A hindered base such as 2,6-di-*tert*-butylpyridine is used routinely as a hydrogen chloride acceptor, as conversion to oxazolone is slow with such bases. Ammonium or potassium salts of *N*-hydroxybenzotriazole (HOBt) have also been applied as bases in solid-phase syntheses involving Fmoc amino acid chlorides.

Fmoc amino acid chlorides have been used as stable, easily accessible derivatives for rapid coupling reactions, without any racemization. Their general application is, however, somewhat limited as not all Fmoc-protected amino acid derivatives are accessible. This holds true especially for trifunctional amino acids with side chain-protecting groups of the *tert*-butyl type (Boc, *tert*-butyl ester, *tert*-butyl ether).

In contrast, the corresponding acid fluorides do not suffer from such limitations. Urethane-type protected amino acid fluorides of trifunctional amino acids with a side-chain protection of the *tert*-butyl or trityl type can be obtained and are sufficiently stable.

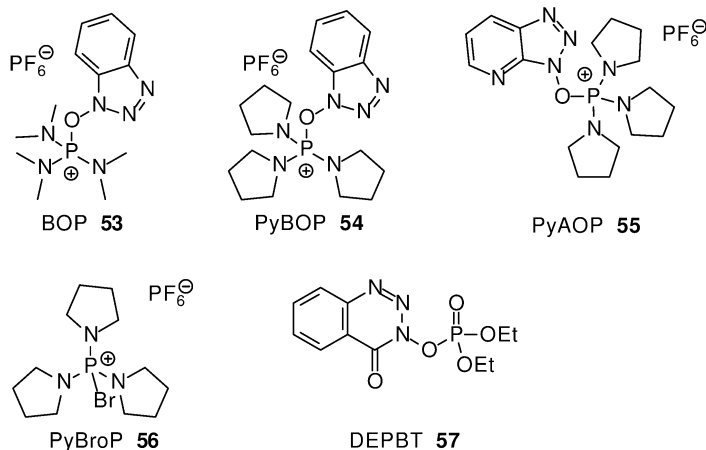
Further advantages of the acid fluorides relative to the chlorides include their greater stability toward water, including moisture in the air, and their relative lack of conversion to the corresponding oxazolones on treatment with tertiary organic bases. Fmoc-protected amino acid fluorides were found to be efficient carboxy-activated derivatives suited both for solution syntheses and for SPPS [294–296]. The highly reactive Fmoc amino acid fluorides are especially recommended for SPPS of complicated longer peptides, and also mainly for the coupling of sterically hindered amino acid building blocks [297]. The advantage of amino acid fluorides in the coupling of sterically hindered amino acid components, compared to other coupling methods, is clearly based on an efficient stabilization of the tetrahedral transition state by the highly polarized C–F dipole and the relatively small size of the fluoride leaving group. Acylations with Fmoc amino acid fluorides also occur in the absence of a tertiary base – a fact which is important with respect to minimal racemization [298]. Fmoc amino acid fluorides are usually stable crystalline derivatives and can be synthesized from the *N*-protected amino acid with 2-diethylamino sulfur trifluoride (DAST), cyanuric fluoride, or tetramethyl fluoroformamidinium hexafluorophosphate $[(Me_2N)_2CF]^+PF_6^-$ (TFFH). The latter reagent is a non-hygroscopic, stable salt that is also suitable for generating the acyl fluoride *in situ* on treatment with ethyl diisopropyl amine and is, therefore, also regarded as a coupling reagent for solution- or solid-phase syntheses [299].

4.3.6

Phosphonium Reagents

Coupling reagents allowing for the *in situ* generation of active esters are of considerable importance in peptide chemistry.

The HOBt-derived benzotriazol-1-yloxy-tris(dimethylamino)phosphonium hexafluorophosphate (BOP) **53**, also called Castro's reagent [300], is a very efficient reagent which has also been applied successfully in SPPS. The main disadvantage of this reagent, however, is that the highly toxic and carcinogenic HMPA is formed during the course of the reaction, whilst an unacceptably high risk of racemization has also been observed [301]. Benzotriazol-1-yloxy-tris-pyrrolidinophosphonium hexafluorophosphate, termed PyBOP **54** [302], has therefore been developed. The PyBOP coupling reagent and its HOAt analogue PyAOP **55** [303], in which the dimethylamino groups are replaced by pyrrolidine substituents, represent viable alternatives to BOP which do not produce the hazardous by-product. Sterically hindered amino acids can often be coupled successfully using bromo-tris(pyrrolidino)phosphonium hexafluorophosphate PyBroP **56** [304]. These reagents do not react with α -amino groups, and so may be added directly to a mixture of the amino and carboxy component to be coupled. A tertiary amine is usually added in order to form the anion of the carboxy component. Nucleophilic attack of the carboxylate leads to a highly reactive acylphosphonium species that is immediately transformed into the HOBt ester in the presence of HOBt.

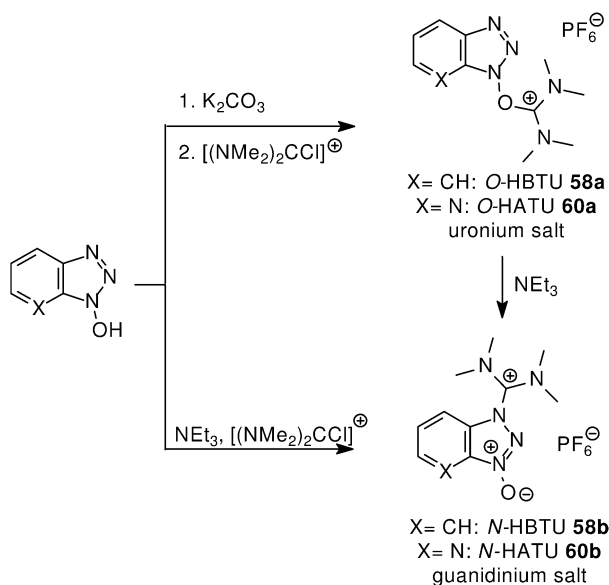


3-(Diethoxyphosphoryloxy)-1,2,3-benzotriazin-4(3H)-one (DEPBT) **57** is a coupling reagent with superior performance reported by the Goodman group [305]. DEPBT mediates peptide bond formation with remarkable resistance to racemization, and it is not necessary to protect the hydroxy group (Ser, Thr, Tyr) of the amino component and the imidazole function of histidine [306].

4.3.7

Guanidinium/Uronium Reagents

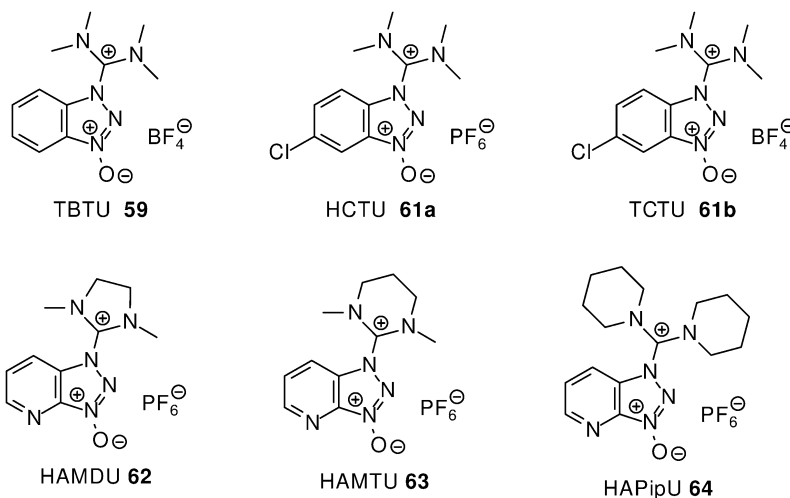
Several guanidinium/uronium salts have found widespread application as coupling reagents in SPPS, for segment condensations, and other purposes of peptide and protein chemistry. Among them is the progenitor of uronium reagents HBTU **58** and its tetrafluoroborate TBTU **59** introduced in 1978 [307, 308]. Interestingly, the counterion has no significant influence on the reactivity and racemization of the reagent. By analogy with phosphonium salts, HBTU was originally assumed to be the uronium salt **58a** (*O*-(1*H*-benzotriazol-1-yl)-*N,N,N'*-tetramethyluronium hexafluorophosphate, *O*-HBTU) as shown in Scheme 4.10. However, Carpino *et al.* [309] elucidated the true structure of HBTU, HATU, and similar derivatives as they are formed according to the originally described procedure. They exist in the solid state and in solution as guanidinium salts (*N*-HBTU **58b**, *N*-HATU **60b**) [310]. Carpino *et al.* could also obtain the true uronium salts *O*-HBTU **58a** and *O*-HATU **60a** by using the potassium salt of the corresponding benzotriazole (KOBt, KOAt). According to the IUPAC nomenclature, *N*-HATU, the commercially available form of HATU must be named 1-[bis-(dimethylamino)methylumyl]-1*H*-1,2,3-triazolo[4,5-*b*]pyridine-3-oxide hexafluorophosphate or as 1-[(dimethylamino)-(dimethyliminium)methyl]-1*H*-1,2,3-triazolo[4,5-*b*] pyridine-3-oxide hexafluorophosphate. In the presence of tertiary amines the uronium salts isomerize to give the guanidinium salts (Scheme 4.10) [310]. It is very likely that other guanidinium/uronium reagents [311] display the same behavior. However, the given acronyms are, maintained. HATU **59**, *N*-[(dimethyl-amino)-1*H*-1,2,3-triazolo



Scheme 4.10

[4,5-*b*]pyridine-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate [312, 313] has been proven to be very efficient in difficult sterically hindered coupling reactions and usually causes a minimal level of racemization [280]. However, because of its high price this most reactive guanidinium/uronium salt has not found wide application in industry. HCTU **61a** and TCTU **61b** have been found to act as useful alternatives to HBTU **58**, TBTU **59**, and HATU **60**. HCTU is a nontoxic, nonirritating, and noncorrosive coupling reagent. It is stable in DMF, matches HATU in efficiency and represents a suitable coupling reagent for fast Fmoc SPPS [314].

From the practical point of view it should be mentioned that the counterion tetrafluoroborate improves the solubility of guanidinium salts in contrast to hexafluorophosphate. Further examples of guanidinium salts of HOAt are HAMDU **62**, HAMTU **63** and HAPipU **64**.



Mechanistically, the guanidinium/uronium reagents function similarly to the phosphonium reagents mentioned above. Nucleophilic attack of the carboxylate results in an acyloxy-phosphonium or O-acylisouronium salt intermediate that reacts intermediately with the benzotriazole derivative to give the appropriate active ester which reacts with the amino component forming the peptide bond (Figure 4.31). Up to now, intermediate acyloxy-phosphonium species or O-acylisouronium salts have not been detected.

Unlike carbodiimides or phosphonium salt reagents, guanidinium/uronium salt reagents may undergo a reaction with free amino groups present. Consequently, the formation of tetramethylguanidinium derivatives of free amines (e.g., N-terminal amino groups) has been observed [315, 316]. The replacement of the dimethylamino moieties by pyrrolidino groups has also been suggested [317], and other active ester types, such as in PfPyU, have been considered [318].

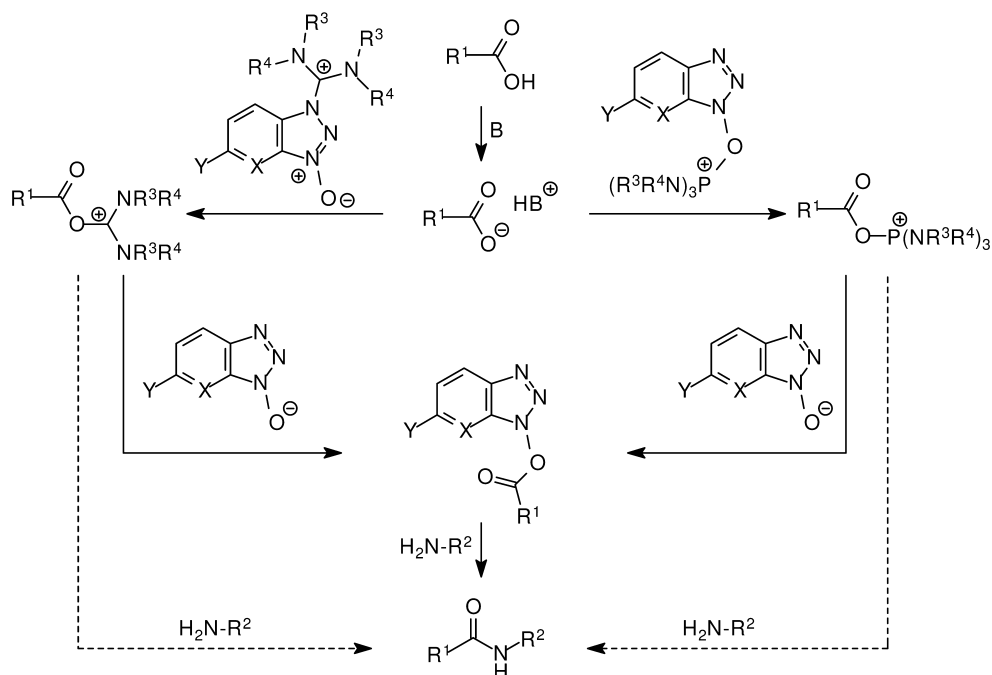


Figure 4.31 Mechanism of peptide bond formation via activation by onium (phosphonium and aminium/uronium) salts of hydroxybenzotriazole derivatives.

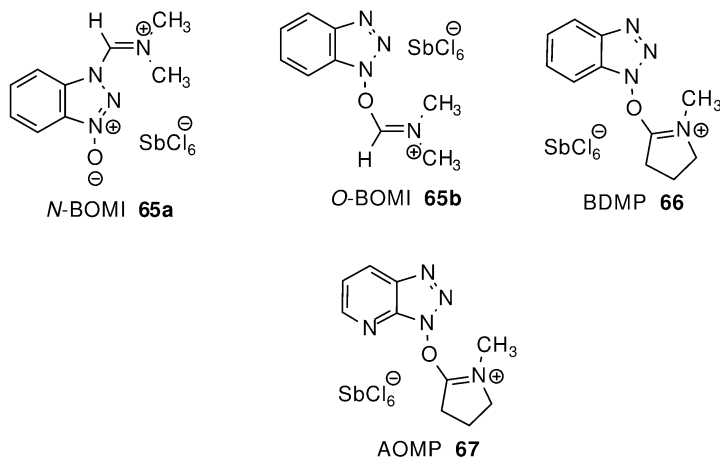
All phosphonium and guanidinium/uronium salt reagents based on HOBT or HOAt serve as direct coupling reagents. A comparative study [316] and theoretical investigations [319] on the mechanism have been published. For example, BOP **53**, HATU **59**, and HAPipU **64** are especially suited for peptide cyclizations [320]. 2-Trifluoroacetylthiopyridine-HOBT as a highly efficient reagent for segment couplings must also be mentioned in this context [321]. A mixture of trifluoroacetylthiopyridine, the sodium salt of HOBT, and a carboxyl component was used for the acylation of amino component esters in high yields and with minimal racemization.

4.3.8

Immonium Type Coupling Reagents

Immonium type coupling reagents have been designed by replacement of the amino group of the central carbon atom in known guanidinium/uronium reagents with a hydrogen, an aryl or an alkyl group [322]. They are an extension of the appropriate uronium-based counterparts and analogous mechanisms can be assumed. BOMI **65**, 5-(1*H*-benzotriazol-1-yloxy)-*N,N*-dimethylmethaniminium hexachloroantimonate, and BDMP **66**, 5-(1*H*-benzotriazol-1-yloxy)-3,4-dihydro-1-methyl-2*H*-pyrrolonium hexachloroantimonate, seem to be more reactive than, for example, AOMP **67**. As in the

case of HBTU and HATU it was shown by X-ray analysis that the amidinium form *N*-BOMI **65a** prevails in the crystalline form rather than *O*-BOMI **65b**.



It has been reported that immonium type reagents gave better results than guanidinium/uronium reagents because the electron density of their entral carbon is lower than that of the corresponding guanidinium/uronium salts. BOMI found application both in solution phase and solid phase peptide synthesis.

4.3.9

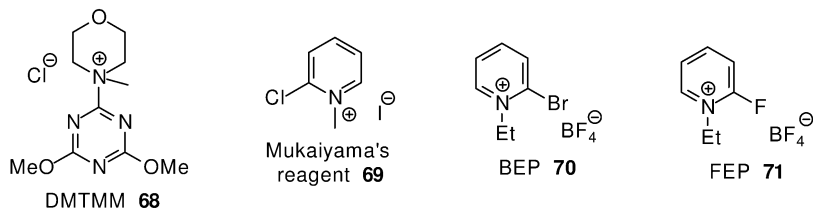
Further Special Methods for Peptide Synthesis

Ammonium salts such as triazinyl esters and Mukaiyama's reagent have also found application in amide bond formation beside guanidinium/uronium- and iminium-based coupling reagents. The very cheap commercially available 2-chloro-4,6-dimethoxy-1,3,5-triazine has also been used. 4-(4,6-Dimethoxy-(1,3,5) triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM) **68** as a very efficient activating agent for an one-pot peptide coupling procedure was first described by Kamiński *et al.* [323, 324]. DMTMM initially undergoes S_NAr reaction forming an activated triazinyl ester that reacts with the amino component [325]. No additional base is required since *N*-methylmorpholine is liberated in the first step of the reaction.

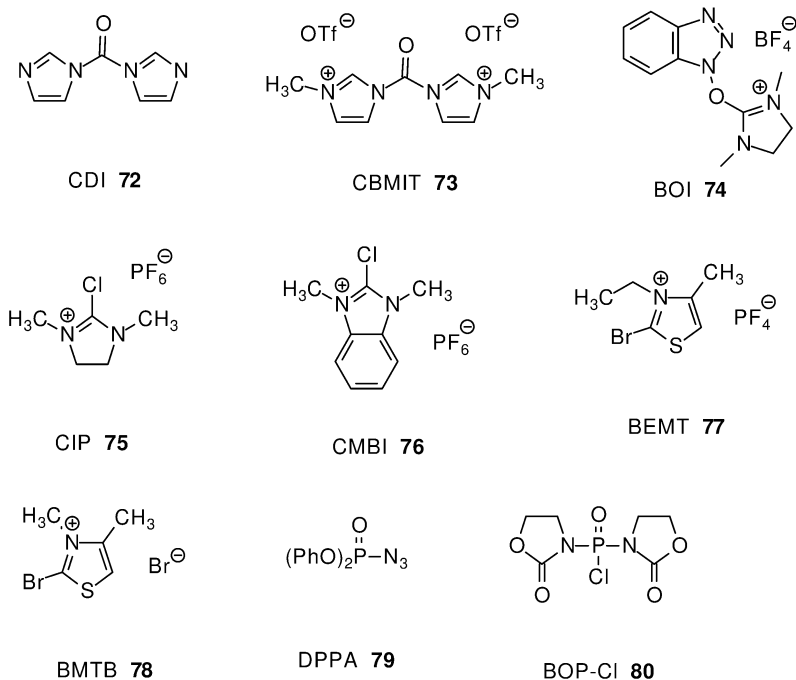
Mukaiyama's reagent, 2-chloro-1-methylpyridinium iodide **69**, gives as an intermediate an activated pyridinium ester that reacts with the amino component in a one-pot coupling procedure [326]. This method has not often been used because of the poor solubility of the pyridinium iodides in conventional solvents.

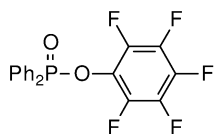
α-Halopyridinium-type coupling reagents have been described by Xu's group [327] as alternatives to **69**, such as BEP **70**, 2-bromo-1-ethyl pyridinium tetrafluoroborate and FEP **71**, 2-fluoro-1-ethyl pyridinium tetrafluoroborate, which found application in SPPS, especially for coupling of *N*-methyl amino acid. Tetrafluoroborate or hexa-

chloroantimonate was chosen as the counterion to improve the solubility of the pyridinium reagents.

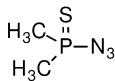


Imidazolium reagents are rather old coupling reagents because 1,1'-carbonyldiimidazole, CDI **72**, was introduced as early as 1958 [328]. Nucleophilic attack of the carboxylate moiety at the carbonyl carbon of CDI leads to the active imidazolide intermediate. Interestingly, it has been reported that in the synthesis of analogs of mosapride [329] CDI was more efficient than 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide (EDC). Bismethylating of CDI using methyl triflate led to CBMIT **73** [330] which allows coupling reactions with sterically hindered amino acids. The imidazolium reagent BOI **74** and its precursor CIP **75** were developed by Kiso's group [331]. Especially, CIP/HOAt combined coupling reagents provided the best results in the synthesis of Aib-containing peptides. The benzene ring-fused derivate of CIP, termed CMBI **76**, is also an efficient coupling reagent [332] as is the thiazolium-type reagent BEMT **77** [333]. In contrast to the latter, BMTB **78** [334] is a crystalline and non-hygroscopic compound.

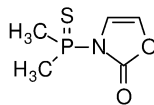




FDPP 81



MPTA 82



MPTO 83

Organophosphorus reagents have been developed for peptide coupling since the introduction of the mixed carboxylic-phosphoric anhydride approach by Yamada and coworkers in 1972 [335]. Propanephosphonic anhydride has also been reported as a coupling reagent [264]. DPPA **79** is used to synthesize acyl azides and an improved DPPA approach replaces triethylamine with sodium bicarbonate as the base. The DPPA/ NaHCO_3 procedure has been successfully used for peptide cyclization [336]. Notable variations in organophosphorus reagents were the development of BOP-Cl **80** [337] and FDPP **81** [338] which have found wide application in the macrocyclization of cyclic depsipeptides. In contrast to DPPA, the thiophosphonic-type coupling reagents MPTA **82** and MPTO **83** are crystalline substances that are stable for long-term storage [339].

Ionic liquid-mediated peptide synthesis (LMPS) peptide synthesis uses ionic liquids [340] as non-conventional media or as a soluble support. The application of ionic liquids as non-aqueous reaction media in biocatalysis has been described, including an enzyme-catalyzed synthesis of aspartame in 2000 [341]. Some years later, chemical peptide synthesis in room-temperature ionic liquids using coupling agents such as HATU and BOP was demonstrated [342]. The new approach presents some advantages, especially in the coupling of sterically hindered amino acid derivatives. Furthermore, *ionic liquid-supported synthesis* [343] has been developed. As in liquid-phase peptide synthesis the first amino acid is anchored to a suitable ionic liquid, e.g., 3-hydroxyethyl-(1-methylimidazolium)-tetrafluoroborate using common coupling agents. After solvent washing, the next Boc-amino acid was coupled step-by-step followed by detachment from the ionic liquid and phase-separation.

Simultaneous protection/activation approach is termed a concept that is based on the generation of N^α -protected amino acids with concomitant activation of the carboxy group. For such a purpose the well-known N-carboxy anhydrides (NCA) may be employed. NCAs are derived from an amino acid, and formally composed of a carbamic acid (N-carboxy group) and a carboxylic acid (carboxy group) to give a mixed anhydride. Their practical application in peptide synthesis suffers from the high tendency towards ring-opening polycondensation, with the consequence that N-carboxy anhydrides can only be employed under carefully controlled reaction conditions (cf. Section 4.3.2.3). Alternatively, in the *bis(trifluoromethyl)-1,3-oxazolidin-5-one approach* [344] hexafluoroacetone is used as a reagent that simultaneously protects the N^α -group and activates the carboxy group of an amino acid. This bidentate protecting/activating concept, developed by Burger and coworkers, with a minimum number of synthetic steps can be efficiently employed in peptide synthesis. Reaction of

the activated lactone moiety with the N-terminus of a peptide releases hexafluoroacetone from the N,O-ketal. Racemization does not occur during this procedure, and the activated derivatives are stable for storage.

The *O-propargyl-O'-pentafluorophenyl carbonate approach* [345] is a different example of the simultaneous N $^{\alpha}$ -protection and carboxy group activation concept. The reaction of *O*-propargyl-*O'*-pentafluorophenyl carbonate with amino acids brings about Poc-protection (propargyloxycarbonyl-protection) of the N $^{\alpha}$ -group with intermediate formation of a mixed anhydride that readily forms the pentafluorophenyl ester in a one-pot reaction. After peptide coupling, the Poc-group can be cleaved using benzyl triethylammonium tetrathiomolybdate.

Forced peptide synthesis [346] is termed an approach of amide bond formation by physically pushing the reagents together. A plasma-oxidized flat PDMS stamp inked with a Boc amino acid was pressed into contact with an amine monolayer on gold to form an amide bond. Following this procedure, an RGD peptide was synthesized and shown to support cell adhesion, and a 20-mer PNA strand was synthesized step-by-step on the surface and proven to hybridize with a 16-mer dsDNA.

4.4

Racemization During Synthesis

An inherent risk of racemization is imposed on all reactions involving functional groups directly connected to the stereogenic center of an α -amino acid. This may lead to a partial or total loss of the stereochemical information. While the expression "racemization" in organic chemistry is used for the complete conversion of a single enantiomer into the racemate, it is often used in peptide chemistry also for partial or total epimerization at one chiral center [347], irrespective of whether a mixture of diastereomers or enantiomers is formed in the course of the process. The optical purity and stereochemical integrity of a synthetic peptide consequently depends very much on the degree of partial epimerization during coupling steps or other types of reactions sensitive toward racemization. If only 1% epimerization occurs on each amino acid residue during peptide synthesis of a 10-peptide, the mixture of diastereomers contains only 90.4% of the peptide with the correct stereochemistry. In the case of a 50-peptide, only 60.5% of the desired diastereomer is obtained in theory. As the biological activity of a peptide or a protein depends strongly on the configuration of the chiral centers of the building blocks, racemization must be minimized during peptide synthesis [347, 348].

4.4.1

Direct Enolization

Although free amino acids are usually configurationally stable, the situation may be different when carboxy-activated derivatives **84** are involved. In an equilibrium, reversible carbanion formation **85** leading to a partial loss of the stereo-

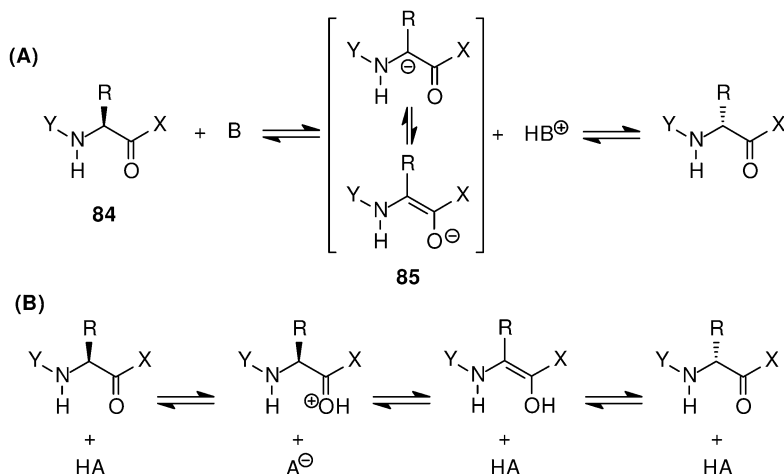


Figure 4.32 Mechanisms of base-catalyzed (A) and acid-catalyzed epimerization (B) of activated amino acid and peptide derivatives. X = activating group; Y = N^α-protecting group of peptide chain; B = base; HA = acid; R = amino acid side chain.

chemical integrity may be observed, because re-protonation occurs without facial selectivity.

The acidity of the proton H^α of an amino acid depends on the nature of the substituents. Only a base-catalyzed enolization mechanism (Figure 4.32A) plays a significant role under peptide coupling conditions. The rate of racemization depends on the electron-withdrawing effect of the groups R, Y, and X, on the solvent, temperature, and on the chemical properties of the base B used. Base-catalyzed racemization has been observed under different coupling conditions, during ester saponification, etc., while an acid-catalyzed enolization (Figure 4.32B) results in the racemization of *N*-substituted *N*-methyl amino acids upon reaction with HBr/HOAc.

4.4.2

5(4H)-Oxazolone Mechanism

Racemization during coupling reactions is mainly caused during amino acid activation by the formation of stereochemically labile 5(4*H*)-oxazolones (2-oxazolin-5-ones, azlactones) D/L-**88** as reaction intermediates (Figure 4.33).

The propensity toward 5(4*H*)-oxazolone formation strongly correlates with the activation potential of the activating group X in **86**, and with the electronic properties of the N-acyl residue R¹-CO-. Acyl residues such as acetyl, benzoyl, or trifluoroacetyl favor oxazolone formation due to the nucleophilic potential of the carbonyl oxygen. The mesomeric electron donation by the amino group enhances the nucleophilic attack of the oxygen atom on the activated carboxy group (Figure 4.33). The oxazolone L-**88** formed by this process from an L-configured amino acid, is prone to epimerization in the tautomeric equilibrium between L-**88**, **90**, and D-**88**. An oxazolone of type **88**

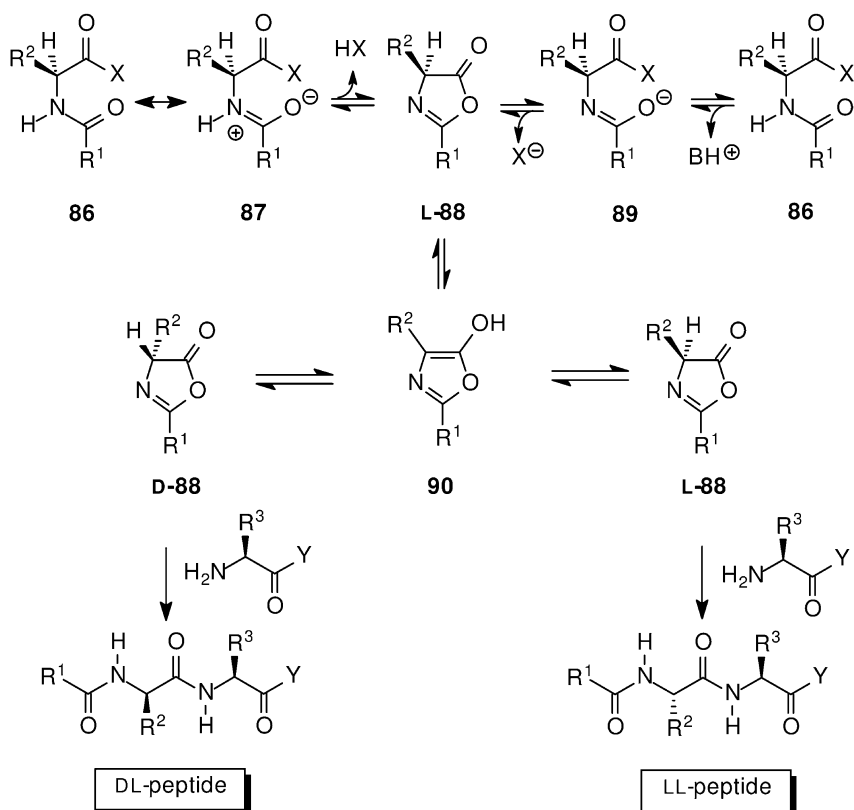
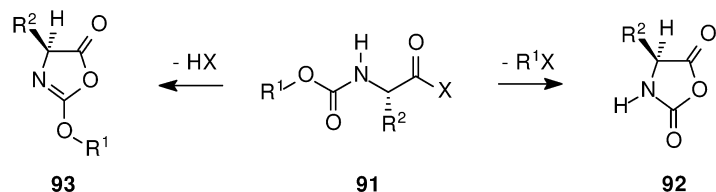


Figure 4.33 Racemization via 5(4H)-oxazolone formation. R^1 = alkyl, aryl or alkoxy residue; R^2 , R^3 = amino acid side chain; B = base; X = activating group.

may easily racemize under coupling conditions in a base-catalyzed reaction via the aromatic intermediate **90**, giving rise to a mixture of both enantiomers **L-88** and **D-88**. In the past, oxazolone formation was not supposed to occur with N-alkoxycarbonyl-protected amino acid derivatives. If at all, activated N-alkoxycarbonyl amino acids **91** (Scheme 4.11) were believed to undergo cyclization giving N-carboxy anhydrides **92**, but not to form 2-alkoxy-5(4H)-oxazolone (**93**). The synthetic tactics of peptide chain assembly starting from the C-terminus using urethane-type protecting groups,



Scheme 4.11

which is regarded to be devoid of racemization, relies on this assumption. However, it was shown in 1977 that 2-alkoxy-5(4*H*)-oxazolones may in fact be formed upon activation of the carboxy group [260, 261, 349, 350].

These derivatives seem to be much less prone to racemization than their 2-alkyl or 2-aryl counterparts, and also undergo aminolysis much more rapidly [256, 348]. Although the reasons for that behavior have not yet been fully clarified, the decreased acidity of a urethane-protected NH group (R–O–CO–NH) compared to an amide group (R–CO–NH) seems to be a crucial factor. As depicted in Figure 4.33, the base abstracts the amide proton in a rapid reaction step. The rate-determining ring closure is initiated by nucleophilic attack of the oxygen atom of the anionic amide group on the activated carboxy group (**89**). Hence, a decrease in the NH acidity – as is the case in the urethane-protected derivatives – lowers the rate of oxazolone formation. Preactivation of amino acids with coupling reagents is sometimes accompanied by racemization, and hence should be avoided when, for example, racemization-prone building blocks are to be incorporated.

Despite their tendency to racemize, the oxazolones still represent carboxy-activated amino acid derivatives and so are incorporated into the peptide. The degree of racemization observed after peptide coupling reactions depends on both the propensity for oxazolone formation and the rate of oxazolone aminolysis. The nucleophilicity/basicity relationship of the amino component is also a crucial parameter.

Asymmetric induction may also be involved in this kind of epimerization process. In 1966, Weygand *et al.* reported that so-called positive diastereomers (L,L-peptides) are predominantly formed upon reaction of racemic 5(4*H*)-oxazolones with L-configured amino acid esters [351]. This phenomenon relies on the reaction of a configurationally labile molecule present in two enantiomeric forms L-**88**, D-**88** with a chiral compound and may, in a positive sense, suppress or in a negative sense, enhance racemization in peptide synthesis. Consequently, the degree of epimerized product observed after a peptide synthesis is the result of both the racemization according to the oxazolone mechanism and asymmetric induction.

4.4.3

Racemization Tests: Stereochemical Product Analysis [352]

Determination of the degree of racemization occurring in one discrete coupling step, and evaluation of the diastereomeric purity of peptide products, is a necessary measure because of the imminent risk of racemization. Subsequent to the proof of sequence homogeneity of the product using different analytical techniques (e.g., high-performance liquid chromatography (HPLC), capillary electrophoresis, mass spectrometry), the absence of undesired diastereomers in the final product has to be proven. This may be done by complete proteolytic degradation, as reported by Finn and Hofmann [353]. Total hydrolysis of the peptide, followed by chiral resolution of all amino acid components with analytical techniques is a viable alternative, but is impeded by the uncertainty of partial epimerization during hydrolysis. The application of isotope-labeled hydrolytic agents (deuterium- or tritium-labeled HCl) circumvents this disadvantage [347, 354]. The configurational integrity of amino acids

may be determined by incubation with L-amino acid oxidase, the transformation into diastereomeric dipeptides, or gas-liquid chromatography (GLC) on enantiomerically pure stationary phases.

Information on the degree of racemization during a coupling reaction may be obtained by the synthesis of model peptides, applying either an activated N-acyl amino acid or an activated N-protected peptide. Sometimes, the additional chiral center of isoleucine is utilized in a simple test method, as the D-allo-isoleucine residue formed by racemization at the isoleucine C α is a diastereomer of L-isoleucine and hence can easily be separated. The variety of methods described has been summarized in review articles [347, 350]. Despite the value of such model systems [313, 355–357] in the development of new coupling methods, a generalization of the results is commonly not possible, because racemization is also sequence-dependent. A HPLC-based racemization test utilizing fluorescence detection has been published by Griehl *et al.* (Figure 4.34) [356]. In this investigation, the coupling of Z-Ala-Trp-OH to H-Val-OMe was examined using different coupling methods. The diastereomeric tripeptides L,L,L-Z-Ala-Trp-Val-OMe and L,D,L-Z-Ala-Trp-Val-OMe were separated by reversed-phase (RP) HPLC. Fluorimetric detection provides 100 times higher sensitivity compared to the UV-detection commonly used.

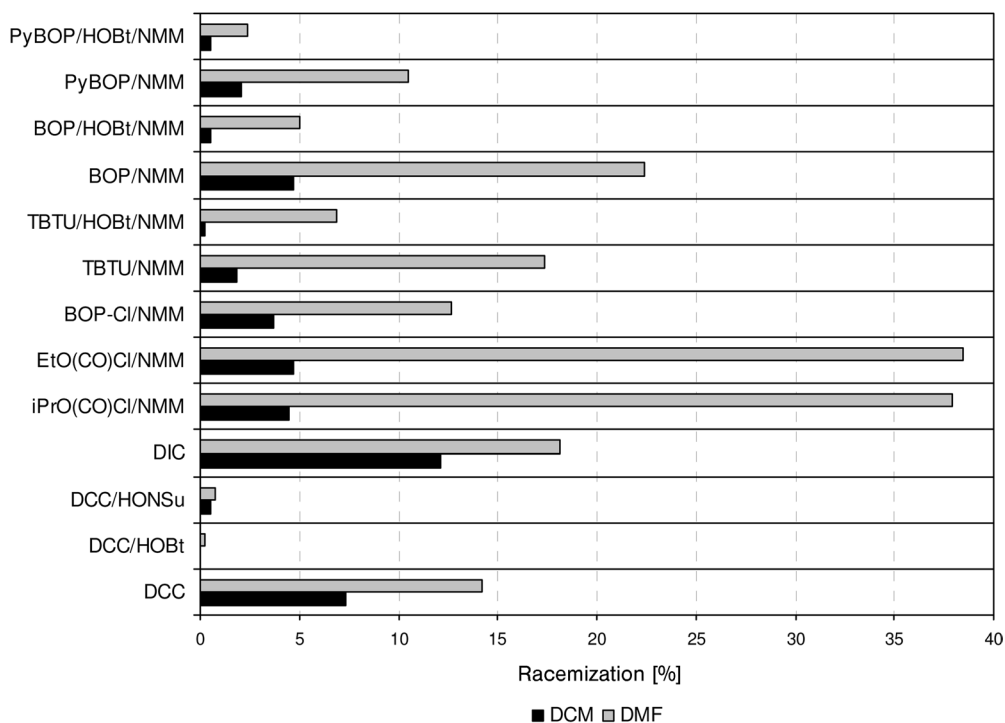


Figure 4.34 Percentage of racemization upon coupling of Z-Ala-Trp-OH to H-Val-OMe with different coupling reactions.

In summary, the problem of racemization/epimerization still impedes the chemical synthesis of peptides.

4.5

Solid-Phase Peptide Synthesis (SPPS)

Ever since the early days of peptide chemistry, reactions involving peptide synthesis have been traditionally performed in solution, and many impressive results – including the total synthesis of small proteins – have been achieved using this conventional technique. In 1981, ribonuclease A – which consists of 124 amino acids – was synthesized by Yajima and Fujii [358], obtained in crystalline form, and shown to have a high specific activity. Unfortunately, peptide synthesis in solution is highly labor-intensive and requires extensive knowledge with regard to the strategy and tactics in choosing protecting groups and coupling methods, as well as solving problems of solubility.

One major advantage of solution-based synthesis is the high purity of the final product, though this is clearly dependent upon the purification of intermediate compounds. Despite these restrictions, highly specialized and experienced research teams have successfully completed many ambitious syntheses following this technique.

The ingenious concept of peptide synthesis on a solid support – which is now known as SPPS – was developed by Robert Bruce Merrifield in 1963 [359], and provided a major breakthrough in peptide chemistry. Merrifield was awarded the Nobel prize in 1984 for this unique invention that has revolutionized organic chemistry during the past 20 years. Today, the concept has been extended not only to peptides, oligonucleotides and oligosaccharides [360] but has been already generalized to organic synthesis on polymeric supports, which includes not only heterogeneous reactions involving an insoluble polymer, but also the application of soluble polymeric materials which allow homogeneous reactions (liquid-phase peptide synthesis) to be conducted.

In SPPS the peptide chain is assembled in the usual manner, starting from the C-terminus. The amazingly simple concept is that the first amino acid of the peptide to be synthesized is connected via its carboxy group to an insoluble polymer that may be easily separated from either reagents or dissolved products by the use of filtration. The general principle is shown in Figure 4.35.

A necessary prerequisite is that anchoring groups (linkers) are introduced into the polymeric material (Figure 4.35, step 1). An amino acid which is protected at N^α is then reacted with the functional group of the linker (step 2). Subsequently, the temporary protecting group is removed (step 3) and the next amino acid component is coupled (step 4). Steps 3 and 4 are then repeated (step 5) until the required peptide sequence has been assembled. Finally, the covalent bond between the linker moiety and the peptide chain is cleaved. In many cases the semipermanent side chain-protecting groups may be simultaneously removed (step 6). The insoluble polymeric support is then separated from the dissolved product by filtration.

When syntheses are carried out on a polymeric support there is no need to perform the tedious and time-consuming isolation and purification of all

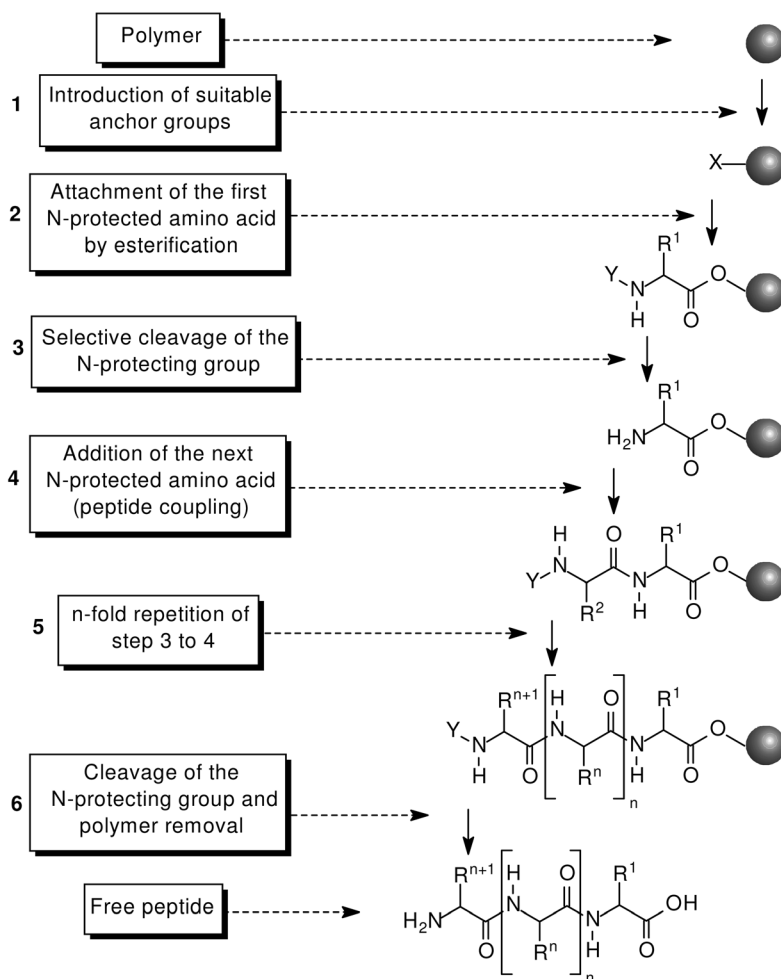


Figure 4.35 Process of solid-phase peptide synthesis. Y = temporary protecting group; R¹, R², Rⁿ, Rⁿ⁺¹ = amino acid side chains, if necessary, protected with semipermanent protecting group.

intermediates, as would be necessary in solution synthesis. The product of all the reactions (the growing peptide chain) remains bound to the support during steps 3 to 5, and excess reagents and by-products are removed by filtration. The simple technical operation and the potential for automation initially led to the euphoric conclusion that the chemical synthesis of polypeptides and proteins had, in principle, been solved by the solid-phase concept. Clearly, the Merrifield synthesis had a major influence on chemical peptide and protein synthesis as well as on solid-phase organic synthesis. The concept of chemical synthesis on a polymeric support also promoted the chemical synthesis of many other biomolecules, including oligo- and polynucleotides.

Unfortunately, despite the high expectations that were initially imposed on the Merrifield method, it does suffer from several limitations:

- The final product of a synthesis carried out on a polymeric support is only a homogeneous compound if all deprotection and coupling steps proceed quantitatively.
- A large excess of each amino acid component is required in the corresponding coupling reaction in order to achieve complete conversion.
- There is a permanent risk of undesirable side reactions during activation, coupling, and deprotection.
- Monitoring the reaction progress and analysis of complete conversion are difficult to perform in heterogeneous reaction systems, and are hampered by experimental error.
- Swelling properties of the polymeric resin and diffusion of the reagents are important parameters for the success of a solid-phase synthesis.
- Aggregation phenomena of the growing peptide chain may complicate the synthesis.
- On occasion, drastic conditions required to cleave the peptide from the polymer may also damage the final product.

Although all methods described for the synthesis of peptides and organic molecules on a polymeric support rely on the basic principle introduced by Merrifield, the expression “Merrifield synthesis” has mainly been coined for SPPS. More specialized information is available in a series of review articles and monographs [361–368].

4.5.1

Solid Supports and Linker Systems

The polymer support plays a decisive role in the outcome of any SPPS. The polymeric support must inevitably be chemically inert, mechanically stable, completely insoluble in the solvents used, and easily separated by filtration. It must contain a sufficient number of reactive sites where the first amino acid of the peptide chain to be synthesized can be attached. Interactions between the peptide chains bound to the resin should be minimal. As with a carboxy-protecting group, the resin must be stable under the chemical conditions of the synthetic procedures, and cleavage under mild conditions must be possible. Furthermore, it has to display mechanical stability in order for the necessary separation procedures to be carried out.

Initially, a copolymer of polystyrene with 1–2% divinyl benzene as cross-linker was used in SPPS. The dry resin beads (Figure 4.36) are normally 20–80 μm in diameter, and are able to swell to five- or six-fold volume in the different organic solvents used for peptide synthesis (e.g., dichloromethane; Table 4.6). Consequently, the polymeric support – when suspended in these solvents – is not a static solid matrix but a well-solvated gel with mobile polymeric chains. This facilitates diffusional access of the reagents to all reaction sites. Polystyrene with a lower degree of cross-linking (1% divinylbenzene rather than 2%) was found to have a superior performance, mainly because of the improved swelling properties.

Table 4.6 Swelling properties (based on dry volume) of polystyrene/1% divinylbenzene resin beads in different solvents.

Solvent	Swelling factor	Solvent	Swelling factor
Tetrahydrofuran	5.5	<i>N,N</i> -Dimethyl acetamide	3.4
Dichloromethane	5.1	Diethylether	2.5
Dioxan	4.6	Acetonitrile	2.0
Toluene	4.5	Ethanol	1.05
<i>N,N</i> -Dimethyl formamide	3.5	Methanol	0.95

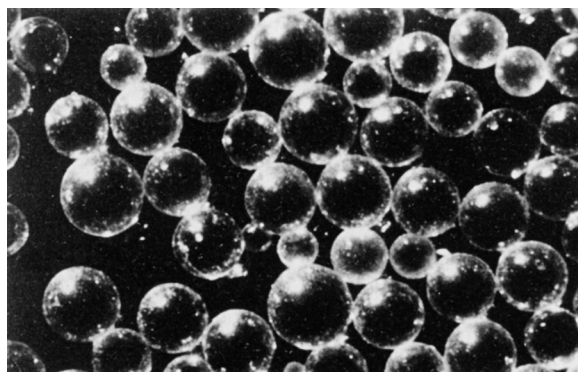


Figure 4.36 Polystyrene/divinylbenzene cross-linked resin.

By using autoradiography with tritium-labeled building blocks, it has been shown that the peptides are distributed uniformly within the gel matrix (Figure 4.37). This finding contradicts the previous assumption that the reactions proceeded only on the surface of the resin beads.

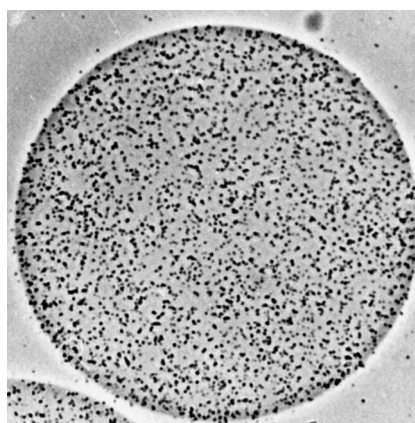


Figure 4.37 Distribution of ^3H -labeled peptides in a single polystyrene/divinylbenzene bead, as revealed by autoradiography.

Approximately 10^{12} polypeptide chains can be located on one polystyrene/divinylbenzene bead of 50 μm diameter, and with a resin loading of 0.3 mmol peptide per gram resin.

The heterogeneous reactions used in SPPS are usually two- or three-fold slower than homogeneous reactions. The reaction sites are uniformly distributed within the polymeric matrix, and the reaction rates of the early coupling and deprotection steps are usually comparable with those at the end of a synthesis. An optimum degree of resin loading with the first amino acid ranges from 0.2 to 0.5 mmol per gram resin. On the basis of this assumption, it can be calculated that the peptide, after coupling of 15 amino acids, contributes more than 50% of the total mass. The growing mass ratio of peptide/resin material does not usually reduce the efficiency of further chain elongation, although the swelling behavior in nonpolar solvents is significantly reduced. The addition of higher percentages of DMF or urea is reported to be advantageous in such cases. Support materials that are more polar (e.g., polyamide-based resins) have also been recommended for the synthesis of longer peptides.

The classical Merrifield resin complies well with the requirements for suitable polymeric support materials, but it is not optimal with respect to mechanical stability, loading capacity, diffusional problems, and differences in solvation between the polymer and the peptide.

Cross-linked poly(dimethylacrylamide) was introduced as a more hydrophilic support by Atherton and Sheppard [369]. This forms a gel and is highly solvated by solvents appropriate for peptide synthesis (e.g., DMF) which favors peptide synthetic reactions. The swelling properties of this polymer are superior to those of polystyrene-based resins, and this allows improved accessibility of the functional groups present on the polymer matrix, even in the micropores. This type of support has also been used in further developments such as polymerization in the pores of a macroporous rigid silica (kieselguhr) matrix, providing improved mechanical stability. Hence, syntheses can be performed under moderate pressure in a column, while the reaction progress can be monitored continuously by spectrophotometry (see Figure 4.41).

Hydrophilic tentacle polymers are obtained by grafting polyethylene glycol (PEG) chains with an arbitrary degree of polymerization onto polystyrene beads. Grafted copolymers using porous polystyrene are best suited for this purpose [370], with the high solvation of the PEG chains conferring high mobility on the peptides bound to the copolymer. The term “tentacle polymer gel” (TentaGel) was coined for these materials, which allow solid-state NMR techniques such as magic angle spinning (MAS) ^1H -NMR or even solution-phase ^{13}C -NMR techniques to be used to monitor reaction progress. Different types of conformation have been found in relaxation time measurements that distinguish between solvated tentacles and nonsolvated random coils. While the latter are characterized by broad lines, the former give rise to sharp NMR absorptions, even though they are present in the solid state. The swelling factors in different solvents are quite similar; therefore, these polymers are stable to moderate pressure and may be used in columns for continuous-flow peptide synthesis.

Polyethylene-based materials are finding increasing application in SPPS, e.g., amphiphilic resins such as PEG-PS and other cross-linked resins entirely based on PEG, especially for the synthesis of difficult or long peptides [371]. Cross-linked ethoxylate acrylate resin (CLEAR) is another type of PEG-based support material. Recently, paramagnetic versions of the CLEAR supports were synthesized by entrapment of magnetite particles with the aim being to use magnetic CLEAR beads in the purification, separation, and analysis of target molecules from complex mixtures [372].

Other types of polymers have been tested for applications in synthesis. The polystyrene resin itself has been chemically modified, but structurally different polymers have also been studied. Non-swelling, highly cross-linked polystyrene/divinylbenzene matrices are rigid and have a high inner surface area; pellicular and brush-type resins have also been developed. While the former are obtained by the polymerization of a thin polymeric layer on the surface of inert glass beads, the latter consist of a linear polystyrene brush polymer grafted onto a polyethylene film surface. These present the anchoring linker moieties in the form of a brush on the surface. Polymers in the shape of strips, films, and fibers have also been tested in automated syntheses.

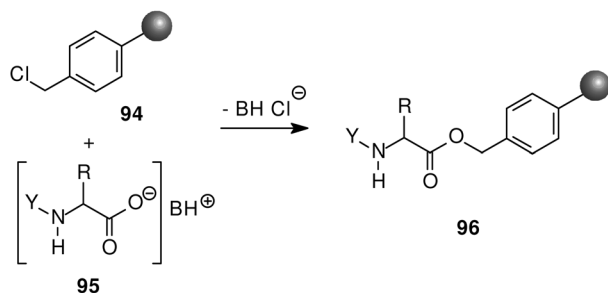
Sintered polyethylene, cellulose, silica, controlled pore glass (CPG), and chitin have also been used as planar support materials.

Nowadays, in line with the growing development of combinatorial chemistry and solid-phase bioassays [373], further requirements must be imposed on resin materials for synthetic purposes. Among these are biocompatibility, good swelling properties in either aqueous or buffered aqueous solutions, and a reduced nonspecific binding affinity to biomolecules.

The introduction of anchoring groups (linkers) on a polymeric support is the precondition for application in peptide synthesis, and the many different linker systems currently available for SPPS and solid-phase organic synthesis [374]. Additional handles in the form of bifunctional linker moieties may be attached to the polymer matrix. One of the functional groups fulfils the requirements for a protecting group allowing for cleavage under mild conditions, while the other is used for attachment to the resin. This concept provides improved control of the level of resin loading, and also allows greater freedom in the choice of polymer matrix type (e.g., polystyrene, polyacrylamides, CPG, chitin, cellulose). Handles often contain internal reference amino acids (IRAA); these improve the monitoring of reaction progress, provide an exact yield determination, and help to control the integrity of the resin-bound peptide chain [369].

Because linkers must be considered as protecting groups for the C-terminal carboxy group of the peptide to be synthesized, their design and implementation must consider especially their chemical properties. Depending on whether the C-terminus of the desired peptide is required as a carboxylate, a carboxamide, a hydrazide (for backing-off strategy), an ester, a thioester [375], or an alcohol, the linker system must be chosen appropriately. In most cases, the first amino acid is connected to the resin via an ester bond, though amide or hydrazide bonds may also be involved.

Correct selection of the handle allows yield optimization with respect to the attachment and cleavage steps, as well as suppression of racemization during attachment of the first amino acid and diketopiperazine formation on the dipeptide



Scheme 4.12

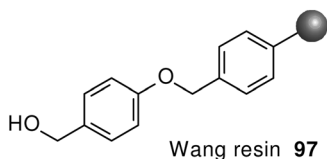
stage. For certain applications (e.g., peptide backbone cyclizations) the linker must behave orthogonally to the other semipermanent and temporary protecting groups.

Meanwhile, many different variants for the chemical functionalization of polymers have been identified, and the number is steadily increasing as the solid-phase synthesis of peptides and organic molecules undergoes rapid development.

Chloromethyl resin (Merrifield resin) [376, 377] The chloromethyl group in **94** (Scheme 4.12) is the classical anchoring moiety present in the Merrifield resin, and it is introduced into polystyrene/divinylbenzene resins by Friedel-Crafts-type chloromethylation with an alkoxy-substituted chloromethane in the presence of tin(IV) chloride. Attachment of the first amino acid to the resin to give **96** is performed as a nucleophilic substitution reaction of chloride by the amino acid carboxylate **95** under reflux conditions in an organic solvent such as ethanol, tetrahydrofuran, or dioxan. Often, cesium salts or alkyl ammonium salts are used. Other bases such as tertiary amines or anhydrous potassium fluoride may also be applied.

This protocol prevents racemization that might otherwise occur with an activated amino acid derivative. Therefore, the chloromethyl resin is preferred over the corresponding hydroxymethyl resin. The resulting benzyl-type peptidyl ester can be cleaved on completion of the peptide chain assembly only with very strong acids such as liquid HF, TFMSA, or HBr/TFA. Consequently, this type of resin is used in combination with Boc as a temporary protecting group and semipermanent TFA-stable side chain protection of the benzyl or cyclohexyl type to obtain peptides with a free carboxy group at the C-terminus (peptide acids). One disadvantage of the Merrifield resin is that 1–2% of the growing peptide is cleaved from the resin during each of the repetitive acidolytic deprotection steps.

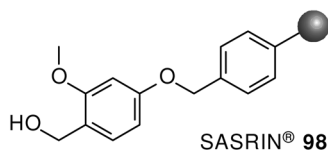
4-Benzyloxybenzyl Alcohol Resin (Wang Resin) [378] The Merrifield resin can easily be converted into the Wang resin **97** by etherification with methyl 4-hydroxybenzoate, followed by reduction of the methyl ester with LiAlH_4 .



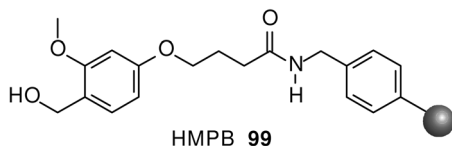
This anchoring group comprises a 4-alkoxy-substituted benzyl alcohol moiety, which confers increased acid-lability onto the linker. Hence, the Wang resin is used routinely in batch Fmoc chemistry.

Active esters (e.g., pentafluorophenyl esters) or carbodiimides (e.g., DCC or carbonyldiimidazol) are used for the direct attachment of the first amino acid on the resin. 4-(Dimethylamino)pyridine is often added as an acylation catalyst in carbodiimide couplings, but in these cases racemization may occur and the addition of HOBt is recommended. Recently, a new method for racemization-free loading of Fmoc-amino acids to Wang resin has been described [379]. Cleavage of the final product proceeds smoothly with 95% TFA, and yields peptide acids with concomitant removal of *tert*-butyl-type side chain-protecting groups.

2,4-Dialkoxybenzyl Alcohol Resin (Super acid-Sensitive Resin, SASRIN®) [380] The introduction of one more methoxy group in the 2-position of the benzyl alcohol moiety (**98**) results in a further increased sensitivity towards acid, because the benzyl cation intermediate of acidolysis is highly stabilized by both electron-donating alkoxy groups in the 2- and 4-positions. Cleavage occurs even in the presence of 0.5–1% TFA in dichloromethane, or by treatment with the acidic alcohol 1,1,1,3,3,3-hexafluoroisopropanol in dichloromethane. This property makes the resin very well suited for the synthesis of peptide acids that are fully protected at the side chains. Attachment of the first amino acid component is performed analogously, as discussed for the Wang resin. Racemization during this esterification may be a severe problem.



4-(4-Hydroxymethyl-3-Methoxyphenoxy)butyric Acid Resin (HMPB Resin) [381] The HMPB handle type **99** is chemically related to the 2,4-dialkoxybenzyl alcohol resin, and displays similar cleavage characteristics. Detachment of the final product, even in the form of fully protected peptides, is achieved with dilute acid (1% TFA/dichloromethane) to yield peptide acids.



4-(Hydroxymethyl)phenylacetamidomethyl Resin (PAM Resin) [382] The PAM resin (Figure 4.38) was developed as an alternative to the Merrifield resin, where loss of peptides during acidolytic Boc deprotections and insufficient reactivity in the nucleophilic displacement of the linker benzyl chloride by the first amino acid carboxylate is sometimes observed.

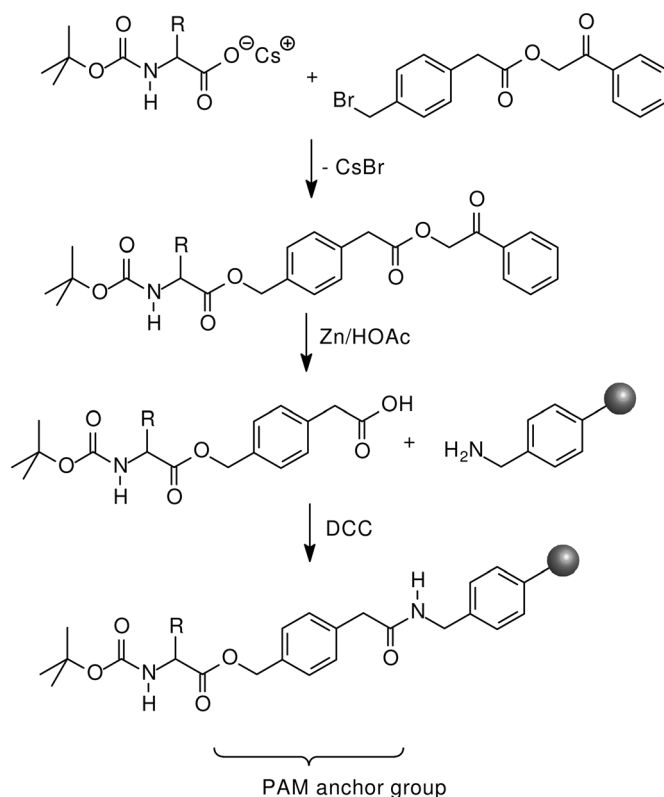


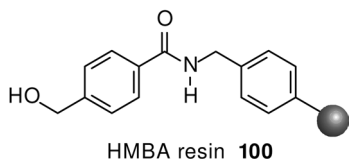
Figure 4.38 Coupling of the PAM linker, loaded with the first Boc-protected amino acid, to aminomethylated polystyrene. PAM = phenylacetamidomethyl.

Aminomethyl groups can also be attached to the aromatic moieties of polystyrene. The electron-withdrawing carboxamidomethyl group renders the PAM resin about 100 times more stable towards treatment with TFA. It is introduced by acylation of aminomethyl polystyrene with (4-hydroxymethylphenyl)acetic acid. The first amino acid is introduced by carbodiimide or active ester chemistry forming a resin-bound ester.

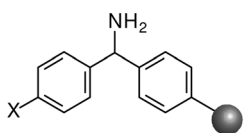
Alternatively, a PAM anchoring group, where the first amino acid residue is already attached to it, may be introduced by the reaction sequence displayed in Figure 4.38 (pre-formed handle strategy). The bromide substituent in 4-(bromomethyl)phenylacetic acid phenacyl ester undergoes nucleophilic displacement upon treatment with the cesium carboxylate of the corresponding amino acid. Subsequently, the phenacyl ester is cleaved by reduction with zinc in acetic acid and the loaded PAM anchoring group is attached to aminomethyl polystyrene.

The PAM resin is fully compatible with the Boc/Bzl protection scheme, and is cleaved with strong acids such as liquid HF, TFMSA, or HBr/TFA to give peptides with a free acid located C-terminally.

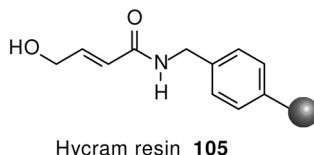
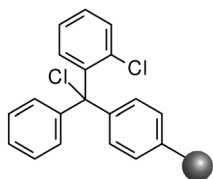
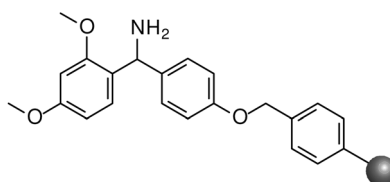
Hydroxymethylbenzoic Acid Resin (HMBA Resin) [383] The hydroxymethylbenzoic acid resin **100** is closely related to the PAM resin. The only difference is the missing benzylic CH_2 group, but this results in a completely altered chemical reactivity. It is obtained by acylation of aminomethyl polystyrene with 4-(hydroxymethyl)benzoic acid and is completely resistant towards treatment with acids (even liquid HF). Consequently, on-resin side-chain deprotection can be accomplished. The first amino acid is introduced using carbodiimide or active ester chemistry. The highly versatile HMBA linker is cleaved by a variety of nucleophiles such as hydroxide ions (to give peptide acids), alcohols (to give peptide esters), ammonia or amines (to give peptide amides), hydrazine (to give peptide hydrazides), or LiBH_4 (to give peptide alcohols).



Benzhydrylamine Resin (BHA Resin) and 4-Methylbenzhydrylamine Resin (MBHA Resin) [384, 385] The acid lability of a benzyl ester can be increased not only by the introduction of additional electron-donating groups on the aromatic ring, but also by the attachment of one (BHA resin **101**, MBHA resin **102**, Rink amide resin **103**) or two additional aromatic rings (Barlos resin **104**) to the benzylic carbon atom. The benzhydrylamine resin (BHA resin) was designed accordingly. Acidolysis with HF, TFMSA, or HBF_4/TFA yields the peptide amide. This fact renders the BHA resin suitable for the Boc/Bzl protection scheme. Increased acid lability is observed after introduction of an electron-donating methyl substituent in the 4-position of the second aromatic ring (MBHA resin **102**).



MBHA resin X = CH_3 **102**

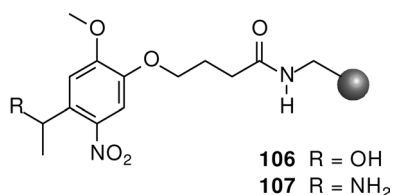


(2,4-Dimethoxy)benzhydrylamine Resin (Rink amide Resin) [386] The combination of a benzhydrylamine group with two additional methoxy substituents on the second aromatic ring renders the (2,4-dimethoxy)benzhydrylamine resin **103** even more labile towards acidolysis, and allows cleavage of peptide amides with 95% TFA.

***o*-Chlorotritylchloride Resin (Barlos Resin)** [387] Attachment of the first amino acid to the *o*-chlorotritylchloride resin **104** is accomplished upon reaction with an amino acid carboxylate salt (e.g., diisopropylethylammonium or triethylammonium salts). This procedure does not suffer from racemization, as the amino acid acts as the nucleophile and not as an electrophilic species. The steric constraints of the *o*-chlorotrityl group impedes diketopiperazine formation on the dipeptide stage [388]. The trityl group has already been discussed in the context of side chain-protecting groups (cf. Section 4.2.4). *o*-Chloro substitution slightly increases stability towards acids, but the *o*-chlorotrityl linker remains very sensitive to acid. Cleavage occurs upon treatment with 0.5% TFA in dichloromethane or with 1,1,1,3,3,3-hexafluoroisopropanol in dichloromethane. The *o*-chlorotrityl handle is also suitable for the attachment of C-terminal alcohols, thiols, and amines.

Hydroxycrotonoyl Aminomethyl Resin (Hycram Resin) [389] Allyl esters are completely orthogonal to other types of protecting groups, because they are completely stable towards acids or bases usually employed in peptide synthesis. Consequently, an allyl-based handle represents a useful alternative to other linker types. The Hycram resin **105**, loaded with the first amino acid component is obtained upon acylation of aminomethyl polystyrene with esters of, for example, Boc- or Fmoc-protected 4-hydroxycrotonyl derivatives. Cleavage of the final product from the resin is accomplished under neutral conditions by treatment with a Pd(0) catalyst that effects allyl transfer to a suitable nucleophile. Fully protected peptides or very sensitive derivatives (e.g., glycopeptides) can be synthesized according to this concept.

Photolabile Linkers Photolysis of the peptide-resin bond provides an additional dimension in orthogonality. Similarly to photolabile protecting groups, linker moieties (e.g., **106**, **107**) are available that allow cleavage by photolysis [390]. The 4-bromomethyl-3-nitrobenzoic acid handle is incorporated in the Nbb resin that is fully compatible with Boc/Bzl SPPS and partly compatible with the Fmoc/tBu SPPS (see Chapter 5, Scheme 5.1).



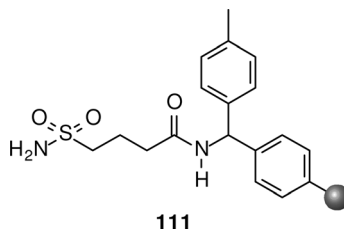
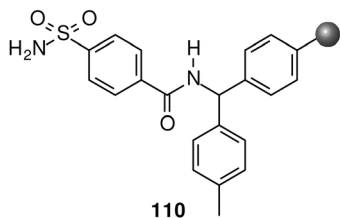
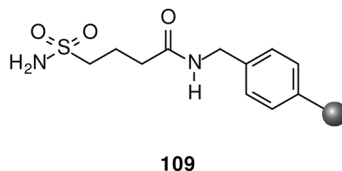
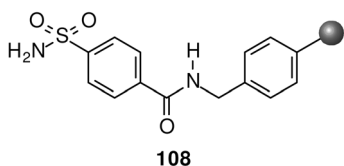
4.5.2

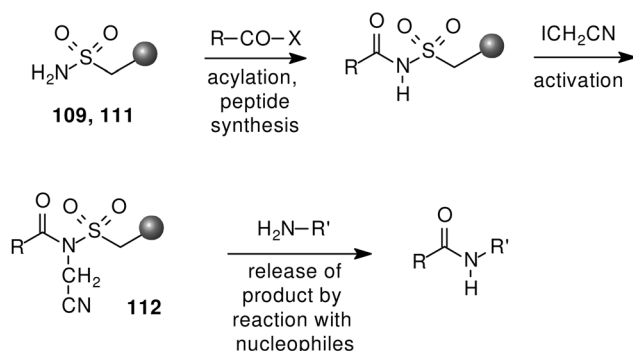
Safety-Catch Linkers

Safety-catch linker strategies [374, 391, 392] rely on a special linker moiety that connects the polymeric support and the growing peptide chain. This linker is stable under the conditions of peptide synthesis but is finally activated for cleavage by a discrete chemical modification. In addition, the activation step usually provides an activated C-terminus of the peptide which may react with diverse nucleophiles. Intramolecular reaction of this activated position with the free N-terminus of the peptide leads to cyclic peptides and simultaneously liberates the target compound from the solid support. Two major types of safety-catch linkers may be distinguished:

- Those characterized by a hidden orthogonality of the stable and labile form, which means that cleavage mechanisms of the two forms are different.
- Those which rely on different kinetics of the cleavage reaction in both forms.

Kenner's sulfonamide safety-catch linker [392, 393] tethers a carboxylic acid to a solid support (e.g., aminomethyl polystyrene, **108**, **109** or MHBA, **110**, **111**), which is stable towards acidic and basic reaction conditions. Alkanesulfonamide-type linkers (**109**, **111**) are superior to arenesulfonamide linkers (**108**, **110**) because of the greater nucleophilicity of the alkanesulfonamide nitrogen [394]. Coupling conditions have been developed to load Boc- or Fmoc-protected amino acids with high efficiency and minimal racemization. After acylation of the alkanesulfonamide-type linkers (**109**, **111**) and activation, preferably by treatment with iodoacetoneitrile, the secondary sulfonamide **112** can be cleaved by aminolysis, peptidolysis, hydrazinolysis, or saponification to provide the primary amides, peptides, hydrazides, and carboxylic acids, respectively [393, 392] (Scheme 4.13). Backbone cyclizations by simultaneous cleavage and ring closure are also feasible (Section 6.1.1) [392].





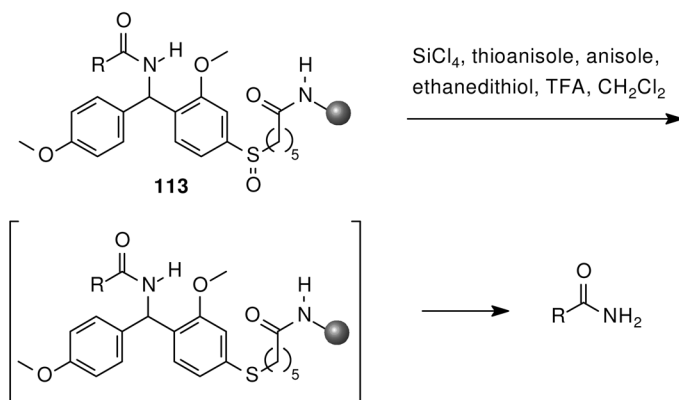
Scheme 4.13

In contrast to diazomethane, the application of iodoacetone nitrile as alkylating agent led to a significant improvement because the *N*-cyanomethyl *N*-acylsulfonamide **112** is extremely reactive in nucleophilic displacement reactions under mild conditions.

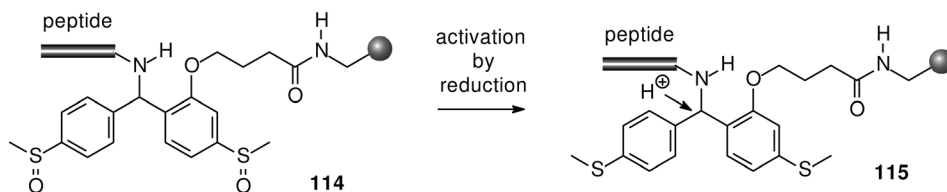
Near-racemization-free displacement of the activated linker moiety with various nucleophiles has been demonstrated. Even C-terminal thioesters, that may be subsequently used for ligation purposes, can be obtained on thiolysis of peptides bound to a safety-catch linker [392, 395].

Unverzagt *et al.* developed an orthogonal combination of the sulfonamide safety-catch linker with an acidolytically cleavable Rink-amide linker. The latter allows facile detachment of the peptide chain for step-by-step mass spectrometric synthesis monitoring [396].

Safety-catch linkers relying on oxidative activation have been employed in the synthesis of linear and cyclic peptides [397] but their application is limited to the synthesis of peptides that do not contain any oxidation-sensitive amino acid. The sulfoxide-containing safety-catch linker DSA **113** [398] resists treatment with acid but undergoes acidolysis after reduction of the sulfoxide group; this generates a third electron-releasing group on the aromatic system (Scheme 4.14).



Scheme 4.14



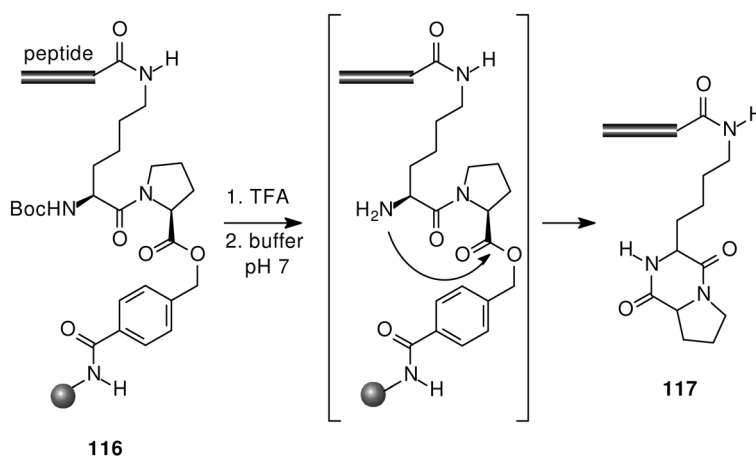
Scheme 4.15

The DSA linker is structurally related to the safety-catch amide linker (SCAL) developed by Lebl [399]. In the oxidized form **114**, SCAL is stable to a wide range of chemical conditions (TFA, HF, thiophenol, reducing agents, basic conditions, and Hg^{2+}). Upon reduction of the two aryl sulfoxide groups to give the corresponding aryl sulfide **115** by $\text{PPh}_3/\text{Me}_3\text{SiCl}/\text{CH}_2\text{Cl}_2$ or $(\text{EtO})_2\text{P}(\text{S})\text{SH}/\text{DMPU}$, SCAL becomes labile towards treatment with TFA (Scheme 4.15). SCAL has also been used for SPPS by chemical ligation [400].

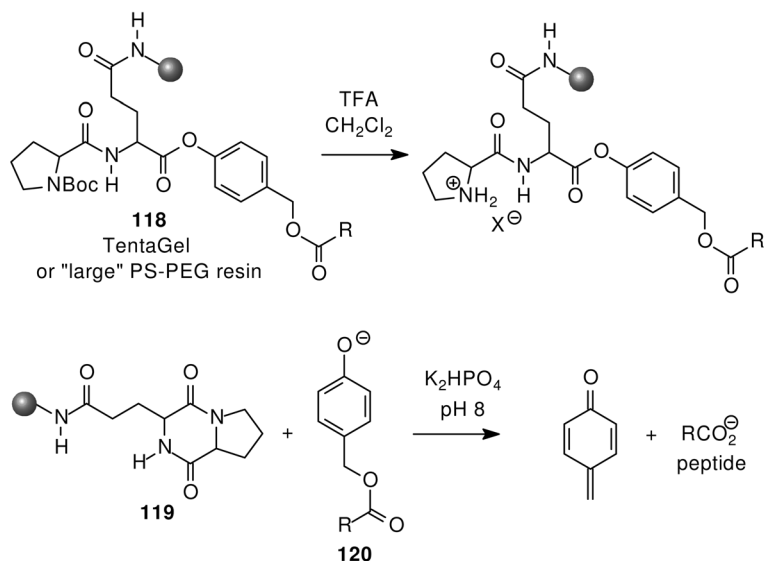
Another concept for a safety-catch linker **116**, developed by Geysen *et al.* [401] relies on the triggered formation of peptide with a C-terminal diketopiperazine **117** that is accompanied by cleavage from the resin (Scheme 4.16).

A similar linker for TentaGel or “large” PS-PEG resin has been designed by Atrash and Bradley [402]; this combines diketopiperazine formation with subsequent fragmentation (Scheme 4.17). Cleavage of the Boc group at proline in **118** triggers diketopiperazine **119** formation, which in turn releases a peptidylloxymethylphenolate **120** that spontaneously undergoes fragmentation, liberating the peptide assembled on the resin as the free carboxylate. These linkers have been shown to be suitable for a direct biological evaluation of a peptide released from the resin *in situ*.

A concept involving linker cleavage by neighboring group participation has been presented by Frank and coworkers [403].



Scheme 4.16



Scheme 4.17

4.5.3

Protection Schemes

Each of the repetitive synthesis cycles (deblocking–coupling) is initiated by selective cleavage of the N^α -protecting group. The anchoring group (linker) and the semipermanent amino acid side chain-protecting groups must be correctly chosen with respect to the chemical properties of the temporary N^α -protecting group, which is usually held constant for all amino acids used throughout the synthesis. The selection and optimization of protecting group chemistry doubtless holds a key position in the success of a peptide synthesis.

In SPPS, as in liquid-phase synthesis, either the acid-labile *tert*-butoxycarbonyl group or the base-labile fluorenyl-9-methoxycarbonyl group are applied preferably as temporary protecting groups (cf. Section 4.2.1.1).

4.5.3.1 Boc/Bzl-protecting Groups Scheme (Merrifield Tactics)

The standard Merrifield system is based on Boc-protecting group tactics, which rely on selective acidolytic cleavage. The *tert*-butoxycarbonyl group is used as the temporary N^α -protecting group in combination with benzyl-type semipermanent side-chain protection, as Boc is usually cleaved with TFA (20–50%). Figure 4.39 shows an example of the protecting group combination for a tripeptide amide synthesis ($\text{H-Asp-Gly-Tyr-NH}_2$) on MBHA resin.

The semipermanent benzyl-type groups must be stable under the conditions of repetitive Boc cleavage. In order to safeguard the stability of the side-chain protection, benzyl-type groups with electron acceptors may be applied (e.g., 2,6-dichlorobenzyl, 2,6-Dcb). Cyclohexyl esters are often applied preferentially for the protection of side-

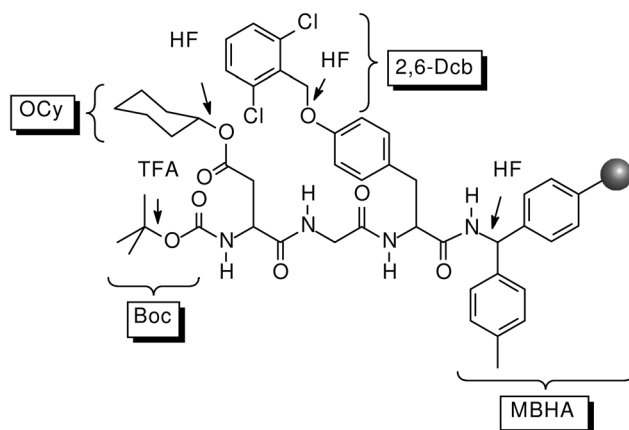


Figure 4.39 Boc/Bzl protecting group tactics in solid-phase peptide synthesis according to Merrifield. MBHA = 4-methylbenzhydrylamine resin.

chain carboxy groups. The semipermanent protecting groups and the MBHA linker group are cleaved simultaneously by treatment with liquid HF on completion of the synthesis, as shown in Figure 4.39. Peptide assembly on the PAM resin would result in the peptide with C-terminal acid under similar cleavage conditions.

As discussed previously, scavengers such as anisole, thioanisole, dimethyl sulfide, or triisopropylsilane must be added to the deprotection reaction in order to avoid side reactions of the intermediate carbenium ions. Other scavengers have been especially recommended in the context of the two-stage, low–high HF cleavage method that usually provides improved product purity [404, 405].

Absolute stability of the side chain-protecting groups is generally required for the success of a peptide synthesis. In cases where N^α -deprotection relies on acidolysis, this is especially critical because cleavage of the temporary and semipermanent protecting groups as well as of the linker proceed mechanistically in a very similar manner. Fine-tuning of the cleavage kinetics is required.

Special attention must be paid to selection of the side-chain protection. As any protected amino acid residue employed in the synthesis must be strictly compatible with the protecting group tactics chosen for the synthesis of a particular peptide, some protecting group schemes are more appropriate than others. The following semipermanent protecting groups are very compatible with Boc/Bzl tactics (final cleavage with liquid HF):

- Arg(Tos), Arg(Mts)
- Asp(OBzl), Asp(OCy), Glu(OBzl), Glu(OCy)
- Cys(Acm), Cys[Bzl(4-Me)], Cys(Mob)
- His(Bom), His(Dnp) – pre-cleavage treatment is necessary, His(Z), His(Tos)
- Lys[Z(2-Cl)]
- Ser(Bzl), Thr(Bzl)
- Trp(For) – pre-cleavage treatment is necessary
- Tyr[Z(2-Br)].

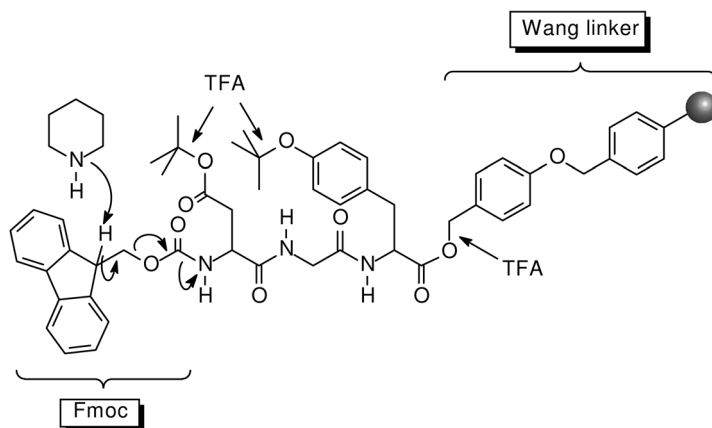


Figure 4.40 Fmoc/tBu protecting group tactics in solid-phase peptide synthesis according to Sheppard.

4.5.3.2 Fmoc/tBu-Protecting Groups Scheme (Sheppard Tactics)

The Fmoc-protecting group tactics make use of the base lability of the fluorenyl-9-methoxycarbonyl group (Fmoc). It is a widely applied alternative to the Boc/Bzl scheme with two-dimensional orthogonality (Figure 4.40) [369]. Fmoc is cleaved by base-catalyzed elimination where the secondary amine (piperidine) also traps the dibenzofulvene initially formed in the reaction. The semipermanent side-chain-protecting groups are mostly of the *tert*-butyl type, and can be cleaved under relatively mild reaction conditions with TFA. Linker moieties displaying comparable acid lability are mainly used.

Although introduced into peptide chemistry as early as 1970, the Fmoc/tBu scheme has been widely applied in SPPS only since 1978. Many different coupling methods are compatible with the Fmoc group, among which are the highly reactive amino acid chlorides and amino acid fluorides [298]. Figure 4.41 shows a schematic view of a semi-automatic continuous-flow peptide synthesizer developed by Atherton and Sheppard [369]. Pressure-stable resin material is required for application in a column reactor.

The following side-chain-protecting groups have been used preferentially in Fmoc tactics:

- Arg(Pmc), Arg(Pbf)
- Asp(OtBu), Glu(OtBu)
- Asn(Trt), Asn(Tmb), Gln(Trt), Gln(Tmb)
- Cys(Trt)
- His(Trt)
- Lys(Boc)
- Ser(tBu), Thr(tBu)
- Trp(Boc)
- Tyr(tBu).

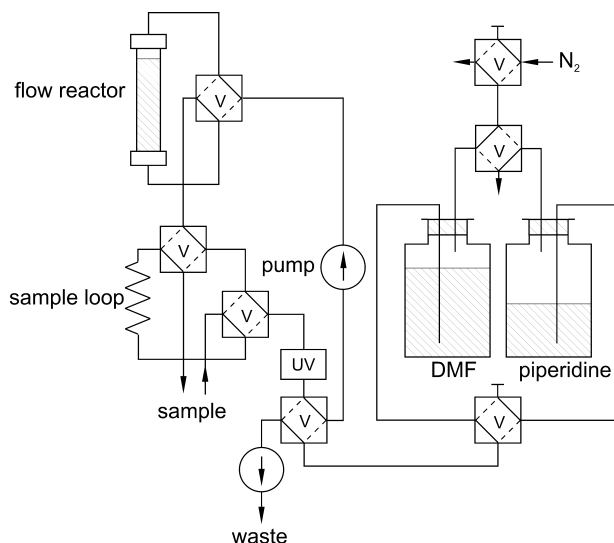


Figure 4.41 Schematic view of a semiautomatic continuous-flow peptide synthesizer according to Atherton and Sheppard [369]. DMF = dimethylformamide; UV = ultraviolet detection; V = valve.

4.5.3.3 Three- and More-Dimensional Orthogonality

As alternatives, completely orthogonal protection schemes are available [1, 361]. These are characterized by a set of protecting groups where deprotection reactions are chemically independent. The third dimension of orthogonality relies on the application of protecting groups or linkers labile to, for example, photolysis, thiolysis (Dts), hydrazinolysis, or cleavage with transition metal catalysts (Alloc) [41, 42].

The acid-stable N^α -dithiasuccinoyl group (Dts) represents an elegant example of three-dimensional orthogonality in protection schemes of SPPS [361, 362], and can be cleaved by thiolysis under neutral reaction conditions. A possible protecting group combination is shown in Figure 4.42.

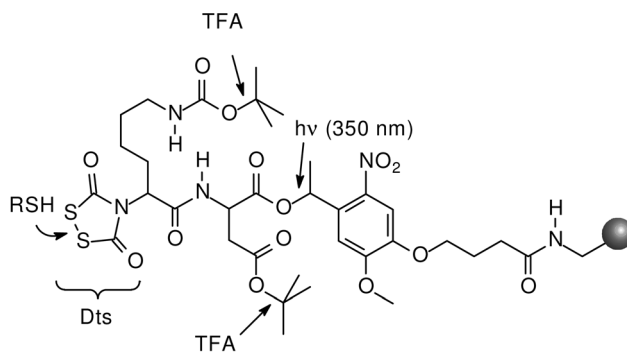


Figure 4.42 Completely orthogonal protecting group tactics (three-dimensional orthogonality) in solid-phase peptide synthesis [362].

While free thiols cleave the intramolecular disulfide bond of the temporary protecting group Dts, the semipermanent side-chain protection is removed by acidolysis. The linker containing a photolabile 2-nitrobenzyl moiety can be cleaved photolytically at 350 nm. The sequence of cleavage reactions can be chosen arbitrarily, and as a consequence fully or partially protected peptide acids can be obtained.

4.5.4

Chain Elongation

4.5.4.1 Coupling Methods

The coupling reaction is a process of eminent importance for the success of a SPPS, because complete conversion is the basic precondition for the formation of a homogeneous final product. Coupling reagents are, therefore, applied in excess (usually three-fold); this increases the conversion but may also give rise to several undesired side reactions. Initially carbodiimides – especially DCC, but also DIC, because of the higher solubility of the corresponding urea – were used preferentially as coupling reagents. Peptide couplings in solid-phase syntheses are usually performed in dichloromethane, sometimes with addition of DMF, and at ambient temperature. The carbodiimide/HOBt method again prevails due to its major advantages, and not only with respect to the detrimental dehydration of Asn or Gln carboxamide side chains. Symmetrical anhydrides likewise proved to be efficient acylating agents. Active esters, especially of the pentafluorophenyl type, have also become very popular. The rather new coupling reagents of the onium-type, such as TBTU, HBTU, BOP, PyBOP, UNCA derivatives, and Fmoc-protected amino acid fluorides encompass high coupling yields in nonpolar solvents and minimize or even exclude racemization and other side reactions.

4.5.4.2 Undesired Problems During Elongation

Incomplete conversion in solid-phase synthesis leads to a mixture of mismatch and core sequences, as shown in Figure 4.43.

Considering the synthesis of a pentapeptide A-B-C-D-E, incomplete coupling may lead to four core sequences and three mismatch sequences. The latter arise when acylation or deprotection are incomplete and one or more amino acid components are skipped in the chain elongation. Mismatch sequences may also be formed when the amino acid sequence is correct, but acylation of nucleophilic side-chain functionalities, for example, occurs after partial deblocking.

The separation of undesired side products from the target peptide on completion of the synthesis is very tedious, and often impossible on a preparative scale. Hence, all possible preventative measures must be considered to safeguard quantitative reaction. Double coupling steps represent one of these measures. The application of large excesses of reagents used may, however, result in other side reactions, such as aminoacyl insertions. If an acylation reaction is found to remain incomplete even after multiple repetition of a single coupling step, capping is a viable alternative. Capping means acylation of the unreacted amino groups with acetic anhydride, *N*-acetyl imidazole, or other acylating agents in order to avoid further chain elongation

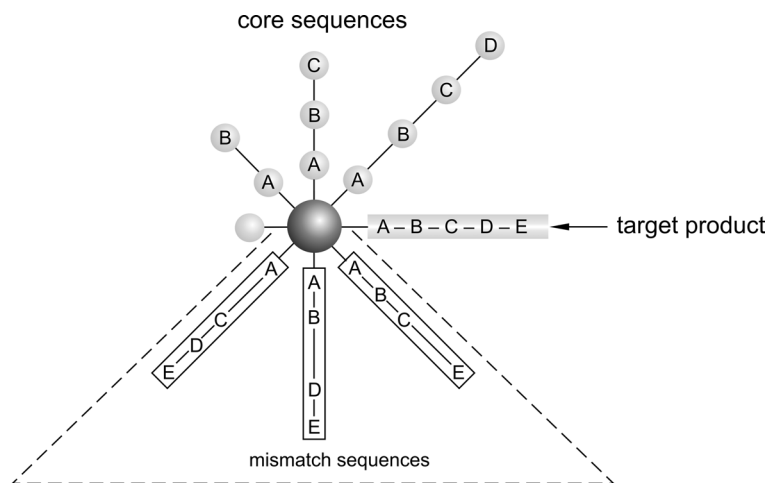


Figure 4.43 Schematic view of side-chain reactions in solid-phase peptide synthesis.

of the mismatch sequence at a later stage. These capped core sequences can usually easily be separated from the final product.

Peptides with a chain-terminating pyroglutamyl residue may be formed during N^α -deprotection of glutamine or glutamate residues located at the N-terminus.

The permanent and irreparable risk of side reactions on all synthetic steps is, despite all advantages, one of the peculiarities of SPPS. Mismatch and incomplete core sequences are caused by incomplete conversions in each of the repetitive steps of the synthesis cycle. If the yield of each cycle is hypothetically assumed to be 99%, the desired product is formed after 10 cycles in a maximum total yield of 90%, while after 100 cycles only 37% of the product is present in the mixture. An individual yield of 95% per cycle causes the total product yield after 10 or 100 cycles to decrease to 60% or 0.6%, respectively. Hence, the desired product may easily be contaminated by a series of structurally and chemically very similar compounds such as diastereomers, formed by epimerization, or mismatch and incomplete core sequences. Isolation and purification of one product from such a microheterogeneous mixture is a very difficult, though not unsolvable, problem. Despite all attempts that have been undertaken during past years to investigate most of the possible side reactions and to elucidate the mechanisms, single-cycle yields exceeding 99.5% do not appear attainable. Consequently, the remaining 0.5% adds up to a cumulative product contamination.

Mismatch and incomplete core sequences may be identified and quantified by a modified Edman degradation, also known as preview analysis [406]. In a successful synthesis, these are present in an amount below 0.3%. Mismatch or even wrong sequences are often the result of harmful or sometimes destructive reaction conditions during the repetitive deblocking steps or the final deprotection reaction.

Furthermore, N^α -Boc derivatives from commercial sources, for example, might in the past have been contaminated with $\sim 0.3\%$ of the corresponding N^α -*sec*-butylox-

ycarbonyl derivative. Once incorporated into the growing peptide chain, these derivatives resist acidolytic deprotection and so terminate the peptide chain.

Fmoc-protected amino acids formerly obtained by the classical methods might contain 1–4% of Fmoc-protected dipeptides as a consequence of competing mixed anhydride activation of the amino acid during the protection step. Incorporation of one or more of these contaminating dipeptides leads to mismatched elongated peptides. Nowadays, most N-protected amino acid derivatives are commercially available with chemical and optical purities >99.5%. Indeed, it is essential that all reagents and solvents, as well as the amino acid building blocks, be of high purity when used for peptide synthesis.

The application of N^α-urethane-type protected amino acids usually maintains the degree of racemization below the analytical limits of the test systems employed (<0.03%). Consequently, in normal cases this process does not significantly impede the outcome of a synthesis. Nevertheless, the esterification of amino acids, for example in the course of an attachment to a hydroxymethyl linker moiety or the activation of the C-terminus of a peptide, may lead to racemization of between 0.2 and 10%.

Diketopiperazine formation may compete efficiently with the coupling of the third amino acid to a resin-bound dipeptide. The diketopiperazine is cleaved from the resin during the attack of the free amino terminus on the C-terminal ester bond. Some amino acids such as glycine, proline, D-amino acids, and N-alkyl amino acids in the second position especially favor this process.

4.5.4.3 Difficult Sequences

One major precondition for complete peptide couplings is diffusion of the acyl component through the matrix, which only occurs to a satisfactory extent if the peptidyl-resin is sufficiently solvated. The most serious problem in SPPS is associated with sequence-related incomplete acylation reactions. In this context, random and nonrandom 'difficult sequences' may be distinguished. Both types are characterized by reproducible incomplete acylation, limited improvement by repetitive coupling or the introduction of capping steps, and even inferior results when sterically hindered amino acids or higher resin loading are applied. Random difficult aminoacylation is associated with the incorporation of sterically hindered amino acids, for example β-branched derivatives or building blocks with bulky side-chain-protecting groups. Nonrandom problems occur due to the sequence-specific formation of stable β-sheets that tend to aggregate and prevent further acylation [407, 408]. Secondary structures stabilized by intermolecular hydrogen bonds impede the diffusion and accessibility of the reactive N-terminus.

Merrifield reported that a SPPS proceeded well up to the eleventh amino acid, but provided low yields at residues 12 to 17 [409]. According to Kent [407], such sequences may occur at a distance of 5–15 residues from the C-terminal residue at the resin. Despite the considerable research carried out in order to understand this phenomenon [410–412], it remains a major obstacle in peptide and protein synthesis.

Nonrandom difficult sequences may be predicted to a certain extent [413]. The propensity of the amino acids to occur in a random coil structure, facilitating synthesis,

Table 4.7 Empirical conformational parameters P_c^* of all amino acids.

Amino acid	P_c^*	Amino acid	P_c^*
Ala	0.75	Leu	0.75
Arg	0.96	Lys	0.90
Asn	1.26	Met	0.70
Asp	1.07	Phe	0.76
Cys	1.08	Pro	1.64
Gln	0.83	Ser	1.21
Glu	0.80	Thr	0.98
Gly	1.47	Trp	0.79
His	0.96	Tyr	0.98
Ile	0.74	Val	0.73

or in ordered secondary structures, hampering acylation, has been expressed by values. The inherent tendency of a peptide chain to form a random coil conformation is then expressed by the average of the single P_c^* values of the n amino acids:

$$\langle P_c^* \rangle = \left(\sum P_c^* \right) / n$$

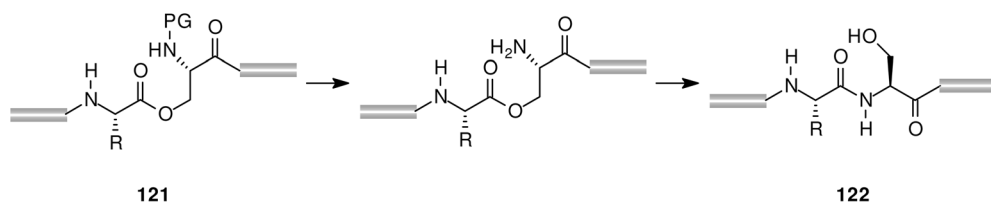
Sequences with $\langle P_c^* \rangle$ values greater than 1.0 are associated with near-quantitative acylation reactions, while sequences with values of 0.9–1.0 require longer reaction times or multiple couplings (Table 4.7). $\langle P_c^* \rangle$ values smaller than 0.9 indicate that persistent acylation problems may occur ('nonrandom difficult sequences'). β -Sheet formation can be monitored using FT-IR spectroscopy [414].

Aggregation is not a severe problem in Boc chemistry, because it is abolished by repetitive treatment with TFA during cleavage of the temporary protecting group.

Fmoc chemistry does not encompass this advantage, but sustains β -sheet formation. This usually leads to prolonged reaction times, and either an increase in the reaction temperature or solvent changes often become necessary; this in turn may favor racemization [256, 415] and premature Fmoc cleavage [416, 417].

The β -sheet aggregates can be disintegrated by the application of:

- solvent additives that break up stable secondary structures, such as 2,2,2-trifluoroethanol [418], 1,1,1,3,3,3-hexafluoroisopropanol [418, 419], DMSO [420], or ethylene carbonate and Triton X [421];
- chaotropic salts in aprotic solvents [422, 423];
- temporary backbone amide-protecting groups (cf. Section 4.2.3) such as *N*-phenylthiomethyl [424], Hmb [97, 98, 425], Hnb [99], SiMb [100], or serine or threonine-derived pseudoproline dipeptide building blocks Fmoc-Xaa-Ser(Ox)-OH and Fmoc-Xaa-Thr(Ox)-OH [101–104];
- HOAt-based coupling reagents;
- polymeric supports based on polyoxyethylene-polystyrene copolymers [426], that diminish interaction between growing peptide chains;
- *in situ* neutralization in the context of a modified Boc/Bzl synthesis scheme [427, 428];



Scheme 4.18

- introduction of a pre-sequence [429];
- sonication [430];
- reduced resin loading;
- coupling at elevated temperature [431].

4.5.4.4 Chemical Strategies for SPPS Methodological Improvements

As mentioned in the previous section, difficult sequences require special consideration when planning a synthesis. Besides the introduction of backbone amide protecting groups or serine/threonine derived pseudo-proline building blocks, the application of the O-acyl isopeptide method provides appropriate measures for obtaining difficult sequences. It relies on $O \rightarrow N$ intramolecular acyl migration reaction of O-acyl isopeptides (Scheme 4.18). The concept has been developed independently by three different groups and has been named the click peptide [432], switch peptide [433, 434], or depsipeptide method [435, 436].

During the peptide synthesis a serine or threonine residue protected at N^α is incorporated involving the β -hydroxy functionality, giving rise to a depsipeptide bond (N^α -protected O-acyl isopeptide **121**). Such a single isopeptide moiety prevents undesired formation of secondary structures. Cleavage of the intermediate N^α -protecting group triggers an $O \rightarrow N$ acyl migration to give the peptide sequence **122**. The method allowed, for example, the synthesis of the Alzheimer's disease-related amyloid β peptide (1-42) [$A\beta$ (1-42)]. The water soluble $A\beta$ (1-42) isopeptide precursor with Gly²⁵-Ser²⁶ replacement by the corresponding β -depsipeptide upon Ser²⁶ N^α -deprotection undergoes an $O \rightarrow N$ acyl migration forming the target $A\beta$ (1-42). The O-N acyl migration can not only be triggered by acidolytic cleavage of the N^α -Boc group but also, for example, by photolysis of photolabile N^α -protecting groups. Besides $A\beta$ (1-42), other difficult sequences like the Jung-Redemann 10-peptide and 26-peptide, H-(VT)₁₀NH₂ and the 37-peptide of the FEP28 WW-domain have been synthesized.

4.5.4.5 On-Resin Monitoring

Stringent control of the coupling reactions is another indispensable prerequisite in SPPS. The ninhydrin test is one of the simplest and most frequently used methods for this purpose [437]. A positive color reaction, performed with a small aliquot of the resin material, indicates unconverted amino groups. This so-called Kaiser test is simple, reliable (in most cases), and requires only minutes to perform. Titrimetric methods are also quite easy to perform. Further monitoring methods that rely on

color reactions include the TNBS test [438], the acetaldehyde/chloroanil test [439], and the bromophenol blue test [440]. Dorman developed a protocol involving the determination of chloride ions after protonation of unconverted amino groups of the peptide with pyridinium chloride and subsequent elution of bound chloride with triethylamine [441]. Free amino groups may also be directly titrated with 0.1N HClO_4 , for example. Although these methods could be integrated in an automated process, they are too time-consuming. A similar situation applies to gel-phase NMR spectroscopy, which allows a direct determination of resin-bound groups.

In the Fmoc synthesis scheme the coupling of the N^α -Fmoc-protected amino acids can be monitored directly and quantitatively by the decrease in absorbance of the reaction solution at 300 nm. Similarly, cleavage of the Fmoc group with piperidine can be monitored by the increase in absorbance at the same wavelength.

Cleavage and total hydrolysis of an aliquot of resin-bound peptide after a certain number of synthetic steps is also an important and sensible tool to check the integrity of the product. Analysis may also be performed on each synthetic step by mass spectrometry (MALDI-ToF MS) [442]. Exact yield determination and control of the integrity of the peptides bound to the resin can reliably be performed using the IRAA technique [369, 443, 444]. Dual linker analytical constructs comprising additional analytical components for analysis have also been developed [445].

4.5.5

Automation of the Process

The first automated peptide synthesizer was developed by Merrifield in 1966 for batchwise peptide synthesis (Figure 4.44). The original machine, which is now in the Smithsonian Museum, consists of a reactor unit and a controlling unit. The former comprises the reaction vessel and valve systems for solvents, reagents, and amino acid derivatives as well as reservoirs for all these components. Nowadays, different types of synthesizers are commercially available.

Solid-phase peptide syntheses may also be performed in a continuous-flow mode using resin-filled columns. In some instances this method has proven to be superior to the batchwise synthesis developed by Merrifield. Major advantages lie in the reduced reagent and solvent consumption, and in the very short coupling cycles (1–2 min for TentaGel polymers of size 8 μm). The reaction progress may be monitored on a real-time basis by recording the conductivity of the solution [446]. However, the usual types of polymeric support material cannot be used, because of their insufficient pressure stability and the solvent-dependent swelling properties. Hence, chemically modified silica gel [447] and kieselguhr/polydimethylacrylamide hybrid materials have been successfully applied to automated continuous-flow synthesis [448, 449].

TentaGel-type polymers are also appropriate solid support materials for continuous-flow syntheses [370] as they are chemically and physically inert and also stable towards moderate pressure. Their uniform spherical shape and homogenous swelling behavior are further advantages.

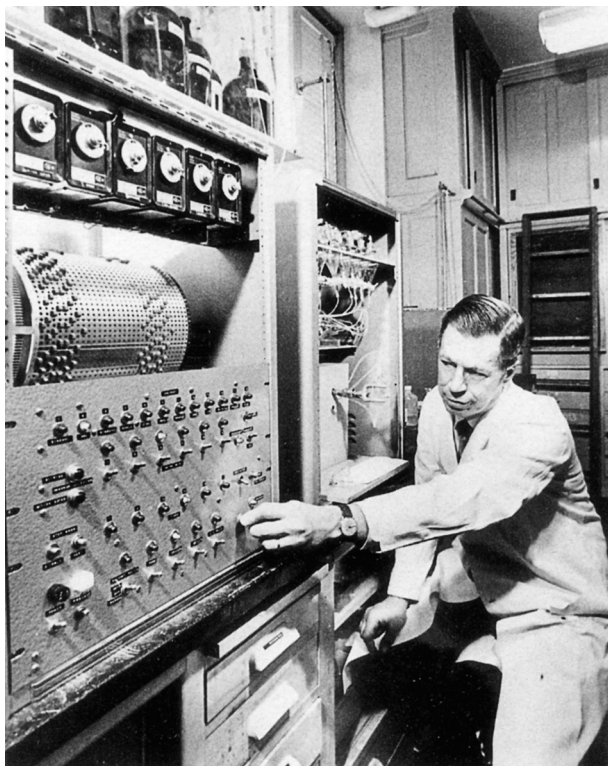


Figure 4.44 Robert Bruce Merrifield and the first-generation solid-phase peptide synthesizer.

4.5.6

Peptide Cleavage from the Resin

Final cleavage of the peptide from the polymeric support usually represents the last step of SPPS. As the anchoring group (linker) is usually chosen to be compatible with the peptide synthetic operations, selective cleavage between the C-terminus of the peptide and the solid support occurs upon treatment with reagents that may concomitantly effect partial or complete deprotection of the peptide side chains. In most cases, acidolytic methods are used, though other methods (transition metal catalysis for allyl-based linkers, photolysis [450, 451]) are available for the synthesis of sensitive compounds or special applications.

4.5.6.1 Acidolytic Methods

The Boc/Bzl scheme often relies on final cleavage by anhydrous liquid HF at 0 °C. This reaction must be carried out in a special apparatus, because HF decomposes glassware. Scavengers such as anisole or others must be added. Both the anchoring group and all nitrogen- or oxygen-protecting groups based on the *tert*-butyl type or benzyl type are cleaved. The nitro group of nitroarginine and the methoxybenzyl group of cysteine for example are also labile towards these conditions. Other

functionalities, such as N^{im}-benzyl groups, N^{im}-2,4-dinitrophenyl groups (His), S-alkylthio groups (Cys), and 4-nitrobenzylesters (Asp, Glu) are resistant to these cleavage conditions. The two-step, high–low HF protocol often provides the final product at greater purity.

Other acidolytic cleavage reactions do not require a special apparatus and may be performed in the reactor that is used for solid-phase synthesis. Cleavage reactions with TFMSA in TFA as the solvent can be improved considerably by the addition of scavenging sulfur compounds such as thioanisole or methionine [452, 453]. Several trialkylsilyl compounds have been selected as cleavage reagents on the basis of the HSAB (*hard and soft acids and bases*) concept developed by Pearson. Hard Lewis acids such as trimethylsilyl trifluoromethane sulfonate (TMSOTf) or trimethylsilyl bromide (TMSBr) are combined in TFA as the solvent with soft bases (e.g., thioanisole) as a highly efficient deblocking system both for solution-phase and for solid-phase purposes [454]. Cleavage with this “cocktail” (e.g., 1 M TMSOTf/thioanisole in TFA; 0 °C for 60 min) is much faster than the conventional method applying 1 M TFMSA/thioanisole in TFA. The different methods for acidolytic sulfide-assisted cleavage in nonaqueous solvents have been reviewed with respect to mechanism and applications [455].

The Fmoc/tBu scheme allows much milder final deprotection conditions, for example with TFA or acetic acid in dichloromethane as the solvent (cf. Section 4.5.1).

In the case of special linker systems (e.g., HMBA), cleavage with different nucleophiles is possible, and results in the formation of C-terminal peptide acids (cleavage with aqueous or alcoholic hydroxide solutions, thiophenolate in DMF, or solvated cyanide), peptide esters (alcohols in the presence of tertiary amines), peptide amides (ammonia or amines), peptide hydrazides (hydrazine), or even peptide aldehydes and alcohols (reductive cleavage with LiBH₄).

Without doubt, the final cleavage reaction is the most critical step of an SPPS, though on occasion even the separation of the two steps, linker cleavage and side-chain deprotection, may be advantageous.

4.5.6.2 Side Reactions

Sensitive peptide bonds such as Asp-Pro or Aib-Pro may be cleaved under the strongly acidic cleavage conditions. Asp-Gly sequences are often prone to succinimide formation and concomitant isoaspartyl peptide ($\alpha \rightarrow \beta$ shift) formation, especially when treated with strong acids such as HF or TFMSA. Similar reaction conditions facilitate an N $\rightarrow$ O acyl shift involving serine or threonine residues. The products formed are thermodynamically labile, and the process may be reversed in aqueous bicarbonate solution (pH 7.5).

Tryptophan or tyrosine residues may, under certain circumstances, suffer from oxidative degradation. Tyrosine is prone to electrophilic C-alkylation, which may be prevented by the 2-bromobenzyl protecting group in Boc tactics. Arginine residues are reported to undergo deguanylation to give ornithine residues. As mentioned previously, methionine and cysteine residues are susceptible to oxidation and alkylation. In Fmoc tactics, most side reactions (e.g., side-chain alkylation or lactam formation) and others are sufficiently reduced by appropriately protected amino acid

building blocks, for example Fmoc-Arg(Pbf)-OH [456], Fmoc-Arg(Pmc)-OH [457], Fmoc-Trp(Boc)-OH [458], and Fmoc-Asn(Trt)-OH [459].

4.5.6.3 Advantages and Disadvantages of the Boc/Bzl and Fmoc/tBu Schemes

Although, as mentioned previously, one advantage of the Boc tactics lies in the termination of β -sheet structures during repetitive acidolysis of the temporary Boc protection, the major drawbacks are to be seen in the harsh reaction conditions for the final deprotections that give rise to different side reactions (e.g., side-chain alkylation, cleavage of sensitive peptide bonds, rearrangements, cf. Section 4.5.7.2). Moreover, special equipment is needed for the final HF cleavage reaction.

Fmoc tactics are characterized by the possibility of synthesizing side-chain-protected fragments for segment condensations and peptide cyclizations, the relatively mild cleavage conditions of the side-chain-protecting groups and of the linker, and by UV-spectroscopic monitoring of the reaction progress. A detrimental feature is the favored formation of β -sheets and the propensity for diketopiperazine formation in the dipeptide stage. Spontaneous cleavage of Fmoc has also been observed [416, 417].

Systematic investigations slightly favor the Fmoc/tBu scheme over the Boc/Bzl scheme [460]. However, this conclusion does not in the main rely on the synthetic performance, but rather on the greater practicability and easy technical application. Rapid continuous-flow protocols [461] and optimized support materials [426] are regarded as beneficial.

Nonetheless, a general decision regarding which protecting group tactics would be preferred cannot be made. All significant parameters in the synthesis of longer peptides or peptides containing difficult sequences should be carefully analyzed, and the selected synthesis scheme should first be investigated in a small-scale synthesis (preview synthesis) combined with appropriate analytical methods in order to identify potential weak points in the synthetic plan.

4.5.7

Examples of Syntheses by Linear SPPS

Since its invention in 1963, the Merrifield concept has been applied to the synthesis of thousands of biologically active peptides and peptide analogues and, in this respect, a large number of improved variants have been used besides the original protocol. It must be mentioned that this synthetic concept has had an enormous impact on the development of peptide chemistry and organic chemistry during the past few decades. Indeed, the manifold approaches to combinatorial chemistry would not have been developed without the foundations having been set by Bruce Merrifield.

SPPS underwent extreme activity between 1968 and 1972, when the commercialization of automated synthesizers fostered its widespread application. In those early days, the chemical synthesis of small proteins had its limitations, and an attempted synthesis of lysozyme resulted in a complex mixture with a specific activity of only 0.5–1%. The total synthesis of ribonuclease A [462] proved to be a benchmark in that it was the first synthesis of an enzyme with a satisfactory activity of 3.5%. The application of HPLC methods in the purification of synthetic peptides during the early 1980s again led to an enormous increase in solid-phase synthesis activities that

continues to this day. Initially, the solid-phase syntheses of molecules such as conotoxin G1, thymosin α , glucagon, β -endorphin, parathyrin, β -lipotropin, interleukin-3, HIV protease, and green fluorescent protein provided an impressive announcement of the vast potential of this synthetic method. Usually the commercial production of peptides includes the preparation of compounds such as corticotropin, atriopeptin, parathyrin, and calcitonin on a scale up to 200 g [463]. However, since the drug enfuvirtide (T20, Fuzeon™) has been developed and approved, large-scale peptide synthesis even of such a long peptide (36 amino acids) is performed on a metric ton scale (cf. Sections 5.3.4 and 9.1)

4.5.8

Special Methods of Polymer-supported Synthesis

Liquid Phase Peptide Synthesis (LPPS) is based on the application of a soluble polymeric support with the advantage of homogeneous reaction conditions. This may overcome some disadvantages of SPPS. Although the use of non-cross-linked soluble polystyrene ($M_r \sim 200$ kDa) as a solid support allows synthesis to be carried out in the homogeneous phase [464], the excess reagents present after each coupling step can only be separated by quite tedious precipitation operations. The use of heterogeneous preparations of low-molecular weight polyethyleneglycol (PEG) as a polymeric support and a C-terminal-protecting group for the growing peptide chain led to the important improvement of liquid-phase peptide synthesis [465]. In this method, excess low-molecular weight coupling reagents can be separated by ultrafiltration. Further progress was achieved in the development of the so-called crystallization method of Bayer and Mutter [466]. The addition of a suitable organic solvent (diethyl ether) rapidly induces spontaneous and quantitative crystallization of the peptidyl-PEG by the formation of helical structures. Low-molecular weight starting materials and reagents are usually not co-precipitated, and may be easily separated. The principle of liquid-phase peptide synthesis is shown in Figure 4.45. Although coupling times correspond to those of conventional synthesis, an excess of acylating agents and double or multiple couplings are necessary in order to obtain maximum conversion. The growing polypeptide chain influences the solubility, and viscous solutions may result even when DMF is used as solvent. The recently available homogenous preparations of PEG could be used to convert LPPS to a true handle approach (cf. Section 5.4.3.2). The preferred area for the application of liquid-phase peptide synthesis is for polypeptides containing fewer than 30 amino acid building blocks. The clear disadvantages of liquid-phase peptide synthesis are the long time taken for the operation, and the difficulties in its automation. In 2000, the versatile protection strategy based on the Fmoc group was adapted to LPPS [467].

Ionic liquid-supported peptide synthesis is based on anchoring the first N-protected amino acid to a suitable ionic liquid, e.g., 3-hydroxyethyl-(1-methylimidazolium)-tetrafluoroborate using common coupling agents [468]. After solvent washing, the next Boc-amino acid is coupled step-by-step followed by detachment from the ionic liquid and phase-separation. Up to now, this approach has not found further applications.

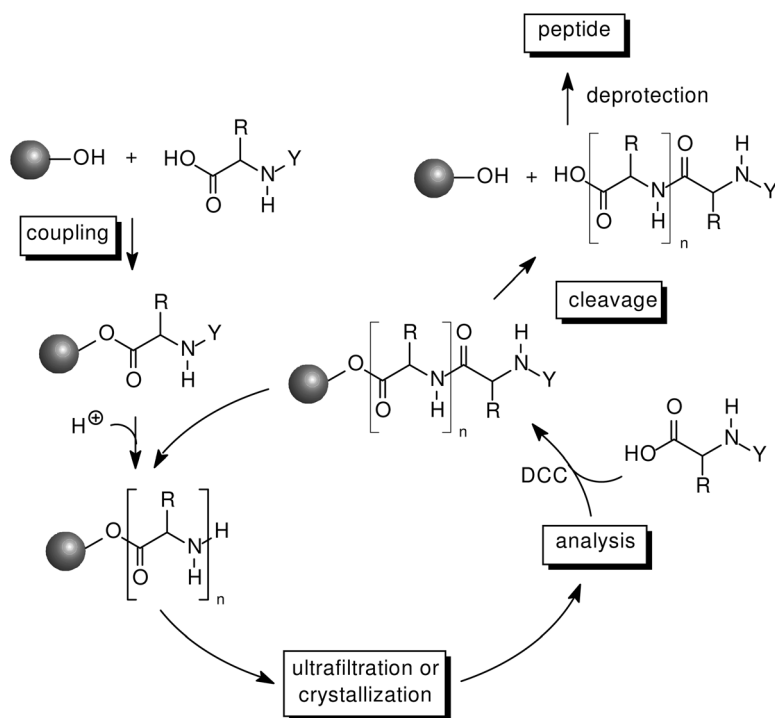


Figure 4.45 Scheme of liquid-phase peptide synthesis on a soluble polymer.
Y = N^α-protecting group.

Alternating solid-phase/liquid-phase peptide synthesis [469] combines the strategic aspects of both SPPS, LPPS, and of the polymer reagent peptide synthesis. The basic idea comprises a physical modification of the growing polypeptide chain during the coupling reaction in order to facilitate separation of the starting material from the product. However, this interesting concept has not yet received practical application.

Polymer-supported activating reagents such as polymer-supported *N,N'*-dicyclohexyl carbodiimide (PS-DCC) are characterized by the main advantage that the coupling reagent-related by-products, in this special case *N,N'*-dicyclohexyl urea, are polymer bound and easy to eliminate by filtration.

Catch and release approach is nowadays termed a continuous process for peptide synthesis by applying polymer-bound active esters already developed by Wieland and Birr [470] in 1966. This polymer reagent peptide synthesis does not utilize synthesis on a polymeric support, because the growing polypeptide chain is retained permanently in solution. Instead, the polymer-bound reagents – and especially the activated carboxy components – are reacted with amino components with liberation of the polymer and of the protected peptide derivative. The polymer-bound reagents can be applied in excess, and the peptide formed may be separated – usually without difficulty – avoiding an excess of amino component in the final mixture. The columns contain a polymer cross-linked with *p,p'*-dihydroxydiphenylsulfone where the active

esters of N-protected amino acids are attached. However, the repetitive acidolytic deblocking reagents with subsequent neutralization steadily increase the salt concentration of the reaction solution. Photolabile N^α-amino-protecting groups have also been recommended for this column operation. Following synthesis, these are cleaved simply by photolysis, after which the resulting peptide with the free amino-terminus may be subjected to the next synthetic cycle. Further alternatives of polymer-supported active esters using tetrafluorophenol (TFP) resin [471], HOBT resin [472] and oxime resin [473] have been described. The appropriate carboxy component is loaded onto the resin by classic ester condensation approaches.

Soluble handle approaches including the application of positively charged handles and the excluded protecting group (EPG) method [474] are discussed in Chapter 5, Section 5.4.3.

Finally, during the last decades, solid phase strategies have yielded numerous applications, especially in the field of parallel and combinatorial chemistry [475].

4.5.9

Microwave-Enhanced Peptide Synthesis

Interestingly, peptide synthesis may be accelerated by using microwave irradiation as the energy source. The first example of microwave-enhanced solid-phase peptide synthesis was published in 1992. The authors noted that microwave heating reduced amino acid coupling time 2–4 fold, especially when side-chain hindered amino acids were used. Surprisingly, racemization was usually not observed [476]. For coupling of racemization-prone amino acids (e.g. His, Cys), epimerization can be suppressed by cooling the contents of the reaction vessel in an ice bath before applying microwave irradiation for a short period and then repeating this procedure several times. Optimized conditions for synthesizing longer peptides involve deprotection steps of three microwave pulses of 30 s at 100 W with intermediate cooling ($T_{\text{max}} = 40^{\circ}\text{C}$) and coupling steps of five microwave pulses of 30 s at 50 W, again with intermediate cooling.

Automatic microwave-enhanced SPPS provides even longer peptides in a very good crude purity within a short period of time. A 41-peptide has been prepared in a total time of 31 h. The reason why microwave heating exerts an accelerating effect on the deprotection and peptide coupling reactions is still under debate. Besides de-aggregation of the peptide leading to increased coupling rates, continuous re-orientation of the amide dipoles in the microwave field might be one reason for efficient energy transfer. Energy transmission is obviously due to dielectric losses and not to conduction and convection processes like in conventional heating.

α -Aminoisobutyric acid (Aib) was efficiently coupled to sterically hindered natural amino acids with microwave heating using either PyBOP/HOBt or HBTU/HOBt as the coupling reagents [477]. The main constituent of the fibrillar aggregate responsible for Alzheimer's disease, A β (1-42), was prepared in 69% purity within 19 h in an automated microwave peptide synthesiser [478]. Microwave heating has also been used for the coupling of glycosylated amino acids in SPPS [479]. Microwave irradiation has also been successfully applied to the synthesis of peptide arrays in combinatorial synthesis [480].

4.6

Biochemical Synthesis

Cells are capable of assembling amino acids and modifying peptides and proteins in a highly regulated and chemically efficient manner (cf. Section 3.2). These *in vivo* reactions occur under mild conditions, and the catalyzing enzymes are the keys to the specificity and efficiency of the process. Before the ribosomal synthesis pathway had been elucidated, general peptide and protein biosynthesis was believed to be performed by a number of simple enzymatic processes. Moreover, despite the fact that the ribosomal cycle of protein biosynthesis (cf. Section 3.2.1) is now known to comprise complex machinery, the formation of a peptide bond was originally thought to be catalyzed by an enzyme that was tentatively named peptidyl transferase. Indeed, the peptidyl transferase center (PTC) is a ribosomal complex catalyzing peptide bond formation and peptide release, but the molecular details of the ribozyme mechanism have yet to be revealed.

Regiospecificity and stereospecificity are two major advantages of biocatalysts. In chemical peptide synthesis, the obligatory protection and deprotection of the α -amino function, the carboxy group and the various side-chain functionalities before and after a synthetic step are the most fundamental time-consuming operations. Furthermore, as many of the synthetic operations take place at groups adjacent to a center of chirality, a permanent risk of racemization must be taken into consideration. Therefore, the chemical synthesis of peptides continues to be a formidable chemical effort.

Although there are more than 150 chemical methods for peptide bond formation, an ideal coupling method should allow rapid reaction, without racemization or other side reactions, and result in quantitative coupling when the carboxy and amino components are applied stoichiometrically.

Recombinant DNA techniques, genetic engineering, genome sequencing, biochemical protein ligation, and the application of enzymes in chemical synthesis to control stereo- and regiospecificity have led to the so-called “biotechnological revolution”. For this reason, new biochemically inclined methodologies are required on the synthetic front, both for the functional analysis of peptides and proteins, and the discovery of new therapeutic agents.

4.6.1

Recombinant DNA Techniques

The nucleotide sequence of the human genome has now been elucidated. Despite the existence in human cells of about 150 000 different genes each containing approximately 5×10^9 base pairs, the isolation of gene sequences is a realistic approach. The methods used for sequence analysis of DNA according to Maxam and Gilbert (1977) or Sanger (1977) can be applied reliably. Today, with the advent of nano-sequencing methodology like the *454 technology*, the sequence analysis of an isolated gene is more efficient than protein sequence analysis, and consequently protein sequences are often derived from DNA sequences.

Only a limited number of specialized peptides (cyclic peptides, peptide antibiotics) are synthesized nonribosomally by multi-enzyme complexes. A preparative procedure for the ribosomal synthesis of polypeptides and proteins is based on the principles of gene technology. Proteins and peptides can be expressed on a large scale by the recombination and expression of genetic material, for example in bacteria, and pharmacologically active peptides and proteins form a special focal point of interest in this context [481–483]. Indeed, based on their endogenous functions in the human body, these compounds may find application as therapeutic drugs, as well as influencing both pathophysiology and healing processes (cf. Chapter 9).

4.6.1.1 Principles of DNA Technology

Since the elucidation of the molecular structure of DNA by Watson and Crick, genetics and molecular biology have experienced almost unbelievable development. There is no species barrier for DNA, which serves as the “blueprint” for protein biosynthesis. For example, bacteria such as *Escherichia coli* may not only accept genetic material from other microorganisms and stably transmit it to their successors, but may also serve as recipients of genetic material from either plants or animals.

Gene expression comprises synthesis of the corresponding mRNA (transcription) and synthesis of the protein (translation). The biosynthesis of a foreign gene product (protein) in an organism relies on a recombination of the genetic material of the microorganism with the DNA fragment encoding for the desired protein.

Eukaryotic genes contain noncoding elements (introns) in addition to the sequence information encoding for a protein. The precursor messenger RNA (mRNA) formed after transcription is subsequently processed into mature mRNA by splicing, at which point the introns are removed from the sequence.

Consequently, the genomic DNA of eukaryotes is not appropriate for direct transfection; instead, DNA which is complementary to the mature mRNA (cDNA) must be synthesized and used for transfection. The formation of cDNA is a prerequisite in order to obtain the correct protein sequence derived from a human gene, for example. Mature mRNA isolated from the donor cells is used as a template to synthesize cDNA *in vitro* with the enzyme reverse transcriptase, which occurs naturally in retroviruses. This enzyme requires a synthetic primer (of about 12 nucleotides) that is complementary to the sequence of the 3'-end of mRNA. In eukaryotic cells, this nucleotide sequence is often a poly-adenosine tail, but it may also encode for the C-terminal amino acid sequence of the protein. The hybrid composed of cDNA and mRNA is then subjected to alkaline hydrolysis, which destroys the mRNA. The cDNA is subsequently replicated and amplified by polymerase I. Blunt ends are formed by S1-nuclease, and synthetic oligonucleotide sequences containing cleavage points for restriction endonucleases are attached using DNA ligase for insertion into the vector. Alternatively, homopolymeric sequences are added using terminal transferase. The procedure is shown schematically in Figure 4.46. Finally, the encoding DNA may be sequenced for verification purposes and subsequently be inserted into expression vectors of suitable host cells. The process comprises the following steps (Figure 4.47):

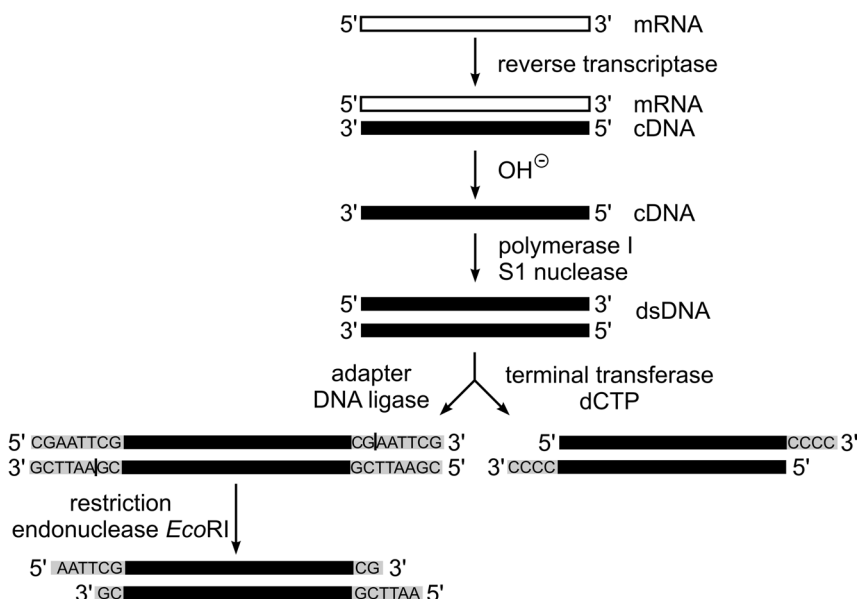


Figure 4.46 Reverse transcription of mRNA into cDNA, followed by attachment of synthetic oligonucleotides.

1. Isolation of the encoding DNA fragment from the donor organism.
2. Insertion of the DNA into a vector.
3. Transfection of the vector into the host organism.
4. Cultivation of the host organism (cloning), which leads to gene amplification, mRNA synthesis, and protein synthesis.
5. Isolation of the recombinant protein.

Originally, the term “cloning” meant the amplification of cells, with formation of identical daughter cells, but nowadays the term is also used for the amplification of DNA. Initially, minute amounts of the protein are identified using biological assays. Often even microgram amounts are sufficient in order to raise antibodies or to perform partial sequencing, starting from the N-terminus. By using this sequence information, the corresponding genes can be identified and either isolated or obtained from gene banks.

The DNA fragments containing the genetic material to be transfected can be introduced into bacteria or other host organisms using so-called vectors. Usually, either plasmids (circular double-stranded DNA molecules that exist extrachromosomally in bacteria) or phages (viruses that infect bacteria) are used as vectors. The discovery and isolation of restriction endonucleases capable of cleaving double-stranded DNA at well-defined positions was the major precondition in the development of gene technology.

The DNA encoding for a protein can be obtained by isolation of the corresponding gene, by enzymatic synthesis from mRNA, or, in special cases, by chemical synthesis. Chemical DNA synthesis [484] can be performed on commercial

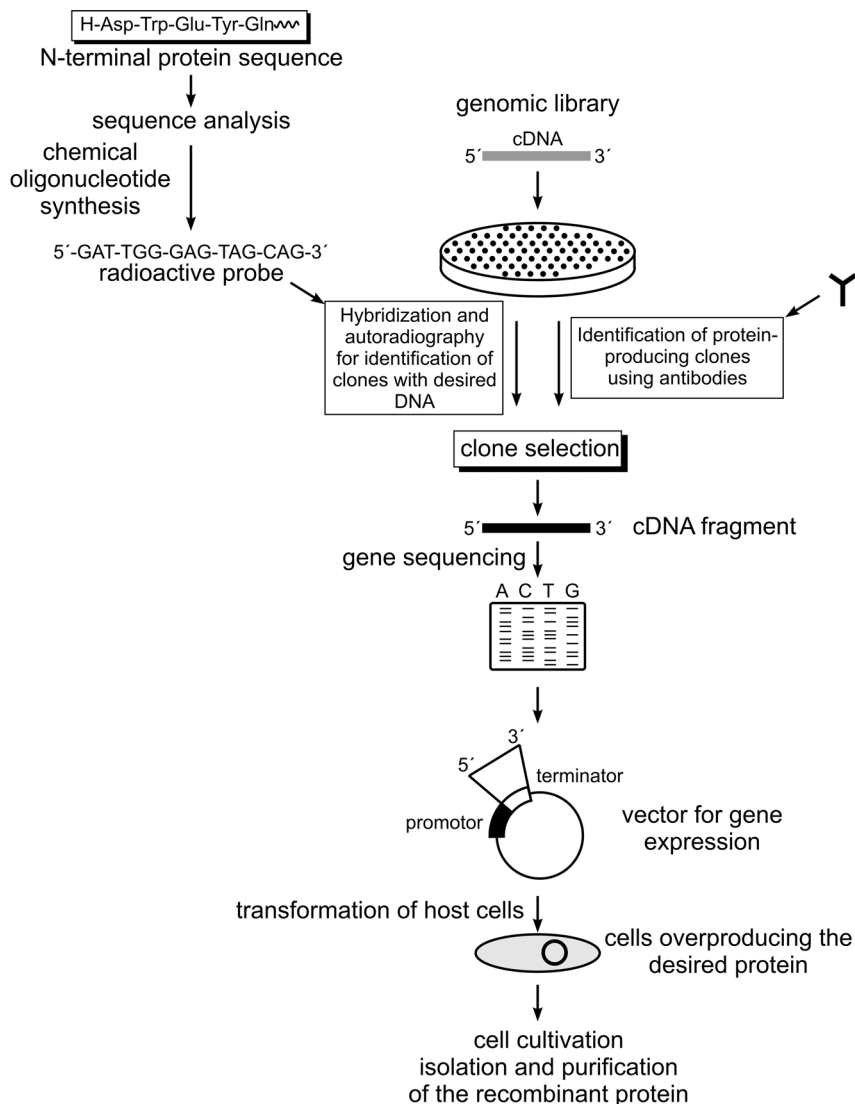


Figure 4.47 Procedure to obtain a recombinant protein from the gene.

automatic synthesizers more efficiently than peptide synthesis. Genes of different proteins (e.g., the genes of human interferons) that contain more than 500 base pairs, or the DNA sequences encoding for the two insulin chains, have been synthesized chemically. The chemical synthesis of oligonucleotides containing special cleavage sequences for restriction endonucleases is also of great importance. Such synthetic linker groups are attached to the ends of a DNA molecule (Figure 4.46) in order to allow the insertion of this DNA fragment into the vector. The DNA to be amplified and expressed, as well as the vector DNA, are cut by an

appropriate restriction endonuclease, and the desired DNA fragment is inserted. The phosphate linkage between the DNA portions is formed by ligases to give the intact recombinant vector DNA molecule, which then is introduced into the host organism. Successful transfection can be monitored with labeled oligonucleotides (*in situ* hybridization), or on the basis of the protein product using antibodies against this specific protein.

The successful expression of a recombinant protein in a host system requires not only the DNA sequence that encodes for the corresponding protein, but also additional regulatory sequences on the DNA inserted into a plasmid or another vector. Usually, highly efficient regulatory DNA domains are inserted preceding the DNA fragment encoding for the gene product in order to provide a high protein yield.

The selection of new host organisms for recombinant DNA is a major task, because *E. coli* as a host suffers from several disadvantages. Very often, eukaryotic proteins are modified post-translationally (e.g., by glycosylation), and hence contain carbohydrate chains of different lengths. Microorganisms do not modify proteins post-translationally, and usually do not export proteins into the cell culture medium. Furthermore, *E. coli* produces endotoxins that must be separated from the desired protein before its use as a pharmacologically active compound.

In the meantime, recombinant techniques for *Bacillus*, *Rhizobium* and *Streptomyces* have been described. Eukaryotic cells (from yeasts, insects, or mammals) can also be used as host organisms. Baker's yeast (*Saccharomyces cerevisiae*) or other yeasts are ideal expression systems that also allow post-translational glycosylation. Stable plasmids are known for baker's yeast. Yeast applied in biotechnological processes is usually allowed to reproduce asexually, which safeguards stable transmission of the genetic material. The genes for interferon and insulin have been expressed in yeast.

Complex post-translational modifications such as glycosylation patterns specific for humans, or γ -carboxylations on glutamate residues, require mammalian cells as expression host systems.

4.6.1.2 Examples of Synthesis by Genetic Engineering

As outlined in the preceding section, microorganisms may be genetically programmed to produce human hormones (e.g., insulin, somatotropin), enzymes, or vaccines. The potential of this area of applied gene technology has quickly been recognized, and many new applications are currently under investigation in medicinal chemistry, biotechnology, and agriculture. Currently, many therapeutic peptides and proteins are obtained by the fermentation of recombinant cells (cf. Chapter 9).

Somatostatin was the first peptide to be obtained by the use of recombinant DNA techniques [485]. The DNA encoding for the amino acid sequence of somatostatin was synthesized chemically using the amino acid triplet codons used most frequently in *E. coli*, after which the DNA sequence was inserted into an expression plasmid. The relatively small peptide was protected against proteolysis by fusion to the protein β -galactosidase. Cleavage of the 14-peptide from the hybrid protein with cyanogen bromide was only possible because somatostatin does not contain methionine (Figure 4.48).

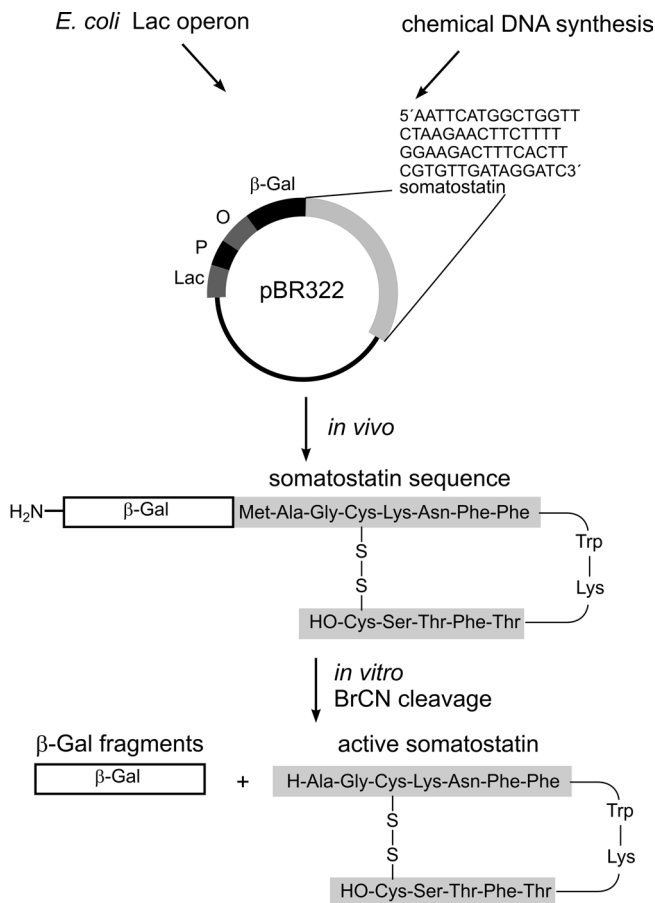


Figure 4.48 Recombinant synthesis of somatostatin.

Interestingly, the genes encoding for the interferons from human leukocytes do not contain any introns, and so interferons can be produced in a straightforward manner in bacteria.

The production of human growth hormone using recombinant DNA techniques represents a major milestone in the substitution therapy of dwarfism. Before 1985, this hormone could be obtained only from human hypophyses at autopsy. All analogues produced natively by other mammalian organisms have different amino acid sequences, and the potential for viral contamination has also been a point of criticism. The gene family of the human growth hormone was discovered during molecular biology studies. The gene contains four introns, and the regulatory N-terminal pre-sequence is composed of 26 amino acids, which is typical for an export protein. This sequence is cleaved during secretion by a signal peptidase. Expression of the cDNA in *E. coli* originally provided the human growth hormone in a yield of 2.4 mg L^{-1} , which is far from optimum [486]. Protropin® (Genentech), which was approved for therapeutic use in 1985, additionally contains one nonse-

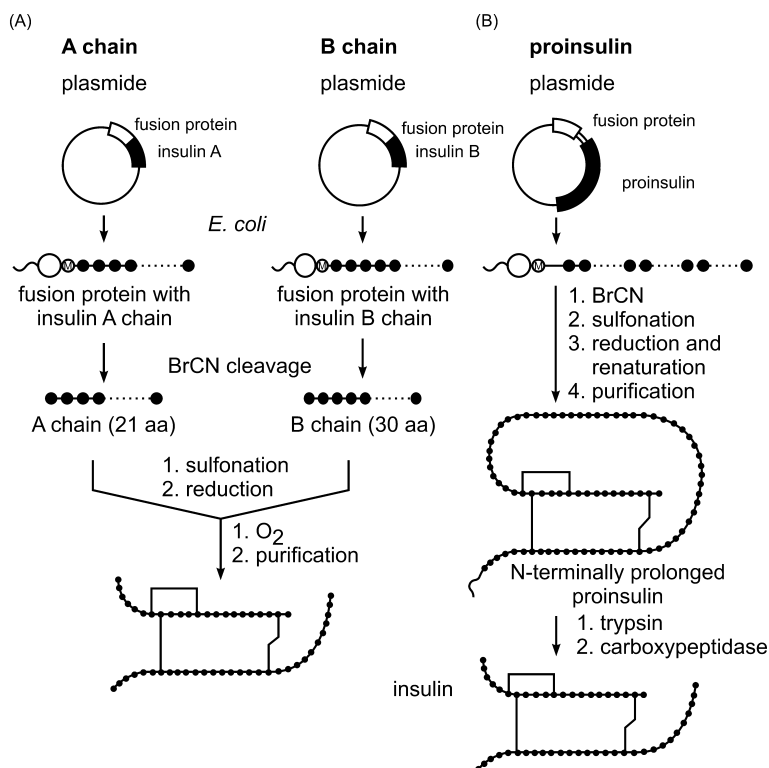


Figure 4.49 Biotechnological insulin synthesis. (A) Separate synthesis of A and B chain, followed by oxidation to give active insulin. (B) Synthesis of proinsulin with subsequent enzymatic processing.

quence-specific N-terminal methionine residue, while the peptide with the original sequence was marketed as Humatrop® (Eli Lilly) [487].

Two alternative procedures are described for the overexpression of insulin. Eli Lilly expressed human insulin by separate overexpression of the A and B chains in the form of fusion proteins in *E. coli* (Figure 4.49A) [488]. Following the isolation of both chains, they were connected by reductive thiolysis and subsequent oxidation (in air) to produce the active hormone. In the second variant (Figure 4.49B), proinsulin was synthesized analogously to biosynthesis as a fusion protein in *E. coli* and transformed enzymatically into active insulin [489].

Insulin analogues designed to improve the therapeutic properties of the hormone have been produced by recombinant DNA technology [490]. The first clinically available insulin analogue, lispro, was marketed in the U.S. in 1996 [491].

Insulin glargine (HOE 901, Lantus®) is produced using a nonpathogenic genetically engineered strain of *E. coli* and has found application as a long-acting agent for the treatment of type 1 and 2 diabetes mellitus (cf. Section 9.3.1). A typical DNA technology production plant is shown in Figure 4.50.

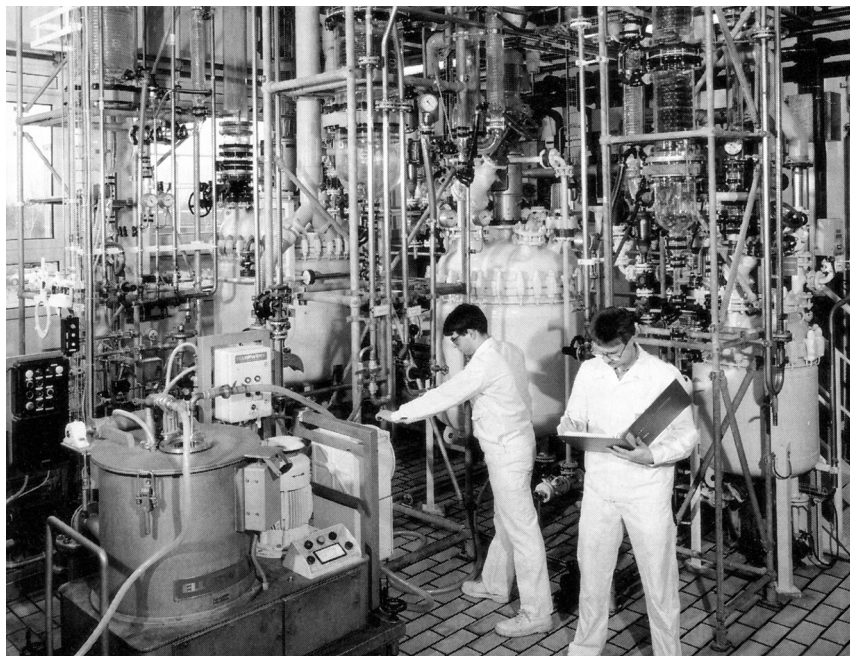


Figure 4.50 Production plant for industrial gene technological insulin synthesis (Photo: Walter Kloos/Hoechst AG).

4.6.1.3 Cell-free Translation Systems

As shown above, the biotechnological synthesis of natural products is mainly based on cellular systems. Methods for the expression of foreign genetic material in living cells often suffers from limitations, however, and this is especially true for the production of polypeptides of a comparatively short length (up to 50–60 amino acid residues). Since most bioactive peptides and polypeptides are released after the complex processing of biosynthetic precursors, it is difficult to create vectors that result in the formation of active peptide products when expressed in either prokaryotic or eukaryotic cells. Even when an expression system has been successfully created, the peptide products may be unstable or toxic to the cells. Other problems also result from insoluble peptide products forming aggregates and inclusion bodies. These general limitations may be avoided when translation is performed in cell-free systems.

Zamechnik *et al.* [492] described the first systems for cell-free protein synthesis, and demonstrated that peptide synthesis proceeds on ribosomes and requires ATP, GTP, and tRNA. The coat protein of coliphage F2 was the first protein synthesized in a cell-free translation system using an extract from *E. coli* [493]. The main disadvantage of cell-free translation systems is the low peptide yield, as under these conditions only two to three polypeptide chains are formed per mRNA molecule used [494, 495].

A continuous cell-free translation system capable of producing polypeptides in high yield was described by Spirin *et al.* [496]. This continuous-flow, cell-free (CFCF) system allows continuous feeding of amino acids, ATP, and GTP to the reaction

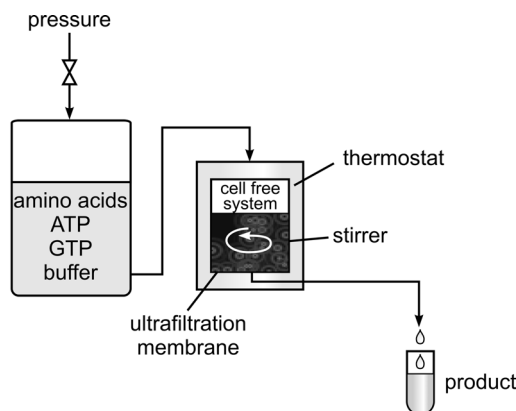


Figure 4.51 Schematic diagram of a bioreactor for continuous-flow translation, fitted with an ultrafiltration membrane.

mixture containing the cell extract and mRNA, and continuous removal of the polypeptide product (product removal proved to be essential to maintain prolonged ribosomal activity). The simplest device for performing a continuous-flow translation contained an Amicon 8 MC microfiltration system which was installed in a thermostatically controlled chamber (Figure 4.51).

Later, a thermostatically controlled chromatography column supplied with an ultrafiltration membrane at the output and a standard column adapter at the input was used. These continuous-flow bioreactors have been applied to cell-free translation using both prokaryotic (*E. coli* lysate) and eukaryotic (wheat embryos, *Triticum* sp.) systems. The mRNA of human [Val⁸]calcitonin was obtained by *in vitro* SP6 phage polymerase transcription of the synthetic calcitonin gene inserted into a plasmid under SP6 promoter, and was used for human [Val⁸]calcitonin synthesis. Between 150 and 300 copies of calcitonin were synthesized per mRNA in both types of translation systems during 40 h. Pulsed bioreactors fitted with a hollow fiber ultrafiltration membrane and with a flat ultrafiltration membrane, respectively, have also been used. The inclusion of the translation mixture in granules (e.g., alginate) or polysaccharide vesicles was found to be advantageous. In this configuration, the translation of calcitonin mRNA proceeded for 100 h and produced 250 µg calcitonin from 5 mL of the incubation mixture. The CFCF system enables the production of rather pure polypeptides based on the fact that the ultrafiltration membrane prevents leakage of most proteins present in the reaction mixture. The CFCF system has doubtless opened new fields of application, and many of the limitations originally imposed on cellular systems have now been overcome using cell-free gene expression. An ongoing problem is that the productivity and efficiency of cell-free systems are still low, and newer bioreactors that incorporate efficient systems to monitor the translation course are required. Moreover, strict maintenance of ATP and GTP concentration, and the development of efficient regeneration systems for both cofactors, are indispensable. However, several promising approaches to increase the efficiency of the CFCF system have been suggested [497, 498].

One major breakthrough in recent years was the development of a new generation of cell-free transcription–translation systems. The open and flexible character of these systems allows direct control over expression conditions via the addition of supplements to the expression reaction. The *Rapid Translation System* (RTS) is a cell-free protein production system employing an enhanced *E. coli* lysate to perform coupled *in vitro* transcription–translation reactions. The wheat germ based system is of special interest for producing eukaryotic multidomain proteins in a folded state. A continuous supply of energy substrates, nucleotides and amino acids, combined with the removal of by-products, guarantees a high yield of protein production. Straight-forward downstream processing, circumventing the need for cell lysis, renders such systems ideally suited for high-throughput screening applications [499–501].

4.6.1.4 Proteins Containing Non-Proteinogenic Amino Acids – The Expansion of the Genetic Code

An interesting concept to expand the genetic code has been introduced by Schultz and coworkers [502]. This uses chemically synthesized 3'-aminoacyl suppressor tRNA that allows the incorporation of nonproteinogenic amino acids of any type into a protein. The codon UAG on mRNA usually associates with nonacylated suppressor tRNA and leads to protein chain termination. The method is based on chemoenzymatic aminoacylation of the suppressor tRNA with the nonproteinogenic aminoacyl residue (Figure 4.52), which is subsequently incorporated into the protein. More than

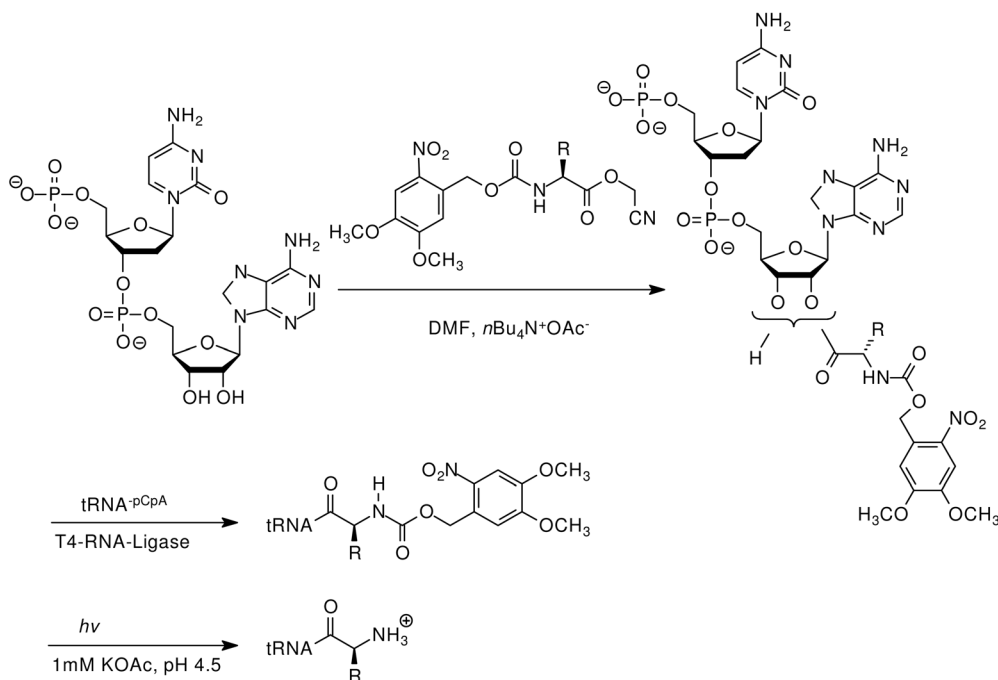


Figure 4.52 Chemoenzymatic Aminoacylation of tRNA with Noncoded Amino Acids According to Schultz *et al.* [502].

30 different non-encoded amino acids have been incorporated into proteins, including derivatives with fluorophors, caged compounds, photoaffinity labels, or spin labels in the side chain [503]. The usage of the stop codon UAG enables site-selective incorporation of the non-proteiniogenic building block.

4.6.2

Enzymatic Peptide Synthesis

Although enzymes have, in general, become valuable tools in medium- to large-scale synthetic organic chemistry [504], a universal C–N ligase with high catalytic efficiency for all possible combinations of the 22 proteinogenic amino acids to be coupled has yet to be developed. Indeed, the demands on specificity are so severe that not even nature could solve this problem. Hence, protein biosynthesis (cf. Section 3.2.1) has been developed as a stepwise strategy, starting from the N-terminus of the growing peptide chain. The process is catalyzed by the ribosomal peptidyl transferase, and followed by limited proteolysis and other post-translational modifications of the precursor molecules (cf. Section 3.2.2). According to studies by the Nobel laureate Thomas R. Cech *et al.* [505], an *in vitro* selected ribozyme is capable of catalyzing the same type of peptide bond formation as a ribosome, and its sequence and secondary structure appear strikingly similar to the “helical wheel” portion of 23S rRNA involved in the activity of ribosomal peptidyl transferase. These findings provided evidence for the feasibility of the “RNA world” hypothesis, by demonstrating that RNA itself is capable of catalyzing peptide bond formation. In comparison to the specificity requirements of proteases, peptidyl transferase appears to be an old, unspecific ribozyme, which is in accordance with its function in evolution.

Even if it were possible to separate ribozyme activity from the ribosome, or to isolate an *in vitro* selected ribozyme that catalyzes peptide bond formation like a ribosome, such a biocatalyst does not seem suitable for simple practical use in peptide synthesis.

Consequently, those enzymes which usually act as hydrolases, catalyzing the cleavage of peptide bonds, should be considered as ‘peptide ligases’. The reaction of proteolysis is irreversible, as peptide bond hydrolysis is exergonic. Although the ionic hydrolysis products are thermodynamically more stable, the fundamental suitability of proteases for catalysis of the formation of peptide bonds is based on the principle of microscopic reversibility that was predicted by J. H. van’t Hoff in 1898. Despite this, about 40 years elapsed before the first experimental proof of van’t Hoff’s prediction was produced, this being based on the clear-cut protease-catalyzed synthesis of an amide bond [506]. Unfortunately, another 40 years elapsed before this approach gained any industrial importance, though during the past few decades considerable effort has been made to determine the optimum conditions for protease-catalyzed peptide synthesis (for reviews, see [507–511]).

4.6.2.1 Approaches to Enzymatic Synthesis

According to the short notation of protease–substrate interactions [512], proteases have binding grooves ($S_n \dots S_3$ – S_2 – S_1 – S'_1 – S'_2 – S'_3 – $\dots$ – S'_n) on both sides of the scissile

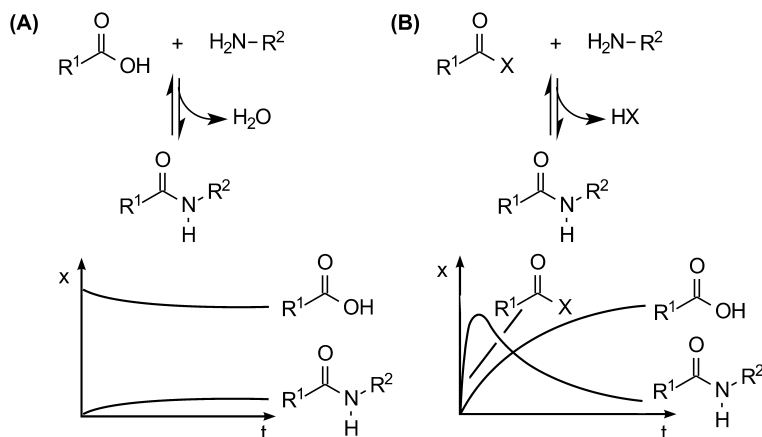


Figure 4.53 Comparison of the equilibrium controlled approach (A) and the kinetically controlled approach (B) of protease-catalyzed peptide synthesis. R^1-COOH = carboxy component; R^1-CO-X = slightly activated carboxy component; H_2N-R^2 = amino component; HX = leaving group.

bond $-P_1-P'_1-$ of the substrate, whereas substrate binding ($P_n \dots P_3-P_2-P_1-P'_1-P'_2-P'_3 \dots P'_n$) in proteolysis is mostly reduced to the P_1 amino acid residue. Peptide bond formation is, in contrast to proteolysis, a two-substrate reaction that not only requires a specificity-dependent insertion of the carboxy component into the S subsites of the active site, but also optimal binding of the amino component in the S' region. Therefore, the $S'-P'$ interactions essentially determine the outcome of the C–N ligation reaction.

As mentioned above, the equilibrium of a protease-catalyzed reaction normally lies on the side of the thermodynamically more stable cleavage products. Several mechanistically different manipulations are possible to shift the equilibrium in favor of peptide bond formation. Approaches towards protease-catalyzed peptide bond formation are generally classified according to the type of carboxy component used (Figure 4.53).

In the equilibrium-controlled approach (Figure 4.53A), the carboxy component contains a free carboxy group, while in the kinetically controlled approach (Figure 4.53B) the carboxy component is employed in activated form, mainly as an alkyl ester. Both strategies are fundamentally different due to the energy required for the conversion of the starting components into the peptide products.

Equilibrium-controlled synthesis This equilibrium-controlled or thermodynamic approach (Figure 4.53A) represents the direct reversal of proteolysis. Consequently, all proteases can be used – independent of their mechanisms. Besides this advantage, the large amount of enzyme required and the low reaction rate are drawbacks of this approach. The ionization equilibrium must be manipulated to shift the equilibrium towards peptide bond formation. Addition of water-miscible organic solvents to the aqueous reaction mixture is a viable approach, however, the catalytic activity of the enzyme may decrease under such conditions. The lower dielectric constant of the

medium reduces the acidity of the carboxy group and, to a lesser extent, the basicity of the amino group of the nucleophilic amino component. In addition to the manipulations mentioned above, reverse micelles, anhydrous media containing minimal water concentrations, water mimics, and reaction conditions promoting product precipitation are often employed.

Kinetically controlled synthesis In contrast to the equilibrium-controlled approach, the kinetically controlled protease-catalyzed peptide synthesis [513] requires much less enzyme, the reaction time to obtain maximum product yield is significantly shorter, and the product yield depends both on the properties of the enzyme used and its substrate specificity. Whereas the equilibrium-controlled approach ends up in equilibrium (Figure 4.53A), the concentration of the product formed in the kinetic approach passes a maximum before the slower hydrolysis of the product becomes important. Amidase activity of most proteases is lower than esterase activity. The product inevitably is hydrolyzed, if the reaction is not stopped once the acyl donor ester is consumed. Figure 4.54 indicates the kinetic model of a protease-catalyzed acyl transfer reaction.

The kinetic approach requires an acyl donor ester (Ac-X) as the carboxy component, and is limited to proteases that rapidly form an acyl enzyme intermediate (Ac-E), for example serine and cysteine proteases. The enzyme (EH) acts as a transferase catalyzing the transfer of the acyl moiety to the nucleophilic amino component (HN). The amino component reacts, in competition with water, with the acyl enzyme to form the peptide (Ac-N). The ratio between aminolysis and hydrolysis is of decisive importance for successful preparative peptide synthesis. However, serine and cysteine proteases are not perfect acyltransferases and undesired reactions may take place due to their limited specificity (e.g., hydrolysis of the acyl enzyme, secondary hydrolysis of the ligation product and proteolytic cleavage of other protease-labile peptide bonds present in the segments to be ligated). An understanding of the molecular interactions between the acyl enzyme and the attacking nucleophilic amino component allows optimization of the acyl transfer efficiency. Efficient nucleophilic attack of the amino component essentially depends on optimal

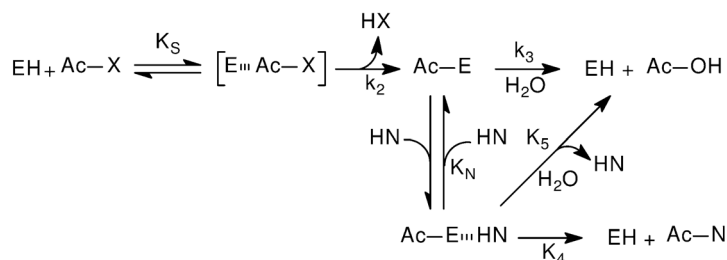


Figure 4.54 Kinetic model for protease-catalyzed acyl transfer reaction. EH = free enzyme; Ac-X = acyl donor (carboxy component); HX = leaving group; Ac-E = acyl enzyme complex; Ac-OH = hydrolysis product; HN = acyl acceptor (amino component); Ac-N = aminolysis product (synthesized peptide).

binding within the active site by S' – P' interactions. More information on the specificity of the S' subsites of serine and cysteine proteases has been obtained by systematic acyl transfer studies with libraries of nucleophilic amino components, known as S' subsite mapping [513].

Esterase activity can be further positively manipulated by varying the kind of leaving group. For preparative peptide synthesis, such manipulation is very important as it allows complete conversion of the acyl donor ester before the product is hydrolyzed. There is no doubt that the course of kinetically controlled protease-catalyzed peptide synthesis can be more efficiently influenced than the equilibrium approach.

In contrast to the chemical approach, peptide synthesis using stepwise enzymatic chain assembly may start either from the N-terminus (cf. Figure 5.2) or the C-terminus, because enzyme-catalyzed couplings are devoid of racemization. However, stepwise protease-catalyzed assembly should not be directly compared to automatic solid-phase technique (cf. Section 4.5). Selected di- and tripeptides (cf. Section 9.2), which are important intermediates in many pharmaceutical products, can be synthesized enzymatically on a large scale even in a continuous process, using solubilizing protecting groups [514–516].

In addition, the application of cross-linked enzyme crystals (CLECs) of thermolysin [517] and subtilisin [518] has been recommended for the synthesis of small peptides and peptide alkylamides. As will be shown below, solid-to-solid conversion has been proven as a very useful method for the synthesis of selected short peptides that comply with the requirements for this special synthetic procedure. Methodological support for practical application of protease-catalyzed peptide synthesis is available [519].

Although the kinetic approach should be preferred, the final decision depends on the overall synthesis concept. The largest industrial-scale application of the equilibrium approach is most likely the enzymatic synthesis of Z-Asp-Phe-OMe, the precursor of the peptide sweetener aspartame [520]. Transpeptidation technology on a large industrial scale is well documented for the conversion of porcine insulin into human insulin by trypsin [521] – a process that was widely used before the gene technology production of insulin.

4.6.2.2 Manipulations to Suppress Competitive Reactions

The most important factors limiting the widespread routine application of proteases in kinetically controlled peptide synthesis are hydrolysis of the acyl donor ester and proteolysis within the starting fragments to be coupled, or the final condensation product. These undesired reactions can be minimized by various manipulations of the reaction medium, the enzyme, the substrate, and on mechanistic features of the process.

Medium engineering is based on the replacement of water by organic solvents, and uses mono- or biphasic aqueous-organic and organic monophasic mixtures or microaqueous systems [522]. Competing hydrolysis can be minimized by decreasing the water concentration in the reaction medium of enzymatic synthesis in microaqueous systems that do not employ organic solvents, termed solid-to-solid conversion [523–525], and in frozen aqueous systems [526].

Enzyme engineering describes a range of techniques, from deliberate chemical modification to genetic remodeling of a wild-type enzyme [527]. A mutant of subtilisin BNP', termed "subtiligase", was obtained by Jackson *et al.* [528] by protein design and used in a further synthesis of RNase A (cf. Section 5.2.2.2, Figure 5.9).

Substrate engineering has been verified in irreversible C–N ligations by mimicking enzyme specificity. The substrate mimetic approach will be discussed separately in Section 4.6.2.3 because of its significance in the field of protease-catalyzed peptide synthesis.

4.6.2.3 Substrate Mimetic Approach

Despite the development of various possibilities to suppress competing reactions, proteolytic cleavage of the peptide bond formed cannot be completely suppressed. In this context, it must be considered that the newly formed peptide can be cleaved during the course of the catalytic process by the same enzyme. The peptide bond-forming step and the proteolytic cleavage step do not differ in substrate specificity requirements. Since the specificity of proteases is manifested by the side chain of the P₁ amino acid residue, an irreversible peptide bond formation does not seem possible. In ribosomal peptide synthesis (cf. Section 3.2.1), the selection of the amino acid residues to be coupled is performed as a specific recognition process prior to the acyl transfer reaction, catalyzed by the unspecific ribozyme peptidyl transferase.

An irreversible C–N ligation in a biomimetic fashion would be possible also in enzymatic peptide synthesis if the specificity selection could be separated from the actual peptide bond-forming step. This proved to be feasible by transferring the moiety that determines the specificity from the P₁ amino acid side chain to the leaving group of the acyl donor ester. Hence, the enzyme is capable of recognizing, in a kinetically controlled acyl transfer reaction, the acyl donor ester by the specificity determinant present in the leaving group. The specificity-bearing leaving group is released upon formation of the acyl enzyme followed by peptide bond-forming aminolysis. Consequently, the newly formed peptide bond can no longer be cleaved by the enzyme, because the moiety responsible for recognition by the protease is then lacking. This irreversible synthesis concept was for the first time confirmed in a model peptide synthesis catalyzed by trypsin involving various nonspecific N^α-protected amino acid 4-guanidinophenyl esters (OGp) as acyl donors, and various amino acid and peptide derivatives as nucleophilic acyl acceptors. Trypsin accepts the guanidino group of the 4-guanidinophenyl ester as a substitute for the arginine residue that is usually recognized by the S₁ subsite [529]. This concept was exploited in further examples by another group, who named this type of acyl donor "ester inverse substrate" [530].

The more correct and nowadays used term 'substrate mimetic' was introduced by Jakubke and coworkers [531], together with an extension of this approach to irreversible peptide segment condensations with other proteases. The substrate-mimetic approach represents a very specific and efficient form of substrate engineering. The mechanism of the substrate mimetic concept has been later elucidated by Bordusa's group [510, 532]. As shown in Figure 4.55, a common acyl donor A binds its acyl residue to the S-binding site, where the leaving group is present at the S'₁

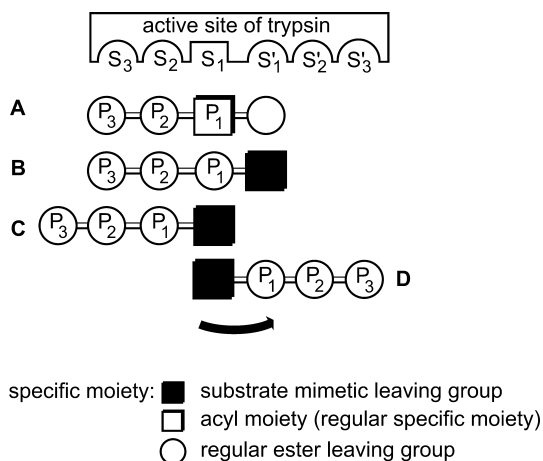


Figure 4.55 Comparison of the binding of a common acyl donor ester (A) and a substrate mimetic in different binding positions (B–C) to the active site of trypsin with the hypothetical binding mode of a substrate mimetic (D).

subsite. The scissile ester bond may be attacked by Ser¹⁹⁵ in the case of trypsin, followed by acylation of the enzyme. However, applying the same binding principle, a substrate mimetic with the leaving group located either in S'₁ (B) or in S₁ (C) leads to catalytically unproductive binding. Recognition of the specificity determinant in the leaving group by the S₁ subsite fails in binding arrangement B, whereas in the binding position C the scissile ester bond would be far away from the active Ser¹⁹⁵ and neither hydrolysis nor acylation would occur. Hence, the binding of a substrate mimetic and the subsequent acylation leads to an acyl enzyme having the acyl residues at the S'-subsites (D). Prior to the deacylation step, the acyl residue has to flip to the S-subsite of the enzyme. The common kinetic model has to be extended by a rearrangement step between the two acyl enzyme species E–Ac, with the acyl moiety in the S' subsite and the common acyl enzyme Ac–E. Peptide nucleophiles that bind specifically to the S' subsite should be able to react with the acyl moiety and cleave it from the S' region more efficiently than water, which decreases unwanted hydrolysis of the acyl enzyme.

In summary, cationic, anionic, and hydrophobic types of substrate mimetics have been designed, characterized and applied for various peptides synthesis, including peptide [533] and peptoid [534] segment condensations. The C–N ligation strategy based on the substrate mimetic concept allows, for the first time, an irreversible peptide bond formation catalyzed by proteases. This specific programming of enzyme specificity by molecular mimicry corresponds in practice to a conversion of a protease into a C–N ligase, and results in a biocatalyst which has not been developed evolutionarily by nature. When combined with frozen-state enzymology (as demonstrated in model reactions [535]) or with protease mutants lacking amidase activity, the substrate mimetics C–N ligation approach will undoubtedly contribute to significant progress in enzymatic peptide synthesis. Bearing in mind that proteases

have already proven their value for organic synthesis, even up to metric ton scale, and including the fact that these enzymes are generally recognized as normal bench reagents for reactions based on their native hydrolytic activity, a general approach to the reverse enzymatic polypeptide synthesis remains to be formulated. However, by building on promising strategies, such as the substrate mimetic approach, and site-specific chemically or ionic liquid-modified enzymes, significant improvements have been made, and more can be expected in the near future. The final breakthrough may be reached by using a combination of these strategies. Additional input can be expected from the combination of enzyme engineering with site-directed and, in particular, with directed evolution techniques that currently appear to be the most fertile approaches to the design of enzymes with tailored selectivities and synthetically relevant activities in essentially any suitable reaction medium [536, 537]. The substrate mimetic approach combined with expressed protein ligation led to expressed enzymatic ligation (EEL) (cf. Section 5.5.5.3) underlining the importance of this useful variant of protease-catalyzed peptide synthesis.

4.6.3

Further Selected Biochemical Methods

4.6.3.1 Non-ribosomal Peptide Synthesis

The application of the *non-ribosomal peptide synthesis* (NRPS) apparatus (cf. Section 3.2.3) displays certain advantages in biotechnological processes, including peptide bond formation in water, a lack of any protection needs for the amino acid building blocks, and epimerization domains that enable selective incorporation of D-amino acids into the growing peptide chain, ultimately to obtain cyclic peptides [538].

4.6.3.2 Peptide Bond Formation by LF-Transferase

Recently, the molecular basis for the *non-ribosomal protein-based peptide bond formation by aminoacyl-tRNA protein transferase* has been reported [539]. Eubacterial leucyl/phenylalanyl-tRNA protein transferase (LF-transferase) catalyzes peptide bond formation between Leu-tRNA^{Leu} (or Phe-tRNA^{Phe}) and an N-terminal Arg (or Lys) of a protein acceptor substrate. Interestingly, the mechanism is similar to the reverse of the acylation step of proteolytic peptide bond cleavage catalyzed by serine proteases, which was originally proposed for the ribosome [540]. However, the peptide bond forming step on the ribosome proceeds in a different manner (cf. Section 3.2.1). In the mechanism of the LF-transferase-catalyzed reaction a protein-based chemical reaction is involved, in which the electron relay from Asp¹⁸⁶ to Gln¹⁸⁸ is the underlying step (Figure 4.56).

4.6.3.3 Antibody-catalyzed Peptide Bond Formation

Antibody-catalyzed peptide bond formation was first described by the groups of Hirschman and Schultz in the mid-90s. According to Linus Pauling [541], the action of an enzyme depends to a large extent on the complementarity between the active site of the enzyme and the transition state of the enzyme-catalyzed reaction, as shown schematically for the peptide bond-forming step (Figure 4.57).

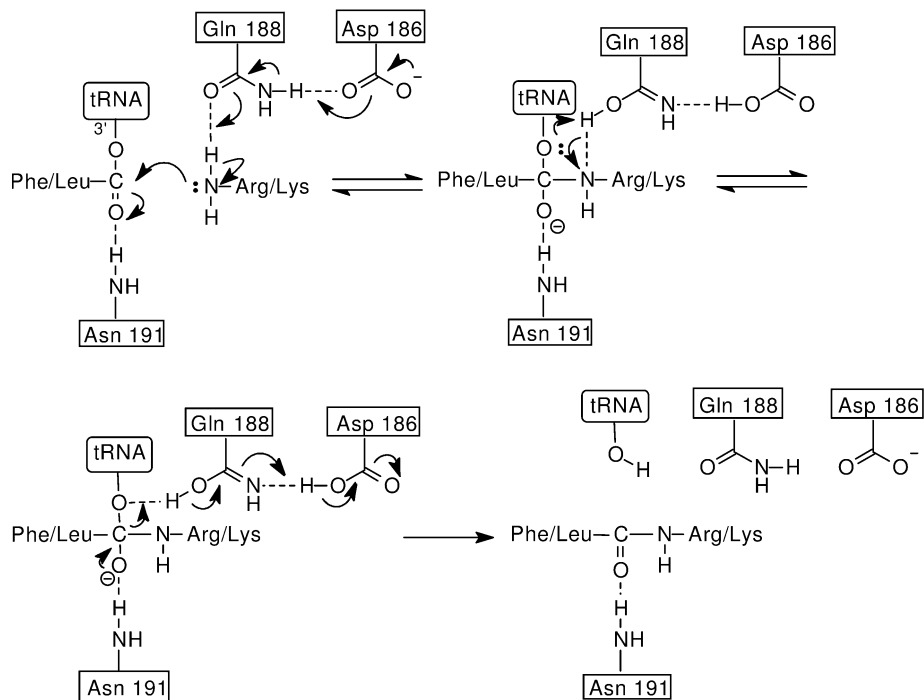


Figure 4.56 A proposed mechanism for the LF-transferase-catalyzed peptide bond formation [539].

In 1969, it was proposed by Jencks [542] that antibodies could be produced to function as catalysts. According to this idea, structures that mimic the stereoelectronic properties of transition states might be capable of eliciting antibodies that accelerate the corresponding chemical reaction. Consequently, if an antibody could

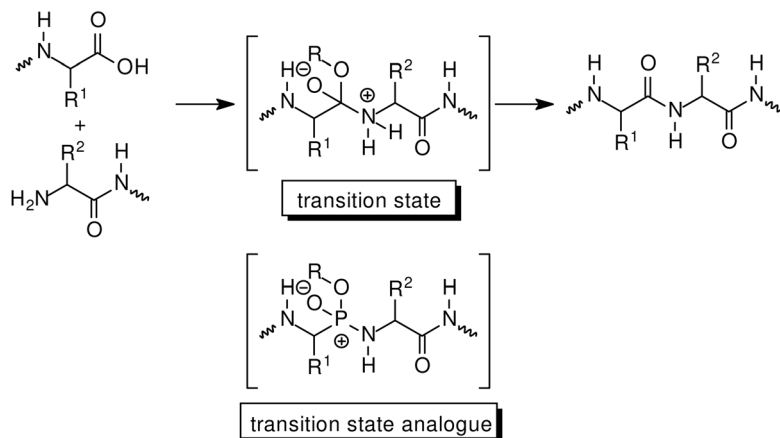


Figure 4.57 Comparison of the transition-state structure of peptide synthesis by ester aminolysis with the structure of a designed transition-state analogue.

bind a substrate already in the transition state conformation, it might act as an enzyme catalyzing the reaction to which the transition state conformation is predisposed. For this reason, catalytic antibodies have also been termed “abzymes”. Investigations concerning requirements for the catalysis of peptide bond formation using catalytic antibodies was the consequence of progress in antibody catalysis for organic transformations [543, 544].

The normal approach to generate catalytic antibodies requires the design of haptens that are topologically and structurally near-identical to the transition state of the appropriate reaction, as shown in Figure 4.57. In particular, peptide synthesis should require a large number of catalytic antibodies in order to consider the specificity requirements of the different amino acid residues in the C-terminal position of the carboxy component, and the N-terminal position of the amino component to be coupled.

As shown in Figure 4.58, Hirschmann *et al.* [545] synthesized the phosphonamide shown in Figure 4.58A as a transition-state analogue which was used to generate antibodies for catalyzing the model peptide synthesis (Figure 4.58B).

The cyclohexyl moiety of the transition-state analogue should create a binding site in the antibody that could accommodate various hydrophobic side-chain moieties of the carboxy component. The *p*-nitrobenzyl residue in the phosphonate was designed in order to facilitate the dissociation of both the leaving group and peptide from the antibody. The authors described antibody-catalyzed dipeptide syntheses starting from all possible stereoisomeric combinations of the ester and amide substrates in yields between 44% and 94%. Two monoclonal antibodies (16G3 and 18C10) were found to cause a ~220-fold acceleration of the reaction. Further studies with the so-called antibody ligase 16G3 have demonstrated that this abzyme catalyzes not only the synthesis of dipeptides but also the coupling of an activated amino acid with a dipeptide to form a tripeptide, as well as a simple

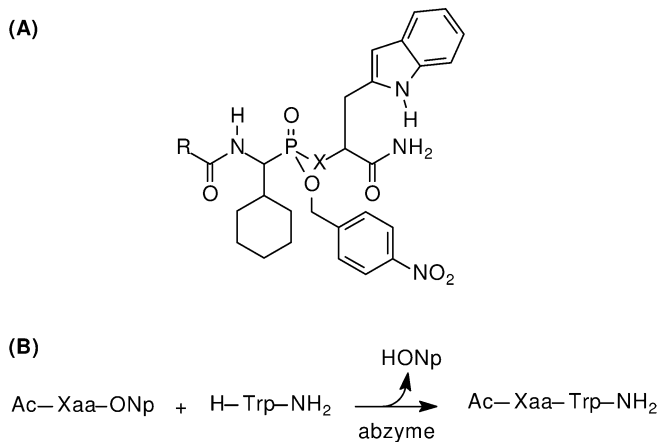


Figure 4.58 Structural variants of the transition-state analogue (A) for an abzyme-catalyzed model peptide synthesis, (B) Xaa = Val, Leu, Phe; X = NH; R = $-(\text{CH}_2)_3\text{COOH}$ /X = O; R = $-(\text{CH}_2)_3\text{COOH}$ /X = O; R = CH_3 .

(2 + 2) segment condensation with average yields of 80% within a reaction time of 20 min [546].

Jacobsen and Schultz [547] created an antibody against a neutral phosphonate diester transition-state analogue that was capable of catalyzing the reaction between *N*-phenylethoxycarbonyl-L-alanyl azide and *N*-(4-nitrobenzoyl)-*N'*-L-phenylalanyl-1,3-diaminopropane at a 10 000-fold rate relative to the uncatalyzed reaction, and in 65% yield.

Interestingly, in both cases the catalytic antibodies do not catalyze the hydrolysis of both the active ester and the azide to an appreciable degree. Product hydrolysis and racemization have not been observed. As mentioned earlier, the main disadvantage of the catalytic antibody approach is the requirement for a large number of antibody catalysts to accommodate the wide specificity pattern of amino acids in coupling reactions. On the other hand, these results provide an impetus for producing further generations of antibodies capable of catalyzing the ligation of longer unprotected fragments, combined with a general strategy for the development of sequence-specific antibody ligases.

4.7

Review Questions

- Q4.1. Explain what is meant by “discontinuous epitope”
- Q4.2. Why is it necessary to protect functional group of amino acids for chemical peptide synthesis?
- Q4.3. Name the requirements that a temporary protecting group has to fulfil.
- Q4.4. What is protecting group orthogonality?
- Q4.5. Why are acyl residues not suitable as N^α -protecting groups?
- Q4.6. Name three important N^α -protecting groups and their cleavage conditions and mechanisms.
- Q4.7. Classify protecting moieties for the carboxy group.
- Q4.8. Why are C-terminal hydrazides useful for further couplings?
- Q4.9. What is backbone protection and when is it required? Name different backbone protecting groups and their cleavage conditions.
- Q4.10. Write the mechanism for undesired lactam formation of carboxy-activated Arg.
- Q4.11. Name at least three protecting groups for the Arg guanidino group and comment on their lability/applicability for different synthesis conditions.
- Q4.12. How is it possible to selectively protect either the α - or the ϵ -amino group of Lys?
- Q4.13. What is the most prominent side reaction of Asp(OR) residues and Asn residues?
- Q4.14. What are the advantages of Bom and Bum as the N^{im} -protecting groups in His?
- Q4.15. Is it necessary to protect the thioether in Met and the amide in Asn/Gln?
- Q4.16. Outline the application of different anhydrides in peptide coupling.

- Q4.17. Give three examples of carbodiimides used as coupling reagents. What is the mechanism of carbodiimide activation of the carboxy group? Which side reactions may occur?
- Q4.18. What is an active ester? Why is it reactive towards nucleophiles?
- Q4.19. Discuss the role of HOBt and HOAt in peptide coupling reactions.
- Q4.20. Why have acid halides been considered as unsuitable for carboxy-activation in peptide synthesis? Which acid halides are nowadays being used for that purpose?
- Q4.21. What is the disadvantage of BOP in peptide coupling? What are the structural differences between BOP, PyBOP, and PyAOP?
- Q4.22. Write the structures of HBTU and HATU? What was the unexpected result of their structure elucidation?
- Q4.23. How does peptide bond formation mediated by HATU proceed?
- Q4.24. Which functional group is generated upon treatment of a carboxylic acid with DPPA?
- Q4.25. Discuss the concept of racemization/epimerization. Why is it undesired? How does it occur? How can it be suppressed?
- Q4.26. Explain the concept of SPPS.
- Q4.27. Name some resin materials. Why are the swelling properties of a resin important?
- Q4.28. Name three SPPS linkers with different acid lability.
- Q4.29. Explain the concept of safety-catch resins. Name some examples.
- Q4.30. What are Merrifield tactics?
- Q4.31. What are Sheppard tactics?
- Q4.32. Give a definition of “core sequence” and “mismatch sequence”.
- Q4.33. What are random and non-random difficult sequences?
- Q4.34. How can the problem of synthesizing difficult sequences be solved?
- Q4.35. How can peptides be obtained by recombinant synthesis?
- Q4.36. What is cell-free translation?
- Q4.37. How can proteins with non-encoded amino acids be obtained?
- Q4.38. What enzymes are used in enzymatic peptide synthesis?
- Q4.39. What is the difference between the equilibrium approach and the kinetic approach in enzymatic peptide synthesis?
- Q4.40. Explain the substrate mimetic approach.
- Q4.41. What is an abzyme?

References

1 F. Albericio, <i>Biopolymers</i> 2000 , 55, 123.	5 A. Patchornik, B. Amit, R.B. Woodward, <i>J. Am. Chem. Soc.</i> 1970 , 92, 6333.
2 M. Bergmann, L. Zervas, <i>Chem. Ber.</i> 1932 , 65, 1192.	6 D.T. Gish, F.H. Carpenter, <i>J. Am. Chem. Soc.</i> 1953 , 75, 950.
3 S.C. McKay, N.F. Albertson, <i>J. Am. Chem. Soc.</i> 1957 , 79, 4686.	7 F.H. Carpenter, D.T. Gish, <i>J. Am. Chem. Soc.</i> 1952 , 74, 3818.
4 F. Weygand, E. Nintz, <i>Z. Naturforsch. B</i> 1965 , 20, 429.	8 R.A. Boissonnas, G. Preitner, <i>Helv. Chim. Acta</i> 1955 , 36, 875.

- 9 L. Kisfaludy, S. Dualszky, *Acta Chim. Acad. Sci. Hung.* **1960**, 24, 301.
- 10 K. Blaha, J. Rudinger, *Coll. Czech. Chem. Commun.* **1965**, 30, 985.
- 11 N. Izumiya, K. Noda, S. Terada, *Bull. Chem. Soc. Japan* **1970**, 43, 1883.
- 12 B.W. Erickson, R.B. Merrifield, *J. Am. Chem. Soc.* **1973**, 95, 3757.
- 13 J.W. Chamberlin, *J. Org. Chem.* **1966**, 31, 1658.
- 14 C. Birr, W. Lochinger, G. Stahnke, P. Lang, *Ann. Chem.* **1972**, 763, 162.
- 15 R. Schwyzler, *Angew. Chem.* **1959**, 71, 742.
- 16 G.R. Matsueda, J.M. Stewart, in: *Peptides: Chemistry, Structure and Biology*, R. Walter, J. Meienhofer (Eds.), Ann Arbor Scientific Publishers, Ann Arbor, **1975**.
- 17 D.S. Kemp, C.F. Hoyng, *Tetrahedron Lett.* **1975**, 4625.
- 18 P. Sieber, B. Iselin, *Helv. Chim. Acta* **1968**, 51, 622.
- 19 D. Wildemann, M. Drewello, G. Fischer, M. Schutkowski, *J. Chem. Soc., Chem. Commun.* **1999**, 1809.
- 20 A. Tun-Kyi, R. Schwyzler, *Helv. Chim. Acta* **1976**, 59, 1642.
- 21 D.F. Veber, W.J. Paleveda Jr., Y.C. Lee, R. Hirschmann, *J. Org. Chem.* **1977**, 42, 3286.
- 22 G.W. Anderson, A.C. McGregor, *J. Am. Chem. Soc.* **1957**, 79, 6180.
- 23 E. Wünsch, R. Spangenberg, *Chem. Ber.* **1971**, 104, 2427.
- 24 H. Eckert, M. Listl, I. Ugi, *Angew. Chem. Int. Ed.* **1978**, 17, 361.
- 25 W.L. Haas, E.V. Krumkalns, K. Gerzon, *J. Am. Chem. Soc.* **1966**, 88, 1988.
- 26 H. Kalbacher, W. Voelter, *Angew. Chem. Int. Ed.* **1978**, 17, 944.
- 27 G. Jäger, R. Geiger, in: *Peptides 1971*, H. Nesvadba (Ed.), North-Holland, Amsterdam, **1971**, 78.
- 28 L.A. Carpino, *Acc. Chem. Res.* **1987**, 20, 401.
- 29 L.A. Carpino, G.Y. Han, *J. Am. Chem. Soc.* **1970**, 92, 5748.
- 30 B. Henkel, E. Bayer, *J. Peptide Sci.* **2001**, 7, 152.
- 31 A.T. Kader, C.J.M. Stirling, *J. Chem. Soc.* **1964**, 258.
- 32 E.T. Wolters, G.I. Tesser, R.J.F. Nivard, *J. Org. Chem.* **1974**, 39, 3388.
- 33 C.G.J. Verhart, G.I. Tesser, *Recl. Trav. Chim. Pays-Bas* **1988**, 107, 621.
- 34 V.V. Samukov, A.N. Sabirov, P.I. Pozdnyakov, *Tetrahedron Lett.* **1994**, 35, 7821.
- 35 A.N. Sabirov, Y.D. Kim, H.J. Kim, V.V. Samukov, *Protein Peptide Lett.* **1997**, 4, 307.
- 36 L.A. Carpino, M. Philbin, *J. Org. Chem.* **1999**, 64, 4315.
- 37 L.A. Carpino, M. Philbin, M. Ismail, G.A. Truran, E.M.E. Mansour, S. Iguchi, D. Ionescu, A. El-Faham, C. Riemer, R. Warrass, M.S. Weiss, *J. Am. Chem. Soc.* **1997**, 119, 9915.
- 38 L.A. Carpino, M. Ismail, G.A. Truran, E.M.E. Mansour, S. Iguchi, D. Ionescu, A. El-Faham, C. Riemer, R. Warrass, *J. Org. Chem.* **1999**, 64, 4324.
- 39 L.A. Carpino, E.M.E. Mansour, *J. Org. Chem.* **1999**, 64, 8399.
- 40 C.M. Stevens, R. Watanabe, *J. Am. Chem. Soc.* **1950**, 72, 725.
- 41 A. Loffet, H.X. Zhang, *Int. J. Peptide Protein Res.* **1993**, 42, 346.
- 42 F. Guibe, *Tetrahedron* **1998**, 54, 2967.
- 43 L.A. Carpino, A.C. Sam, *J. Chem. Soc., Chem. Commun.* **1979**, 514.
- 44 B.H. Lipshutz, P. Papa, J.M. Keith, *J. Org. Chem.* **1999**, 64, 3792.
- 45 D. Stevenson, G.T. Young, *J. Chem. Soc., Chem. Commun.* **1967**, 900.
- 46 P. Wang, R. Layfield, M. Landon, R.J. Mayer, R. Ramage, *Tetrahedron Lett.* **1998**, 39, 8711.
- 47 L.A. Carpino, H.N. Parameswaran, R.K. Kirkley, J.W. Spiewak, E. Schmitz, *J. Org. Chem.* **1970**, 35, 3291.
- 48 H.A. Moynihan, W.P. Yu, *Synth. Commun.* **1998**, 28, 17.
- 49 S. Routier, L. Sauge, N. Ayerbe, G. Coudert, J.-Y. Merour, *Tetrahedron Lett.* **2002**, 43, 589.
- 50 R. Pascal, R. Sola, *Tetrahedron Lett.* **1998**, 39, 5031.

- 51 A. Isidro-Llobet, X. Just-Baringo, A. Ewenson, M. Alvarez, F. Albericio, *Biopolymers (Peptide Sci.)* **2007**, *88*, 733.
- 52 P. Wessig, S. Czaplá, K. Möllnitz, J. Schwarz, **2006**, *Synlett* 2235.
- 53 R. Ramage, L. Jiang, Y.-D. Kim, K. Shaw, J.-L. Park, H.-J. Kim, *J. Peptide Sci.* **1999**, *5*, 195.
- 54 C. Carreno, M.E. Mendz, Y.D. Kim, H.-J. Kim, S.A. Kates, D. Andreu, F. Albericio, *J. Peptide Res.* **2000**, *56*, 63.
- 55 P.M. Balse, H.J. Kim, G. Han, V.J. Hruby, *J. Peptide Res.* **2000**, *56*, 70.
- 56 P. Gmez-Martnez, M. Dessolin, F. Guib, F. Albericio, *J. Chem. Soc., Perkin Trans. I* **1999**, 2871.
- 57 N. Thieriet, P. Gmez-Martnez, F. Guib, *Tetrahedron Lett.* **1999**, *40*, 2505.
- 58 G. Barany, R.B. Merrifield, *J. Am. Chem. Soc.* **1977**, *99*, 7363.
- 59 G. Barany, F. Albericio, *J. Am. Chem. Soc.* **1985**, *107*, 4936.
- 60 M. Planas, E. Bardaji, K.J. Jensen, G. Barany, *J. Org. Chem.* **1999**, *64*, 7281.
- 61 J.S. Choi, H. Kang, N. Jeong, H. Han, *Tetrahedron* **2005**, *61*, 2493.
- 62 S.C. Miller, T.S. Scanlan, *J. Am. Chem. Soc.* **1997**, *119*, 2301.
- 63 E. Vedejs, C. Kongkittingam, *J. Org. Chem.* **2000**, *65*, 2309.
- 64 S.C. Miller, T.S. Scanlan, *J. Am. Chem. Soc.* **1998**, *120*, 2690.
- 65 L.A. Carpino, D. Ionescu, A. El-Faham, P. Henklein, H. Wenschuh, M. Bienert, M. Beyermann, *Tetrahedron Lett.* **1998**, *39*, 241.
- 66 T. Fukuyama, C.K. Jow, M. Cheung, *Tetrahedron Lett.* **1995**, *36*, 6373.
- 67 T. Fukuyama, M. Cheung, C.K. Jow, Y. Hidai, K. Toshiyuki, *Tetrahedron Lett.* **1997**, *38*, 5831.
- 68 L. Zervas, D. Borovas, E. Gazis, *J. Am. Chem. Soc.* **1963**, *85*, 3660.
- 69 E. Wünsch, R. Spangenberg, *Chem. Ber.* **1972**, *105*, 740.
- 70 A. Tun-Kyi, *Helv. Chim. Acta* **1978**, *61*, 1086.
- 71 G.B. Bloomberg, G. Askin, A.R. Gargaro, M.J. Tanner, *Tetrahedron Lett.* **1993**, *36*, 4709.
- 72 P. Dumy, I.M. Eggleston, S. Cervigni, U. Sila, X. Sun, M. Mutter, *Tetrahedron Lett.* **1995**, *36*, 1255.
- 73 L.A. Chhabra, B. Hothi, D.J. Evans, P.D. White, B.W. Bycroft, W.C. Chan, *Tetrahedron Lett.* **1998**, *39*, 1603.
- 74 B. Kellam, B.W. Bycroft, W.C. Chan, S.R. Chhabra, *Tetrahedron* **1998**, *54*, 6817.
- 75 M.G. Ryadnov, L.V. Klimenko, Y.V. Mitin, *J. Peptide Res.* **1999**, *53*, 322.
- 76 N. Thieriet, F. Guib, F. Albericio, *Organic Lett.* **2000**, *2*, 1815.
- 77 A. Joansson, E. Åkerblom, K. Ersmark, G. Lindeberg, A. Hallberg, *J. Comb. Chem.* **2000**, *2*, 496.
- 78 M. Brenner, W. Huber, *Helv. Chim. Acta* **1953**, *36*, 1109.
- 79 J.R. McDermott, N.L. Benoiton, *Can. J. Chem.* **1973**, *51*, 2555.
- 80 B. Löhr, S. Orlich, H. Kunz, *Synlett* **1999**, 1136.
- 81 H. Kessler, R. Siegmeyer, *Tetrahedron Lett.* **1983**, *24*, 281.
- 82 S.A.M. Merette, A.P. Burd, J.J. Deadman, *Tetrahedron Lett.* **1999**, *40*, 753.
- 83 P. Sieber, *Helv. Chim. Acta* **1977**, *60*, 2711.
- 84 H. Gerlach, *Helv. Chim. Acta* **1977**, *60*, 3039.
- 85 M. Wagner, H. Kunz, *Synlett* **2000**, 400.
- 86 L.A. Carpino, H.-G. Chao, S. Ghassemi, E.M.E. Mansour, C. Riemer, R. Warrass, D. Sadat-Aalae, G.A. Truran, H. Imazumi, A. El-Faham, D. Ionescu, M. Ismail, T.L. Kowaleski, C.H. Han, H. Wenschuh, M. Beyermann, M. Bienert, H. Shroff, F. Albericio, S.A. Triola, M.A. Soile, S.A. Kates, *J. Org. Chem.* **1995**, *60*, 7718.
- 87 W.C. Chan, B.W. Bycroft, D.J. Evans, P.D. White, *J. Chem. Soc., Chem. Commun.* **1995**, *21*, 2209.
- 88 B. Briand, N. Kotzur, V. Hagen, M. Beyermann, *Tetrahedron Lett.* **2008**, *49*, 85.
- 89 M. Goodman, A. Felix, L. Moroder, C. Toniolo, *Synthesis of Peptides and Peptidomimetics in: Houben-Weyl, Methoden der organischen Chemie*,

- Vol. E22a, K.H. Büchel (Ed.), Thieme, Stuttgart, 2002.
- 90 H.-D. Jakubke, H. Jeschkeit, *Aminosäuren, Peptide, Proteine*, Verlag Chemie, Weinheim, 1982.
- 91 M. Bodanszky, A. Bodanszky, *The Practice of Peptide Synthesis*, Springer Verlag, Berlin, 1994.
- 92 T.W. Greene, P.G.M. Wuts, *Protective Groups in Organic Synthesis*, Wiley, New York, 1999.
- 93 G.N. Müller, H. Waldmann, *Tetrahedron Lett.* **1999**, 40, 3549.
- 94 F. Weygand, W. Steglich, J. Bjarnason, *Chem. Ber.* **1968**, 101, 3642.
- 95 M. Narita, M. Doi, H. Sugawara, K. Ishikawa, *Bull. Chem. Soc. Jpn.* **1985**, 58, 1473.
- 96 R. Bartl, K.-D. Klöppel, R. Frank, in: *Peptides 1992*, C.H. Schneider, A.N. Eberle (Eds.), Escom, Leiden, **1993**.
- 97 C. Hyde, T. Johnson, D. Owen, M. Quibell, R.C. Sheppard, *Int. J. Peptide Protein Res.* **1994**, 43, 431.
- 98 T. Johnson, M. Quibell, R.C. Sheppard, *J. Peptide Sci.* **1995**, 1, 11.
- 99 L.P. Miranda, W.D.F. Meutermans, M.L. Smythe, P.F. Alewood, *J. Org. Chem.* **2000**, 65, 5460.
- 100 P. May, M. Bradley, D.C. Harrowven, D. Pallin, *Tetrahedron Lett.* **2000**, 41, 1627.
- 101 T. Haack, M. Mutter, *Tetrahedron Lett.* **1992**, 33, 1589.
- 102 T. Wöhr, M. Mutter, *Tetrahedron Lett.* **1995**, 36, 3847.
- 103 T. Wöhr, F. Wahl, A. Nefzi, B. Rohwedder, T. Sato, X. Sun, *J. Am. Chem. Soc.* **1996**, 118, 9218.
- 104 W.R. Sampson, H. Patsiouras, N.J. Ede, *J. Peptide Sci.* **1999**, 5, 403.
- 105 S. Bajusz, *Acta Chim. Acad. Sci. Hung.* **1965**, 44, 23.
- 106 G. Jäger, R. Geiger, *Chem. Ber.* **1970**, 103, 1727.
- 107 H. Rink, P. Sieber, F. Raschdorf, *Tetrahedron Lett.* **1984**, 25, 621.
- 108 O. Nishimura, M. Fujino, *Chem. Pharm. Bull.* **1976**, 24, 1568.
- 109 H. Yajima, M. Takeyama, J. Kanaki, K. Mitani, *J. Chem. Soc., Chem. Commun.* **1978**, 482.
- 110 M. Fujino, O. Nishimura, M. Wakimazu, C. Kitada, *J. Chem. Soc., Chem. Commun.* **1980**, 668.
- 111 R. Ramage, J. Green, A.J. Blake, *Tetrahedron* **1991**, 47, 6353.
- 112 L.A. Carpino, H. Shroff, S.A. Triolo, E.-S.M.E. Mansour, H. Wenschuh, F. Albericio, *Tetrahedron Lett.* **1993**, 34, 7829.
- 113 C.G. Fields, G.B. Fields, *Tetrahedron Lett.* **1993**, 34, 6661.
- 114 H.B. Arzeno, D.S. Kemp, *Synthesis* **1988**, 32.
- 115 T. Johnson, R.C. Sheppard, R. Valerio, in: *Peptides 1990* E. Giralt, D. Andreu (Eds.), Escom, Leiden, **1991**, 34.
- 116 M. Bodanszky, J.C. Tolle, S.S. Bdeshtane, A. Bodanszky, *Int. J. Peptide Protein Res.* **1978**, 12, 57.
- 117 R. Dölling, M. Beyermann, J. Haenel, F. Kernchen, E. Krause, M. Brudel, M. Bienert, *J. Chem. Soc., Chem. Commun.* **1994**, 853.
- 118 M. Quibell, D. Owen, L.C. Packman, T. Johnson, *J. Chem. Soc., Chem. Commun.* **1994**, 2343.
- 119 J.L. Lauer, C.G. Fields, G.B. Fields, *Lett. Peptide Sci.* **1994**, 1, 197.
- 120 J.-L. Fauchre, R. Schwyzler, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 3, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1981**.
- 121 A.H. Karlström, A.E. Uden, *Tetrahedron Lett.* **1995**, 36, 3909.
- 122 R.G. Hiskey, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 3, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1981**, 137.
- 123 V. Du Vigneaud, L.F. Audrieth, H.S. Loring, *J. Am. Chem. Soc.* **1930**, 52, 4500.
- 124 N. Fujii, A. Otaka, S. Funakoshi, K. Bessho, T. Watanabe, K. Akaji, H. Yajima, *Chem. Pharm. Bull.* **1987**, 35, 2339.

- 125 S. Akabori, S. Sakakibara, Y. Shimonishi, Y. Nobuhara, *Bull. Chem. Soc. Jpn.* **1964**, 37, 433.
- 126 M.C. Munson, C. Garcia-Echeverria, F. Albericio, G. Barany, *J. Org. Chem.* **1992**, 57, 3013.
- 127 L. Zervas, I. Photaki, *Chimia* **1960**, 14, 375.
- 128 L. Zervas, I. Photaki, *J. Am. Chem. Soc.* **1962**, 84, 3887.
- 129 G. Amiard, R. Heyms, L. Velluz, *Bull. Soc. Chim. Fr.* **1956**, 192, 698.
- 130 I. Photaki, J. Taylor-Papadimitriou, C. Sakarellos, P. Mazarakis, L. Zervas, *J. Chem. Soc., Chem. Commun.* **1970**, 2683.
- 131 P. Sieber, B. Kamber, A. Hartmann, A. Jöhl, B. Riniker, W. Rittel, *Helv. Chim. Acta* **1974**, 57, 2617.
- 132 S.N. McCurdy, *Peptide Res.* **1989**, 2, 147.
- 133 A. Chimiak, in: *Peptides* 1963, G.T. Young (Ed.), Pergamon Press, Oxford, **1963**, 37.
- 134 D.F. Veber, J.D. Milkowski, R.G. Denkwalter, R. Hirschmann, *Tetrahedron Lett.* **1968**, 3057.
- 135 D.F. Veber, J.D. Milkowski, S.L. Varga, R.G. Denkwalter, R. Hirschmann, *J. Am. Chem. Soc.* **1972**, 94, 5456.
- 136 Y. Kiso, M. Yoshida, Y. Fujiwara, T. Kumura, M. Shimokura, K. Akaji, *Chem. Pharm. Bull.* **1990**, 38, 673.
- 137 E. Wünsch, R. Spangenberg, in: *Peptides* 1969, E. Scoffone (Ed.), North-Holland, Amsterdam, **1971**, 30.
- 138 M. Bodanszky, M.A. Bednarek, *Int. J. Pept. Protein Res.* **1982**, 20, 434.
- 139 E. Wünsch, R. Spangenberg, in: *Peptides* 1969, E. Scoffone (Ed.), North-Holland, Amsterdam, **1971**, 30.
- 140 U. Weber, P. Hartter, *Hoppe-Seylers Z. Physiol. Chem.* **1970**, 351, 1384.
- 141 E. Atherton, M. Pinori, R.C. Sheppard, *J. Chem. Soc., Perkin Trans I* **1985**, 2057.
- 142 R. Eritja, J.P. Ziehler-Martin, P.A. Walker, T.D. Lee, K. Legesse, F. Albericio, B. Kaplan, *Tetrahedron* **1987**, 43, 2675.
- 143 R. Matsueda, R. Walter, *Int. J. Peptide Protein Res.* **1980**, 16, 392.
- 144 A.M. Kimbonguila, A. Merzouk, F. Guib, A. Loffet, *Tetrahedron* **1999**, 55, 6931.
- 145 Y. Han, G. Barany, *J. Org. Chem.* **1997**, 62, 3841.
- 146 H.J. Musiol, F. Siedler, D. Quarzago, L. Moroder, *Biopolymers* **1994**, 34, 1553.
- 147 Y. Han, F. Albericio, G. Barany, *J. Org. Chem.* **1997**, 62, 4307.
- 148 T. Kaiser, G.J. Nicholson, H.J. Kohlbau, W. Voelter, *Tetrahedron Lett.* **1996**, 37, 1187.
- 149 Y.X. Han, *J. Org. Chem.* **1997**, 62, 4307.
- 150 J.J. Pastuszak, A. Chimiak *J. Org. Chem.* **1981**, 46, 1868.
- 151 B. Kamber, W. Rittel, *Helv. Chim. Acta* **1968**, 51, 2061.
- 152 R.G. Hiskey, F.I. Carroll, R.M. Babb, J.O. Bledsoe, R.T. Puckett, B.W. Roberts, *J. Org. Chem.* **1961**, 26, 1152.
- 153 B. Kamber, *Helv. Chim. Acta* **1973**, 56, 1370.
- 154 H. Lamthanh, C. Roumestand, C. Deprun, A. Menez, *Int. J. Peptide Protein Res.* **1993**, 41, 85.
- 155 P. Sieber, *Helv. Chim. Acta* **1977**, 60, 2711.
- 156 H. Lamthanh, H. Virelizier, D. Frayssinhes, *Peptide Res.* **1995**, 8, 316.
- 157 M. Weinert, D. Brandenburg, H. Zahn, *Hoppe-Seylers Z. Physiol. Chem.* **1969**, 350, 1556.
- 158 H. Immer, I. Eberle, W. Fischer, E. Moser, in: *Peptides* 1988, G. Jung, E. Bayer (Eds.), de Gruyter, Berlin, **1989**, 94.
- 159 M. Bodanszky, J. Martinez, *Synthesis* **1981**, 333.
- 160 S. Sakakibara, T. Fujii, *Bull. Chem. Soc. Jpn.* **1969**, 42, 1466.
- 161 V. du Vigneaud, O.K. Behrens, *J. Biol. Chem.* **1937**, 117, 27.
- 162 S. Shaltiel, *Biochem. Biophys. Res. Commun.* **1967**, 29, 178.
- 163 J.R. Bell, J.H. Jones, *J. Chem. Soc., Perkin Trans. I* **1974**, 2336.
- 164 T. Brown, J.H. Jones, *J. Chem. Soc., Chem. Commun.* **1981**, 648.
- 165 T. Brown, J.H. Jones, J.D. Richards, *J. Chem. Soc., Perkin Trans. I* **1982**, 1553.
- 166 E. Schnabel, H. Herzog, P. Hoffmann, E. Klauke, I. Ugi, *Ann. Chem.* **1968**, 716, 175.

- 167 L. Moroder, A. Hallet, E. Wünsch, O. Keller, G. Wersin, *Hoppe-Seyler's Z. Physiol. Chem.* **1976**, 357, 1655.
- 168 K. Barlos, D. Papaioannou, D. Theodoropolous, *J. Org. Chem.* **1982**, 47, 1324.
- 169 P. Sieber, B. Riniker, *Tetrahedron Lett.* **1987**, 28, 6031.
- 170 G. Losse, U. Krychowski, *J. Prakt. Chem.* **1970**, 312, 1097.
- 171 E.V. Kudryavtseva, M.V. Sidorova, R.P. Evstigneeva, *Russ. Chem. Rev.* **1998**, 67, 545.
- 172 S. Sakakibara, T. Fujii, *Bull. Chem. Soc. Jpn.* **1970**, 43, 3954.
- 173 J.M. van der Eijk, R.J.M. Nolte, J.W. Zwikker, *J. Org. Chem.* **1980**, 45, 547.
- 174 K. Kitagawa, K. Kitade, Y. Kiso, T. Akita, S. Funakoshi, N. Fujii, H. Yajima, *J. Chem. Soc., Chem. Commun.* **1979**, 955.
- 175 R. Colombo, F. Colombo, J.H. Jones, *J. Chem. Soc., Chem. Commun.* **1984**, 292.
- 176 A.M. Kimbonguila, S. Boucida, F. Guib, A. Loffet, *Tetrahedron* **1997**, 53, 12525.
- 177 S.J. Harding, J.H. Jones, *J. Peptide Sci.* **1999**, 5, 399.
- 178 S.J. Harding, J.H. Jones, A.N. Sabirov, V.V. Samukov, *J. Peptide Sci.* **1999**, 5, 368.
- 179 P. Sieber, *Tetrahedron Lett.* **1987**, 28, 6147.
- 180 T. Brown, J.H. Jones, J.D. Richards, *J. Chem. Soc., Perkin Trans. I* **1982**, 1553.
- 181 R. Pipkorn, B. Ekberg, *Int. J. Peptide Protein Res.* **1986**, 27, 583.
- 182 K. Yoshizawa-Kumagaye, T. Ishizu, S. Isaka, M. Tamura, R. Okihara, Y. Nishiuchi, T. Kimura, *Protein Peptide Lett.* **2005**, 12, 579.
- 183 T. Mizoguchi, G. Levin, D.W. Wooley, J.M. Stewart, *J. Org. Chem.* **1968**, 33, 903.
- 184 E. Wünsch, G. Fries, A. Zwick, *Chem. Ber.* **1958**, 91, 542.
- 185 H. Sajiki, K. Hirota, *Tetrahedron* **1998**, 54, 13981.
- 186 E. Schröder, *Ann. Chem.* **1964**, 670, 127.
- 187 K. Okawa, A. Tani, *J. Chem. Soc. Jpn.* **1950**, 75, 1197.
- 188 B.W. Erickson, R.B. Merrifield, *J. Am. Chem. Soc.* **1973**, 95, 3757.
- 189 L. Lapatsanis, in: *Peptides 1978*, I.Z. Siemion, G. Kupryszewski (Eds.), University Press, Wroclaw, **1979**, 105.
- 190 Y. Nishiyama, S. Shikama, K.-I. Morita, K. Kurita, *J. Chem. Soc., Perkin Trans. I* **2000**, 1949.
- 191 D. Yamashiro, C.H. Li, *J. Org. Chem.* **1973**, 38, 591.
- 192 J.P. Tam, R.B. Merrifield, in: *The Peptides*, Volume 9, S. Udenfriend, J. Meienhofer (Eds.), Academic Press, New York, **1987**, 185.
- 193 F.M. Callahan, G.W. Anderson, R. Paul, J.E. Zimmermann, *J. Am. Chem. Soc.* **1963**, 85, 201.
- 194 F. Weygand, W. Steglich, I. Lengyel, F. Fraunberger, A. Maierhofer, W. Oettmeier, *Chem. Ber.* **1966**, 99, 1944.
- 195 P.M. Pojer, S.J. Angyal, *Tetrahedron Lett.* **1976**, 3067.
- 196 A.D. Morley, *Tetrahedron Lett.* **2000**, 41, 7401.
- 197 J. Bodi, Y. Nishiuchi, H. Nishio, T. Inui, T. Kimura, *Tetrahedron Lett.* **1998**, 39, 7117.
- 198 Y. Nishiuchi, H. Nishio, T. Inui, J. Bodi, T. Kimura, *J. Peptide Sci.* **2000**, 6, 84.
- 199 A. Di Fenza, M. Tancredi, C. Galoppini, P. Povero, *Tetrahedron Lett.* **1998**, 39, 8529.
- 200 H. Kunz, *Chem. Ber.* **1976**, 109, 2670, 3693.
- 201 H. Huang, D.L. Rabenstein, *J. Peptide Res.* **1999**, 53, 548.
- 202 M. Girard, F. Cavelier, J. Martinez, *J. Peptide Sci.* **1999**, 5, 457.
- 203 M. Ohno, S. Tsukamoto, S. Sato, N. Izumiya, *Bull. Chem. Soc. Jpn.* **1973**, 46, 3280.
- 204 D. Yamashiro, C.H. Li, *J. Org. Chem.* **1973**, 38, 2594.
- 205 P. White, in: *Peptides, Chemistry, Structure and Biology*, J.A. Smith, J.E. Riviere (Eds.), Escom, Leiden, **1992**, 537.
- 206 Y. Nishiuchi, H. Nishio, T. Inui, T. Kimura, S. Sakakibara, *Tetrahedron Lett.* **1996**, 37, 7529.
- 207 Y. Kiso, T. Kumura, M. Shimokura, T. Narukami, *J. Chem. Soc., Chem. Commun.* **1988**, 287.

- 208 H.M. Franzen, K. Nagren, L. Grehn, B. Langström, U. Ragnarsson, *J. Chem. Soc., Perkin Trans. I* **1988**, 497.
- 209 H.A. Guillaume, J.W. Perrich, J.D. Wade, G.W. Tregear, R.B. Johns, *J. Org. Chem.* **1989**, 54, 5731.
- 210 S. Capasso, P. Di Cerbo, *J. Peptide Res.* **2000**, 56, 382.
- 211 M. Xie, J. Aub, R.T. Borchardt, M. Morton, E.M. Topp, D. Vander Velde, R.L. Schowen, *J. Peptide Res.* **2000**, 56, 165.
- 212 R. Li, A.J. D'Souza, B.B. Laird, R.L. Schowen, R.T. Borchardt, E.M. Topp, *J. Peptide Res.* **2000**, 56, 326.
- 213 P. Sieber, B. Riniker, *Tetrahedron Lett.* **1991**, 32, 739.
- 214 B. Riniker, A. Flörsheimer, H. Fretz, P. Sieber, B. Kamber, *Tetrahedron* **1993**, 49, 9307.
- 215 Y. Shimonishi, S. Sakakibara, S. Akabori, *Bull. Chem. Soc. Jpn.* **1962**, 35, 1966.
- 216 W. König, R. Geiger, *Chem. Ber.* **1970**, 103, 2041.
- 217 R.W. Holley, *J. Am. Chem. Soc.* **1955**, 77, 2552.
- 218 J.-D. Glass in: *The Peptides: Analysis, Synthesis, Biology*, Volume 9, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1987**, 167.
- 219 P. Hermann, *Biomed. Biochim. Acta* **1991**, 50, 19.
- 220 T. Pathak, H. Waldmann, *Curr. Opin. Chem. Biol.* **1998**, 2, 112.
- 221 F. Eisele, D.J. Owen, H. Waldmann, *Bioorg. Med. Chem.* **1999**, 7, 193.
- 222 M. Schelhaas, H. Waldmann, *Angew. Chem. Int. Ed.* **1996**, 35, 2056.
- 223 H. Waldmann, D. Sebastian, *Chem. Rev.* **1994**, 94, 911.
- 224 T. Kappes, H. Waldmann, *Liebigs Ann./Recueil* **1997**, 803.
- 225 B. Sauerbrei, T. Kappes, H. Waldmann, *Top. Curr. Chem.* **1997**, 186, 65.
- 226 T. Kappes, H. Waldmann, *Carbohydr. Res.* **1998**, 305, 341.
- 227 A.G. Gum, T. Kappes-Roth, H. Waldmann, *Chem. Eur. J.* **2000**, 6, 3714.
- 228 H. Waldmann, E. Nägele, *Angew. Chem. Int. Ed.* **1995**, 34, 2259.
- 229 E. Nägele, M. Schelhaas, N. Kuder, H. Waldmann, *J. Am. Chem. Soc.* **1998**, 120, 6889.
- 230 T. Pohl, H. Waldmann, *Angew. Chem. Int. Ed.* **1996**, 35, 1720.
- 231 P. Braun, H. Waldmann, H. Kunz, *Bioorg. Med. Chem.* **1993**, 197.
- 232 P. Braun, H. Waldmann, H. Kunz, *Synlett* **1992**, 39.
- 233 H. Kunz, D. Kowalczyk, P. Braun, G. Braun, *Angew. Chem. Int. Ed.* **1994**, 33, 336.
- 234 J. Eberling, P. Braun, D. Kowalczyk, M. Schultz, H. Kunz, *J. Org. Chem.* **1996**, 61, 2638.
- 235 F. Eisele, J. Kuhlmann, H. Waldmann, *Angew. Chem. Int. Ed.* **2001**, 40, 369.
- 236 J. Sander, H. Waldmann, *Chem. Eur. J.*, **2000**, 6, 1564.
- 237 M. Schelhaas, S. Glomsda, M. Hänsler, H.-D. Jakubke, H. Waldmann, *Angew. Chem. Int. Ed.* **1996**, 35, 106.
- 238 J. Sander, H. Waldmann, *Angew. Chem. Int. Ed.* **1999**, 38, 1250.
- 239 M. Schelhaas, E. Nägele, N. Kuder, B. Bader, J. Kuhlmann, A. Wittinghofer, H. Waldmann, *Chem. Eur. J.* **1999**, 5, 1239.
- 240 A. Cotte, B. Bader, J. Kuhlmann, H. Waldmann, *Chem. Eur. J.* **1999**, 5, 922.
- 241 S. Flohr, V. Jungmann, H. Waldmann, *Chem. Eur. J.* **1999**, 5, 669.
- 242 M.W. Pennington, B.M. Dunn, *Peptide Synthesis Protocols*, Humana Press, Totowa, NJ, **1995**.
- 243 P. Lloyd-Williams, F. Albericio, E. Giralt, *Chemical Approaches to the Synthesis of Peptides and Proteins*, CRC Press, Boca Raton, **1997**.
- 244 S.-Y. Han, Y.-A. Kim, *Tetrahedron* **2004**, 60, 2447.
- 245 C.A.G.N. Montalbetti, V. Falque, *Tetrahedron* **2005**, 61, 10827.
- 246 J. Howl (Ed.), *Peptide Synthesis and Applications*, Humana Press, Totowa, NJ, **2005**.
- 247 N.L. Benoiton, *Chemistry of Peptide Synthesis*, CRC, Boca Raton, **2006**.

- 248 H.-D. Jakubke, N. Sewald, *Peptides from A-Z, A Concise Encyclopedia*, Wiley-VCH, Weinheim, **2008**.
- 249 F. Albericio, R. Chinchilla, D. Dodsworth, C. Najera, *Org. Rep. Proceed. Int.* **2001**, 33, 203.
- 250 J. Meienhofer, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 1, E. Gross, J. Meienhofer Academic Press, New York, **1979**, 197.
- 251 T. Curtius, *Ber. Dtsch. Chem. Ges.* **1902**, 35, 3226.
- 252 T. Shioiri, K. Ninomiya, S. Yamada, *J. Am. Chem. Soc.* **1972**, 94, 6203.
- 253 J. Honzl, J. Rudinger, *Coll. Czech. Chem. Commun.* **1961**, 26, 2333.
- 254 J. Meienhofer in: *The Peptides: Analysis, Synthesis, Biology*, Volume 1, E. Gross, J. Meienhofer (Eds.) Academic Press, New York, **1979**, 263.
- 255 F.M.F. Chen, R. Steinauer, N.L. Benoiton, in: *Peptides 1984*, U. Ragnarsson (Ed.), Almqvist and Wiksell, Stockholm, **1984**.
- 256 F.M.F. Chen, Y. Lee, R. Steinauer, N.L. Benoiton, *Can. J. Chem.* **1987**, 65, 613.
- 257 F.M.F. Chen, N.L. Benoiton, *Can. J. Chem.* **1987**, 65, 619.
- 258 N.L. Benoiton, Y. Lee, F.M.F. Chen, *Int. J. Peptide Protein Res.* **1988**, 31, 443.
- 259 N.L. Benoiton, Y.C. Lee, F.M.F. Chen, *Int. J. Peptide Protein Res.* **1993**, 42, 278.
- 260 J.H. Jones, M.J. Witty, *J. Chem. Soc., Chem. Commun.* **1977**, 281.
- 261 J.H. Jones, M.J. Witty, *J. Chem. Soc., Perkin Trans. I* **1979**, 3203.
- 262 R. Ramage, D. Hopton, M.J. Parrot, R.S. Richardson, G.W. Kenner, G.A. Moore, *J. Chem. Soc., Perkin Trans. I* **1985**, 461.
- 263 S.J. Wittenberger, M.A. McLaughlin, *Tetrahedron Lett.* **1999**, 40, 7175.
- 264 P. Koch, C. Vedder, T. Schaffer, *Chim. Oggi* **2008**, 26, 6.
- 265 A. Pelter, T.E. Levitt, P. Nelsoni, *Tetrahedron* **1970**, 26, 1539.
- 266 J.C. Sheehan, J.J. Hlavka, *J. Org. Chem.* **1958**, 23, 635.
- 267 R. Hirschmann, R.G. Denkwalter, H. Schwam, R.G. Strachan, T.E. Beesley, D.F. Veber, E.F. Schoenewaldt, H. Barkemeyer, W.J. Paleveda, J.T.A. Jacob, *J. Am. Chem. Soc.* **1966**, 88, 3163.
- 268 R. Hirschmann, R.G. Denkwalter, *Naturwissenschaften* **1970**, 57, 145.
- 269 W.D. Fuller, M.P. Cohen, M. Shabankareh, R. Blair, M. Goodman, F.R. Naider, *J. Am. Chem. Soc.* **1990**, 112, 7414.
- 270 T.B. Sim, H. Rapoport, *J. Org. Chem.* **1999**, 64, 2532.
- 271 D.H. Rich, J. Singh, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 1, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1979**, 241.
- 272 J. Izdebski, D. Kuncze, *J. Peptide Sci.* **1997**, 3, 141.
- 273 S. Nozaki, *J. Peptide Res.* **1999**, 54, 162.
- 274 H. Ito, N. Takamatsu, I. Ichikizaki, *Chem. Lett.* **1975**, 577.
- 275 E. Wünsch, F. Dress, *Chem. Ber.* **1966**, 99, 110.
- 276 F. Weygand, D. Hoffmann, E. Wünsch, *Z. Naturforsch. B* **1966**, 21, 426.
- 277 W. König, R. Geiger, *Chem. Ber.* **1970**, 103, 788.
- 278 K. Barlos, D. Papaioannou, S. Voliotis, R. Prewo, J.H. Bieri, *J. Org. Chem.* **1985**, 50, 696.
- 279 M. Ueki, T. Yanagihara, in: *Peptides 1998*, S. Bajusz, F. Hudecz (Eds.), Akademiai Kiado, Budapest, **1998**, 252.
- 280 L.A. Carpino, *J. Am. Chem. Soc.* **1993**, 115, 4397.
- 281 L.A. Carpino, A. El-Faham, C.A. Minor, F. Albericio, *J. Chem. Soc., Chem. Commun.* **1994**, 201.
- 282 Y.M. Angell, C. Garcia-Echeverria, D.H. Rich, *Tetrahedron Lett.* **1994**, 35, 5981.
- 283 L.A. Carpino, A. El-Faham, *Tetrahedron* **1999**, 55, 6813.
- 284 N. Robertson, L. Jiang, R. Ramage, *Tetrahedron* **1999**, 55, 2713.
- 285 J.C. Spetzler, M. Meldal, J. Felding, P. Veds, M. Begtrup, *J. Chem. Soc., Perkin Trans. I* **1998**, 1727.
- 286 M. Bodanszky, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 1, E. Gross,

- J. Meienhofer Academic Press, New York, 1979, 105.
- 287 R.B. Woodward, R.A. Olofson, *J. Am. Chem. Soc.* **1961**, 83, 1010.
 - 288 S.M. Beaumont, B.O. Handford, G.T. Young, *Acta Chim. Sci. Hung.* **1965**, 44, 37.
 - 289 H.-D. Jakubke, A. Voigt, *Chem. Ber.* **1966**, 99, 2944.
 - 290 L.A. Carpino, H. Imazumi, B.M. Foxman, M.J. Vela, P. Henklein, A. El-Faham, J. Klose, M. Bienert, *Organic Lett.* **2000**, 2, 2253.
 - 291 C. Kitada, M. Fujino, *Chem. Pharm. Bull.* **1978**, 26, 585.
 - 292 L.A. Carpino, M. Beyermann, H. Wenschuh, M. Bienert, *Acc. Chem. Res.* **1996**, 29, 268.
 - 293 E. Falb, T. Yechezkel, Y. Salitra, C. Gilon, *J. Peptide Res.* **1999**, 53, 507.
 - 294 L.A. Carpino, B. Cohen, K.E. Stephens, D. Sadat-Aalae, J.H. Tien, D.C. Langridge, *J. Org. Chem.* **1986**, 51, 3732.
 - 295 L.A. Carpino, D. Sadat-Aalae, H.G. Chao, R.H. De Selms, *J. Am. Chem. Soc.* **1990**, 112, 9651.
 - 296 H. Wenschuh, M. Beyermann, S. Rothmund, L.A. Carpino, M. Bienert, *Tetrahedron Lett.* **1995**, 36, 1247.
 - 297 H. Wenschuh, M. Beyermann, R. Winter, M. Bienert, D. Ionescu, L.A. Carpino, *Tetrahedron Res.* **1996**, 37, 5438.
 - 298 H. Wenschuh, M. Beyermann, A. El-Faham, S. Ghassemi, L.A. Carpino, M. Bienert, *J. Chem. Soc. Chem., Commun.* **1995**, 669.
 - 299 L.A. Carpino, A. El-Faham, *J. Am. Chem. Soc.* **1995**, 117, 5401.
 - 300 B. Castro, J.R. Dormoy, G. Evin, C. Selve, **1975**, *Tetrahedron Lett.* 1219.
 - 301 R. Steinauer, F.M.F. Chen, N.L. Benoiton, *Int. J. Peptide Protein Res.* **1989**, 34, 295.
 - 302 J. Coste, D. Le-Nguyen, B. Castro, *Tetrahedron Lett.* **1990**, 31, 205.
 - 303 F. Albericio, M. Cases, J. Alsina, S.A. Triolo, L.A. Carpino, S.A. Kates, *Tetrahedron Lett.* **1997**, 38, 4853.
 - 304 C. Auvin-Guette, E. Frerot, J. Coste, S. Rebuffat, P. Jouin, B. Bodo, *Tetrahedron Lett.* **1993**, 34, 2481.
 - 305 H. Li, X. Jiang, Y.-H. Ye, C. Fan, T. Romoff, M. Goodman, *Organic Lett.* **1999**, 1, 91.
 - 306 Y. Ye, H. Li, X. Jiang, *Biopolymers (Peptide Sci.)* **2005**, 80, 172.
 - 307 V. Dourtoglou, J.-C. Ziegler, B. Gross, *Tetrahedron Lett.* **1978**, 15, 1269.
 - 308 V. Dourtoglou, B. Gross, V. Lambropoulou, C. Ziodrou, *Synthesis* **1984**, 572.
 - 309 I. Abdelmoty, F. Albericio, L.A. Carpino, B.M. Foxman, S.A. Kates, *Lett. Peptide Sci.* **1994**, 1, 57.
 - 310 L.A. Carpino, H. Imazumi, A. El-Faham, F.J. Ferrer, C. Zhang, Y. Lee, B.M. Foxman, P. Henklein, C. Hanay, C. Mügge, H. Wenschuh, J. Klose, M. Beyermann, M. Bienert, *Angew. Chem. Int. Ed.* **2002**, 41, 441.
 - 311 R. Knorr, A. Trzeciak, W. Bannwarth, D. Gillissen, *Tetrahedron Lett.* **1989**, 30, 1927.
 - 312 L.A. Carpino, A. El-Fahan, C.A. Minor, F. Albericio, *J. Chem. Soc., Chem. Commun.* **1994**, 201.
 - 313 L.A. Carpino, A. El-Fahan, F. Albericio, *Tetrahedron Lett.* **1994**, 35, 2279.
 - 314 C.A. Hood, G. Fuentes, H. Patel, K. Page, M. Menakuru, J.H. Park, *J. Peptide. Sci.* **2008**, 14, 97.
 - 315 H. Gausepohl, M. Kraft, R.W. Frank, *Int. J. Peptide Protein Res.* **1989**, 34, 287.
 - 316 F. Albericio, J.M. Bofill, A. El-Faham, S.A. Kates, *J. Org. Chem.* **1998**, 63, 9678.
 - 317 S. Chen, J. Xu, *Tetrahedron Lett.* **1992**, 33, 647.
 - 318 J. Habermann, H. Kunz, *Tetrahedron Lett.* **1998**, 39, 265.
 - 319 J.M. Bofill, F. Albericio, *Tetrahedron Lett.* **1999**, 40, 2641.
 - 320 A. Ehrlich, S. Rothmund, M. Brudel, M. Beyermann, L.A. Carpino, M. Bienert, *Tetrahedron Lett.* **1993**, 34, 4781.
 - 321 U. Schmidt, H. Grieser, *J. Chem. Soc., Chem. Commun.* **1993**, 1461.
 - 322 P. Li, J.C. Xu, *Tetrahedron.* **2000**, 56, 4437.
 - 323 Z.J. Kamiński, P. Paneth, J. Rudziński, *J. Org. Chem.* **1998**, 63, 4248.
 - 324 Z.J. Kamiński, B. Kolesińska, J. Kolesińska, G. Sabatino, M. Chelli,

- P. Rovero, M. Błaszczuk, M.L. Głowska, A.M. Papini, *J. Am. Chem. Soc.* **2005**, *127*, 16912.
- 325** A. Falchi, G. Giacomelli, A. Porcheddu, M. Taddei, *Synlett* **2000**, 275.
- 326** T. Mukaiyama, *Angew. Chem. Int. Ed.* **1979**, *18*, 707.
- 327** P. Li, J.-C. Xu, *Tetrahedron* **2000**, *56*, 8119.
- 328** G.W. Anderson, R. Paul, *J. Am. Chem. Soc.* **1958**, *80*, 4423.
- 329** S. Kato, T. Morie, N. Yoshida, *Chem. Pharm. Bull.* **1995**, *43*, 699.
- 330** A.K. Saha, P. Schultz, H. Rapoport, *J. Am. Chem. Soc.* **1989**, *111*, 4856.
- 331** K. Akaji, N. Kuriyama, Y. Kiso, *J. Org. Chem.* **1996**, *61*, 3350.
- 332** P. Li, J.C. Xu, *Tetrahedron* **2000**, *56*, 9949.
- 333** P. Li, J.C. Xu, *Tetrahedron Lett.* **1999**, *40*, 8301.
- 334** R. Wischnat, J. Rudolph, R. Hanke, R. Kaese, A. May, H. Theis, U. Zuther, *Tetrahedron Lett.* **2003**, *44*, 4393.
- 335** T. Shioiri, K. Ninomiya, S. Yamada, *J. Am. Chem. Soc.* **1972**, *94*, 6203.
- 336** D.L. Boger, J.-H. Chen, K.W. Saionz, *J. Am. Chem. Soc.* **1996**, *118*, 1629.
- 337** J. Diago-Meseguer, A.L. Palomo-Coll, *Synthesis* **1980**, 547.
- 338** S. Chen, J. Xu, *Tetrahedron Lett.* **1991**, *32*, 6711.
- 339** T. Katoh, M. Ueki, *Int. J. Peptide Protein Res.* **1993**, *42*, 264.
- 340** P. Welton, *Ionic Liquids in Synthesis* Wiley-VCH, Weinheim, Germany, **2003**.
- 341** M. Erbdinger, A.J. Mesiano, A.J. Russel, *Biotechnol. Progr.* **2000**, *16*, 1129.
- 342** H. Vallete, L. Ferron, G. Coquerel, A.-C. Gaumont, J.-C. Plaquevent, *Tetrahedron Lett.* **2004**, *45*, 1617.
- 343** W. Miao, T.-H. Chan, *J. Org. Chem.* **2005**, *70*, 3251.
- 344** J. Spengler, C. Böttcher, F. Albericio, K. Burger, *Chem. Rev.* **2006**, *106*, 4728.
- 345** R. Ramesh, S. Rajasekaran, R. Gupta, S. Chandrasekaran, *Org. Lett.* **2006**, *8*, 1933.
- 346** T.P. Sullivan, M.L. van Poll, P.Y.W. Dankers, W.T.S. Huck, *Angew. Chem. Int. Ed.* **2004**, *43*, 4190.
- 347** D.S. Kemp, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 1, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1979**, 315.
- 348** N.L. Benoiton, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 5, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1983**, 217.
- 349** N.L. Benoiton, *Biopolymers* **1996**, *40*, 245.
- 350** N.L. Benoiton, *Int. J. Peptide Protein Res.* **1994**, *44*, 399.
- 351** F. Weygand, W. Steglich, X. Barocio de la Lama, *Tetrahedron Suppl.* **1966**, *8*, 9.
- 352** C. Czerwenka, W. Lindner, *Anal. Bioanal. Chem.* **2005**, *382*, 599.
- 353** F.M. Finn, K. Hofmann, in: *The Proteins*, Volume 2, H. Neurath, R.L. Hill (Eds.), Academic Press, New York, **1976**, 183.
- 354** M. Manning, E. Coy, W.H. Sawyer, *Biochemistry* **1970**, *9*, 3925.
- 355** M. Goodman, P. Keogh, H. Anderson, *Bioorg. Chem.* **1977**, *6*, 239.
- 356** C. Griehl, J. Weight, H. Jeschkeit, *J. High. Res. Chromatogr.* **1994**, *17*, 700.
- 357** C. Griehl, J. Weight, H. Jeschkeit, Z. Palacz, in: *Peptides 1992*, C.H. Schneider, A.N. Eberle (Eds.), Escom, Leiden, **1993**, 459.
- 358** H. Yajima, N. Fujii, *J. Chem. Soc., Chem. Commun.* **1980**, 115.
- 359** R.B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149.
- 360** P.H. Seeberger, *Chem. Soc. Rev.* **2008**, *37*, 19.
- 361** G. Barany, R.B. Merrifield, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 2, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1980**, 1.
- 362** R.B. Merrifield, *Angew. Chem. Int. Ed.* **1985**, *24*, 799.
- 363** M.W. Pennington, B.M. Dunn, *Peptide Synthesis Protocols* Humana Press, Totowa, **1994**.
- 364** G.B. Fields, *Solid-Phase Peptide Synthesis* Academic Press, New York, **1997**.
- 365** W.C. Chan, P.D. White, *Fmoc Solid Phase Peptide Synthesis: A Practical*

- Approach*, Oxford University Press, Oxford, 2000.
- 366 A. Kates, F. Albericio, *Solid-Phase Synthesis: A Practical Guide*, CRC Press, 2000.
 - 367 J. Alsina, F. Albericio, *Biopolymers* 2003, 71, 454.
 - 368 K. Hojo, M. Maeda, N. Tanakamaru, K. Mochida, K. Kawasaki, *Protein Peptide Lett.* 2006, 13, 189.
 - 369 E. Atherton, R.C. Sheppard, *Solid Phase Peptide Synthesis: A Practical Approach* IRL Press, Oxford, 1989.
 - 370 E. Bayer, *Angew. Chem. Int. Ed.* 1991, 30, 113.
 - 371 B.G. de la Torre, A. Jakab, D. Andreu, *Int. J. Peptide Res. Therap.* 2007, 13, 265.
 - 372 P.G. Sasikumar, M. Kempe, *Int. J. Peptide Res. Therap.* 2007, 13, 129.
 - 373 M. Meldal, I. Svendsen, K. Breddam, F. Auzanneau, *Proc. Natl. Acad. Sci. USA* 1994, 91, 3314.
 - 374 I.W. James, *Tetrahedron* 1999, 55, 4855.
 - 375 D. Swinnen, D. Hilvert, *Org. Lett.* 2000, 2, 2439.
 - 376 J.H. Jones, W.I. Ramage, *J. Chem. Soc., Chem. Commun.* 1978, 472.
 - 377 J.H. Jones, A.V. Stachulsky, *J. Chem. Soc., Perkin Trans. I* 1979, 2261.
 - 378 S.S. Wang, *J. Am. Chem. Soc.* 1973, 95, 1328.
 - 379 K. Sandhya, B. Ravindranath, *Tetrahedron Lett.* 2008, 49, 2435.
 - 380 M. Mergler, R. Tanner, J. Gosteli, P. Grogg, *Tetrahedron Lett.* 1988, 29, 4005.
 - 381 B. Riniker, A. Flörsheimer, H. Fretz, P. Sieber, B. Kamber, *Tetrahedron* 1993, 49, 9307.
 - 382 A.R. Mitchell, S.B.H. Kent, M. Engelhardt, R.B. Merrifield, *J. Org. Chem.* 1978, 43, 2845.
 - 383 R.C. Sheppard, B.J. Williams, *Int. J. Peptide Protein Res.* 1982, 20, 451.
 - 384 P.G. Pietta, P.F. Cavallo, K. Takahashi, G.R. Marshall, *J. Org. Chem.* 1974, 39, 44.
 - 385 G.R. Matsueda, J.M. Stewart, *Peptides* 1981, 2, 45.
 - 386 H. Rink, *Tetrahedron Lett.* 1987, 28, 3787.
 - 387 K. Barlos, O. Chatzi, D. Gatos, G. Stavropoulos, *Int. J. Peptide Protein Res.* 1991, 37, 513.
 - 388 M. Gairi, P. Lloyd-Williams, F. Albericio, E. Giralt, *Int. J. Peptide Protein Res.* 1995, 46, 119.
 - 389 H. Kunz, B. Dombo, *Angew. Chem. Int. Ed.* 1988, 27, 711.
 - 390 C.P. Holmes, D.G. Jones, *J. Org. Chem.* 1995, 60, 2318.
 - 391 M. Patek, M. Lebl, *Biopolymers* 1998, 47, 353.
 - 392 P. Heidler, A. Link, *Bioorg. Med. Chem.* 2005, 13, 585.
 - 393 G.W. Kenner, J.R. McDermott, R.C. Sheppard, *J. Chem. Soc., Chem. Commun.* 1971, 636.
 - 394 B.J. Backes, J.A. Ellman, *J. Org. Chem.* 1999, 64, 2322.
 - 395 R. Ingenito, E. Bianchi, D. Fattori, A. Pessi, *J. Am. Chem. Soc.* 1999, 121, 11369.
 - 396 S. Mezzato, M. Schaffrath, C. Unverzagt, *Angew. Chem. Int. Ed.* 2005, 44, 1650.
 - 397 D.L. Marshall, I.E. Liener, *J. Org. Chem.* 1970, 35, 867.
 - 398 T. Kimura, T. Fukui, S. Tanaka, K. Akaji, Y. Kiso, *Chem. Pharm. Bull.* 1997, 45, 18.
 - 399 M. Patek, M. Lebl, *Tetrahedron Lett.* 1991, 32, 3891.
 - 400 A. Brik, E. Keinan, P.E. Dawson, *J. Org. Chem.* 2000, 65, 3829.
 - 401 A.M. Bray, N.J. Maeji, R.M. Valerio, R.A. Campbell, H.M. Geysen, *J. Org. Chem.* 1991, 56, 6659.
 - 402 B. Atrash, M. Bradley, *J. Chem. Soc., Chem. Commun.* 1997, 1397.
 - 403 G. Panke, R. Frank, *Tetrahedron Lett.* 1998, 39, 17.
 - 404 J.P. Tam, W.F. Health, R.B. Merrifield, *J. Am. Chem. Soc.* 1983, 105, 6442.
 - 405 W.F. Health, J.P. Tam, R.B. Merrifield, *Int. J. Peptide Protein Res.* 1986, 28, 498.
 - 406 G.W. Tregear, J. van Rietschoten, R. Sauer, H.D. Niall, H.T. Keutmann, *Biochemistry* 1977, 16, 2817.
 - 407 S.B.H. Kent, *Annu. Rev. Biochem.* 1988, 57, 959.

- 408 J. Bedford, C. Hyde, T. Johnson, W. Jun, D. Owen, M. Quibell, R.C. Sheppard, *Int. J. Peptide Protein Res.* **1992**, 40, 300.
- 409 R.B. Merrifield, in: *Recent Progress in Hormone Research*, Volume 23, G. Pincus, (Ed.), Academic Press, New York, **1967**, 451.
- 410 M. Beyermann, M. Bienert, *Tetrahedron Lett.* **1992**, 33, 3745.
- 411 V. Krchnak, Z. Flegelova, J. Vagner, *Int. J. Peptide Protein Res.* **1993**, 42, 450.
- 412 M. Narita, S. Ouchi, *J. Org. Chem.* **1994**, 52, 686.
- 413 R.C. de, L. Milton, S.C.F. Milton, P.A. Adams, *J. Am. Chem. Soc.* **1990**, 122, 6039.
- 414 B. Henkel, E. Bayer, *J. Peptide Sci.* **1998**, 4, 461.
- 415 R. Nagaraj, P. Balaram, *Tetrahedron* **1981**, 37, 2001.
- 416 M. Bodanszky, S.S. Deshmane, J. Martinez, *J. Org. Chem.* **1979**, 44, 1622.
- 417 H. Wenschuh, M. Beyermann, E. Krause, M. Brudel, R. Winter, M. Schumann, L.A. Carpino, M. Bienert, *J. Org. Chem.* **1994**, 59, 3275.
- 418 M. Narita, S. Honda, H. Umeyama, S. Obana, *Bull. Chem. Soc. Jpn.* **1988**, 61, 281.
- 419 H. Kuroda, Y.-H. Chen, T. Kumura, S. Sakakibara, *Int. J. Peptide Protein Res.* **1992**, 40, 294.
- 420 E. Oliveira, L. Miranda, F. Albericio, D. Andreu, A.C.M. Paiva, C.R. Nakaie, M. Tominaga, *J. Peptide Res.* **1997**, 49, 300.
- 421 L. Zhang, C. Goldammer, B. Henkel, G. Panhaus, F. Zühl, G. Jung, E. Bayer, in: *Innovation and Perspectives in Solid Phase Synthesis*, R. Epton (Ed.), Intercept, Andover, **1994**, 711.
- 422 D. Seebach, A. Thale, A.K. Beck, *Helv. Chim. Acta* **1989**, 72, 857.
- 423 M.W. Pennington, I. Zaydenberg, M.E. Byrnes, R.S. Norton, W.R. Kem, *Int. J. Peptide Protein Res.* **1994**, 43, 463.
- 424 R. Bartl, K.-D. Klöppel, R. Frank, in: *Peptides 1992*, C.H. Schneider, A.N. Eberle (Eds.), Escom, Leiden, **1993**, 277.
- 425 T. Johnson, M. Quibell, D. Owen, R.C. Sheppard, *J. Chem. Soc., Chem. Commun.* **1993**, 369.
- 426 W. Rapp, L. Zhang, R. Häbich, F. Bayer, in: *Peptides 1988*, G. Jung, E. Bayer (Eds.), de Gruyter, Berlin, **1989**.
- 427 M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S.B.H. Kent, *Int. J. Peptide Protein Res.* **1992**, 40, 180.
- 428 M. Schnölzer, S.B.H. Kent, *Science* **1992**, 256, 221.
- 429 B.D. Larsen, A. Holm, *J. Peptide Res.* **1998**, 52, 470.
- 430 H.G. Chao, M.S. Bernatowicz, G.R. Matsueda, *J. Org. Chem.* **1993**, 58, 2640.
- 431 L.M. Varanda, M.T. Miranda, *J. Peptide Res.* **1997**, 50, 102.
- 432 Y. Sohma, T. Yoshiya, A. Taniguchi, T. Kimura, Y. Hayashi, Y. Kiso, *Biopolymers* **2007**, 88, 253.
- 433 R. Mimna, M.S. Camus, A. Schmid, G. Tuchscherer, H.A. Lashuel, M. Mutter, *Angew. Chem. Int. Ed.* **2007**, 46, 2681.
- 434 M. Mutter, A. Chandravarkar, C. Boyat, J. Lopez, S. Dos Santos, B. Mandal, R. Mimna, K. Murat, L. Patiny, L. Saucède, G. Tuchscherer, *Angew. Chem. Int. Ed.* **2004**, 43, 4172.
- 435 I. Coin, R. Dölling, E. Krause, M. Bienert, M. Beyermann, C.D. Sferdean, L.A. Carpino, *J. Org. Chem.* **2006**, 71, 6171.
- 436 I. Coin, M. Beyermann, M. Bienert, *Nat. Protoc.* **2007**, 2, 3247.
- 437 E. Kaiser, R.L. Colescott, C.D. Bossinger, P.I. Cook, *Anal. Biochem.* **1970**, 34, 595.
- 438 W.S. Hancock, J.E. Battersby, *Anal. Biochem.* **1976**, 71, 260.
- 439 T. Voikovskiy, *Peptide Res.* **1995**, 8, 236.
- 440 V. Krchnák, J. Vágnér, M. Lebl, *Int. J. Peptide Protein Res.* **1988**, 32, 415.
- 441 L.C. Dorman, *Tetrahedron Lett.* **1969**, 2319.
- 442 G. Talbo, J.D. Wade, N. Dawson, M. Manoussios, G.W. Tregear, *Lett. Peptide Sci.* **1999**, 4, 121.
- 443 E. Atherton, D.L.J. Clive, R.C. Sheppard, *J. Am. Chem. Soc.* **1975**, 97, 6584.

- 444 F. Albericio, G. Barany, *Int. J. Peptide Protein Res.* **1993**, 41, 307.
- 445 G.M. Williams, R.A.E. Carr, M.S. Congreve, C. Kay, S.C. McKeown, P.J. Murray, J.J. Scicinski, S.P. Watson, *Angew. Chem. Int. Ed.* **2000**, 39, 3293.
- 446 S.A. Salisbury, E.J. Treemer, J.W. Davies, D.E.I.A. Owen, *J. Chem. Soc., Chem. Commun.* **1990**, 538.
- 447 E. Bayer, G. Jung, I. Halasz, I. Sebastian, *Tetrahedron Lett.* **1979**, 4503.
- 448 R. Frank, H. Gausepohl, in: *Modern Methods in Protein Chemistry*, Volume 3, H. Tschesche (Ed.), de Gruyter, Berlin, **1988**, 41.
- 449 A. Dryland, R.C. Sheppard, *J. Chem. Soc., Chem. Commun.* **1986**, 125.
- 450 N. Kneib-Gordonier, F. Albericio, G. Barany, in: *Peptides: Chemistry, Structure and Biology* J.E. Rivier, G.R. Marshall (Eds.), Escom, Leiden, **1990**, 895.
- 451 D.L. Whitehouse, S.N. Savinov, D.J. Austin, *Tetrahedron Lett.* **1997**, 38, 7851.
- 452 H. Irie, N. Fujii, H. Ogawa, H. Yajima, M. Fujino, S. Shinagawa, *J. Chem. Soc., Chem. Commun.* **1976**, 922.
- 453 Y. Kiso, S. Nakamura, K. Ito, K. Kawa, K. Kitagawa, T. Akita, H. Moritoki, *J. Chem. Soc., Chem. Commun.* **1979**, 971.
- 454 N. Fujii, A. Otaka, O. Ikemura, K. Akaji, S. Funakoshi, Y. Hayashi, Y. Kuroda, H. Yajima, *J. Chem. Soc., Chem. Commun.* **1987**, 274.
- 455 J.P. Tam, in: *Peptide Chemistry 1987*, T. Shiba, S. Sakakibara (Eds.), Protein Research Foundation, Osaka, **1988**, 199.
- 456 L.A. Carpino, H. Shroff, S.A. Triolo, E.M.E. Mansour, H. Wenschuh, F. Albericio, *Tetrahedron Lett.* **1993**, 34, 7829.
- 457 R. Ramage, J. Green, *Tetrahedron Lett.* **1987**, 28, 2287.
- 458 P. White, in: *Peptides: Chemistry and Biology*, J.A. Smith, J.E. Rivier (Eds.), Escom, Leiden, **1992**, 537.
- 459 P. Sieber, B. Riniker, *Tetrahedron Lett.* **1991**, 32, 739.
- 460 A.J. Smith, J.D. Young, S.A. Carr, D.R. Marshak, L.C. Williams, K.R. Williams, in: *Techniques in Protein Chemistry III*, R.H. Angeletti (Ed.), Academic Press, Orlando, **1992**, 219.
- 461 W. Rapp, in: *Peptides 1992*, C.H. Schneider, A.N. Eberle (Eds.), Escom, Leiden, **1993**, 243.
- 462 B. Gutte, R.B. Merrifield, *J. Am. Chem. Soc.* **1969**, 91, 501.
- 463 G. Barany, N. Kneib-Cordonier, D.G. Mullen, *Int. J. Peptide Protein Res.* **1987**, 30, 705.
- 464 M.N. Shemyakin, Y.A. Ovchinnikov, A.A. Kiryushkin, I.V. Kozhevnikova, *Tetrahedron Lett.* **1965**, 2323.
- 465 E. Bayer, M. Mutter, *Nature* **1972**, 237, 512.
- 466 M. Mutter, E. Bayer, *Angew. Chem. Int. Ed.* **1974**, 13, 88.
- 467 M. Fischer, D.I. Zheleva, *J. Peptide Sci.* **2002**, 8, 529.
- 468 W. Miao, T.-H. Chan, *J. Org. Chem.* **2005**, 70, 3251.
- 469 H. Frank, H. Hagenmair, *Experientia* **1975**, 31, 131.
- 470 T. Wieland, C. Birr, *Angew. Chem. Int. Ed.* **1966**, 5, 310.
- 471 J.M. Salvino, N.V. Kumar, E. Orton, J. Airey, T. Kiesov, K. Crawford, *et al.*, *J. Comb. Chem.* **2000**, 2, 691.
- 472 I.R. Baxendale, S.V. Ley, *Bioorg. Med. Chem. Lett.* **2000**, 10, 1983.
- 473 M.A. Scialdone, *Tetrahedron Lett.* **1996**, 37, 8141.
- 474 D.B. Head, J.Z. Dong, J.A. Burton, *J. Peptide Res.* **2005**, 65, 384.
- 475 B.A. Bunin, *The Combinatorial Index* Academic Press, London, **1998**.
- 476 H.-M. Yu, S.-T. Chen, K.-T. Wang, *J. Org. Chem.* **1992**, 57, 4781.
- 477 V. Santagada, F. Fiorino, E. Perissutti, B. Severino, V. De Filippis, B. Vivencio, G. Caliendo, *Tetrahedron Lett.* **2001**, 42, 5171.
- 478 J.M. Collins, N.E. Leadbeater, *Org. Biomol. Chem.*, **2007**, 5, 1141.
- 479 T. Matsushita, H. Hinou, M. Fumoto, M. Kuroguchi, N. Fujitani, H. Shimizu,

- S.-I. Nishimura, *J. Org. Chem.* **2006**, 71, 3051.
- 480 J.K. Murray, S.H. Gellman, *Nature Protocols* **2007**, 2, 624.
- 481 D. Blohm, C. Bollschweiler, H. Hillen, *Angew. Chem. Int. Ed.* **1988**, 27, 207.
- 482 G. Gellissen (Ed.), *Production of Recombinant Proteins*, Wiley-VCH, Weinheim, **2004**.
- 483 F.M. Wurm, *Nature Biotechnol.* **2004**, 22, 1393.
- 484 J.W. Engels, E. Ullmann, *Angew. Chem. Int. Ed.* **1989**, 28, 716.
- 485 K. Itakura, T. Hirose, R. Crea, A.D. Riggs, H.L. Heyneker, F. Bolivar, H.W. Boyer, *Science* **1977**, 198, 1056.
- 486 D.V. Goeddel, H.L. Heyneker, T. Hozumi, R. Arentzen, K. Itakura, D.G. Yansura, M.J. Ross, G. Miozzari, R. Crea, P.H. Seeburg, *Nature* **1979**, 281, 544.
- 487 H.M. Hsiung, N.G. Mayne, G.W. Becker, *Bio/Technology* **1986**, 4, 991.
- 488 D.V. Goeddel, D.G. Kleid, F. Bolivar, H.L. Heyneker, D.G. Ansura, R. Crea, T. Horose, A. Kraszewski, K. Itakura, A.D. Riggs, *Proc. Natl. Acad. Sci. USA* **1979**, 76, 106.
- 489 L. Villa-Komaroff, A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S.P. Nabes, W.L. Chick, W. Gilbert, *Proc. Natl. Acad. Sci. USA* **1978**, 75, 3727.
- 490 A.H. Barnett, D.R. Owens, *Lancet* **1997**, 349, 47.
- 491 B. Zinman, H. Tildesley, J. Chiasson, E. Tsui, F. Strack, *Diabetes* **1997**, 46, 440.
- 492 M.B. Hoagland, M.L. Stephenson, J.F. Scott, L.I. Hecht, P.C. Zamechnik, *J. Biol. Chem.* **1958**, 231, 241.
- 493 D. Nathans, G. Notani, J.H. Schwatz, N.D. Zinder, *Proc. Natl. Acad. Sci. USA* **1962**, 48, 1424.
- 494 B.E. Roberts, B.M. Paterson, *Proc. Natl. Acad. Sci. USA* **1973**, 70, 2330.
- 495 T. Suryanarayana, A.R. Subramanian, *Biochemistry* **1983**, 22, 2715.
- 496 A.S. Spirin, V.I. Baranov, L.A. Ryabova, S.Y. Ovodov, Y.B. Alakhov, *Science* **1988**, 242, 1162.
- 497 Y. Wu, D.Y. Zhang, F.R. Kramer, *Proc. Natl. Acad. Sci. USA* **1992**, 89, 11789.
- 498 I.Y. Morozov, V.I. Ugarov, A.B. Chetverin, A.S. Spirin, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 9325.
- 499 M. Hoffmann, C. Nemetz, K. Madin, B. Buchberger, *Biotechnol. Annu. Rev.* **2004**, 10, 1.
- 500 V.A. Shirokov, A. Kommer, V.A. Kolb, A.S. Spirin, *Methods Mol. Biol.* **2007**, 375, 19.
- 501 Y. Endo, T. Sawasaki, *Curr. Opin. Biotechnol.* **2006**, 17, 373.
- 502 D. Mendel, V.W. Cornish, P.G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **1995**, 24, 435.
- 503 L. Wang, P.G. Schultz, *Angew. Chem. Int. Ed.* **2005**, 44, 34.
- 504 W. Gerhartz, *Enzymes in Industry* VCH, Weinheim, **1990**.
- 505 B. Zhang, T.R. Cech, *Chem. Biol.* **1998**, 5, 539.
- 506 M. Bergmann, H. Fraenkel-Conrat, *J. Biol. Chem.* **1937**, 119, 707.
- 507 W. Kullmann, *Enzymatic Peptide Synthesis* CRC Press, Boca Raton, **1987**.
- 508 H.-D. Jakubke in: *The Peptides: Analysis, Synthesis, Biology*, Volume 9, Chapter 3, S. Udenfriend, J. Meienhofer (Eds.), Academic Press, New York, **1987**, 103.
- 509 F. Bordusa, H.-D. Jakubke, in: *Houben-Weyl: Synthesis of Peptides and Peptidomimetics*, Volume E22a, M. Goodman, A. Felix, L. Moroder, C. Toniolo (Eds.), Georg Thieme Verlag, Stuttgart, **2002**, 642.
- 510 F. Bordusa, *Chem. Rev.* **2002**, 102, 4817.
- 511 C. Lombard, J. Saulnier, J.M. Wallach, *Protein Peptide Lett.* **2005**, 12, 621.
- 512 I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, 27, 157.
- 513 V. Schellenberger, H.-D. Jakubke, *Angew. Chem. Int. Ed.* **1991**, 30, 1437.
- 514 A. Fischer, A. Bommarius, K. Drauz, C. Wandrey, *Biocatalysis* **1994**, 8, 289.
- 515 A. Flörshheimer, M.-R. Kula, H.J. Schütz, C. Wandrey, *Biotechnol. Bioeng.* **1989**, 33, 1400.

- 516 G. Hermann, A. Schwarz, C. Wandrey, G. Knaup, K. Drauz, H. Berndt, *Biotechnol. Appl. Biochem.* **1991**, 13, 346.
- 517 R.A. Persichetti, N.L.S. Clair, J.P. Griffith, M.A. Navia, A.L. Margolin, *J. Am. Chem. Soc.* **1995**, 117, 2732.
- 518 Y.-F. Wang, K. Yakolevsky, A.L. Margolin, *Tetrahedron Lett.* **1996**, 37, 5317.
- 519 D. Ullmann, H.-D. Jakubke, in: *Proteolytic Enzymes, Tools and Targets*, E.S. Sterchi, W. Stöcker (Eds.), Springer, Berlin, **1999**, 312.
- 520 T. Harada, S. Irino, Y. Kunisawa, K. Oyama, EP 768 384 **1997**, *Chem. Abstr.* **1997**, 126, 316.
- 521 J. Markussen, *Human Insulin by Tryptic Transpeptidation of Porcine Insulin and Biosynthetic Precursors* MTP Press, Lancaster, **1987**.
- 522 J.S. Dordick, *Enzyme Microb. Technol.* **1989**, 11, 194.
- 523 P.J. Halling, U. Eichhorn, P. Kuhl, H.-D. Jakubke, *Enzyme Microb. Technol.* **1995**, 17, 601.
- 524 U. Eichhorn, A.S. Bommarius, K. Drauz, H.-D. Jakubke, *J. Peptide Sci.* **1997**, 3, 261.
- 525 M. Erbeltinger, X.W. Ni, H.J. Halling, *Enzyme Microb. Technol.* **1998**, 23, 141.
- 526 M. Hänsler, H.-D. Jakubke, *Amino Acids* **1996**, 11, 379.
- 527 C.-H. Wong, *Chimia* **1993**, 47, 127.
- 528 D.Y. Jackson, J. Vurnier, C. Quan, M. Stanley, J. Tom, J.A. Wells, *Science* **1994**, 266, 243.
- 529 V. Schellenberger, H.-D. Jakubke, N.P. Zapevalova, Y.V. Mitin, *Biotechnol. Bioeng.* **1991**, 38, 104.
- 530 H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Tetrahedron Lett.* **1997**, 38, 1777.
- 531 F. Bordusa, D. Ullmann, C. Elsner, H.-D. Jakubke, *Angew. Chem. Int. Ed.* **1997**, 36, 2473.
- 532 R. Günther, S. Thust, H.-J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, 267, 3496.
- 533 V. Cerovsky, F. Bordusa, *J. Peptide Res.* **2000**, 55, 325.
- 534 B. Yoo, K. Kirshenbaum, *J. Am. Chem. Soc.* **2005**, 127, 17132.
- 535 N. Wehofskey, S.W. Kirbach, M. Hänsler, J.-D. Wissmann, F. Bordusa, *Organic Lett.* **2000**, 2, 2027.
- 536 F. Bordusa, in: *Enzymatic Formation of C-N Bonds in: Bioorganic Chemistry Highlights II: From Chemistry to Biology* C. Schmuck, H. Wennemers (Eds.), Wiley-VCH, Weinheim, **2004**.
- 537 F. Bordusa, *ChemFiles* **2005**, 6, 13.
- 538 S.A. Sieber, M.A. Marahiel, *Chem. Rev.* **2005**, 105, 715.
- 539 K. Watanabe, Y. Toh, K. Suto, Y. Shimizu, N. Oka, T. Wada, K. Tomita, *Nature* **2007**, 449, 867.
- 540 P. Nissen, J. Hansen, N. Ban, P.B. Moore, T.A. Steitz, *Science* **2000**, 289, 920.
- 541 L. Pauling, *Am. Sci.* **1948**, 36, 51.
- 542 W.P. Jencks, *Catalysis in Chemistry and Enzymology* McGraw-Hill, New York, **1969**.
- 543 R.A. Lerner, S.J. Benkovics, P.J. Schultz, *Science* **1991**, 252, 659.
- 544 D. Hilvert, G. MacBeath, J.A. Shin, in: *Bioorganic Chemistry: Peptides and Proteins*, S.M. Hecht Oxford University Press, Oxford, **1998**.
- 545 R. Hirschmann, A.B. Smith III, C.M. Taylor, P.A. Benkovic, S.D. Taylor, K.M. Yager, P.A. Sprengeler, S.J. Benkovic, *Science* **1994**, 265, 234.
- 546 D.B. Smithrud, P.A. Benkovic, S.J. Benkovic, C.M. Taylor, K.M. Yager, J. Witherington, B.W. Philips, P.A. Sprengler, A.B. Smith III, R. Hirschmann, *J. Am. Chem. Soc.* **1997**, 119, 278.
- 547 J.R. Jacobsen, P.G. Schultz, *Proc. Natl. Acad. Sci. USA* **1994**, 91, 5888.

5

Synthesis Concepts for Peptides and Proteins

5.1

Strategy and Tactics

Before synthesizing larger peptides (and even proteins), one of the most important prerequisites to be performed is a thorough planning with regard to the order of peptide bond-forming steps, and also the choice of protection scheme(s) to be used. It may also be useful to consider the general method of peptide chain assembly. Bodanszky made his first recommendations for the design of schemes for peptide synthesis in 1966, but it was not until 15 years later that he suggested the terminology used, and especially the terms “strategy” and “tactics”, which may have caused some “blurring” of his original concept [1].

Originally, the term “strategy” was coined for the overall planning of peptide synthesis, whilst “tactics” was thought to comprise the selection of protecting groups and coupling methods. There is, however, no consensus among peptide chemists concerning the usage of the term, strategy. In principle, peptides can be synthesized either by stepwise chain assembly, starting preferentially from the C-terminus, or by the condensation of peptide segments. The Merrifield method in particular has often been referred to as a “solid-phase strategy”, although this technique in its original form is simply a variant of the linear (stepwise) coupling of amino acids in the C → N direction. In addition, the two major protection schemes in solid-phase polypeptide synthesis (SPPS) – the Boc/Bzl and Fmoc/tBu approaches – have occasionally been raised to the status of a strategy, despite being only categories of protecting group tactics. In general, there is no advantage in exaggerating the discussions regarding terminology; rather, it is more important to circumvent unforeseen problems of peptide and protein synthesis by utilizing not only the extreme flexibility of the global concept, but also the experience and creativity of peptide chemists in the laboratory.

5.1.1

Linear or Stepwise Synthesis

This strategy is defined as the incremental addition of each amino acid in the sequence, until the entire sequence of the target peptide has been assembled.

Independently of the medium in which a synthesis is performed, the strategy is considered to be “linear” when stepwise chain elongation starts at either the C-terminal or N-terminal residue (Figure 5.1).

The stepwise chain elongation starting at the N-terminal residue may occasionally be referred to as “inverse peptide synthesis” (Figure 5.1), and is in agreement with Nature’s way of synthesizing proteins (cf. Section 3.2.1). Moreover, it has the advantage that only the N-terminal amino acid residue requires an N^α -protecting group, and this is removed on the completion of chain assembly. The permanent risk of racemization or epimerization (cf. Section 4.4) at the stage of the dipeptide or longer peptides remains a serious drawback of chemical coupling procedures, but this disadvantage can be completely circumvented using enzymatic coupling procedures (cf. Section 4.6.2), as shown in Figure 5.2.

Bordusa *et al.* [2] synthesized the tetrapeptide Z-Lys-Tyr-Arg-Ser-OBzl as a model system for stepwise enzymatic synthesis in the $N \rightarrow C$ direction, without any side chain protection, and starting with Z-Lys-SBzl and repetitive enzymatic coupling (Figure 5.2, reactions 1 to 3) of the following three amino acid alkyl esters. The terminal blocking groups were cleaved by catalytic hydrogenation, and the free tetrapeptide H-Lys-Tyr-Arg-Ser-OH could be obtained in an overall yield of 62% and purity >98% (by HPLC).

Stepwise chemical peptide synthesis in the $N \rightarrow C$ direction has been described both in solution and as a solid-phase variant. An interesting approach to synthesis

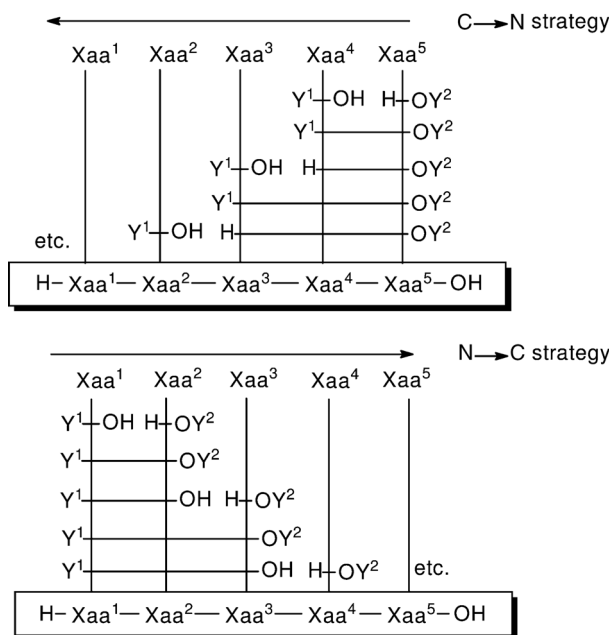


Figure 5.1 Strategy of linear (step-wise) chain elongation. Y^1 = N^α -amino protecting group, Y^2 = C^α -carboxy protecting group.

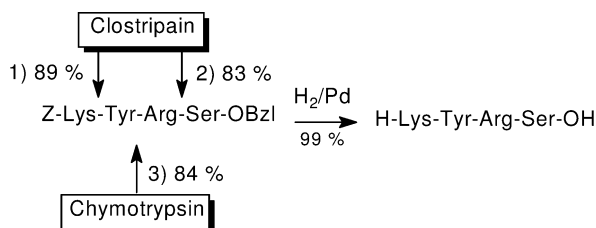


Figure 5.2 Enzymatic synthesis of a tetrapeptide in the $N \rightarrow C$ direction without any side-chain protection.

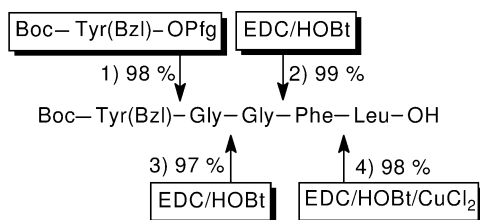
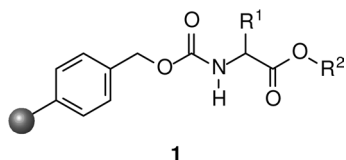


Figure 5.3 Synthesis of [Leu]enkephalin by a linear $N \rightarrow C$ strategy using mainly free amino acids as amino components.

in solution using free amino acids as amino components has been described by Mitin *et al.* [3, 4].

As shown in Figure 5.3, protected [Leu]enkephalin was synthesized in a model reaction starting with Boc-Tyr(Bzl)-OPfg followed by coupling of the next three free amino acids (Gly, Phe and Leu) using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) in the presence of HOBt. This coupling agent achieves high yields in every coupling step and, in the case of the last coupling reaction, $CuCl_2$ as an additional additive suppresses racemization to a significant degree. The problem of the low solubility of free amino acids in aprotic solvents has been solved by the application of a 3 M solution of $Ba(ClO_4)_2$, $Ca(ClO_4)_2$ or $Ca(ClO_4)_2$ in dimethylformamide, yielding 0.4 to 10 M solutions of the amino acids, depending on their structure. The application of this principle in solid-phase synthesis should offer several interesting advantages, provided that the general procedure can be applied for all amino acids without side reactions.

A solid-phase variant of a linear synthesis starting from the N-terminus was described by Letsinger and Kornet [5] as early as 1963. For this purpose, the N-terminal starting amino acid ester was linked to the polymeric resin via a benzyloxycarbonyl moiety 1.



After saponification of the ester, the next amino acid residue was coupled via a mixed anhydride. Despite further modifications, this procedure did not find wide application. A new variant of inverse synthesis using HOBt salts of amino acid 9-fluorenylmethyl esters has been described [6]. The first amino acid was attached via a trityl linker on Tentagel resin. The lowest degree of racemization was observed using *N,N'*-diisopropylcarbodiimide (DIC) as coupling agent, but the application of TBTU/NMM led to increased racemization. Another interesting attempt at SPPS in the reverse ($N \rightarrow C$) direction was described by Albericio's group [7]. This approach is based on the application of 2-Cl-trityl resin, allyl ester as the temporary protecting group, and either $\text{Cu}(\text{OEt})_2/\text{DIPCDI}$ or HATU/DIPEA as the coupling method.

Stepwise chain elongation starting with the C-terminal residue (Figure 5.1) requires the incorporation of each residue in blocked and activated form, and removal of the N^α -protecting group after each chain elongation cycle. Although this type of strategy seems to be too laborious for practical purposes, the advantages are: (i) the application of urethane-type N^α protection generally prevents racemization of the amino acid residue it is attached to; and (ii) the use of an acylating agent in high excess drives the reaction to near-completion. Bodanszky and du Vigneaud [8] demonstrated the efficiency of this approach in a stepwise synthesis of oxytocin. The application of active esters (*p*-nitrophenyl ester in the case of the oxytocin synthesis) of blocked amino acids for the incorporation offers the advantage that the excess reagent remains mainly unchanged and can be recovered. Despite the successful application of the incremental chain elongation in the syntheses of oxytocin, and even for the 27-peptide porcine secretin [9], the need for long series of coupling and deblocking procedures did not convince the majority of users about such advantages. In practice, under large-scale conditions peptides of up to five or six amino acids in length are generally synthesized by the stepwise approach in solution. However, the consequence has been the development of the solid-phase synthesis by Merrifield (cf. Sections 4.5 and 5.3.1). The superior performance of the stepwise $C \rightarrow N$ strategy under solid-phase conditions resulted in the method's full appreciation and led to broad application.

5.1.2

Convergent Synthesis

Convergent synthesis, in the past also termed segment condensation (fragment condensation), is defined as the construction of the target structure by final assembly of separately synthesized intermediate segments. Fragment assembly has been further subdivided into

- convergent synthesis of fully protected fragments
- chemoselective ligation of unprotected fragments.

For constructing proteins of novel design and structure a third approach, termed directed assembly, has been developed, in which individual peptide strands are non-covalently driven to associate into protein-like structures.

Assuming, for example, that the formation of each peptide bond could be achieved with a yield of 80%, in the stepwise synthesis of an octapeptide the overall yield is 21%, whereas a segment condensation strategy, involving four dipeptides that are coupled pairwise to give two tetrapeptides followed by final assembly to the target octapeptide, results in an overall yield of 51%. Although, based on this formal consideration, the segment condensation approach seems superior, such a conclusion is not valid because other factors also determine the success of the synthesis.

The convergent synthesis generally allows greater flexibility in the choice of protecting groups and coupling methods. However, in the synthesis of larger peptides this strategy is sometimes impeded by:

- an increased risk of racemization
- poor solubility of larger protected intermediate segments
- low coupling rate, with a concurrent risk of side reactions.

In order to exclude or minimize racemization, it is advantageous to have Gly or Pro as the C-terminal residue of a protected peptide segment, or to use methods of peptide bond formation without or with minimal risk of racemization, including racemization-free enzymatic approaches.

In convergent synthesis, a judicious dissection of the target sequence will greatly improve the results. As mentioned above, optimum segments should contain C-terminal Gly or Pro residues, but if this is not possible then Ala or Arg should be chosen as C-terminal residues as they are less prone to racemization. With regard to the size of the peptides synthesized by linear or convergent synthesis, there is a difference between laboratory-scale synthesis and large-scale synthesis performed on an industrial level. As a rule, in large-scale synthesis the segments should (preferably) be no longer than approximately five amino acids, which is about the same size as peptides generally manufactured by linear synthesis [10].

The ‘convergence’ of suitable fragments can be performed either in solution or on a solid phase. The benefit of convergent synthesis is that fragments of the desired peptide target are synthesized, purified and characterized, ensuring that each fragment is of high integrity. In this way the cumulative effects of stepwise synthetic errors are minimized.

5.1.3

Tactical Considerations

5.1.3.1 Selected Protecting Group Schemes

The tactics of peptide synthesis comprise selection of the optimum combination of protecting groups and the most suitable coupling method for each peptide bond-forming step. Even when synthesizing the simplest peptide in a controlled manner, it is essential that certain functional groups are protected. Tactical issues in segment condensation strategies are much more complex. First, it is of great importance to introduce – and, more importantly, to remove – all protecting groups under conditions that do not damage the integrity, and especially the stereochemical purity, of the peptide to be synthesized. As mentioned in Sections 4.2 and 4.5.3, a differentiation

must be made between temporary amino-protecting groups and semipermanent side-chain-protecting groups. The latter must be stable both during the repetitive cleavage reactions of the temporary protecting groups and during the coupling reactions. The third category of protection applies to the C-terminus of the peptide. The demands on C-terminal protecting groups are very high. On the one hand, they must be stable throughout the whole synthesis route, but on the other hand C-terminal protecting groups are sometimes required which can be removed in the presence of all other protecting groups. This is particularly important if a fully protected peptide segment is required to be coupled at its C-terminal amino acid residue. Especially, in large-scale solution synthesis the most suitable C-terminal protecting group is the free acid itself, though this limits the choice of coupling agents.

The selection of methods for protection and deblocking in linear synthesis requires special attention due to the repetitive character. Acidolysis at two widely different levels of acidity is commonly applied. Accordingly, the removal of temporary α -amino-protecting groups is often performed with trifluoroacetic acid (TFA), while the deblocking of side-chain-protecting groups is performed with liquid HF. However, this system of differential acidolysis generally presents increasing problems in the synthesis of larger peptides.

A better compatibility of protecting groups is based on the so-called orthogonality of protecting groups, although the term “orthogonal” has nothing to do with the absolute or relative geometries of protecting groups according to the normal meaning of “right-angled” or “situated at right-angles”. Orthogonal protecting groups will be completely removed by one reagent, without affecting the other groups. For example, complete chemical selectivity allows for catalytic hydrogenolysis of the Z group in the presence of *tert*-butyl-type side-chain-protecting groups, that in turn are cleaved by mild acidolysis. Furthermore, the Boc group can be used as an N^α -protecting group in combination with benzyl-based side-chain protection. Another orthogonal protecting group combination that relies on two different chemical cleavage mechanisms is the Fmoc/*t*Bu scheme, characterized by the base-labile Fmoc group as temporary protective group and acid-labile permanent *tert*-butyl side-chain protection. However, the N^α -Fmoc protection is also not free of problems due to sequence-dependent Fmoc deprotection inefficiency and premature deprotection in the course of chain assembly [11]. It is of great advantage that several research groups [12–16] have been engaged in the “fine-tuning” of side-chain-protecting groups, such that both types of chemistry are undergoing steady refinements.

The solubility of intermediate segments in organic solvents plays a fundamental role in the synthesis of large peptides and proteins. Besides the safe maximum protection scheme with a global masking of all functional groups, the minimum protection tactics have been developed for reducing solubility problems and minimizing the number of synthesis steps. There is no doubt that the minimal side-chain protection approach is desirable for large-scale synthesis as it minimizes the number of synthesis steps, though in practice it is not possible to omit side-chain protection completely. Unprotected ϵ -amino functions cannot be used in chemical segment condensations. Furthermore, it is necessary to protect also the thiol function

of cysteine, preferentially with the Ac_m group. The latter can be removed by iodine with concomitant disulfide formation. In addition, the trityl residue is another thiol-protecting group that, in combination with Ac_m, allows site-directed cyclizations. The carboxy groups of Asp and Glu must also be protected. From a practical point of view, it should be stated that the more hydrophilic properties of minimal side-chain-protected peptides do not promote extractive work-up procedures. Generally, there is no doubt that the formation of unexpected side products in the course of chemical operations with minimally protected peptides cannot be excluded. Taking this into consideration, the maximum protection approach seems to be the preferred tactical variant both in the chemical solution-phase synthesis and in solid-phase synthesis.

There are two main protection schemes (protocols) that have been used in SPPS:

- Boc/Bzl protection protocol, also termed Boc/Bzl chemistry or Merrifield tactics
- Fmoc/tBu protection protocol, also termed Fmoc/tBu chemistry or Sheppard tactics.

In the Boc/Bzl scheme side-chain protecting groups include ether, ester, and urethane-type derivatives based on benzyl alcohol, fine-tuned either with electron-donating methoxy or methyl groups, or with electron-withdrawing halogens, resulting in a proper level of acid stability or lability. Alternatively, ether and ester derivatives based on cyclopentyl or cyclohexyl alcohol have been used in special cases. In summary, the side-chain protecting groups must be stable to repeated cycles of Boc detachment, yet be completely cleaved by anhydrous hydrogen fluoride [17] or trifluoromethane sulfonic acid (TFMSA) [18]. From the mechanistic point of view, a distinction can be made between the high HF procedure (S_N1 mechanism) and the low HF procedure (S_N2 mechanism) [19].

The Fmoc/tBu chemistry uses the Fmoc group for N^α-amino protection, which is usually removed with piperidine in DMF or *N*-methylpyrrolidine. Compatible side-chain protecting groups are primarily ether, ester and urethane-type derivatives based on *tert*-butanol. The side-chain-protecting groups are removed at the same time as the appropriate anchoring linkages by TFA acidolysis [20]. The chemistry of both procedures has different features and different problems [21]. The milder conditions of the Fmoc/tBu protocol have led to its being preferred by a majority of peptide chemists [22]. However, certain deleterious side reactions are more prevalent in the Fmoc/tBu chemistry [23]. For a large-scale solution synthesis the Fmoc group is less attractive, because of the lack of volatility and the reactivity of the dibenzofulvene by-product. Selected protecting group (PG) schemes used in the maximum protection approach are listed in Table 5.1.

The maximum Boc/Bzl/Pac protection scheme, pioneered by Sakakibara [24], is characterized by simple deprotection with HF; moreover, the by-products are easily separated. The Ac_m group is stable to HF and allows additional purification of a peptide intermediate before folding takes place. The cyclohexyl ester (OCy) group resists aspartimide formation during coupling reactions [25]. The 4-methylbenzyl [Bzl(4-Me)] moiety as a protecting group for Cys permits the isolation of peptides with free thiol groups after a single treatment with

Table 5.1 Selected protecting group schemes for the maximum protection approach.

	Fmoc/tBu protocol	Boc/Bzl protocol	Boc/Bzl/Pac protocol
Temporary protecting groups	Fmoc	Boc	Boc
Permanent protecting groups			
Asp/Glu	OtBu	OBzl	OCy
Arg	Mtr/Pmc	Tos/Mts	Tos
Lys	Boc	Z(2-Cl)	Z(2-Cl)
His	Boc/Bom/Trt	Tos/Dnp	Bom
Cys	Trt/Tmb	Npys/Fm/Bzl(4-Me)	AcM
Ser	tBu	Bzl	Bzl
Thr	tBu	Bzl	Bzl
Tyr	tBu	Z(2-Br), 2,6-Cl ₂ Bzl	Z(2-Br), 3-Pn ^a
Trp	—	For	For, Hoc
Asn/Gln	Trt/Tmb	Xan	Xan ^a

^aEssential for the HMFS resin method.

HF [26, 27]. The Z(2-Br) group for the aromatic hydroxy function prevents the benzylation of Tyr during HF cleavage [28], and the Z(2-Cl) group is stable to TFA during Boc deprotection [29].

5.1.3.2 Preferred Coupling Procedures

A veritable arsenal of coupling methods exists for the coupling of the individual amino acids, as described in Section 4.3. The classical procedures, such as mixed carboxylic–carbonic anhydrides, symmetrical anhydrides, carbodiimides, azides and commercially available pre-activated amino acid derivatives, such as various active esters, N-carboxyanhydrides (NCAs) and urethane type-protected NCAs, known as UNCAs [30], have not lost their importance as coupling reagents. In particular, the pre-activated derivatives are compatible with unprotected C-terminal amino acid residues in the amino component, and have the benefit that by-products are not generated from the activating agent. More recently, onium (phosphonium and guanidinium/uronium) salts of benzotriazole derivatives have found increasing popularity. This has been primarily based on their superior reactivity, convenience and efficiency, as well as their minimal side reactions. These coupling reagents have been used especially in segment coupling reactions. The carbodiimide method, in combination with auxiliary nucleophiles as additives, such as HOBt, Cl-HOBt and HOAt, has also been used in large-scale segment coupling. However, the mixed anhydride method (with or without auxiliary nucleophiles) and the classical azide coupling procedure, that has advantages due to the low racemization sensitivity, have not lost their importance. Furthermore, amino components with unprotected C-terminal carboxy functions are compatible with the azide method, though the by-product N₃[−] gives rise to safety concerns in large-scale syntheses. The special features of large-scale peptide synthesis [10, 31] are discussed in Chapter 9, but details with regard to tactical coupling methods are outlined in the following sections.

5.2

Solution Phase Synthesis (SPS)

The first synthesis of oxytocin was performed by the Nobel laureate V. du Vigneaud and his coworkers [32] in 1953, and today the majority of peptide pharmaceuticals are prepared commercially using classical synthesis in solution. Indeed, despite the dominance of SPPS, peptide synthesis in solution remains a major approach to the preparation of peptides, and even proteins. Peptide synthesis in solution can be performed by both linear and convergent strategies. In principle, a stepwise synthesis in solution can be applied to small oligopeptides and segments up to about five amino acids, whereas SPPS is much more successful for the synthesis of longer peptides. However, in the solid/solution phase-hybrid approach (see Section 5.3.3) peptide segments of medium size are also synthesized on solid supports. Various features such as the complexity of the target molecule, the protection scheme chosen, and the economical as well as logistical considerations, determine the strategic route. Convergent peptide synthesis (CPS) in solution has been used for the large-scale production of small to medium-length peptides in quantities of up to hundreds of kilograms, or even metric tons per year; examples include inhibitors of angiotensin-converting enzymes (ACE) and HIV protease, analogues of luteinizing hormone-releasing hormone (LH-RH), oxytocin, desmopressin, and aspartame. Calcitonins obtained from various species and containing 32 amino acid residues are among the longest peptides manufactured for commercial application.

Using the classical solution procedure, the product can be purified and characterized at each step in the reaction. As assembly of the entire molecule can be performed with purified and well-characterized segments, the desired final product is isolated in a highly homogeneous form. Unfortunately, this major advantage requires the investment of a considerable amount of work and time although, when experienced, teams of workers can simultaneously synthesize several different segments. Within the context of convergent peptide synthesis in solution, it should be borne in mind that individual segments may also be synthesized using SPPS.

5.2.1

Convergent Synthesis Using Maximally Protected Segments

The low solubility of intermediate segments in the usual organic solvents is a major difficulty when synthesizing large peptides in solution. Indeed, Erich Wünsch, a pioneer of segment condensation in solution using fully protected segments, predicted about 30 years ago that this ideal strategy might involve an insolubility problem when synthesizing peptides in the range between 30 and 50 amino acid residues [33].

In 1981, Fujii and Yajima [34, 35] described the solution synthesis of the 124-residue protein ribonuclease A (RNase A) by coupling 30 relatively small-sized segments. The azide procedure was used for segment coupling, as at that time the method was believed to be devoid of racemization. After deprotection and protein folding, the enzyme was isolated in crystalline form and shown, after

chromatography, to have full biological activity. The procedure was problematic, however, because during the segment elongation process of every coupling step, a large excess (between 3- and 30-fold) of each carboxy segment was required. In addition to causing serious problems of insolubility of the intermediates, the necessary excess of carboxy segments and separation of by-products formed via Curtius rearrangement greatly increased purification difficulties.

Following the impressive synthesis of RNase A, Sakakibara's group at the Peptide Institute in Osaka began to investigate a general procedure for the solution synthesis of peptides containing more than 100 residues. In an impressive review, Sakakibara [24] provided a comprehensive description of this effective approach to the chemical synthesis of proteins, and predicted that linear peptide sequences consisting of more than 200 residues might be synthesized using this approach.

This actual concept of solution synthesis of proteins is based on new types of solvent systems for dissolving fully protected segments, in which segment condensation reactions can be performed smoothly in solution. A mixture of chloroform (CHL) and 2,2,2-trifluoroethanol (TFE) have been used for the solution synthesis of, e.g., midkine, a 121-aa protein [36], and human pleiotrophin, a 136-aa protein [37]. The entire target molecule is assembled from fully protected segments with a size of about 10 residues. This size fits the requirement for purifying segments with currently available methods (e.g., HPLC, column chromatography), and for characterizing homogeneity. Furthermore, each segment is designed to have a common structure of Boc-peptide-OPac (Pac, phenacyl), and all side-chain functions are protected by benzyl-type groups (cf. Table 5.1). Pro and Hyp are not suitable as N-terminal residues in the segments, but Val and Ile are still acceptable. However, Gly, Ala, Leu, Pro, and Hyp(Bzl) are suitable C-terminal residues, while Gln, Asn, Lys [Z(Cl)], Glu(Ocy), Asp(Ocy), Trp(For), and Ser(Bzl) might be acceptable. In contrast, His(Bom), Cys(Acm), Arg(Tos), Tyr[Z(2-Cl)], Tyr(3-Pn), Trp(Hoc), Ile, Val, Phe, and Thr(Bzl) must not occur in this position. An alternative removal of the Boc or OPac group yields the segments, which are coupled using the water-soluble 1-ethyl-3-(3'-dimethylaminopropyl)-carbodiimide (EDC) in the presence of 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (HODhbt) to give the entire sequence. The general principle of the Sakakibara approach is shown schematically in Figure 5.4.

A mixture of trifluoroethanol (TFE) in chloroform or dichloromethane (DCM) in a ratio of 1:3 (v/v) displays highly suitable properties as the coupling solvent, especially, when HODhbt is used as additive instead of HOBt.

The C $\rightarrow$ N strategy for the synthesis of the protected segments is shown schematically in Figure 5.5. The synthesis of the protected segments starts from Boc-Xaa¹-OPac, and the first coupling reaction is performed after removal of the Boc group.

As mentioned previously, once the Boc-peptide-OPac is obtained, it can be used alternately as an amino component after removal of the Boc group or as the carboxy component after cleavage of the OPac group. The Pac group can be removed under mild conditions by reduction with zinc in acetic acid at 40 or 50 °C [38], and it is stable to TFA.

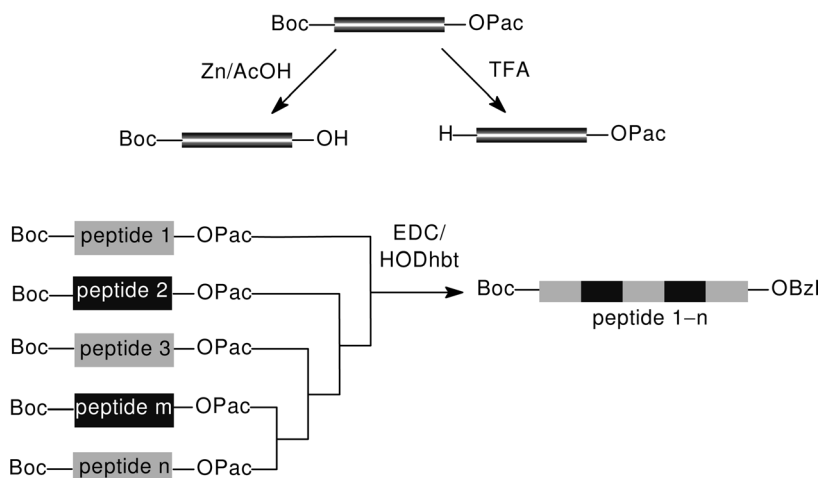


Figure 5.4 Alternative deblocking of Boc or Pac groups and segment condensation using EDC/HODhbt according to the Sakakibara approach.

In order to refine the strategy of this solution synthesis the Sakakibara group introduced a standard solid phase method on an *N*-[9-(hydroxymethyl)-2-fluorenyl] succinamic acid (HMFS) linker developed by Rabanal *et al.* [39] for preparing protected peptide segments (see Section 5.3.4).

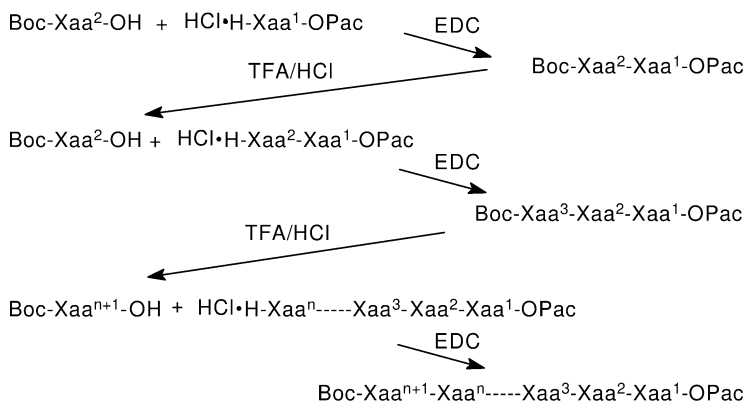


Figure 5.5 Sakakibara approach to segment synthesis starting from an amino acid phenacyl ester.

5.2.2

Convergent Synthesis Using Minimally Protected Segments

5.2.2.1 Chemical Approaches

As early as 1969, Hirschmann *et al.* [40] succeeded in synthesizing ribonuclease S (a protein containing 104 residues), using minimally protected segments, and in which only the side-chain functions of Cys and Lys were blocked. With the

exception of Trp, the sequence of the ribonuclease S protein contains all other trifunctional amino acids. For the synthesis of the segments in the range between tri- and nona-peptides, only the NCA/NTA method and HOSu procedure could be used. The azide method was mainly used for assembling the segments to larger ones. Only the final azide coupling of the two large segments with 44 and 60 residues led to a significantly lower yield.

As already observed in the classical ribonuclease S protein synthesis, the chemical coupling of minimally protected segments is associated with a permanent risk of unexpected side reactions. In principle, coupling of two unprotected segments requires the selective activation of the C-terminal carboxylic acid of one of the segments, without affecting any of the other carboxy groups present in either segment. Moreover, all other amino functions except the one involved in peptide bond formation must be blocked or deactivated. Some attempts have been made but an ideal solution of this very complicated problem has not been found.

The *thiocarboxy segment condensation method* pioneered by Blake [41] has an interesting feature, in which selective activation of the thiocarboxy group at the C-terminus by silver ions is achieved without affecting side-chain carboxy groups (Figure 5.6). Both segments to be coupled can be synthesized using Boc/Bzl SPPS.

In general, thioglycine is mainly chosen as the C-terminus in order to avoid racemization during the following segment condensation. After standard SPPS, the segments are detached from the resin by treatment with liquid HF. However, the N-terminus of the carboxy component must bear a protecting group Y that is stable under the strongly acidic cleavage conditions, such as Fmoc, Msc, Troc, iNoc, TFA, or Ac. The protection of other side-chain functions is not necessary, with the exception of free amino groups. Both α -inhibin containing 92 residues [42, 43], and β -lipotropin [44], were synthesized using this method. This elegant method has some limitations:

- thiocarboxylic acids are not stable toward oxidation and hydrolysis
- free side-chain amino groups in the segments can undergo undesired amide bond formation with the activated thiocarboxy group
- no procedure could be developed to synthesize cysteine-containing polypeptides based on the thiocarboxy segment condensation approach.

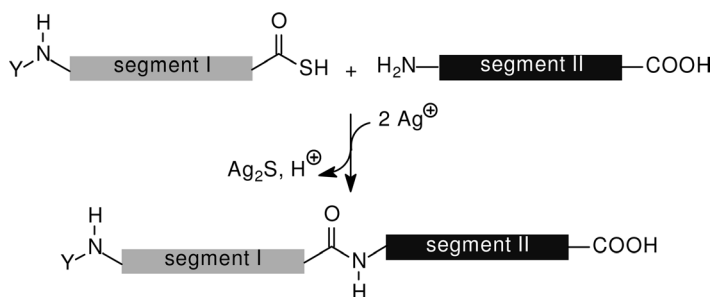


Figure 5.6 Schematic description of the thiocarboxy segment condensation method according to Blake [41].

Boc-OSu cannot be used for the essential blocking of side-chain amino groups due to side reactions with the highly nucleophilic thiol moiety of the thiocarboxy group, and the alternative citraconyl group lacks stability, even under conditions of mild acidity.

Aimoto *et al.* [45] have described a conceptually similar, but different and significantly improved, approach to polypeptide synthesis. The *Aimoto thioester approach* [46] is characterized by converting an S-alkyl thioester moiety in the presence of a silver salt (AgNO_3 or AgCl) into an active ester derived from HOBT or HODhbt, followed by segment condensation of partially protected segments. Peptide segment thioesters can be prepared via a Boc solid-phase method. The insertion of one amino acid between a thioester linker and an MBHA resin diminishes loss of the growing peptide during TFA treatment for Boc deblocking. The segments, except for the N-terminal ones, must be blocked by the iNoc group, this being removable by zinc dust in aqueous acetic acid, even in the presence of Boc groups. Following HF treatment, Boc-OSu was used as the reagent for protecting side-chain amino groups.

H-Ala-Gln-Val-Ile-Asn⁵-Thr-Phe-Asp-Gly-Val¹⁰-Ala-Asp-Tyr-Leu-Gln¹⁵-Thr-Tyr-His-Lys-Leu²⁰-
 -Pro-Asp-Asn-Tyr-Ile²⁵-Thr-Lys-Ser-Glu-Ala³⁰-Gln-Ala-Leu-Gly-Trp³⁵-Val-Ala-Ser-Lys-Gly⁴⁰-
 -Asn-Leu-Ala-Asp-Val⁴⁵-Ala-Pro-Gly-Lys-Ser⁵⁰-Ile-Gly-Gly-Asp-Ile⁵⁵-Phe-Ser-Asn-Arg-Glu⁶⁰-
 -Gly-Lys-Leu-Pro-Gly⁶⁵-Lys-Ser-Gly-Arg-Thr⁷⁰-Trp-Arg-Glu-Ala-Asp⁷⁵-Ile-Asn-Tyr-Thr-Ser⁸⁰-
 -Gly-Phe-Arg-Asn-Ser⁸⁵-Asp-Arg-Ile-Leu-Tyr⁹⁰-Ser-Ser-Asp-Trp-Leu⁹⁵-Ile-Tyr-Lys-Thr-Thr¹⁰⁰-
 -Asp-His-Tyr-Gln-Thr¹⁰⁵-Phe-Thr-Lys-Ile-Arg¹¹⁰-OH

2

For the synthesis of barnase 2, a bacterial RNase containing 110 amino acid residues, the sequence was divided into four partially protected segments:

- (A) Boc-[Lys(Boc)^{19,27}]Barnase(1–34)-S-C(CH₃)₂-CH₂-CO-Nle-NH₂
- (B) iNoc-[Lys(Boc)^{39,49}] Barnase(35–52)-S-C(CH₃)₂-CH₂-CO-Nle-NH₂
- (C) iNoc-[Lys(Boc)^{62,66}] Barnase(53–81)-S-C(CH₃)₂-CH₂-CO-Nle-NH₂
- (D) H-[Lys(Boc)^{98,108}]Barnase (82–110)-OH

These segments were coupled stepwise in DMSO using HOSu, AgNO_3 and NMM starting with C + D, followed by B + (C–D), and A + (B–C–D), respectively. Finally, barnase was obtained in 11% yield, based on the fragment D. Further examples of polypeptide synthesis are detailed in a review [46].

5.2.2.2 Enzymatic Approaches

A mutant of subtilisin BNP, termed subtiligase, was obtained by Jackson *et al.* [47] by protein design. Subtiligase was used in a further synthesis of RNase A. This approach combines chemical solid-phase synthesis of the segments and enzymatic

segment condensation. Starting from the C-terminal fragment RNase A(98–124), the further fragments bearing the N-terminal iNoc group (77–97, 64–76, 52–63, 21–51, and 1–20) were chosen such that the appropriate C-terminal amino acid residues (P_1 residues) of the fragments (Tyr⁹⁷, Tyr⁷⁶, Val⁶³, Leu⁵¹, and Ala²⁰) closely matched the substrate specificity for large and hydrophobic residues of the protease mutant. The first segment condensation is shown schematically in Figure 5.7.

In particular, an efficient leaving group of the acyl donor ester can provide high reaction rates, in combination with a decreasing risk of possible proteolysis of starting segments and the final product. The excellent leaving group of the Phe-NH₂-modified carboxamido methyl ester, which even in unmodified form was shown to be a highly efficient acyl donor moiety [48], and the application of a considerable excess of acyl donor segments ensured that, in the RNase fragment condensations, most of the possible side reactions could be minimized.

It is clear that the substrate mimetic approach (cf. Section 4.6.2.4) should also be useful for combining irreversible enzymatic ligation of segments with suitable methods of SPPS, especially in the preparation of acyl component peptide segments in the form of appropriate esters [49]. This methodology has also been used by Bordusa's group [50] in the semisynthesis of the biologically active 493–515 sequence of human thyroid regulatory subunit anchoring protein Ht31 via a-chymotrypsin-catalyzed (8 + 16) segment condensation via phenyl ester and 4-guanidinophenyl ester, respectively (Figure 5.8).

The resulting 24-peptide represents a minimum region of Ht31 required to bind to the regulatory subunit of cAMP-dependent protein kinase. The Ht31-(493–515)

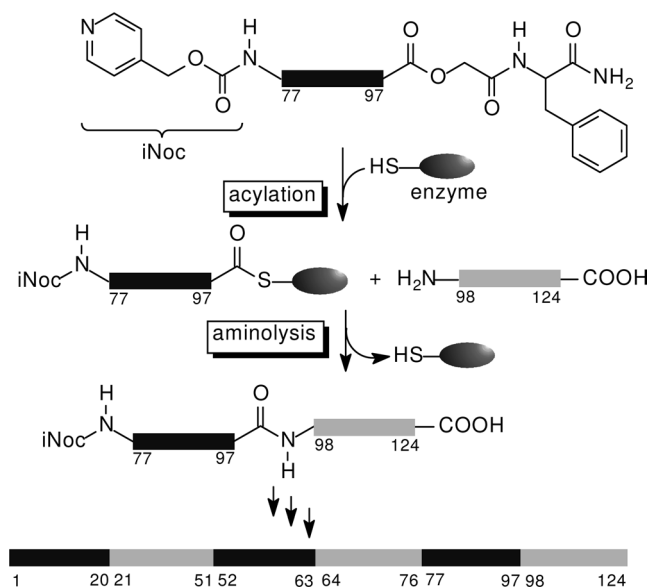
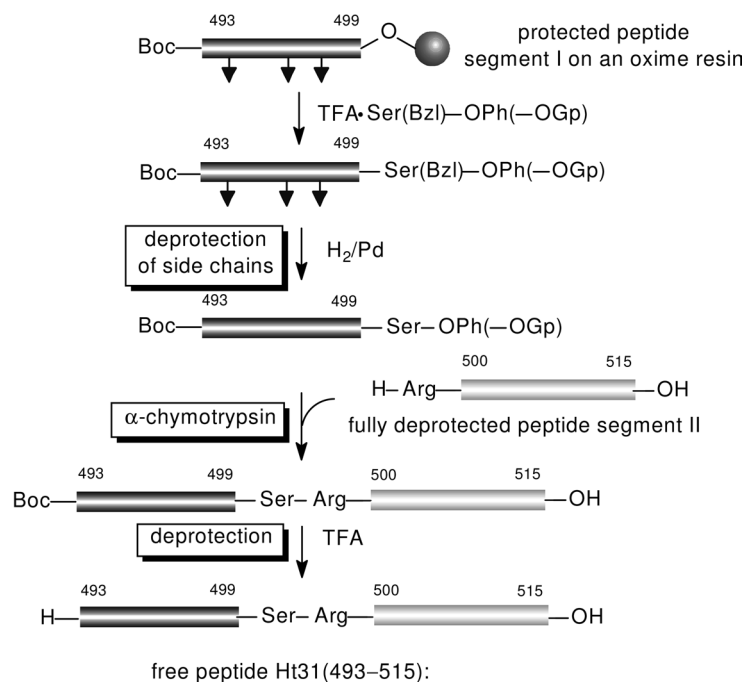


Figure 5.7 The first subtiligase catalyzed segment condensation within the enzymochemical synthesis of RNase A.



H-Asp-Leu-Ile-Glu-Glu-Ala-Ala-**Ser-Arg**-Ile-Val-Asp-Ala-Val-
-Ile-Glu-Gln-Val-Lys-Ala-Ala-Gly-Ala-Tyr-OH

Figure 5.8 Chemoenzymatic synthesis of the 24-peptide Ht31(493–515) using the substrate mimetic approach. O Gp = 4-guanidophenylxy.

peptide inhibited forskolin-dependent activation of a chloride channel in mammalian heart cells, thus suggesting an involvement of protein kinase A-anchoring proteins in this process. From a synthetic point of view, this example represents the first synthesis of a longer biologically active peptide that was prepared by the substrate mimetics-mediated semisynthetic approach.

An ingenious combination of chemical and enzymatic strategies, as was demonstrated in the synthesis of RNase A, should at present represent the state of the art in this field. However, the C–N ligation strategy based on the substrate mimetic concept allows, for the first time, the irreversible formation of peptide bond, catalyzed by proteases. When combined with frozen-state enzymology, as shown for model reactions [51], or using protease mutants with either minimal [52] or absent amidase activity, the substrate mimetics C–N ligation approach will contribute towards significant progress in enzymatic segment coupling. Fragment condensations using recombinant polypeptides generated as mimetic-based carboxy components, together with chemically synthesized or recombinant amino components, will doubtlessly broaden the application of this approach in the near future. This specific programming of enzyme specificity by molecular mimicry corresponds in practice

to the conversion of a protease into a C–N ligase, resulting in a biocatalyst that nature was incapable of developing evolutionarily.

5.3

Solution Phase/Solid Phase-Hybrid Approaches

In order to circumvent solubility problems in solution-phase segment condensation, some combined solution phase/solid phase approaches (SPS/SPPS-hybrid approaches) have been developed. In the early 1990s Riniker *et al.* [53–55] have described the first hybrid approach. Before going into details, it is necessary to describe the features of solid phase synthesis of protected segments.

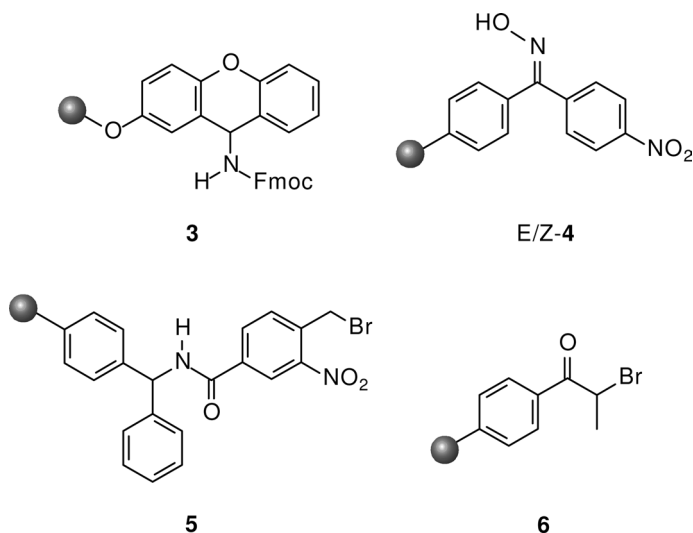
5.3.1

Solid Phase Synthesis of Protected Segments

The synthesis of protected segments on a solid support requires some modification of the chemistry involved in comparison to the linear SPPS, since in the latter case it is mainly the free target peptide that results following final detachment from the resin. Hence, procedures are required for detachment of the protected segment from the solid support, under mild conditions, at high yield and without affecting the configurational integrity of the C-terminal amino acid. A high degree of compatibility between the blocking groups of the segment and the peptide resin anchorage is of particular interest in this respect.

For this purpose handles and resins are necessary that can be cleaved by extremely mild acidolysis (cf. Section 4.5.1). The highly acid-labile 4-(4-hydroxymethyl)-3-methoxyphenoxybutyric acid (HMPB) handle found application in the first SPS/SPPS-hybrid approach [53]. Also Fmoc/tBu-based convergent SPPS (CSPPS) (cf. Section 5.4.2) is compatible with highly acid-labile resins and linkers. The 2-chlorotrityl resin allows the detachment of protected segments by treatment with mixtures of AcOH/TFE/DCM, or even, in the complete absence of acids, with hexafluoroisopropanol in DCM. The latter reagent excludes the contamination of the protected segment with a carboxylic acid, as such contamination can cause capping of the free amino group of the protected segment in coupling reactions. Furthermore, resins and linkers of the trityl type suppress diketopiperazine formation at the dipeptide stage of the synthesis – an unwanted side reaction that is especially pronounced with C-terminal Pro or Gly residues. A further highly acid-labile resin is known as SASRIN (Super Acid Sensitive ResIN, cf. Section 4.5.1). The detachment of protected peptides from SASRIN can be performed at high yields using 1% TFA in DCM. The highly acid-labile xanthenyl resin [56] **3** can be used as the C-terminal segment if a protected peptide amide is required. Boc/Bzl CSPPS requires base-labile solid supports, but photolysis and allyl transfer are two further options. The 4-nitrobenzophenone oxime resin **4**, also termed “Kaiser oxime resin” [57, 58], has been used extensively for the synthesis of Boc/Bzl-protected segments. Several methods have been described for the cleavage of protected

segments from the resin, including hydrazinolysis, ammonolysis, or aminolysis using suitable amino acid esters. Transesterification of the peptide resin with *N*-hydroxypiperidine, followed by treatment of the hydroxypiperidine ester with zinc in acetic acid, yielding the free carboxylic acid, is the preferred procedure. The only disadvantage is the lability of **4** towards nucleophiles (including the free N^α amino group of the growing peptide chain) during neutralization with base after the acidolytic deprotection of the Boc group. Photolytic cleavage of protected peptide segments from the solid support is possible using, for example, nitrobenzyl and phenacyl resins. 3-Nitro-4-bromomethylbenzhydrylamido-polystyrene **5**, known as Nbb-resin [59, 60], is fully compatible with the Boc/Bzl CSPPS, which allows detachment of the protected segment by irradiation at 360 nm in a mixture of DCM and TFE. The photolabile α -methylphenacyl ester resin **6** [61] is only compatible with the Boc/Bzl approach.



Although the Fmoc/tBu approach very often provides protected segments in high purity after cleavage from the resin, in other cases a thorough purification of the segments is essential before segment condensation. Preparative RP-HPLC using acetonitrile/water systems as the eluant is a suitable method for purification of the Fmoc/tBu-blocked segments [53, 62], as well as silica gel column chromatography. The poor solubility of the segments is generally a serious problem for effective purification; however, the solubility of protected peptide segments may be enhanced by backbone protection (see Section 4.5.4.3).

5.3.2

SPS/SPPS-Hybrid Condensation of Lipophilic Segments

This lipophilic segment coupling approach [53–55] has been applied to the synthesis of some medium-sized peptides, such as human calcitonin-(1–33), human neuropeptide Y, and the sequence 230–249 of mitogen-activated 70 K S6 kinase. Lipophilic protected

segments, that are relatively soluble in organic solvents (e.g., chloroform, tetrahydrofuran, ethyl acetate), can be obtained using a special, maximum protection chemistry. For this purpose the side chains of Asn, Gln, and His are always blocked with the Trt group, Arg with the Pmc group, and Trp with the Boc group. All other side-chain functions, with the exception of the thiol group of Cys, are protected with *tert*-butyl-type groups. Cys may also be blocked with the Trt residue, but if acid-stable Cys protection is required then AcM protection is necessary.

The synthesis of the lipophilic segments is performed by Fmoc/tBu SPPS on a resin bearing the highly acid-labile 4-(4-hydroxymethyl-3-methoxyphenoxy)-butyric acid (HMPB) handle. The length of the segments can vary up to approximately 20 amino acid residues, and preferably they should contain, if possible, Gly or Pro at the carboxy terminus.

The principle of the lipophilic segment coupling approach is shown schematically in Figure 5.9. As most peptide segments are soluble in *N*-methylpyrrolidone, the segment coupling reactions have been performed in this solvent using TPTU/HOBt. Most segment condensations reach completion within a few minutes at room temperature, and the resultant protected peptides are purified by flash or medium-pressure chromatography on silica gel. Finally, acidolytic deprotection is usually carried out in TFA/H₂O/ethanedithiol (76:4:20).

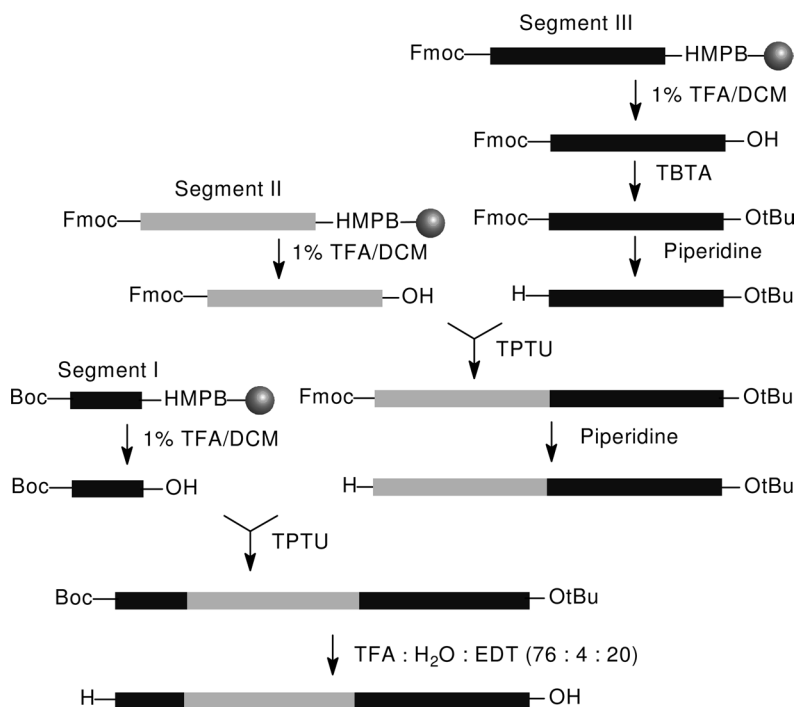


Figure 5.9 Principle of the lipophilic segment coupling approach according to Riniker *et al.* [53]. The amino acid side chains are protected as the tBu ethers (Ser, Tyr, Thr), tBu esters (Asp, Glu), Boc derivatives (Lys, Trp), Trt derivatives (Asn, Gln, His, Cys), or Pmc derivatives (Arg).

5.3.3

Phase Change Synthesis

The phase change synthesis method according to Barlos and Gatos [63] is another interesting hybrid approach. A benzyl alcohol linker attached to the xanthenyl resin allows C-terminal-protected esters to be prepared. Otherwise, segments synthesized by SPPS are cleaved from the resin, followed by subsequent esterification at the C-terminal carboxy group.

The phase change synthesis of ProT α -(59–75) according to Barlos and Gatos is shown in Figure 5.10. After cleavage of the protected segment (69–75) from the 2-chlorotrityl resin, the esterification is performed by treatment with 2-chlorotritylchloride (Clt-Cl) and diisopropylethylamine (DIPEA) in DMF/DCM mixtures [64]. The N-terminal Fmoc group is cleaved with piperidine, yielding the C-terminal segment for the condensation with the segment Fmoc-ProT α -(59–68)-OH in solution. The peptide is re-attached to the support after selective cleavage of the 2-chlorotrityl group from Fmoc-ProT α -(59–75)-Clt with 1% TFA in DCM, or alternatively with AcOH/TFE/DCM.

5.3.4

SPS/SPPS-Hybrid Approach to Protein and Large Scale Peptide Synthesis

The synthesis of the green fluorescent protein (GFP) precursor molecule with 238 residues by Sakakibara's group was the masterly performance of a highly sophisticated synthesis technique, and belongs to the state-of-the-art protein syntheses [24]. The entire sequence of GFP was divided into 26 segments, from which only four were synthesized by the classical solution procedure, whilst the major proportion

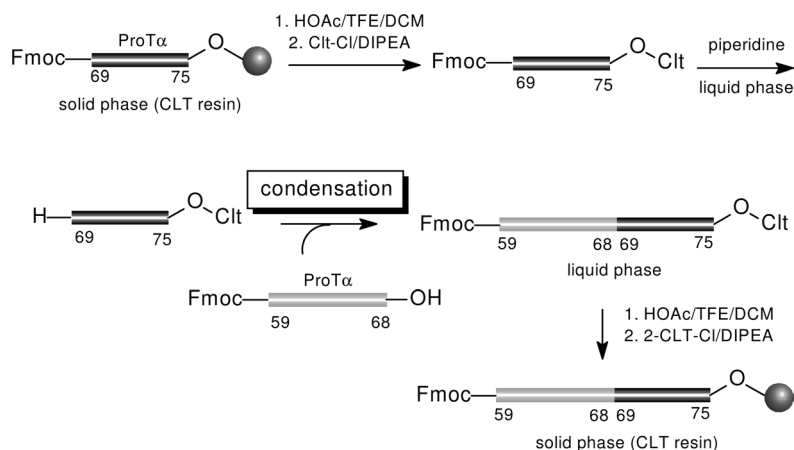
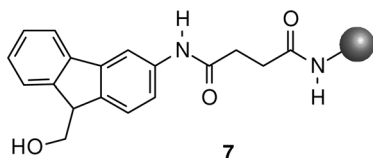


Figure 5.10 Phase change synthesis of the protected prothymosin α -(59–75) segment according to Barlos and Gatos [63].

was assembled on HFMS-resin [39] 7. The HMFS linker belongs to the 9-hydroxymethylfluorene type protecting groups which are designed to be cleaved by nucleophiles, e.g., piperidine and morpholine, via a β -elimination process. Consequently, a high compatibility between protecting groups of the segments and the anchoring group which is cleavable with morpholine or pyridine in DMF is required.



The coupling of the four large segments (Boc-(1–51)-OPac, Boc-(52–116)-OPac, Boc-(117–174)-OPac and Boc-(175–238)-OBzl) with preceding removal of the corresponding protecting groups resulted in the two final segments, Boc-(1–116)-OPac and Boc-(117–238)-OBzl. The last segment condensation in chloroform/TFE (3:1) $\text{Boc-(1–116)-OH} + \text{H-(117–238)-OBzl} \rightarrow \text{Boc-(1–238)-OBzl}$ yielded the fully protected target molecule. Following cleavage of the terminal Boc group with TFA, the remaining product (465 mg) was treated with HF in the presence of Cys•HCl and *p*-cresol at -5°C for 1 h. Purification by reversed-phase (RP)-HPLC, followed by removal of the two remaining Ac groups, yielded the final product (25 mg). Finally, under these experimental conditions, about 10% of the synthetic GFP precursor was found to fold into the native GFP protein. Beside this protein several other large peptides and proteins have been synthesized, including the muscarinic toxin 1 (MTX1) [65].

An impressive example of a SPS/SPPS-hybrid synthesis process of a peptide drug in bulk is the large scale synthesis of T20 (enfuvirtide, FuzeonTM). This 39-peptide is the first example of a novel class of antiviral agents that inhibits the membrane fusion and specifically blocks the fusion of the AIDS virus from entering into human blood cells [66]. This synthesis, comprising over 100 chemical steps, is an example of a very large peptide drug production partly produced on a solid support. The 36-peptide has been divided into three segments (Figure 5.11) which are synthesized individually on Barlos resin (*o*-chlorotriptylchloride resin) (cf. Section 4.5.1) according to the Fmoc/tBu protocol using HBTU/Cl-HOBt as coupling agent (Figure 5.12). After SPPS of segment C' lacking the C-terminal Phe-NH₂ the sequence of segment C was completed by solution coupling this amino acid amide to the C-terminus. After cleavage of the Fmoc group the resulting peptide was

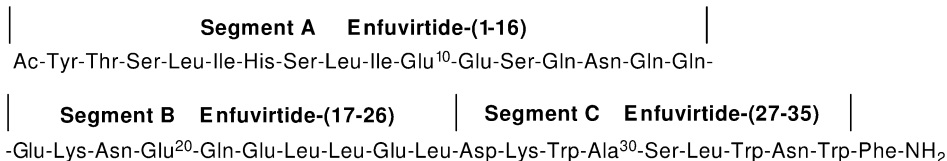


Figure 5.11 Primary structure of Enfuvirtide (Fuzeon, T20) and its segments.

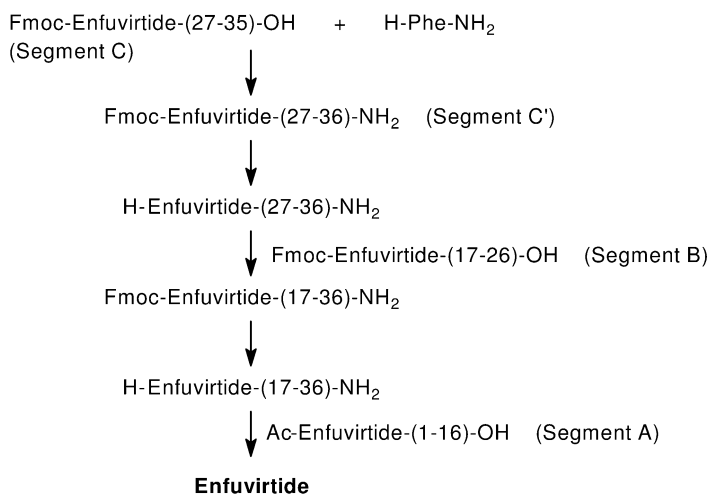


Figure 5.12 Simplified synthesis scheme of the large scale synthesis of Enfuvirtide (Fuzeon, T20).

coupled again in solution with segment B followed by N^α -amino group detachment and final SPS coupling with segment A completing the sequence of the fully protected peptide. All side-chain protecting groups were removed with TFA in the presence of the scavenger DTT (dithiothreitol) followed by isolation and purification.

5.4

Optimized Strategies on a Polymeric Support

5.4.1

Standard SPPS

The main repeating steps in stepwise elongation of a peptide chain are coupling and deprotection using preformed blocked amino acids. In conventional solution synthesis with equimolar amounts of reactants, the basic operations entail a large number of reactions at each stage, together with purification procedures (e.g., washing, crystallization), and the collection of solid products by filtration or centrifugation, and chromatography.

During the mid-twentieth century the need to facilitate the solution synthesis process had been clear for some time and it was in 1963 that R. B. Merrifield [67] first introduced stepwise peptide synthesis. This entailed having the C-terminal starting amino acid attached covalently to polystyrene beads in order to simplify the classical synthesis in solution (a detailed description is provided in Chapter 4). The original synthesizer developed by Merrifield (see Figure 4.44) is now located in the Smithsonian Museum. Developments in the field have been very rapid and today a number of commercial instruments are available which perform most of the synthesis steps under computer control.

The mechanization and automation of chain assembly are the ultimate goals of SPPS. Consequently, during the past 40 years a series of continuous developments and improvements have led to a revolution, not only in peptide synthesis, but also in organic synthesis – and especially in combinatorial synthesis [68–71].

In its early stages, Merrifield's concept roused a great degree of skepticism, based mainly on two objections. The first objection was that synthetic intermediates could not be purified during chain assembly, and purification and characterization of the synthesized product would be possible only on completion of the synthesis. The second objection was that nonquantitative reactions in N^α -deprotection and coupling, in connection with incomplete orthogonality between temporary and permanent protecting groups, would cause the production of truncated and deleted sequences. Since deletion peptides are structurally closely related to the desired peptide, their separation from the target product at the final purification step might be very difficult. The need for yields approaching 100% at every step is demonstrated with the data compiled in Table 5.2.

Until now, the chemistry involved in SPPS has been refined to such an extent that most of the reactions are performed repetitively and reproducibly in near-quantitative yields. This, together with a parallel refinement of analytical and purification techniques, underlines why, at present, the vast majority of peptides are synthesized by SPPS. It might be misleading however, to believe that most polypeptides (and even small proteins) can – either now or in the future – be synthesized efficiently and safely using automated synthesizers running standardized reaction protocols. With few exceptions, it can be assumed that carefully implemented stepwise SPPS protocols are effective in producing peptides which are up to 50 residues in length, and which can be further purified using the power of RP-HPLC. In general, this purification technique has been increasingly exploited as a manufacturing tool by the major pharmaceutical companies to produce a range of commercial products, including somatostatin, LH-RH, salmon calcitonin, and other analogues. By contrast, the simple principle of SPPS and its subsequent technical improvements have brought peptide synthesis within the scope of the undergraduate chemist or biochemist,

Table 5.2 Relationship between average yield per step and total yield depending on the number of residues of the peptide in SPPS.

Peptide	Number of residues	Total yield [%]			
		Yield [%] per step ^a			
		95	98	99	99.9
Growth hormone	191	0.006	2.2	15.0	82.8
Ribonuclease	124	0.2	8.3	29.1	88.4
Trypsin inhibitor (bovine)	58	5.4	31.6	56.4	94.5
Insulin A chain	21	35.8	66.8	81.8	98.0

^aRelated to the C-terminal amino acid.

albeit with dangerous consequences. Clearly, many syntheses of longer peptides up to the size of small proteins have been attempted, but the data have not been published because of failure or inconclusive results. Perhaps it should also be pointed out that, on occasion, commercially prepared custom-synthesized peptides have also been produced using multiple synthesis machines that are not under adequate analytical control. Consequently, it is possible that such products might contain undetected erroneous structures, or be contaminated with undesirable by-products [72, 73].

Despite SPPS being based on a simple principle, its operation requires the knowledge and expertise of an experienced peptide chemist in order to be successful. An example is in the area of the so-called “difficult sequences” [74–78], which clearly demonstrates the intense effort required to tackle SPPS under problematic conditions (cf. Section 4.5.4.3).

In summary it is clear that, while stepwise SPPS represents a very useful means of producing small quantities of peptides, solution synthesis is preferable for the preparation of larger quantities of peptides, and especially of the kilogram quantities required in industrial production. The accumulation of by-products that occurs during a long stepwise synthesis may be avoided by subdivision of the target polypeptide or protein into several segments, each of which is synthesized by linear SPPS. Following detachment of the segments from the resin in suitable protected form and subsequent purification, they can be used for assembly of the complete sequence (cf. Section 5.3).

5.4.2

Convergent Solid-Phase Peptide Synthesis

In addition to linear SPPS, segment condensation in solution and chemical ligation (cf. Section 5.5), convergent solid-phase peptide synthesis (CSPPS) has been developed to circumvent the problems caused by the poor solubility of protected segments in solution [79–82]. The assembly of segments to produce the target polypeptide or protein can be performed both in the $C \rightarrow N$ and $N \rightarrow C$ directions, or it may also start from a middle region and proceed in both directions [83]. However, the $C \rightarrow N$ direction has until now been the preferred method for the synthesis of proteins, as shown schematically in Figure 5.13.

Generally, the target protein must be divided into segments bearing (preferentially) Gly and Pro as the C-terminal amino acid in order to minimize the risk of epimerization during the coupling reaction.

In CSPPS, the resin-bound C-terminal segment is a prerequisite for high efficiency. In principle, it can be synthesized directly on the resin by stepwise SPPS, but to ensure the highest possible purity, the re-attachment of an independently synthesized, purified and characterized segment onto the resin offers a better approach. Concerning the nature of the resin, it has been reported that those that give the best results in standard SPPS are also well suited for CSPPS. Polystyrene and polyacrylamide resins give good results as well as polyethylene glycol-grafted polystyrene supports. The loading of the resin should be chosen to be lower than that used

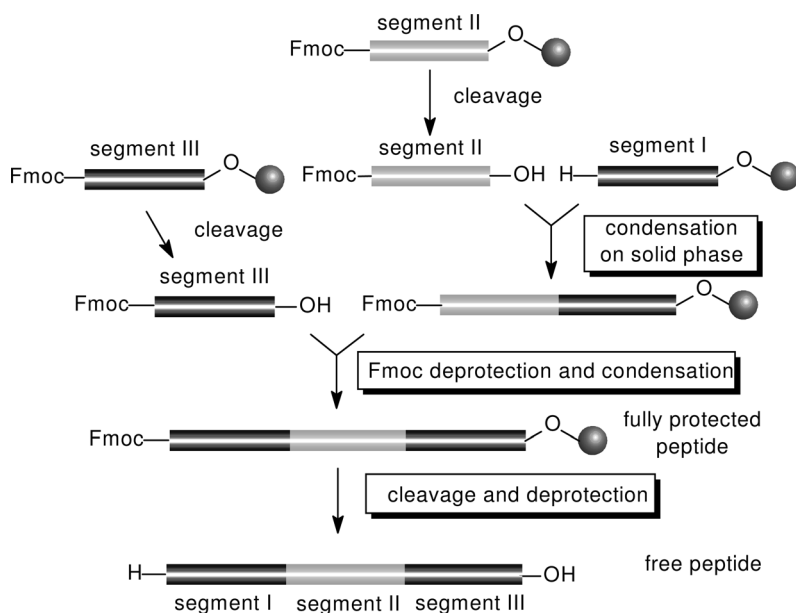


Figure 5.13 Principle of convergent solid phase peptide synthesis (CSPPS) starting from the C-terminus.

in linear SPPS [62, 84, 85]. As a rule, at the end of the segment condensation the weight ratio between the protected target polypeptide and the resin should exceed 1:2, and loading values between 0.04 and 0.2 mmol g^{-1} should fit these requirements. With a higher loading, the resin loses its swelling and polarity properties during peptide chain elongation.

As the condensation of protected segments with the resin-bound C-terminal segment depends to a greater extent on the concentration of the carboxy component than on the excess applied, solutions with the highest possible concentration should be used. The purity of solvents and reagents must be carefully checked for the very often long-lasting segment coupling reactions. There is no significant difference to linear SPPS with respect to coupling methods. Nowadays carbodiimides, usually in the presence of additives such as HOSu, HOBt, HOAt, and reagents based on onium salts in the presence of HOBt and HOAt, have been used successfully in CSPPS.

There is no doubt that epimerization at the C-terminal-activated amino acid of the segment is the most important side reaction (cf. Section 4.4). Placing Gly in this position excludes epimerization, which is also normally minimized in the case of Pro. Epimerization remains the main problem of CSPPS for all other residues, however [86]. Methods to quantify the extent of epimerization that might occur in a given segment coupling have been described [87, 88]. In a simple CSPPS model system, the lowest epimerization was found using DIC/HOBt as coupling reagent [89]. Barlos and Gatos [63] studied the extent of epimerization of the C-terminal residue Glu⁹⁴ during condensation of the prothymosin α (ProT α) segments ProT α -(87–94)

and ProT α -(95–109) using various coupling reagents and solvent systems. Epimerization was efficiently suppressed using carbodiimides and acidic additives in DMSO (D-Glu: <0.2%), whereas the use of onium salts, also in DMSO but in the presence of bases (e.g., diisopropylethylamine), resulted in an unacceptably high degree of epimerization (D-Glu: ~34%). The use of DMF as solvent or DCM as cosolvent in DMSO (50:50) for DCC/HOBt segment couplings led to a similar D-Glu percentage. Furthermore, it must be considered that epimerization is also heavily dependent on the reaction time and so long-lasting segment condensations will be accompanied by a high degree of epimerization, even though model studies conducted under identical conditions and normal reaction times have shown a lack of effect on the chiral integrity of the C-terminal residue of the carboxy component. For this reason, a short double coupling with the carboxy component, followed by acetylation, seems to be more effective than slow and incomplete condensation with high epimerization of the C-terminal residue of the carboxy component.

Monitoring of the coupling reaction during CSPPS can be performed using the Kaiser test, but this is less sensitive with increasing length of the peptide chain. Edman sequencing offers an alternative method to determine the yields of the segment couplings. The removal of an aliquot of resin, and cleavage of the peptide followed by analysis with RP-HPLC, mass spectrometry, capillary electrophoresis, or NMR, should be the most straightforward monitoring strategy.

The above-mentioned 109-residue protein prothymosin α is one of the largest target peptides synthesized by CSPPS [85]. The C-terminal 34-residue segment, ProT α -(76–109)-2-chlorotrityl resin was synthesized by linear Fmoc/tBu SPPS, and then assembled with nine segments, containing on average 7–10 amino acids each, to the fully protected Fmoc-ProT α -(1–109)-2-chlorotrityl resin. Prothymosin α was obtained in 11% overall yield after cleavage from the resin, deprotection and purification, and showed identical biological activity as the natural protein. Further examples of peptides synthesized by CSPPS are rat atrial natriuretic factor [90], β -amyloid protein [91], the N-terminal repeat region of γ -zein [92], and hirudin [93]. The latter hirudin variant 1 (HV1) containing 65 aa and three disulfide bonds was synthesized according to the Fmoc-based convergent protocol on 2-chlorotrityl resin. Seven protected segments were assembled on the resin-bound C-terminal segment HV1-(55–65).

5.4.3

Handle Approaches

Alternative approaches to simplify solution synthesis, and also to make it less time-consuming, were examined already before the invention of SPPS. Attempts have been directed towards using “soluble handles” in order to facilitate isolation and to avoid the full characterization of intermediates at each step.

5.4.3.1 Positively Charged Handles

In 1968, Young's group introduced an interesting combination of solid-phase synthesis and synthesis in solution [94]. These authors used the picolyl ester

moiety (see Table 4.2), which is quite resistant to acids, as a semipermanent blocking group for the C-terminal carboxy group. Hence, the Boc group could be used for protection of the α -amino function. The protonated picolyl handle increases the solubility of peptide intermediates in an aqueous medium, and allows isolation from the reaction mixture. This can be easily performed by passing the crude reaction mixture through an appropriate ion-exchange resin. The by-products are not retained, while the desired peptide can be recovered by elution with a suitable buffer. Both the coupling reaction and the deprotection are carried out in solution by well-established procedures. The picolyl ester method has been used in several syntheses, including that of complex peptides [95], but was superseded by SPPS, mainly because the process was not readily amenable to mechanization and automation.

Independently, Wieland and Racky [96] also developed positively charged handles with colored, ion-exchangable benzyl esters which allowed isolation of the desired peptide. A purification method which is the opposite of that introduced by the groups of Young and Wieland has been developed by Carpino *et al.* [97]. Tris-(2-aminoethyl)amine has been used as a substitute for 4-(aminomethyl)piperidine in the Fmoc/polyamine approach to rapid peptide synthesis.

5.4.3.2 Liquid Phase Method

In 1965, Shemyakin's group [98] proposed a polymeric support that is soluble in organic solvents as an alternative resin material for SPPS. High-molecular weight non-cross-linked polystyrene ($M_r \sim 200$ kDa) is chloromethylated in the usual way and esterified with the first Boc amino acid. The stepwise chain elongation from the C-terminal end is achieved with Boc amino acid *N*-hydroxysuccinimide esters. After completion of the necessary steps of one synthesis cycle, the intermediates are separated from solution by dilution with water. The resulting precipitate is intensively washed to remove low-molecular weight by-products. Although effective, this procedure suffers from some disadvantages, mainly related to the properties of the polystyrene.

In 1971, Mutter *et al.* [99] proposed polyethylene glycol (PEG), $\text{HO}-\text{CH}_2-\text{CH}_2-(\text{O}-\text{CH}_2\text{CH}_2)_n-\text{O}-\text{CH}_2-\text{CH}_2-\text{OH}$, or the monofunctional monoethyl ether of PEG as more efficient polymers that allow precipitation of the intermediates by addition of diethyl ether, thereby providing a simple method of purification after the coupling of each amino acid [100]. PEG is soluble in the majority of organic solvents. Deblocking and coupling reactions take place in homogeneous solution (cf. Figure 4.45). The solubilizing effect of PEG even permits coupling in aqueous solution by water-soluble carbodiimides (EDC), and allows monitoring of the intermediates by NMR spectroscopy. Although the liquid-phase method is versatile for stepwise and convergent synthesis, it has some inherent drawbacks and hence has not attracted the attention of automated synthesizer manufacturers. Furthermore, since the protected peptide intermediates are attached to heterogenous preparations of soluble, low-molecular weight PEG this approach should not be defined as a handle method. Liquid-phase synthesis intermediates cannot be characterized by standard chemical methods such as HPLC or mass spectrometry. However, the use of homogenous PEG preparations would convert this approach to a true handle method.

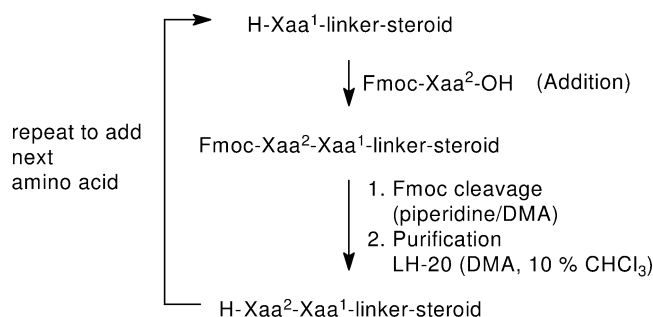


Figure 5.14 Principle of the EGP method.

5.4.3.3 Excluded Protecting Group Method

The excluded protecting group (EPG) method [101] is a handle approach in which the growing peptide is attached to a steroid or other homogenous high-molecular weight compound, making it different from other compounds of the reaction mixture. It offers the possibility of synthesizing homogenous protected segments at moderate prices. The principle of this approach is shown schematically in Figure 5.14.

In this method the C-terminal amino acid is esterified to the 2-position of cholestane as the (2*S*,3*S*)-iodohydrin ester and the penultimate amino acid is added to the aminoacyl-steroid as the Fmoc-pentafluorophenyl ester. After removal of the Fmoc group with Et₂NH/DMF (~ 15% v/v) and evaporation to ~ 10 ml the solution is chromatographed on Sephadex LH-20 in DMF. The dipeptidyl-steroid elutes as the free amine well separated from other components of the reaction mixture. Repetition of the coupling/deprotection/purification cycle yields the fully protected peptide. On completion of the synthesis, the cholestane iodohydrin ester can be selectively removed by treatment with Zn/AcOH yielding the peptide with intact α -amino and side-chain-protecting groups. HF can be used for global deprotection.

5.5

Chemical Ligation Strategies

As shown above, several chemical synthesis strategies have achieved success in the preparation of proteins of 75 to ~ 150 amino acids. Although approaches have been described that are technologically feasible to synthesize proteins exceeding this range, they have not reached the point of becoming routine and are not (yet) applicable to larger proteins. As unprotected peptides and proteins of different sizes can be obtained from either chemical or recombinant sources, their ligation to larger polypeptides and proteins should provide a unifying operational strategy for both total and semisynthesis of proteins.

As a consequence, suitable ligation chemistry is required to fill the gap that currently exists in the synthesis of proteins starting from unprotected segments. As shown above, the controlled formation of a peptide bond between two peptide

segments normally requires minimal or maximal protection schemes. Despite the progress in convergent protein synthesis, this approach faces several difficulties. The protected segments are often poorly soluble in aqueous systems used for liquid-chromatographic purification and, additionally, in organic solvents traditionally used for coupling procedures. Furthermore, the individual rates for segment coupling are substantially lower than those in stepwise synthesis, and racemization in segment coupling is a permanent risk when constructing segments that have neither a Gly nor a Pro C-terminus.

About two decades ago, an alternative to the convergent approach methods was developed that was capable of linking peptide segments. Protecting groups are less essential when a nonamide bond is used to link two peptide segments. The price to be paid for realizing this chemoselective ligation principle is the formation of a nonpeptide bond (non-natural structure) at the ligation site. Formally, in the sense of rational nomenclature, the resulting ligation products should be defined as protein analogues lacking one peptide bond. However, in practice, the non-natural structures are often well tolerated within a folded protein, and numerous examples of fully active proteins prepared by such a method underline this assumption. This approach is often called “chemical ligation” [102], though this term is somewhat confusing as ligation or chemical ligation describes, without any doubt, the preferential classical condensation of peptide segments joined by peptide bonds. Finally, protecting groups can be omitted totally if the reaction is carried out chemoselectively. This approach is based on the chemical linkage of unprotected peptides which are easy to handle because of their high solubility in the solvents used for the ligation. These segments are functionalized with groups that react chemoselectively with only one group of the acceptor peptide, preserving the integrity of the unprotected side-chain functionalities. Nonetheless, on the basis of this chemical ligation, Kent and coworkers [103] later extended this term to “native chemical ligation” in order to encompass true peptide bond-forming ligation. Contrary to historical and didactical considerations the native chemical ligation (NCL) will be discussed first since in recent years there has been an explosion of interest in this protein-assembly technology.

5.5.1

Native Chemical Ligation

In 1953, Theodor Wieland [104] described a chemoselective reaction of a peptide thioester in aqueous buffer systems with a cysteinyl peptide derivative, resulting in the formation of a ‘natural’ peptide bond between the two reactants. In this case, both segments bear unique and complementary functional groups that are mutually reactive with each other, but are nonreactive with all other functionalities present in the segments to be coupled.

The N-terminal cysteine in the amino component functions as an “interior” thioester capture, thus forming a new thioester through transesterification. The covalent S-acyl intermediate undergoes spontaneously S- to N-acyl migration forming a peptide bond via a five-membered ring intermediate.

This basic principle of segment coupling in unprotected form by chemical ligation does not require an external template as the thiol moiety acts as an interior capture. In 1994, Kent and colleagues [103] used this principle (Figure 5.15) in order to ligate unprotected peptide segments in the presence of an N-terminal Cys residue in the amino component giving a “native” amide (i.e., peptide) bond at the site of ligation. In contrast to original types of ligation chemistry resulting in nonpeptide bonds (cf. Sections 5.5.3 and 5.5.4), these authors named this technique native chemical ligation (NCL). The synthesis of human interleukin 8 starting with IL8-(1–33) and IL8-(34–72) was the first example of NCL published by Kent’s group [103]. This thioester-mediated covalent joining of unprotected peptide segments at a Cys residue is without doubt the most practical and most widely used chemoselective ligation method [105–108].

Despite these advantages, certain limitations of NCL persist. First, the access to peptide thioesters has been a long-standing problem. Second, cysteine is a relatively rare amino acid building block in proteins (1.4% content) with the consequence that, in a certain target protein, an artificial cysteine must be inserted to provide a suitable coupling side. Furthermore, NCL is not fully convergent and multiple peptide segments are joined in a linear fashion from C-terminus to N-terminus. The reason is that the C-terminal thioester is not generally compatible with the N-terminal Cys when present in the same molecule and under typical reaction conditions. Therefore, for assembling very long peptide sequences multiple ligation and deprotection cycles are required. The following sub-sections will address these limitations.

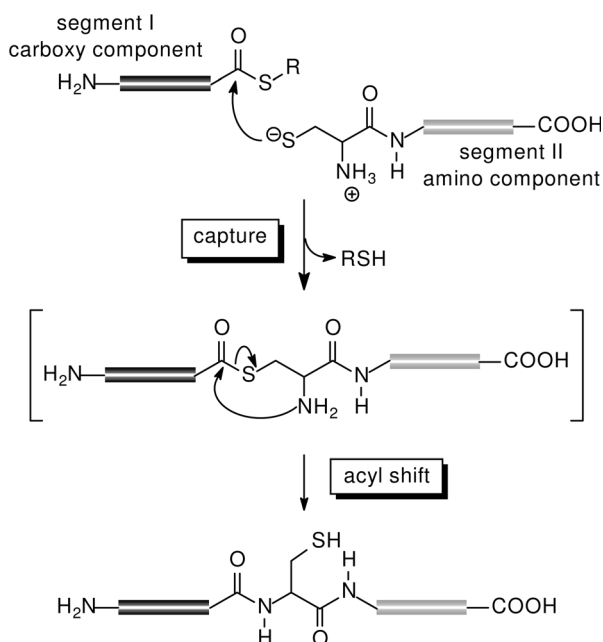


Figure 5.15 Principle of native chemical ligation.

5.5.1.1 Facile Peptide Thioester Synthesis

Peptide thioesters were traditionally synthesized on a thioester resin by using Boc-based SPPS [109]. The strongly acidic N^α -deprotection and resin-cleavage conditions were generally not compatible with acid-sensitive side-chain modifications such as glycosylation and phosphorylation in the segments to be ligated. On the other hand, synthesis of C-terminal thioester peptide segments using Fmoc chemistry has been hampered by the base lability of the thioester structure. Therefore, alternative synthetic schemes that are compatible with Fmoc-based SPPS are required.

Different examples of peptide thioester preparation have been described. First, attempts to attenuate the basicity of cleavage conditions should be discussed [110]. A combination of 1-methylpyrrolidine-hexamethyleneimine-HOBt for Fmoc detachment to prevent cleavage of the thioester from the resin has been proposed [111]. However, racemization of the peptide thioester under these conditions could not be excluded [112]. The second approach is based on Kenner's safety-catch type linker technology [113] (cf. Section 4.5.2). The conceptually elegant method allows the cleavage of the corresponding thioester by thiolate [114] and also found application for glycopeptide synthesis [115]. Some difficulties and limitations of this method have been reported together with an alternative aryl hydrazine safety-catch linker [116]. The creation of the thioester group after cleaving the peptide from the resin is an additional interesting approach [117, 118], allowing conventional SPPS technology. Further possibilities for peptide thioester preparation have been published, e.g., using a trithioorthoester [119], by $O \rightarrow N$ and $N \rightarrow S$ acyl shift strategies, respectively, [120, 121], via solution-phase tosylamide preparation [122], or using a self-purification effect [123]. A new approach to *in situ* generation of thioesters for application in native ligation was described starting with 2,2-dithiodiethanol as a linker for SPPS on 2-chlorotrityl resin where the remaining hydroxy group was loaded with the first amino acid [124].

Interestingly, a peptide building block, not having a thioester moiety but a cysteinyl prolyl ester (CPE) autoactivation unit at the C-terminus has been used in native chemical ligation [125]. Peptides containing the CPE unit can be synthesized by standard Fmoc SPPS and applied as building blocks for ligation with a cysteinyl peptide according to the procedure of NCL.

5.5.1.2 Extended Native Chemical Ligation

Extended native chemical ligation (ENCL) strategies have been developed because of the requirement for an N-terminal Cys. Cys residues are seldom conveniently distributed throughout a segment sequence to be ligated. For this reason, it would be desirable to have the option to apply thioester-supported ligation at residues other than Cys.

Auxiliary-mediated ligation is based on a thiol appendage that orchestrates the NCL-type reaction and is cleaved after the ligation reaction. The thiol appendage permits ligations at junctions other than Xaa-Cys. The first classes were those of the 1-aryl-2-mercaptoethyl auxiliary type 8 [126] and the Tmb auxiliary type 9 [127].



They contain an electron-rich N-benzyl substitution at the thiol function which allows the release of the peptide bond by acidolysis. The synthesis of cytochrome b526 (106 aa) was carried out using the 1-(2,4-dimethoxyphenyl)-2-mercaptoethyl auxiliary group **8** forming a His-Gly peptide bond between the two starting segments [128]. Attempts to use the 4,5,6-trimethoxy-2-mercaptobenzyl (Tmb) thiol auxiliary **9** to ligate complex glycopeptides indicated that the Gly-Gln ligation product was contaminated with a thioester by-product that had been formed via a N $\rightarrow$ S acyl shift [129]. However, methylation of the SH function of the auxiliary prior to auxiliary cleavage prevented this undesired acyl migration. In addition, nonparticipating Cys residues must be protected. Further approaches to extend the principle of thioester ligation to non-cysteinyll residues are, for example, Met ligation [130] and His ligation [131].

Wong and colleagues [132] described a new glycopeptide ligation approach utilizing a carbohydrate template to facilitate the ligation. This elegant method, called *sugar-assisted ligation* (SAL), utilizes a glycopeptide in which the carbohydrate is derivatized with a thiol auxiliary at the 2-position. In the presence of a peptide thioester and suitable conditions thioester exchange is followed by an S $\rightarrow$ N acyl transfer resulting in a ligation product with a native peptide backbone (Figure 5.16). However, the thiol

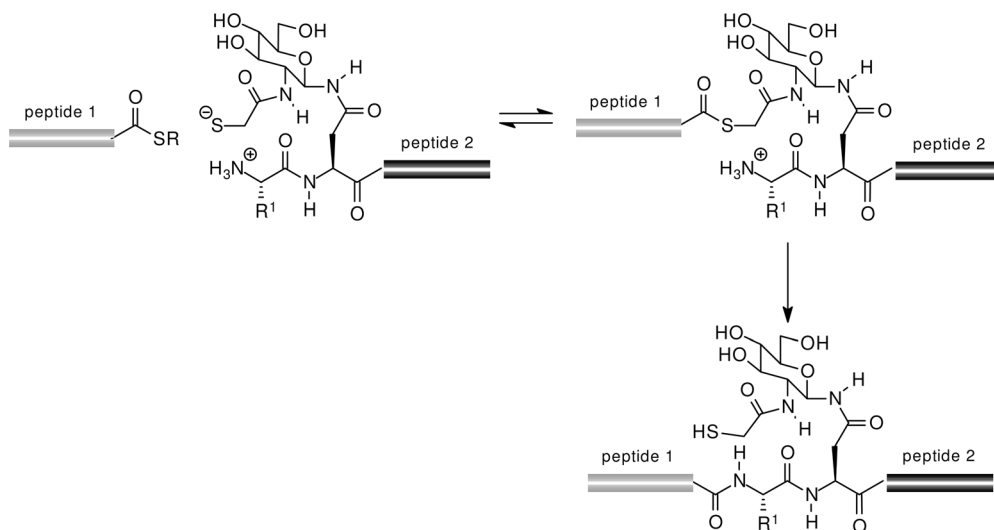


Figure 5.16 Principle of sugar-assisted ligation (SAL).

handle needs to be desulfurized to generate the carbohydrate N-acetyl group. To circumvent this problem Wong's group developed a second-generation SAL using an auxiliary which can be removed in the presence of other Cys residues after ligation [133].

Recently, a new peptide ligation strategy based on a side-chain auxiliary with similar characteristics to the sugar moiety used in SAL was described by Brik and coworkers [134]. As a removable auxiliary a thiol-modified cyclohexane is attached via an ester bond with the carboxylate side-chain of an Asp or Glu residue. This side-chain auxiliary mimics the sugar function in SAL and facilitates $S \rightarrow N$ acyl transfer for peptide ligation. Finally, the auxiliary can be removed, without product isolation, via a saponification step. Similar to SAL, the peptide is extended with an extra amino acid next to the residue bearing the auxiliary. In summary, this approach results in higher flexibility at the ligation junction, therefore expanding the scope of NLC beyond the Cys residue.

Desulfurization strategies are further useful means to open up the access to tailor-made proteins and glycoproteins. Cysteine with its optimal properties for coupling in NCL can be subsequently converted into alanine [135]. The latter amino acid occurs frequently in proteins and Xaa-Ala is, with the exception of Xaa = Pro, a suitable coupling site. In this case nonparticipating Cys residues must be protected from desulfurization, preferably with the acetamidomethyl (Acm) group [136]. This strategy was verified in the synthesis of the trypsin inhibitor EETI-II. After ligation of the two segments the sulfur atom was removed by reaction with an excess of Raney nickel, followed by the removal of the Acm groups by reaction with I_2 . Furthermore, instead of Cys, β -mercaptophenylalanine has also been used [137, 138]. After desulfurization with nickel boride Phe resulted in the ligation site. Phe occurs in proteins with a frequency of 4.1%. However, metal-mediated desulfurizations have some limitations. The metal-free desulfurization approach of Danishefsky and coworkers [139] has extended the scope of NCL for the implementation of Xaa-Ala as a potential ligation site in protein synthesis. These authors used the water soluble VA-044 (2,2'-azo-bis[2-(2-imidazolin-2-yl)propane] dihydrochloride) as radical initiator, together with a large excess of tBuSH and EtSH as hydride donors, to accelerate the desulfurization reaction. Seitz and coworkers developed a chemical ligation strategy at valine by using penicillamine (Pen, β,β -dimethylcysteine) residues in peptide synthesis and as the thiol nucleophile in the ligation. They employed Danishefsky's metal-free desulfurization to convert the Pen residues in the ligated peptide into Val residues [140].

5.5.1.3 Kinetically Controlled Ligation

The concept of kinetically controlled ligation (KCL), developed by Kent and coworkers [141], enables the reaction of a peptide thioarylester and a Cys-peptide thioalkylester to yield a single product. Therefore, KCL enables a highly practical, fully convergent strategy for the synthesis of a target protein. The majority of proteins synthesized so far by NCL have been constructed from merely two peptide segments, thus limiting the synthetic access to target proteins of about 100 or fewer amino acids. Before the introduction of KCL a larger number of peptide segments were typically

ligated sequentially in the C-to-N terminal direction. Such a strategy by sequential reactions is inefficient, even in the case of so-called *one-pot ligations* [142].

The KCL approach is based on the inherent differences in reactivity between alkyl thioesters and aryl thioesters. Aryl thioesters are more reactive but are often formed *in situ* by the addition of, e.g., excess thiophenol as a catalyst to an alkyl thioester. The difference in reactivity is mainly attributed to the pK_a of the thiol component of the thioester [pK_a thiophenol = 6.6, pK_a 2-mercaptoethanesulfonic acid (MESNA) = 9.2], explaining the better leaving group capacity of aryl thiols. Interestingly, the 4-mercaptophenyl acetic acid was observed to be the most efficient aryl thiol component. Beside the total synthesis of crambin employing five sequential chemical ligations [141], the modular total chemical synthesis of a human immunodeficiency virus type 1 protease [143] and the convergent chemical synthesis of human lysozyme [144] have been described by Kent's group. The 130aa lysozyme was assembled from four segments of comparable length, as shown in Figure 5.17.

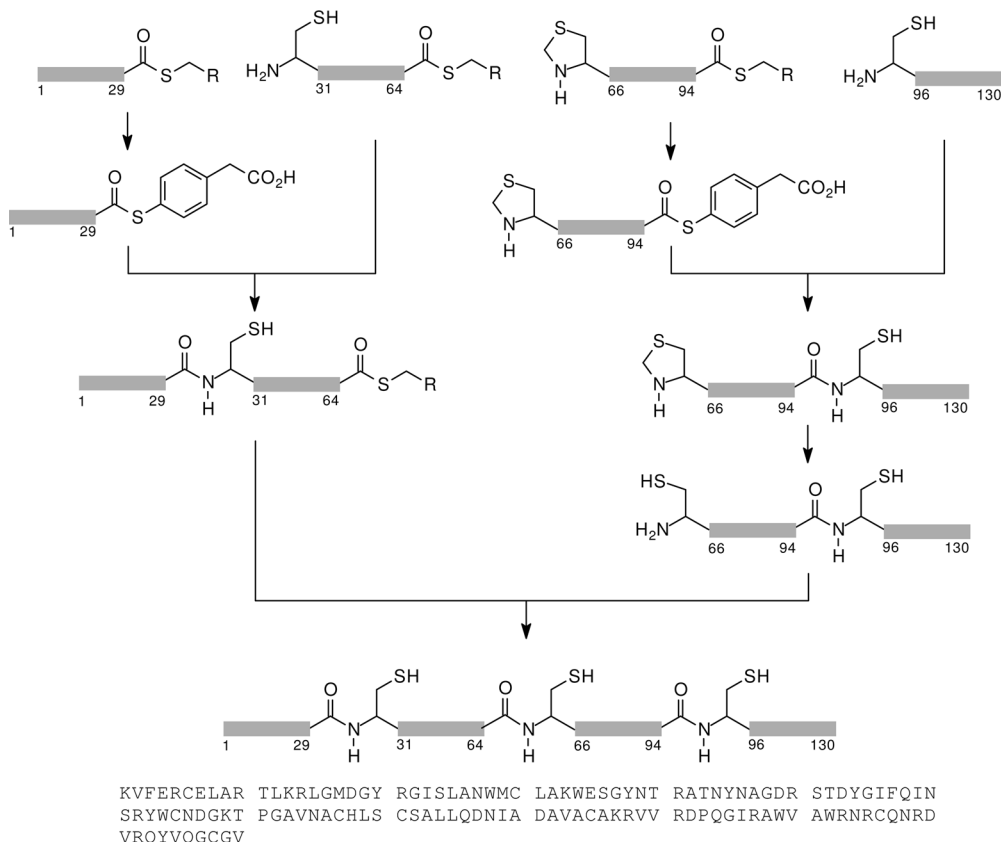


Figure 5.17 Convergent synthesis of human lysozyme.

Key to the assembly strategy was the KCL methodology. The peptide α -thioarylester was generated by transthioesterification using an excess of 4-mercaptophenyl acetic acid [107].

The full-length enzyme was obtained by ligating the two halves (1–64) + (65–130), after which folding and disulfide formation are used to yield the native enzyme. The structure was confirmed by mass spectrometry, 2D TOCSY NMR, and high-resolution (1.04 Å) X-ray diffraction.

Finally, the convergent chemical synthesis of a 203 aa ‘covalent dimer’ HIV-1 protease enzyme molecule ($M_r = 21\,870$ Da) with full enzymatic activity and correct three-dimensional structure is the largest linear polypeptide chain synthesized to date by this method [145].

5.5.1.4 Solid Phase Chemical Ligation

The solid-phase chemical ligation (SPCL) [146, 147] of unprotected peptide segments in aqueous solution can be performed in both $N \rightarrow C$ and $C \rightarrow N$ directions, and has special advantages in relay ligation procedures. The C-terminal segment of the target protein is attached via a cleavable linker to a water-compatible solid support. This strategy requires an N-terminal-protecting group to prevent cyclization/polymerization reactions with the internal (N-terminal Cys and thioester-containing) segment during the condensation reaction. A wide range of N-terminal protecting-groups which are cleavable by acidic, basic, photolytic, reductive, or oxidative conditions can be used for this purpose. The omission of otherwise necessary chromatographic separations and lyophilization steps results in a saving of time, and also considerably decreases the yield. By using this approach, it was possible to assemble some unprotected segments of 35–50 residues to the target polymer-bound protein by consecutive ligation on a water-compatible, cellulose-based polymer support. A complementary approach utilizing a highly stable, safety catch acid-labile (SCAL) linker [148] (cf. Section 4.5.2) producing an amide C-terminus has been described for a three-segment synthesis of vMIP I, a 71-amino acid chemokine [149]. Despite the advantages of this procedure, the relatively low chemical stability of cellulose-based supports toward HF and prolonged TFA treatment, for example, is a drawback that may be circumvented using more robust resins, perhaps of the superpermeable organic combinatorial chemistry resin (SPOCC) type [150].

5.5.1.5 Alternative Approaches to Native Chemical Ligation

One year after Kent’s first NCL publication [103], Tam and coworkers described alternative ligation approaches [151] which were termed ‘chemoselective capture activation ligation’, ‘intramolecular ligation’, ‘biomimetic ligation’, or ‘orthogonal ligation’ (for review see [152]). The *reverse thioester ligation approach* [151] is characterized by formation of a Cys bond via nucleophilic attack of the C-terminal thiocarboxy group of an unprotected peptide segment on the bromomethyl group of an N-terminal β -bromoalanine residue of the amino component (Figure 5.18). In the initial step a covalent thioester is formed that undergoes rearrangement forming the Cys residue at the ligation site.

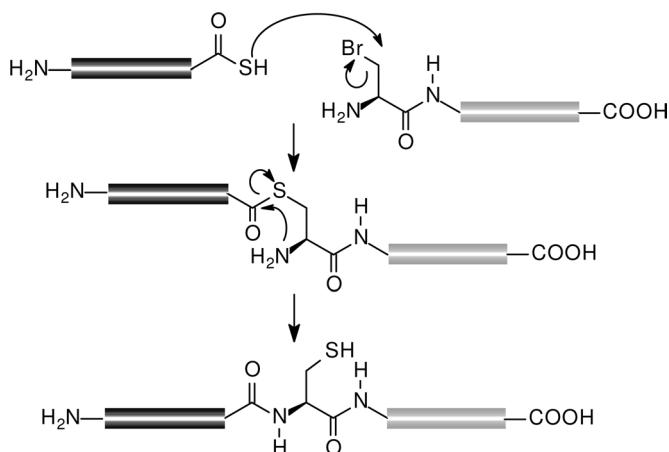


Figure 5.18 Schematic representation of the reverse thioester ligation approach.

The *imine capture ligation approach* [151, 152] is based on the reaction of an acyl-aldehyde with a suitable N-terminal amino acid (Cys, Ser, Thr, His, Trp or Asn) of an appropriate amine segment leading to an imine intermediate that undergoes rapid ring tautomerization followed by $O \rightarrow N$ migration. Thiol and hydroxy amino acids as N-terminal residues of the amino component form pseudoproline, e.g., 2-hydroxymethyl thiaproline (in the case of Cys) and appropriate oxaproline derivatives (in the case of Ser and Thr), respectively.

DNA template-controlled ligation has been termed an approach to peptide bond formation between peptide nucleic acid (PNA) precursors that associate in a sequence-specific manner with a DNA template [153]. This condensation is performed during the course of a thioester ligation reaction between one PNA molecule with a C-terminal glycyl thioester and a second PNA molecule bearing an N-terminal Cys residue.

5.5.2

Expressed Protein Ligation (Intein-mediated Protein Ligation)

In 1998 chemoselective ligation was extended to a semisynthesis version of NCL. The gene for a desired protein is cloned into an intein vector and expressed. The occurring splicing process at the recombinant-protein-intein junction is intercepted by a thiol, generating the α -thioester recombinant protein. The latter is then ligated to a peptide segment or protein that bears an N-terminal Cys residue. This procedure was described independently by Muir and coworkers [154, 155] who called it *expressed protein ligation* (EPL), and Xu and coworkers [156, 157] who called it *intein-mediated protein ligation* (IPL).

The EPL/IPL approach (Figure 5.19) uses a genetically engineered intein and a chitin binding domain (CBD) as fusion partners in order to express a protein or protein segment of interest.

clone gene into intein vector

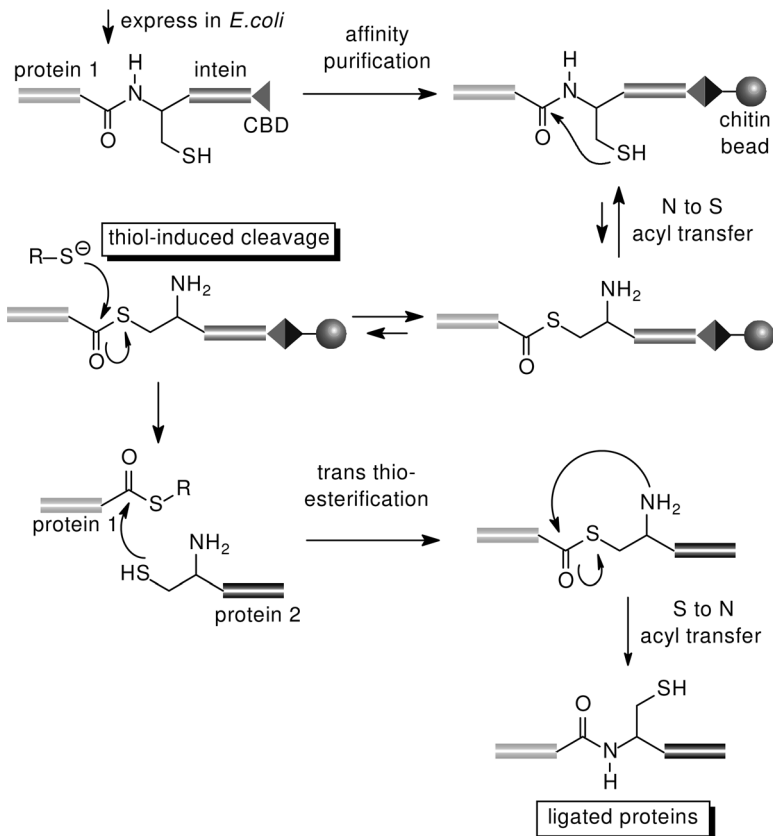


Figure 5.19 Principle of the expressed protein ligation/intein-mediated protein ligation.

CBD allows separation of the target protein of interest by binding to a chitin resin. Incubation of the chitin-bound protein with a suitable thiol reagent results in cleavage between the target protein and the intein, yielding the appropriate α -thioester. The latter can be ligated with chemically synthesized peptide segments, which also allows unnatural amino acids to be site-specifically introduced into proteins [158]. In addition, according to a modified protein splicing reaction developed by Xu and coworkers [157], specially generated recombinant proteins bearing an N-terminal cysteine residue can be used.

EPL/IPL represents a novel semisynthetic approach to extend greatly the native chemical ligation strategy. This approach facilitates site-specific protein labeling with a broad range of physical probes, e.g., fluorophores, post-translational modifications, stable isotopes, spin labels, and unnatural amino acids, and has been applied to a wide range of protein-engineering problems [159–161]. The power of protein ligation to study protein function is evident and the technology is constantly improving, aiming towards minimally invasive methods allowing analysis of target

proteins with the ensemble of all other proteins *in vivo* [162]. Mootz and coworkers [163] demonstrated that protein *trans*-splicing can be exploited for the preparation of N-terminally modified semisynthetic proteins. This approach requires neither an α -thioester at the synthetic peptide nor high reactant concentration. The so-called conditional protein splicing developed by the groups of Mootz and Muir is based on the fusion of inactive target protein domains to a split intein [164, 165].

5.5.3

Prior Capture-mediated Ligation

At the very beginning of chemical peptide synthesis Theodor Wieland [104] and Max Brenner [166] independently had the idea to convert the bimolecular reaction between two peptides to be ligated into an intramolecular reaction by bringing together the respective C- and N-termini on a template in order to facilitate intramolecular acyl transfer reaction. A similar mechanism is realized practically in the reverse ligase action of serine and cysteine proteases in which O- or S-acyl intermediates undergo an O- or S- to N-acyl transfer reaction forming a peptide bond. Furthermore, in protein splicing [167] a series of acyl transfer reactions leads to a final covalent O- or S-acyl intermediate, resulting in a spontaneous uncatalyzed O- or S- to N-acyl transfer forming the corresponding peptide bond. All these chemical and biological advancements might have stimulated several groups to vary this fundamental basic principle in the development of both template-mediated amide bond formations and segment condensation by prior capture-mediated ligation. Native chemical ligation (NCL), as the most important part of prior capture-mediated ligation, has already been discussed in Section 5.5.1

5.5.3.1 Template-mediated Ligation

The principle of a template-mediated amide bond formation according to Brenner et al. [166] using a salicylamide template (Figure 5.20) is based on a base-catalyzed intramolecular rearrangement [168]. Demonstrated in a general sense, the carboxy function of salicylic acid is acylated with an amino acid alkyl ester, whereas the hydroxy group is esterified with a protected amino acid.

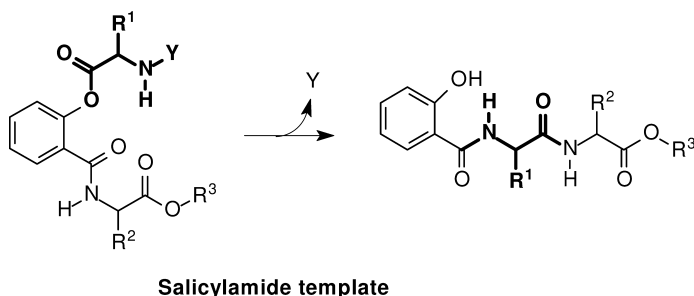


Figure 5.20 Principle of template-mediated amide bond formation.

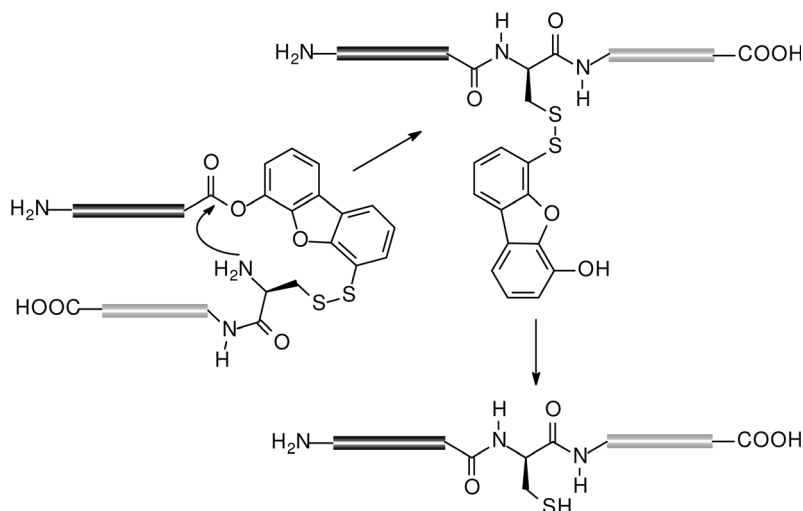


Figure 5.21 Thiol capture ligation according to Kemp.

After removal of the N-protecting group (Y), base-catalyzed intramolecular rearrangement gives the N-salicylpeptide alkyl ester that bears the newly incorporated amino acid in the N-terminal position. The next amino acid can then be esterified onto the free phenolic hydroxy group of the template, thereby continuing chain elongation. However, this very interesting concept suffers from certain limitations, namely problems with selective removal of the N-salicyl group on completion of the synthesis, and bringing together longer segments on the template. Nevertheless, this inherent template-assisted synthesis concept based on entropic activation facilitates intramolecular acylation reactions.

The *thiol capture ligation*, introduced by Kemp [169], is an elegant current variation of template-promoted amide bond formation between two segments using 4-hydroxy-6-mercaptodibenzofuran as a template (Figure 5.21).

The template is attached by disulfide bridge formation with a suitable solid support that is used for Bpoc/tBu-SPPS of the C-terminal peptide segment to be ligated [170, 171]. After deprotection with TFA, Cys residues may be kept protected (Acm). The second peptide segment, acting as an amino component, must bear an N-terminal Cys residue and can be synthesized on Wang resin by standard Fmoc/tBu-SPPS. After appropriate deprotection the latter is coupled to the template by formation of a disulfide bridge, thus yielding the two segments ready for intramolecular acyl transfer reaction. This procedure could be used for the synthesis of a 39-peptide [172]. In contrast to conventional segment condensations, the very rapid intramolecular acyl transfer reaction does not require an excess of one segment, and the solubility of the intermediates in aqueous solvent systems is significantly improved due to mostly unprotected amino acid side-chain functions.

Levacher's amine-capture method [173] is an interesting traceless auxiliary approach to amine capture peptide ligations (Figure 5.22) using a quinolium thioester. Though it remains to be seen if this procedure can be extended to

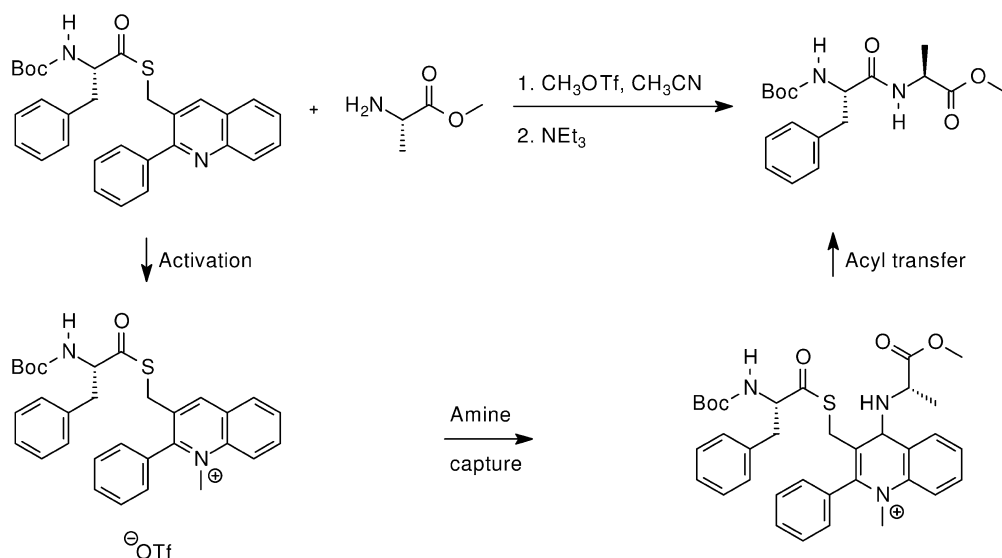


Figure 5.22 Amine capture approach according to Levacher and coworkers [173].

unprotected segments, it shows an interesting advance in template-mediated ligations.

5.5.4

Non-native Chemoselective Ligation

Non-native chemical ligation techniques were first developed for polypeptide and protein synthesis starting from unprotected segments. When a non-amide bond is used to link two peptide segments, protecting groups are less essential.

5.5.4.1 Thioester- and Thioether-forming Ligations

The first example of this type of chemical ligation, also termed backbone-engineered ligation, was published by Schnolzer and Kent in 1992 [174].

The principle of this thioester-forming ligation is shown schematically in Figure 5.23.

Segment I with a thiocarboxy group at its C-terminus (thiocarboxy glycine) reacts at acidic pH with the N-terminal bromoacetyl group of segment II, forming the condensation product with the thioester group as a non-natural bond. Under this condition, all free amino groups are protonated. This approach was first applied to the synthesis of an HIV-1 protease analogue. The protein analogue is stable in the pH range 3–6, but undergoes base-catalyzed hydrolysis above pH 7. Analogously, the mirror image enzyme D-HIV-1 protease was prepared [175]. Other examples for chemoselective ligations using thioesters have been described by Gärtner *et al.* [176] and Wallace [177]. Thioether-forming ligation [178–180] and directed disulfide formation [181] are further possibilities of this approach.

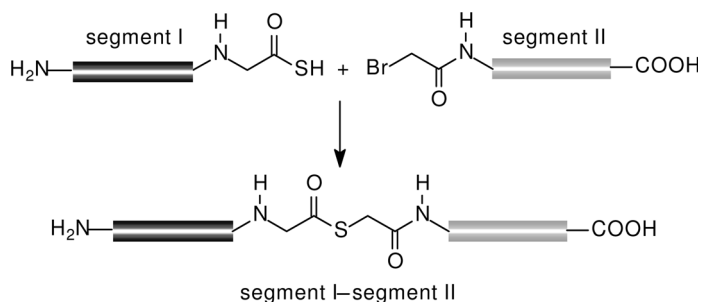
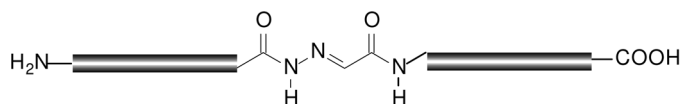


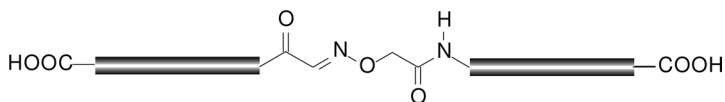
Figure 5.23 Thioester forming ligation approach.

5.5.4.2 Hydrazone- and Oxime-forming Ligations

Hydrazone-forming ligation [182–184] starts from C-terminal peptide hydrazides that are chemically ligated at pH 4.6 with amino components bearing an aldehyde function at the N-terminus obtained by mild periodate oxidation of N-terminal Ser or Thr residues.



Backbone engineered peptide hydrazone **10**



Backbone engineered peptide oxime **11**

The resulting backbone-engineered peptide hydrazone **10** can be reduced with sodium cyanoborohydride to produce the more stable peptide hydrazide. Oxime-forming ligation [185, 186] is based on the coupling of a O-peptidyl hydroxylamine as the amino component with a peptide aldehyde forming the peptide oxime **11**.

Furthermore, the application of more than one type of ligation chemistry allows the ligation of several unprotected segments in a specific fashion. The synthesis of the cMyc-Max transcription factor-related protein could be performed using both the thioester and oxime approaches [187].

5.5.5

Alternative Ligation Approaches

The need for a specific amino acid at the ligation site can be considered as a limitation that might be overcome by the development of more general ligation

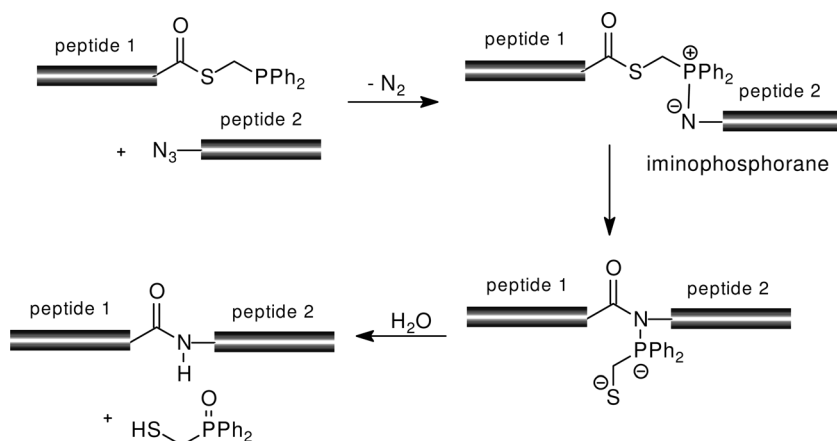


Figure 5.24 Mechanism of the traceless Staudinger ligation.

strategies. Especially interesting are those that operate under equally attractive conditions. A couple of very interesting methods have been described though most of them have not been extended to segment condensations of large peptides.

The amine capture approaches based on chemoselective mechanisms, already discussed in Section 5.5.3, are useful examples of alternative ligation strategies.

5.5.5.1 Staudinger Ligation

This bio-orthogonal method is, beyond NCL, the first general chemoselective procedure for coupling of unprotected segments. The ‘traceless Staudinger ligation’ has been developed as a new ligation strategy by Raines and coworkers [188–190]. This method is based mechanistically on the Staudinger reduction, wherein a phosphine reduces an azide through an iminophosphorane intermediate [191]. As shown in Figure 5.24 the key iminophosphorane intermediate undergoes an intramolecular amidation reaction yielding a peptide bond via the amidophosphonium salt [192].

(Diphenylphosphino)methanethiol has been used for the ligation of peptides forming a junction at a Gly residue, whereas bis(4-methoxy-diphenylphosphino)methanethiol is suitable for ligations at non-glycyl residues. Bis(4-dimethylaminoethyl)phosphinoethanethiol was the most efficient reagent since it mediates rapid ligations of equimolar segments in water [193]. Since the final ligation product has no residual atoms, this method is often termed as ‘traceless’.

5.5.5.2 Ketoacid-hydroxylamine Amide Ligations

A mechanistically unique, highly chemoselective peptide bond forming reaction has been developed by Bode *et al.* [194]. The ligation of unprotected segments is performed by an unusual decarboxylative condensation of *N*-alkylhydroxylamines with α -ketoacids. Peptide-derived α -ketoacids and appropriate hydroxylamine

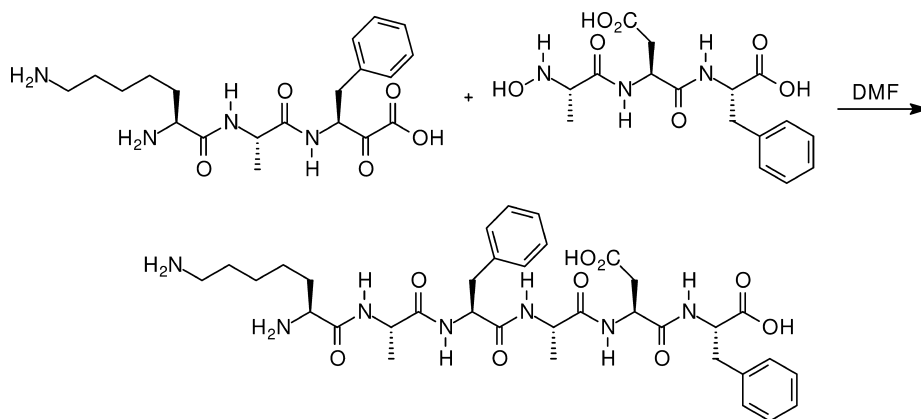


Figure 5.25 Principle of α -ketoacid-hydroxylamine amide ligation according to Bode [194].

derivatives react, without the need for additional reagents and with concurrent loss of carbon dioxide and water, to form a peptide bond between the two segments (Figure 5.25). The key advantage of this ligation strategy is that there is no need for a particular amino acid to ensure successful ligation. This methodology offers an exciting new ligation approach in peptide and protein synthesis. However, reliable methods for the synthesis of key amino acid and peptide precursors in enantiopure form must be established.

5.5.5.3 Expressed enzymatic ligation

Expressed enzymatic ligation (EEL) [195] combines the advantages of expressed protein ligation (cf. Section 5.5.2) and the substrate mimetic approach (cf. Section 4.6.2.4) of protease-catalyzed ligation. In this case the requirement for a Cys at the ligation site is lacking. Especially for peptide – but in particular for protein synthesis – the approach works best when, in addition to substrate mimetics, engineered proteases with minimized proteolytic activity are used. The latter ensure successful ligations, even in the presence of cleavage-sensitive amino acid moieties within the starting peptide fragments. In the original report describing expressed enzymatic ligation, the Glu/Asp-specific V8 protease from *Staphylococcus aureus* was non-covalently modified by thioglycolic acid, which inhibited the native proteolytic activity of the protease completely. The thioester substrate mimetic bearing the specific leaving group $-S-CH_2-COOH$ was generated by standard EPL using the intein-chitin binding tag. V8 protease-mediated ligation with a synthetic peptide containing an N-terminal Ser provided the desired final ligation product. Alternatively, the undesired cleavage activity of proteases could be suppressed by manipulating the enzyme directly via both site-directed and directed evolution approaches. Appropriate examples are reported for both cases. Finally, the enzyme can be manipulated in terms of its substrate specificity by covalent chemical modifications, mainly within the active site of the biocatalyst.

5.5.5.4 Sortase-mediated Ligation

Sortase-mediated protein ligation is an enzyme-based variant of native protein ligation [196]. Sortases are transpeptidases catalyzing a cell wall sorting reaction that attaches surface proteins to the cell wall envelope. The enzyme isoform Sortase A cleaves surface proteins at a LPXTG-motif (X is any amino acid) between Thr and Gly to form a threonyl thiol ester. The latter reacts with the N-terminal amino group of the pentaglycine branch of Lipid II [197]. Using the principle of sortase A-mediated transpeptidation, peptides with one or more N-terminal glycines can function as nucleophilic amino components in the ligation process. This enzymatic ligation procedure is highly selective due to both the low tolerance of Sortase for deviation in the LPXTG recognition motif, and the limited occurrence of the latter in proteins. However, as the ligation product contains the crucial recognition motif (just like the carboxy component), the former is also a potential substrate along with the by-product bearing the P₁-Gly residue in the N-terminal position. Under these conditions it must be assumed that it is possible for the reaction to attain equilibrium. In order to reach yields >50%, the equilibrium should be shifted to the product side by manipulations that are equivalent to those applied in equilibrium-controlled enzymatic peptide synthesis (cf. Section 4.6.2.2). The synthesis of various polypeptides, including biologically active nucleic acid-peptide conjugates [198], underlines that this approach might be an interesting technique for the ligation of unprotected peptides and proteins.

5.6

Review Questions

- Q5.1. Discuss the strategic differences in ribosomal peptide and protein synthesis and chemical peptide synthesis.
- Q5.2. What is inverse peptide synthesis and what might be the problems connected with it?
- Q5.3. Describe convergent peptide synthesis and discuss possible problems and their solution.
- Q5.4. What are the differences between Merrifield tactics and Sheppard tactics?
- Q5.5. What is the SPS/SPPS-hybrid approach? Where is it used?
- Q5.6. What kind of reactions are used for chemical ligation? Describe the concept of native chemical ligation.
- Q5.7. Why is it difficult to synthesize peptide thioesters using Fmoc chemistry?
- Q5.8. Is a cysteine residue necessary for native chemical ligation? Do you know alternatives?
- Q5.9. Why are thioesters composed of thiophenol more reactive than those of alkyl thiols like 2-mercaptoethansulfonic acid.
- Q5.10. Describe the scope of expressed protein ligation.
- Q5.11. Write the mechanism of the traceless Staudinger ligation.

References

- 1 M. Bodanszky, in: *Perspectives in Peptide Chemistry*, A. Eberle, R. Geiger, T. Wieland (Eds.), Karger, Basel, **1981**, p. 15.
- 2 F. Bordusa, D. Ullmann, H.-D. Jakubke, *Angew. Chem. Int. Ed.* **1997**, 36, 1099.
- 3 Y. V. Mitin, M. G. Ryadnow, *Protein Peptide Lett.* **1999**, 6, 87.
- 4 M. G. Ryadnov, L. V. Klimenko, Y. V. Mitin, *J. Peptide Res.* **1999**, 53, 322.
- 5 R. L. Letsinger, M. J. Kornet, *J. Am. Chem. Soc.* **1963**, 85, 3045.
- 6 B. Henkel, L. Zhang, E. Bayer, *Liebigs Ann./Recueil* **1997**, 2161.
- 7 N. Thieriet, F. Guibe, F. Albericio, *Org. Lett.* **2000**, 2, 1815.
- 8 M. Bodanszky, du Vigneaud V., *J. Am. Chem. Soc.* **1959**, 81, 5688.
- 9 M. Bodanszky, M. A. Ondetti, S. D. Levine, N. J. Williams, *J. Am. Chem. Soc.* **1967**, 89, 6753.
- 10 L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, B. Nilsson, M. Verlander, *Biopolymers* **2000**, 55, 227.
- 11 B. D. Larsen, A. Holm, *Int. J. Peptide Protein Res.* **1994**, 43, 1.
- 12 A. H. Karlstrom, A. Unden, *Tetrahedron Lett.* **1995**, 36, 3909.
- 13 A. H. Karlstrom, A. Unden, *Int. J. Peptide Protein Res.* **1996**, 48, 305.
- 14 A. H. Karlstrom, A. Unden, *Chem. Commun.* **1996**, 959.
- 15 M. Royo, J. Alsina, E. Giralt, U. Slomczynska, F. J. Albericio, *J. Chem. Soc., Perkin Trans. I* **1995**, 1095.
- 16 M. C. Munson, Garcia Echeverria C. F. Albericio, G. F. Barany, *J. Org. Chem.* **1992**, 57, 3013.
- 17 S. Sakakibara, Y. Shimonishi, *Bull. Chem. Soc. Jpn.* **1965**, 38, 1412.
- 18 H. Yajima, N. Fujii, M. Shimokura, K. Akaji, S. Kiyama, M. Nomizu, *Chem. Pharm. Bull.* **1983**, 31, 1800.
- 19 J. P. Tam, W. F. Health, R. B. Merrifield, *J. Am. Chem. Soc.* **1983**, 105, 6442.
- 20 R. Schwyzler, H. Kappeler, *Helv. Chim. Acta* **1963**, 46, 1550.
- 21 S. B. H. Kent, *Annu. Rev. Biochem.* **1988**, 57, 957.
- 22 R. H. Angeletti et al., *Methods Enzymol.* **1997**, 289, 697.
- 23 H. A. Remmer, G. B. Fields in: *Peptide and Protein Drug Analysis*, R. E. Reid (Ed.), Marcel Dekker, **2000**, p. 133.
- 24 S. Sakakibara, *Biopolymers* **1999**, 51, 279.
- 25 J. P. Tam, T. W. Wong, M. W. Riemen, F. S. Tjoeng, R. B. Merrifield, *Tetrahedron Lett.* **1979**, 4033.
- 26 B. W. Erickson, R. B. Merrifield, *J. Am. Chem. Soc.* **1973**, 95, 3750.
- 27 Y. Nakagawa, Y. Nishiuchi, J. Emura, S. Sakakibara in: *Peptide Chemistry 1980*, K. Okawa (Ed.), Protein Research Foundation, Osaka, **1981**, 41.
- 28 D. Yamashiro, C. H. Li, *J. Org. Chem.* **1973**, 38, 591.
- 29 B. W. Erickson, R. B. Merrifield, *Isr. J. Chem.* **1974**, 12, 79.
- 30 W. D. Fuller, M. Goodman, F. Naider, Y.-F. Zhu, *Biopolymers* **1996**, 40, 183.
- 31 T. Bruckdorfer, O. Marder, F. Albericio, *Curr. Pharm. Biotechnol.* **2004**, 5, 29.
- 32 V. du Vigneaud, C. Ressler, J. M. Swan, C. W. Roberts, P. G. Katsoyannis, S. Gordon, *J. Am. Chem. Soc.* **1953**, 75, 4879.
- 33 E. Wünsch, *Angew. Chem. Int. Ed.* **1971**, 10, 786.
- 34 N. Fujii, H. Yajima, *J. Chem. Soc., Perkin Trans. I*, **1981**, 831.
- 35 N. Fujii, H. Yajima, *Chem. Pharm. Bull.* **1981**, 29, 600.
- 36 T. Inui, J. Bodi, S. Kubo, H. Nishio, T. Kimura, S. Kojima, H. Matuta, T. Muramatsu, S. Sakakibara, *J. Peptide Sci.* **1996**, 2, 28.
- 37 T. Inui, M. Nakao, H. Nishio, Y. Nishiuchi, S. Kojima T. Maramatsu, T. Kimura, S. Sakakibara, in: *Proceedings of the 1st International Peptide Symposium*, Y. Shimonishi (Ed.), Kluwer, Dordrecht, **1999**, p. 552.
- 38 J. B. Hendrickson, C. Kandall, *Tetrahedron Lett.* **1970**, 343.

- 39 F. Rabanal, E. Giralt, F. Albericio, *Tetrahedron* **1995**, 51, 1449.
- 40 R. Hirschmann, R. Nutt, D. F. Veber, R. A. Vitali, S. L. Varga, T. A. Jacob, F. W. Holly, R. G. Denkwalter, *J. Am. Chem. Soc.* **1969**, 91, 507.
- 41 J. Blake, *Int. J. Peptide Protein Res.* **1981**, 17, 273.
- 42 J. Blake, D. Yamashiro, K. Ramasharma, C. H. Li, *Int. J. Peptide Protein Res.* **1986**, 28, 468.
- 43 D. Yamashiro, C. H. Li, *Int. J. Peptide Res.* **1988**, 31, 322.
- 44 J. Blake, C. H. Li, *Proc. Natl. Acad. Sci. USA* **1983**, 80, 1556.
- 45 S. Aimoto, N. Mizoguchi, H. Hojo, S. Yoshimura, *Bull. Chem. Soc. Jpn.* **1989**, 62, 524.
- 46 S. Aimoto, *Biopolymers* **1999**, 51, 247.
- 47 D. Y. Jackson, J. Vurnier, C. Quan, M. Stanley, J. Tom, J. A. Wells, *Science* **1994**, 266, 243.
- 48 P. Kuhl, U. Zacharias, H. Burckhardt, H.-D. Jakubke, *Monatsh. Chem.* **1986**, 117, 1195.
- 49 V. Cerovsky, F. Bordusa, *J. Peptide Res.* **2000**, 55, 325.
- 50 V. Cerovsky, J. Kockskämper, H. Glitsch, F. Bordusa, *Chembiochem.* **2000**, 2, 126.
- 51 N. Wehofskey, S. W. Kirbach, M. Hänslar, J.-D. Wissmann, F. Bordusa, *Organic Lett.* **2000**, 2, 2027.
- 52 R. Grünberg, I. Domgall, R. Günter, K. Rall, H.-J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, 267, 7024.
- 53 B. Riniker, A. Flörsheimer, H. Fretz, P. Sieber, B. Kamber, *Tetrahedron* **1993**, 49, 9307.
- 54 B. Riniker, A. Flörsheimer, H. Fretz, B. Kamber in: *Peptides* 1992, C. H. Schneider, A. N. Eberle (Eds.), Escom, Leiden, **1993**.
- 55 B. Kamber, B. Riniker, in: *Peptides, Chemistry and Biology, Proc. 12th American Peptide Symposium*, J. A. Smith, J. E. Rivier (Eds.), Escom, Leiden, **1992**, 525.
- 56 M. Mergler, R. Nyfelder, R. Tanner, J. Gosteli, P. Grogg, *Tetrahedron Lett.* **1988**, 29, 4005, 4009.
- 57 W. DeGrado, E. T. Kaiser, *J. Org. Chem.* **1982**, 47, 3258.
- 58 E. T. Kaiser, *Acc. Chem. Res.* **1989**, 22, 47.
- 59 D. H. Rich, S. K. Gurwara, *J. Chem. Soc., Chem. Commun.* **1973**, 610.
- 60 N. Kneib-Cordonier, F. Albericio, G. Barany, *Int. J. Peptide Protein Res.* **1990**, 35, 527.
- 61 S. S. Wang, *J. Org. Chem.* **1976**, 41, 3258.
- 62 M. Quibell, L. C. Packman, T. J. Johnson, *J. Am. Chem. Soc.* **1995**, 117, 11656.
- 63 K. Barlos, D. Gatos, *Biopolymers* **1999**, 51, 266.
- 64 P. Athanassopoulos, K. Barlos, O. Hatzi, D. Gatos, C. Tzavara in: *Solid Phase Synthesis and Combinatorial Libraries*, R. Epton (Ed.), Mayflower Scientific, Birmingham, **1996**, 243.
- 65 Y. Nishiuchi, H. Nishio, T. Inui, J. Bodi, T. Kimura, *J. Peptide Sci.* **2000**, 6, 84.
- 66 B. L. Bray, *Nat. Rev. Drug Discov.* **2003**, 2, 587.
- 67 R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, 85, 2149.
- 68 E. Atherton, R. C. Sheppard, *Solid Phase Peptide Synthesis: A Practical Approach*, Oxford University Press, Oxford, **1989**.
- 69 G. B. Fields, R. L. Noble, *Int. J. Peptide Protein Res.* **1990**, 35, 161.
- 70 R. B. Merrifield in: *Peptides, Synthesis, Structures, and Applications*, B. Gutte (Ed.), Academic Press, New York, **1995**, p. 94.
- 71 P. Lloyd-Williams, F. Albericio, E. Giralt, *Chemical Approaches to the Synthesis of Peptides and Proteins*, CRC Press, New York, **1997**.
- 72 J. D. Fontenot, J. M. Ball, M. A. Miller, C. M. David, R. C. Montelaro, *Peptide Res.* **1991**, 4, 19.
- 73 A. J. Smith, J. D. Young, S. A. Carr, D. R. Marshak, L. C. Williams, K. R. Williams, in: *Techniques in Protein Chemistry III*, R. H. Angeletti (Ed.), Academic Press, New York, **1992**, p. 219.
- 74 R. B. Merrifield, in: *Recent Progress in Hormone Research*, Volume 23, G. Pincus (Ed.), Academic Press, New York, **1967**, p. 451.

- 75 S. B. H. Kent, *Annu. Rev. Biochem.* **1988**, 57, 957.
- 76 M. Beyermann, M. Bienert, *Tetrahedron Lett.* **1992**, 33, 3745.
- 77 V. Krchnak, Z. Flegelova, J. Vagner, *Int. J. Peptide Protein Res.* **1993**, 42, 450.
- 78 M. Narita, S. Ouchi, *J. Org. Chem.* **1994**, 52, 686.
- 79 E. Pedrose, A. Grandas, M. A. Saralegui, E. Giralt, C. Granier, van Rietschoten J. *Tetrahedron* **1982**, 38, 1183.
- 80 P. Lloyd-Williams, F. Albericio, E. Giralt, *Tetrahedron* **1993**, 49, 11065.
- 81 H. Benz, *Synthesis* **1994**, 337.
- 82 F. Albericio, P. Lloyd-Williams, E. Giralt, *Methods Enzymol.* **1997**, 289, 313.
- 83 P. Athanassopoulos, K. Barlos, O. Hatzi, D. Gatos, C. Tzavara in: *Solid Phase Synthesis and Combinatorial Libraries*, R. Epton (Ed.), Mayflower Scientific, Birmingham, **1996**, 243.
- 84 K. Barlos, D. Gatos, G. Papaphotiou, W. Schäfer, *Liebigs Ann. Chem.* **1993**, 215.
- 85 K. Barlos, D. Gatos, W. Schäfer, in: *Peptides 1990*, E. Giralt, D. Andreu (Eds.), Escom, Leiden, **1991**.
- 86 D. A. Tomalia, A. M. Naylor, W. A. Goddard III, *Angew. Chem. Int. Ed.* **1990**, 29, 138.
- 87 M. Goodman, P. Keogh, H. Anderson, *Bioorg. Chem.* **1977**, 6, 239.
- 88 R. Steinauer, F. M. F. Chen, N. L. Benoiton, *J. Chromatogr.* **1985**, 325, 111.
- 89 M. Quibell, L. C. Packman, T. J. Johnson, *J. Chem. Soc., Perkin Trans. I*, **1996**, 1219.
- 90 T. A. Lyle, S. F. Brady, T. M. Ciccarone, C. D. Colton, W. J. Palevada, D. F. Veber, R. F. Nutt, *J. Org. Chem.* **1987**, 52, 3752.
- 91 J. C. Hendrix, K. J. Halverson, P. T. Lansbury, *J. Am. Chem. Soc.* **1992**, 114, 7930.
- 92 I. Dalcol, F. Rabanal, M.-D. Ludevid, F. Albericio, E. Giralt, *J. Org. Chem.* **1995**, 60, 7575.
- 93 S. Goulas, D. Gatos, K. Barlos, *J. Peptide Sci.* **2006**, 12, 116.
- 94 R. Camble, R. Garner, G. T. Young, *Nature* **1968**, 217, 247.
- 95 R. Macrae, G. T. Young, *J. Chem. Soc., Perkin Trans. I*, **1975**, 1185.
- 96 T. Wieland, W. Racky, *Chimia* **1968**, 22, 375.
- 97 L. A. Carpino, D. Sadat-Aalae, M. Beyermann, *J. Org. Chem.* **1990**, 55, 1673.
- 98 M. M. Shemyakin, Y. A. Ovchinnikov, A. A. Kiryushkin, I. V. Kozhevnikova, *Tetrahedron Lett.* **1965**, 2323.
- 99 M. Mutter, H. Hagenmaier, E. Bayer, *Angew. Chem. Int. Ed.* **1971**, 10, 811.
- 100 M. Mutter, E. Bayer, *Angew. Chem. Int. Ed.* **1974**, 13, 88.
- 101 D. B. Head, J. Z. Dong, J. A. Burton, *J. Peptide Res.* **2005**, 65, 384.
- 102 T. W. Muir, S. B. H. Kent, *Curr. Opin. Biotechnol.* **1993**, 4, 420.
- 103 P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, 266, 776.
- 104 T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, *Liebigs Ann. Chem.* **1953**, 583, 129.
- 105 P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, 69, 923.
- 106 D. Macmillan, *Angew. Chem. Int. Ed.* **2006**, 45, 7668.
- 107 E. C. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, 128, 6640.
- 108 C. Haase, O. Seitz, *Angew. Chem. Int. Ed.* **2008**, 47, 1553.
- 109 H. Hojo, S. Aimoto, *Bull. Chem. Soc. Jpn.* **1991**, 64, 111.
- 110 A. B. Clippingdale, C. J. Barrow, J. D. Wade, *J. Peptide Sci.* **2000**, 6, 225.
- 111 X. Li, T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **1998**, 39, 8669.
- 112 K. Hasegawa, Y. L. Sha, J. K. Bang, T. Kawakami, K. Akaji, S. Aimoto, *Lett. Peptide Sci.* **2002**, 8, 277.
- 113 G. W. Kenner, J. R. McDermott, R. C. Sheppard, *J. Chem. Soc., Chem. Commun.* **1971**, 636.
- 114 R. Ingenito, E. Bianchi, D. Fattori, A. Pessi, *J. Am. Chem. Soc.* **1999**, 121, 11369.
- 115 Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman, C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, 121, 11684.

- 116 J. A. Camarero, B. J. Hackel, De Yoreo J. J. A. R. Mitchel, *J. Org. Chem.* **2004**, 69, 4145.
- 117 R. von Eggelkraut-Gottanka, A. Klose, A. G. Beck-Sickinger, M. Beyermann, *Tetrahedron Lett.* **2003**, 44, 3551.
- 118 A. Sewing, D. Hilvert, *Angew. Chem. Int. Ed.* **2001**, 40, 3395.
- 119 J. Brask, F. Albericio, K.-J. Jensen, *Org. Lett.* **2003**, 5, 2951.
- 120 G. Chen, J. D. Warren, J. Chen, B. Wu, Q. Wan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, 128, 7460.
- 121 T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr, S. Aimoto, *Tetrahedron Lett.* **2005**, 46, 8805.
- 122 S. Manabe, T. Sugioka, Y. Ito, *Tetrahedron Lett.* **2007**, 48, 849.
- 123 F. Mende, O. Seitz, *Angew. Chem. Int. Ed.* **2007**, 46, 4577.
- 124 A. P. Tofteng, K. J. Jensen, T. Hoeg-Jensen, *Tetrahedron Lett.* **2007**, 48, 2105.
- 125 T. Kawakami, S. Aimoto, *Chem. Lett.* **2007**, 36, 76.
- 126 P. Botti, M. R. Carrasco, S. B. H. Kent, *Tetrahedron Lett.* **2001**, 42, 1831.
- 127 J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, 124, 4642.
- 128 D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 6554.
- 129 B. Wu, J. H. Chen, J. D. Warren, G. Chen, Z. H. Hua, S. J. Danishefsky, *Angew. Chem. Int. Ed.* **2006**, 45, 4116.
- 130 J. P. Tam, Q. Yu, *Biopolymers* **1998**, 46, 319.
- 131 L. Zhang, J. P. Tam, *Tetrahedron Lett.* **1997**, 38, 3.
- 132 A. Brik, Y. Y. Yang, S. Ficht, C.-H. Wong, *J. Am. Chem. Soc.* **2006**, 128, 5626.
- 133 S. Ficht, R. J. Payne, A. Brik, C.-H. Wong, *Angew. Chem. Int. Ed.* **2007**, 46, 5975.
- 134 M.-Y. Lutsky, N. Nepomniaschi, A. Brik, *Chem. Commun.* **2008**, 1229.
- 135 L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, 123, 526.
- 136 B. L. Pentelute, S. B. H. Kent, *Org. Lett.* **2007**, 9, 687.
- 137 D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, 129, 10064.
- 138 P. Botti, S. Tchertchian, *WO/2006/133962*, **2006**.
- 139 Q. Wan, S. J. Danishefsky, *Angew. Chem. Int. Ed.* **2007**, 46, 9248.
- 140 C. Haase, H. Rohde, O. Seitz, *Angew. Chem. Int. Ed.* **2008**, 47, 6807.
- 141 D. Bang, B. L. Pentelute, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2006**, 45, 3985.
- 142 D. Bang, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2004**, 45, 3985.
- 143 E. C. B. Johnson, E. Malito, Y. Shen, D. Rich, W.-J. Tang, S. B. H. Kent, *J. Am. Chem. Soc.* **2007**, 129, 11480.
- 144 T. Durek, V. Y. Torbeev, S. B. H. Kent, *Proc. Natl. Acad. Sci. USA* **2007**, 104, 4846.
- 145 V. Y. Torbeev, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2007**, 46, 1667.
- 146 L. E. Canne, P. Botti, R. J. Simon, Y. Chen, E. A. Dennis, S. B. H. Kent, *J. Am. Chem. Soc.* **1999**, 121, 8720.
- 147 E. C. Johnson, T. Durek, S. B. H. Kent, *Angew. Chem. Int. Ed.* **2006**, 45, 3283.
- 148 M. Patek, M. Lebl, *Tetrahedron Lett.* **1991**, 32, 3891.
- 149 A. Brik, E. Keinan, P. E. Dawson, *J. Org. Chem.* **2000**, 65, 3829.
- 150 J. Rademann, M. Grotli, M. Meldal, K. Bock, *J. Am. Chem. Soc.* **1999**, 121, 5459.
- 151 J. P. Tam, Y.-A. Lu, C. F. Liu, J. J. Shao, *Proc. Natl. Acad. Sci. USA* **1995**, 92, 12485.
- 152 J. P. Tam, J. Xu, K. D. Eom, *Biopolymers (Pept. Sci.)* **2001**, 60, 194.
- 153 C. Dose, S. Ficht, O. Seitz, *Angew. Chem. Int. Ed.* **2006**, 45, 5369.
- 154 T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 6705.
- 155 V. Severinov, T. W. Muir, *J. Biol. Chem.* **1998**, 273, 16205.
- 156 T. C. Evans, Jr., J. Benner, M.-Q. Xu, *Protein Sci.* **1998**, 7, 2256.
- 157 T. C. Evans, Jr., J. Benner, M.-Q. Xu, *J. Biol. Chem.* **1999**, 274, 3923.
- 158 B. Ayers, U. K. Blaschke, J. A. Camarero, G. J. Cotton, M. Holford, T. W. Muir, *Biopolymers* **1999**, 51, 343.
- 159 T. W. Muir, *Annu. Rev. Biochem.* **2003**, 72, 249.
- 160 D. Rauh, H. Waldmann, *Angew. Chem. Int. Ed.* **2007**, 46, 826.

- 161 C. P. R. Hackenberger, *ChemBioChem* **2007**, *8*, 1221.
- 162 V. Muralidharan, T. M. Muir, *Nat. Methods* **2006**, *3*, 429.
- 163 C. Ludwig, M. Pfeiff, U. Linne, H. D. Mootz, *Angew. Chem. Int. Ed.* **2006**, *45*, 5218.
- 164 H. D. Mootz, E. S. Blum, T. W. Muir, *Angew. Chem. Int. Ed.* **2004**, *43*, 5189.
- 165 H. D. Mootz, E. S. Blum, A. B. Tyszkiewicz, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 10561.
- 166 M. Brenner, J. P. Zimmermann, J. Wehrmüller, P. Quitt, A. Hartmann, W. Schneider, U. Beglinger, *Helv. Chim. Acta* **1957**, *40*, 1497.
- 167 L. Saleh, F. B. Perler, *Chem. Rev.* **2006**, *106*, 183.
- 168 P. L. Russel, R. M. Topping, D. E. Tutt, *J. Chem. Soc. (B)* **1971**, 657.
- 169 D. S. Kemp, *Biochemistry* **1981**, *20*, 1793.
- 170 D. S. Kemp, N. G. Galakatos, *J. Org. Chem.* **1986**, *51*, 1821.
- 171 D. S. Kemp, N. G. Galakatos, B. Bowen, K. Tam, *J. Org. Chem.* **1986**, *51*, 1829.
- 172 D. S. Kemp, R. I. Carey, *J. Org. Chem.* **1993**, *58*, 2216.
- 173 S. Leleu, M. Penhoat, A. Bouet, G. Dupas, C. Papamichael, F. Marsais, V. Levacher, *J. Am. Chem. Soc.* **2005**, *127*, 15668.
- 174 M. Schnolzer, S. B. H. Kent, *Science* **1992**, *256*, 221.
- 175 R. deLisle-Milton, S. C. F. Milton, M. Schnolzer, S. B. H. Kent in: *Techniques in Protein Chemistry IV*, R. H. Angeletti (Ed.), Academic Press, New York, **1993**, p. 251.
- 176 H. Gärtner, K. Rose, R. Cotton, D. Timms, R. Camble, R. E. Offord, *Bioconj. Chem.* **1992**, *3*, 262.
- 177 C. J. A. Wallace, *FASEB J.* **1993**, *7*, 505.
- 178 F. A. Robey, R. L. Fields, *Anal. Biochem.* **1989**, *177*, 373.
- 179 Y.-A. Lu, P. Clavijo, M. Galatino, Z. -Y. Shen, W. Liu, J. P. Tam, *Mol. Immunol.* **1991**, *6*, 623.
- 180 D. R. Englebretsen, B. G. Garnham, D. A. Bergman, P. F. Alewood, *Tetrahedron Lett.* **1995**, *36*, 8871.
- 181 M. Baca, T. W. Muir, M. Schnolzer, S. B. H. Kent, *J. Am. Chem. Soc.* **1995**, *117*, 1881.
- 182 T. P. King, S. W. Zhao, T. Lam, *Biochemistry* **1986**, *25*, 5774.
- 183 I. Fisch, G. Künzi, K. Rose, R. Offord, *Bioconj. Chem.* **1992**, *3*, 147.
- 184 H. F. Gaertner, R. E. Offord, R. Cotton, D. Timms, R. Camble, K. Rose, *J. Biol. Chem.* **1994**, *269*, 7224.
- 185 S. Pochon, F. Buchegger, A. Pelegrin, J. -P. Mach, R. E. Offord, J. E. Ryser, K. Rose, *Int. J. Cancer* **1989**, *43*, 1188.
- 186 K. Rose, *J. Am. Chem. Soc.* **1994**, *116*, 30.
- 187 L. E. Canne, A. R. Ferre-D'Amare S. K. Burley, S. B. H. Kent, *J. Am. Chem. Soc.* **1995**, *117*, 2998.
- 188 B. L. Nilsson, M. B. Soellner, R. T. Raines, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91.
- 189 M. B. Soellner, B. L. Nilsson, R. T. Raines, *J. Am. Chem. Soc.* **2006**, *128*, 8820.
- 190 B. L. Nilsson, L. L. Kiesling, R. T. Raines, *Org. Lett.* **2001**, *3*, 9.
- 191 J. Meyer, H. Staudinger, *Helv. Chim. Acta* **1919**, *2*, 635.
- 192 M. Köhn, R. Breinbauer, *Angew. Chem. Int. Ed.* **2004**, *43*, 3106.
- 193 A. Tam, M. B. Soellner, R. T. Raines, *J. Am. Chem. Soc.* **2007**, *129*, 11421.
- 194 J. W. Bode, R. M. Fox, K. D. Baucom, *Angew. Chem. Int. Ed.* **2006**, *45*, 1248.
- 195 Z. Machova, von Eggelkraut-Gottanka R. N. Wehofskey, F. Bordusa, A. G. Beck-Sickinger, *Angew. Chem. Int. Ed.* **2003**, *42*, 4916.
- 196 H. Y. Mao, S. A. Hart, A. Schink, B. A. Pollok, *J. Am. Chem. Soc.* **2004**, *126*, 2670.
- 197 B. A. Frankel, R. G. Kruger, D. E. Robinson, N. L. Kelleher, D. G. McCafferty, *Biochemistry* **2005**, *44*, 11188.
- 198 S. Fritz, Y. Wolf, O. Kraetke, J. Klose, M. Bienert, M. Beyermann, *J. Org. Chem.* **2007**, *72*, 3909.

6

Synthesis of Special Peptides and Peptide Conjugates

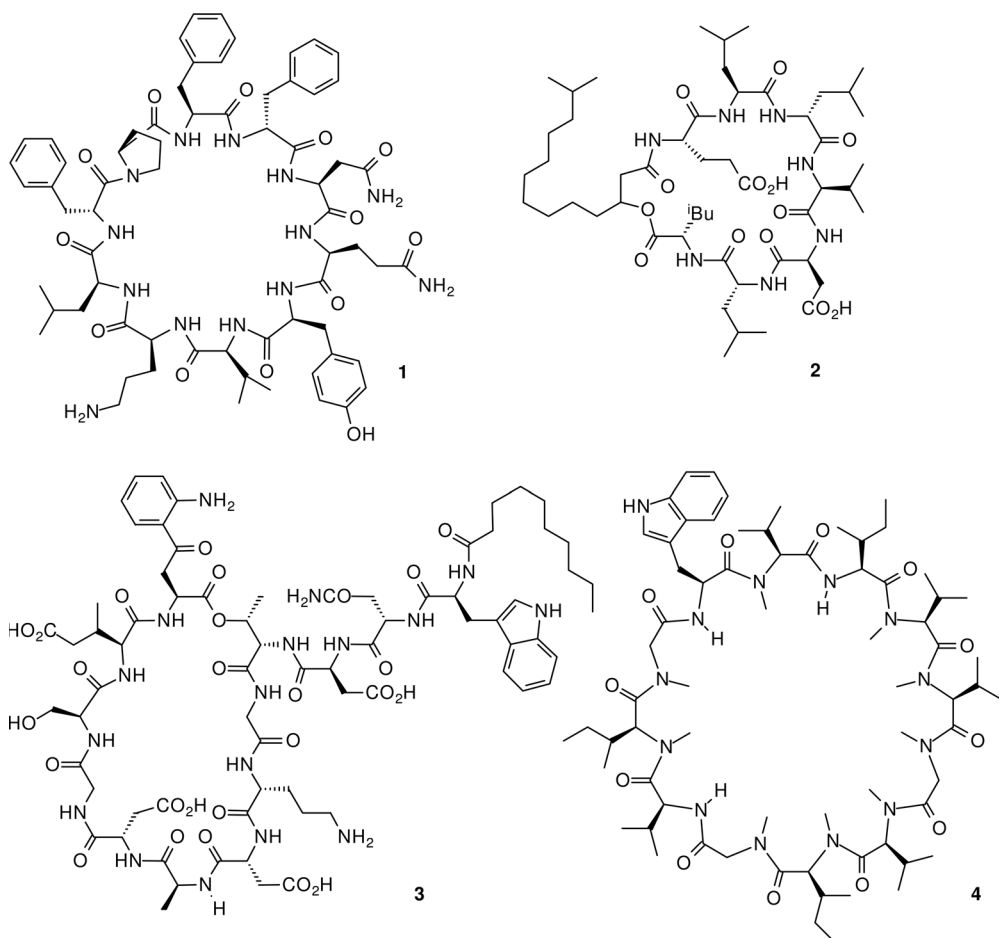
6.1

Cyclopeptides

Cyclopeptides form a large class of naturally occurring or synthetic compounds with a variety of biological activities [1]. In addition, they have gained enormous importance as models for turn-forming peptide and protein structures. They display manifold biological activities in the form of hormones, antibiotics, ion carrier systems, anti-mycotics, cancerostatics, and toxins. Several representatives of this class of compound had been isolated, their structures elucidated, and the compounds produced by total synthesis by the mid-twentieth century. However, during the past four decades there has been a rapid increase in the number of cyclopeptides with unusual structures isolated from plants, fungi, bacteria, and marine organisms [2]. Cyclopeptides and cyclodepsipeptides often contain unusual amino acids, among them D-amino acids, β -amino acids, N-alkyl amino acids, and α,β -didehydro amino acids. Consequently, ribosomal synthesis can be excluded for most cases. Biosynthesis usually proceeds via an activation of the amino acids as the thioester in multienzyme complexes (thiotemplate mechanism, cf. Section 3.2.3) [3].

The cyclosporins, a family of about 25 cyclic peptides, are produced by the fungus *Beauveria nivea* (previously designated *Tolypocladium inflatum*). Cyclosporin O [4] and the well-known relative Cyclosporin A (Section 3.3.8.1, **39**) are sequence-homologous cyclic undecapeptides with anti-fungal, anti-inflammatory, and immunosuppressive activity. Further examples of such peptides include gramicidin S (Section 3.1, **2**), tyrocidin A **1**, surfactin **2**, and many others. Most of them have interesting biological activity profiles. Daptomycin **3**, a novel lipopeptide antibiotic, was approved in 2003 by the FDA in the USA and is used for the treatment of certain infections caused by Gram-positive bacteria (see also Sections 6.5 and 9.4.2). There are indications that synthetic lipopeptides may surpass the natural parents as anti-infective agents [5].

Omphalotin A **4**, which is structurally related to the cyclosporins, belongs to a family of cyclic dodecapeptides formed by *Omphalotus olearius*. It outweighs known nematocides *in vitro* with respect to activity and selectivity.



In contrast, for some classes, such as the lantibiotics [6, 7], a ribosomal precursor synthesis with subsequent processing by decarboxylations, dehydrations, dehydrogenations, and formation of the cyclic sulfide has been proven (cf. Section 3.3.8.2, Figure 3.25).

Cyclopeptides can be classified chemically as either homodetic or heterodetic compounds (cf. Section 2.2). Cyclic depsipeptides contain α -hydroxy acids, and are members of the latter class because the peptide backbone also contains ester bonds. Cyclopeptides, where a disulfide bond is involved in the ring-closure reaction, are also classified as heterodetic cyclopeptides, and will be discussed in Section 6.2.

Results gained from several biological studies involving cyclopeptides have led to the conclusion that this class of compounds is often characterized by increased metabolic stability, improved receptor selectivity, controlled bioavailability, and improved profiles of activity. Moreover, the constrained geometry of cyclopeptides is a favorable precondition for conformational and molecular modeling studies on key secondary structure elements.

Cyclopeptides and cyclodepsipeptides are metabolized in mammals only to a very limited extent because of their resistance towards enzymatic degradation; however, they are excreted via hepatic clearance more easily than are open-chain analogues because of their increased lipophilicity.

Interest in synthetic cyclopeptides is motivated by different objectives that range from the development of synthetic methodology, through physico-chemical studies to investigations on the role of different structural factors, for example ring size, building block configuration, and influence of the side-chain functionalities. Structure–activity relationship studies on bioactive cyclopeptides remain the focus of general interest, and include conformational studies with respect to peptide–receptor interaction. Cyclic peptides serve as models of protein-recognition motifs, and are used to mimic β -sheets, β -turns, or γ -turns.

Besides potential direct application in therapy or diagnosis, cyclopeptides play an important role in the lead optimization process starting from a linear peptide epitope. The fact that cyclopeptides also display decreased flexibility and restriction of conformation in biologically important moieties allows tailored interaction with receptors, and important conclusions on the structural requirements of a receptor–ligand interaction can be derived from these studies.

Cyclopeptides are appropriate to mimic continuous or discontinuous epitopes of proteins and are, hence, suitable agents to interfere with protein–protein interactions [8]. The backbone cyclic (BC) proteinomimetic approach, as coined by Gilon *et al.* [9, 10], combines backbone cyclization of peptides with screening of libraries of BC peptides sharing the same sequence but having different conformations (cycloscan). Cyclopeptides are obtained in this approach e.g. by disulfide formation across cysteine residues or N-(ω -mercaptoalkyl) amino acids. Alternatively, lactam formation involving various N-substituted building blocks, e.g. N-(ω -carboxyalkyl) amino acids or N-(ω -aminoalkyl) amino acids (Figure 6.1) has been employed.

In this context the studies of Kessler *et al.* [11], which relate to the cyclic analogues of antamanide, thymopoietin, somatostatin [12], as well as cyclic RGD [13, 14] and LDT [15, 14] peptides, are of eminent importance. The so-called *spatial screening*

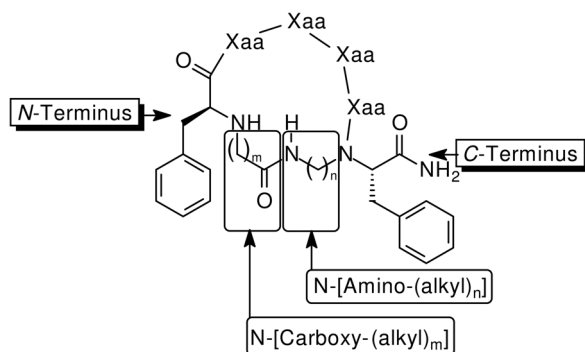


Figure 6.1 Schematic representation of the backbone cyclic (BC) proteinomimetic approach (cycloscan) with lactam ring closure.

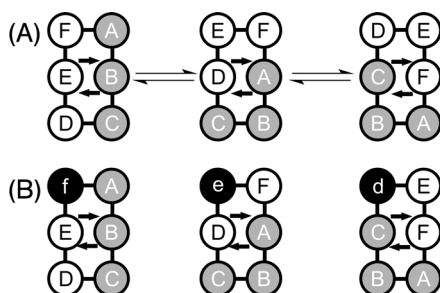


Figure 6.2 Schematic view of the spatial screening method. (A) A cyclic hexapeptide displays an equilibrium of different turns. (B) Incorporation of e.g. a D-amino acid locks the conformation and leads to a predictable presentation of a recognition sequence A-B-C.

approach [13, 14] has proven amenable for the optimization of selectively binding protein ligands.

A cyclopeptide has less conformational flexibility than its linear parent and, hence, fewer degrees of freedom. If binding to a receptor is still possible for the cyclopeptide, the binding event profits from a smaller loss of entropy compared to the linear peptide. While e.g. a cyclic hexapeptide displays several interconverting conformations (Figure 6.2A), *spatial screening* relies on the presence of a turn-inducing element (e.g. D-amino acids, symbolized by non-capital letters in Figure 6.2) locks the overall conformation of the peptide. D-Amino acids are predominantly found in position $i + 1$ of a β II' turn. Consequently, the peptide presents the functionalities of putative recognition elements in predictable three-dimensional arrays, determined by the relative position of the turn-inducing element.

The cyclic RGD peptide c-(-Arg-Gly-Asp-D-Phe- N^{Me} Val-) that resulted from *spatial screening* studies is an antiangiogenic agent that has recently been assessed in clinical trials against otherwise untreatable melanoma, solid tumors, glioma, and glioblastoma [16]. Besides D-amino acids, proline and other N-alkyl amino acids (see also Chapter 2) have been proven useful in this type of peptide conformation design. Cyclic hexapeptides with either one D-amino acid or a proline residue emerged as mimetics for the trisaccharide epitope HNK-1 [17]. Multiple N-methylation of cyclic peptides not only stabilizes the overall peptide conformation, but also improves the bioavailability, as was shown for somatostatin analogues like c-(-Pro-Phe-D- N^{Me} Trp- N^{Me} Lys-Thr- N^{Me} Phe-) [12]. Moreover, β -amino acids have also been identified as reliable inducers of secondary structure in cyclopeptides [18, 19].

To improve the ADME (absorption, distribution, metabolism, excretion) profile, nonpeptidic ligands have been developed using the information of the spatial distances and orientations of the most important pharmacophoric groups (especially the carboxyl group and the basic moiety at the other end of the molecule) [20]. Highly active and selective nonpeptide ligands of the integrins $\alpha_v\beta_3$ and $\alpha_5\beta_1$ that are orally bioavailable have been developed [21, 22].

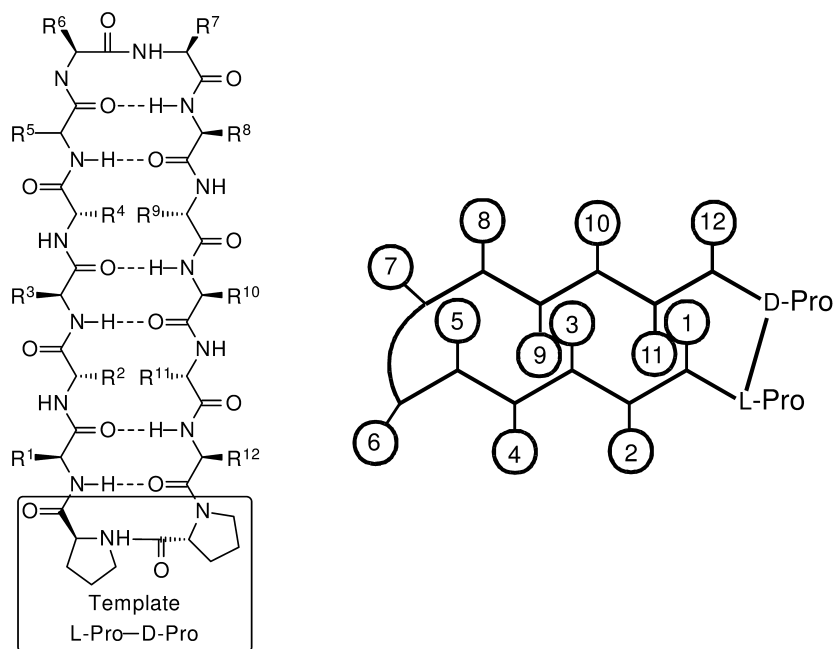


Figure 6.3 Design principle of a β -hairpin loop as a protein epitope mimetic (PEM).

Cyclic tetra-, penta-, and hexapeptides are especially appropriate for *spatial screening* applications, as they are assumed to be conformationally homogenous. Several reviews on the natural occurrence, biological importance, design [13, 23–25] and application – including synthetic aspects [26, 27] – of the cyclopeptides are available [28].

Larger cyclic peptides that mimic a β -hairpin have been successfully employed as protein epitope mimetics (PEM, Figure 6.3). The β -hairpin in such compounds is, for example, stabilized by the turn-inducing sequence D-Pro-L-Pro. The size and shape of β -hairpin PEMs appear well suited for the design of inhibitors of both protein–protein and protein–nucleic acid interactions. β -Hairpin mimetics have been discovered that possess antimicrobial activity, while others are potent inhibitors of the chemokine receptor CXCR4 [29].

Total synthesis is the best method to obtain structural variation and to optimize the biological properties. In most cases, sufficient quantities of the peptide for use in biological investigations can only be provided by synthetic means, as these naturally occurring, highly bioactive compounds occur in only minute amounts and their isolation is extremely difficult. Whilst an exhaustive presentation of all aspects of cyclopeptides is beyond the scope of this book, some principles regarding their synthesis will be discussed, using selected examples, in the following sections.

In addition to backbone head-to-tail cyclization, in principle there are three further general topologies for peptide cyclization (Figure 6.4). Side chain-to-side chain ring closure may be performed by disulfide or amide bond formation between suitable

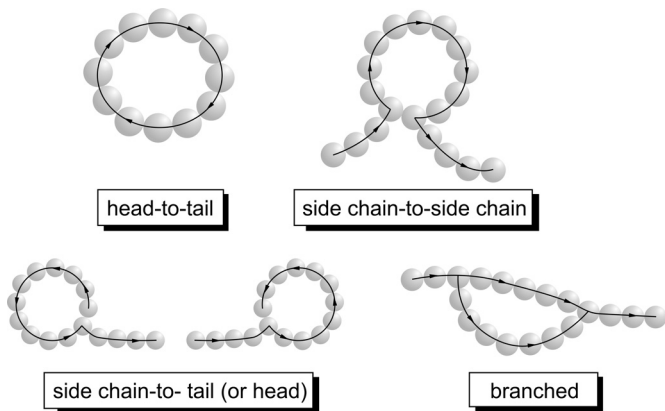
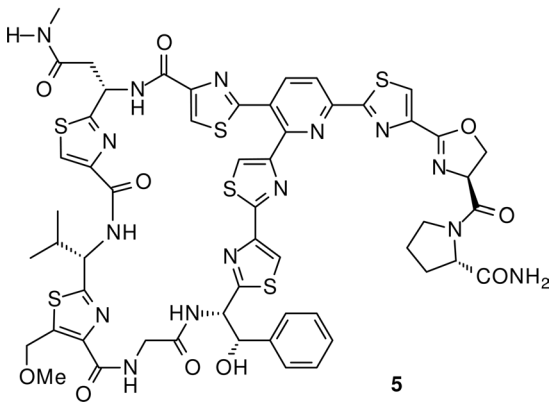


Figure 6.4 General topologies for peptide cyclization.

functional groups. Head-to-side chain or side chain-to-tail cyclizations are performed in a similar manner.

Cyclotides are plant-derived mini-proteins that contain 28–37 amino acids and comprise a head to tail macrocycle together with a knotted arrangement of three conserved disulfide bonds (cystine knot). They are characterized by well-defined secondary structures, adopt a compact three-dimensional fold and show exceptional resistance against chemical, thermal, and proteolytic degradation. The cyclotides show a wide range of biological effects, including HIV inhibitory, antimicrobial, uterotonic, cytotoxic, hemolytic, neurotensin antagonistic, trypsin inhibitory, and insecticidal activities [30].

Besides the cyclization topologies mentioned in Figure 6.4, cyclopeptides are known that comprise also thiazole or oxazole moieties [27], as for example the thiazolylpeptide GE2270 A 5 [31]. Thiazole formation may be biosynthetically explained by the reaction of a Cys residue with a carboxy group. However, the chemical synthesis of such compounds is not straightforward, and relies heavily on intelligent organometallic cross-coupling chemistry, as exemplified for 5 [31].



6.1.1

Backbone Cyclization (Head-to-Tail Cyclization)

As is known from organic chemistry, the formation of medium-size rings with 9–12 atoms is extremely difficult, with carbocyclic ring systems of medium size being energetically disfavored because of eclipsed and transannular interactions. However, the situation changes when heteroatoms are present, as for example in macrocyclic natural products, as these minimize – or even exclude – eclipsed or transannular strain.

The synthesis and conformation of backbone-cyclized cyclopeptides containing more than nine backbone atoms has undergone extensive examination. In principle, all methods suitable for the formation of a peptide bond can be used for the cyclization of linear peptide sequences involving amide bonds. Ring-closure reactions usually proceed much more slowly compared to normal peptide bond formations, and side reactions [32, 33], such as undesired intermolecular peptide bond formation leading to cyclodimerization of linear peptide precursors, may predominate. This can be suppressed by performing the cyclization reaction under high dilution conditions in 10^{-4} – 10^{-3} M solution.

Head-to-tail cyclizations of 7- to 10-peptides are usually not impeded by sequence-specific problems. For shorter sequences, the reaction rate of the ring closure depends on the presence of turn-inducing elements such as δ -amino acids, proline, glycine, or N-alkyl amino acids, which favor turns or *cis*-peptide bonds. Purely L-configured tetra- and pentapeptides that do not contain any glycine, proline or δ -amino acid residues are usually very difficult to cyclize.

The active ester method (see Section 4.3.4) was found to be a straightforward method for peptide cyclization. The pentafluorophenylester ring-closure reactions performed by Schmidt *et al.* proceed smoothly in the presence of DMAP as an acylation catalyst. The principle of the reaction is shown schematically in Figure 6.5 [34].

The linear N-terminally protected peptides are transformed into the pentafluorophenylester, for example by treatment with a carbodiimide and pentafluorophenol. These protected and activated intermediates are not stable for an unlimited period of time. Three variants are possible depending on the N-terminal N^α -amino protective group of the linear peptide pentafluorophenylester. In the case of an N-terminal Z group, the protected peptide pentafluorophenylester is added dropwise slowly to a

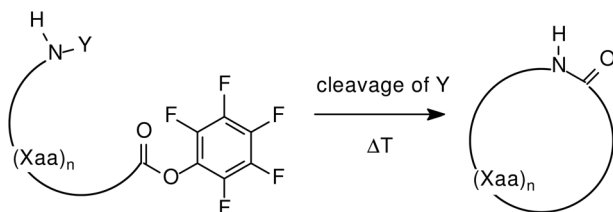
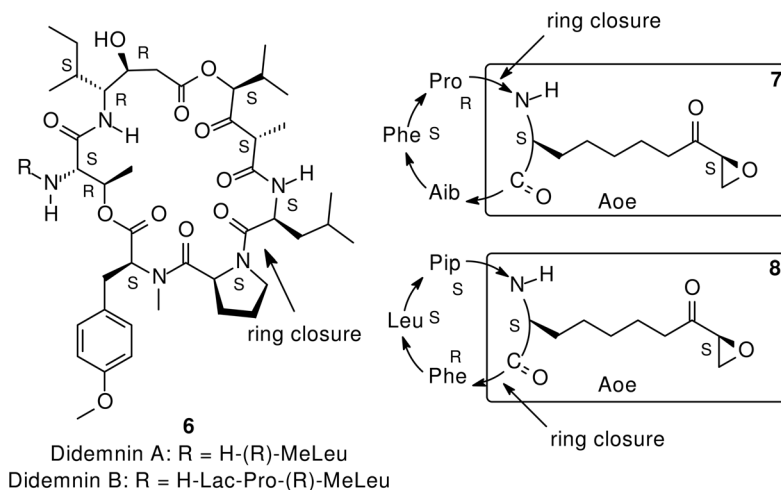


Figure 6.5 Pentafluorophenylester cyclization of Z-, Fmoc-, or Boc-protected peptides. Cleavage conditions: H_2/Pd ($Y = Z$), TFA, then base ($Y = Boc$), piperidine ($Y = Fmoc$).

hot dioxane solution (95 °C) containing Pd/C, some alcohol, and dimethylaminopyridine under a hydrogen atmosphere. The Z group is cleaved under these conditions, and ring closure proceeds on the palladium surface. Yields of 70–80% have been obtained in a series of cyclotetrapeptides containing three L-configured and one D-configured amino acid, or in a series of peptide alkaloids.

The 23-membered didemnins **6** have been obtained in 70% yield using the Boc/OPfp protocol [35]. The cyclotetrapeptides Chlamydocin [c-(Aib-Phe-D-Pro-Aoe-), **7**] and WF 3161 [c-(D-Phe-Leu-Pip-Aoe-), **8**], together with trapoxin A [c-(Phe-Phe-D-Pro-Aoe-)], trapoxin B [c-(Phe-Phe-D-Pip-Aoe-)], HC Toxin [c-(D-Pro-Ala-D-Ala-Aoe-)], Cyl-1 [c-(D-(Tyr(Me))-Ile-Pro-Aoe-)], and Cyl-2 [c-(D-(Tyr(Me))-Ile-Pip-Aoe-)] belong to the chlamydocin group of histone deacetylase (HDAC) inhibitors. HDAC is considered a target in tumor therapy, as its inhibition effects cell-cycle arrest and induces differentiation. Chlamydocin **7** and WF3161 **8**, both of which belong to the most potent cancerostatics *in vitro*, have been obtained in 95 and 70% yields, respectively [36]. The non-natural amino acid (2S,9S)-2-amino-8-oxo-9,10-epoxydecanoic acid (Aoe) was assembled stereoselectively in a seven-step synthesis.



A sometimes unacceptably high risk of epimerization at the C-terminal amino acid (racemization) must be taken into account for cyclization reactions involving carboxy activation of chiral amino acid building blocks, especially because of the prolonged reaction times [37]. Hence, precursors with C-terminal glycine or proline residues that are less prone to racemization should be chosen whenever possible. If this is not possible, a coupling reagent with inherently low racemization potential should be used. Comparative studies regarding the efficiency of peptide cyclizations using different new coupling reagents have revealed that reagents based on HOAt gave the best results with respect to yields and minimization of racemization. C-terminal D-amino acid residues favor the formation of pentapeptide rings [38].

Peptides containing an amino acid building block at the C-terminus which is prone to racemization may be cyclized using the azide method or related variants such as

the diphenylphosphoryl azide (DPPA) protocol [39], though these methods are characterized by prolonged reaction times [33]. The coupling reagents BOP and TBTU permit rapid cyclization reactions to be carried out, but suffer from significant C-terminal racemization [37]. Better results have been obtained using the azide method or DPPA compared to BOP or HBTU in ring-closure reactions of D-Pro-containing β -casomorphin tetra- and pentapeptides [33].

The problem of cyclo-oligomerization in the case of small peptides is attributed to the predominantly trans-configured peptide bonds favoring an extended conformation of the linear precursors. Dimerization and cyclo-dimerization can be reduced by performing the cyclization in solution under “pseudo-high dilution conditions”, where the linear precursor and the coupling reagent HATU (because of its limited half-life under basic conditions) are added simultaneously, using a dual syringe pump at a very slow rate, to a small volume of solvent and base [40].

However, the application of HATU, for example, sometimes suffers from undesired N-tetramethylguanylation if the cyclization reaction is too slow. This problem may be overcome by the addition of HOAt [41] or by replacing HATU by a HOAt-derived phosphonium reagent like AOP or PyAOP [42]. In summary, HOAt-derived reagents without doubt enriched the synthetic repertoire for peptide macrocyclizations.

Cyclopeptides may be synthesized by solution- or solid-phase methods (or by a combination of these) where a linear sequence is synthesized on a polymeric support and cyclization is performed in solution after cleavage from the resin. The side chain-protecting groups are cleaved in the final step. The major obstacles of classical cyclization reactions in solution (cyclo-oligomerization and cyclo-dimerization) must be avoided by high dilution, as discussed previously. A viable route towards cyclopeptides comprises the assembly and cyclization of a peptide while it is still bound to the resin. The so-called pseudo-dilution, a kinetic phenomenon favoring intramolecular reactions of resin-bound peptides over intermolecular side reactions, presents a major advantage [43–45], provided that the resin loading is not too high. The additional application of a larger excess of soluble reagent may favor the cyclization reaction. As the desired product is bound to the resin, simple washing and filtration processes facilitate the whole process and provide the potential for automation. One precondition for on-resin cyclization is the attachment of the first amino acid to the polymeric support via a side-chain functional group (Figure 6.6, Table 6.1). Care has to be taken when choosing the C-terminal amino acid of the linear precursor and the coupling reagent in order to minimize epimerization at the C-terminal residue during backbone cyclization.

This strategy requires one further orthogonal protective group for the C-terminal carboxy group, and the synthesis of cyclo-(Gly-His)₃ in 42% purified yield was a first successful example of this [54]. Such types of cyclization reaction can be performed using orthogonal protecting groups such as the Fmoc/*tert*-butyl/allyl type [55]. Alternatively, resin attachment is possible via a backbone amide linker (Figure 6.7) [56].

The success of the peptide cyclization reaction, either in solution or on a solid support, depends primarily on the conformational preferences of the linear target sequence, and also on the solvent, the bases applied, the concentration, and the temperature.

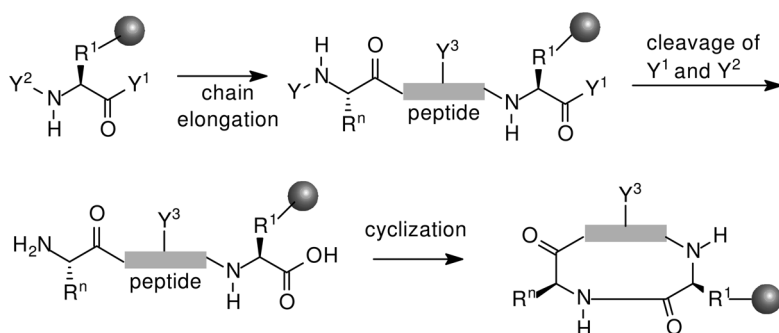


Figure 6.6 On-resin cyclization of peptides with side chain resin attachment. Y¹ = carboxy-protecting group; Y² = temporary amino-protecting group; Y³ = semipermanent side chain-protecting groups.

The first solid-phase head-to-tail cyclization by intramolecular aminolysis of a resin bound *o*-nitrophenyl peptide ester was described as early as 1965 by Patchornik *et al.* The concept of combined cyclization and cleavage from the resin does not necessarily require side chain or backbone anchoring, if special linkers for C-terminal attachment are employed (oxime resin [57, 58], thioester resin [59]). The acid-stable *p*-nitrobenzophenone oxime resin was found to be the most suitable [60]. Whilst in the former case tyrocidin A was obtained in 30% yield after purification, the synthesis of cyclo-(-Arg-Gly-Asp-Phg-), a cyclic tetrapeptide with inhibitory activity against cell adhesion, was achieved with a yield exceeding 50%. In the latter case simultaneous cyclization and cleavage, utilizing the lability of the oxime ester linkage to nucleophilic attack, was performed [61].

Table 6.1 Selected protecting-group/linker tactics for backbone cyclization.

Amino acid	N ^α -Protection	Resin	Orthogonal CO protecting group	Ref.
Asp	Fmoc	Wang	ODmb	[46, 47]
Asp	Fmoc	PAC/PAL	OAl	[48]
Asp/Glu	Fmoc	Pepsyn-K	ODmab	[49]
Asp/Glu	Boc	Merrifield (HM-PS)	OFm	[50]
Asp/Glu	Boc	MBHA	OFm	[50]
Lys/Orn/Dab	Boc	HM-PS, linked via urethane group	OFm	[50]
Ser	Boc	AM-PS, linked via succinyl group	ONbz	[50]
Tyr	Boc	Merrifield (HM-PS)	ONbz	[50]
Lys	Fmoc	Wang (modified)	OAl	[51]
Ser/Thr/Tyr	Boc	Merrifield (HM-PS, modified)	OAl	[52]
His	Fmoc	Trt	OAl	[53]

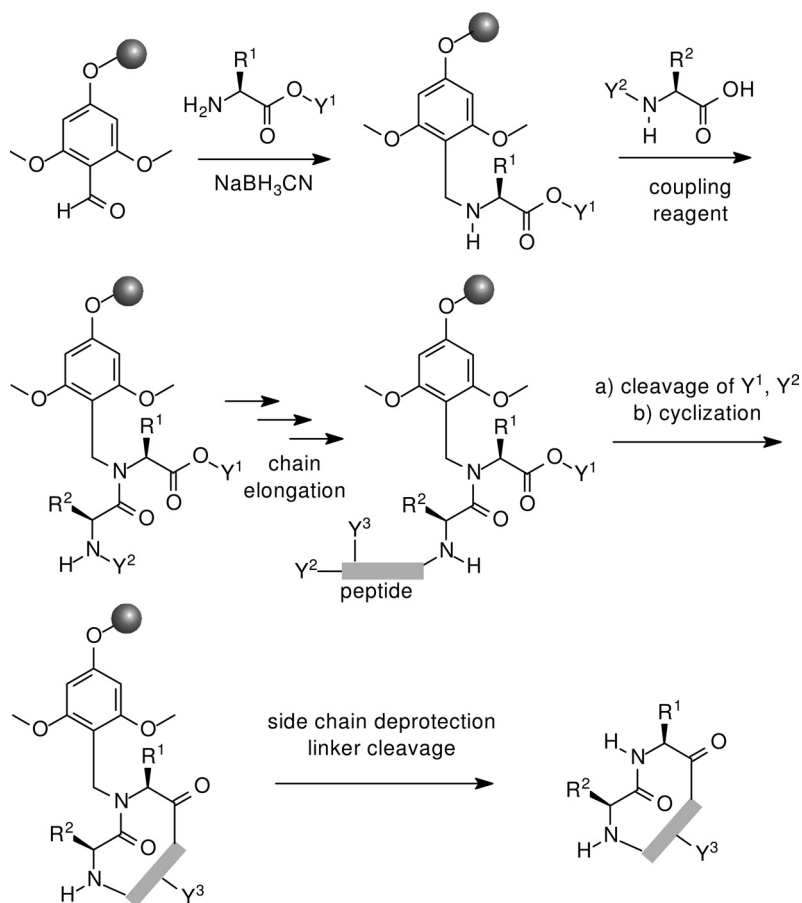
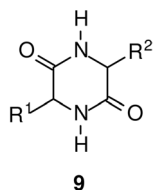
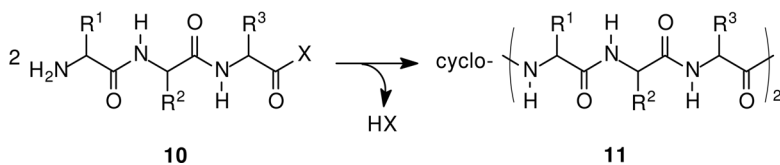


Figure 6.7 On-resin cyclization of peptides with resin attachment via backbone amide linker. Y^1 = carboxy-protecting group; Y^2 = temporary amino-protecting group; Y^3 = semipermanent side-chain protecting groups.

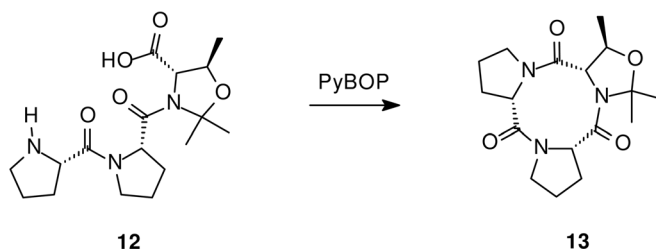
Cyclic dipeptides (diketopiperazines) **9** are very easily formed by intramolecular aminolysis of dipeptide esters. These cyclodipeptides, which comprise a six-membered ring, are usually stable towards proteolysis and are often formed as unwanted side products in solution or solid-phase synthesis of linear peptides at the dipeptide stage. Diketopiperazines may serve as scaffolds for combinatorial synthesis in drug discovery [62–64].



Cyclic tripeptides (nine-membered rings) are not usually formed when ring closure is attempted starting from carboxy activated tripeptides **10**. In most cases, the reaction results in cyclodimerization, forming the corresponding hexapeptides **11**. Only cyclic tripeptides containing N-substituted glycine, proline, or hydroxyproline residues are known as exceptions. Basic investigations by Rothe *et al.* [65] revealed that stable cyclotripeptides must contain at least two secondary amino acid building blocks.



Preorganization of the linear precursor seems to be a major precondition for successful ring closure. The pseudoproline-containing tripeptide **12** is cyclized to give **13** without oligomerization, even at 0.1 M concentration [66]. Besides increasing the propensity towards *cis*-peptide bond formation with secondary amino acids, the incorporation of β -amino acids also leads to improved yields of cyclic tripeptides [67].



Cyclic tetrapeptides (12-membered rings) that are assembled exclusively from L-amino acids are also not readily accessible. The conformation of the linear precursor influences the tendency towards cyclization [68]. If the linear precursor lacks N-alkyl amino acids, usually only traces of product are formed [69]. Polycondensation efficiently competes with ring closure. Cyclotetrapeptides containing N-alkyl amino acids such as sarcosine are formed more easily. Likewise, tetrapeptides with one β -amino acid and/or a D-amino acid are accessible [18, 70]. The synthesis of the tetrapeptides $c\text{-}(\text{-Arg-Gly-Asp-Xaa-})$, Xaa = Ala, Phe, Phg, D-Ala, D-Phe, D-Phg, by on-resin cyclization on Wang resin was investigated in a systematic study. It was shown that the target peptides can be obtained in very good yields without formation of cyclodimerization by-products [71].

When the usual coupling methods are employed, cyclic pentapeptides are usually obtained in acceptable to good yields. However, cyclodimerization, by which the corresponding cyclodecapeptides are formed, may compete with cyclization. Despite the variety of coupling reagents that have been examined, the azide method with DPPA, as well as the coupling reagents HATU and AOP give good results. However, the azide method with DPPA is sometimes accompanied by side-products derived

from Curtius rearrangement. In some cases superior results have been obtained with the cyclic, trimeric propylphosphonic acid anhydride (PPA) as the coupling reagent. Application of PPA with sterically hindered pentapeptides gives only little epimerization (2–6%), while HAPyU as the coupling reagent leads to the formation of 64–87% of diastereomers [72].

Cyclization of hexapeptides, despite having the potential to form two complementary β -turns, is not always straightforward. In reported cases, the presence of a D-amino acid supports the macrocyclization [73].

The linear peptide Ala-Phe-Leu-Pro-Ala, which cannot be converted into monocyclic product under conventional cyclization conditions, can be cyclized into c-(Ala-Phe-Leu-Pro-Ala-) by using a photolabile peptide cyclization auxiliary, a N-terminally located 6-nitro-2-hydroxybenzyl group (Hnb, Figure 6.8, see also Section 4.2.3). Cyclization is accomplished through a cyclic nitrophenyl ester that preorganizes the peptide for lactamization. The auxiliary is subsequently removed photolytically [74]. Likewise, cyclotetrapeptides can be efficiently obtained when the Hnb group is incorporated both at the N-terminus and within the sequence. The N-terminal Hnb group acts as the cyclization auxiliary and performs the ring closure/ring contraction, while the second Hnb group favors *cis* amide bonds to conformationally favor backbone cyclization. Following this route as displayed in Figure 6.8 the all-L cyclic tetrapeptide c-(Tyr-Arg-Phe-Ala-) was successfully prepared [75].

In the case of the cyclodepsipeptides the formation of a peptide bond during cyclization is preferred, because an ester bond is more difficult to form than an amide bond. Cyclization yields generally are rather low, and a yield of 50% is usually regarded as high.

Jung *et al.* reported the synthesis of omphalotin A (4), which contains several N-methyl amino acids, (Figure 6.9) on a solid support [76]. Bis(trichloromethyl) carbonate (BTC, triphosgene) was used as the coupling reagent. This method had originally been introduced by Gilon *et al.* [77], who used BTC together with collidine

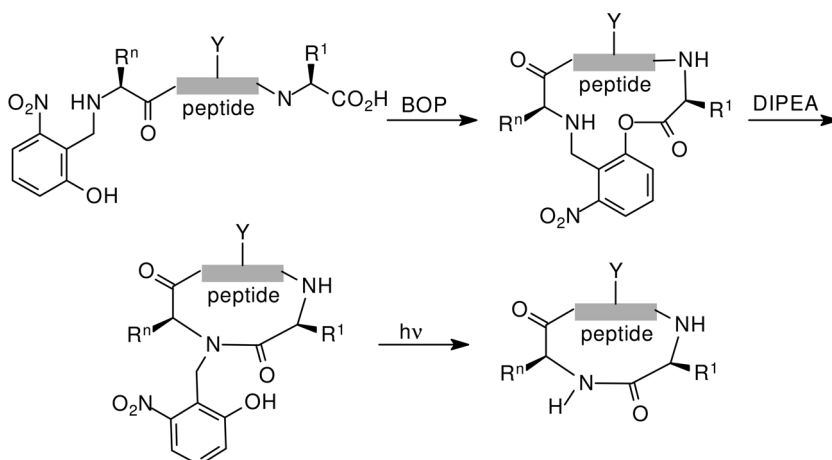


Figure 6.8 Cyclization of tetra- and pentapeptides by ring contraction.

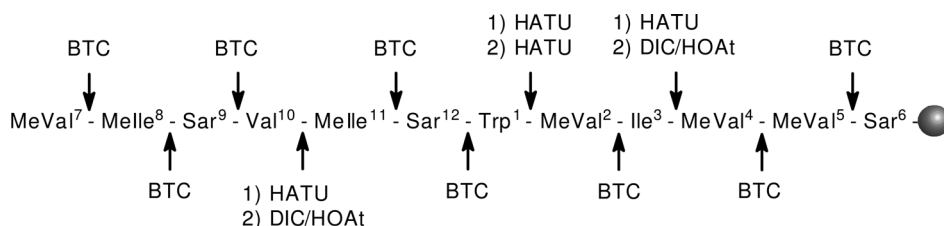


Figure 6.9 Solid phase synthesis of the linear precursor of Omphalotin A suited best for cyclization.

as the base in THF at 50 °C for *in situ* activation of N^α-Fmoc protected amino acids. However, it is still unclear whether carboxy group activation proceeds via acid chloride or mixed anhydride formation. The resin-bound peptide is reacted with these pre-activated derivatives. The method is reported to be devoid of racemization, even for a sequence containing three consecutive N-methyl amino acids.

Head-to-tail cyclization by intramolecular aminolysis on the resin experienced a further renaissance with the development of the so-called safety-catch resins (cf. Section 4.5.2). During the course of this process, the linker moiety is activated so that the bound peptide is able to undergo intramolecular aminolysis (Figure 6.10) [78].

Thiol-mediated backbone cyclization of cysteine-containing peptides (cf. Section 5.4) by S → N acyl migration has also been described [79, 80]. Peptide thioesters may be cyclized enzymatically in a biomimetic fashion using a thioesterase [81, 82] and the *in vivo* synthesis of cyclic peptides has also been described [83].

Application of *split intein-mediated circular ligation of peptides and proteins* (SICLOPPS, Figure 6.11) permits the combinatorial synthesis of homodetic cyclo-

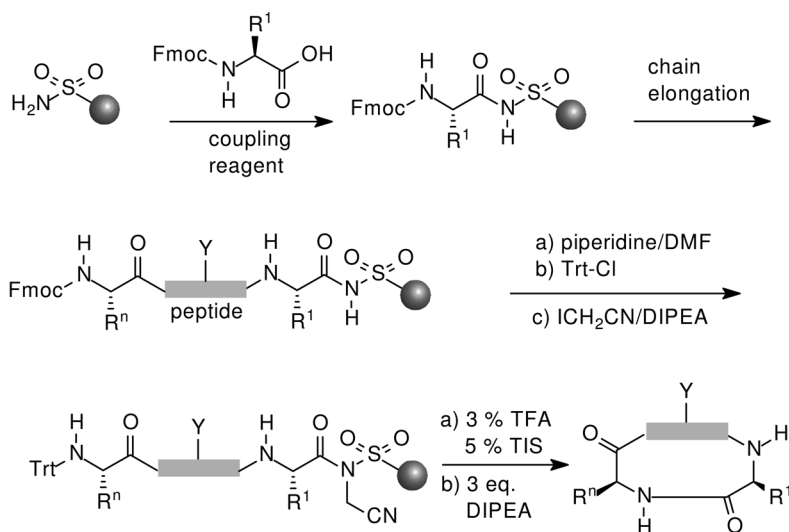


Figure 6.10 Peptide cyclization with simultaneous cleavage from safety-catch resin. Y = semipermanent side-chain protecting groups.

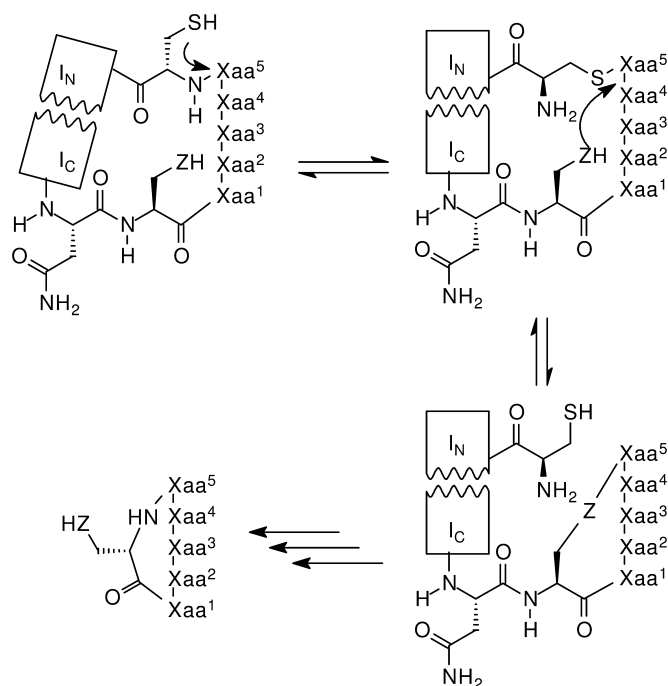


Figure 6.11 *In vivo* cyclopeptide synthesis according to the split intein-mediated circular ligation of peptides and proteins ($Z = O, S$).

peptides in a living cell. Split inteins are a subgroup of self-splicing proteins (see Section 5.5). The N-terminal (I_N) and the C-terminal intein (I_C) associate non-covalently. Homodetic cyclopeptides are obtained, when the target sequence or a combinatorial pool of target sequences is flanked N-terminally by Cys or Ser and C-terminally by Cys. The recombinant intein undergoes a sequence of intramolecular reactions, before the cyclopeptide is finally formed. Nucleophilic attack of the C-terminal Cys side chain thiol provides a thioester that undergoes acyl migration to the N-terminal Cys or Thr residue with formation of the heterodetic cyclopeptide. Subsequent S/O → N acyl migration provides the target cyclopeptide. The SICLOPPS method provides intracellularly homodetic cyclopeptides that are metabolically stable and can be used to screen for inhibitors of protein–protein interactions e.g. by a combination with two-hybrid systems [84, 85].

Moreover, chemoselective cyclopeptide formation by a traceless Staudinger ligation (cf. Section 5.5.1) involving an N-terminal α -azido acid and a C-terminal diphenylphosphinomethylthioester has been reported [86].

Macrocyclization between two additional functional groups connected to backbone amides may also be regarded as backbone cyclization, or as a side-chain to side-chain cyclization. The cyclization may occur either via disulfide or amide bond formation [9, 10, 87] or via ring-closing metathesis, where two N-allyl, N-homoallyl or longer derivatives react intramolecularly under transition metal catalysis to give peptide

cyclization [88, 89]. Cyclopeptide-derived macrocycles have been obtained by ring-closing metathesis with Grubbs catalyst using a strategy that involves simultaneous metathesis macrocyclization and resin cleavage [90]. Van Maarseveen and coworkers succeeded in the synthesis of cyclic tetrapeptide analogues by Cu^I catalyzed 1,3-dipolar azide/alkyne cycloaddition of an N-terminal α -azido acid with the C-terminal proline derivative 2-ethynylpyrrolidine, where the carboxy group of proline was replaced by a C $\equiv$ C triple bond [91].

6.1.2

Side Chain-to-Head and Tail-to-Side Chain Cyclizations

Side chain-to-head or tail-to-side chain cyclizations may proceed via macrolactamization between a side-chain amino group and the C-terminus, or between a side chain carboxy group and the N-terminus. On-resin cyclizations have been reported [92–94], and lactone formation involving side-chain hydroxy groups is also possible. Furthermore, intramolecular thioalkylation of an N-terminal bromoacetyl group with a cysteine residue of the peptide chain is feasible and results in the formation of a cyclic thioether [95, 96]. Head-to-side chain cyclizations and branched cyclopeptides have been reviewed in [1]. On-resin tail to side-chain cyclization by Cu^I catalyzed 1,3-dipolar cycloaddition of propargylglycine and an N-terminal α -azido acid has been reported [97].

6.1.3

Side Chain-to-Side Chain Cyclizations

Macrocyclic disulfides (cf. Section 6.2), thioethers [98], lactams [93, 99, 100] and lactones are obtained upon reactions between appropriate functional groups present in the side chains. Ring-closure reactions between side chain functional groups were described for the first time by Schiller *et al.* [101], and have been examined with respect to methodology by Felix *et al.* [102]. Subsequently, many different types of cyclopeptides have been synthesized since that early investigative period. The approach requires orthogonality of the side-chain protecting groups relative to the temporary and C-terminal protection or linker. Usually, highly acid-sensitive linkers (SASRIN[®], o-Chlorotrityl resin) are combined with acid-labile semi-permanent side-chain protecting groups of the *tert*-butyl type. However, an allyl-type protection scheme (OAll/Alloc) has also been tested [103].

Hruby *et al.* contributed a large variety of sterically constrained cyclopeptides where cyclization is mainly brought about by disulfide formation between cysteine or penicillamine (Pen, β,β -dimethylcysteine) residues. Conformational and topographical constraints in opioid peptides play a major role with respect to selectivity for the opioid receptor subtypes μ , δ , and κ . The cyclic disulfide H-Tyr-c-(D-Pen-Gly-Phe-D-Pen)-OH (DPDPE) represents an important reference compound and, hence, a milestone in the development of conformationally constrained receptor-selective peptides [104]. There are two types of conformational restriction present in DPDPE; cyclization to a 14-membered ring and the presence of two pairs of topographically

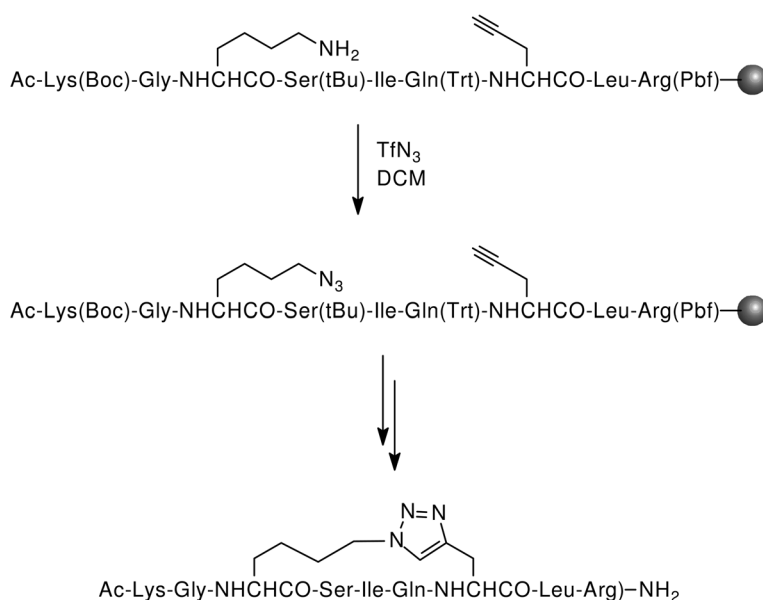


Figure 6.12 Side-chain to side-chain cyclization by “click chemistry” triazole formation.

constraining geminal methyl groups in the β -position of Pen. This results in conformational homogeneity and allowed determination of the solution conformation using a combination of high-resolution NMR and computational methods. The approach of cyclizing peptides across side-chain thiol groups such as disulfides has been successfully employed to other bioactive peptide sequences, like c-(Cys⁴,Cys¹⁰)- α -MSH and other α -MSH analogues [23, 105, 106]

Besides formation of the above-mentioned functional groups, olefin metathesis [107] and azo cyclization [108] have been employed to obtain cyclic peptides. Meldal *et al.* [109] published the successful synthesis of a peptide that was cyclized by a Cu^I-catalyzed 1,3-dipolar cycloaddition (“click chemistry”) [110]. Only recently, the peptide cyclization by a Cu^I-catalyzed 1,3-dipolar cycloaddition between a lysine-derived ω -azide and propargylglycine was reported. The thus-formed 1,2,3-triazole ring is assumed to be quasi-isosteric to a lactam bridge (Figure 6.12) [111].

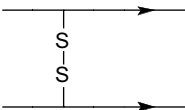
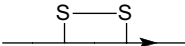
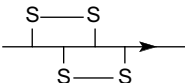
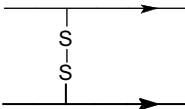
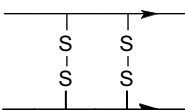
6.2

Cystine Peptides [112–115]

Disulfide bridges, which are formed upon oxidative cross-linking of two thiol groups of cysteine residues, frequently occur in peptides and proteins. As discussed in Chapter 2, disulfide bonds contribute greatly to the formation and stabilization of distinct three-dimensional structures in secreted proteins and peptides.

A distinction can be made between intrachain and interchain disulfide bridges. In the former type, a disulfide bridge links two cysteine residues of one peptide

Table 6.2 Symmetrical and unsymmetrical cystine peptides.

Type	Occurrence
Symmetrical cystine peptides 	This motif does not occur very frequently in nature, but it is sporadically found in dimerized peptides and proteins. The most common representative is oxidized glutathione.
Unsymmetrical cystine peptides - with one intrachain disulfide bridge 	These cyclic heterodetic structures are very common in peptide hormones such as oxytocin, vasopressin, and somatostatin. Furthermore, the structural motif has often been used for the design of conformationally constrained cyclic peptides.
- with two or more intrachain disulfide bridges 	These polycyclic heterodetic structures are found in many bioactive cystine-rich peptides, which are increasingly discovered and isolated from all kingdoms of life. The main representatives are (neuro)hormones, growth factors, protease inhibitors, the large families of cystine-rich toxins in venoms of spiders, snails, scorpions, snakes etc, as well as the antimicrobial peptides and components of the innate immune system.
- with one interchain disulfide bridge 	Two different peptide chains are connected across a disulfide bridge. This motif is produced in enzymatic protein digests.
- with two or more interchain disulfide bridges 	This motif leading to cyclic heterodetic structures is found e.g. in the family of insulin and relaxin peptides as products of post-translational processing of pro-forms with the respective intrachain disulfides. In this case the two different peptide chains may be arranged in a parallel or antiparallel manner.

chain leading to a cyclic peptide; by contrast, interchain disulfides produce bridges between two peptide strands. Disulfide bridged peptides can be classified as shown in Table 6.2.

The synthesis of simple monocyclic disulfide-bridged peptides is usually straightforward [113]. The linear peptide sequence is assembled either in solution or on a solid support, with both thiol groups bearing identical protecting groups. Following cleavage of either all side-chain-protecting groups or, alternatively, of only the thiol-protecting groups, cyclization is generally performed in solution under conditions of high dilution (10^{-3} – 10^{-5} M), but is also increasingly performed even on resin exploiting the pseudo-dilution effect on the solid surface. Generally, air oxygen,

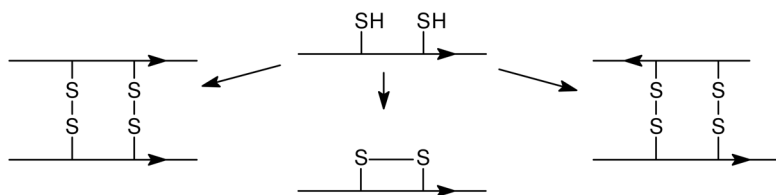


Figure 6.13 Side reactions in peptide cyclization via disulfide bond formation.

iodine, di(2-pyridyl)disulfide or DMSO serve as oxidants, while other reagents such as azodicarboxylates, pyridylsulfenyl chlorides, alkylloxycarbonylsulfenyl chlorides, alkyltrichlorosilane/sulfoxide or thallium(III) trifluoroacetate are applied more frequently in the case of stepwise formation of multiple disulfide bridges. Cyclodimerization and oligomerization may be observed as undesired side reactions (Figure 6.13).

Some thiol-protecting groups allow for simultaneous cleavage and disulfide formation (cf. Section 4.2.4.4) [116]. Indeed, treatment of peptides containing Cys(Acm) or Cys(Trt) residues with iodine leads to concurrent cleavage of the protecting groups and formation of the disulfide via the intermediate sulfenyl iodide that attacks a second Cys(Acm) or Cys(Trt) residue, thus generating the disulfide bond. As the rate of this reaction depends strongly upon the nature of the thiol-protecting group, it can also be exploited for selective disulfide formation in multiple Cys-containing peptides (see below).

The synthesis of single- or double-stranded multiple cystine-peptides is far more challenging. Various basic strategies have evolved which have to be judiciously applied, depending upon the target molecule. These are

- chain assembly by condensation of fragments, which already contain the target disulfide bridges followed by regioselective formation of the additional intra- or interchain disulfides;
- synthesis of the appropriately protected cysteine-rich peptides and their subsequent conversion into the cystine peptides with the correct disulfide connectivities in successive regioselective reaction steps;
- oxidative refolding of the cysteine-rich peptides by exploiting the structural information encoded in the primary sequence that can dictate a prevalent and even exclusive formation of the correct disulfide isomer;
- induction of the desired disulfide connectivities with selenocysteine. The first approach was applied particularly in the past for the assembly of peptide chains in solution by the fragment condensation procedure as elegantly exemplified by the successful synthesis of crystalline human insulin (Figure 6.14A) [117].

For the second strategy the availability of various orthogonal thiol-protecting groups is essential. Despite the large number of such protecting groups developed over the last decades [116], their orthogonal and regioselective conversion into disulfides still represents a challenging task that is severely limited by the number of disulfide bonds to be formed. Indeed, only a few examples with more than three

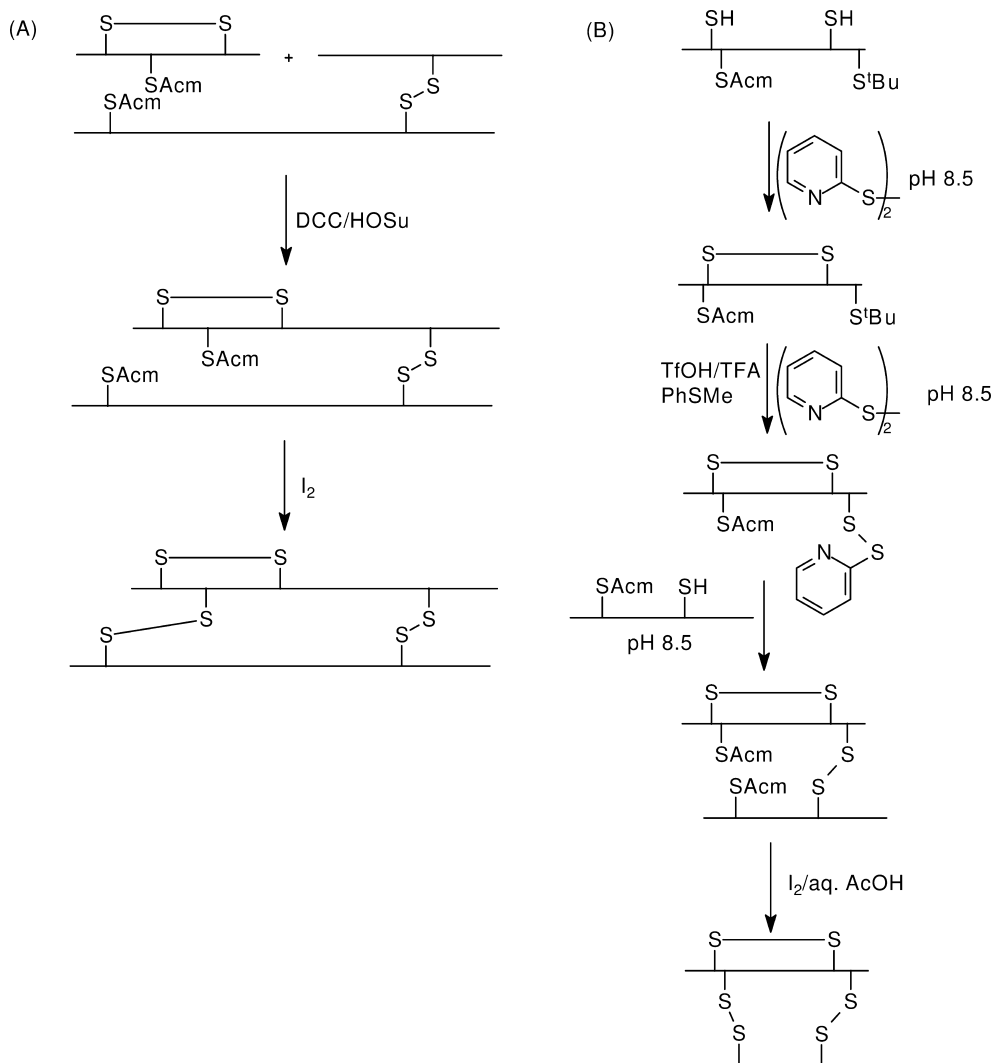


Figure 6.14 (A) Assembly of human insulin by the fragment condensation approach with preformed disulfides [117]. (B) Regioselective disulfide formation for the assembly of relaxin 3 [118].

disulfide bridges have been reported so far. The multiple cysteine peptides are synthesized either by the Fmoc- or Boc-tactics applying optimal overall combinations of protecting groups that allow, upon cleavage of the peptide from the resin, a stepwise pairing of the cysteine residues, usually starting with oxidation of the first pair of deprotected thiol groups. This is followed by stepwise regioselective cross-linking of additional pairs of suitably protected cysteine residues via disulfide bond formation, as shown in Figure 6.14B for human relaxin 3 [118]. Thereby the different

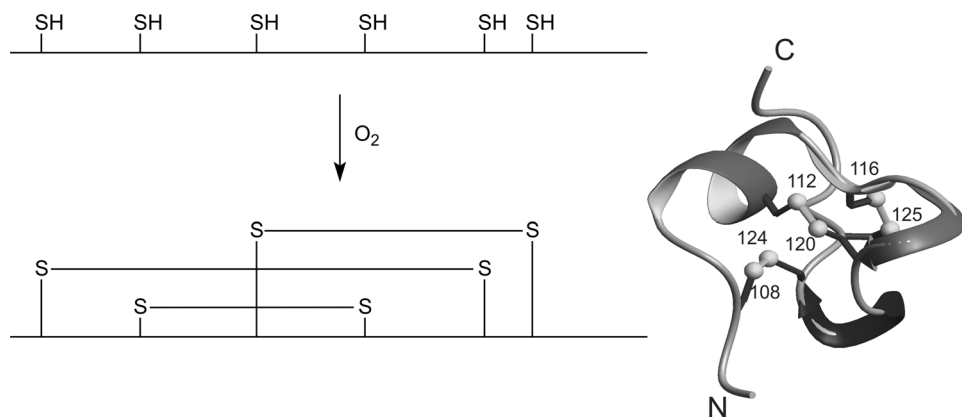


Figure 6.15 Oxidative folding of the C-terminal domain 107–130 of minicollagen-1 from *Hydra* (yield: quantitative) [122].

reactivities of the thiol derivatives towards acidolysis, as well as towards the various oxidizing agents, can be exploited.

More recently, the large experience gained in oxidative folding of proteins has been successfully transferred to cysteine-rich peptides. Indeed, for numerous single-stranded peptides containing two or more disulfides, such an approach was found to generate the correct disulfide isomers in satisfactory to practically quantitative yields [114, 115, 119, 120]. This has been observed even for protein subdomains such as the N-terminal (1–33) [121] and C-terminal (107–130) domains of minicollagen-1 from *Hydra* (Figure 6.15) [122]. The high yields of oxidative refolding of relatively short peptide sequences are mainly attributed to the thermodynamically favored compact structures that originate from burying hydrophobic residues including the disulfides (see Figure 6.16A). This thermodynamically coupled process of folding and disulfide formation is even responsible for the relatively high yields of oxidative

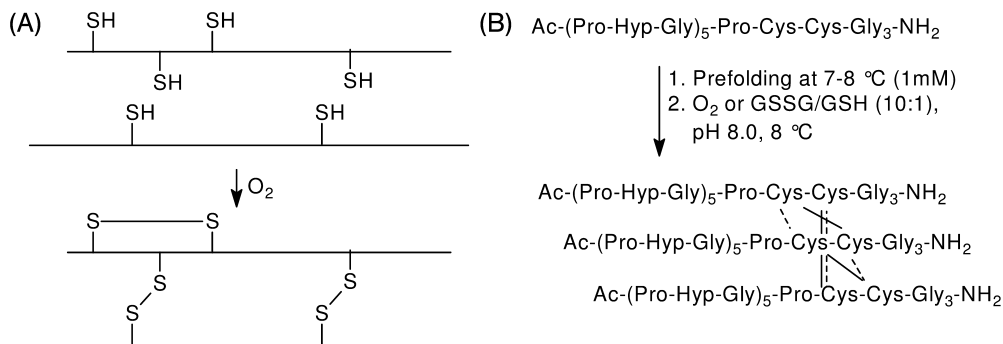


Figure 6.16 (A) Oxidative recombination of the human insulin A- and B-chain [123, 124].

(B) Oxidative recombination of a collagen peptide into the homotrimer using the collagen type III cystine knot [126].

recombination of double-stranded cysteine-rich peptides, e.g. human insulin, that generates the desired disulfide connectivities at extents by far superior to the statistically expected yields (Figure 6.16) [123–125]. In contrast, in the case of triple-stranded homotrimeric collagen peptides related to collagen type III it is the folded state, i.e. the triple helix, that preorganizes the cysteine residues in the correct position for the prevalent formation of the correct cystine-knot, the exact disulfide connectivities of which are still not known (Figure 6.16B) [126]. Alternatively, for the correct pairing of two cysteine residues displaced on different peptide chains, activation of one thiol as an unsymmetrical disulfide of the nitropyridylsulfenyl or alkoxycarbonylsulfenyl type or as sulfenohydrazide is required for efficient thiolysis by the second unprotected cysteine residue to produce the interchain disulfide bond, as shown in Figure 6.14B for the regioselective assembly of human relaxin 3. By this strategy even heterotrimers or architectures of higher order can be constructed [114, 127, 128]. Thereby slightly acidic conditions are required to prevent disulfide scrambling of already established disulfide bonds.

An alternative recently developed strategy is based on selenocysteine as an isosteric replacement of cysteine [114, 129]. Its highly reducing redox potential of -381 mV [130] compared to that of cysteine leads generally to exclusive diselenide formation independently of the presence of additional cysteine thiol groups. This advantageous redox potential can be exploited for induction of the desired additional disulfide connectivities into the native and even non-native isomers.

6.3 Glycopeptides

The various properties and functions of glycoproteins are discussed in Section 3.2.2.4. In order to perform biological studies and therapeutic evaluations, a large amount of material is needed and, consequently, the synthesis of glycopeptides as models for glycoproteins is currently of great interest. With the methods of native protein ligation, even the chemical synthesis of uniformly glycosylated proteins is now within reach. This synthetic task represents a major challenge, however, because, in addition to the requirements of peptide synthesis, the methods of carbohydrate chemistry must also be considered.

Glycopeptide synthesis and the so-called glycopeptide remodeling are of major importance because many glycoproteins are of pharmaceutical interest; examples include juvenile human growth hormone, CD4, and the tissue plasminogen activator. Glycopeptide remodeling requires modification of glycoproteins by the removal or addition of carbohydrate units. Without doubt, enzymes are very much suited to manipulations of the oligosaccharide moiety of glycoproteins, the sensitivity and polyfunctionality of which requires high selectivity. Glycosyltransferases experience increasing application in glycopeptide and glycoprotein synthesis [131]. However, several of the glycosyltransferases that might be used for this purpose are still not available. It should be noted that, in addition to remodeling, the synthesis of new protein oligosaccharide conjugates is of major importance [132, 133]. As well as the

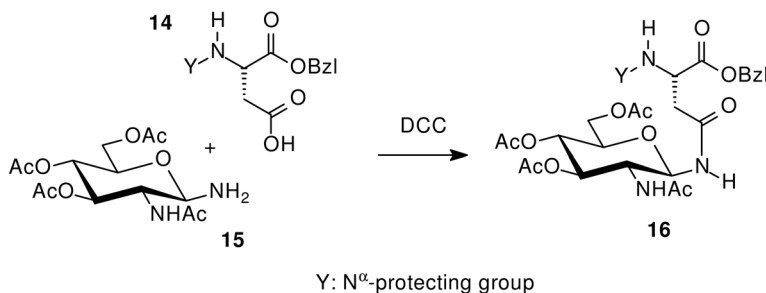
problems of peptide synthesis that have been discussed previously, the chemical synthesis of glycopeptides imposes high requirements with regard to the reversible and highly selective protection of additional functional groups, and also to the stereoselective formation of glycosidic bonds [131, 134–142]. The development of the methodology for glycopeptide synthesis requires special consideration of the additional complexity and lability of the carbohydrate moiety [133]. The protecting groups must be chosen correctly to ensure that their removal is selective, and does not interfere with the acid- and base-labile oligosaccharide groups.

The formation of a glycosidic bond between the carbohydrate moiety and the peptide unit is the crucial step in glycopeptide synthesis and, in general, two strategies for this can be distinguished:

- The most common approach uses glycosylated amino acid building blocks which are appropriately protected for the stepwise synthesis of the glycopeptide.
- The block glycosylation approach utilizes conjugation of the final carbohydrate unit to the full-length peptide. Unfortunately this approach is hampered by side reactions such as aspartimide formation in the case of N-glycosides, or difficulties in forming stereoselective O-glycosylation reactions with complex targets, not to mention the complex protecting-groups strategies necessary.

The hydroxy functions present in the carbohydrates are usually reversibly blocked by either acyl, acetal, or benzyl ether protecting groups. In particular, benzyl ethers and benzylidene groups exclude the simultaneous application of the benzyloxycarbonyl group and of benzyl esters as temporary protecting groups in the peptide moiety. The high chemical lability of the glycosidic bonds greatly complicates peptide synthesis. Glycosidic bonds are usually hydrolyzed under acidic conditions, and a permanent risk of β -elimination exists for all glycosyl serine and threonine derivatives, even under weakly basic conditions [143].

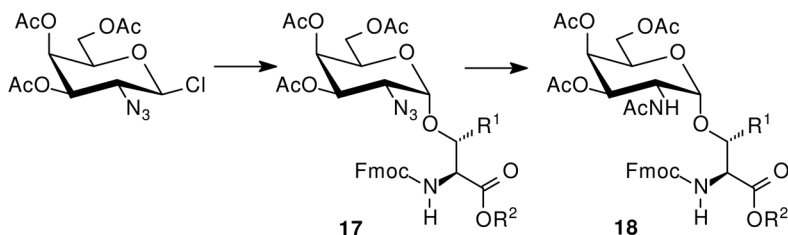
The β -N-glycosidic bond between N-acetylglucosamine and asparagine is formed by carbodiimide coupling of N-protected aspartic acid α -benzyl ester **14** with 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy- β -glucopyranosylamine **15** (Scheme 6.1), the latter molecule being obtained from the corresponding glucosyl chloride via the azide. The azide is selectively reduced to the amine by catalytic hydrogenation on Raney-nickel without cleaving the benzyl-type protecting groups. The conjugate **16** is formed in 49% yield.



Scheme 6.1

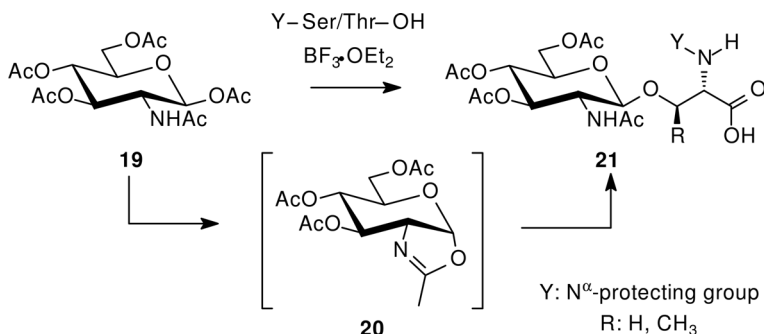
Glycosyl amines readily undergo anomerization under acidic conditions, where the amine is protonated. Equilibration favors the β -anomer as a result of the reverse anomeric effect [136]. For the synthesis of more complex N-glycans, block condensation of the pre-formed full-length glycosylamine with the β -activated aspartic acid derivative is preferred for the introduction of the carbohydrate moiety, as in the chemoenzymatic synthesis of a sialylated undecasaccharide-asparagine conjugate as published by Unverzagt (Figure 6.17) [144].

The formation of an α -O-glycosidic bond between *N*-acetylgalactosamine and serine or threonine very often makes use of the azido group as a precursor for the 2-acetamido moiety. The azido group does not exert neighboring group participation like the 2-acetamido group, which would lead to predominant formation of the β -anomer. Most often, the 1-halogeno carbohydrates are employed in Königs–Knorr type glycosylations of Fmoc-protected serine and threonine esters. Highest α -selectivities are obtained upon application of an insoluble promoter such as silver perchlorate. Subsequently, the 2-azido group of **17** can be reduced and acetylated to give the 2-acetylamido group of **18** (Scheme 6.2). Selective reduction of the azido group in complex glycopeptides is achieved using nickel tetrahydridoborate.



Scheme 6.2

The synthesis of a β -O-glycosidic bond between *N*-acetylglucosamine and serine or threonine makes use of the anchimeric assistance of the 2-acetamido group in **19** (Scheme 6.3). $\text{BF}_3 \cdot \text{Et}_2\text{O}$ activation of glycosyl acetates clearly improves the yields of N-protected glycosyl amino acid derivatives **21** [145, 146]. The application of readily available anomeric acetates usually does not require protection of the amino acid carboxy group, thus rendering protecting-group manipulation unnecessary.



Scheme 6.3

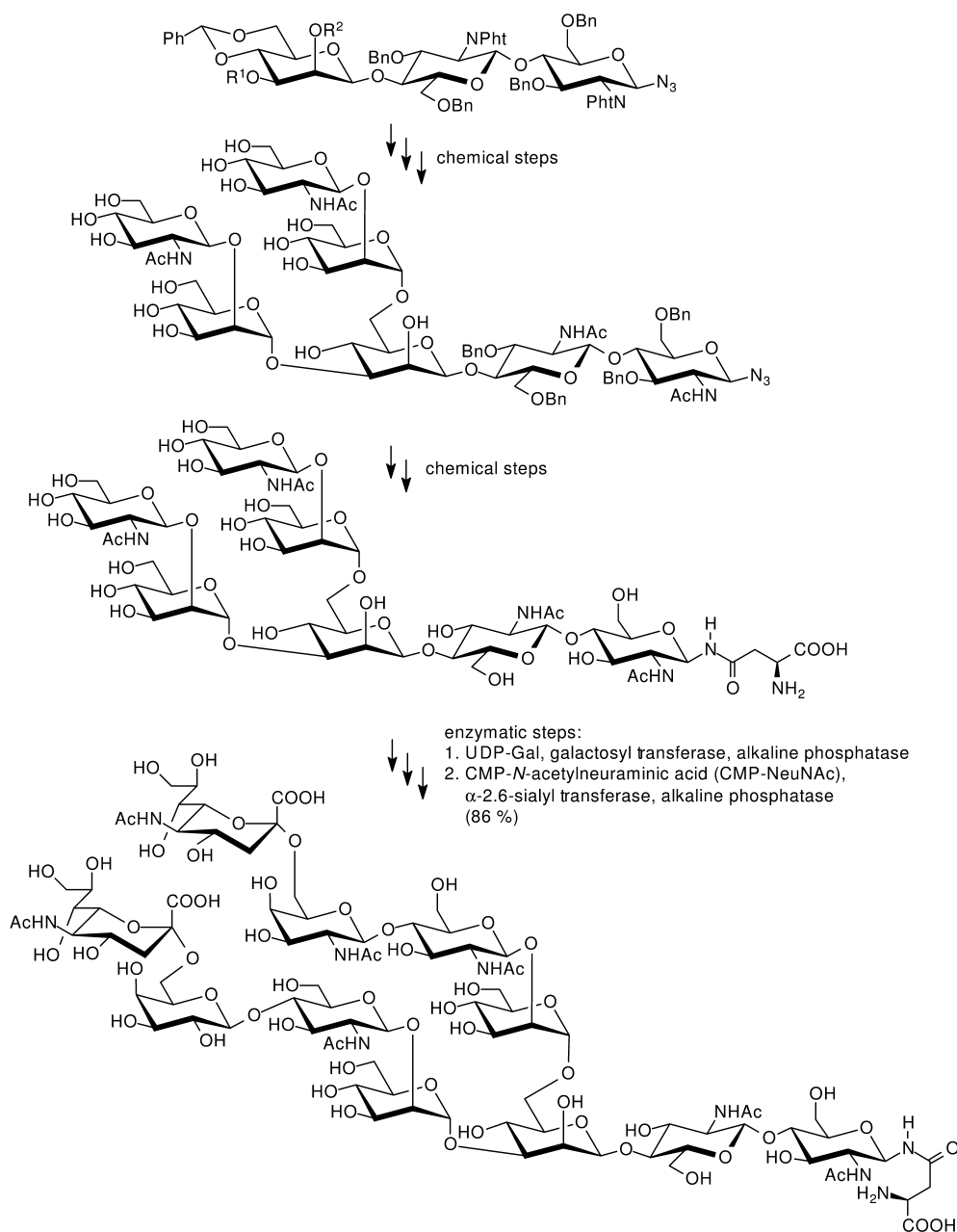
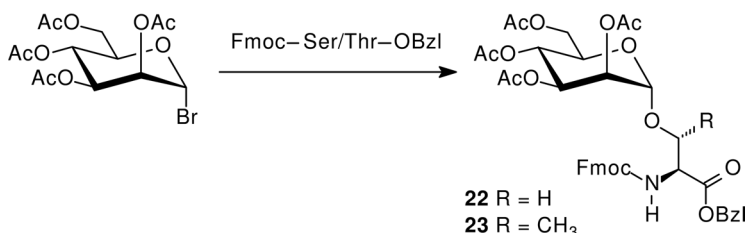


Figure 6.17 Chemoenzymatic synthesis of the sialylated undecasaccharide-asparagine conjugate according to Unverzagt *et al.* [144].

Both the anomeric effect and anchimeric assistance of the acetyl group present on the 2-hydroxy function are exploited for the selective synthesis of α -mannosyl-serine **22** ($R = H$) and threonine **23** ($R = CH_3$) under Königs–Knorr-type conditions (Scheme 6.4) [147]



Scheme 6.4

In principle, glycopeptides may be synthesized either in solution or on a solid phase. An N-glycopeptide cluster containing two sialyl-Lewis^x moieties has been synthesized in solution by Kunz *et al.* [148]. First, the asparagine residue is protected as the allyl ester, and with the Boc group; the glycosylated amino acid is then selectively deprotected at the amino group, without hampering the acetyl-protecting groups of the carbohydrate moiety. The fully protected glycopeptide is then assembled using preformed glycosylated amino acid building blocks. Allyl ester protection of the carboxy group allows selective deblocking by rhodium(I)-catalyzed deallylation. Finally, the *tert*-butyl ester can be cleaved using formic acid, after which the acetyl groups of the carbohydrate moieties are removed using highly diluted sodium methoxide.

A very useful protocol for the synthesis of glycopeptides makes use of enzymatic reactions. Glycosyl amino acids are thus incorporated into oligopeptides by enzymatic coupling steps, and further elaboration of the oligosaccharide part can be achieved using glycosyltransferases. This enzymatic strategy allows synthesis in aqueous solution and requires only minimum protection. The synthesis of a glycopeptide using a thermostable thiolsubtilisin mutant (S221C, M50F, N76D, G169S) is shown in Figure 6.18. This mutant favors aminolysis over hydrolysis by a factor of approximately 10^4 [149]. Glycosyltransferases, as well as exo- and endoglycosidases, are valuable catalysts for the formation of specific glycosidic linkages. The glycosyltransferases transfer a given carbohydrate from the corresponding sugar nucleotide donor to a specific hydroxy group of the acceptor sugar. Currently, a large number of eukaryotic glycosyltransferases (e.g., β 1–4-galactosyl transferase, β 1–4-GalT; Figure 6.18) have been cloned that exhibit exquisite linkage and substrate specificity [131].

In the chemical synthesis of glycopeptides the complexity and lability of the carbohydrate moieties have to be taken into account. The selective removal of glycosyl amino acid or glycopeptide protecting groups has remained an unsolved problem for quite some time, having first been tackled during the late 1970s. A Z group in

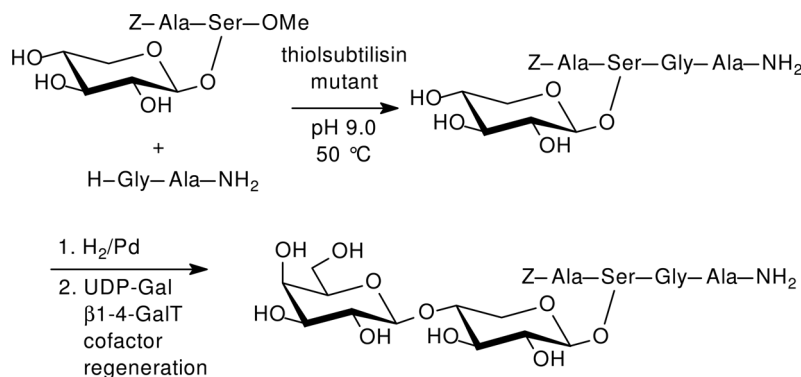
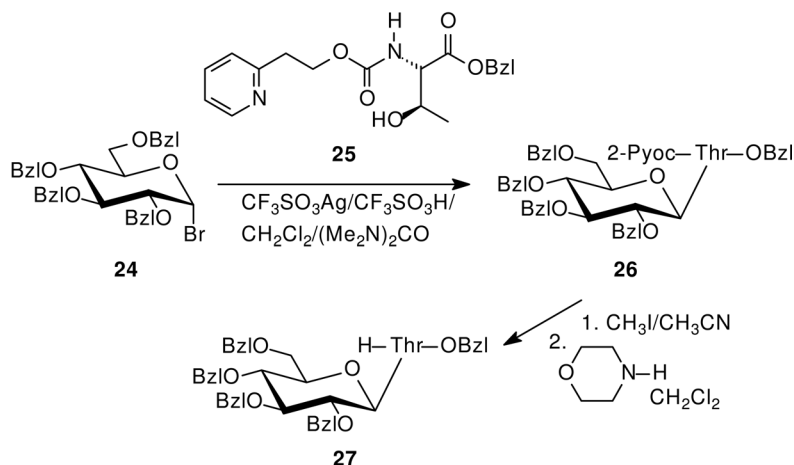


Figure 6.18 Enzymatic synthesis of an O-glycopeptide using a subtilisin mutant for peptide coupling and a galactosyl transferase for glycosylation.

β -glycosyl amino acid derivatives cannot be cleaved using HBr/AcOH without destroying glycosidic bonds. Application of Boc as the temporary protecting group in glycopeptide synthesis is usually not possible, because recurring Boc cleavage requires acidolysis which is incompatible with the acid-labile glycosidic bonds. Although cleavage of the Boc residue is possible under certain reaction conditions, acidolytic deprotection reactions should be used in the synthesis of glycopeptides only after careful consideration. Glycopeptides with an O-linkage to Ser/Thr easily undergo β -elimination in the presence of strong bases. Hence, the base sensitivity of glycosylserine or glycosylthreonine derivatives further restricts the repertoire of deblocking reactions.

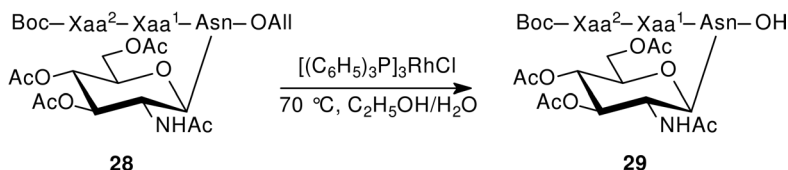
Usually Fmoc is employed as the temporary N^α -protecting group. Iterative treatment with bases like piperidine or morpholine does not normally affect the glycopeptide. The oligosaccharide part is in the majority of cases protected with acetyl groups. However, this is not always free of complications and the O-benzyl protection that is also frequently used in carbohydrate chemistry [150] is incompatible with Cys or Met containing peptides.

Protecting groups must allow cleavage under either mild or neutral reaction conditions. The two-step protecting groups fulfil both requirements, namely stability during synthesis, and lability on deblocking. The 2-pyridylethoxycarbonyl (2-Pyoc) group and its 4-pyridyl analogue (4-Pyoc) are stable under both acidic and basic conditions, but can be converted into labile derivatives upon methylation at the pyridine nitrogen atom. The alkylated derivatives are cleaved under very mild conditions (morpholine/ CH_2Cl_2). Glycosylation of 2-Pyoc-Thr-OBzl **25** starting from the glucopyranosylbromide **24** according to the silver triflate procedure in the presence of an equimolar amount of trifluoromethanesulfonic acid provides access to the glycosyl amino acid **26** (Scheme 6.5). The labilizing reactions for cleavage to give **27** as mentioned above can, however, only be applied if no other amino acids that would readily undergo alkylation reactions are present in the target sequence.



Scheme 6.5

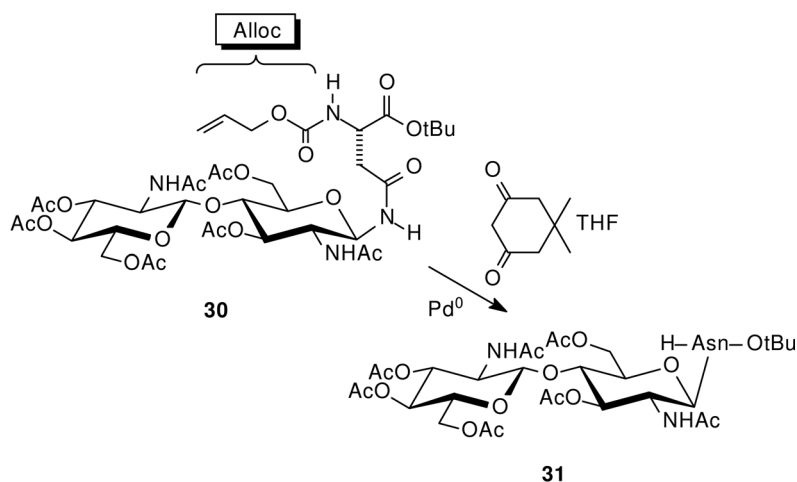
The allyl-type protecting groups are completely orthogonal to most other protecting groups, and provide an excellent method for temporary reversible protection in glycopeptide synthesis. Allyl esters of amino acids are very easy to obtain, are stable under glycosylation conditions, and can be cleaved by Rh^{I} catalysis, as shown in the deprotection of the glycopeptide **28** to give **29** (Scheme 6.6).



Scheme 6.6

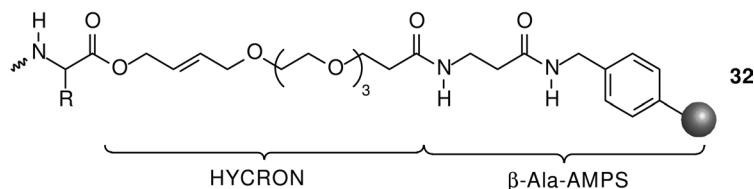
The allyl ester is isomerized under these reaction conditions to produce a vinyl ester (propenyl ester) that immediately undergoes hydrolysis under the reaction conditions.

An even milder method for the selective cleavage of allyl esters utilizes palladium(0)-catalyzed allyl transfer to morpholine. This principle is especially suited for the sensitive, elimination-prone O-glycosylserine and O-glycosylthreonine derivatives, and also allows smooth cleavage of the allyloxycarbonyl (Alloc) group in peptides and glycopeptides [151]. The Alloc residue is stable towards treatment with trifluoroacetic acid (TFA). Cleavage of the Alloc protecting group from the glycosylasparagine derivative **30** results in the formation of **31** in 90% yield (Scheme 6.7). The allyl group in this case is transferred to the 1,3-diketone as the nucleophile.

**Scheme 6.7**

Solid-phase peptide synthesis (SPPS) can also be adapted to the construction of glycopeptides. Basically, most linker systems already described for SPPS (Wang, HMPA, SASRIN, HYCRON, Rink amide, PAL, Sieber; see Section 4.5.1) can be used for the assembly of glycopeptides. Allyl-type linker moieties have proven appropriate for the solid-phase synthesis of complex glycopeptides [152]. Flexible polar oligo-ethyleneglycol spacers are often used to minimize steric hindrance and associations with the hydrophobic polystyrene matrix.

The HYCRON linker **32** was applied successfully in a synthesis of peptide T, an O-glycosylated tetrapeptide which occurs in the sequence of the HIV envelope protein gp120, and of an O-glycosylated glycopeptide of the mucin type [153]. Details of further studies in the solid-phase synthesis of glycopeptides are outlined in several references [134–137, 154].



The swelling parameters of the solid support deserve special attention, because high mass transfer rates for larger reagents or catalysts are also required. For combined application of chemical and enzymatic synthesis, good swelling behavior both in organic solvents and in aqueous buffered solutions is necessary. For the latter purpose, hydrophilic resins like kieselguhr-supported poly(dimethylacrylamide) or polyethylene glycol based resins are preferable.

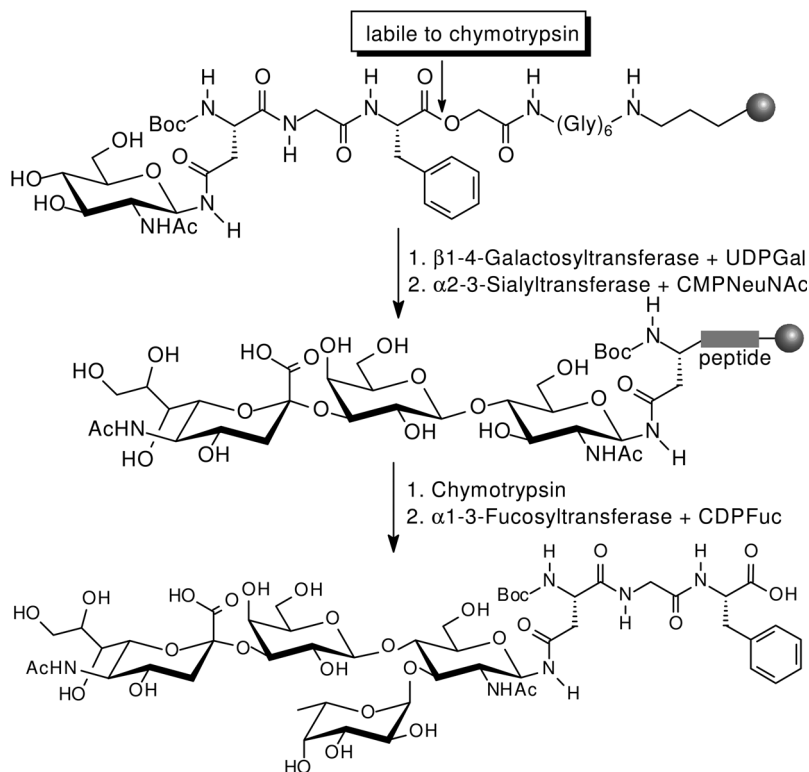


Figure 6.19 Solid-phase glycopeptide synthesis involving enzymatic glycosyl transfer and a chymotrypsin-labile resin anchor.

Enzymatic glycosylation may also be performed on an aminopropyl-modified silica gel support [155]. An enzyme-labile linker molecule was used in the synthesis shown in Figure 6.19.

This elegant strategy allows rapid and iterative formation of peptide and glycosidic bonds in either organic or aqueous solvents, including enzymatic deprotection of the glycopeptide from the solid support [156]. The power of chemoenzymatic synthesis was, for example, elegantly demonstrated by the construction of the glycopeptide P-selectin ligand-1 (PSGL-1) [157].

Chemical glycoprotein synthesis nowadays very much relies on native chemical ligation (Section 5.5). Since the discovery of NCL, a series of glycoproteins has been chemically synthesized. Native chemical ligation relies on mild and selective reactions that are compatible with the presence of glycans. Bertozzi and coworkers synthesized dipteracin, an anti-microbial glycoprotein with 82 amino acids and two N-acetylgalactosamine residues following this approach for the N-terminal glycopeptide thioester [158]. A defined glycoform of GlyCAM-1 with thirteen GalNAc residues was synthesized from three synthetic peptides and a recombinant protein thioester [159]. Glycopeptides with a C-terminal thioester have been synthesized chemically and subsequently ligated with a recombinant protein part. One of the

most attractive methods for the chemical synthesis of peptide thioesters relies on Kenner's safety-catch sulfonamide linker (Chapter 4), which permits the peptide synthesis by using the Fmoc/tBu protection scheme [160].

Safety-catch linkers require chemical activation prior to cleavage. This renders mass spectrometric monitoring of the peptide synthesis very tedious, because for MS analysis a small amount of resin has to be activated and subsequently cleaved. Unverzagt *et al.* [161] developed an orthogonal combination of the sulfonamide safety-catch linker with an acidolytically cleavable Rink-amide linker. The latter allows facile detachment of the glycopeptide chain for step-by-step mass spectrometric synthesis monitoring.

Wong *et al.* [162] developed the concept of native chemical ligation by sugar-assisted ligation for the synthesis of glycopeptides, which was successfully applied to the synthesis of β -O- and N-linked glycopeptides (cf. Section 5.5.1.2). It does not require a Cys residue, but relies on the presence of a mercaptoacetyl group which replaces the acetyl group of e.g. a GlcNAc or GalNAc residue.

6.4

Phosphopeptides

In view of the importance of protein phosphorylation (cf. Section 3.2.2.6), it is clear that phosphorylated peptides are highly valuable tools for the study of protein phosphorylation and dephosphorylation, as well as for the recognition of phosphorylated proteins. Phosphorylated peptides, for example, can be used to determine the specificity of protein phosphatases [163], and may be synthesized via two fundamentally different routes: (i) global phosphorylation; or (ii) a building block approach. Whilst the former method – which is also known as post-assembly phosphorylation – makes use of selectively side-chain-deprotected serine, threonine or tyrosine residues that are phosphorylated on completion of the synthesis, the building block approach utilizes phosphorylated amino acids. This, of course, imposes further problems with respect to protecting-group strategy. Both solution-phase and solid-phase syntheses of phosphorylated peptides have been reported. Different types of protecting groups have been used for the phosphate group: allyl, methyl, benzyl, *tert*-butyl, and 2,2-dichloroethyl. Appropriate phosphoserine, phosphothreonine, or phosphotyrosine derivatives for the building block approach can be synthesized by two different methods.

The first of these methods involves phosphorylation with a dialkyl- or diallylchlorophosphate under alkaline conditions (Figure 6.20A). One advantage of this method is that the phosphorus atom is already in the correct oxidation state P(V). Alternately, Fmoc-Tyr(PO(NMe₂)₂)-OH, for example, may be applied in peptide synthesis [164].

The second method uses phosphorus(III) compounds, like those employed in oligonucleotide synthesis (Figure 6.20B). A phosphoramidite is reacted with a suitably protected amino acid under mildly acidic conditions (tetrazole) to form a phosphite. Subsequently, an oxidation is performed using iodine, meta-chloroperbenzoic acid, or *tert*-butylhydroperoxide. Phosphorylation is usually performed on N $^{\alpha}$ -protected

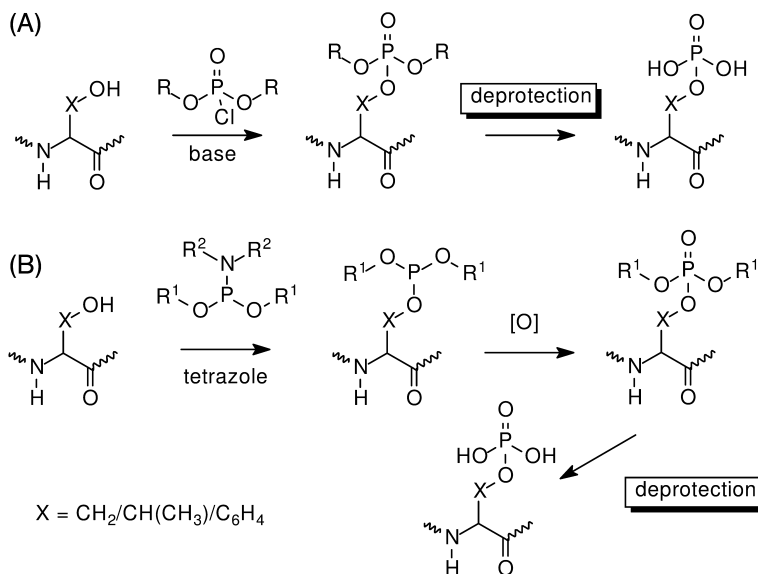


Figure 6.20 Phosphorylation with phosphorus(V) and phosphorus(III) reagents.

amino acids where the carboxy group is also blocked. Following phosphorylation of the side-chain hydroxy group, the carboxy group protection is selectively removed.

Peptide synthesis methodology is somewhat limited for serine and threonine derivatives, because these compounds readily undergo base-mediated β -elimination to produce dehydroamino acid derivatives; hence, the N^α -Fmoc protection scheme is not universally applicable. The N^α -Alloc protection scheme is highly compatible with the synthesis of phosphopeptides. β -Elimination is suppressed when only one hydroxy function of the phosphate group is protected. Hence, under basic conditions this phosphate group is deprotonated and, consequently, converted into a poor leaving group. Under these conditions, Fmoc tactics can be applied to the synthesis of phosphopeptides with Fmoc-Ser(PO(OBzl)OH)-OH [165, 166] and Fmoc-Thr(PO(OBzl)OH)-OH [166]. Partially protected phosphoamino acids, however, suffer from both pyrophosphate formation [167] and incompatibility with PyBOP and carbodiimide coupling reagents [168].

Enzyme-labile protecting group techniques provide in these cases an interesting and advantageous alternative (cf. Section 4.2.5). The selectively phosphorylated pentapeptide (Figure 6.21) has been synthesized by a combination of enzymatic and chemical protecting group operations. In this sequence, both heptyl esters (cleavage with lipases) and allyl esters (cleavage with Pd⁰) have been employed as protecting groups for the C-terminus. The N-terminus may be protected with an allyl-type group (Alloc) or, alternately, with an enzyme-labile protecting group such as Z (OAcPh) (Figure 6.22).

Treatment with piperidine not only leads to β -elimination, but may also cleave one alkyl group from dimethyl- or dibenzyl-protected phosphotyrosine. This dealkylation

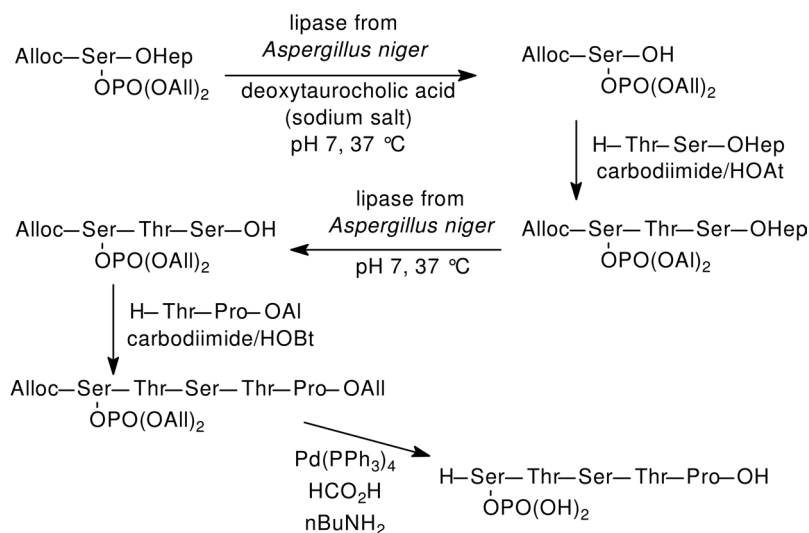


Figure 6.21 Chemoenzymatic synthesis of phosphopeptides involving the enzyme-labile heptyl ester.

can be suppressed by applying an alternative Fmoc deprotection reagent such as 2% 1,8-diazabicyclo [5.4.0]undec-7-ene (DBU) in dimethylformamide (DMF).

Most coupling reagents are compatible with phosphorylated building blocks. If the phosphate group is unprotected, then amino acid active esters may be used. Benzyl and *tert*-butyl protecting groups on the phosphate moiety are labile to TFA-based

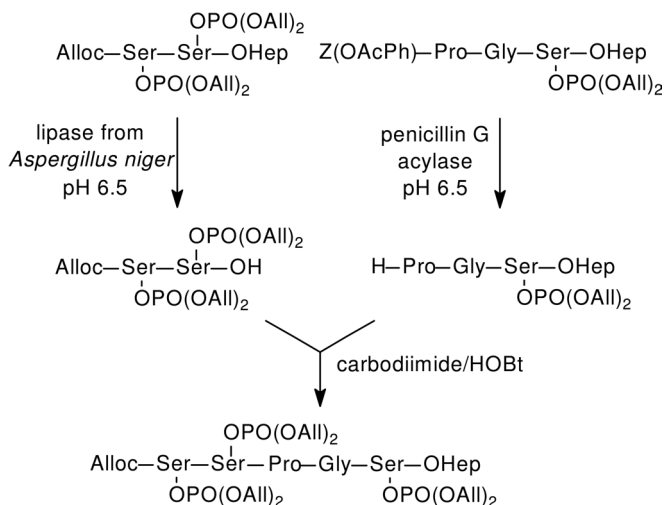


Figure 6.22 Chemoenzymatic synthesis of phosphopeptides involving the enzyme-labile heptyl ester and Z(OAcPh)-protecting groups.

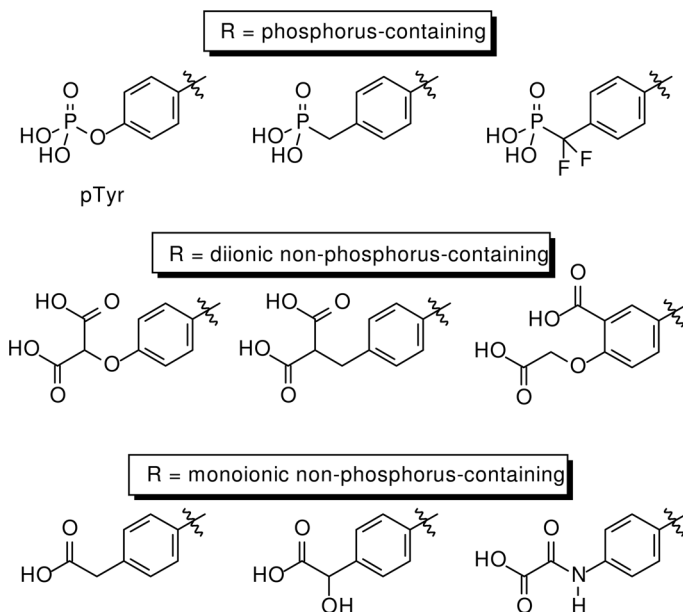


Figure 6.23 Phosphotyrosine (pTyr) mimetics.

cleavage conditions typically used in Fmoc tactics. Treatment of pTyr-containing peptides with liquid HF causes significant dephosphorylation. This also holds true for phosphoserine/phosphothreonine-containing peptides and treatment with HBr/AcOH. Special procedures must be applied when phosphorylated peptides are to be synthesized according to Boc tactics [169].

Several nonhydrolyzable phosphotyrosine mimetics have been developed and applied to peptide synthesis [170] (Figure 6.23).

6.5

Lipopeptides

Lipopeptides form an emerging class of peptides with increasing importance. Antimicrobial lipopeptides are mainly products of bacterial cells, where they are synthesized through the nonribosomal pathway. Many lipopeptides include non-native amino acid residues and are constrained by cyclization, which improves their stability against proteolysis. Cationic lipopeptides include polymyxin B and their related compounds as well as lipopeptaibols [171]. Lipidated glycopeptide antibiotics like teicoplanin [172] have demonstrated effectiveness in the treatment of methicillin-resistant *Staphylococcus aureus* (MRSA) infection, for which a marked increase in the prevalence is being noted worldwide. Daptomycin 3, a cyclic lipopeptide antibiotic that contains a straight C₁₀ lipid side-chain, belongs to a group of 10-membered cyclic lipopeptide antibiotics that are secondary metabolites produced

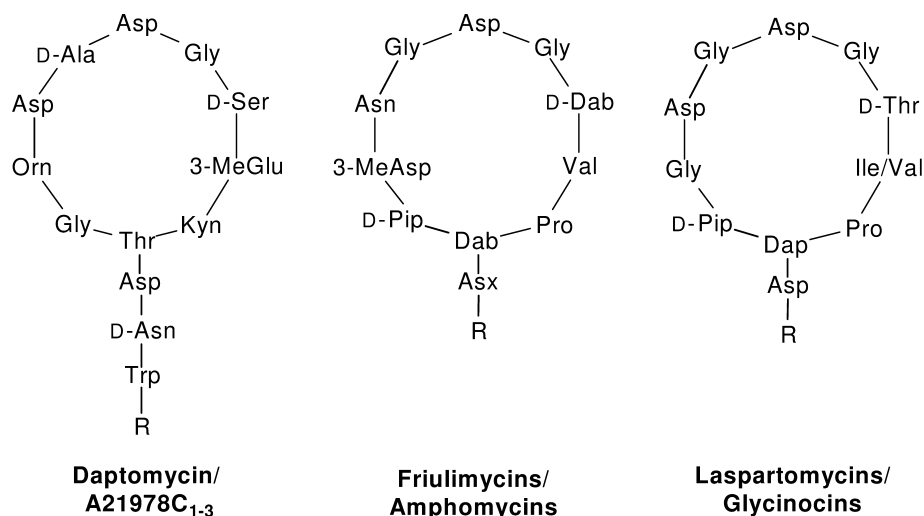


Figure 6.24 Comparison of peptide core structures for cyclic lipopeptides. R = long chain branched or unbranched acyl residue; 3-MeGlu = 3-methylglutamate; 3-MeAsp = 3-methylaspartate; Kyn = L-3-anthranilylalanine.

by actinomycetes. Other members are amphomycins, friulimycins, laspartomycins, etc. (Figure 6.24). Daptomycin (Cubicin), was approved in the USA in 2003 for the treatment of skin infections caused by Gram-positive pathogens. However, oral application is not possible because of insufficient resorption. Daptomycin is produced on a large scale by feeding decanoic acid to the fermentation of *Streptomyces roseosporus*. Studies on the chemical modification of daptomycin rely practically on semisynthesis and have so far mainly focused on peripheral modifications (exocyclic amino acids, fatty acid tail, side-chain derivatizations, e.g. of Orn) [173].

Bacterial lipoproteins, structural components of the bacterial cell wall, typically contain the amino acid *S*-glycerylcysteine, which is acylated by three long-chain acyl groups, e.g. *O,O,N*-tripalmitoyl-*S*-glycerylcysteine (Pam₃Cys, Figure 6.25). Such lipidated structures are highly immunogenic, presumably by addressing Toll-like receptors (TLR) and hence may be employed as synthetic adjuvants in vaccination strategies, when combined with a suitable peptide antigen. Minimized recognition epitopes for such lipopeptide vaccines can easily be prepared by solid-phase peptide synthesis [174–176].

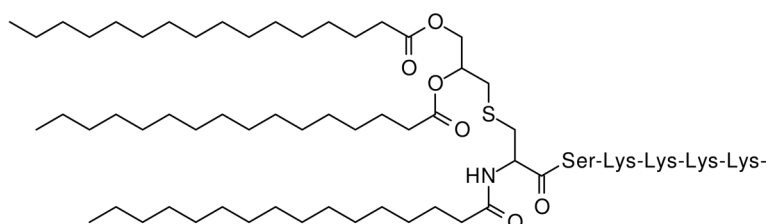


Figure 6.25 The synthetic adjuvant Pam₃Cys-Ser-(Lys)₄.

Membrane-bound proteins are frequently modified covalently by lipid residues (cf. Section 3.2.2.7) [134, 177, 178]. G-protein-coupled receptors are S-palmitoylated at a C-terminal cysteine residue. The α -subunits of heterotrimeric G-proteins and non-tyrosine receptor kinases contain N-myristoylated N-terminal glycine residues together with S-palmitoylation of a neighboring cysteine residue. The γ -subunits of G-proteins are S-farnesylated or S-geranyl-geranylated at cysteine residues.

The lipid moieties are necessary to recruit and anchor proteins to the membrane. Furthermore, it has been postulated that lipidation of proteins represents an event involved in signal transduction [179].

Lipid-modified peptides can be used as a tool in order to study the biosynthesis and processing of lipid-modified proteins, as well as their biological role. A single N-myristoylation or S-farnesylation is not sufficiently hydrophobic to achieve stable membrane attachment of peptides. Only double or multiple lipid modifications lead to stable association of the peptides to membranes [180–182].

Lipid-modified peptides are chemically very sensitive; the thiol ester moiety present in S-palmitoyl derivatives hydrolyzes spontaneously at pH >7 in aqueous solution, and β -elimination might also be problematic during synthesis. The alkene groups of farnesyl residues, for example, may undergo side reactions upon treatment with acids, and consequently all coupling and deprotection reactions in the synthesis of lipid-modified peptides must be carried out under very mild conditions [183, 184]. Classical methods can be applied only to a limited extent to the assembly of such sensitive molecules. If only S-palmitoylation of cysteine residues is required in the target peptide, then Boc tactics may be used as the thioester is acid stable. On the other hand, lipopeptides containing only the acid-labile S-farnesyl cysteine can be synthesized using modified Fmoc tactics. In addition, allyl-type protecting groups have proven useful in the synthesis of sensitive derivatives [180, 185].

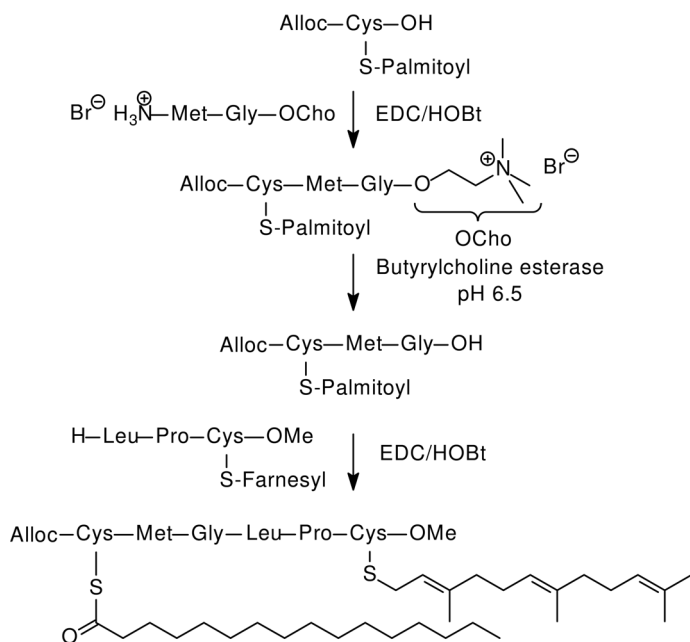
The acid-sensitivity has also to be considered when choosing an appropriate solid support for SPPS. Most linkers are not compatible with the high acid lability of the prenyl groups and consequently the linker systems are restricted to the oxime resin, trityl resin, safety-catch resin, or the newly developed hydrazinobenzoyl resin [186, 187]. The S-lipidated cysteine derivatives are employed in Fmoc- or Alloc-protected form.

Enzyme-labile protection groups are especially appropriate in the synthesis of these highly sensitive derivatives (cf. Section 4.2.5).

Synthesis of the S-palmitoylated and S-farnesylated lipohexapeptide of the C-terminus of human *N*-Ras protein could be accomplished by the application of a choline ester as carboxy-protecting group. This can be cleaved by butyrylcholine esterase at near-neutral pH, without any risk of β -elimination (Figure 6.26).

Alternatively, appropriately protected Cys residues may be employed in solid phase synthesis, combined with on-resin lipidation (Figure 6.27) [184, 187].

Site-specifically lipidated peptides have been conjugated to recombinant proteins by different methods (maleimide-ligation, native chemical ligation, Diels–Alder ligation) to give functionally active proteins [184, 187].



C-terminus of human N-Ras protein

Figure 6.26 Chemoenzymatic synthesis of lipidated peptides using enzyme-labile choline esters. OCho = choline.

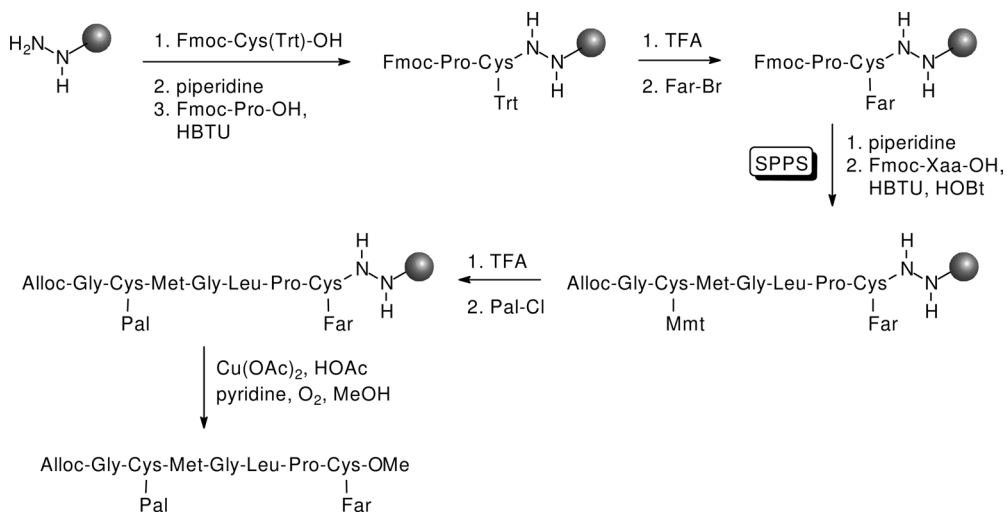


Figure 6.27 Synthesis of an N-Ras lipopeptide by on-resin lipidation.

6.6

Sulfated Peptides [188, 189]

The chemical synthesis of peptides containing tyrosine-O-sulfate (Section 3.2.2.9) is a difficult task for peptide chemistry, mainly because the sulfate moiety is intrinsically labile. When using Fmoc tactics, the final cleavage and deprotection with acid represents the critical step in the synthesis because of the acid sensitivity of the tyrosine sulfate.

The solution synthesis of tyrosine-O-sulfate-containing peptides, applying Z tactics, was described during the early 1980s. The acid lability of the tyrosine sulfate can be attenuated by strong metal counterions, or more efficiently by intramolecular ionic interactions, e.g. with proximal arginine residues [190–192].

The acid lability of amino acid O-sulfate esters has been evaluated [193, 194]. In the building block approach, barium or sodium salts of suitably N-protected tyrosine-O-sulfate are used. The N-terminus, as well as the side chains of Lys, Ser, and Thr, must be appropriately protected, and the synthesis of peptides containing multiple tyrosine sulfate residues remains one of the most challenging problems in peptide chemistry. Tyrosine may be sulfated using concentrated sulfuric acid in the cold, chlorosulfonic acid/pyridine, or with the complexes pyridine- SO_3 , $\text{Me}_3\text{N}\cdot\text{SO}_3$, or $\text{DMF}\cdot\text{SO}_3$. Post-assembly sulfation has also been reported as an alternative to the building block approach, though it suffers greatly from the lack of selectivity of the sulfation reaction. Post-assembly sulfation with pyridine- SO_3 is to be preferred for a highly acidic peptide, because in such cases the loss of sulfation in the acid deprotection step of the final peptide in the building block approach becomes critical [195].

Analogues of the C-terminal nonapeptide amide cholecystokinin-(25–33) have been synthesized in solution using Z as the temporary protecting group, together with side-chain protection of the *tert*-butyl type, with the tyrosine-O-sulfate building block being used as the barium salt [190].

Although the application of Z as the temporary protecting group is suitable for the synthesis of small peptides containing O-sulfation, it is not suitable for larger peptides. Solid phase peptide synthesis with tyrosine-O-sulfate building blocks is feasible according to the Fmoc/tBu protection scheme, because the final treatment with TFA to release the peptide from the resin and cleave the side-chain protecting groups is compatible with the sulfated peptides, if the reaction conditions are carefully controlled. Gastrin-34 [194], as well as CCK-33 [196], and CCK-39 [194] have been synthesized by Fmoc-based solid-phase synthesis on 2-chlorotrityl resin with Fmoc-Tyr(SO_3Na)-OH [194] or Fmoc-Tyr(SO_3H)-OPfp [196] as the tyrosine-O-sulfate building blocks. Final cleavage/deprotection with TFA is possible; this does not cause destruction of the O-sulfate group if the reaction mixture temperature is kept low, and no sulfur-containing scavengers are added [194]. Additional examples for the synthesis of sulfated peptides are the α -conotoxins (EpI, PnIA and PnIB) [197] and N-terminal peptides of the chemokinin receptors CCR5 [198] and CXCR4 [199]. Glycosulfopeptides representing a region of P-selectin glycoprotein ligand-1 (PSGL-1) have been synthesized to explore the roles of individual sulfated tyrosine residues as well as the placement of the glycosyl moiety relative to it [200].

6.7

Review Questions

- Q6.1. What limits the cyclization of smaller peptides (tri-, tetrapeptides)?
- Q6.2. How can the cyclization of tri-, tetrapeptides be achieved?
- Q6.3. Why can ring contraction facilitate the cyclization of small peptides? How could that be achieved?
- Q6.4. Name the possible topologies for peptide cyclization.
- Q6.5. What are possible side reactions upon backbone cyclization?
- Q6.6. What is *spatial screening*?
- Q6.7. Why are cyclic peptides important in drug design?
- Q6.8. Name the motifs of disulfide bridges in peptides.
- Q6.9. How can disulfide bond formation be achieved?
- Q6.10. What are the basic motifs for the linkage between peptide and saccharide moieties in glycoproteins?
- Q6.11. How can the β -N-glycosidic bond between N-acetylglucosamine and asparagine be formed?
- Q6.12. How can the α -O-glycosidic bond between N-acetylgalactosamine and serine or threonine be formed?
- Q6.13. How can the β -O-glycosidic bond between N-acetylglucosamine and asparagine be formed?
- Q6.14. Describe an example of a chemo-enzymatic approach towards glycopeptide synthesis.
- Q6.15. What is the advantage of allyl-type protection and linker chemistry in glycopeptide synthesis?
- Q6.16. What is the major problem with obtaining peptides phosphorylated or sulfated at serine or threonine? How can it be avoided?
- Q6.17. Describe the chemical sensitivity of the different lipid-modified peptides.

References

- 1 S.A. Kates, N.A. Sol, F. Albericio, G. Barany, in: *Peptides: Design, Synthesis, and Biological Activity*, C. Basava, G.M. Anantharamaiah (Eds.), Birkhäuser, Boston, 1994, p. 39.
- 2 Y. Hamada, T. Shioiri, *Chem. Rev.* **2005**, 105, 4441.
- 3 E.A. Felnagle, E.E. Jackson, Y.A. Chan, A.M. Podevels, A.D. Berti, M.D. McMahon, M.G. Thomas, *Mol. Pharm.* **2008**, 5, 191.
- 4 B. Thern, J. Rudolph, G. Jung, *Tetrahedron Lett.* **2002**, 43, 5013.
- 5 R. Jerala, *Expert Opin. Investig. Drugs* **2007**, 16, 1159.
- 6 C. Chatterjee, M. Paul, L. Xie, W.A. van der Donk *Chem. Rev.* **2005**, 105, 633.
- 7 J.M. Willey, W.A. van der Donk *Annu. Rev. Microbiol.* **2007**, 61, 477.
- 8 R. Kasher, D.A. Oren, Y. Barda, C. Gilon, *J. Mol. Biol.* **1999**, 292, 421.
- 9 S. Hess, O. Ovadia, D.E. Shalev, H. Senderovich, B. Qadri, T. Yehezkel, Y. Salitra, T. Sheynis, R. Jelinek, C. Gilon, A. Hoffman, *J. Med. Chem.* **2007**, 50, 6201.
- 10 N. Qvit, H. Reuveni, S. Gazal, A. Zundevich, G. Blum, M.Y. Niv, A. Feldstein, S. Meushar, D.E. Shalev, A.

- Friedler, C. Gilon, *J. Comb. Chem.* **2008**, *10*, 256.
- 11 H. Kessler, *Angew. Chem. Int. Ed.* **1982**, *21*, 512.
- 12 E. Biron, J. Chatterjee, O. Ovadia, D. Langenegger, J. Bruegggen, D. Hoyer, H.A. Schmid, R. Jelinek, C. Gilon, A. Hoffman, H. Kessler, *Angew. Chem. Int. Ed.* **2008**, *47*, 2595.
- 13 R. Haubner, D. Finsinger, H. Kessler, *Angew. Chem. Int. Ed.* **1997**, *36*, 1374.
- 14 T. Weide, A. Modlinger, H. Kessler, *Top. Curr. Chem.* **2007**, *272*, 1.
- 15 J. Boer, D. Gottschling, A. Schuster, M. Semmrich, B. Holzmann, H. Kessler, *J. Med. Chem.* **2001**, *44*, 2586.
- 16 G.C. Tucker, *Curr. Oncol. Rep.* **2006**, *8*, 96.
- 17 D. Bächle, G. Loers, E.W. Guthöhrlein, M. Schachner, N. Sewald, *Angew. Chem. Int. Ed.* **2006**, *45*, 6582.
- 18 F. Schumann, A. Müller, M. Kopsch, G. Müller, N. Sewald, *J. Am. Chem. Soc.* **2000**, *122*, 12009.
- 19 S. Urman, K. Gaus, Y. Yang, U. Strijowski, S. De Pol, O. Reiser, N. Sewald, *Angew. Chem. Int. Ed.* **2007**, *46*, 3976.
- 20 D. Heckmann, H. Kessler, *Methods Enzymol.* **2007**, *426*, 463.
- 21 G.A. Sulyok, C. Gibson, S.L. Goodman, G. Hölzemann, M. Wiesner, H. Kessler, *J. Med. Chem.* **2001**, *44*, 1938.
- 22 D. Heckmann, A. Meyer, L. Marinelli, G. Zahn, R. Stragies, H. Kessler, *Angew. Chem. Int. Ed.* **2007**, *46*, 3571.
- 23 S. Fung, V.J. Hruby, *Curr. Opin. Chem. Biol.* **2005**, *9*, 352.
- 24 D.A. Horton, G.T. Bourne, M.L. Smythe, *J. Comput. Aided Mol. Des.* **2002**, *16*, 415.
- 25 A. Meyer, J. Auernheimer, A. Modlinger, H. Kessler, *Curr. Pharm. Des.* **2006**, *12*, 2723.
- 26 J.N. Lambert, J.P. Mitchell, K.D. Roberts, *J. Chem. Soc., Perkin Trans.* **2001**, *1*, 471.
- 27 J.S. Davies, *J. Peptide Sci.* **2003**, *9*, 471.
- 28 J.M. Humphrey, A.R. Chamberlin, *Chem. Rev.* **1997**, *97*, 2243.
- 29 J.A. Robinson, *Acc. Chem. Res.*, **2008**, *41*, 1278.
- 30 D.J. Craik, M. Čemažar, N.L. Daly, *Curr. Opin. Drug Discov. Dev.* **2007**, *10*, 176.
- 31 H.M. Müller, O. Delgado, T. Bach, *Angew. Chem. Int. Ed.* **2007**, *46*, 4771.
- 32 N. Izumiya, T. Kato, M. Waki, *Biopolymers* **1981**, *20*, 1785.
- 33 R. Schmidt, K. Neubert, *Int. J. Peptide Protein Res.* **1991**, *37*, 502.
- 34 U. Schmidt, A. Lieberknecht, H. Griesser, J. Talbiersky, *J. Org. Chem.* **1982**, *47*, 3261.
- 35 U. Schmidt, M. Kroner, H. Griesser, *Tetrahedron Lett.* **1988**, *29*, 3057.
- 36 U. Schmidt, A. Lieberknecht, H. Griesser, F. Bartkowiak, *Angew. Chem. Int. Ed.* **1984**, *23*, 318.
- 37 N.L. Benoiton, Y.C. Lee, R. Steinauer, F.M.F. Chen, *Int. J. Peptide Protein Res.* **1992**, *40*, 559.
- 38 A. Ehrlich, H.-U. Heyne, R. Winter, M. Beyermann, H. Haber, L.A. Carpino, M. Bienert, *J. Org. Chem.* **1996**, *61*, 8831.
- 39 T. Shioiri, K. Ninomiya, S. Yamada, *J. Am. Chem. Soc.* **1972**, *94*, 6203.
- 40 M. Malešević, U. Strijowski, D. Bächle, N. Sewald, *J. Biotechnol.* **2004**, *112*, 73.
- 41 J. Klose, A. El-Faham, P. Henklein, L.A. Carpino, M. Bienert, *Tetrahedron Lett.* **1999**, *40*, 2045.
- 42 F. Albericio, J.M. Bofill, A. El-Faham, S.A. Kates, *J. Org. Chem.* **1998**, *63*, 9678.
- 43 G. Barany, R.B. Merrifield, in: *The Peptides: Analysis, Synthesis, Biology*, Volume 2, E. Gross, J. Meienhofer (Eds.), Academic Press, New York, **1980**, p. 1.
- 44 L.T. Scott, J. Rebek, L. Ovsyanko, L. Sims, *J. Am. Chem. Soc.* **1977**, *99*, 626.
- 45 S. Mazur, P. Jayalekshmy, *J. Am. Chem. Soc.* **1979**, *101*, 677.
- 46 A. Müller, F. Schumann, M. Kopsch, N. Sewald, *Lett. Peptide Sci.* **1997**, *4*, 275.
- 47 J.S. McMurray, *Tetrahedron Lett.* **1991**, *32*, 7679.

- 48 S.A. Kates, N.A. Solé, C.R. Johnson, D. Hudson, G. Barany, F. Albericio, *Tetrahedron Lett.* **1993**, 34, 1549.
- 49 M. Cudic, J.D. Wade, L. Otvos Jr., *Tetrahedron Lett.* **2000**, 41, 4527.
- 50 P. Romanovskis, A.F. Spatola, *J. Peptide Res.* **1998**, 52, 356.
- 51 J. Alsina, F. Rabanal, E. Giralt, F. Albericio, *Tetrahedron Lett.* **1994**, 35, 9633.
- 52 J. Alsina, C. Chiva, M. Ortiz, F. Rabanal, E. Giralt, F. Albericio, *Tetrahedron Lett.* **1997**, 38, 883.
- 53 G. Sabatino, M. Chelli, S. Mazzucco, M. Gianneschi, A.M. Papini, *Tetrahedron Lett.* **1999**, 40, 809.
- 54 S.S. Isied, C.G. Kuehn, J.M. Lyon, R.B. Merrifield, *J. Am. Chem. Soc.* **1982**, 104, 2632.
- 55 A. Trzeciak, W. Bannwarth, *Tetrahedron Lett.* **1992**, 33, 4557.
- 56 J. Alsina, K.J. Jensen, F. Albericio, G. Barany, *Chem. Eur. J.* **1999**, 5, 2787.
- 57 G. Osapay, A. Profit, J.W. Taylor, *Tetrahedron Lett.* **1990**, 31, 6121.
- 58 M. Xu, N. Nishino, H. Mihara, T. Fujimoto, N. Izumiya, *Chem. Lett.* **1992**, 191.
- 59 L.S. Richter, J.Y.K. Tom, J.P. Brunier, *Tetrahedron Lett.* **1994**, 35, 5547.
- 60 N. Nishino, M. Xu, H. Mihara, T. Fujimoto, Y. Ueno, H. Kumagai, *Tetrahedron Lett.* **1992**, 33, 1479.
- 61 N. Nishino, M. Xu, H. Mihara, T. Fujimoto, *Tetrahedron Lett.* **1992**, 33, 1479.
- 62 A.K. Szardenings, T.S. Burkoth, H.H. Lu, D.W. Tien, D.A. Campbell, *Tetrahedron Lett.* **1997**, 38, 6573.
- 63 C.J. Dinsmore, D.C. Bashore, *Tetrahedron Lett.* **2002**, 43, 3297.
- 64 P.M. Fischer, *J. Peptide Sci.* **2003**, 9, 9.
- 65 M. Rothe, J. Haas, in: *Peptides 1990*, E. Giralt, D. Andreu (Eds.), Escom, Leiden, **1991**, p. 212.
- 66 T. Rückle, P. De Lavallaz, M. Keller, P. Dumy, M. Mutter, *Tetrahedron Lett.* **1999**, 40, 11281.
- 67 K. Gademann, D. Seebach, *Helv. Chim. Acta*, **2001**, 84, 2924.
- 68 D. Besser, R. Olender, R. Rosenfeld, O. Arad, S. Reissmann, *J. Peptide Res.* **2000**, 56, 337.
- 69 M. El Haddadi, F. Cavelier, E. Vives, A. Azmani, J. Verducci, J. Martinez, *J. Peptide Sci.* **2000**, 6, 560.
- 70 M.P. Glenn, M.J. Kelso, J.D.A. Tyndall, D.P. Fairlie, *J. Am. Chem. Soc.* **2003**, 125, 640.
- 71 M.C. Alcaro, G. Sabatino, J. Uziel, M. Chelli, M. Gianneschi, P. Rovero, A.M. Papini, *J. Peptide Sci.* **2004**, 10, 218.
- 72 J. Klose, M. Bienert, C. Mollenkopf, D. Wehle, C. Zhang, L.A. Carpino, P. Henklein, *J. Chem. Soc., Chem. Commun.* **1999**, 1847.
- 73 H. Kessler, B. Haase, *Int. J. Peptide Protein Res.* **1992**, 39, 36.
- 74 W.D.F. Meutermans, S.W. Golding, G.T. Bourne, L.P. Miranda, M.J. Dooley, P.F. Alewood, M.L. Smythe, *J. Am. Chem. Soc.* **1999**, 121, 9790.
- 75 W.D.F. Meutermans, G.T. Bourne, S.W. Golding, D.A. Horton, M.R. Campitelli, D. Craik, M. Scanlon, M.L. Smythe, *Org. Lett.* **2003**, 5, 2711.
- 76 B. Thern, J. Rudolph, G. Jung, *Angew. Chem. Int. Ed.* **2002**, 41, 2307.
- 77 E. Falb, Y. Yechezkel, Y. Salitra, C. Gilon, *J. Peptide Res.* **1999**, 53, 507.
- 78 L. Yang, G. Morriello, *Tetrahedron Lett.* **1999**, 40, 8197.
- 79 L. Zhang, J.P. Tam, *J. Am. Chem. Soc.* **1997**, 119, 2363.
- 80 Y. Shao, W. Lu, S.B.H. Kent, *Tetrahedron Lett.* **1998**, 39, 3911.
- 81 M.A. Marahiel, T. Stachelhaus, H.D. Mootz, *Chem. Rev.* **1997**, 97, 2651.
- 82 J.W. Tranger, R.M. Kohli, H.D. Mootz, M.A. Marahiel, C.T. Walsh, *Nature* **2000**, 407, 215.
- 83 C.P. Scott, E. Abel-Santos, M. Wall, D.C. Wahnnon, S.J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1999**, 96, 13638.
- 84 A. Tavassoli, S.J. Benkovic, *Angew. Chem. Int. Ed.* **2005**, 44, 2760.
- 85 A. Tavassoli, S.J. Benkovic, *Nature Protoc.* **2007**, 2, 1126.

- 86 R. Kleineweischede, C.P.R. Hackenberger, *Angew. Chem. Int. Ed.* **2008**, 47, 5984.
- 87 G. Gellerman, A. Elgavi, Y. Salitra, M. Kramer, *J. Peptide Res.* **2001**, 57, 277.
- 88 J.N. Lambert, J.P. Mitchell, K.D. Roberts, *J. Chem. Soc., Perkin Trans. I* **2001** 471.
- 89 J.F. Reichwein, C. Versluis, R. Liskamp, *J. Org. Chem.* **2000**, 65, 6187.
- 90 J. Pernerstorfer, M. Schuster, S. Blechert, *J. Chem. Soc., Chem. Commun.* **1997**, 1949.
- 91 V.D. Bock, R. Perciaccante, T.P. Jansen, H. Hiemstra, J.H. van Maarseveen, *Org. Lett.* **2006**, 8, 919.
- 92 M. Lebl, V.J. Hruby, *Tetrahedron Lett.* **1984** 25, 2067.
- 93 V. Cavallaro, P. Thompson, M. Hearn, *J. Peptide Sci.* **1998**, 4, 335.
- 94 S. Plane, *Int. J. Peptide Protein Res.* **1990**, 35, 510.
- 95 F.A. Robey, R.L. Fields, *Anal. Biochem.* **1989**, 177, 373.
- 96 K.D. Roberts, J.N. Lambert, N.J. Ede, A.M. Bray, *Tetrahedron Lett.* **1998**, 39, 8357.
- 97 R.A. Turner, A.G. Oliver, R.S. Lokey, *Org. Lett.* **2007**, 9, 5011.
- 98 L. Yu, Y. Lai, J.V. Wade, S.M. Coutts, *Tetrahedron Lett.* **1998**, 39, 6633.
- 99 P.W. Schiller, T.M.-D. Nguyen, J. Miller, *Int. J. Peptide Protein Res.* **1985**, 31, 231.
- 100 P. Grieco, P.M. Gitu, V.J. Hruby, *J. Peptide Res.* **2001**, 57, 250.
- 101 P.W. Schiller, T.M.-D. Nguyen, J. Miller, *Int. J. Peptide Protein Res.* **1985**, 25, 171.
- 102 A.M. Felix, C.-T. Wang, E.P. Heimer, A. Fournier, *Int. J. Peptide Protein Res.* **1988**, 31, 231.
- 103 P. Grieco, P.M. Gitu, V.J. Hruby, *J. Peptide Sci.* **2001**, 57, 250.
- 104 V.J. Hruby, R.S. Agnes, *Biopolymers* **1999**, 51, 391.
- 105 G. Han, C. Haskell-Luevano, L. Kendall, G. Bonner, M.E. Hadley, R.D. Cone, V.J. Hruby, *J. Med. Chem.* **2004**, 47, 1514.
- 106 V.J. Hruby, M. Cai, J.P. Cain, A.V. Mayorov, M.M. Dedek, D. Trivedi, *Curr. Top. Med. Chem.* **2007**, 7, 1085.
- 107 J.A. McCubbin, M.L. Maddess, M. Lautens, *Org. Lett.* **2006**, 8, 2993.
- 108 G. Fridkin, C. Gilon, *J. Peptide Res.* **2002**, 60, 104.
- 109 M. Roice, I. Johannsen, M. Meldal, *QSAR Comb. Sci.* **2004**, 23, 662.
- 110 Y.L. Angell, K. Burgess, *Chem. Soc. Rev.* **2007**, 36, 1674.
- 111 S. Cantel, A. Le Chevalier Isaad, M. Scrima, J.J. Levy, R.D. Dimarchi, P. Rovero, J.A. Halperin, A.M. D'Ursi, A.M. Papini, M. Chorev, *J. Org. Chem.* **2008**, 73, 5663.
- 112 I. Annis, B. Hargittai, G. Barany, *Methods Enzymol.* **1997**, 289, 198.
- 113 K. Akaji, Y. Kiso, in: *Houben-Weyl, Methods of Organic Chemistry*, Vol. E22b, *Synthesis of Peptides and Peptidomimetics*, M. Goodman, A. Felix, L. Moroder, C. Toniolo (Eds.), Georg Thieme Verlag, Stuttgart, **2002**, p. 101.
- 114 L. Moroder, H.-J. Musiol, M. Götz, C. Renner, *Biopolymers* **2005**, 80, 85.
- 115 C. Boulègue, H.-J. Musiol, V. Prasad, L. Moroder, *Chem. Today* **2006**, 24, 24.
- 116 L. Moroder, H.-J. Musiol, N. Schaschke, L. Chen, B. Hargittai, G. Barany, in: *Houben-Weyl, Methods of Organic Chemistry*, Vol. E22b, *Synthesis of Peptides and Peptidomimetics*, M. Goodman, A. Felix, L. Moroder, C. Toniolo (Eds.), Georg Thieme Verlag, Stuttgart, **2002**, p. 384.
- 117 P. Sieber, B. Kamber, A. Hartmann, A. Jöhl, B. Riniker, W. Rittel, *Helv. Chim. Acta* **1974**, 57, 2617.
- 118 R.A.D. Bathgate, F. Lin, N.F. Hanson, L. Otvos, A. Guidolin, C. Giannakis, S. Bastiras, S.L. Layfield, T. Ferraro, S. Ma, C.X. Zhao, A.L. Gundlach, C.S. Samuel, G.W. Tregear, J.D. Wade, *Biochemistry* **2006**, 45, 1043.
- 119 G. Bulaj, B.M. Olivera, *Antioxidants Redox Signaling* **2008**, 10, 141.
- 120 M. Čemažar, C.W. Gruber, D.J. Craik, *Antioxidants Redox Signaling* **2008**, 10, 103.

- 121 C. Boulègue, A.G. Milbradt, C. Renner, L. Moroder, *J. Mol. Biol.* **2006**, 358, 846.
- 122 E. Pokidysheva, A.G. Milbradt, S. Meier, C. Renner, D. Haussinger, H.P. Bächinger, L. Moroder, S. Grzesiek, T.W. Holstein, S. Özbek, J. Engel, *J. Biol. Chem.* **2004**, 279, 30395.
- 123 J.G. Tang, C.L. Tsou, *Biochem. J.* **1990**, 268, 429.
- 124 J.G. Tang, C.C. Wang, C.L. Tsou, *Biochem. J.* **1988**, 255, 451.
- 125 Z.-Y. Guo, Z.-S. Qiao, Y.-M. Feng, *Antioxidants Redox Signaling* **2008**, 10, 127.
- 126 C. Boulègue, H.-J. Musiol, C. Renner, L. Moroder, *Antioxidants Redox Signaling* **2008**, 10, 113.
- 127 J. Ottl, L. Moroder, *Tetrahedron Lett.* **1999**, 40, 1487.
- 128 J. Ottl, L. Moroder, *J. Am. Chem. Soc.* **1999**, 121, 653.
- 129 D. Besse, N. Budisa, W. Karnbrock, C. Minks, H.-J. Musiol, S. Pegoraro, F. Siedler, E. Weyher, L. Moroder, *Biol. Chem.* **1997**, 378, 211.
- 130 D. Besse, F. Siedler, T. Diercks, H. Kessler, L. Moroder, *Angew. Chem. Int. Ed.* **1997**, 36, 883.
- 131 K.M. Koeller, C.-H. Wong, *Chem. Rev.* **2000**, 100, 4465.
- 132 B.D. Livingston, E.M.D. Robertis, J.C. Paulson, *Glycobiology* **1990**, 1, 39.
- 133 O. Seitz, I. Heinemann, A. Mattes, H. Waldmann, *Tetrahedron* **2001**, 57, 2247.
- 134 M. Meldal, P.M. St Hilaire, *Curr. Opin. Chem. Biol.* **1997**, 1, 552.
- 135 J.B. Jones, *Synlett* **1999**, 1495.
- 136 C.M. Taylor, *Tetrahedron* **1998**, 54, 11317.
- 137 O. Seitz, *ChemBiochem.* **2001**, 1, 214.
- 138 T. Bielfeldt, S. Peters, M. Meldal, K. Bock, H. Paulsen, *Angew. Chem. Int. Ed.* **1992**, 31, 857.
- 139 F.E. McDonald, S.J. Danihefsky, *J. Org. Chem.* **1992**, 57, 7001.
- 140 S.T. Cohen-Anisfeld, D.T. Langbury, Jr., *J. Am. Chem. Soc.* **1993**, 115, 10531.
- 141 H. Kunz, *Pure Appl. Chem.* **1993**, 65, 1223.
- 142 A.L. Handlon, B. Fraser-Reid, *J. Am. Chem. Soc.* **1993**, 115, 3796.
- 143 P. Sjölin, J. Kihlberg, *J. Org. Chem.* **2001**, 66, 2957.
- 144 C. Unverzagt, *Angew. Chem. Int. Ed.* **1996**, 35, 2350.
- 145 O. Seitz, C.-H. Wong, *J. Am. Chem. Soc.* **1997**, 119, 8766.
- 146 T. Pol, H. Waldmann, *J. Am. Chem. Soc.* **1997**, 119, 6702.
- 147 H. Vegad, C.J. Gray, P.J. Somers, A.S. Dutta, *J. Chem. Soc., Perkin Trans. I* **1997**, 1429.
- 148 K. von dem Bruch, H. Kunz, *Angew. Chem. Int. Ed.* **1994**, 33, 101.
- 149 C.-H. Wong, M. Schuster, P. Wang, P. Sears, *J. Am. Chem. Soc.* **1993**, 115, 5893.
- 150 H. Hojo, Y. Nakahara, *Biopolymers (Peptide Sci.)* **2007**, 88, 308.
- 151 H. Kunz, C. Unverzagt, *Angew. Chem. Int. Ed.* **1984**, 23, 436.
- 152 W. Kosch, J. März, H. Kunz, *React. Polym.* **1994**, 22, 181.
- 153 O. Seitz, H. Kunz, *Angew. Chem. Int. Ed.* **1995**, 34, 803.
- 154 H. Herzner, T. Reipen, M. Schultz, H. Kunz, *Chem. Rev.* **2000**, 100, 4495.
- 155 M. Schuster, P. Wang, J.C. Paulson, C.-H. Wong, *J. Am. Chem. Soc.* **1994**, 116, 1135.
- 156 C.-H. Wong, R.L. Halcomb, Y. Ichikawa, T. Kajimoto, *Angew. Chem. Int. Ed.* **1995**, 34, 521.
- 157 S.D. Rosen, C.R. Bertozzi, *Curr. Opin. Cell Biol.* **1994**, 6, 663.
- 158 Y. Shin, K.A. Winans, B.J. Backes, S.B.H. Kent, J.A. Ellman, C.R. Bertozzi, *J. Am. Chem. Soc.* **1999**, 121, 11684.
- 159 D. Macmillan, C.R. Bertozzi, *Angew. Chem. Int. Ed.* **2004**, 43, 1355.
- 160 R. Ingenito, E. Bianchi, D. Fattori, A. Pessi, *J. Am. Chem. Soc.* **1999**, 121, 11369.
- 161 S. Mezzato, M. Schaffrath, C. Unverzagt, *Angew. Chem. Int. Ed.* **2005**, 44, 1650.
- 162 A. Brik, C.-H. Wong, *Chem. Eur. J.* **2007**, 13, 5670.

- 163 J.S. McMurray, D.R. Coleman, IV, W. Wang, M.L. Campbell, *Biopolymers* **2001**, 60, 3.
- 164 H.-G. Chao, B. Leiting, P.D. Reiss, A.L. Burkhardt, C.E. Klimas, J.B. Bolen, G.R. Matsueda, *J. Org. Chem.* **1995**, 60, 7710.
- 165 T. Wakamiya, K. Saruta, J. Yasuoka, S. Kusunoto, *Chem. Lett.* **1994**, 1099.
- 166 T. Vorherr, W. Bannwarth, *Bioorg. Med. Chem. Lett.* **1995**, 5, 2661.
- 167 E.A. Ottinger, Q. Xu, G. Barany, *Peptide Res.* **1996**, 9, 223.
- 168 J.W. Perich, N.J. Ede, S. Eagle, A.M. Bray, *Lett. Peptide Sci.* **1999**, 6, 91.
- 169 A. Otaka, K. Miyoshi, M. Kaneko, H. Tamamura, N. Fujii, M. Nomizu, T.R. Burke, P.P. Roller, *J. Org. Chem.* **1995**, 60, 3967.
- 170 T.R. Bruke, Jr., Z.-J. Yao, D.-G. Liu, J. Voigt, Y. Gao, *Biopolymers* **2001**, 60, 32.
- 171 R. Jerala, *Expert Opin. Investig. Drugs* **2007**, 16, 1159.
- 172 K.C. Nicolaou, C.N.C. Boddy, S. Bräse, N. Winssinger, *Angew. Chem. Int. Ed.* **1999** 38 2096.
- 173 R.H. Baltz, V. Miao, S.K. Wrigley, *Nat. Prod. Rep.* **2005**, 22, 717.
- 174 K.H. Wiesmüller, W.G. Bessler, G. Jung, *Int. J. Peptide Protein Res.* **1992**, 40, 255.
- 175 K.H. Wiesmüller, B. Fleckenstein, G. Jung, *Biol. Chem.* **2001**, 382, 571.
- 176 P.M. Moyle, I. Toth, *Curr. Med. Chem.* **2008**, 15, 506.
- 177 P.J. Casey, *Science* **1995**, 268, 221.
- 178 G. Milligan, M. Parenti, A.I. Magee, *Trends Biochem. Sci.* **1995**, 20, 181.
- 179 S. Moffet, B. Mouillac, H. Bonin, M. Bouvier, *EMBO J.* **1993**, 12, 349.
- 180 H. Waldmann, M. Schelhaas, E. Nägele, J. Kuhlmann, A. Wittinghofer, H. Schröder, J.R. Silvius, *Angew. Chem. Int. Ed.* **1997**, 36, 2238.
- 181 H. Schröder, R. Leventis, S. Rex, M. Schelhaas, E. Nägele, H. Waldmann, J.R. Silvius, *Biochemistry* **1997**, 36, 13102.
- 182 H. Schröder, R. Leventis, S. Shahinian, P.A. Walton, J.R. Silvius, *J. Cell Biol.* **1996**, 134, 647.
- 183 D. Kadereit, P. Deck, I. Heinemann, H. Waldmann, *Chem. Eur. J.* **2001**, 7, 1184.
- 184 L. Brunsveld, J. Kuhlmann, H. Waldmann, *Methods* **2006**, 40, 151.
- 185 T. Schmittberger, A. Cott, H. Waldmann, *J. Chem. Soc., Chem. Commun.* **1998**, 937.
- 186 L. Brunsveld, J. Kuhlmann, H. Waldmann, *Methods* **2006**, 40, 151.
- 187 L. Brunsveld, J. Kuhlmann, K. Alexandrov Alfred Wittinghofer, R.S. Goody, H. Waldmann, *Angew. Chem. Int. Ed.* **2006**, 45, 6622.
- 188 H.-J. Musiol, A. Escherich, L. Moroder, Synthesis of sulfated Peptides, in: *Houben-Weyl, Methods of Organic Chemistry*, Vol. E22b, *Synthesis of Peptides, and Peptidomimetics*, M. Goodman, A. Felix, L. Moroder, C. Toniolo (Eds.), Georg Thieme Verlag, Stuttgart, **2002**, p. 425.
- 189 C. Seibert, T.P. Sakmar, *Biopolymers* **2008**, 90, 459.
- 190 E. Wünsch, L. Moroder, L. Wilschowitz, W. Göhring, R. Scharf, J.D. Gardner, *Hoppe Seylers Z. Physiol. Chem.* **1981**, 362, 143.
- 191 L. Moroder, L. Wilschowitz, M. Gemeiner, W. Göhring, S. Knof, R. Scharf, P. Thamm, J.D. Gardner, T.E. Solomon, E. Wünsch, *Hoppe Seylers Z. Physiol. Chem.* **1981**, 362, 929.
- 192 T. Yagami, K. Kitagawa, C. Aida, H. Fujiwara, S. Futaki, *J. Peptide Res.* **2000**, 56, 239.
- 193 B. Penke, F. Hajnal, J. Lonovics, G. Holzinger, T. Kadar, G. Telegdy, J. Rivier, *J. Med. Chem.* **1984**, 27, 845.
- 194 K. Kitagawa, C. Aida, H. Fujiwara, T. Yagami, S. Futaki, M. Kogire, J. Ida, K. Inoue, *J. Org. Chem.* **2001**, 66, 1.
- 195 S. De Luca G. Morelli, *J. Peptide Sci.* **2004**, 10, 265.
- 196 B. Penke, L. Nyerges, *Peptide Res.* **1991**, 4, 289.

- 197** Y. Sekigawa, K. Kitagawa, T. Koide, E. Tachikawa, in: *Peptide Science 2004*, (Y. Shimohigashi, Ed.) The Japanese Peptide Society, Osaka, **2005**, p. 611.
- 198** L. Duma, D. Häussinger, M. Rogowski, P. Lusso, S. Grzesiek, *J. Mol. Biol.* **2007**, 365, 1063.
- 199** T. Inui, H. Sargsyan, P. Cano, I. Ayzenshtat, B. Arshava, J. Anglister, F. Naider, in: *Peptide Science 2005*, (T. Wakamiya, Ed.) The Japanese Peptide Society, Osaka, **2006**, p. 23.
- 200** A. Leppänen, S.P. White, J. Helin, R.P. McEver, R.D. Cummings, *J. Biol. Chem.* **1999**, 275, 39569.

7

Peptide and Protein Design, Pseudopeptides, and Peptidomimetics

Native peptides can be directly applied as pharmacologically active compounds to only a very limited extent. The major disadvantages of the application of a peptide in a biological system – for example, rapid degradation by proteases, hepatic clearance, undesired side effects by interaction of conformationally flexible peptides with different receptors [1–7], and low membrane permeability due to their hydrophilic character – are in most cases detrimental to oral application. Furthermore, most peptides are not able to pass body barriers such as the blood–brain barrier, and also suffer from rapid processing and excretion. The high structural flexibility, especially of linear peptides, has the consequence that the receptor-bound, biologically active conformation represents only one member of a large ensemble of conformers.

The “Lipinski rules of five” [8], also often referred to as the “Pfizer rules of five” provide a rough estimate of whether a compound with a certain pharmacological or biological activity has properties that render it suitable for application as an orally active drug in humans. According to this rule a compound is not suitable as a drug if

- there are more than five hydrogen bond donors (sum of OH and NH),
- the molecular mass exceeds 500 Da,
- the log *P* (octanol/water partition coefficient, a measure of the lipophilicity) is over 5,
- there are more than 10 hydrogen bond acceptors (expressed as the sum of N and O).

All numbers given in the rules are multiples of five, which is the origin of the name. The rules have been established on the basis of physico-chemical parameters. Most peptides do not comply with these rules, which led to the opinion among medicinal chemists that peptides are not good drug candidates. However, as for any rule, there are exceptions. The rule is not applicable to compound classes that are substrates for biological transporters. Moreover, new application strategies do not require oral availability. Peptides are increasingly becoming approved as drugs or are found in the development pipeline of pharmaceutical companies.

Additionally, peptide chemistry may contribute considerably to the drug development process. The interaction of a peptide or a protein epitope with a receptor or an enzyme is the initial event based on molecular recognition [9], and generally elicits a biological response. Likewise, ligand or substrate binding to a protein relies on induced fit and surface complementarity of the ligand/substrate and the protein.

Peptides and peptidomimetics may interfere with protein–protein interaction. Molecular recognition between proteins is often mediated by the recognition of specific, sometimes very short, epitopes presented on surface-exposed loops or hairpins, which may, in many cases, be easily addressed with peptides corresponding to the surface-exposed epitope.

In many other cases protein–protein interaction involves relatively large and shallow contact areas of 7.5–15 nm², often with discontinuous binding epitopes. The development of such inhibitors is especially challenging and often makes use of peptide epitopes attached to scaffolds in order to mimic larger, discontinuous binding epitopes. Compounds able to mimic the structure and function of extended regions of protein surfaces have been proven to modulate protein–protein interactions [10–13]. In this context, the interaction between the tumor suppressor p53 and its natural antagonist MDM2 is a widely studied system with respect to modulating protein–protein interactions [14]. Schepartz *et al.* [15] developed the general methodology of protein grafting, where the functional epitope is attached to a relatively rigid mini-protein.

The problem of interfering with physiological or pathological events that are associated with molecular recognition of discontinuous epitopes has been addressed by using scaffolded and assembled peptides. Such synthetic mimetics have proven to be promising tools for the modulation of protein–ligand interactions [16]. Synthetic mimetics corresponding to a discontinuous protein epitope involved in protein–protein interaction have been designed on a rational basis and synthesized. Three peptides stemming from HIV-1 gp120f with the sequences I⁴²⁴NMWQEVG-KA⁴³³, S³⁶⁵GGDPEIVT³⁷³, and L⁴⁵⁴TRDGGN⁴⁶⁰ were grafted synthetically to different scaffolds and shown to functionally imitate the CD4 binding site of HIV-1 gp120 [17].

Chemically modified peptides with improved bioavailability and metabolic stability may be used directly as drugs (cf. Chapter 9). Alternatively, peptidic receptor ligands or enzyme substrates (peptide leads) may serve as starting points for the development of nonpeptide drugs [18]. The first step of any such investigation is to identify the amino acid side-chain residues responsible for receptor–ligand interaction. Subsequently, the topography of these functional (pharmacophoric) groups is reproduced by similar nonpeptidic functionalities on a rigid scaffold [19]. Many efforts have been made to develop peptide-based, pharmacologically active compounds, including peptide modification and the design of peptidomimetics. Whilst modified peptides (by definition) contain nonproteinogenic or modified amino acid building blocks, peptidomimetics (cf. Section 7.3) are nonpeptidic compounds that imitate the structure of a peptide in its receptor-bound conformation and – in the case of agonists – also the biological mode of action on the receptor level.

According to the definition by Ripka and Rich [20], three different types of peptidomimetics may be distinguished:

- Type I: Peptides modified by amide bond isosteres (cf. Section 7.2.2) and secondary structure mimetics (cf. Section 7.2.4). These derivatives are usually designed to closely match the peptide backbone.

- Type II: Small nonpeptide molecules that bind to a receptor or enzyme (functional mimetics, cf. Section 7.3). However, despite often being presumed to serve as structural analogues of native peptide ligands, these nonpeptide antagonists often bind to a different receptor subsite and, hence, do not necessarily mimic the parent peptide.
- Type III: Compounds to be regarded as ideal mimetics, because they are nonpeptide compounds and contain the functional groups necessary for the interaction of the native peptide with the corresponding protein (pharmacophoric groups) grafted onto a rigid scaffold.

The design of all three types of peptidomimetics may be assisted by X-ray crystallographic or nuclear magnetic resonance (NMR) data [21], computational *de novo* design (“*in silico* screening”) [22], and combinatorial chemistry (cf. Chapter 8) [23].

Furthermore, the construction of novel bioactive polypeptides and artificial protein structures or miniaturized enzymes currently forms the focal point of intensive investigations [24]. Although the secondary structure elements (cf. Section 2.5) of a polypeptide chain with known primary structure can be predicted to a limited extent, our knowledge of protein folding is still incomplete. Therefore, reliable predictions of the composition of secondary structure elements to produce a three-dimensionally folded biologically active peptide or protein are, at present, beyond reach.

7.1

Peptide Design

The development of a drug (peptide drug) may start with the identification and design of the peptide sequence essential for biological activity. Such structure–activity relationships, for example, rely in the first instance on a systematic substitution of each amino acid residue of a native peptide by a simple amino acid such as alanine (“alanine scan”). Studies of the receptor selectivity or of the agonism or antagonism of a peptide hormone require structure–activity relationships to be ascertained, using many synthetic peptide analogues. In the past, this approach was highly labor-intensive, but today it is much more straightforward following the application of combinatorial chemistry and automation (cf. Chapter 8).

A viable design procedure is outlined in Figure 7.1. A peptide sequence **1** is identified as being responsible for interaction with a receptor. In the example shown, the peptide contains the sequence Arg-Gly-Asp (RGD) that is known as a universal recognition motif for cell–cell and cell–matrix interactions. Some peptides which contain the RGD sequence in a well-defined conformation efficiently inhibit the binding of extracellular proteins (e.g., fibrinogen, vitronectin) to cellular receptors (integrins). The corresponding protein–receptor interaction is involved, for example, in blood platelet aggregation, tumor cell adhesion, angiogenesis, and osteoporosis [25]. The peptide conformation is subsequently constrained, e.g. by cyclization (**2**) or by sterically hindered amino acids (type I peptidomimetics). If the biological activity is retained, it is likely that the receptor-bound conformation is still

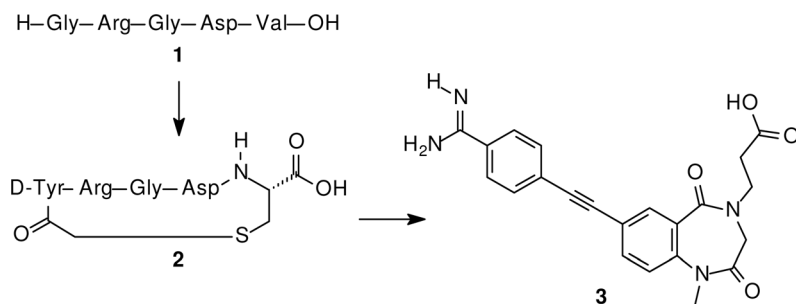


Figure 7.1 Conversion of a peptide sequence into a nonpeptide lead.

accessible [26]. In the next step, the three-dimensional array of the pharmacophoric groups can be deduced and a nonpeptide lead compound **3** (type III peptidomimetic) may be constructed, where the pharmacophoric groups are attached to a rigid scaffold and are presented in the appropriate three-dimensional array [27, 28].

In vitro or *in vivo* biological activity is not sufficient for a drug candidate, however, and many further investigations of toxicology, pharmacokinetics (ADME: absorption, distribution, metabolism, excretion), pharmacology, and clinical trials are required before a drug candidate is approved by national authorities for the treatment of patients. Considerable effort is needed to develop a biologically active compound (lead compound) to a drug candidate, and finally to register it as a pharmacologically active drug that can be marketed. Vast research expense, long development times, and high risks are imposed on innovative drugs (“breakthrough” drugs) where, in contrast to the so-called “me-too” drugs, no other substances with a similar mode of action are being marketed concurrently. In most cases, the development pathway is aimed towards a specific drug, and starts with a pathological phenomenon (biological target). Extensive biochemical and medical knowledge is required to understand the physiological and pathological events involved, and in this respect the recent advances made in biochemistry, molecular biology, gene technology, protein purification, protein crystallography, protein NMR, genomics, proteomics, and computational methods provide the necessary platform for drug development.

Originally, the term “drug design” was coined for the rational design of a molecule that interacts appropriately on a structural or mechanistic basis with the target (enzyme, receptor, or DNA sequence) in a complementary manner, and that ultimately exerts the desired pharmacological effect by activation or blocking of the corresponding target. The complex interactions in drug design are shown schematically in Figure 7.2 [29].

Determination of the three-dimensional (3D) structure of the target protein is a crucial precondition for the rational design of small molecule agonists, antagonists or enzyme inhibitors [30]. In many cases, protein engineering facilitates structure determination because the corresponding protein can be obtained in larger amounts by overexpression (cf. Section 4.6.1). Methods of gene technology may also provide single domains of a protein for structure determination by X-ray crystallography or NMR methods. The 3D structure often results in suggestions for site-directed

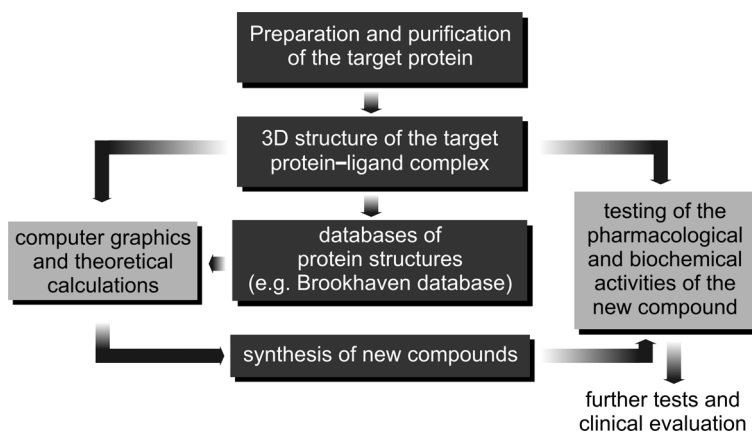


Figure 7.2 Principles of drug design.

mutagenesis, and/or the exchange of single or multiple amino acid residues using molecular biology methods in order, for example, to optimize the stability, substrate specificity, or optimum pH of an enzyme-catalyzed reaction.

Methods of molecular modeling are used to visualize the structures of receptors and peptide drugs on a computer, and then to simulate their interaction. However, as the torsion angles of the peptide backbone and of the side-chain residues may adopt different values, numerous conformations usually result for an unbound peptide ligand that are not biased towards the most efficient conformation with respect to ligand–receptor interaction. Force-field calculations have been developed as a strategy to predict conformations by comparison of their relative potential energy [31]. The single atoms of a protein are considered as hard spheres, and classical mechanics calculations are used to simulate their time-dependent motion. Even solvent molecules may be explicitly included.

Statistical methods (Monte Carlo methods) may also be used to produce a huge number of conformations and to calculate subsequently their potential energy in the local minima [32]. Finally, the conformation with the lowest energy or a family of low-energy conformations is further investigated. This procedure, which initially is based on statistics, can be supplemented by molecular dynamics calculations and may result in an evaluation of the conformational flexibility of molecular structures by force-field methods [31, 33]. In computer-aided molecular design (CAMD) [34], the complementary fitting of a peptide or nonpeptide drug to a receptor plays a crucial role, though sensible application of this concept requires knowledge of the 3D structure of the receptor. During recent years much effort has been given to developing new methods and algorithms (e.g., neuronal network methods, genetic algorithms, machine learning, and graph theoretical methods) in order to predict molecular structures [34–37]. Recent progress in peptide and protein design includes especially new potential functions, more efficient ways of computing energetics, flexible treatments of solvent, and useful energy function

approximations, as well as ensemble-based approaches to scoring designs for inclusion of entropic effects [38].

If the site of action of a biologically active compound is known, then direct computer-aided drug design is used [39]. Tailor-made molecular structures are constructed to optimally fulfil the requirements for binding to the receptor or the active site of an enzyme [40]. In this context, protein–ligand docking is an important method, and many docking programs based on a series of searching algorithms such as surface complementary matching [41, 42], fragment growing [43, 44], random sampling (Monte Carlo), simulated annealing [45–47], and genetic algorithms [48, 49] have been developed.

One NMR-based method has been described to identify and optimize small organic molecules binding to proximal subsites of a single protein in order to produce from these moieties ligands with high affinity to the protein. Di- or multivalent protein–ligand interactions are obtained by covalently linking these pharmacophoric groups. This approach provides experimental structure–activity relationship data, as obtained from NMR [50, 51]. The method relies on ^{15}N -labeled proteins, for instance, because changes in the ^{15}N or ^1H amide chemical shift are observed in ^{15}N heteronuclear single quantum correlation spectra (^{15}N HSQC) upon addition of a ligand to the labeled protein (Figure 7.3).

If no structural data of a target protein are available, the known 3D structure of another, sequence-related protein may be used as the starting point for the so-called homology modeling.

Indirect computer-aided drug design (indirect CADD) is a very useful protocol in the process of rational drug design when the molecular structure of the target is unknown. It allows the evaluation of structural changes upon modification of a lead structure. Systematic substituent variation of a lead structure is correlated with biological data, and these results are used in the identification of crucial structural elements of the ligand that are required for high-affinity receptor binding (receptor mapping). This finally allows the creation of a hypothetical receptor model, and is intensively applied in modern drug design.

The synthesis of chemically modified peptides often employs the incorporation of conformationally constrained amino acids (proline, proline analogues, N-methyl

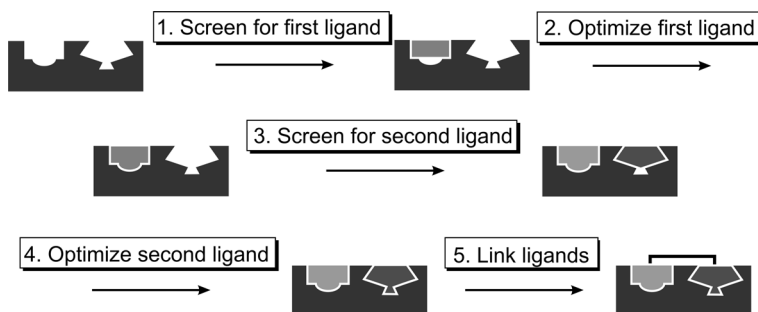


Figure 7.3 Optimization process in establishing structure–activity relationships (SAR) by NMR.

amino acids, C $^{\alpha}$ -dialkyl amino acids, etc.). This type of modification results in rather rigid peptide analogues where the conformational freedom is reduced. A further viable route for the design of peptides with reduced conformational freedom relies on peptide cyclization (cf. Section 6.1).

The design of a peptide depends primarily not only on the targeted application, but also on synthetic considerations [52, 53]. In general, several steps are necessary to obtain a peptide suitable for biochemical or biomedical applications, and several optimization cycles are often required. There is no such thing as a general route for the design, although the procedure is guided by the desired function.

Conformational analysis is a valuable tool in the design of new peptide drugs. Linear peptide chains are usually characterized by high flexibility which renders determination of the 3D structure nontrivial. The basics of conformational analysis were established by Ramachandran *et al.* [54], Scheraga *et al.* [55], and Blout *et al.* [56]. These authors established basic concepts for the classification and description of peptide structures (cf. Section 2.5), and contributed extensively to an improved understanding of peptide and protein conformation and of protein-folding processes. Conformational transitions in polypeptides have been studied using several different methods, for example by Goodman *et al.* [57]. Cyclic peptides offer the advantage of limited conformational freedom, and this makes them particularly suited for conformational studies. Furthermore, many cyclic peptides (Section 6.1) display interesting biological profiles of activity. Peptide conformations can be very efficiently analyzed using NMR spectroscopy (Section 2.5.3), as reviewed in [58–60].

Sequence-based peptide design relies not only on the deletion, exchange, or modification of amino acid-building blocks, but also on structural elements of the peptide backbone. Besides the 22 proteinogenic amino acids, at present more than 1000 nonproteinogenic amino acids are known. Some of these occur naturally in the free form, or are bound in peptides.

The truncation of a sequence by deletion of amino acid building blocks that practically do not contribute to biological activity is an important approach in the design of pharmacologically active peptides starting from longer bioactive peptides. Many peptide hormones and also other bioactive peptides very often contain a minimum sequence responsible for binding to the receptor and the biological activity. Of course, it is sometimes not very easy to identify the essential partial sequence. Often, the spatial arrangement of those residues responsible for the activity depends on the 3D peptide structure where non-neighboring amino acids, with respect to the primary structure, come into close vicinity by peptide folding. In contrast, many peptides are known where the residues essential for the biological activity are localized in well-conserved partial sequences. Messenger segments, responsible for the triggering of physiological events, are located C-terminally in the tachykinins. The C-terminal N $^{\alpha}$ -acetylated 8-peptide sequence Ac-His-Trp-Ala-Val-Gly-His-Leu-Met-NH $_2$ of gastrin-releasing peptide displays higher relative activity compared to the native sequence. Furthermore, this C-terminal sequence is highly homologous to the 14-peptide bombesin isolated from frog skin. Melanin-concentrating hormone (MCH) is a 17-peptide with a disulfide bridge between Cys ^{5}

and Cys¹⁴; only the partial sequence 5–15 is necessary for full biological activity, with Trp¹⁵ being of major importance for interaction with the MCH receptor.

7.2

Modified Peptides

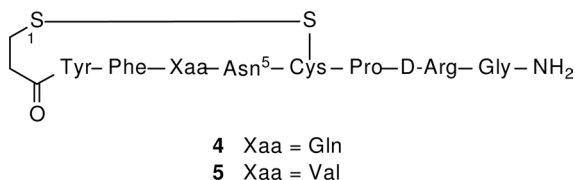
7.2.1

Side-Chain Modification

The most straightforward approach for peptide modification is to introduce changes into the side chains of single amino acids. On this level, a multitude of possibilities for the synthesis of nonproteinogenic amino acids already exists, and useful preparative routes for the asymmetric synthesis of many derivatives have been developed. Among others, the sulfur in cysteine and methionine has been replaced by selenium or tellurium, for example to modify the redox properties of peptides and proteins [61].

This strategy allows the incorporation of amino acids with side chains that do not occur naturally in peptides or proteins, with the aim being to introduce special functional groups, to restrict the conformational flexibility of a peptide, or to enhance its metabolic stability. Furthermore, D-configured amino acids, N^α-alkylated amino acids, C^α-dialkylated amino acids, or α,β-didehydro amino acids may be employed.

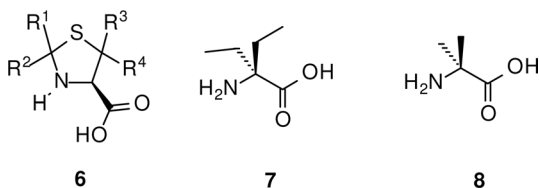
Stereochemical substitutions provide information about the steric requirements in the formation of distinct secondary structures and their influence on peptide–receptor interaction. Furthermore, the incorporation of D-amino acids usually confers higher metabolic stability on the peptide. [1-deamino,D-Arg⁸]Vasopressin (DDAVP) **4** is a modified peptide applied as a selective antidiuretic agent with long-term activity, suitable for the treatment of diabetes insipidus. DDAVP acts as a V₂-receptor agonist in the kidney, with the missing α-amino group increasing proteolytic stability. According to the investigations of Manning *et al.* [62], the exchange of Gln⁴ by Val results in the potent V₂-receptor agonist [Val⁴]vasopressin **5** that simultaneously is a weak V₂-receptor antagonist.



Isosteric and other substitutions similarly provide valuable information on peptide–receptor interactions. Often, methionine is exchanged in this context by norleucine (Nle); indeed, such a substitution in the 14-peptide α-MSH (cf. Section 3.3.3.3) resulted in [Nle⁴]α-MSH displaying two-fold increased activity compared to the native peptide. Oxytocin modification through an exchange of Tyr² by Phe resulted in only weak agonist activity, while [4-MePhe²]oxytocin is an antagonist.

Many substitutions have been performed in gonadoliberein (gonadotropin-releasing hormone, GnRH, cf. Section 3.3.2.2) that can be used either for diagnostic purposes or as a drug against infertility. Interestingly, GnRH superagonist analogues result in receptor down-regulation. Such long-term activity reduces LH and FSH liberation and, in males, this provides therapy for prostate carcinoma.

Conformational and topographical restrictions are particularly suited as manipulations for peptide design targeted towards an increase in receptor selectivity, metabolic stability, and the development of highly potent agonists or antagonists. The incorporation of N-methyl amino acids influences the *cis/trans* ratio of the peptide bond by lowering the relative energy for the *cis*-isomer. Consequently, the torsion angle ω can also be influenced by the incorporation of special amino acids. Pseudoprolines **6** are synthetic proline analogues that can be obtained by a cyclocondensation reaction of the amino acids cysteine, threonine or serine with aldehydes or ketones [63]. These derivatives, which have been mentioned in the context of the synthesis of difficult sequences (cf. Section 4.5.4.3), also influence the *cis/trans* peptide bond ratio [64, 65]. The introduction of different substituents at C² of the pseudoprolines can be used to predetermine the *cis/trans* peptide bond ratio. In particular, the C²-disubstituted analogues **6** ($R^3 = R^4 = \text{CH}_3$) induce up to 100% *cis*-configured peptide bonds in di- or tripeptides [64–66].



The influence of certain naturally occurring amino acids on the secondary structure of proteins and peptides has been discussed in Chapter 2 (Section 2.4). This information on the conformational bias of certain amino acids can be utilized to guide the design of peptide analogues. Besides their ability to influence the *cis/trans* peptide bond ratio, N-alkyl-substituted amino acids can favor the formation of turn structures because they very often occur in position $i + 2$ of a β -turn [67]. Similarly, D-amino acids may also be used for the rational design of peptides with defined secondary structure because they stabilize a β II'-turn where they occur in position $i + 1$. The same position $i + 1$ is often favored by proline in β I- or β II-turns. In these cases, a *trans*-peptide bond is involved. Proline with a *cis*-peptide bond usually occupies position $i + 2$ in a β VIa- or β VIb-turn; this situation is also often observed for N-alkyl amino acids.

The dipeptide sequence D-Pro-Gly especially favors a β -turn, and also promotes hairpin nucleation [68]. Strategies for the stabilization of peptide conformations by secondary structure mimetics will be discussed in Section 7.2.4.

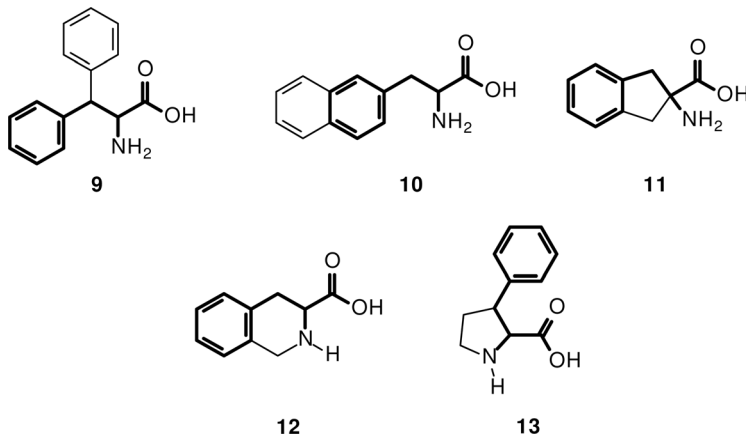
The [N-MeNle³]CCK-8 analogue displays higher specificity for the CCK-B receptor than the CCK-A receptor. This preference is due to a *cis*-peptide bond present in the

[Nle³]CCK-8 analogue, but not in the [N-Me-Nle³]CCK-8 derivative, as has been shown by NMR spectroscopy [69].

Moreover, multiple incorporation of N-methyl amino acids was shown to significantly improve metabolic stability and intestinal permeability of peptides. Incorporation of three N-methyl amino acids into the Veber–Hirschmann peptide, a cyclic somatostatin analog [c-(Pro-Phe-D-Trp-Lys-Thr-Phe-)], resulted in 10% oral bioavailability [70].

Some C^α-dialkyl amino acids, such as diethylglycine (Deg) **7**, α-aminoisobutyric acid (Aib) **8**, and isovaline (Iva, 2-amino-2-methylbutyric acid) occur naturally. These derivatives have frequently been incorporated into peptides to investigate the conformational requirements of receptors [71]. They also play an important role as building blocks for the stabilization of short peptides in a well-defined conformation, depending on the nature of the two substituents attached to the C^α-carbon. 3₁₀-Helices are stabilized by the incorporation of Aib and other C^α-dialkyl-substituted amino acids [72, 73].

Compounds **9–13** are constrained derivatives of phenylalanine. The introduction of a sterically demanding functional group at C^β, as in **9**, mainly constrains conformations around the side-chain dihedral angle χ^1 . This angle, in combination with the backbone dihedrals ϕ and ψ , determines the 3D positioning of side-chain functional groups (e.g., as pharmacophores). Consequently, amino acids with additional substituents at C^α, C^β, or cyclic amino acids are suitable building blocks for the introduction of a restriction in “ χ -space” [18, 74–76].

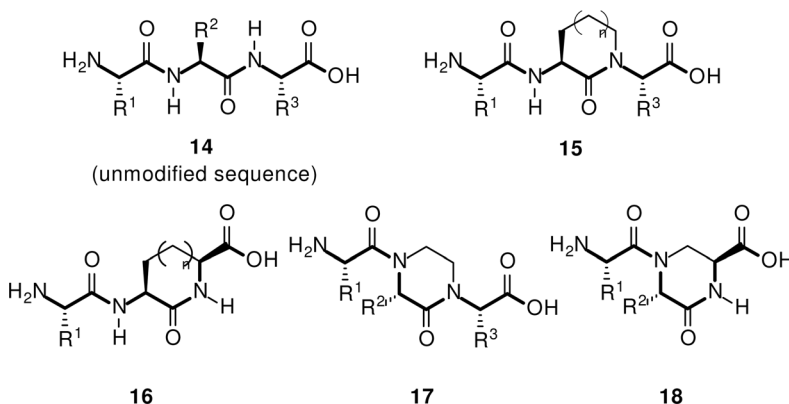


2-Naphthylalanine **10** is a sterically demanding derivative that may also lead to improved interaction of the aromatic system with hydrophobic areas of a protein receptor. Compound **11** is a representative of C^α-dialkyl amino acids. Tetrahydroisquinoline carboxylic acid **12** (Tic) is a phenylalanine analogue where the torsion angle χ^1 is limited to a very small range of values. 3-Phenylproline **13** may be regarded as a chimera containing both proline and phenylalanine substructures.

Incorporation of α,β -didehydrophenylalanine (Δ Phe) residue in a peptide induces in most cases β -turns in shorter peptides and 3_{10} -helices in longer sequences. The achiral and planar residue Δ Phe preferably induces torsion angles (ϕ, ψ) of $(-60^\circ, -30^\circ)$, $(-60^\circ, 150^\circ)$, $(80^\circ, 0^\circ)$, or the mirror image values, respectively [77].

The introduction of bridges between two adjacent amino acid residues leads to the formation of dipeptide mimetics.

Bridges can, for example, be formed between C^α and N^α (**15**, **18**), between two C^α (**16**), or between two N^α (**17**). This type of modification also makes the peptide more rigid.



7.2.2

Backbone Modification

The modification of the peptide chain (backbone) comprises, for instance, the exchange of a peptide bond by amide analogues (e.g., ketomethylene, vinyl, ketodifluoromethylene, amine, cyclopropene) [78]. Some basic types of peptide backbone modifications are displayed in Figure 7.4 [79].

The NH group of one or more amino acids within a peptide chain may be alkylated [80] or exchanged by an oxygen atom (depsipeptide), a sulfur atom (thioester), or a CH_2 group (ketomethylene isostere) [81]. However, peptidylthioesters are quite unstable. The CH moiety (C^α) may be exchanged in a similar manner by a nitrogen atom (azapeptide) [82, 83], by a C-alkyl group (C^α -disubstituted amino acid) [71], or by a BH^- group (borapeptide). The carbonyl groups may be replaced by thiocarbonyl groups (endothiopeptide) [84], CH_2 groups (reduced amide bond) [85], SO_n groups (sulfinamides, $n = 1$; sulfonamides, $n = 2$) [86], $POOH$ groups (phosphonamides), or B-OH groups. The peptide bond of one or more amino acid residues in **19** (cf. Figure 7.5) may be inverted (NH-CO), giving retro peptides. In order to maintain the original side-chain orientation, the retro modification (**20**, **21**) has to be accompanied by appropriate stereochemical compensation (inversion of the

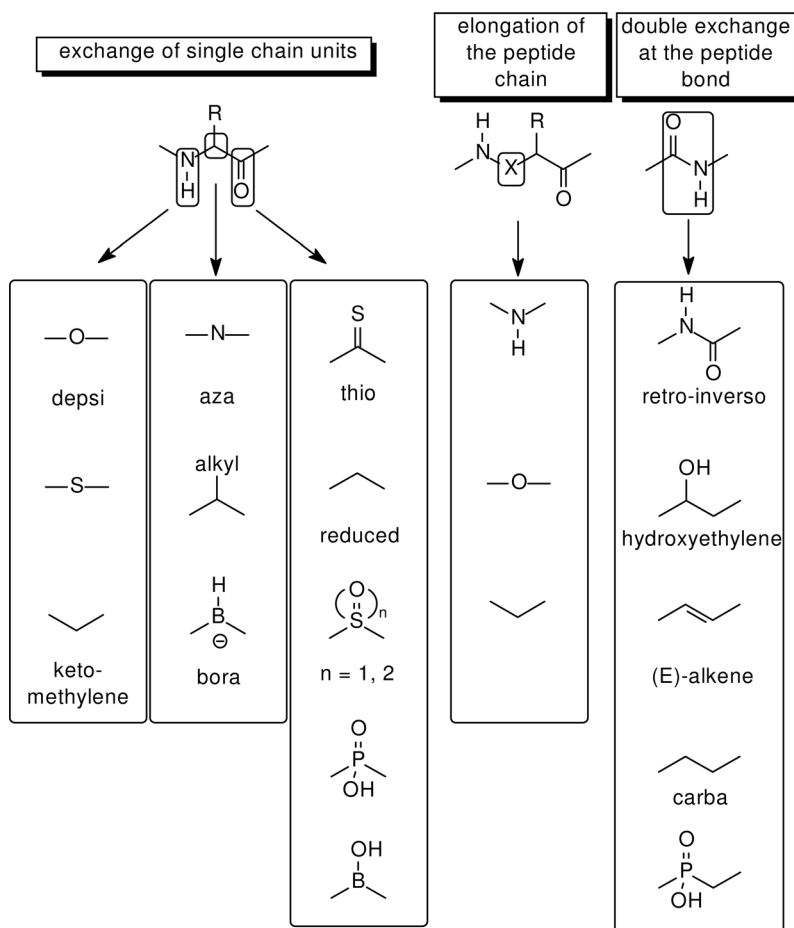


Figure 7.4 Selected types of peptide backbone modification.

configuration at C^α). The consequence is formation of the so-called retro-inverso peptides (Figure 7.5) [87, 88].

Peptides, where an amide bond has been replaced to give hydroxyethylene [89], *E*-alkene [90, 91], or alkane structures, have also been described. Besides these isosteric replacements, the peptide chain may be extended by one atom (O: aminoxo acid [92, 93]; NH: hydrazino acid [94, 95]; CH_2 : β -amino acid [96–98]; Section 7.4).

Vinylogous peptides contain amino acids extended by a vinyl group as building blocks, which is also not an isosteric replacement (Section 7.4). Such peptide analogues have been found in nature, as in the compounds cyclotheonamide A 22 ($R = \text{H}$) and B 23 ($R = \text{Me}$) that occur in the sponge *Theonella*. Besides vinylogous tyrosine, these peptides further contain the nonproteinogenic amino acid β -amino- α -oxohomoarginine [99]. They act as thrombin inhibitors and are of interest in medicinal chemistry as potential antithrombotic agents.

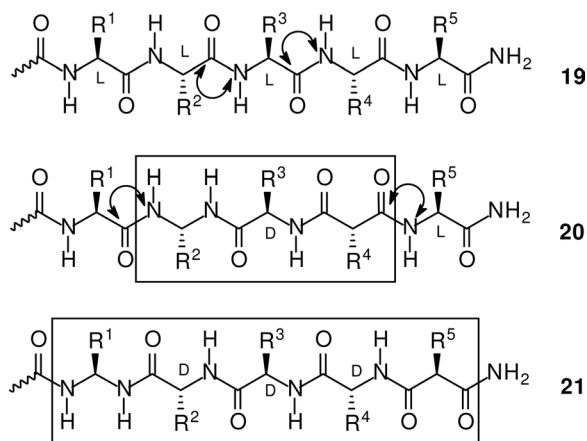
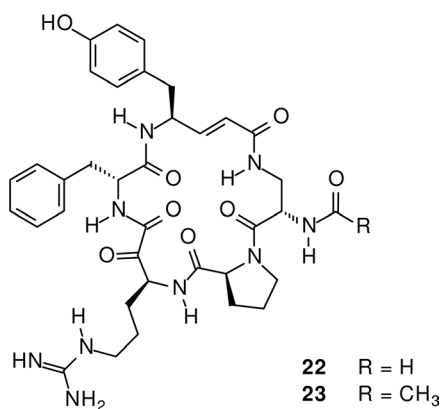


Figure 7.5 Retro-inverso peptide modification.



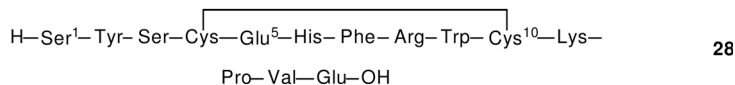
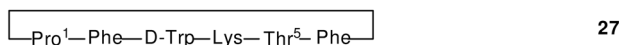
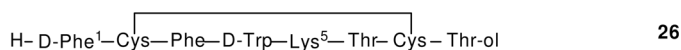
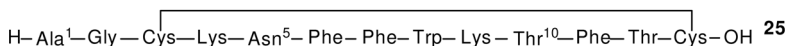
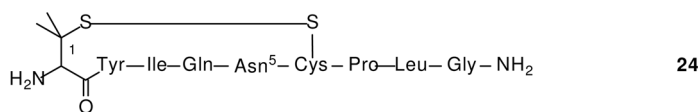
7.2.3

Combined Modification (Global Restriction) Approaches

The introduction of global restrictions in the conformation of a peptide is achieved by limiting the flexibility of the peptide through cyclization (Section 6.1). This may eventually lead to a stabilization of peptidic secondary structures and, consequently, this is an important approach to drug development. Turns (or loops) are important conformational motifs of peptides and proteins, besides α -helix or β -sheet structures. Reverse turns comprise a diverse group of structures with a well-defined 3D orientation of amino acid side chains. β -Turns constitute the most important subgroup, and are formed by four amino acids (see Section 2.5.1.3). Turns lead to a reversion of the peptide chain, and may be mimicked by cyclic peptides with well-

defined conformation. Depending on the type of cyclization (Sections 6.1 and 6.2) and on the ring size, the angles ϕ , ψ , ω and χ will have limited flexibility. This reduction in the degrees of freedom may eventually lead to receptor binding with high affinity because of entropic reasons, provided that the receptor-bound conformation is still accessible to the modified peptide.

In heterodetic cyclic peptides where the cyclization is obtained by cysteine disulfide bridges, cysteine residues may also be replaced by sterically hindered cysteine analogues such as penicillamine (3-mercapto-valine, Pen). Such an exchange may lead to a differentiation between agonistic and antagonistic analogues, as has been shown by Du Vigneaud *et al.* for [Pen¹]oxytocin **24** and by Hruby *et al.* for cyclic enkephalin analogues [74]. [D-Pen¹,D-Pen⁵]enkephalin (DPDPE), a cyclic disulfide, is a highly potent and selective δ -opioid receptor antagonist.



Ring size and the formation of stable conformations play important roles in the design of cyclic peptide analogues. This has been intensively investigated in the oxytocin and vasopressin series. In these cases cyclization to produce a 20-membered ring is essential for biological activity. However, this cyclization does not necessarily have to occur by disulfide formation, because the carba analogues, where the sulfur atoms are replaced by CH₂ groups, also display biological activity.

Somatostatin **25** (cf. Section 3.3.1.4) which is a regulatory peptide with a broad spectrum of activity, likewise contains an intrachain disulfide bridge. Potential applications have been envisaged in the treatment of acromegaly or of retinopathy in diabetes because **25** suppresses the liberation of growth hormone and glucagon,

an increased level of which contributes especially to organ damage in diabetes. Native somatostatin is rapidly metabolized and therefore is not suited to the treatment of juvenile diabetes; hence, metabolically stable analogues have to be developed. The excision of six amino acid residues from the ring in **25** produced the 20-membered somatostatin analogue **26** developed by Bauer *et al.* [100]. This peptide drug, named octreotide, has been approved for the treatment of acromegaly and of patients with metastasizing carcinoid and vasoactive tumors. Finally, the ring size may be further reduced to produce an 18-membered cyclic hexapeptide **27** that induces the liberation of growth hormone, insulin and glucagon with the same potency as somatostatin [101].

Peptides that natively are linear sequences can also be modified by the incorporation of cyclic structural elements in order to improve receptor interaction. In this context, the superagonist **28** of the α -melanocyte-stimulating hormone should be mentioned, where a β -turn predicted for the linear peptide may be stabilized by the replacement of Met⁴ and Gly¹⁰ with two cysteine residues and disulfide bridge formation [102].

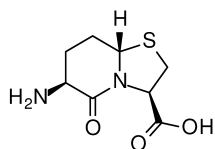
Kessler *et al.* [25] first established the concept of “spatial screening”, whereby small libraries of stereoisomeric peptides with conformational constraints (e.g., cyclic peptides) are used for different 3D presentations of pharmacophoric side-chain groups (Section 6.1). The correlation of biological activity with peptide conformation provides useful information about the peptide conformation with the best fit to the corresponding target receptor.

7.2.4

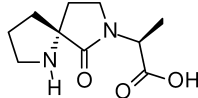
Modification by Secondary Structure Mimetics

It is desirable to have a repertoire of building blocks that reliably induce a desired conformation in a peptide. Consequently, numerous efforts have been undertaken to synthesize secondary structure mimetics that induce a well-defined secondary structure [103]. The desired conformation should be imitated as closely as possible, and the synthetic route for the secondary structure mimetic should permit the introduction of appropriate side chains onto the mimetic scaffold. While certain α -amino acids also exert conformational bias on the peptide chain (see Sections 2.5.1.3, 6.1, and 7.2.1), this section focuses on synthetic nonproteinogenic modules that induce and stabilize secondary structure. A large variety of conformationally constrained dipeptide mimetics that may be introduced into a strategic position in a peptide, using methods of peptide synthesis, in order to induce a defined secondary structure has been compiled in reviews [104–107]. Derivatives of this type may also be used as scaffolds for peptidomimetics (Section 7.3).

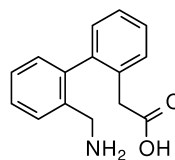
The β -turn is a structural motif that is often postulated to occur in the biologically active form of linear peptides [108]. Consequently, most secondary structure mimetics imitate a β -turn. Compounds **29–37** are examples of β -turn mimetics (**29** [109], **31** [110], **32** [111], **33** [112], **34** [113], **35** [114], **36** [115], **37** [116]).



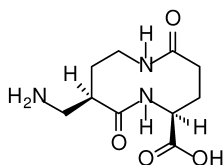
29



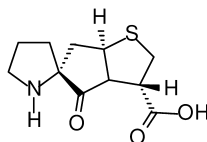
30



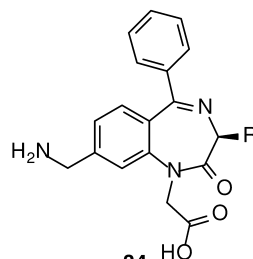
31



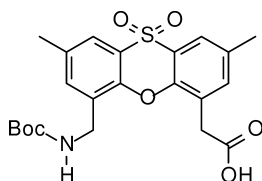
32



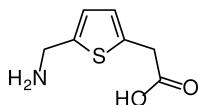
33



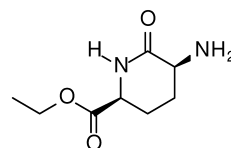
34



35



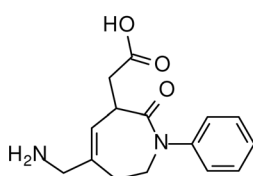
36



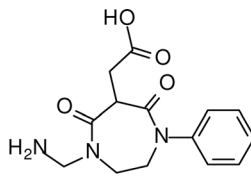
37

However, as has been shown by both experimental [117] and theoretical methods [118], not all scaffolds designed as β -turn mimetics actually display the desired properties.

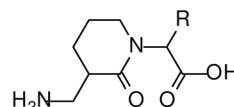
γ -Turns, which occur less frequently than β -turns in proteins and peptides, may also be mimicked by a series of compounds. β -Amino acids are most likely the simplest γ -turn mimetics; they induce stable conformations, for example in cyclic tetra- and pentapeptides composed of four α -amino acids and one β -amino acid [98]. Other γ -turn mimetics are compounds 38 [119], 39 [120], and 40 [121].



38



39



40

α -Helix initiators [122, 123] and ω -loop mimetics have also been described [124]. ω -Loops are larger reverse turns in a peptide chain comprising 6–16 amino acid residues.

7.2.5

Transition State Inhibitors

Structural complementarity is a key issue in enzyme–substrate interactions, and both substrates and substrate analogues compete for binding to the active site and formation of the enzyme–substrate complex. It was first recognized in the 1920s that the affinity of a substrate towards a catalyst increases concomitantly with the distortion of the reactant towards its structure in the transition state. The transition state of an enzyme reaction can be assumed to differ significantly from the substrate ground state conformation. For example, a transition state with a tetrahedral configuration is formed in nucleophilic displacement reactions starting from a sp^2 -hybridized carboxy group, as is present in peptide bond hydrolysis. Pauling [125] suggested that a potent enzyme antagonist might be developed that mimics the enzyme–substrate transition state despite being unreactive (transition state analogue, Figure 7.6). Substrate and substrate analogues are relatively weakly bound in

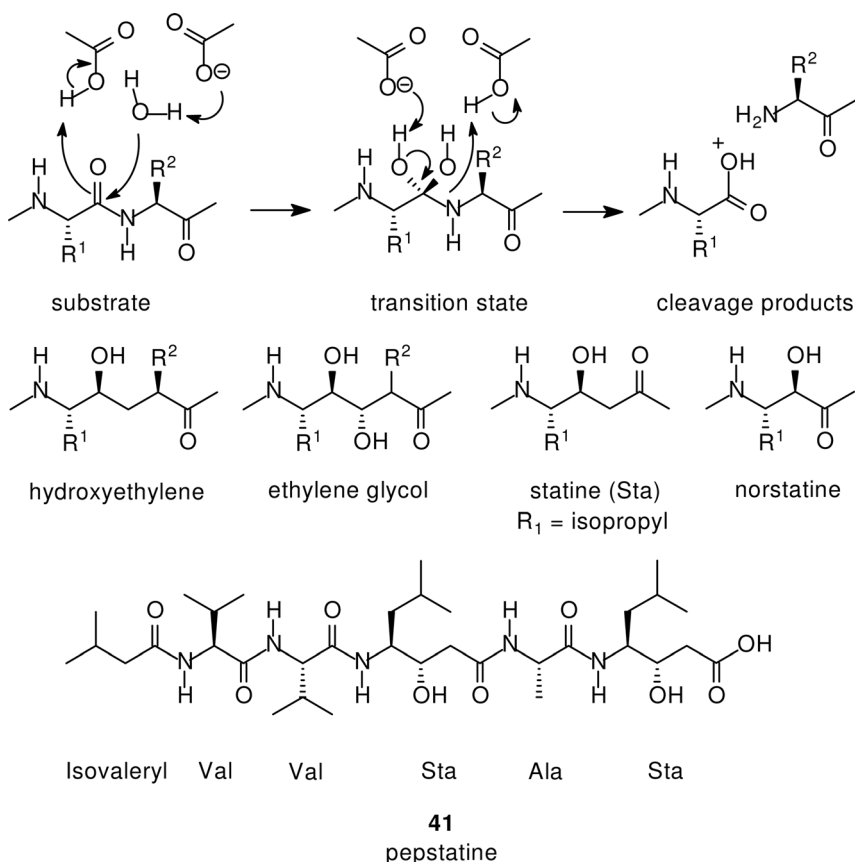


Figure 7.6 Mechanism of proteolysis by an aspartyl protease and structures of some transition state inhibitors.

their ground state, because tight binding would be counterproductive to efficient catalysis, while the substrate in its transition state is bound much more closely [126].

Based on this concept, transition state inhibitors of enzyme reactions (e.g., protein or peptide cleavage by a protease) have been developed (Figure 7.6) [5, 127, 128]. Nature once more provided a prototype for this kind of inhibitor with pepstatine **41**, an inhibitor of aspartyl proteases. This peptide (isovaleryl-Val-Val-Sta-Ala-Sta-OH), which is isolated from culture filtrates of *Streptomyces* species, contains – besides the N-terminal acyl moiety – the unusual amino acid statine (Sta), which is regarded as an amino acid analogue mimicking the transition state of peptide bond cleavage. Hydroxyethylene, ethylene glycol, or norstatine represent other building blocks for transition state inhibitors. This amino acid and its analogues have been applied in the development of many mechanism-based inhibitors of proteases and other enzymes [5]. The concept has also been utilized for the design of suitable haptens in the creation of catalytic antibodies (see Section 4.6.3) [129, 130].

7.3

Peptidomimetics

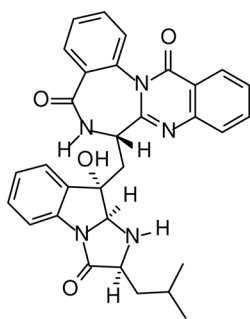
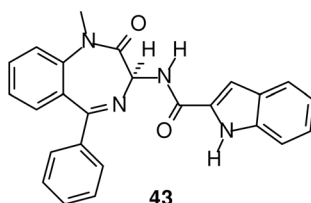
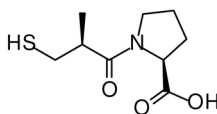
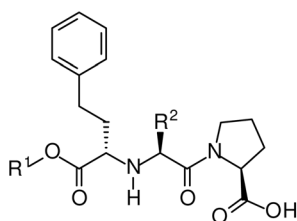
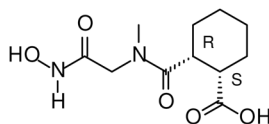
As the therapeutic application of peptides is limited because of certain disadvantages (see above), small organic molecules remain, in most cases, the most viable approach for the identification and optimization of potential drugs [131]. A major goal of modern medicinal chemistry is to find rational first principles for systematically transforming the information present in a natural peptide ligand into a nonpeptide compound of low molecular weight (type II and type III peptidomimetics). Chemical structures designed to convert the topographical information present in a peptide into small nonpeptide structures are referred to as peptidomimetics. Today, these compounds – which combine bioavailability and stability superior to that of bioactive peptides with increased receptor selectivity – are the subject of major interest by pharmaceutical companies. Peptidomimetics range from peptide-isosteric molecules to compounds where similarities to peptides can barely be identified. Information obtained from the structure–activity relationships and conformational properties of peptide structures can enhance the design of nonpeptide compounds [132]. The design of a peptidomimetic relies primarily on knowledge of the conformational, topochemical, and electronic properties of the native peptide and its corresponding receptor. Structural effects, such as a favorable fit to the binding site, the stabilization of a conformation by introducing rigid elements, and the placement of structural elements (functional groups, polar or hydrophobic regions) into strategic positions favoring the required interactions (hydrogen bonds, electrostatic bonds, hydrophobic interactions) must be taken into account.

The major objective in the development of small molecules displaying pharmacological activity is to mimic the complex molecular interactions between natural proteins and their ligands. Often, the mode of action of a biologically active peptide on the receptor level can be imitated by a small molecule (agonism), or can be blocked (antagonism) [1, 4, 18, 133]. A peptidomimetic may also be designed as an enzyme inhibitor.

The design of peptidomimetics exceeds the mere application of peptide modifications, and is targeted towards nonpeptidic compounds, characterized by a high degree of structural variation. Both random screening and design using molecular modeling have proved to be helpful in drug development. Peptidomimetic receptor antagonists or enzyme inhibitors usually imitate the binding of nonadjacent peptide substructures (pharmacophoric groups) to a protein, with this interaction often also relying on hydrophobic protein–ligand interactions. However, small organic molecules containing hydrophobic moieties may undergo a conformational transition in aqueous solution because of an undesired intramolecular hydrophobic interaction (hydrophobic collapse). Consequently, the pharmacophoric groups should be presented on a rather rigid scaffold in order to avoid hydrophobic collapse of the bioactive conformation in an aqueous environment [5].

Peptidomimetics may be identified or discovered by extensive screening of natural or synthetic product collections (compound libraries). The combinatorial synthesis (Chapter 8) of a multitude of different peptidic or nonpeptidic compounds, combined with careful evaluation of receptor binding, are promising approaches for the discovery of new lead structures that subsequently must be further optimized with respect to their pharmacological properties.

Asperlicin **42** was discovered during a screening of fungal metabolites, and was found to be a lead structure as a cholecystokinin A (CCK-A) antagonist. **42** is structurally very similar to diazepam, which consequently was combined with a D-tryptophan structural motif; this eventually led to the design and synthesis of a selective orally administered peptidomimetic antagonist **43** of the peptide hormone cholecystokinin [134].

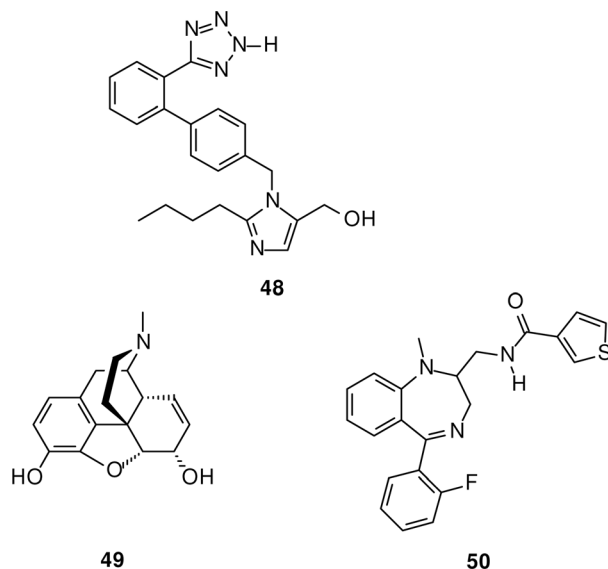
**42****43****44****45** $R^1 = \text{H}$, $R^2 = (\text{CH}_2)_4\text{NH}_2$ **46** $R^1 = \text{CH}_2\text{CH}_3$, $R^2 = \text{CH}_3$ **47**

The angiotensin-converting enzyme (ACE, cf. Section 3.3.6.1) inhibitor captopril **44** [135] was the first peptidomimetic compound to have been developed by rational design. Wolfenden introduced (R)-2-benzylsuccinic acid as an inhibitor for carboxypeptidase A (CPA), the enzyme mechanism of which is closely related to that of ACE [136, 137]. (R)-2-Benzylsuccinic acid belongs to the class of transition state analogues (Section 7.2.5).

The 9-peptide teprotide Pyr-Trp-Pro-Arg-Pro-Gln-Ile-Pro-Pro-OH isolated from snake venoms possesses bradykinin-potentiating activity that is based on the inhibition of ACE. While the modified C-terminal dipeptide sequence Ala-Pro displays only weak inhibition of ACE, exchange of the N-terminal amino group by a carboxy group resulting in a more potent ACE inhibitor. This modification was based on the finding by Wolfenden concerning the inhibitory activity of benzylsuccinic acid on CPA. Replacement of the carboxy group by a thiol function, which strongly coordinates metal ions (e.g., Zn^{2+}), resulted in captopril, which is an inhibitor of the Zn-dependent metalloprotease ACE, and has been approved as an orally administered drug. The highly potent analogues lisinopril **45** and enalapril **46** have been synthesized by variation of different regions of captopril [138]. The hydroxamate **47** displays extremely low toxicity and an activity for ACE inhibition which is comparable to that of **44** ($\text{IC}_{50} = 7 \text{ nM}$) [139].

Losartan **48** is a highly active angiotensin II antagonist that was optimized by molecular design. It is the first nonpeptide angiotensin II antagonist, and is currently used for the treatment of hypertonia.

Morphine **49**, as a representative of the opioid alkaloids, is the classic example of a nonpeptidic compound that was found to be a mimetic of endogenous peptides. Morphine imitates the biological effect of β -endorphin on the corresponding receptor. The endogenous opioids resemble morphine very much from a pharmacological point of view, even in the side effects such as addictive potential and respiratory depression. Important developments following the discovery of opioid peptides included evidence of the existence of three different opioid receptor classes (μ , δ , κ), each with various subclasses (see Section 3.3.2.1). High expectations were imposed on this field of research, which was aimed towards the development of an opioid compound with high analgesic potency but without any adverse side effects, and as a result the field of opioid peptide and peptidomimetics became popular with regard to the application of new design principles. Concepts, such as conformational restriction, peptide bond replacements, the incorporation of turn mimetics, and library screening for the identification of novel ligands, were extensively tested in the area of opioids. However, despite many interesting developments, the original aim has not (yet) been achieved. For example, the benzodiazepine derivative tifluadom **50** is an agonist of the κ -opioid receptor and an antagonist of the CCK-A receptor. Although animal experiments have shown it to be devoid of addictive potential and respiratory depression [140], its administration is associated with locomotor incoordination [141].



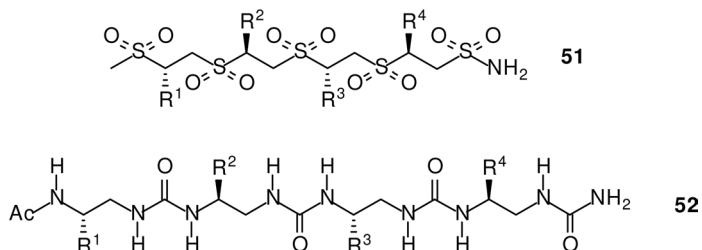
7.4

Pseudobiopolymers

Antisense oligonucleotides form one of the first classes of non-natural biopolymers [142]. Several other types of pseudobiopolymers have been described, and these are composed of nonpeptidic molecules in a similar, repetitive manner as peptides. Moreover, some pseudobiopolymers fold to give stable, reproducible secondary structures that mimic protein secondary structures such as helices, sheets, or turns [103, 143–145]. They are often characterized by metabolic stability, resemble peptidic structures, and some of them may be used therapeutically. Gellman first coined the term foldamers for any polymer that reproducibly adopts a specific ordered conformation [143, 146]. The study of foldamers, that display intrinsic folding propensities, has revealed different synthetic backbones with conformational behavior like a biopolymer.

The peptoids (Section 7.4.1) [147], peptide nucleic acids (Section 7.4.2) [148], and β -peptides (Section 7.4.3) [149–154] are important representatives of this class of compounds. Oligomers of aza-amino acids (hydrazine carboxylates) have been named azatides [155, 156]. Oligosulfonamides [157, 158], hydrazinopeptides, composed of α -hydrazino acids [94, 159] and aminoxypeptides, composed of α -aminooxy acids [92], may be regarded as β -peptide analogues with respect to the backbone atom pattern (Section 7.4.3). Oligocarbamates (Section 7.4.4) [160], oligopyrrolinones (Section 7.4.5) [161], oligosulfones **51**, γ -peptides [162, 163], oligoureases **52** [145], vinylogous peptides [164–166], vinylogous oligosulfonamides [167], and

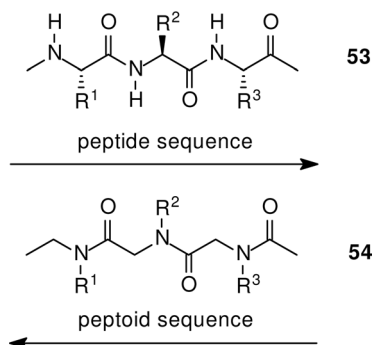
carbopeptides (oligomers of carbohydrate-derived amino acids) [168, 169] form further classes of pseudobiopolymers. Many of these non-natural oligomers (e.g. peptoids, β -peptides [170], γ -peptides [170], oligocarbamates) are completely resistant towards proteolysis. Some representatives of these classes will be briefly introduced in the following sections.



7.4.1

Peptoids

Peptoids are oligomers composed of N-substituted glycine building blocks [147, 171, 172]. The side chain of each amino acid in the peptides is formally shifted in the peptoids by one position from C^α to the amino group nitrogen. Comparison of the peptide chain 53 with the peptoid chain 54 shows that the direction of the peptide bond in 54 should be reversed (retro-sequence) in order to provide the same relative arrangement of side-chain residues R and carbonyl groups. Peptoids are achiral, and resemble the structures of the retro-inverso-peptidomimetics [87]. Helical peptoid structures are favored when bulky N-alkyl side chains are present [173]. Interestingly, these helices are not stabilized by hydrogen bonds, and the helicity depends on side-chain chirality [174] and oligomer length. N-Substituted glycine derivatives can be synthesized easily. The principle of solid-phase peptoid synthesis is shown in Figure 7.7 [7].



Among the different types of polymeric support which may be used, the Rink amide resin has provided good results. Variant A utilizes the Fmoc/tBu tactics where

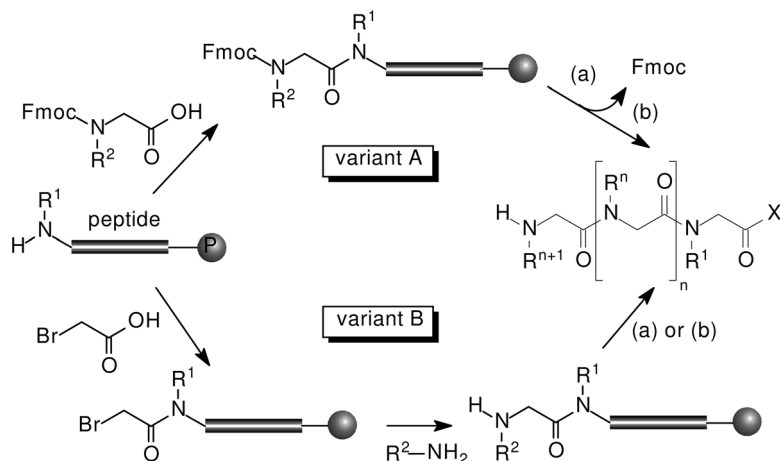


Figure 7.7 Solid-phase synthesis of peptoids. Variant A, building block approach; variant B, submonomer approach. (a) chain elongation (n repetitions); (b) deprotection and cleavage from solid support; $R^1 \dots R^n$ = side-chain residues; $X = \text{NH}_2, \text{OH}$.

the suitably protected N-substituted glycine derivatives are coupled to polymer-bound NHR groups using PyBOP or PyBroP [147]. The submonomer solid-phase synthesis (variant B) [171] does not apply the N-substituted glycine monomer. Instead, bromoacetyl building blocks are coupled to the growing peptoid chain and then subjected to a nucleophilic substitution reaction of the bromide substituent by a primary amine carrying the side-chain substituent. Additional protection of the amino group is not necessary, and no different monomeric building blocks have to be synthesized. Peptoid libraries may be synthesized in a similar manner to peptide libraries, and this facilitates high-throughput screening for the identification of novel nonpeptide drugs.

Besides α -peptoids formed from N-substituted glycine residues, β -peptoids with achiral side chains [175] and chiral side chains [176] have also been reported. The submonomer approach towards β -peptoids closely resembles variant B in Figure 7.7, but uses acryloyl moieties instead of bromoacetyl groups as the electrophile [175, 176]. Oligomers of alternating repeats of α -amino acids and β -peptoid residues with chiral N-alkyl moieties have been synthesized by fragment condensation of (Xaa- β -Yaa) dimers and investigated with respect to antimicrobial and antiplasmodial activity [177, 178].

Conformational studies revealed that peptoids display higher conformational flexibility than peptides. Theoretical studies have shown that both α -peptoid and β -peptoid structures should be able to adopt helical structures with both *trans*- and *cis*-peptide bonds despite lacking hydrogen bond donor capabilities. There are close relationships between the helices of α -peptoids and polyglycine and polyproline helices observed for α -peptides, while β -peptoid helices are predicted to correspond to the 3_{14} -helix of β^3 -peptides [179].

Peptoids resist proteolysis and are, therefore, metabolically stable. Preliminary studies show that peptoids with structural resemblance to biologically active peptides have, for example, a similar activity to that of enzyme inhibitors compared to the native peptide sequence. The peptoid analogue of the hepatitis A 3C-protease inhibitor Ac-Glu-Leu-Ala-Thr-Gln-Ser-Phe-Ser-NH₂ displays, in the retro-sequence, a similar inhibition to the original peptide. An additional important finding is that peptoid analogues of peptide sequences from HIV-Tat protein bind to the corresponding RNA. This class of peptoids represents a promising concept in the search for new lead structures to influence pathological events. A series of peptoids with biomimetic sequences that additionally contain chiral N-alkyl groups fold into helical structures and display low-micromolar antimicrobial activities, similar to naturally occurring cationic antimicrobial peptides [180].

7.4.2

Peptide Nucleic Acids (PNA)

Peptide nucleic acids are nonionic analogues of oligonucleotides containing a peptide backbone [148, 181–183]. They are mimetics of DNA that can be employed as antisense nucleotides. PNA originally were designed to mimic an oligonucleotide binding to double-stranded DNA in the major groove, but they have eventually succeeded as good structural analogues of nucleic acids. The antisense concept is based on the application of oligonucleotides to hybridize with DNA or RNA in order to prevent transcription by the formation of a triple helix, or to bind mRNA by duplex formation. However, the first examples – which were described as DNA/RNA backbone analogues [184, 185] – did not display any ability to hybridize with oligonucleotides. Today, this concept is of therapeutic importance in cases where diseases can be treated on the DNA or mRNA level. The nonionic oligonucleotides of the PNA type are totally inert towards degradation by nucleases, and can also be synthesized in larger amounts by solid-phase synthesis.

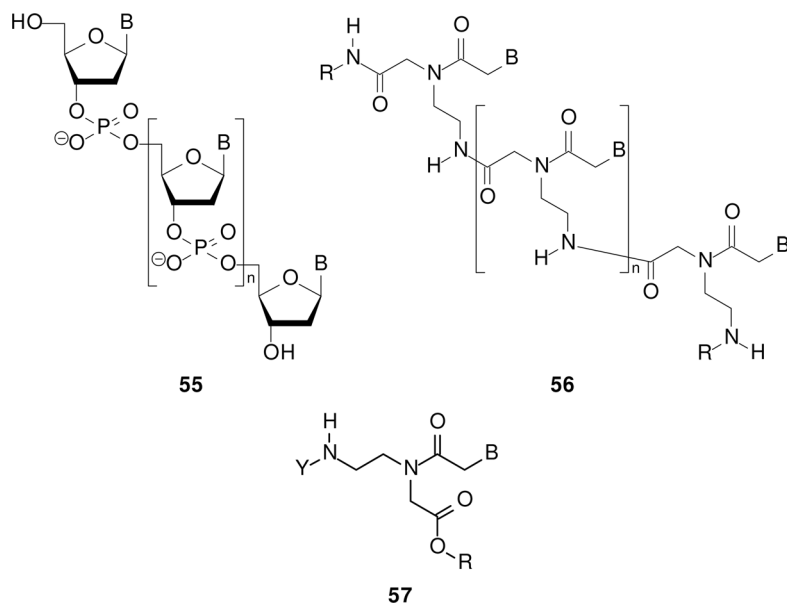
Originally designed to recognize double-stranded DNA in a sequence-specific manner, the unique physico-chemical properties of PNA have paved the way for the development of a variety of diagnostic assays [186–190].

The desoxyribose-phosphate backbone of an oligodesoxynucleotide **55** is replaced in the PNA **56** by a pseudopeptidic 2-(aminoethyl)glycine unit. Besides the 2-(aminoethyl)glycine moiety, other monomers may be used for PNA synthesis [148].

The N-protected monomer unit **57** (Y = Boc or Fmoc) comprises an amino acid [2-(aminoethyl)glycine] that is connected via a methylenecarbonyl spacer to the corresponding nucleobase B (thymine, cytosine, adenine, or guanine).

In principle, most of the coupling reagents used in peptide chemistry can be applied to PNA synthesis. The active ester **57** (R = Pfp) may be used. The building blocks of type **57** can be obtained by alkylation of the corresponding base (B) with methyl-2-bromoacetate in DMF in the presence of K₂CO₃. After saponification of the methyl ester and conversion of the free carboxylic acid to the pentafluorophenylester, the nucleobase derivative is reacted with N-(N'-Boc-aminoethyl)glycine (N'-Boc-Aeg), which again is subsequently converted into the pentafluorophenylester **57** using

DCC. An alternative synthesis of Fmoc-PNA-OPfp **57** (Y = Fmoc) has been described [191].



7.4.3

β-Peptides, Hydrazino Peptides, Aminoxy Peptides, and Oligosulfonamides

One characteristic feature of β-amino acids is the additional carbon atom between the amino group and the carboxy group. β-Amino acids derived from α-amino acids by homologation are called β-homoamino acids, as there is an additional C₁ unit inserted between the carboxy group and C^α of a natural amino acid. Side chains may be attached either to C^α (β²-homoamino acid, **59**) or to C^β (β³-homoamino acid, **58**), or to both of them. This fact considerably influences the secondary structure of the corresponding oligomer. The first structural investigations on homo-oligomers of β-amino acids were performed over 40 years ago. Kovacs *et al.* postulated, in 1965, that poly-(L-β-aspartic acid) forms a 3.4₁₄-P-helix [right-handed helix with 3.4 amino acid residues per turn and a 14-membered hydrogen bonded “ring” from NH (*i*) to CO (*i* + 2)] [192]. Homopolymers of poly-(L-β-aspartic acid) have been investigated by Muñoz-Guerra *et al.* [149, 193], while Gellman *et al.* [143, 150, 151] and Seebach *et al.* [152, 153, 163] profoundly examined the conformational behavior of a series of different well-defined β-peptides. All these compounds adopt predictable and reproducible helical conformations with hydrogen bond patterns depending on the substituent position (Figure 7.8). Indeed, many β-peptides prefer 3₁₄-helices even in aqueous solutions [153, 194–196]. Helix formation occurs already in β-peptide hexamers, while in the case of α-amino acids usually more than 10–12 amino acid

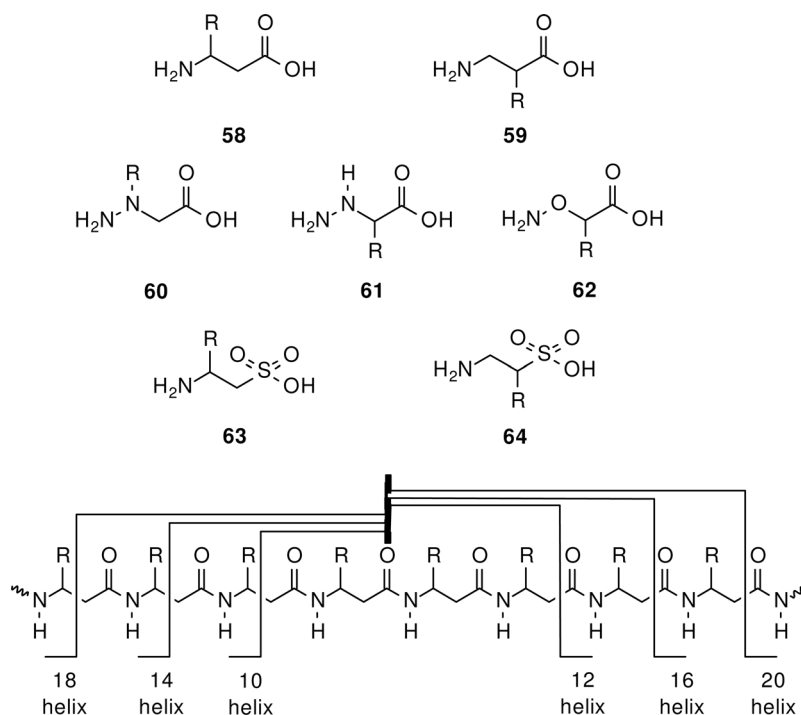


Figure 7.8 β -Amino acids, β -amino acid analogues and hydrogen bond patterns in β -peptide helices.

residues are required to form a stable helix. In molecular dynamics simulations, β -peptide helices have been shown to refold spontaneously after thermal defolding [197]. An alternating combination of β^2 - (59) [198] and β^3 -amino acids (58) leads to an irregular helix with 10- and 12-membered hydrogen bond “rings” [199, 200]. Both Boc- [201, 202] and Fmoc-protected β -amino acids [96] are available for peptide synthesis. β -Peptides display besides interesting conformational properties [163, 200, 203, 204] also unexpected biological activities, e.g. as somatostatin analogues [163, 205], as agonists, or ligands of major histocompatibility complex (MHC) proteins in the immune response, of the lipid transport protein SR-B1, of the HIVgp41 fusion protein, and others [206], or as antibacterial agents [207].

Recent investigations have uncovered the potential of mixed or “heterogeneous” backbones to expand the structural and functional repertoire of foldamers. In such compounds either sequences of alternating α -amino acids and β -amino acids or block oligomers of these building blocks are present [208]. β -Peptides able to assemble into defined, cooperatively folded, quaternary structures resembling small proteins composed of β -amino acids have been reported [209].

Hydrazino peptides [94, 159] and aminoxy peptides [92, 93], composed of α -hydrazino acids (60, 61) or α -aminoxy acids (62) are structurally related to β -peptides, because they also contain an additional skeleton atom between the amino and the carboxy groups. Foldamers of α -, β - and γ -aminoxy acids have been studied and it was shown that such peptides consisting of aminoxy acids adopt

well-defined secondary structures [93, 210]. α -Aminoxy peptides composed of α -aminoxy acids comprise analogs of β -amino acids with replacement of C^β by an oxygen atom. They form helices stabilized by hydrogen bonds in eight-membered “rings” [92]. The lone pair repulsion of the nitrogen and oxygen atoms renders the backbone of aminoxy peptides more rigid than that of β -peptides [93].

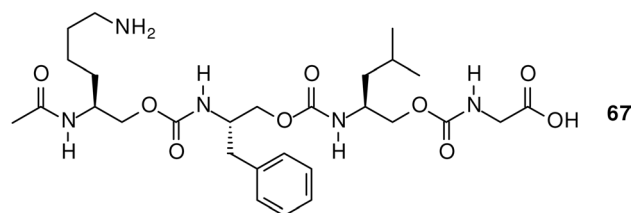
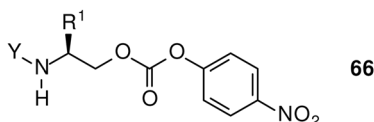
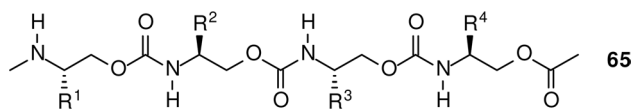
N-Hydroxy peptides contain N^α -hydroxy amino acids of type $-N(OH)-C^\alpha HR-(CO)-$. The hydroxamate group in such a modified peptide potentially acts as a metal coordination site. The N^α -hydroxy group confers different hydrogen bonding capabilities compared to native peptides, which influences the peptide conformation and potentially also molecular recognition [211].

7.4.4

Oligocarbamates

Oligocarbamates **65**, which form another class of pseudobiopolymers (biopolymer mimetics), contain a carbamate moiety instead of a peptide bond [160]. The oligocarbamates may be considered as γ -peptide analogues, with three skeleton atoms being found between the amino group and the carboxy group of one monomer. Oligocarbamates are stable towards proteolysis and are more hydrophobic than the peptides; hence, they may be able to permeate body barriers such as the blood–brain barrier.

The monomeric N-protected aminoalkylcarbonates are accessible starting from the corresponding amino alcohols that are converted into a carbamate active ester (e.g., *p*-nitrophenylester **66**). The oligocarbamates may be obtained in solid-phase synthesis using either the base-labile Fmoc group (**66**, Y = Fmoc) or the photolabile nitroveratryloxycarbonyl group (**66**, Y = Nvoc). Cho *et al.* [160] succeeded in synthesizing a library containing 256 different oligocarbamates, with **67** being one representative. The superscript “c” indicates the carbamate.

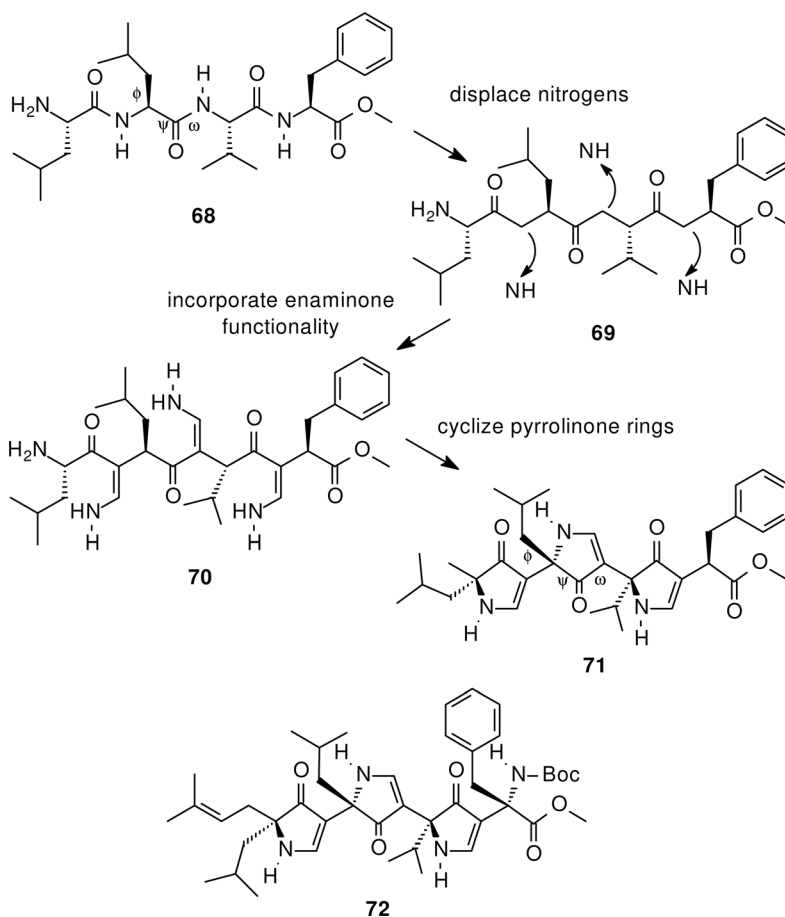


Ac-Lys^c-Phe^c-Leu^c-Gly-OH

7.4.5

Oligopyrrolinones

Oligopyrrolinones contain a novel peptidomimetic principle where a strongly modified peptide backbone is integrated in cyclic structures also containing vinyllogous amino acids [161]. A structural comparison between the oligopyrrolinone **71** and the tetrapeptide **68** with a parallel β -sheet structure clearly shows that **71** comprises comparable dihedral angles ϕ , ψ , and ω as well as a similar orientation of the side chains (Scheme 7.1). Formally, the amide nitrogen is displaced (**69**), incorporated in an enaminone (**70**), and then cyclized (**71**). Preconditions for the formation of hydrogen bonds, as in a β -sheet, are fulfilled. The 3D similarity of both the side chains and carbonyl groups in **68** and **72** has been proven by X-ray crystallography. **72** is present in an antiparallel β -sheet structure.



Scheme 7.1

7.5

Macropeptides and *de novo* Design of Peptides and Proteins

The *de novo* design of peptides and proteins has emerged as a challenging approach to study of the relationship between the structure and function of a protein. Solving the protein folding problem remains as the “Holy Grail” for computational structural biology, and this is still out of reach. *De novo* design can be regarded as an alternate route to tackle the problem of protein folding, the term having been coined for approaches that involve the construction of a protein with a well-defined 3D structure that does not consist of a sequence directly related to that of any natural protein.

7.5.1

Protein Design

Proteins are characterized by an impaired variety of shape and function. Consequently, protein design is not easy to realize compared to the design of smaller model compounds, for example the active sites of enzymes. Protein design represents a major challenge, and provides valuable information on the complex interplay between the structure and function of proteins. The principles and methods for the design of proteins have been the topics of several reviews [212–218].

Most *de novo* design strategies rely on knowledge obtained from an increasing amount of crystallographic protein structure data. These data provide basic information on the propensity of a single amino acid (or of a short amino acid sequence) to form a particular secondary structure. The assembly of supersecondary and tertiary folds relies on further hydrophobic or electrostatic interactions. Recent success in protein design is not merely the result of a simple hierarchic procedure, but has relied heavily on computer-based strategies [219–221].

Although the assembly of proteins by use of chemical synthesis, chemoenzymatic strategies and gene technological approaches is feasible (see also Chapters 4 and 5), it is not a straightforward task. Different concepts have been developed that mainly follow two important strategies [222]:

- The assembly of a larger peptide with known secondary structure motifs that forms a well-defined tertiary structure.
- The design of smaller peptide moieties of known secondary structure followed by assembly on a template.

De novo protein design comprises the arrangement of simple secondary structures (β -turn, α -helix, β -sheet) to tertiary structures (folds) of increasing complexity (β -hairpin [223, 224], helix–turn–helix [225], coiled coils, helix bundles; Section 2.4) [212, 213, 226–228].

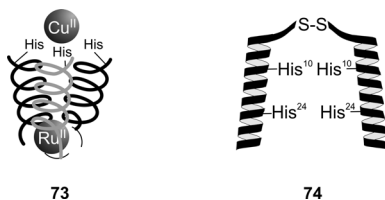
The design of non-natural redox proteins according to the former strategy has been realized by a sensible combination of complex chemistry and protein chemistry [213].

An interesting concept utilizes metal-ion-induced self-association for the difficult organization of small amphiphilic peptide subunits in aqueous solution to give topologically well-determined protein structures. Such a *de novo* design of functional

protein requires the assembly of molecular scaffolds with predetermined structure, where the desired functional groups of an active site of an enzyme can be introduced. Using this approach, Ghadiri *et al.* [229] obtained a Ru^{II} metalloprotein with an exactly defined metal coordination site. Within this metalloprotein, three histidine residues were located proximal to the C-terminus of the three-helix protein bundle and formed a coordinating site for a metal, thereby favoring the incorporation of Cu^{II} to give the $\text{Ru}^{\text{II}}\text{Cu}^{\text{II}}$ protein **73**.

Two different metal ions and three peptide subunits assemble in a highly selective process of self-aggregation to a heterodinuclear metalloprotein with three parallel helices. This achievement is one of the preconditions for the design of potentially redox-active metalloproteins and of artificial light-collecting metalloproteins.

DeGrado *et al.* [230] synthesized helical 31-peptides that assembled to give dimeric bundles of four helices that are connected pairwise via disulfide bridges. The model peptide **74** contains two histidine building blocks in each subunit (His^{10} , His^{24}), dimerizes upon addition of Fe^{II} -heme to give four redox centers at the putative binding sites, and also serves as a mimetic for the cytochrome b subunit of cytochrome bc_1 . Furthermore, miniaturized metalloproteins have been designed and synthesized *de novo* [213, 214, 231].



Peptide-based nanotubes comprise well-ordered supramolecular nanoparticles formed from linear or cyclic oligopeptides by self-assembly. This process is directed by the formation of an extensive network of intersubunit hydrogen bonds. The Ghadiri group has also rationally designed tubular peptide nanostructures (peptide nanotubes) and transmembrane ion channels (Figure 7.9). This concept is based on the assumption that cyclopeptides with alternating D- and L-configured amino acid building blocks form a flat ring where the planes of the amide groups in the scaffold are arranged perpendicularly to the ring plane [232].

Furthermore, it was postulated that this special arrangement is energetically favored under certain conditions, leading to the formation of tubular associates that are open at both ends. This process is favored by intermolecular hydrogen bonds together with interactions leading to a stacking of the cyclopeptide rings. Eventually, a selectively N-methylated cyclic peptide, $\text{c}-(\text{-Phe-D-N}^2\text{-Me-Ala-})_4$ could be synthesized that assembles in apolar organic solvents by self-organization to produce a discrete, soluble, cylinder-shaped associate. In the solid state, the peptide

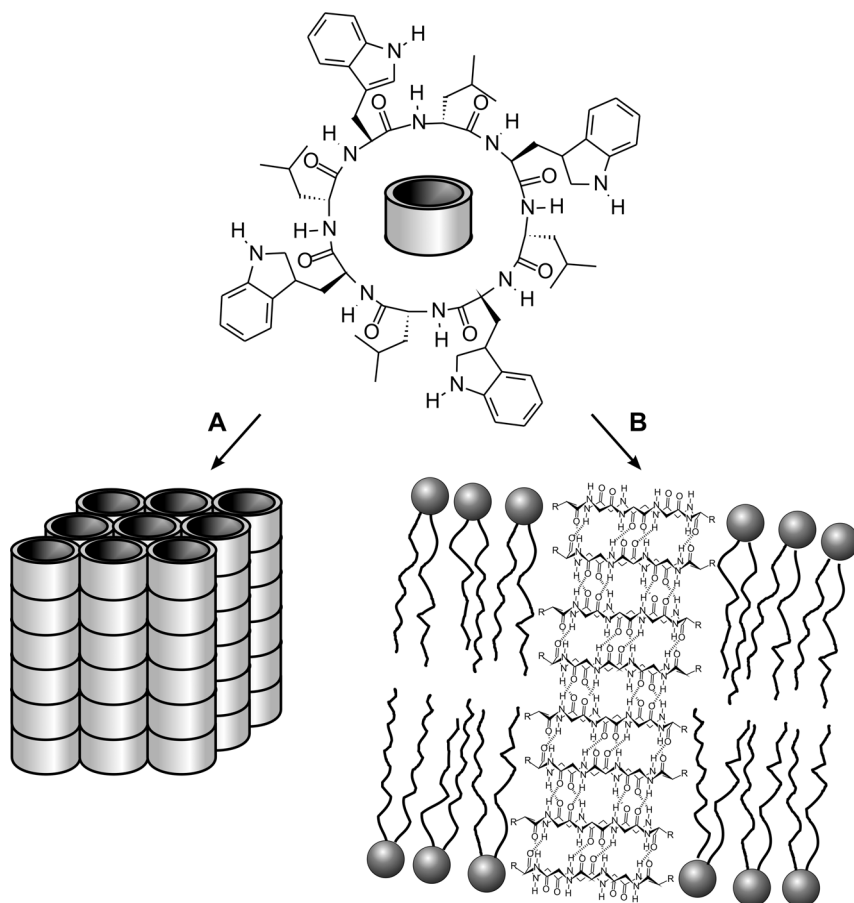


Figure 7.9 Assembly of peptide nanotubes (A) and transmembrane ion channels. (B) starting from cyclic peptides with alternating D- and L-amino acids.

forms, by self-organization, a crystal that is characterized by a pore structure in the shape of an ordered parallel arrangement of water-filled channels with a diameter of 7–8 Å.

The rational design of such structures is of considerable interest in the investigation into molecular transport processes, for the chemistry of inclusion compounds, and for catalytic processes, as well as the manufacture of new optical and electronic units.

The preparation of new artificial proteins and enzymes with predetermined 3D structure and tailor-made chemical, biological, and catalytic properties represents the ultimate goal for peptide and protein chemists.

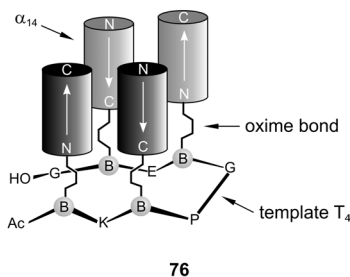
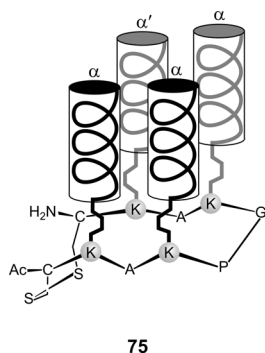
The second strategy referred to at the beginning of this chapter is the template-associated synthetic proteins (TASP) concept, developed by Mutter *et al.* [233, 234].

These TASP with predetermined 3D structure are obtained by synthesis. The *de novo* design of polypeptide sequences is limited by the protein-folding problem, though this can be avoided by constructing proteins of non-natural chain architecture with predetermined intramolecular folding. TASP molecules combine the structural features of natural proteins such as peptide moieties with predetermined secondary structure and synthetic elements such as topological templates; the result is branched structures with different possible folding topologies [215–217] (Figure 7.10).

By selecting the correct template, TASP with, for example, $\beta\alpha\beta$, α -helix bundle, or β -barrel tertiary structures can be obtained, whilst protein-like folding motifs may be constructed from a “molecular kit”. The individual building blocks (α -helix, β -sheet, loops, turns) are assembled in a regioselective manner by chemoselectively addressable groups at both ends of each building block [216, 235]. The schematic view of a TASP molecule **75** shows that the peptide chains attached to a cyclic octapeptide template via the lysine side chains form α -helical structures, and also associate intramolecularly to a four α -helix bundle that closely resembles protein structures.

The computational design of a TASP helical bundle, for example, requires the following steps:

- The design of amphiphilic helical peptide blocks based on the principles for secondary structure formation.
- Self-association of the peptide blocks by optimal intramolecular interaction.
- The design of tailor-made templates and functional attachment of the peptide building blocks.
- Stabilization of a hypothetical TASP structure by minimization of the conformational energy.



Solid-phase synthesis is used for the assembly of the peptides. The nonconsecutive mode of connection of the subunits excludes the application of synthetic methods by recombinant DNA techniques. Regioselectively addressable functionalized

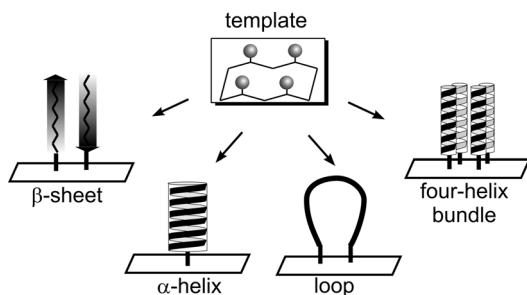


Figure 7.10 Structural motifs of a template-assembled synthetic protein (TASP).

templates (RAFT) are used for chemoselective ligation [215, 236], and allow the attachment of four different peptide building blocks or the assembly of binding loops as receptor mimetics. The antiparallel four-helix bundle TASP T_4 -($2\alpha 14\downarrow, \alpha 14\uparrow$) **76** was designed as a mimetic for the lysozyme epitope mAb HyHel-10 [237]. The TASP concept has also been used for the *de novo* synthesis of four-helix bundle metalloproteins in a combinatorial approach [238, 239]. It has also been adapted to an orthogonal assembly of small libraries of purified peptide building blocks and their application in protein design [240].

The α -helical coiled coil is another simple but highly versatile protein folding motif. Approximately 2–3% of all protein residues form coiled coils [241], the latter consisting of two to five amphipathic α -helices that are wrapped around each other in a left-handed twist with a seven-residue periodicity. The stability of this fold is mainly achieved by a packing of apolar side chains into a hydrophobic core (“knobs-into-holes”) [242]. In the seven-residue repeat the first and fourth positions are occupied by hydrophobic uncharged (U) residues forming a hydrophobic core at the interhelical interface. The fifth and seventh positions of the heptad repeat are often occupied by charged residues (C). The residues 2, 3, 6 in the outer boxes (X) do not contribute to interhelical contacts (Figure 7.11).

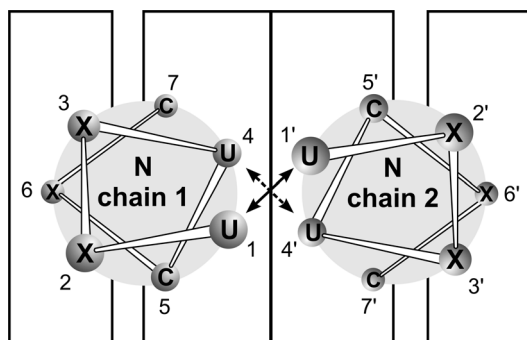


Figure 7.11 Interactions in coiled coils. U, uncharged amino acid; C, charged amino acid; X, any amino acid.

This α -helical coiled coil motif is often regarded as an ideal model system for investigations into the principles of protein stability and for *de novo* design [243–247]. Many studies have elucidated the influence of sequence variations on structure, orientation and aggregation state of coiled coils [248–251]. The design principle of coiled coil proteins has also been used for the concept of a *de novo* designed peptide ligase [252].

7.5.2

Peptide Dendrimers [253–256]

Dendrimers are highly ordered, highly branched polymers with a wide variety of potential chemical applications. Dendrimers bearing peptide moieties with predetermined secondary structure are useful tools in the *de novo* design of proteins that may be applied as antigens and immunogens, for serodiagnosis, and for drug delivery [253]. Dendrimers are formed by successive reactions of polyfunctional monomers around a core and, consequently, have many terminal groups.

Dendrimeric macromolecules differ from normal polymers as they are constructed from AB_n monomers in an iterative fashion. They may be synthesized either by a divergent or by a convergent strategy (Figure 7.12). The divergent strategy makes use of the assembly starting from the initiator core to the periphery. Each new layer of monomer units attached to the growing molecule is called a “generation”. In the convergent strategy, the assembly of dendrimers starts from the periphery and continues towards the central core [257].

Tam *et al.* [258] utilized branched oligolysine cores as synthetic carriers of fully synthetic antigens. The *multiple antigen peptides* (MAP) developed by Tam *et al.* comprise a simple amino acid such as glycine or β -alanine as an internal standard, an inner oligolysine core, and multiple copies of the synthetic peptide antigen (Figure 7.13). One advantage of MAP is that they have a well-defined and high antigen: carrier ratio. The derivatives are stable, and no conjugation to an antigenic protein is necessary; hence, undesired epitopes are not present. MAP have proven in

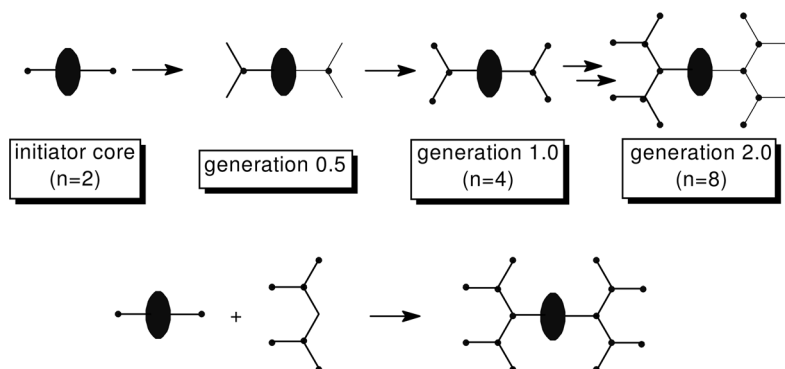


Figure 7.12 Divergent and convergent synthesis of dendrimers.

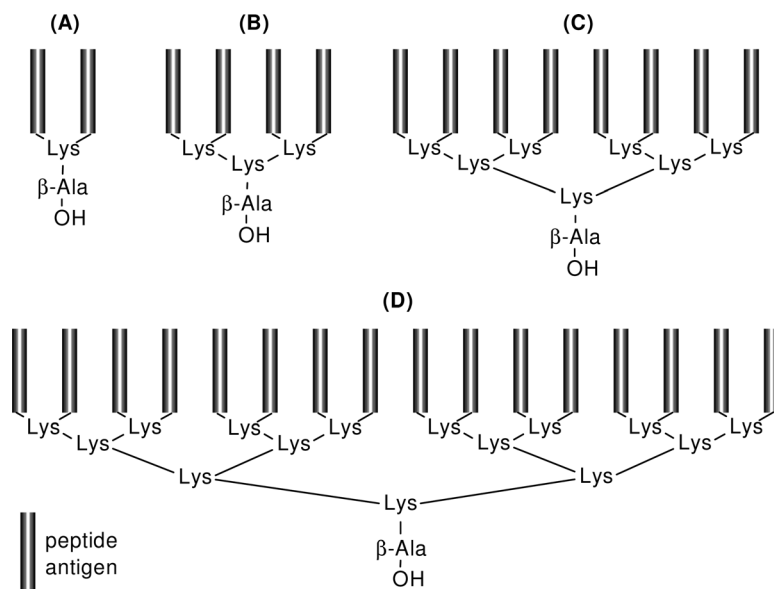


Figure 7.13 Schematic structure of (A) di-, (B) tetra-, (C) octa-, and (D) hexadeca-valent multiple antigen peptides (MAP).

numerous cases to be useful in diagnostic and therapeutic applications as well as for antibody production [255].

Dendrimers may be synthesized either in solution or on a solid phase. While the former approach is often very challenging and requires long reaction times and nontrivial purification steps, the application of solid-phase strategies provides the possibility of driving the reactions to completion by use of a large excess of reagents. Furthermore, a distinction may be made between direct and indirect approaches for the synthesis (Figure 7.14).

In the direct approach, multiple copies of the peptide antigen are synthesized by stepwise solid-phase peptide synthesis (SPPS) on a solid-phase-bound dendrimer core. Both Fmoc and Boc tactics have been successfully applied [259, 260]. Low resin loading is necessary in the synthesis of peptide dendrimers in order to minimize interchain interaction that may limit coupling efficiency [261]. The efficiency must be monitored, and when necessary repetitive couplings should be performed using different activating agents. Similarly, galactose has been used as a core moiety, where four identical peptide strands were assembled by solid-phase synthesis [262].

The indirect approach (Figure 7.14) is characterized by the separate synthesis of a suitably functionalized core to where the full-length peptide is ligated in the final step. A classical fragment condensation in solution using fully protected peptides may be applied, but this often suffers from low yield, slow coupling reactions, poor solubility, and problems of racemization. On the other hand, the possibility of purifying the peptide segments before ligation excludes the occurrence of mismatched sequences.

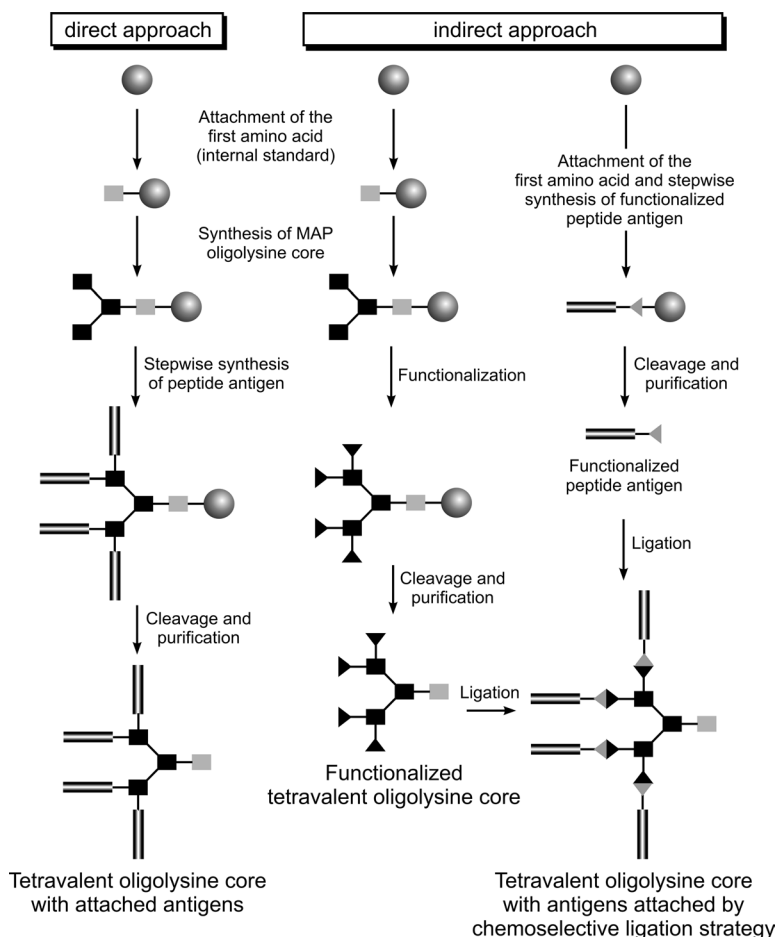


Figure 7.14 Methods for the preparation of multiple antigen peptides (MAP).

Alternatively, unprotected or partially protected peptide epitopes, prepared by SPPS, may be conjugated to the dendrimer core in aqueous solution under mild conditions. In this approach, either thiol chemistry, hydrazone chemistry, or oxime chemistry [263] may be applied (for further information, see [253, 256]).

Using this approach, either lipidated or glycosylated peptide dendrimers may also be synthesized [253, 264, 265]. Glycopeptide dendrimers are branched structures that contain both carbohydrate and peptide moieties [266, 267].

Cationic dendrimers may also be used for gene delivery because they form stable complexes with DNA. Polylysine dendrimers containing a polyethylene glycol moiety have been reported to form a spherical water-soluble complex with DNA [268].

Peptide dendrimers containing a porphyrin core form another important group of functional dendrimers that may be used as either chemosensors or catalysts [269–271].

7.5.3

Peptide Polymers

As oligo- or polypeptides are polymers by definition, three other types of polymers will be discussed in this section. The first of these comprises homopolymers composed of only one type of amino acid, which may be obtained from N-carboxy anhydrides [272]. In particular, polylysine or polyarginine, both of which are present under physiological conditions as polycations, have a unique ability to cross the plasma membrane of cells [273]. Consequently, they may be used to transport a variety of biopolymers and small molecules into cells [274]. For example, polylysine interacts electrostatically with the negatively charged phosphate backbone of DNA and may, therefore, be used for gene transfer [275].

Polylysine is synthesized by polymerization of the N-carboxy anhydride of lysine, and this is subsequently fractionated and characterized with respect to the average molecular weight [276]. However, the heterogeneity of commercially available polylysine in terms of degree of polymerization is a major obstacle in the preparation of reproducible, stable formulations. Therefore, rationally designed synthetic peptides have been suggested as DNA delivery systems.

A second type of peptide-based polymer (also called protein-based polymers or sequential polypeptides) is formed by compounds composed of repeating peptide sequences. In contrast to other polymers, these can be prepared either by chemical synthesis (solution or solid phase) or alternatively by recombinant technology.

Peptide-based polymers can be transformable hydrogels, elastomers, regular thermoplasts, or inverse thermoplasts that are postulated to be molecular machines [277].

The third type of peptide polymers comprises peptide drugs that are chemically conjugated to nanoscale polymer particles such as hydroxypropylmethacrylamide or polyethyleneglycol; this may be carried out using a formulation approach in order to enhance drug stability and targeting possibilities. It has been shown that a vinyl acetate derivative of luteinizing hormone-releasing hormone (LHRH) can be copolymerized with butylcyanoacrylate to form particles of a mean size of 100 nm that are stable *in vitro*. In this way, the average half-life of LHRH in blood could be increased from 2–8 min (peptide) up to 12 h [278].

Co-polymers of peptides and polyethyleneglycol or polyacrylates have an intermediate position between the second and third types referred to above [279, 280].

7.6

Review Questions

- Q7.1. Name four disadvantages of peptides that usually prevent direct application as a drug.
- Q7.2. Name three different classes of peptidomimetics according to Ripka and Rich.

- Q7.3. What is ADME and why is it important in the drug development process?
- Q7.4. What is SAR by NMR?
- Q7.5. What types of backbone modification do you know?
- Q7.6. What are retro-inverso peptides?
- Q7.7. What are secondary structure mimetics?
- Q7.8. What is a transition state inhibitor?
- Q7.9. When designing a peptoid on the basis of a bioactive peptide, the sequence should be inverted. Why?
- Q7.10. What is the sub-monomer synthesis of peptoids? How could it be applied to synthesize β -peptoids?
- Q7.11. What is the TASP concept?
- Q7.12. Describe the approach and utilisation of MAP (multiple antigenic peptides)?

References

- 1 B.A. Morgan, J.A. Gainor, *Annu. Rep. Med. Chem.* **1989**, 24, 243.
- 2 R.M.J. Liskamp, *Recl. Trav. Chim. Pays-Bas* **1994**, 113, 1.
- 3 V.J. Hruby, *Biopolymers* **1993**, 33, 1073.
- 4 A. Giannis, T. Kolter, *Angew. Chem. Int. Ed.* **1993**, 32, 1244.
- 5 R.A. Wiley, D.H. Rich, *Med. Res. Rev.* **1993**, 13, 327.
- 6 A.F. Spatola, in: *Methods in Neurosciences*, Volume 13, P.M. Conn (Ed.), Academic Press, San Diego, **1993**, p. 19.
- 7 J. Gante, *Angew. Chem. Int. Ed.* **1994**, 33, 1699.
- 8 C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv. Drug. Del. Rev.* **1997**, 23, 3.
- 9 M.W. Peczu, A.D. Hamilton, *Chem. Rev.* **2000**, 100, 2479.
- 10 T. Berg, *Angew. Chem. Int. Ed.* **2003**, 42, 2462.
- 11 S. Fletcher, A.D. Hamilton, *Curr. Opin. Chem. Biol.* **2005**, 9, 632.
- 12 H. Yin, A.D. Hamilton, *Angew. Chem. Int. Ed.* **2005**, 44, 4130.
- 13 S.J. Hersberger, S. -G. Lee, J. Chmielewski, *Curr. Top. Med. Chem.* **2007**, 7, 928.
- 14 J.K. Murray, S.H. Gellman, *Biopolymers* **2007**, 88, 657.
- 15 S.E. Rutledge, H.M. Volkman, A. Schepartz, *J. Am. Chem. Soc.* **2003**, 125, 14336.
- 16 J. Eichler, *Comb. Chem. High Throughput Screen.* **2005**, 8, 135.
- 17 R. Franke, T. Hirsch, H. Overwin, J. Eichler, *Angew. Chem. Int. Ed.* **2007**, 46, 1253.
- 18 V.J. Hruby, P.M. Balse, *Curr. Med. Chem.* **2000**, 7, 945.
- 19 J. Boer, D. Gottschling, A. Schuster, B. Holzmann, H. Kessler, *Angew. Chem. Int. Ed.* **2001**, 40, 3870.
- 20 A.S. Ripka, D.H. Rich, *Curr. Opin. Chem. Biol.* **1998**, 2, 441.
- 21 R.E. Babine, S.L. Bender, *Chem. Rev.* **1997**, 97, 1359.
- 22 R.S. Bohacek, C. McMartin, *Curr. Opin. Chem. Biol.* **1997**, 1, 157.
- 23 K.S. Lam, M. Lebl, V. Krchnák, *Chem. Rev.* **1997**, 97, 411.
- 24 K.H. Mayo, *Trends Biotechnol.* **2000**, 18, 212.
- 25 R. Haubner, D. Finsinger, H. Kessler, *Angew. Chem. Int. Ed.* **1997**, 36, 1374.
- 26 J.B. Ball, R.A. Hughes, P.F. Alewood, P. R. Andrews, *Tetrahedron* **1993**, 49, 3467.
- 27 R.S. McDowell, T.R. Gadek, P.L. Barke, D.L. Burdick, K.S. Chan, C.L. Quan, N. Skelton, M. Struble, E.D. Thorsett, M.

- Tischler, J.Y. Tom, T.R. Webb, J.P. Burnier, *J. Am. Chem. Soc.* **1994**, *116*, 5069.
- 28 R.S. McDowell, B.K. Blackburn, T.R. Gadek, L.R. McGee, T. Rawson, M.E. Reynolds, K.D. Robarge, T.C. Somers, E.D. Thorsett, M. Tischler, R.R. Webb II, M.C. Venuti, *J. Am. Chem. Soc.* **1994**, *116*, 5077.
- 29 W.G.J. Hol, *Angew. Chem. Int. Ed.* **1986**, *25*, 767.
- 30 L.M. Amzel, *Curr. Opin. Biotechnol.* **1998**, *9*, 366.
- 31 W. Wang, O. Donini, C.M. Reyes, P.A. Kollman, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.
- 32 R.E. Bruccoleri, M. Karplus, *Biopolymers* **1987**, *26*, 137.
- 33 W.F. van Gunsteren, *Protein Eng.* **1988**, *2*, 5.
- 34 T.J. Marrone, J.M. Briggs, J.A. McCammon, *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71.
- 35 G. Böhm, *Biophys. Chem.* **1996**, *59*, 1.
- 36 P.J. Whittle, T.L. Blundell, *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 349.
- 37 H.J. Böhm, *Prog. Biophys. Mol. Biol.* **1996**, *66*, 197.
- 38 S.M. Lippow, B. Tidor, *Curr. Opin. Biotechnol.* **2007**, *18*, 305.
- 39 K. Gubernator, H. -J. Böhm, (Eds.), *Structure-based Ligand Design*, Wiley-VCH, Weinheim, **1998**.
- 40 C. Quan, N.J. Skelton, K. Clark, D.Y. Jackson, M.E. Renz, H.H. Chiu, S.M. Keating, M.H. Beresini, S. Fong, D.R. Artis, *Biopolymers* **1998**, *47*, 265.
- 41 D.J. Bacon, J. Moulton, *J. Mol. Biol.* **1992**, *225*, 849.
- 42 V. Sobolev, R.C. Wade, G. Vriend, M. Edelman, *Proteins* **1996**, *25*, 120.
- 43 M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261*, 470.
- 44 M. Rarey, B. Kramer, T. Lengauer, *J. Comput. Aid. Mol. Des.* **1997**, *11*, 369.
- 45 S.Y. Yue, *Protein Eng.* **1990**, *4*, 177.
- 46 G.M. Morris, D.S. Goodsell, R. Huey, A.J. Olson, *J. Comput. Aid. Mol. Des.* **1996**, *10*, 293.
- 47 M. Liu, S. Wang, *J. Comput. Aid. Mol. Des.* **1999**, *13*, 435.
- 48 C.M. Oshiro, I.D. Kuntz, J.S. Dixon, *J. Comput. Aid. Mol. Des.* **1995**, *9*, 113.
- 49 G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727.
- 50 S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, *Science* **1996**, *274*, 1531.
- 51 T. Diercks, M. Coles, H. Kessler, *Curr. Opin. Chem. Biol.* **2001**, *5*, 285.
- 52 M.L. Moore, in: *Synthetic Peptides: A User's Guide*, G.A. Grant (Ed.), Freeman, New York, **1992**, p. 9.
- 53 D. Ward (Ed.), *Peptide Pharmaceuticals, Approaches to the Design of Novel Drugs*, Open University Press, Milton Keynes, **1991**.
- 54 G.N. Ramachandran, V. Sasisekharan, *Adv. Protein Chem.* **1968**, *23*, 283.
- 55 H.A. Scheraga, *Chem. Rev.* **1971**, *71*, 195.
- 56 S.M. Bloom, G.D. Fasman, C. Deloze, E. R. Blout, *J. Am. Chem. Soc.* **1961**, *84*, 458.
- 57 M. Goodman, R.P. Saltman, *Biopolymers* **1981**, *20*, 1929.
- 58 H. Kessler, *Angew. Chem. Int. Ed.* **1982**, *21*, 512.
- 59 W.R. Croasmun, R.M.K. Carlson, *Two-Dimensional NMR-Spectroscopy: Application for Chemists and Biochemists*, VCH, Weinheim, **1994**.
- 60 D.J. Craik, N.L. Daly, *Mol. BioSyst.* **2007**, *3*, 257.
- 61 L. Moroder, *J. Peptide Sci.* **2005**, *11*, 187.
- 62 M. Manning, L. Balaspiri, M. Agosta, *J. Med. Chem.* **1973**, *16*, 975.
- 63 T. Haack, M. Mutter, *Tetrahedron Lett.* **1992**, *33*, 1589.
- 64 P. Dumy, M. Keller, D.E. Ryan, B. Rohwedder, T. Wöhr, M. Mutter, *J. Am. Chem. Soc.* **1997**, *119*, 918.
- 65 M. Keller, C. Sager, P. Dumy, M. Schutkowski, G.S. Fischer, M. Mutter, *J. Am. Chem. Soc.* **1998**, *120*, 2714.
- 66 A. Wittelsberger, M. Keller, L. Scarpellino, P. Patiny, H. Acha-Orbea, M. Mutter, *Angew. Chem. Int. Ed.* **2000**, *39*, 1111.

- 67 Y. Takeuchi, G.R. Marshall, *J. Am. Chem. Soc.* **1998**, 120, 5363.
- 68 H.E. Stanger, S.H. Gellman, *J. Am. Chem. Soc.* **1998**, 120, 4236.
- 69 V.J. Hruby, S. Fang, R. Knapp, W.M. Kazmierski, G.K. Lui, H.I. Yamamura, *Int. J. Peptide Protein Res.* **1990**, 35, 566.
- 70 J. Chatterjee, C. Gilon, A. Hoffman, H. Kessler, *Acc. Chem. Res.* **2008**, 44, 1331.
- 71 P. Balaram, T.S. Sudha, *Int. J. Peptide Protein Res.* **1983**, 21, 381.
- 72 I.L. Karle, *Biopolymers* **1996**, 40, 157.
- 73 B. Pispisa, L. Stella, M. Venzani, A. Palleschi, F. Marchiori, A. Polese, C. Toniolo, *Biopolymers* **2000**, 53, 169.
- 74 V.J. Hruby, G. Li, C. Haskell-Luevano, M. Shenderovich, *Biopolymers* **1997**, 43, 219.
- 75 S.E. Gibson, N. Guillo, M.J. Tozer, *Tetrahedron* **1999**, 55, 585.
- 76 S.M. Cowell, Y.S. Lee, J.P. Cain, V.J. Hruby, *Curr. Med. Chem.* **2004**, 11, 2785.
- 77 P. Mathur, S. Ramakumar, V.S. Chauhan, *Biopolymers* **2004**, 76, 150.
- 78 J.-M. Ahn, N.A. Boyle, M.T. MacDonald, K.D. Janda, *Mini-Rev. Med. Chem.* **2002**, 2, 463.
- 79 J. Gante, *Angew. Chem. Int. Ed.* **1994**, 33, 1699.
- 80 P. Manavalan, F.A. Momany, *Biopolymers* **1980**, 19, 1943.
- 81 J.V.N. Vara Prasad, D.H. Rich, *Tetrahedron Lett.* **1990**, 31, 1803.
- 82 J. Gante, *Synthesis* **1989**, 405.
- 83 J. Magrath, R.H. Abeles, *J. Med. Chem.* **1992**, 35, 4279.
- 84 K. Clausen, M. Thorsen, S.O. Lawesson, A.F. Spatola, *J. Chem. Soc., Perkin Trans. I* **1984**, 785.
- 85 L. El Masdouri, A. Aubry, C. Sakarellos, E.J. Gomex, M.T. Cung, M. Marraud, *Int. J. Peptide Protein Res.* **1988**, 31, 420.
- 86 R.M.J. Liskamp, J.A.W. Kruijtz, *Mol. Div.* **2004**, 8, 79.
- 87 M. Chorev, M. Goodman, *Acc. Chem. Res.* **1993**, 26, 266.
- 88 P.M. Fischer, *Curr. Protein Pept. Sci.* **2003**, 4, 339.
- 89 M. Szelke, D.M. Jones, B. Atrash, A. Hallett, B.J. Leckie, in: *Peptides: Structure and Function*, V.J. Hruby, D. H. Rich (Eds.), Pierce, Rockford, **1980**, 579.
- 90 M.M. Hann, P.G. Sammes, P.D. Kennewell, J.B. Taylor, *J. Chem. Soc., Chem. Commun.* **1980**, 234.
- 91 M. Kranz, H. Kessler, *Tetrahedron Lett.* **1996**, 37, 5359.
- 92 D. Yang, J. Qu, B. Li, F. -F. Ng, X. -C. Wang, K.K. Cheung, D. -P. Wang, Y. -D. Wu, *J. Am. Chem. Soc.* **1999**, 121, 589.
- 93 X. Li, D. Yang, *Chem. Commun.* **2006**, 3367.
- 94 A. Aubry, J. -P. Mangeot, J. Vidal, A. Collet, S. Zerkout, M. Marraud, *Int. J. Peptide Protein Res.* **1994**, 43, 305.
- 95 A. Cheguillaume, F. Lehardy, K. Bouget, M. Baudy-Floc'h, P. Le Grel *J. Org. Chem.* **1999**, 64, 2924.
- 96 A. Müller, C. Vogt, N. Sewald, *Synthesis* **1998**, 837.
- 97 A. Müller, F. Schumann, M. Koksche, N. Sewald, *Let. Peptide Sci.* **1997**, 4, 275.
- 98 F. Schumann, A. Müller, M. Koksche, G. Müller, N. Sewald, *J. Am. Chem. Soc.* **2000**, 122, 12009.
- 99 M. Hagihara, S.L. Schreiber, *J. Am. Chem. Soc.* **1992**, 114, 6570.
- 100 W. Bauer, U. Boiner, W. Haller, R. Huguenin, R. Marbach, P. Petcher, J. Pless, in: *Peptides 1982*, K. Blaha, P. Malon (Eds.), de Gruyter, Berlin, **1983**.
- 101 D.F. Veber, R.M. Freidinger, D.S. Prelow, W.J. Poleveda, F.M. Holly, R.G. Strachan, R.F. Nutt, B.H. Arison, C. Homnick, W.C. Randall, M.S. Glitzer, R. Saperstein, R. Hirschmann, *Nature* **1981**, 292, 55.
- 102 T.K. Sayer, V.J. Hruby, P.S. Darman, M.E. Hadley, *Proc. Natl. Acad. Sci. USA* **1982**, 79 1751.
- 103 K.D. Stigers, M.J. Soth, J.S. Nowick, *Curr. Opin. Chem. Biol.* **1999**, 3, 714.
- 104 S. Hanessian, G. McNaughton-Smith, H.-G. Lombart, W.D. Lubell, *Tetrahedron* **1997**, 53, 12789.

- 105 P. Gillespie, J. Cicariello, G.L. Olson, *Biopolymers* **1997**, 43, 191.
- 106 L. Halab, F. Gosselin, W.D. Lubell, *Biopolymers* **2000**, 55, 101.
- 107 M. Eguchi, M. Kahn, *Mini Rev. Med. Chem.* **2002**, 2, 447.
- 108 G.D. Rose, L.M. Gierasch, J.A. Smith, *Adv. Protein Chem.* **1985**, 37, 1.
- 109 C. Nagai, K. Sato, *Tetrahedron Lett.* **1982**, 23, 3759.
- 110 V. Brandmeier, M. Feigel, *Tetrahedron* **1989**, 45, 1365.
- 111 D.S. Kemp, W.E. Stites, *Tetrahedron Lett.* **1988**, 29, 5057.
- 112 M.J. Genin, R.L. Johnson, *J. Am. Chem. Soc.* **1992**, 114, 8778.
- 113 W.C. Ripka, G.V. DeLucca, A.C. Bach II, R.S. Pottorf, J.M. Blaney, *Tetrahedron* **1993**, 49, 3593.
- 114 M. Feigel, *Liebigs Ann. Chem.* **1989**, 459.
- 115 M. Feigel, G. Lugert, C. Heichert, *Liebigs Ann. Chem.* **1987**, 367.
- 116 D.S. Kemp, E.T. Sun, *Tetrahedron Lett.* **1982**, 23, 3759.
- 117 R. Haubner, W. Schmitt, G. Hölzemann, S.L. Goodman, A. Jonczyk, H. Kessler, *J. Am. Chem. Soc.* **1996**, 118, 7881.
- 118 G. Müller, G. Hessler, H.Y. Decornez, *Angew. Chem. Int. Ed.* **2000**, 39, 894.
- 119 J.F. Callahan, J.W. Bean, J.L. Burgess, D.S. Eggleston, S.M. Hwang, K.D. Kopple, P.F. Koster, A. Nichols, C.E. Peishoff, J.M. Samanen, J.A. Vasko, A. Wong, W.F. Huffman, *J. Med. Chem.* **1992**, 35, 3970.
- 120 J.F. Callahan, K.A. Newlander, J.L. Burgess, D.S. Eggleston, A. Nichols, A. Wong, W.F. Huffman, *Tetrahedron* **1993**, 49, 3479.
- 121 M. Sato, J.Y.H. Lee, H. Nakanishi, M.E. Johnson, R.A. Chrusciel, M. Kahn, *Biochem. Biophys. Res. Commun.* **1992**, 187, 1000.
- 122 D.S. Kemp, T.P. Curran, J.G. Boyd, T.J. Allen, *Biochem. Biophys. Res. Commun.* **1991**, 56, 6683.
- 123 D.S. Kemp, T.P. Curran, W.M. Davis, J.G. Boyd, C. Muendel, *J. Org. Chem.* **1991**, 56, 6672.
- 124 R. Sarabu, K. Lovey, V.S. Madison, D.C. Fry, D.W. Greeley, C.M. Cook, G.L. Olson, *Tetrahedron* **1993**, 49, 3629.
- 125 L. Pauling, *Chem. Eng. News* **1946**, 24, 1375.
- 126 R. Wolfenden, *Bioorg. Med. Chem.* **1999**, 7, 647.
- 127 H. -J. Böhm, G. Klebe, H. Kubinyi, *Wirkstoffdesign*, Spektrum Verlag, Heidelberg, Berlin, Oxford, **1996**.
- 128 M.L. Moore, G.B. Dreyer, *J. Comput. Aided Mol. Des.* **1993**, 85.
- 129 K.D. Janda, D. Schloeder, S.J. Benkovic, R.A. Lerner, *Science* **1988**, 241, 1188.
- 130 J.D. Stewart, P.A. Benkovic, *Chem. Soc. Rev.* **1993**, 213.
- 131 T. Kieber-Emmons, R. Murali, M.I. Green, *Curr. Opin. Biotechnol.* **1997**, 8, 435.
- 132 D.C. Horwell, *Trends Biotechnol.* **1995**, 13, 132.
- 133 D.F. Veber, in: *Peptides: Chemistry and Biology*, J.A. Smith, J.E. Rivier (Eds.), Escom, Leiden, **1992**, 3.
- 134 B.E. Evans, M.G. Bock, K.E. Rittle, R.M. DiPrado, W.L. Whitter, D.F. Veber, P.S. Anderson, R.M. Freidinger, *Proc. Natl. Acad. Sci. USA* **1986**, 83, 4918.
- 135 M.A. Ondetti, B. Rubin, D.W. Cushman, *Science* **1977**, 196, 441.
- 136 L.D. Byers, R. Wolfenden, *J. Biol. Chem.* **1972**, 247, 606.
- 137 L.D. Byers, R. Wolfenden, *Biochemistry* **1973**, 12, 2070.
- 138 D.H. Rich, *Compr. Med. Chem.* **1990**, 2, 400.
- 139 L. Turbanti, G. Cerbai, C. Di Bugno, R. Giorgi, G. Garzelli, M. Criscuoli, A. R. Renzetti, A. Subissi, G. Bramante, *J. Med. Chem.* **1993**, 36, 699.
- 140 D. Römer, H.H. Büscher, R.C. Hill, R. Maurer, T.J. Petcher, H. Zeugner, W. Benson, E. Finner, W. Milkowski, P.W. Thies, *Nature* **1982**, 298, 759.
- 141 L.A. Dykstra, D.E. Gmerek, G. Winger, J.H. Woods, *J. Pharmacol. Exp. Ther.* **1987**, 242, 413.
- 142 A. Peyman, *Chem. Rev.* **1990**, 90, 543.
- 143 S.H. Gellman, *Acc. Chem. Res.* **1998**, 31, 173.

- 144 K. Kirshenbaum, R.N. Zuckermann, K.A. Dill, *Curr. Opin. Struct. Biol.* **1999**, 9, 530.
- 145 J.S. Nowick, *Acc. Chem. Res.* **1999**, 32, 287.
- 146 D.J. Hill, M.J. Mio, R.B. Prince, T.S. Hughes, J.S. Moore, *Chem. Rev.* **2001**, 101, 3893.
- 147 R.J. Simon, R.S. Kania, R.N. Zuckermann, V.D. Huebner, D.A. Jewell, S. Banville, S. Wang, S. Rosenberg, C.K. Marlowe, D.C. Spellmeyer, R. Tan, A.D. Frankel, D.V. Santi, F.E. Cohen, P.A. Bartlett, *Proc. Natl. Acad. Sci. USA* **1992**, 89, 9367.
- 148 P.E. Nielsen, *Acc. Chem. Res.* **1999**, 32, 624.
- 149 J.M. Fernandez-Santn, J. Aymam, A. Rodriguez-Galn, S. Muoz-Guerra, J.A. Subirana, *Nature* **1984**, 311, 53.
- 150 D.H. Appella, L.A. Christianson, I.L. Karle, D.R. Powell, S.H. Gellman, *J. Am. Chem. Soc.* **1996**, 118, 13071.
- 151 D.H. Appella, L.A. Christianson, D.A. Klein, D.R. Powell, X. Huang, J.J. Barchi Jr., S.H. Gellman, *Nature* **1997**, 387, 381.
- 152 D. Seebach, P.E. Ciceri, M. Overhand, B. Jaun, D. Rigo, L. Oberer, U. Hommel, R. Amstutz, H. Widmer, *Helv. Chim. Acta* **1996**, 79, 2043.
- 153 D. Seebach, J.L. Matthews, *J. Chem. Soc., Chem. Commun.* **1997**, 2015.
- 154 W.F. DeGrado, J.P. Schneider, Y. Hamuro, *J. Peptide Res.* **1999**, 54, 206.
- 155 H. Han, K.D. Janda, *J. Am. Chem. Soc.* **1996**, 118, 2539.
- 156 H. Han, J. Yoon, K.D. Janda, *Bioorg. Med. Chem. Lett.* **1998**, 8, 117.
- 157 C. Gennari, M. Gude, D. Potenza, U. Piarulli, *Chem. Eur. J.* **1998**, 4, 1924.
- 158 R.M.J. Liskamp, J.A.W. Kruijtzter, *Mol. Div.* **2004**, 8, 79.
- 159 R. Günther, H.J. Hofmann, *J. Am. Chem. Soc.* **2001**, 123, 247.
- 160 C.Y. Cho, E.J. Moran, S.R. Cherry, J.C. Stephans, S.P.A. Fodor, C.L. Adams, A. Sundaram, J.W. Jacobs, P.G. Schultz, *Science* **1993**, 261, 1303.
- 161 A.B. Smith, III, T.P. Keenan, R.C. Holcomb, P.A. Sprengeler, M.C. Guzman, J.L. Wood, P.C. Carroll, R. Hirschmann, *J. Am. Chem. Soc.* **1992**, 114, 10672.
- 162 S. Hanessian, X. Luo, R. Schaum, *Tetrahedron Lett.* **1999**, 40, 4925.
- 163 D. Seebach, A.K. Beck, D.J. Bierbaum, *Chem. Biodiv.* **2004**, 1, 1111.
- 164 M. Hagihara, N.J. Anthony, T.J. Stout, J. Clardy, S.L. Schreiber, *J. Am. Chem. Soc.* **1992**, 114, 6568.
- 165 C. Baldauf, R. Günther, H. -J. Hofmann, *J. Org. Chem.* **2005**, 70, 5351.
- 166 C. Grison, P. Coutrot, S. Genève, C. Didierjean, M. Marraud, *J. Am. Chem. Soc.* **1992**, 114, 6568.
- 167 C. Gennari, C. Longari, S. Ressel, B. Salom, U. Piarulli, S. Ceccarelli, A. Mielgo, *Eur. J. Org. Chem.* **1998**, 2437.
- 168 M.D. Smith, T.D.W. Claridge, G.E. Tranter, M.S.P. Sansom, G.W.J. Fleet, *J. Chem. Soc., Chem. Commun.* **1998**, 2041.
- 169 E. Locardi, M. Stöckle, S. Gruner, H. Kessler, *J. Am. Chem. Soc.* **2001**, 123, 8189.
- 170 J. Frackenpohl, P.I. Arvidsson, J.V. Schreiber, D. Seebach, *ChemBioChem.* **2001**, 2, 445.
- 171 R.N. Zuckermann, J.M. Kerr, S.B.H. Kent, W.H. Moos, *J. Am. Chem. Soc.* **1992**, 114, 10646.
- 172 H. Kessler, *Angew. Chem. Int. Ed.* **1993**, 32, 543.
- 173 K. Kirshenbaum, A.E. Barron, R.A. Goldsmith, P. Armand, E.K. Bradley, K.T.V. Truong, K.A. Dill, F.E. Cohen, R.N. Zuckermann, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 4303.
- 174 P. Armand, K. Kirshenbaum, R.A. Goldsmith, S. Farr-Jones, A.E. Barron, K.T.V. Truong, K.A. Dill, D.F. Mierke, F.E. Cohen, R.N. Zuckermann, E.K. Bradley, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 4309.
- 175 B.C. Hamper, S.A. Kolodziej, A.M. Scates, R.G. Smith, E. Cortez, *J. Org. Chem.* **1998**, 63, 708.
- 176 A.S. Norgren, S. Zhang, P.I. Arvidsson, *Org. Lett.* **2006**, 8, 4533.

- 177 C.A. Olsen, G. Bonke, L. Vedel, A. Adersen, M. Witt, H. Franzyk, J.W. Jaroszewski, *Org. Lett.* **2007**, 9, 1549.
- 178 L. Vedel, G. Bonke, C. Foged, H. Ziegler, H. Franzyk, J.W. Jaroszewski, C.A. Olsen, *ChemBioChem* **2007**, 8, 1781.
- 179 C. Baldauf, R. Günther, H. -J. Hofmann, *Phys. Biol.* **2006**, 3, S1.
- 180 N.P. Chongsiriwatana, J.A. Patch, A.M. Czyzewski, M.T. Dohm, A. Ivankin, D. Gidalevitz, R.N. Zuckermann, A.E. Barron, *Proc. Natl. Acad. Sci. USA* **2008**, 105, 2794.
- 181 P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, *Science* **1991**, 25, 1497.
- 182 M. Egholm, O. Buchardt, P.E. Nielsen, R.H. Berg, *J. Am. Chem. Soc.* **1992**, 114, 1895.
- 183 C. Meier, J.W. Engels, *Angew. Chem. Int. Ed.* **1992**, 31, 1008.
- 184 H. De Koning, U.K. Pandit, *Recl. Trav. Chim. Pays-Bas* **1971**, 90, 1069.
- 185 J.D. Buttrey, A.S. Jones, R.T. Walker, *Tetrahedron* **1975**, 31, 73.
- 186 A. Porcheddu, G. Giacomelli, *Curr. Med. Chem.* **2005**, 12, 2561.
- 187 S. Pensato, M. Saviano, A. Romanelli, *Expert Opin. Biol. Ther.* **2007**, 7, 1219.
- 188 S. Thurley, L. Röglin, O. Seitz, *J. Am. Chem. Soc.* **2007**, 129, 12693.
- 189 K. Petersen, U. Vogel, E. Rockenbauer, K.V. Nielsen, S. Kølvræ, L. Bolund, B. Nexø, *Mol. Cell Probes* **2004**, 18, 117.
- 190 C. Dose, O. Seitz, *Bioorg. Med. Chem.* **2008**, 16, 65.
- 191 S.A. Thomson, J.A. Josey, R. Cadilla, M.D. Gaul, C.F. Hassman, M.J. Luzzio, A.J. Pipe, K.L. Reed, D.J. Ricca, R.W. Wiethe, S.A. Noble, *Tetrahedron* **1995**, 51, 6179.
- 192 J. Kovacs, R. Ballina, R. Rodin, O. Balasubramanian, J. Applequist, *J. Am. Chem. Soc.* **1965**, 87, 119.
- 193 F. López-Carrasquero, M. García-Alvarez, J.J. Navas, C. Alemán, S. Muñoz-Guerra, *Macromolecules* **1996**, 29, 8449.
- 194 D.H. Appella, J.J. Barchi, S.R. Durell, S.H. Gellman, *J. Am. Chem. Soc.* **1999**, 121, 2309.
- 195 P.I. Arvidsson, M. Rueping, D. Seebach, *J. Chem. Soc., Chem. Commun.* **2001**, 649.
- 196 R.P. Cheng, W.F. DeGrado, *J. Am. Chem. Soc.* **2001**, 123, 5162.
- 197 X. Daura, W.F. van Gunsteren D. Rigo, B. Jaun, D. Seebach, *Chem. Eur. J.* **1997**, 3, 1410.
- 198 G. Lelais, D. Seebach, *Biopolymers* **2004**, 76, 206.
- 199 D. Seebach, K. Gademann, J.V. Schreiber, J.L. Matthews, T. Hintermann, B. Jaun, L. Oberer, U. Hommel, H. Widmer, *Helv. Chim. Acta* **1997**, 80, 2033.
- 200 C. Baldauf, R. Günther, H. -J. Hofmann, *Angew. Chem. Int. Ed.* **2004**, 43, 1594.
- 201 E. Juaristi, V.A. Soloshonok, (Eds.), *Enantioselective Synthesis of β -Amino Acids*, Wiley-Interscience, New York, **2005**.
- 202 J. Podlech, D. Seebach, *Liebigs Ann. Chem.* **1995**, 1217.
- 203 D. Seebach, D.F. Hook, A. Glättli, *Biopolymers* **2006**, 84, 23.
- 204 F. Fülöp, T.A. Martinek, G.K. Tóth, *Chem. Soc. Rev.* **2006**, 35, 323.
- 205 K. Gademann, M. Ernst, D. Hoyer, D. Seebach, *Angew. Chem. Int. Ed.* **1999**, 38, 1223.
- 206 D. Seebach, J. Gardinier, *Acc. Chem. Res.* **2008**, 41, 1366.
- 207 E.A. Porter, X. Wang, H.S. Lee, B. Weisblum, S.H. Gellman, *Nature* **2000**, 404, 565.
- 208 W.S. Horne, S.H. Gellman, *Acc. Chem. Res.* **2008**, 41, 1399.
- 209 J.X. Qiu, E.J. Petersson, E.E. Matthews, A. Schepartz, *J. Am. Chem. Soc.* **2006**, 128, 11338.
- 210 D. Yang, Y. -H. Zhang, B. Li, D. -W. Zhang, *J. Org. Chem.* **2004**, 69, 7577.
- 211 J. Lawrence, L. Cointeaux, P. Maire, Y. Vallée, V. Blandin, *Org. Biomol. Chem.* **2006**, 4, 3125.
- 212 P. Balaran, *J. Peptide Res.* **1999**, 54, 195.
- 213 W.F. DeGrado, C.M. Summa, V. Pavone, F. Nastro, A. Lombardi, *Annu. Rev. Biochem.* **1999**, 68, 779.
- 214 R.B. Hill, D.P. Raleigh, A. Lombardi, W.F. DeGrado, *Acc. Chem. Res.* **2000**, 33, 745.

- 215 G. Tuchscherer, L. Scheibler, P. Dumy, M. Mutter, *Biopolymers* **1998**, 47, 63.
- 216 G. Tuchscherer, G. Grell, M. Mathieu, M. Mutter, *J. Peptide Res.* **1999**, 54, 185.
- 217 J. Fernandez-Carneado, D. Grell, P. Durieux, J. Hauert, T. Kovacsics, G. Tuchscherer, *Biopolymers* **2000**, 55, 451.
- 218 L. Baltzer, H. Nilsson, J. Nilsson, *Chem. Rev.* **2001**, 101, 3153.
- 219 P. Koehl, M. Levitt, *J. Mol. Biol.* **1999**, 293, 1161.
- 220 P. Koehl, M. Levitt, *J. Mol. Biol.* **1999**, 293, 1183.
- 221 D.N. Woolson, *Curr. Opin. Struct. Biol.* **2001**, 11, 464.
- 222 H.-B. Kraatz, *Angew. Chem. Int. Ed.* **1994**, 33, 2055.
- 223 M. Favre, K. Moehle, L. Jiang, B. Pfeiffer, J.A. Robinson, *J. Am. Chem. Soc.* **1999**, 121, 2679.
- 224 L. Jiang, K. Moehle, B. Dhanapal, D. Obrecht, J.A. Robinson, *Helv. Chim. Acta* **2001**, 83, 3097.
- 225 J.M. Travins, F.A. Etzkorn, *J. Org. Chem.* **1997**, 8387.
- 226 A. Banerjee, S. Raghothama, P. Balaram, *J. Chem. Soc., Perkin Trans. 2* **1997**, 2087.
- 227 C.K. Smith, L. Regan, *Acc. Chem. Res.* **1997**, 30, 153.
- 228 C. Micklatcher, J. Chmielewski, *Curr. Opin. Chem. Biol.* **1999**, 3, 724.
- 229 M.R. Ghadiri, M.A. Case, *Angew. Chem. Int. Ed.* **1993**, 32, 1594.
- 230 D.E. Robertson, R.S. Farid, C.C. Moser, J.L. Urbauer, S.E. Mulholland, R. Pidikiti, J.D. Lear, A.J. Wand, W.F. DeGrado, P.L. Dutton, *Nature* **1994**, 368, 425.
- 231 A. Lombardi, C.M. Summa, S. Geremia, L. Randaccio, V. Pavone, W.F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2000**, 97, 6298.
- 232 M.R. Ghadiri, *Adv. Mater.* **1995**, 7, 675.
- 233 M. Mutter, G. Vuilleumier, *Angew. Chem. Int. Ed.* **1989**, 28, 535.
- 234 G. Tuchscherer, M. Mutter, *J. Biotechnol.* **1995**, 41, 197.
- 235 G. Tuchscherer, C. Lehmann, M. Mathieu, *Angew. Chem. Int. Ed.* **1998**, 37, 2990.
- 236 K. Rose, L.A. Vilaseca, R. Werlen, A. Meunier, I. Fisch, R.M.L. Jones, R.E. Offord, *Bioconj. Chem.* **1991**, 2, 154.
- 237 O. Nyanguile, M. Mutter, G. Tuchscherer, *Lett. Peptide Sci.* **1994**, 1, 9.
- 238 H.K. Rau, N. DeJonge, W. Haehnel, *Angew. Chem. Int. Ed.*, **2000**, 39, 250.
- 239 R. Schnepf, P. Hörth, E. Bill, K. Wieghardt, P. Hildebrandt, *J. Am. Chem. Soc.* **2001**, 123, 2186.
- 240 W. Haehnel, *Mol. Diversity* **2004**, 8, 219.
- 241 P. Burkhard, J. Stetefeld, S.V. Strelkov, *Trends Cell Biol.* **2001**, 11, 82.
- 242 F.C.H. Crick, *Acta Crystallogr.* **1953**, 6, 689.
- 243 J.A. Talbot, R.S. Hodges, *Acc. Chem. Res.* **1982**, 15, 224.
- 244 N.E. Zhou, B.-Y. Zhu, C.M. Kay, R.S. Hodges, *Biopolymers* **1992**, 32, 419.
- 245 J.G. Adamson, N.E. Zhou, R.S. Hodges, *Curr. Opin. Biotechnol.* **1993**, 4, 428.
- 246 W.D. Kohn, C.M. Kay, R.S. Hodges, *Protein Sci.* **1995**, 4, 237.
- 247 K.S. Thompson, C.R. Vinson, E. Freire, *Biochemistry* **1993**, 32, 5491.
- 248 R.B. Hill, W.F. DeGrado, *J. Am. Chem. Soc.* **1998**, 120, 1138.
- 249 N.E. Zhou, C.M. Kay, R.S. Hodges, *J. Mol. Biol.* **1994**, 237, 500.
- 250 D.M. Eckert, V.N. Malashkevich, P.S. Kim, *J. Mol. Biol.* **1998**, 284, 859.
- 251 H. Wendt, C. Berger, A. Baici, R.M. Thomas, H.R. Bosshard, *Biochemistry* **1995**, 34, 4097.
- 252 A.J. Kennan, V. Haridas, K. Severin, D.H. Lee, M.R. Ghadiri, *J. Am. Chem. Soc.* **2001**, 123, 1797.
- 253 P. Vepřek, J. Ježek, *J. Peptide Sci.* **1999**, 5, 5.
- 254 P. Vepřek, J. Ježek, *J. Peptide Sci.* **1999**, 5, 203.
- 255 P. Niederhafner, J. Šebestík, J. Ježek, *J. Peptide Sci.* **2005**, 11, 757.
- 256 L. Crespo, G. Sandimens, M. Pons, E. Giralt, M. Royo, F. Albericio, *Chem. Rev.* **2005**, 105, 1663.
- 257 N.J. Wells, A. Basso, M. Bradley, *Biopolymers* **1998**, 47, 381.

- 258 J.P. Tam, *Proc. Natl. Acad. Sci. USA* **1988**, 85, 5409.
- 259 A. Basak, A. Boudreault, A. Chen, M. Chretien, N.G. Seidah, C. Lazure, *J. Peptide Sci.* **1995**, 1, 385.
- 260 B. Nardelli, Y. -A. Lu, R.D. Shiu, C. Delpiere-Defoort, A.T. Profy, J.P. Tam, *J. Immunol.* **1992**, 148, 914.
- 261 J.P. Tam, Y.A. Lu, *J. Am. Chem. Soc.* **1995**, 117, 12058.
- 262 K.J. Jensen, G. Barany, *J. Peptide Res.* **2000**, 56, 3.
- 263 J.P. Mitchell, K.D. Roberts, J. Langley, F. Koentgen, J.N. Lambert, *Bioorg. Med. Chem. Lett.* **1999**, 9, 2785.
- 264 A.T. Florence, T. Sakthivel, I. Toth, *J. Controlled Release* **2000**, 65, 253.
- 265 G. Purohit, T. Sakthivel, T. Florence, *Int. J. Pharm.* **2001**, 214, 71.
- 266 P. Niederhafner, J. Šebestík, J. Ježek, *J. Peptide Sci.* **2008**, 14, 2.
- 267 P. Niederhafner, J. Šebestík, J. Ježek, *J. Peptide Sci.* **2008**, 14, 44.
- 268 J.S. Choi, E.J. Lee, Y.H. Choi, Y.J. Jeong, J.S. Park, *Bioconj. Chem.* **1999**, 10, 62.
- 269 M. Sakamoto, A. Ueno, H. Mihara, *Chem. Eur. J.* **2001**, 7, 2449.
- 270 S.A. Vinogradov, D.F. Wilson, *Chem. Eur. J.* **2000**, 6, 2456.
- 271 M. Sakamoto, T. Kamachi, I. Okura, A. Ueno, H. Mihara, *Biopolymers* **2001**, 59, 103.
- 272 H.R. Kricheldorf, *Angew. Chem. Int. Ed.* **2006**, 45, 5752.
- 273 H. Ryser, R. Hancock, *Science* **1965**, 150, 501.
- 274 D.J. Mitchell, D.T. Kim, L. Steinman, C.G. Fathman, J.B. Rothbard, *J. Peptide Res.* **2000**, 56, 318.
- 275 A.D. Miller, *Angew. Chem. Int. Ed.* **1998**, 37, 1768.
- 276 L.C. Smith, J. Duguid, M.S. Wadhwa, M.J. Logan, C. -H. Tung, V. Edwards, J.T. Sparrow, *Adv. Drug Deliv. Rev.* **1998**, 30, 115.
- 277 D.W. Urry, *Biopolymers* **1998**, 47, 167.
- 278 E. Allmann, J. -C. Leroux, R. Gurny, *Adv. Drug Deliv. Rev.* **1998**, 34, 171.
- 279 N. Belcheva, S.P. Baldwin, W.M. Saltzman, *J. Biomater. Sci. Polym. Ed.* **1998**, 9, 207.
- 280 K. Kataoka, G.S. Kwon, M. Yokoyama, T. Okano, Y. Sakurai, *J. Controlled Release* **1993**, 24, 119.

8

Combinatorial Peptide Synthesis

Pharmaceutical research of the past four decades has been ruled by five different paradigms [1]. Initially, the drug discovery process relied on direct pharmacological screening in animals. The increased understanding of biochemical processes and methodological advances in organic synthesis in the 1980s introduced the second paradigm, which made use of drug design and optimization on a rational and molecular basis. During the mid-1980s, several different methods for the synthesis of compound libraries emerged that were inspired by the concept of selection, as verified by nature. This branch of preparative chemistry has been termed combinatorial chemistry [2–8]. Combinatorial chemistry represents an alternative to traditional approaches in pharmaceutical drug development and optimization. Compound libraries of different types [9] eventually introduced a complete change of paradigm in pharmaceutical chemistry. In 2001 the sequencing of the human genome was completed and, hence, in this fourth period, drug discovery focused on high-throughput identification of disease-related genes. Currently, systems biology, the detailed understanding of the complex biological networks in a living organism, represents the challenge for the coming decades with respect to the drug development process.

Combinatorial peptide synthesis and combinatorial organic syntheses – which can be seen as a further development of the former – have nowadays gained widespread acceptance and application. It has been estimated that the chemical space of compounds with a molecular mass below 500 Da would embrace 10^{60} structures. Hence, combinatorial chemistry aims to create and test high numbers of structures, but care has to be taken to generate the highest possible synthetic and structural diversity [10]. However, the major caveat in combinatorial chemistry is that the diversity and biological relevance of the compound libraries is not determined by the feasibility of the reactions and availability of a scaffold. There should be a biochemistry-based rationale in the background and great care should be taken that the compounds are not biochemically naïve.

Modern medicinal chemistry relies on different types of compounds (e.g., peptides, natural products, or small organic molecules) for biological testing, and one source of these is synthesis. The development track proceeds sequentially from test compounds over “lead compounds” to eventually a “drug candidate” and an IND (investigational new drug, Figure 8.1).

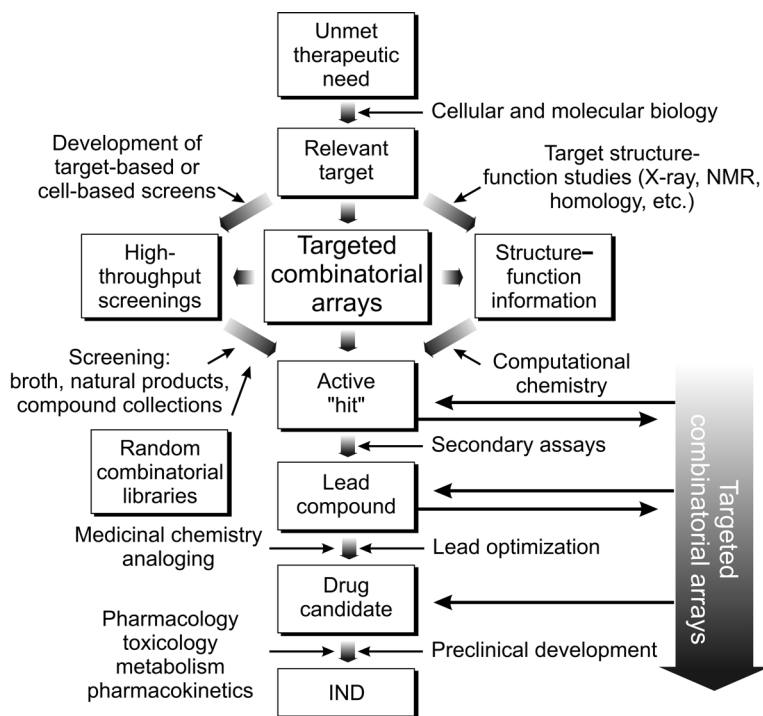


Figure 8.1 The role of combinatorial technology in the drug discovery process (adapted from [10]).

Formerly, most of the compounds had to be tested in animal models at a very early development stage. However, the past 30 years have witnessed enormous advances in biochemistry and biology (including structural biology) that have led to an improved understanding of the role of proteins and their ligands in pathological processes. A drug candidate is often a surrogate for a natural ligand that possesses the appropriate structural complementarity such that it may selectively mimic the natural ligand with a therapeutically beneficial result. Consequently, protein-based primary screening methods have now replaced testing in animal models [11].

Usually, several thousands of different compounds have to be synthesized and biologically tested in order to identify a pharmacological lead compound. Therefore, the pace of synthesis had to be increased considerably to satisfy the demand for structural diversity of chemical compounds. The fascination of combinatorial chemistry basically lies in its potential to synthesize several hundreds or thousands of test compounds, and to subject them to biological testing in order to significantly reduce the time necessary for the development of a drug. High-throughput screening enables scientists to test thousands of compounds per day in a biological assay. Combinatorial syntheses of peptides (peptide libraries) and organic compounds (diversomer libraries) are especially suited for the development of lead structures in pharmaceutical research. The expression “library” was originally coined exclusively for compound mixtures, whilst collections of single compounds are often called “arrays”.

Methods for combinatorial synthesis were developed independently by several groups on the basis of different concepts following two major strategies:

- Parallel (multiple) synthesis: this comprises the targeted simultaneous multiple peptide synthesis (SMPS). Arrays of single substances (peptides) are obtained in a parallel manner, for instance as analogues of a biologically active peptide [12].
- The synthesis of complex mixtures of compounds (libraries) [13].

The synthetic peptide combinatorial libraries (SPCL) are synthesized and tested as mixtures. Many bioassays used for the development of pharmacologically active compounds display sufficient sensitivity and an inherent biological specificity for active compounds present in a mixture. Arguments against this combinatorial synthesis of complex mixtures with respect to the demanding chemical criteria of purity have been opposed by this fact. However, it has been postulated that, for example, the antagonist activity of one compound may override an agonist activity of another compound present in the mixture, leading to a “null” (false-negative) response. Furthermore, synergism between different library members (false-positive response) cannot be excluded.

While parallel syntheses are usually limited to arrays comprising several hundreds to thousands of compounds, depending on the reaction scale required and the method used, the diversity of a mixture-based combinatorial library is, in principle, only restricted by the detection limit of the biological assay and by the total mass of substance to be produced. Full peptide libraries are often constructed using 19 of the proteinogenic amino acids (Cys, Sec, and Pyl are normally omitted). However, many nonproteinogenic or even non-natural amino acids are commercially available nowadays, leading to a high diversity. The numbers (V^P) of different peptides composed of 19 L-amino acids (19 variables V) present in a library, depending on the number of positions P to be varied, are listed in Table 8.1.

Table 8.1 Library diversity correlates with the number of variable positions and the number of diversity elements (e.g. 19 amino acids).

Number of variable positions	Number of peptides	Average mixture weight (1 pmol per peptide)
2	361	92 ng
3	6 859	2.58 µg
4	130 321	64.76 µg
5	2 476 099	1.53 mg
6	47 045 881	34.64 mg
7	893 871 739	765.25 mg
8	16 983 563 041	16.57 g
9	322 687 697 779	353.52 g
10	6 131 066 257 801	7.45 kg

It can easily be calculated that, for a high number of variable positions or a high number of different building blocks (variables), the total mass of the product mixture (likewise, the resin mass in the case of a solid-phase synthesis) increases exponentially. Consequently, if the reaction scale is limited, such combinatorial mixtures may remain incomplete. Moreover, mixture-based libraries may suffer from non-equimolar incorporation of building blocks because of differences in reactivities. Methods to overcome this problem will be discussed in Section 8.2.

Synthetic strategy is another criterion to distinguish between two different approaches in combinatorial chemistry: solid-phase synthesis versus solution synthesis.

In solid-phase combinatorial synthesis, reagents can be used in excess without separation problems in order to attain complete conversion. Facile purification and automation of the process represent further advantages that have already been discussed in Chapter 4. While solid-phase peptide synthesis (SPPS) is now well developed, the solid-phase synthesis of organic molecules imposes different requirements, additional reaction steps, and often involves further development, for example of the linker or protecting group chemistry [14]. Monitoring of the reaction progress in most cases is difficult. The solution-phase synthesis of organic molecules [15] may utilize all known organic reactions, without the need for special linker systems and for any adaptation of known reaction conditions. However, reagents cannot usually be used in excess, and process automation is much more difficult [16]. Solid-supported reagents and scavengers greatly facilitate solution-phase combinatorial synthesis [17].

Which strategy and which degree of diversity (library size) is chosen depends on the problem to be investigated. If nothing is known about a target, then maximum diversity is desirable, and the highest possible number of different compounds should be screened in such a case. However, the increasing knowledge of biochemical concepts and biopolymer structures also contributes considerably to lead structure detection. Structure-based molecular design can be employed to guide the lead-finding process [18]. Structural data on a target protein are applied in a rational approach using computational methods to identify potential low-molecular mass ligands. In such a case a smaller, focused library (e.g., an array of single compounds) may be preferred, which is called the targeted diversity approach [11]. Once a lead compound has been identified, the diversity requirements change fundamentally, and structure–activity relationships are determined by substituent variation in order to optimize the compound properties with the aim of developing a drug candidate. Consequently, a sensible combination of different methodologies and technologies will eventually succeed [19].

8.1

Parallel Synthesis

Multiple peptide synthesis (MPS) means the simultaneous (parallel) synthesis of a multitude of peptide sequences, irrespective of the chain length and amino acid composition.

Peptide chemists first recognized that compound libraries could be synthesized in a combinatorial manner. Many strategies of multiple peptide synthesis (or multiple synthesis in general) are special variants of synthesis on a polymeric support [8]. Combinations of protection groups and activation methods of solid-phase strategy relevant to peptide chemistry can be applied to multiple synthesis. The methodological arsenal ranges from peptide synthesis in the so-called teabags developed by Houghten *et al.* [20], multipin synthesis (polyethylene rods as polymeric support) developed by Geysen *et al.* [21, 22], and spot synthesis developed by Frank [23], to methods with pipetting robots and fully automated synthesis. Furka *et al.* [24, 25] contributed greatly to the development of combinatorial chemistry with the development of the “split and combine” technique. This technique provides a library in the form of a spatially resolved compound mixture (“one bead – one compound”) and will be described in detail in Section 8.2.

There are differences between the variants of multiple peptide synthesis with respect to the polymeric support used, the number of different peptides to be synthesized with this method, and the amount of products formed in the synthesis. Besides the classical Merrifield support polystyrene/divinylbenzene (DVB), polystyrene/polyethylene films, polyethylene rods and cellulose sheets are also used. A full account of combinatorial methods is far beyond the scope of this book, but some of the different methods will be introduced briefly.

8.1.1

Synthesis in Teabags [20]

A certain amount of the polymeric support is placed in a porous polypropylene bag marked with solvent-resistant ink. Usually, about 100 mg of polystyrene cross-linked with 1% divinylbenzene is used per teabag [26]. Other support materials may also be used, but the particle size of the polymeric resin should always be larger than the porosity of the polypropylene bag (64–74 μm). The number of different compounds to be synthesized is equal to the number of teabags used. Cleavage reactions of, for example, the α -amino-protecting groups may be performed in one common vessel for all teabags. The teabags where the same amino acid component has to be coupled are combined in one reaction vessel and reacted with the activated amino acid component, followed by washing. Vigorous shaking is a necessary precondition for complete reaction. After each coupling step, the teabags are removed from the reaction vessel, pooled, and the N-protecting groups then cleaved. The teabags are then sorted again, combined appropriately, and treated with the next activated amino acid in separate reaction vessels. In order to remove excess reagents, the teabags are washed and combined again for the next deblocking reaction. Coupling, washing, and deprotection are the repetitive steps (as discussed for SPPS in Chapter 4). Before the last coupling step, the teabags are carefully washed and dried, after which cleavage of the peptides from the polymeric support is performed. The teabag method is appropriate for the Boc tactics using, for example, *N,N'*-diisopropylcarbodiimide as the coupling reagent, and employing cleavage from the resin with liquid HF [20, 26], and also for the Fmoc tactics, as demonstrated in Figure 8.2.

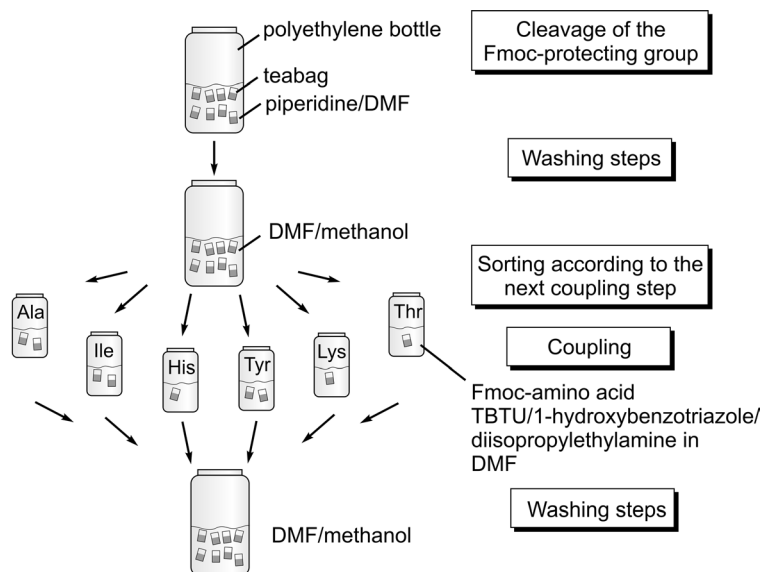


Figure 8.2 Multiple peptide synthesis according to the tea-bag method.

The teabag method is usually suitable for the synthesis of amounts between 30 and 50 mg, and a peptide length of more than 10 amino acids. In contrast to other methods of parallel synthesis, purification and characterization prior to biological testing are possible. This protocol is characterized by a high variability and relatively low costs, but it is highly labor-intensive because of the multitude of manual steps. Computer-assisted calculation of the synthetic cycles and the sorting of the resin bags is essential. Automation of the washing steps by using a combination of the teabag method with suitable semi-automatic peptide synthesizers is advantageous [27]. Peptides synthesized by the teabag method have been applied to the investigation of structure–activity relationships, antigen–antibody interactions, and for the conformational mapping of proteins.

8.1.2

Synthesis on Polyethylene Pins (Multipin Synthesis)

Parallel peptide synthesis on polyethylene rods [21], according to the multipin method, does not usually provide larger amounts of peptides. In many cases, after the synthesis only the protecting groups are cleaved and the peptides remain bound to the polyethylene rods, for instance for antibody binding studies. Peptide assembly is performed using amino functionalized polyethylene rods (pins) about 4 mm in diameter and 40 mm in length (Figure 8.3). However, larger rods are also commercially available. Usually, 96 of these pins are assembled on one block in 8 rows, with 12 pins each. This arrangement allows application in enzyme-linked immunosorbent assay (ELISA) after the synthesis is complete, because the pins fit exactly into the wells

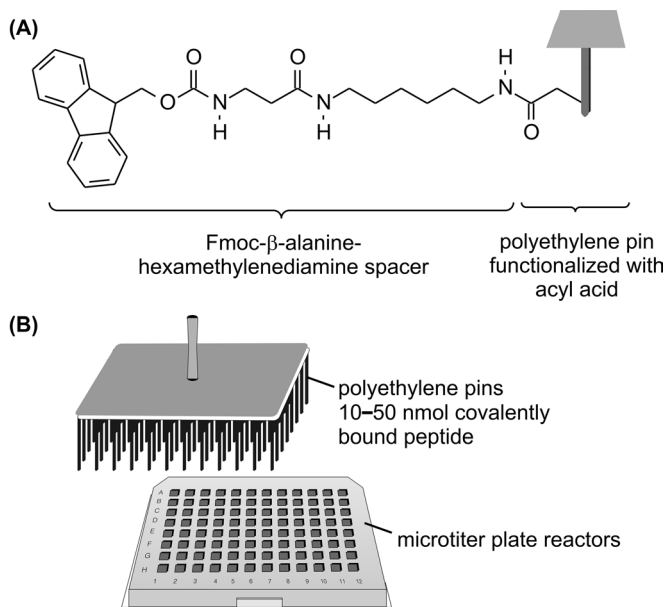


Figure 8.3 Schematic view of parallel peptide synthesis on polyethylene rods (multipin synthesis). (A) Polyethylene rod ($4 \times 40 \text{ mm}^2$) with Fmoc protected β -Ala-hexamethylene spacer. (B) Polyethylene rods are mounted on a plate. The parallel reactions are performed in a complementary microtiter plate.

of the microtiter plates usually applied in this method. Computer software enables calculation of the amounts and distribution of the protected amino acid components in the different wells. Peptide coupling reactions are performed in parallel in the wells of microtiter plates. The activated amino acid component is filled into the corresponding well, and the coupling reaction is started by dipping the whole array of pins, mounted on one plate, into the complementary array of wells on the microtiter plate. Cleavage and washing steps can usually be performed in one bulk solution. Instead of pins permanently grafted to the pin holder, removable “crowns” of different sizes can be used [28]. This method is shown schematically in Figure 8.3. A pin is equipped with Fmoc-protected linker moieties (Figure 8.3A) and, after Fmoc cleavage, the first amino acid is attached. An array of pins uses a complementary array of microtiter plate wells as reactors (Figure 8.3B).

This protocol, which usually provides peptides without any information on their purity, is mainly oriented towards immune analysis. Epitope mapping, which means the systematic screening of partial amino acid sequences of a protein, can be efficiently performed using multipin technology. For example, with this method the amino acid sequence required to be recognized by an antibody can be determined. Between 10 nmol and 2 μmol of a single peptide can be synthesized on each pin, depending on the pin size; hence, this procedure is virtually impractical for preparative synthesis. The pin method also allows the synthesis of either peptides

permanently bound to the pins or free peptides, after cleavage of the final product from the polyethylene rod.

8.1.3

Parallel Synthesis of Single Compounds on Cellulose or Polymer Strips

The spot synthesis [29] is, like the pin method, suitable for the synthesis of peptides in minute amounts, and in most cases is restricted to the final products remaining anchored to the matrix. The spot synthesis utilizes a planar sheet of cellulose as the polymeric support. The first (C-terminal) amino acid is connected via an ester bond and a linker molecule to the hydroxy groups of cellulose. Residual hydroxy functions of cellulose are subsequently deactivated. The amino-protecting group is then cleaved, and the next Fmoc-protected amino acid component is attached to the free amino group. Amino acids and coupling reagents are distributed by a pipetting robot, in 1 μ L aliquots, onto the cellulose sheet at a distance of about 1 cm or less. All washing operations, the capping reaction with acetic anhydride, and the cleavage of the Fmoc protecting group are performed by dipping the whole sheet of cellulose into the appropriate reagent solutions. On completion of the synthesis, side-chain protecting groups are cleaved with 20% trifluoroacetic acid/1% triisobutylsilane. This low-cost technology can also be used for screening the biological activity. The peptides are usually not cleaved from the cellulose sheet and, consequently, the sheet with the bound peptides can be treated and tested like Western blot membranes. An array of between 96 and several thousands of peptides can be created on a single cellulose sheet in a parallel manner. Despite the fact that only 100 nmol of peptide per spot can be synthesized, free peptides may be obtained. For this purpose, a potential cleavage site (usually in the form of a cleavable linker) must be incorporated before attaching the first amino acid [30]. The single spots can be punched out and the peptides may be cleaved in separate vessels.

Even before the development of spot synthesis, an interesting procedure for multiple peptide synthesis on cellulose as polymeric support had been developed [31, 32], based on previous experience in oligonucleotide synthesis. Small cellulose disks are numbered and used for the semi-automatic synthesis of a peptide. Before each coupling, the cellulose disks (1.55 cm diameter) are appropriately sorted with respect to common building blocks to be coupled, and are piled up to form columns. Each column is then separately subjected to the coupling of the amino acid component. Correct sorting is assisted by computer software, which also minimizes the number of synthesis cycles. The peptides are obtained in crude yields of 8–10 mg, and can also be cleaved from the support when suitable anchoring groups have been used. Immunological tests may be performed without cleavage from the solid support.

Paper strips (Whatman 540) have also been used as a solid support. Loading of up to 1–2 μ mol cm^{-2} has been described [33]. Peptides can be synthesized using this approach according to the Boc or Fmoc tactics, and can be cleaved from the support. The peptides bound to the paper may also be subjected directly, after side-chain deprotection, to antibody binding assays in an ELISA. The process is somewhat

related to the teabag strategy in terms of methodology, but cutting of the paper is easier than filling and properly closing the teabags. One obvious disadvantage is the low mechanical stability and the polyfunctionality of the cellulose support.

Cotton-wool displays higher mechanical stability and, therefore, disks made from cotton with a diameter of 3 cm proved advantageous as a support in multiple synthesis [34, 35]. In this case, the solvent can be removed after the washing steps by centrifugation.

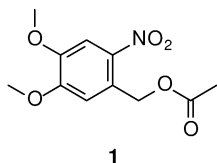
Chemically functionalized polystyrene–polyethylene films may also be used as the solid support material in the form of small pieces ($1.5 \times 3 \text{ cm}^2$). The films can be subjected to deblocking reactions and washing steps together and then separated into different flasks for the coupling reaction [36]. Although the peptide yields are similar to those obtained using the teabag technology, this variant represents a clear simplification with respect to handling because no tedious filling and closing of the teabags is necessary. The copolymer polystyrene–polyethylene also ensures minimal loss of resin. Novel polymeric surfaces (e.g. polypropylene disks) together with new linker/cleavage chemistry and assay systems and automated robot systems extended the scope of the parallel synthesis on planar supports with respect to the number of compounds obtainable by this technique. Thus, highly complex spatially addressed compound arrays have become accessible [37].

Further automation of multiple peptide synthesis is accompanied by a significant increase in efficiency. The different variants of multiple peptide synthesis described above are clearly characterized by a high percentage of manual operations. On the other hand, repetitive steps, such as deprotections, washing and coupling, offer possibilities for automation. Using software-controlled pipetting robots, fully automatic multiple peptide synthesis machines have been developed [38, 39]. The number of tasks in manual multiple peptide synthesis may also be reduced by using semiautomatic machines.

8.1.4

Light-Directed, Spatially Addressable Parallel Synthesis

A photolithographic technique, which is well established in microchip technology, enables the parallel synthesis of a compound array on a glass surface. As one certain peptide sequence corresponds to a discrete x/y -coordinate on the disk, peptides binding with high affinity to a protein can be directly identified in the biological test. Further selective and sensitive analytical methods are not required in this case. The light-directed, spatially addressable parallel chemical synthesis is based on a combination of photolithographic techniques with solid-phase synthesis using photolabile protecting groups such as Nvoc (6-nitroveratryloxycarbonyl) **1** [40].



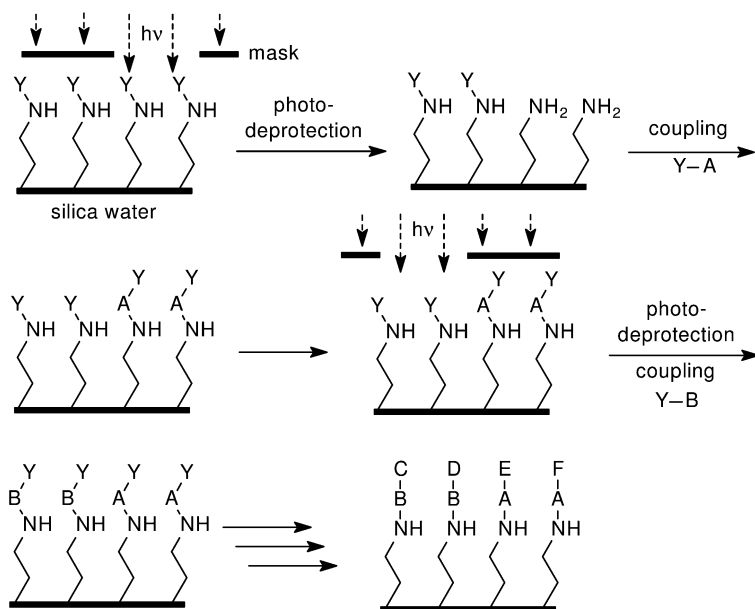


Figure 8.4 Light-directed, spatially addressable multiple synthesis. Y: photolabile N^α-protecting group; A, B, C, D, E, F: amino acids.

Amino-functionalized glass plates are used as the solid support for this technique [41, 42]. Using suitable masks, photodeprotection of the amino terminus is possible with a spatial resolution of $50\text{ }\mu\text{m} \times 50\text{ }\mu\text{m}$. Light of wavelength 365 nm is used for cleavage (Figure 8.4), and subsequently the next Nvoc-protected amino acid is coupled. An array of 2^n different compounds may be synthesized using n different photolithographic masks, each covering half of the surface area and allowing deprotection in the uncovered area.

An appropriate choice and layout of the lithographic mask enables the synthesis of up to 10 000 different peptides per cm^2 . The photolithographic technique allows an exact positioning. Screening with biological test systems can be performed, using readout by fluorescence microscopy or laser confocal fluorescence detection of fluorescently labeled antibodies.

8.1.5

Liquid-Phase Synthesis using Soluble Polymeric Support

The principle of liquid-phase peptide synthesis on a soluble polymeric support, which was first described by Mutter and Bayer in 1974 [43] (see Section 5.3.4.2), was applied to combinatorial synthesis in 1995 by Janda, and has been termed liquid-phase combinatorial synthesis [15, 44]. Polyethyleneglycol monomethylester with a molecular mass of 5 kDa was used as the soluble polymeric support. This material proved useful in peptide, oligonucleotide, and oligosaccharide synthesis. The crys-

tallization method according to Bayer and Mutter is utilized most favorably at each step of the combinatorial synthetic process. The addition of a suitable organic solvent leads to the crystallization/precipitation of the peptidyl polymer with formation of helical structures, thus avoiding inclusions. A further advantage of the synthesis in solution is the application of a recursive deconvolution strategy [45] in order to obtain and identify the most interesting library. Furthermore, the synthetic progress during this combinatorial synthesis can be monitored using NMR spectroscopy.

8.2

Synthesis of Mixtures

Peptide libraries (mixtures) [12] are obtained using special methods of combinatorial peptide synthesis. In contrast to classical peptide synthesis, it is not the carefully purified and characterized single peptide that is the focal point in this case. A library should be a mixture of peptides with an optimum degree of heterogeneity. Such peptide libraries can be composed of from hundred thousands to millions of different peptides. The advantages of combinatorial synthesis can be seen in the relatively small number of synthetic steps. The application of suitable testing methods allows the identification and isolation of peptides with a specific mode of action. Sequencing and characterization of these compounds provides the precondition for the synthesis of this single compound on a larger scale. Combinatorial synthesis also permits the assembly of organic compounds that no longer resemble the original peptidic structure, hence providing perspectives for the synthesis of nonoligomeric compound libraries and screening for lead structures.

Suitable biochemical or biological testing methods in combination with synthetic peptide libraries enable investigations to be made on diverse biochemical issues in a much more efficient manner compared to the methods used previously. Epitope mapping of proteins of differing size for diagnostics, as well as the development of vaccines, is of major importance in this context. Furthermore, the systematic variation of amino acid building blocks allows a rapid and exact study of the folding tendency of longer peptides. Peptide libraries have also contributed significantly to the development of enzyme inhibitors, as well as to screening for new enzymes or the characterization of the specificity of known enzymes. Peptide-specific monoclonal antibodies can be obtained starting from lipopeptide–antigen conjugates containing complete protein sequences in the form of overlapping peptide epitopes. These conjugates allow targeted screening for epitopes of, for example, B-, T-helper, and T-killer cells.

The combination of only 20 amino acids provides 20^6 , or 64 million hexapeptides. The synthesis of such a diverse library has been made possible with the development of modern synthesis robots, whilst the testing of such diverse compounds using high-throughput screening, has opened up new dimensions in drug research compared to the classical procedure. One problem might be that the components of a library are not usually checked in terms of their purity, and so the library may be incomplete. In biological testing, the active compound is identified and characterized; subsequently its structure must be elucidated in order to validate the potential lead structure.

8.2.1

Reagent Mixture Method

The “reagent mixture” method allows a mixture of building blocks to be incorporated into the molecule anywhere within the reaction scheme. This method uses a mixture of reagents with a predefined ratio in excess with respect to the second reactand to achieve nearly equimolar incorporation of each building block (mixture member) at the position of diversity [46–49]. Equal incorporation of each diversity element (building block) requires profound knowledge of the mechanism and kinetics involved in the specific reaction. Equimolar incorporation of the diversity elements may also be achieved by applying a large excess of each building block, where the molar ratio of the building blocks is adjusted according to their different reactivity (isokinetic mixture) [46]. Consequently, a building block with higher reactivity will be applied in the mixture at a lower concentration compared to a building block with lower reactivity. Another method uses double couplings of equimolar building block mixtures, without considering the different reactivities. Consequently, some sequences may be formed in greater amounts than others.

8.2.2

Split and Combine Method

Special procedures are required for the assembly of peptide libraries (mixtures) that contain more than 1000 peptides. The split and combine method [24, 50–52] uses chemically functionalized polymeric beads applied in a solid-phase synthesis. Both Boc and Fmoc tactics may be applied. The method is based on the separation of the beads into equal portions prior to coupling, whereas deprotection of the N^α -protecting groups and the washing steps are performed together. The resin is divided into a number of portions of about equal size and corresponding to the number of different amino acids to be coupled in this position. One amino acid component is coupled to each of these portions in separate containers. The number of reaction vessels is equal to the number of different building blocks to be coupled in one position. Then, all portions are combined and the N^α -protecting group is cleaved. Following this, the whole amount of beads is evenly split again into a new number of separate portions. The principle of the split and combine method is shown schematically in Figure 8.5.

If, for instance, all possible tripeptides comprising Leu, Phe, and Lys are to be synthesized, the resin is split into three portions. The first amino acid is coupled to each portion, the portions are combined, deprotected, washed and split again. Then in each container, the second amino acid component is coupled, the three portions are combined again, protecting groups are cleaved, and the bulk of the resin is split into three portions again. According to this method, all 27 possible tripeptide sequences are synthesized. One single bead contains only peptides with the same sequence (“one bead, one compound”; Figure 8.6). There may be up to 10^{13} identical molecules on one single bead, which is 100 μm in diameter.

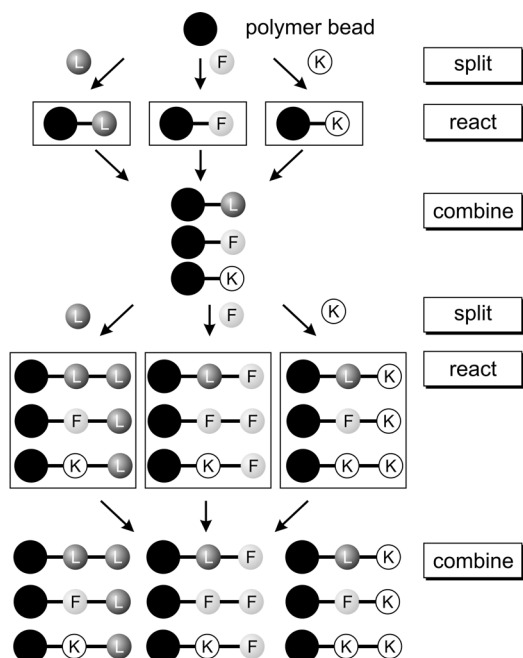


Figure 8.5 Split and combine strategy.

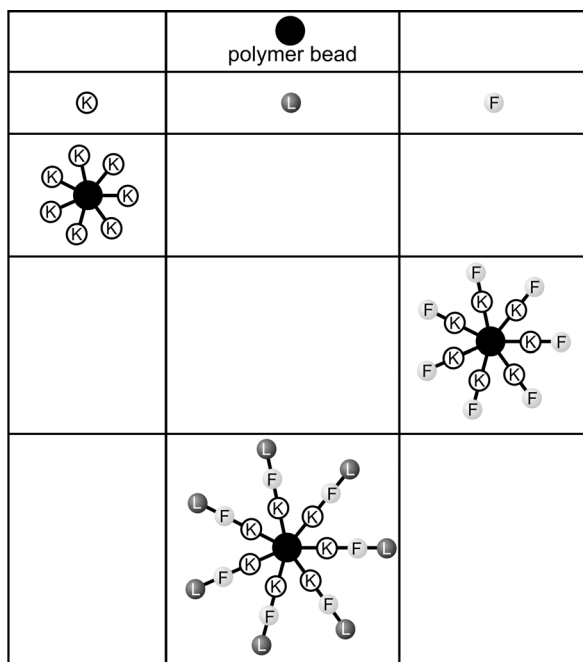


Figure 8.6 The split and combine method provides single compounds on single beads.

Care must be taken to use the appropriate amount of resin in the synthesis to ensure a statistical representation of all compounds in the library; otherwise, the library may be incomplete. Furthermore, sequence mismatches depending on problems that occur during synthesis cannot be excluded. Although each bead contains peptides with the same sequence, the primary structure of these peptides is not known. Peptides libraries synthesized according to this protocol may be used for screening purposes with resin-bound peptides, and the peptides displaying the desired biochemical activity can be identified using suitable bioassays. Special enzyme-labeled antibodies can be used to catalyze color reactions and, consequently, this method may be used for antibody epitope mapping. The peptide sequence that binds with high affinity to an antibody will give rise to an intensive color of the bead after incubation with the enzyme-labeled antibody. The labeled resin beads are subsequently separated, the peptide is cleaved, and the sequence is then determined by microsequencing. As the split and combine method is somewhat limited with respect to the size of the library (variability), sublibraries are often created with fixed (defined) amino acid residues in certain positions.

The three repetitive tasks in the split and combine method (coupling, mixing, and splitting) can be easily automated. For the combination of three different building blocks to one peptide, only nine separate synthetic steps are required in combinatorial synthesis to obtain the $3^3 = 27$ possible tripeptides. This compares favorably with a linear synthesis, where 81 steps are required.

About 50 000 cyclic disulfide peptides with random sequences have been synthesized and tested as antagonists for the platelet glycoprotein receptor GP IIb/IIIa (integrin $\alpha_{IIb}\beta_3$). This receptor binds the tripeptide sequence RGD present in the corresponding proteins. The cyclic peptide with the sequence CRGDC was found to be the most active compound, with a binding affinity of $1\ \mu\text{M}$ [53]. In a different study, 2×10^6 peptides were tested for affinity to the SH3-domain of phosphatidylinositol-3-kinase (PI3K). The peptides were still resin-bound, and the SH3-domain was labeled with a fluorescent marker [54]. The peptides which bound with highest affinity to the protein were consequently present on the brightest beads when examined by fluorescence microscopy. These beads were then selected and the primary structure of the peptide was established by sequencing. The peptide RKLPPRPRR showed the highest affinity towards the SH3-domain of the kinase (binding affinity $7.6\ \mu\text{M}$). A tetradecameric peptide library eventually provided a thrombin inhibitor with an inhibitory constant (K_i) of 20 nM [54]. The fluorescence-activated cell sorting (FACS) technology or special bead-sorting machines which are based on this principle greatly facilitate the high-speed selection of fluorescent beads and, consequently, accelerate the detection of high-affinity ligands [55].

8.2.3

Encoding Methods

The advantages gained in the synthesis of a peptide library and in the testing of the compound mixture are compensated by the necessary deconvolution in order to identify an active compound. Peptide libraries obtained according to the split and

combine method may be analyzed using Edman microsequencing. Binary encoding of compound libraries provides one possible means of overcoming this problem [56]. Brenner and Lerner first described a concept for the generation of an encoded compound library by application of the split and combine method which comprised two consecutive combinatorial synthetic steps. Earlier methods used peptides or oligonucleotides for encoding purposes, but these suffered from limited chemical stability.

Still *et al.* [57–59] developed nonsequential binary encoding of peptide libraries. In this technique, the encoding molecules are not connected to each other in a sequential manner. Each coupling step of an amino acid component is preceded by the tagging step which is necessary to identify the peptide sequence present on the beads (Figure 8.7).

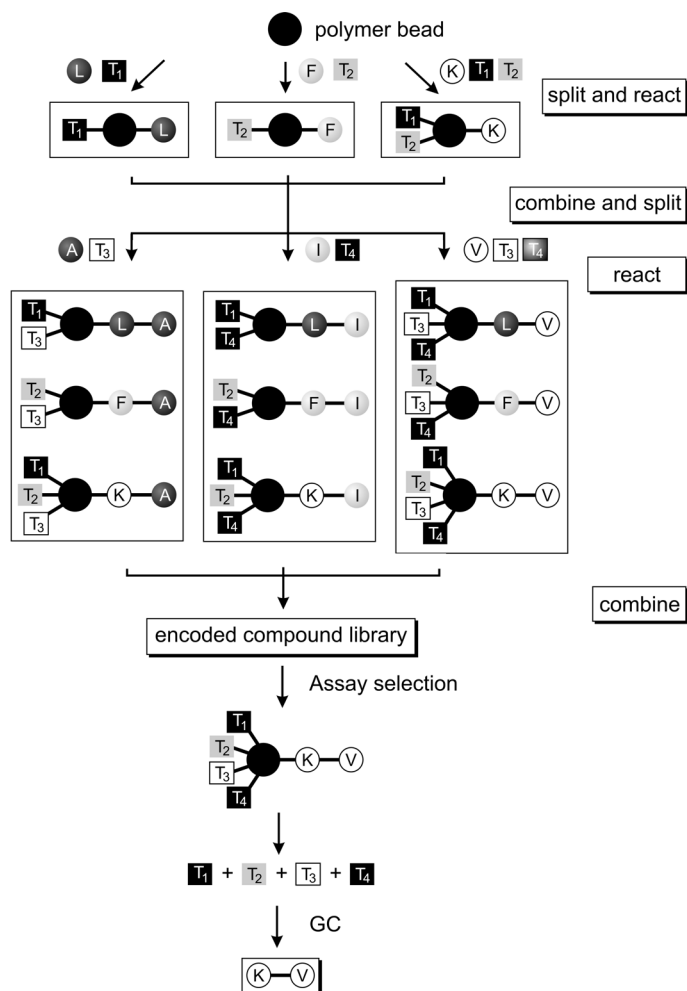


Figure 8.7 Binary encoding using tag molecules.

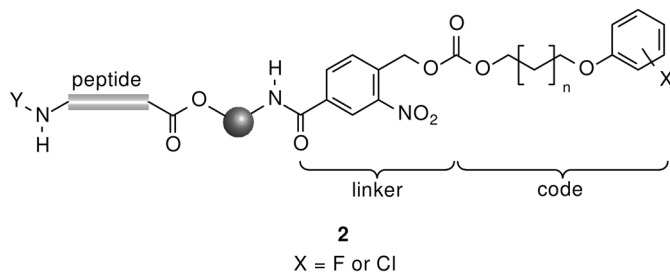
Table 8.2 Binary code of compound libraries.

Educt ^a	Step	Binary code	Code molecule ^a
L	1	10	T ₁
F	1	01	T ₂
K	1	11	T ₁ + T ₂
A	2	10	T ₃
I	2	01	T ₄
V	2	11	T ₃ + T ₄

^aEducts are symbolized by the one-letter code (cf. Fig. 8.7), code molecules by T₁ to T₄.

One disadvantage in the screening of peptides bound to a solid support might be a negative influence of the polymeric matrix on the binding of the peptide to its receptor. Therefore, orthogonal linker systems for the library member and the tag that allow independent cleavage of the library member and the tag from the solid support for an assay in solution are often required [60]. Decoding relies on the formal assignment of a binary code for single encoding molecules and their mixtures, as shown in Table 8.2.

$(2^x - 1)$ building blocks in the library can be encoded with x code molecules. Considering a reaction sequence with n steps, the encoding of a library of $(2^x - 1)^n$ components requires x^n encoding molecules using combinatorial principles. For example, 20 code molecules are necessary to tag 923 521 different compounds obtained from a four-step reaction sequence. Encoding is usually performed before the coupling, as shown schematically in Figure 8.7. Only a small amount of encoding molecules and of free amino groups of the solid support are used for the attachment. A photolabile *o*-nitrobenzylester group, for example, may be used as a linker for the attachment of the encoding molecules such as **2** to safeguard their detachment independently of the library members.



Photolabile linkers, for example, may be combined with linkers allowing oxidative cleavage. For the identification of the compound present on a bead, the tagging molecules must be detached and analyzed, and should also usually be volatile in order to allow gas chromatographic separation (if required). Halogenated aromatic compounds containing alkoxy spacers of varying length and a cleavable linker are coupled to the bead by Rh(II)-catalyzed carbene insertion. The ratio library member/tag molecule is usually between 200 : 1 and 100 : 1. Once bioactivity has been detected

in an enzyme assay or receptor binding studies, single active beads are selected and oxidative or photolytic cleavage of the encoding molecule is performed. The resulting alcohols **3** are silylated, and the volatile silyl derivatives are separated by gas chromatography (electron capture detection, ECD, provides high sensitivity). By using this tagging concept, the identity of a bioactive compound present on a bead may be recognized by selective, highly sensitive methods without the need for peptide sequencing, as has been shown for several examples (e.g. [61, 62]).

Binary encoding (Figure 8.8) is also very efficient for the synthesis of compound libraries of organic molecules. These have been named as “diversomers” in order to distinguish them from the oligomer libraries (peptide libraries) that can be sequenced.

Although peptide libraries have been useful in lead structure finding, the transformation of a biologically active peptide into a nonpeptide drug often incurs high costs for the necessary modifications. Furthermore, oligomer libraries contain repetitive backbone bonds (peptide bonds, nucleotide bonds) that are detrimental to the concept of diversity. One target in the development of diversomers is that of lead structure finding. Diversomer libraries can be subjected to automatic biological screening systems (high-throughput screening); for example, the synthesis of nonpeptide, nonoligomeric libraries of hydantoins and benzodiazepines with a high degree of chemical diversity has been performed automatically [63]. Clearly, the number of organic molecules obtained in this way is small when compared to the diversity achieved with solution-phase organic chemistry. Nonetheless, the adaptation of further organic chemical reactions to this method will ultimately lead to a high number of diversomer libraries.

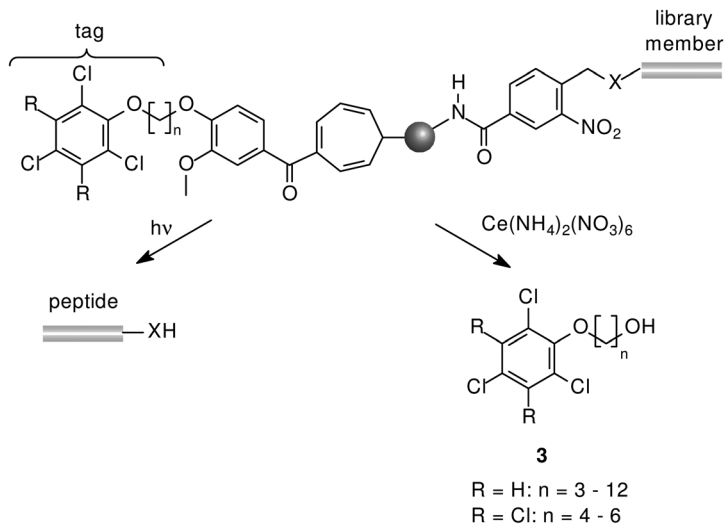


Figure 8.8 Binary encoding of a library with halogenated aromatic compounds bearing different alkoxy spacers. The orthogonal linkers allow independent cleavage of the library member and the tag, respectively.

Incomplete reactions in solid phase chemistry may also provide means of encoding split and combine libraries, which is especially valuable for building blocks that cannot be identified by Edman degradation. For such a purpose the bilayer bead partial amine protection (PAP) or the bilayer bead partial Alloc deprotection (PAD) methods have been developed. They both rely on the application of polyethylene glycol-based resins like TentaGel, which is first swollen in water and then reacted with a dichloromethane solution of substoichiometric amounts of a protecting (PAP) or deprotecting (PAD) reagent. This brings about partial protection (PAP) or deprotection (PAD), predominantly in the outer sphere of the resin bead, which can be used for encoding purposes [64].

Furthermore, the approach of PNA-encoded protease substrate microarrays was developed [65]. A library of 192 protease substrates with a latent fluorophor was prepared by split and combine combinatorial synthesis. Proteolysis of the bond connecting the substrate to the latent fluorophor occurs upon incubation with proteases, protease cocktails, cell lysates, or blood samples, if the substrate is accepted by the protease. Cleavage changes the electronic properties of the fluorophor and results in a large increase in fluorescence. The substrates are encoded with PNA tags, eventually addressing each of the substrates to a predefined location on an oligonucleotide microarray through hybridization. This allows the deconvolution of multiple signals from a solution.

Nonchemical encoding methods have also been developed [66–68] in which each “giant” resin bead contains a microchip to store and read all synthetic steps performed with this particular bead, using high-frequency signals. Further developments of this technique have been reviewed [69].

8.2.4

Peptide Library Deconvolution

The deconvolution of a compound library is the process where it must be determined which discrete substance (library member) has the desired property observed in the biological test. Two major strategies are available for this purpose: (i) the iteration method, and (ii) positional scanning.

The iteration method (Figure 8.9) relies on the creation of sublibraries, once the desired property (e.g., biological activity) has been detected in a mixture library. If a tetrapeptide library has been created where each position is varied by five different amino acid building blocks, the total library consists of $5^4 = 625$ compounds. In the first deconvolution round this total library is re-synthesized in the form of 25 sublibraries, each containing 25 different compounds. In each library, two positions are defined and two positions are varied. Once the sublibrary displaying the highest biological activity has been revealed, the second deconvolution round follows where five further sublibraries are generated on the basis of this result. Each sublibrary comprises five different sequences. Three positions of the tetrapeptide are now fixed, and one position is varied using the five different amino acid building blocks. Again, the biological test is performed and the sublibrary with the highest biological activity is subjected to further deconvolution. Finally, five single

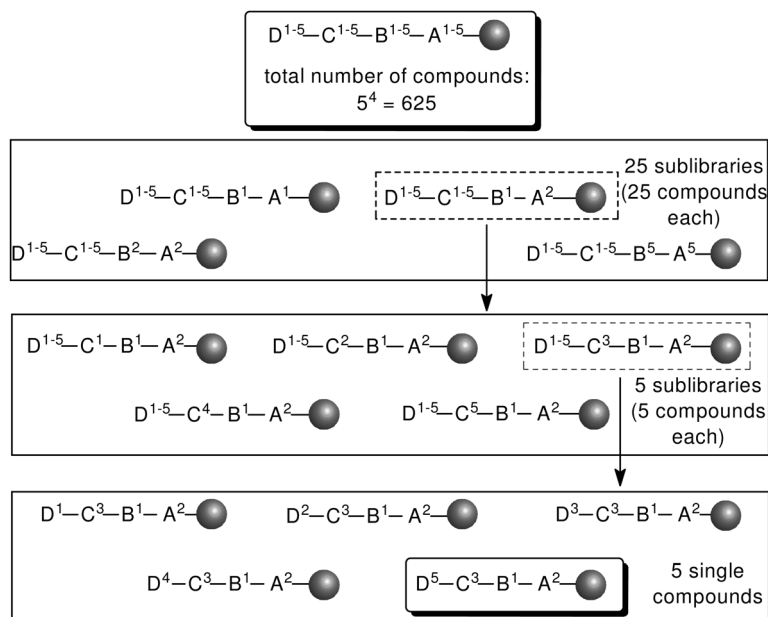


Figure 8.9 Identification of the most active library component by deconvolution. Sublibraries of the most active library (dashed frame) are synthesized until finally the most active single compound ($D^5C^3B^1A^2$) is identified.

compounds are synthesized, corresponding to the variation of the fourth tetrapeptide position, and the active component should be revealed.

One disadvantage of this iterative procedure is that the most active compound may not necessarily be detected. This protocol is also applicable to the “split and combine” method where the sublibraries, for example of the first deconvolution round, are generated by omitting the last mixing step at the end of the synthesis.

Recursive deconvolution has been described by Janda [44, 45]. After each cycle in the split synthesis an aliquot is retained, the major advantage of this approach being that no new syntheses are necessary because the aliquots secured after each split synthesis step correspond to the necessary sublibraries.

The non-iterative method of positional scanning was suggested by both Houghten *et al.* [70, 71] and Furka *et al.* [72]. The partial libraries of positional scanning represent first-order sublibraries in each of which one position is kept invariant, while all other positions are varied (Figure 8.10). In the tetrapeptide example used above, this would mean that 20 first-order sublibraries would be generated, each containing 125 compounds, which corresponds to the variation of three positions by five different building blocks. The sublibrary with the highest biological activity reveals which amino acid residue in the fixed position has the highest contribution to biological activity.

Omission libraries and amino acid test mixtures have been developed by Furka [73].

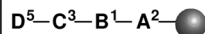


Figure 8.10 Identification of the most active library component by the method of “positional scanning” (positional scanning synthetic combinatorial libraries, PS-SCL). The complete library consists of $5 \times 4 = 20$ sublibraries, whose activities are determined. The most active libraries (framed) indicate the optimal substituents in the sequence position. In the illustrated example these are A², B¹, C³, and D⁵. The most active compound is thus D⁵C³B¹A².

8.2.5

Dynamic Combinatorial Libraries

The concept of dynamic combinatorial chemistry has been developed during the past decade [74, 75]. While the chemical approaches described in the preceding sections rely on discrete, stable compounds or compound mixtures, a dynamic combinatorial library is based on reversible reactions and equilibria. The individual members of the library are interconnected by a network of equilibria, and under thermodynamic

conditions the concentration of each member is dictated by its stability. The building blocks are connected through reversible linkages and, hence, the library is dynamic and adaptive. Any influences that affect this stability will induce shifts in the equilibria, changing the composition of the library. Consequently, a template (receptor) would influence the overall geometry or structure of the predominant product rather than the intrinsic chemistry. The biological target of these mixtures selects the best binders and shifts the equilibrium by binding the best-fitting library structure. Despite being an intriguing concept, dynamic combinatorial chemistry is so far limited to a small number of reversible reactions (e.g. formation of disulfides, imines, hydrazones, acetals).

8.2.6

Biological Methods for the Synthesis of Peptide Libraries

Biological methods for the synthesis of peptide libraries have already been developed [76–78]. Antibody libraries with a high number of binding specificities are of major importance as they enable a more complete therapeutic and diagnostic application of antibody technology, without being dependent on the immunization of animals or the application of eukaryotic cells. Proof of this principle has been shown in investigations on the specific recognition of double-stranded DNA by semisynthetic antibodies [79]. However, the antibodies selected by panning do not yet satisfy the requirement for technical and medical application. In general, tailor-made antibodies can be used for the recognition of certain DNA sequences [80], a finding which may be important for the synthesis of artificial biocatalysts for the neutralization of viruses or toxins, for the identification of certain DNA sequences in gene diagnostics, or in genome assignments as the antibodies block certain DNA sequences and prevent access to other enzymes. Furthermore, the application may be useful in the selection and improvement of catalytic antibodies [81–83].

Phage display is a technology which links the phenotype of a peptide displayed on the surface of a bacteriophage with the genotype encoding for this peptide. This molecular biology technique permits the generation of a peptide library by site-directed mutagenesis. Some 10^7 – 10^9 different oligopeptide sequences can be generated and expressed on the surface of phages, whilst subsequent screening using affinity-based techniques allows the selection of phages that present high-affinity peptides on the surface and which may subsequently be amplified. The uniqueness of a filamentous phage is that three of its five virion proteins tolerate the insertion of foreign peptides. The most studied phages are f1, fd, and M13, all of which infect *E. coli* cells through their f-pili. Among the five coat proteins, pIII, pVI, and pVIII have been used for displaying peptides on the phage surface. The peptide cDNA sequences are inserted into the phage genomes in such a way that they join, or neighbor, the gene for the structural protein pIII. Consequently, the peptides are presented at the N-terminus of pIII after infection of *E. coli* with the phage. pIII is the longest coat protein, and will tolerate the insertion of larger, foreign polypeptides. Unlike pIII, pVIII is a small protein of 50 amino acids where the length of the peptides displayed is limited to five to six amino acids [84]. In early examples of

phage display, polypeptides were fused to the amino terminus of either pIII or pVI on the viral genome level [85], though in several cases the displayed peptides were shown to affect coat protein function. This problem was solved by the development of phagemid display systems.

Peptide libraries with thousands of different peptides can thus be obtained using techniques of molecular biology. Screening is performed by selection and subcloning of phages, and recognition by antibodies. The identification of the active peptide sequence is performed on the DNA level by sequencing of the phage genome [76–78].

Biological compound libraries – and especially phage display libraries – are characterized by the advantage that each library member is able to replicate itself, and that each member carries unique encoding (DNA sequence) [76, 78, 86–90]. However, only genetically encoded amino acids can be employed. Interestingly, phage display techniques have been recently employed not only for retrieving peptide ligands for proteins and other biomolecules, but also for solid inorganic surfaces, which might support assembly, nucleation, and solid phase topology, with implications in bionanofabrication [91].

The concept of *mirror-image phage display* addresses the problem that the peptide sequences selected in the conventional phage display, which are composed of L-amino acids, are not metabolically stable for application as a drug. In mirror-image phage display, biopanning is carried out against the mirror image of the target protein, which has to be synthesized chemically from D-amino acids and has the same sequence as the L-amino acids in the original L-target. As in conventional phage display, peptides consisting of L-amino acids are retrieved that bind to the D-target. Once an L-peptide sequence has been identified as a ligand for the D-target, the corresponding D-peptide will be synthesized. Taking into account the mirror symmetry, the L-target (natural target protein = mirror image of the D-target) will be addressed by the chemically synthesized D-peptide (mirror image of the L-peptide, identified in the phage display experiments) [92, 93].

8.3

Review Questions

- Q8.1. Name the different paradigms of the drug-finding process of the past years.
- Q8.2. Explain the split and combine method. Why is there only one sort of compound on one bead?
- Q8.3. What is the principle of the tea-bag method?
- Q8.4. How does spot synthesis work?
- Q8.5. What is meant by deconvolution? When is it necessary?
- Q8.6. Explain the expression “encoding” in combinatorial chemistry. What methods for encoding do you know?
- Q8.7. Why is photolithography of use in combinatorial chemistry? What is the basic chemistry it relies on?
- Q8.8. What is phage display?
- Q8.9. Explain mirror image phage display.

References

- 1 H.P. Nestler, *Curr. Drug Disc. Technol.* **2005**, 2, 1.
- 2 M.C. Densai, R.N. Zuckermann, W.H. Moos, *Drug Dev. Res.* **1994**, 33, 174.
- 3 R.L. Liu, P.G. Schultz, *Angew. Chem. Int. Ed.* **1999**, 38, 36.
- 4 M.C. Pirrung, *Chem. Rev.* **1997**, 97, 473.
- 5 M. Lebl, *Biopolymers* **1998**, 47, 397.
- 6 A. Nefzi, J.M. Ostresh, R.A. Houghten, *Chem. Rev.* **1997**, 97, 449.
- 7 E.M. Gordon, R.W. Barrett, W.J. Dower, S.P.A. Fodor, M.A. Gallop, *J. Med. Chem.* **1994**, 37, 1385.
- 8 R.M.J. Liskamp, *Angew. Chem. Int. Ed.* **1994**, 33, 633.
- 9 P.C. Andres, D.M. Leonard, W.L. Cody, T.K. Sayer, in: *Multiple and Combinatorial Peptide Synthesis*, B.M. Dunn, M.W. Pennington, (Eds.), Humana Press, Totowa, **1994**.
- 10 L. Weber, *Curr. Opin. Chem. Biol.* **2000**, 4, 295.
- 11 J.C. Hogan, Jr., *Nature Biotechnol.* **1997**, 15, 328.
- 12 G. Jung, A.G. Beck-Sickinger, *Angew. Chem. Int. Ed.* **1992**, 31, 367.
- 13 R.A. Houghten, C. Pinilla, J.R. Appel, S.E. Blondelle, C.T. Dooley, J. Eichler, A. Nefzi, J.M. Ostresh, *J. Med. Chem.* **1999**, 42, 3743.
- 14 C. Blackburn, *Biopolymers* **1998**, 47, 311.
- 15 D.J. Gravert, K.D. Janda, *Chem. Rev.* **1997**, 97, 489.
- 16 F. Balkenhohl, C. von dem Bussche-Hünnefeld, A. Lansky, C. Zechel, *Angew. Chem. Int. Ed.* **1996**, 35, 2288.
- 17 R.J. Booth, J.C. Hodges, *Acc. Chem. Res.* **1999**, 32, 18.
- 18 C. Falciani, L. Lozzi, A. Pini, L. Bracci, *Chem. Biol.* **2005**, 12, 417.
- 19 H.-J. Böhm, M. Stahl, *Curr. Opin. Chem. Biol.* **2000**, 4, 283.
- 20 R.A. Houghten, *Proc. Natl. Acad. Sci. USA* **1985**, 82, 5131.
- 21 H.M. Geysen, R.H. Meloen, S.J. Barteling, *Proc. Natl. Acad. Sci. USA* **1984**, 81, 3998.
- 22 H.M. Geysen, S.J. Barteling, R.H. Meloen, *Proc. Natl. Acad. Sci. USA* **1985**, 82, 178.
- 23 R. Frank, *Tetrahedron* **1992**, 48, 9217.
- 24 Á. Furka, F. Sebestyén, M. Asgedom, G. Dibó, in: *Highlights of Modern Biochemistry*, Proceedings of the 14th International Congress of Biochemistry, Volume 5, VSP, Utrecht, **1988**.
- 25 Á. Furka, F. Sebestyén, M. Asgedom, G. Dibó, *Int. J. Peptide Protein Res.* **1991**, 37, 487.
- 26 R.A. Houghten, *Trends Biotechnol.* **1987**, 5, 322.
- 27 A.G. Beck-Sickinger, H. Dürr, G. Jung, *Peptide Res.* **1991**, 4, 88.
- 28 N.J. Maeji, A.M. Bray, R.M. Valerio, W. Wang, *Peptide Res.* **1995**, 8, 33.
- 29 R. Frank, S. Güler, S. Krause, W. Lindemaier, in: *Peptides 1990*, E. Giralt, D. Andreu (Eds.), Escom, Leiden, **1991** p. 151.
- 30 A.M. Bray, N.J. Maeji, H.M. Geysen, *Tetrahedron Lett.* **1990**, 31, 5811.
- 31 B. Blankenmeyer-Menge, R. Frank, *Tetrahedron Lett.* **1988**, 29, 5871.
- 32 R. Frank, R. Döring, *Tetrahedron* **1988**, 44, 6031.
- 33 J. Eichler, M. Beyermann, M. Bienert, M. Lebl in: *Peptides 1988*, G. Jung, E. Bayer (Eds.), de Gruyter, Berlin, **1989**.
- 34 J. Eichler, M. Bienert, A. Stierandova, M. Lebl, *Peptide Res.* **1991**, 4, 296.
- 35 M. Schmidt, J. Eichler, J. Odarjuk, E. Krause, M. Beyermann, M. Bienert, *Bioorg. Med. Chem. Lett.* **1993**, 3, 441.
- 36 R.H. Berg, K. Almdal, W. Batsberg Pedersen, A. Holm, J.P. Tam, R.B. Merrifield, *J. Am. Chem. Soc.* **1989**, 111, 8024.
- 37 H. Wenschuh, R. Volkmer-Engert, M. Schmidt, M. Schulz, J. Schneider-Mergener, U. Reineke, *Biopolymers* **2000**, 55, 188.
- 38 G. Schnorrenberg, H. Gerhardt, *Tetrahedron* **1989**, 45, 7759.

- 39 H. Gausepohl, M. Kraft, C. Boulín, R.W. Frank, in: *Peptides: Chemistry, Structure and Biology*, J.E. Rivier, G.R. Marshall (Eds.), Escom, Leiden, 1990, 1003.
- 40 A. Patchornik, B. Amit, R.B. Woodward, *J. Am. Chem. Soc.* **1970**, 92, 6333.
- 41 S.P.A. Fodor, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, D. Solas, *Science* **1991**, 251, 767.
- 42 S.P.A. Fodor, R.P. Rava, X.C. Huang, A.C. Pease, C.P. Holmes, C.L. Adams, *Nature* **1993**, 364, 555.
- 43 M. Mutter, E. Bayer, *Angew. Chem. Int. Ed.* **1974**, 13, 88.
- 44 H. Han, M.M. Wolfe, S. Brenner, K.D. Janda, *Proc. Natl. Acad. Sci. USA* **1995**, 92, 6419.
- 45 E. Erb, K.D. Janda, S. Brenner, *Proc. Natl. Acad. Sci. USA* **1994**, 91, 11422.
- 46 J.M. Ostresh, J.H. Winkle, V.T. Hamashin, R.A. Houghten, *Biopolymers* **1994**, 34, 1681.
- 47 K. Burgess, A.I. Liaw, N. Wang, *J. Med. Chem.* **1994**, 19, 2985.
- 48 H.M. Geysen, S.J. Rodda, T.J. Mason, *Mol. Immunol.* **1986**, 23, 709.
- 49 C. Pinilla, J.R. Appel, P. Blanc, R.A. Houghten, *Biotechniques* **1992**, 13, 901.
- 50 K.S. Lam, E. Salmon, E.M. Hersh, V.J. Hruby, W.M. Kazmierski, R.J. Knapp, *Nature* **1991**, 354, 82.
- 51 A. Furka, W.D. Bennett, *Comb. Chem. High Throughput Screen.* **1999**, 2, 105.
- 52 K.S. Lam, M. Lebl, V. Krchnk, *Chem. Rev.* **1997**, 97, 411.
- 53 S.E. Salmon, K.S. Lam, M. Lebl, A. Kandola, P.S. Khattri, S. Wade, M. Pátek, P. Kocis, V. Krchnák, D. Thorpe, S. Felder, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 11708.
- 54 J.K. Chen, W.S. Lane, A.W. Brauer, A. Tanaka, S.L. Schreiber, *J. Am. Chem. Soc.* **1993**, 115, 12591.
- 55 C. Christensen, T. Groth, C.B. Schiødt, N. T. Foged, M. Meldal, *QSAR Comb. Sci.* **2003**, 22, 737.
- 56 S. Brenner, R.A. Lerner, *Proc. Natl. Acad. Sci. USA* **1992**, 89, 5381.
- 57 M.H.J. Ohlmeyer, R.N. Swanson, L.W. Dillard, J.C. Reader, G. Asouline, R. Kobayashi, M. Wigler, W.C. Still, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 10922.
- 58 A. Borchardt, W.C. Still, *J. Am. Chem. Soc.* **1994**, 116, 373.
- 59 H.P. Nestler, P.A. Bartlett, W.C. Still, *J. Org. Chem.* **1994**, 59, 4723.
- 60 C. Chen, L.A.A. Randall, R.B. Millen, A.D. Jones, M.J. Kurth, *J. Am. Chem. Soc.* **1994**, 116, 2661.
- 61 J. Nielsen, S. Brenner, K.D. Janda, *J. Am. Chem. Soc.* **1993**, 115, 9812.
- 62 J.M. Kerr, S.C. Banville, R.N. Zuckermann, *J. Am. Chem. Soc.* **1993**, 115, 2529.
- 63 S. Hobbs De Witt, J.S. Kiely, C.J. Stankovic, M.C. Schröder, D.M. Reynolds Cody, R.R. Pavia, *Proc. Natl. Acad. Sci. USA* **1993**, 90, 6909.
- 64 O.H. Aina, R. Liu, J.L. Sutcliffe, J. Marik, C.-X. Pan, K.S. Lam, *Mol. Pharm.* **2007**, 4, 631.
- 65 N. Winssinger, R. Damoiseaux, D.C. Tully, B.H. Geierstanger, K. Burdick, J.L. Harris, *Chem. Biol.* **2004**, 11, 1351.
- 66 R.L. Affleck, *Curr. Opin. Chem. Biol.* **2001**, 5, 257.
- 67 K.C. Nicolaou, X.-Y. Xiao, Z. Parandoosh, A. Senyei, M.P. Nova, *Angew. Chem. Int. Ed.* **1995**, 34, 2289.
- 68 E.J. Moran, S. Sarshar, J.F. Cargill, M.M. Shahbaz, A. Lio, A.M.M. Mjalli, R.W. Armstrong, *J. Am. Chem. Soc.* **1995**, 117, 10787.
- 69 C. Barnes, S. Balasubramanian, *Curr. Opin. Chem. Biol.* **2000**, 4, 346.
- 70 R.A. Houghten, J.R. Appel, S.E. Blondelle, J.H. Cuervo, C.T. Dooley, C. Pinilla, *Biotechniques* **1992**, 13, 412.
- 71 C.T. Dooley, R.A. Houghten, *Life Sci.* **1993**, 52, 1509.
- 72 Á. Furka, F. Sebestyén, WO93/24517 **1993**.
- 73 Á. Furka, *Drug Dev. Res.* **1994**, 33, 90.
- 74 S. Otto, R.L.E. Furlan, J.K.M. Sanders, *Curr. Opin. Chem. Biol.* **2002**, 6, 321.

- 75 P.T. Corbett, J. Leclaire, L. Vial, K.R. West, J.-L. Wietor, J.K.M. Sanders, S. Otto, *Chem. Rev.* **2006**, 106, 3652.
- 76 S.E. Cwirla, E.A. Peters, R.W. Barrett, W.J. Dower, *Proc. Natl. Acad. Sci. USA* **1990**, 87, 6378.
- 77 J.A. Wells, H.B. Lowman, *Curr. Opin. Biotechnol.* **1992**, 3, 355.
- 78 J.K. Scott, G.P. Smith, *Science* **1990**, 249, 386.
- 79 S.M. Barbas, P. Ghazal, C.F. Barbas III, D.R. Burton, *J. Am. Chem. Soc.* **1994**, 116, 2161.
- 80 M. Famulok, D. Faulhammer, *Angew. Chem. Int. Ed.* **1994**, 33, 1827.
- 81 Y.C.J. Chen, T. Danon, L. Sastry, M. Mubarak, K.D. Janda, R.A. Lerner, *J. Am. Chem. Soc.* **1993**, 115, 357.
- 82 B. Posner, J. Smiley, I. Lee, S. Benkovic, *Trends Biochem. Sci.* **1994**, 19, 145.
- 83 J.D. Stewart, S.J. Benkovic, *Int. Rev. Immunol.* **1993**, 10, 229.
- 84 S. Cabilly, *Mol. Biotechnol.* **1999**, 12, 143.
- 85 S.S. Sidhu, *Curr. Opin. Biotechnol.* **2000**, 11, 610.
- 86 M.B. Zwick, J. Shen, J.K. Scott, *Curr. Opin. Biotechnol.* **1998**, 9, 427.
- 87 J.J. Devlin, L.C. Panganiban, P.E. Devlin, *Science* **1990**, 249, 404.
- 88 G.P. Smith, V.A. Petrenko, *Chem. Rev.* **1997**, 97, 391.
- 89 A. Pini, A. Giuliani, C. Ricci, Y. Runci, L. Bracci, *Curr. Protein Pept. Sci.* **2004**, 5, 487.
- 90 M. Paschke, *Appl. Microbiol. Biotechnol.* **2006**, 70, 2.
- 91 F. Baneyx, D.T. Schwartz, *Curr. Opin. Biotechnol.* **2007**, 18, 312.
- 92 T.N. Schumacher, L.M. Mayr, D.L. Minor, Jr., M.A. Milhollen, M.W. Burgess, P.S. Kim, *Science* **1996**, 271, 1854.
- 93 K. Wiesehan, D. Willbold, *ChemBioChem* **2003**, 4, 811.

9

Application of Peptides and Proteins

Almost all biological processes in living cells are controlled by molecular recognition. Regulatory processes comprise initiation or inhibition via specific protein–protein complex formation. In particular, peptides and proteins possess an enormous potential for diversity, and therefore these compounds are well suited for such complicated control functions. Therefore, peptides and proteins have the potential to be potent pharmaceutical agents for the treatment of many diseases with a broad range of clinical benefits. Actual peptide therapies comprise metabolic diseases, viral indications, cancer, cardiovascular diseases, neurological disorders, microbial and fungal diseases, as well as immune system disorders.

9.1

General Production Strategies

Therapeutic peptides and proteins have, traditionally, been obtained from various sources:

- Isolation from Nature providing bioactive peptides and proteins;
- Chemical synthesis using different strategies (cf. Chapters 4 and 5) including chemical libraries (cf. Chapter 8) and design of peptidomimetics (cf. Chapter 7);
- Recombinant synthesis (cf. Section 4.6.1) including recombinant display technologies (phage, yeast, bacteria, DNA/RNA);
- Expression by transgenic animals and plants;
- Monoclonal antibodies and fusion proteins;

At present, the majority of peptide and protein therapeutics are derived from natural sources. The advantage is that bioactive peptides have undergone natural selection resulting in enhanced *in vivo* stability [1]. Furthermore, they are highly functional, acting as potent agonists and antagonists against different receptors involved in pathological settings. Representative examples are ghrelin (see Section 3.3.1.7) to treat obesity, gastrin-releasing peptide applied in cancer treatment, and glucagon-like peptide-1 (see Section 3.3.1.2) used for the control of diabetes. Native peptides show higher affinity/specificity to target receptors and

lower toxicity profiles compared to small molecules, and display, in contrast to proteins, like for example antibodies, better tissue penetration and room temperature storage behavior. Despite this potential, their low metabolic stability, rapid renal elimination, and the need for effective and patient-friendly delivery technologies are important shortcomings.

Chemical synthesis is at present a useful option for peptides in development comprising SPS, SPPS and the hybrid approach of both strategies. During R&D and early-stage clinical trials chemical synthesis is the preferred production procedure. Cost estimates in small-scale production run from \$ 20 to \$ 60 per amino acid residue for 10 mg final product. However, depending on the total quantity to be synthesized, the price of a synthetic peptide (per amino acid residue) may range from \$ 300 to \$ 500 per g for 300- to 500-g quantities, from \$ 100 to \$ 200 per g for 1- to 2-kg quantities, from \$ 25 to \$ 50 per g for 50- to 100-kg quantities, and less than \$ 10 per g on higher production scales. Many companies are capable of carrying out custom-made peptide synthesis in a short time (within days or weeks) at moderate cost [2]. It can be assumed that chemical manufacture of peptides over 35 aa is generally not economically feasible. One exception seems to be the 36-peptide enfuvirtide (T20, Fuzeon™) which had already been synthesized by Trimeris at lower cost several years ago. More information on large-scale peptide synthesis is given in Section 9.4.1. The most recent boost in large-scale peptide production is connected with Roche's collaboration with Trimeris that led to the industrial manufacturing of enfuvirtide (T20; Fuzeon™). This initiated a revolution in peptide drug manufacturing and led to the installation of large-scale synthesis capacity at Roche with tremendous effects on an entire industrial branch. More details of the enfuvirtide production on a metric ton scale are given later in this chapter and in Section 5.3.4. It was reported that the global market for peptide-based active pharmaceutical ingredients (APIs) is expected to expand at a growth rate approximately double the growth rate of APIs overall. Today the US holds 65% of the worldwide therapeutic peptide market followed by Europe, led by Germany and the UK, with 30% of the total, whereas Japan dominates activities in Asia.

In the past, it was difficult to discover stable and potent peptides from peptide libraries since a large proportion of random peptides in a library were missing stable folds and had other intrinsic disadvantages of non-native peptides. With the creation of new recombinant display technologies characterized by ever-increasing levels of diversity, it is now possible to find high-affinity peptides against most protein targets that are pathologically relevant [3, 4]. Large peptide diversity could be achieved by the increasing rigidity and lack of entropic freedom of highly structured peptides. Nowadays, peptides are available that even match antibodies with respect to affinity.

Recombinant synthesis is an alternative for developing peptide therapeutics longer than 50 aa that are difficult to obtain economically by chemical procedures. In the past, many proteins, and particularly human growth hormone, were prepared from bacteria. However, this production technique was not without shortcomings since bacteria cannot synthesize complex proteins such as monoclonal antibodies or coagulation blood factors, which require post-translational modifications to be active and/or stable *in vivo*. These modifications include mainly folding, cleavage, subunit

association, γ -carboxylation and glycosylation. As discussed in Section 4.6.1.4, the concept of expanding the genetic code established by Schultz and coworkers allows the incorporation of non-proteinogenic amino acids of any type of proteins and therefore circumvents an important limitation of the recombinant manufacture. Ambrx calls this process ReCODE, reconstituting chemically orthogonal directed engineering.

Transgenic expression by eukaryotic cells or species even adds another level of complexity to recombinant synthesis [5]. Mammalian cells can be cultured in fermentors to produce peptides and proteins on an industrial scale. Transgenic animal species can produce recombinant proteins. Currently two systems are being implemented. Milk from transgenic mammals as a source for recombinant proteins has been studied for two decades and the protein human antithrombin III was approved by the European Medicines Agency EMEA in 2006 to be launched on the market. The second system is chicken egg white which recently became more attractive after essential improvement of the methods used to generate transgenic birds. For example, two monoclonal antibodies and a human interferon- β could be recovered from chicken egg white. A broad variety of recombinant proteins such as monoclonal antibodies, vaccines, blood factors, hormones, growth factors, cytokines, enzymes, milk proteins, collagen, fibrinogen and others have been produced experimentally by using these systems and a few others. Although these possibilities have not yet been optimized and are still being improved, a new era in the production of recombinant pharmaceutical proteins was initiated in the mid-eighties and became a reality two decades later. Transgenic plants [6] have proven to be a promising tool for production of recombinant proteins. Transgenic plants as bioreactors will allow the development of a large number of potential therapeutic proteins into successful pharmaceuticals by enabling large scale production at low cost. A first non-therapeutic tumor imaging antibody has been produced in transgenic maize by Monsanto in cooperation with NeoRx who provided the antibody. Furthermore, plant bioreactors for production of a variety of different peptides are under development and will dramatically change the cost and availability of these new materials. One transgenic plant-derived product, hirudin, is now being commercially produced in Canada for the first time. Plants have considerable potential for the production of biopharmaceutical proteins and peptides because they are easily transformed and provide a cheap source of protein.

Initially, the pharmaceutical community was excited about the market potential of peptides and proteins for diagnostic and therapeutic purposes. However, decades after the first chemical synthesis of a peptide hormone, very little advantage has yet been taken of the potential of peptides and proteins as pharmaceuticals and tools for application in basic and clinical research. Only relatively few peptides have been approved as drugs in the past, most likely because the application of proteins in therapeutic use may be hampered by factors such as antigenicity, immunogenicity, and stability of the protein. An important drawback of peptides has been their low bioavailability and the need to use other than oral routes for administration.

9.2

Improvement of the Therapeutic Potential

Peptides and proteins have seen limited use as clinically viable drugs. An important shortcoming is their low metabolic stability and their rapid renal elimination. They are degraded by a variety of proteases and often have very short *in vivo* half-lives. Type and concentration of proteases varies with body compartment. Peptides and proteins within the systemic circulation are generally eliminated through both renal and hepato-biliary routes, based on the molecular size and lipophilicity of the respective biomolecule. Generally, the size and structure of these biomolecules can have a significant impact on their bioavailability. The engineering of therapeutic peptides and proteins [7] provides a valuable means of circumventing some of the disadvantages mentioned above. The goals of engineering techniques are to prolong and/or enhance the biological activity of the appropriate drug at the target site. This can be accomplished by

- minimizing immunogenicity,
- increasing bioavailability,
- reducing elimination,
- improving pharmacokinetics,
- improving affinity/selectivity for the receptor.

Such a design approach is best characterized as interdisciplinary, comprising the introduction of conformational constraints, bond replacements and a variety of other modifications directed towards stability, receptor affinity, membrane permeability, elimination, and a couple of independent attributes to the human body.

9.2.1

Peptide and Protein Drug Modifications

The therapeutic utility of many bioactive proteins and peptides is limited by their short serum half-life. Polymer conjugations are applied to reduce immunogenicity and elimination and to increase stability. Tools for chemical modification of peptides and proteins are macromolecular polymers such as poly(ethylene glycol) (PEG) [8], and poly(styrene maleic acid) [9].

PEGylation is commonly used to reduce renal elimination and to reduce the dosing frequency of protein therapeutics. The positive properties of PEG–peptide conjugates include high water solubility, high mobility in solution and low immunogenicity [10]. PEG can increase the hydrodynamic radius of peptide conjugates, thus preventing renal clearance, can produce improved physical and thermal stability, and minimize enzymatic degradation. Only 1% of the 69 kDa protein albumin is found in the glomerular filtrate. In order to mimic the size of albumin and other larger proteins, PEG peptide conjugates have been prepared with a single large 30–40 kDa PEG or with multiple small PEGs of about 5 kDa in size. PEG is often attached to the N or C termini of peptides, but can also be attached within the peptide sequence to side-chain functionalities of cysteine, lysine, or unnatural

amino acids. A principal disadvantage of PEGylation is the potential loss of activity in the case of improper choice of PEG with regard to branching, length, chemical design, or attachment site. The US FAD (US Food and Drug Administration) has regarded PEG-derivatives as compounds safe for application as a vehicle in pharmaceuticals, food and cosmetics. Roche's deal with Gryphon Therapeutics to promote an erythropoietin analogue with PEG modification underlines the increasing activities of big pharmaceutical companies to put exciting developments of biotech companies into commercial practice.

However, chemical conjugation of PEG to proteins has a number of limitations, mainly high manufacturing costs and the generation of product mixtures including inactive isomers. Recently, it has been reported that Amunix have developed recombinant polypeptide chains with PEG-like properties (called rPEGs) that can be directly fused to therapeutic proteins [11]. rPEG-fusion proteins are well-defined chemical species, thus avoiding the product heterogeneity associated with chemical PEGylation. rPEG fusion proteins can be produced in high yield microbial fermentation. The unique chemical properties of rPEG greatly facilitate product purification. The hydrophilic nature of rPEG improves product solubility and prevents aggregation.

Chemical modifications of the biomolecule itself are another alternative besides conjugation to PEG. Considerable efforts have been made to design peptide-derived compounds with improved stability and the ability to mimic peptide function by peptidomimetics (cf. Chapter 7). Further possibilities include N-terminal (glycosylation, acetylation) or C-terminal (amidation) modifications, incorporation of unnatural building blocks such as D-amino acids, β -amino acids and C $^{\alpha,\alpha}$ -disubstituted amino acids, and cyclization to decrease the conformational flexibility of linear peptides and increase the stability against proteolytic degradation.

Fusion proteins (FP), which are available by genetic fusion of peptides to the Fc domain of human gamma immunoglobulin (IgG), are an interesting option to increase peptide molecular size. This approach takes advantage of the IgG protection function of the neonatal Fc receptor [12], which has been used for the development of a novel drug delivery platform (cf. Section 9.2.2). For example, with alefacept (Amevive) for chronic plaque psoriasis [13], abatacept (Orencia) [14], and etanercept (Enbrel) for rheumatoid arthritis, three proteins fused to Fc have been approved as pharmaceuticals. In principle, it is a disadvantage that Fc fusions are dimeric, which might result in lower drug potency caused by steric hindrance. Strategies for monomeric fusion to the dimeric Fc or optimization of the linker length have been described [12]. Fusion proteins are an alternative to the engineering of humanized mAb (cf. Section 9.3.3.3).

Peptibodies are hybrids consisting of a small peptide moiety and an antibody. AMG 531 is Amgen's first peptibody and potentially represents a new approach to the treatment of idiopathic thrombocytopenic purpura (ITP), an autoimmune bleeding disorder. AMG 531 is a mimic of thrombopoietin fused to Fc [15]. An analgesic peptibody from Amgen targeted to nerve growth factor (NGF) reduced thermal hyperalgesia and tactile allodynia – over-sensitized pain states – in rat models of neuropathic pain [16].

Serum albumin-peptide fusion conjugates are useful degradation-resistant therapeutics, based on the properties of albumin, providing stability as well as greatly reduced renal clearance. Serum albumin has been used for fusion of peptide drugs to the C-terminus in order to maintain drug potency. The interferon- α peptide Albuferon has been modified according to this procedure and shows antiviral activity in patients with chronic hepatitis C [17]. AlbuBNP is a BNP peptide fused to albumin which is preclinically tested for the therapy of congestive heart failure [18]. The albumin-GLP-1 fusion peptide Albugen has been investigated for the treatment of type-2 diabetes [19]. Further chemical ligations of peptides to albumin are performed by linkage to the free Cys³⁴ of albumin, resulting in an extended half-life of peptide therapeutics.

Avimers are a class of binding proteins that overcome the limitations of antibodies and other immunoglobulin-based therapeutic proteins [20]. They are multimers of serum-stable constrained peptide domains derived from a family of human receptor domains. Avimers are smaller than antibodies and were first discovered using exon shuffling and phage display. They bind protein targets with high affinity, and improve thermostability and resistance to proteases. Avimers with sub-nM affinities were obtained against various targets.

9.2.2

Peptide Drug Delivery Systems

During recent years, progress in the areas of formulation and delivery systems has led to the development of several highly successful peptide drugs. A most difficult challenge for peptide therapeutics is the need for effective and patient-friendly delivery technologies. The main goal of these improvements was not only to overcome a lack of oral bioavailability, but also to avoid the need for subcutaneous injection, which often leads to poor patient compliance. Several possibilities of administering peptides and proteins by either oral, pulmonary, mucosal membrane or transcutaneous routes have been reported. These routes of administration very often require specific delivery vehicles and/or permeability enhancers to assist transfer of the drug across the delivery site and into the systemic circulation. Interesting alternatives include nasal sprays for LH-RH (Buserelin), calcitonin, oxytocin and vasopressin, and rectal suppositories for calcitonin. Ointments are often used for the transdermal application of peptides, but sublingual administration is another possibility. Structure–function studies have led to the design of new peptide derivatives suitable for oral administration, such as the vasopressin analogue desmopressin. Modified analogues of somatostatin are available which retain the pharmacological properties of the parent hormone but exhibit a significantly prolonged duration of action. Following administration, many peptide and protein biopharmaceuticals exert their intended action in the systemic circulation, and must therefore resist clearance by conventional mechanisms, including molecular filtration by the kidney and clearance by the reticuloendothelial system. As shown above, PEGylation of peptides and proteins yields PEG-conjugated derivatives with reduced renal clearance and a more than 50-fold enhancement of circulatory half-life.

Further developments achieved in drug delivery systems for peptide and protein pharmaceuticals will continue to increase the therapeutic application of these materials. Pettit and Gombotz [21] defined site-specific drug delivery as delivery through a specific site (i.e., the route of administration), as well as delivery to a specific site (i.e., the site of action). The physical and chemical characteristics of both the peptide to be delivered, and the site to be targeted, must be especially considered in development of the appropriate technology. A synthetic polymer, device or carrier system may be introduced to target the biopharmaceutical to a specific site within the body. Selected examples of site-specific drug delivery are listed in Table 9.1.

The delivery of large therapeutic proteins is currently performed by injection since they are generally absorbed poorly across epithelial surfaces. An interesting drug delivery platform based on a naturally occurring receptor-mediated transport pathway has been developed in order to deliver large protein pharmaceuticals non-invasively. The neonatal Fc receptor (FcRn) is specific for the Fc fragment of IgG and is expressed in epithelial cells where it functions to transport immunoglobulins

Table 9.1 Selected examples of site-specific drug delivery according to Pettit and Gombotz [21].

Site targeted	Remarks
Route of administration	
Transdermal	Assisted by iontophoresis or ultrasound
Pulmonary	Liquid and dry-powder aerosol delivery
Mucous membranes	Aerosol-mucin charge interactions
Oral/intestinal	Small particles, protein-carrier complexes
Specific tissues or organs	
Tumors	Neovascularization markers are targeted
Lungs	Aerosol, liposomal delivery
Brain	Target the transferrin receptor
Intestines	Protect against proteolysis and acid hydrolysis
Eyes	Mucin charge interactions
Uterine horns	Form biodegradable gel <i>in situ</i>
Bones	Hydroxyapatite binds bone-promoting growth factor
Skin	Methylcellulose gels
Cellular/intracellular	
Macrophages	Small particles are phagocytosed
Tumor cells	Fusogenic liposomes to deliver intracellular toxins
Molecular targets	
Tumor antigens	Antibody–enzyme conjugates activate prodrugs
Fibrin/site of clot formation	Fusion proteins combine targeting with toxin
Carbohydrate receptors	Mannose and galactose used to target receptor
Systemic circulation	
Injection	Prolong or sustain circulation

across these cell barriers. FcRn is expressed in both the upper and central airways in non-human primates as well as in humans. Previously, monomeric erythropoietin-Fc (EpoFc) has been successfully administered by inhalation [22].

Generally, one of the main problems with most known delivery technologies is the resulting requirement for higher dosages. Despite the fact that these dosages are normally well tolerated, the resulting higher cost has been a real problem.

Cell-penetrating peptides (CPP), also called *Trojan horse peptides* and *protein transduction domains*, are peptides of different structural classes that are capable of crossing the plasma membranes of mammalian cells in an apparently energy- and receptor-independent fashion [23–25]. CPP translocate rapidly into cells and act as peptidic delivery factors. They have found application for the intracellular delivery of macromolecules with molecular weights several times greater than their own. In order to differentiate them from larger proteins that have been shown to function as transporters across biological membranes, CPP on average contain no more than 30 aa residues. According to the proposed classification, CPPs are arranged into three classes: (i) protein-derived CPPs, (ii) model peptides, and (iii) designed CPPs. *Protein-derived CPPs*, also designated as *protein transduction domains* or *membrane translocation sequences*, usually consist of the minimal effective partial sequence of the parent translocation protein. To the first group belong *penetratin*, RQIKIWFQNRR¹⁰RMKWKK, corresponding to *Drosophila* antennapedia homeodomain-(43–58), *tat fragment*-(48–60), GRKKRRQRRR¹⁰PPQ, derived from human immunodeficiency virus 1 protein-(48–60), and *pVEC*, LLIILRRRIR¹⁰KQAHAAHSKa, derived from murine vascular endothelial cadherin. *Model CPPs* consist of sequences that have been designed with the aim of obtaining well-defined amphipathic α -helical structures, or to mimic the structures of known CPPs. Members of this group are (Arg)₇, RRRRRRRR, and MAP, KLALKLALKA¹⁰LKAALKLAa. *Designed CPPs* are usually chimeric peptides comprising hydrophilic and hydrophobic domains of different origin. MPG, GALFLGFLGA¹⁰AGSTMGAWSP²⁰KSKLRKV, derived from the fusion sequence of HIV-1 gp41 protein coupled to a peptide derived from the nuclear localization sequence of SV40 T-antigen, and *transportan*, GWTLNSAGYL¹⁰LGKINLKALA²⁰ALAKISILa, derived from the minimally active part of galanin-(1–12) coupled to mastoparan via Lys¹³, are further members of this class. The penetration is to some degree an energy-independent mechanism of peptide translocation across the cell membrane. The sequence of CPP allows the addressing of cargoes into the cytoplasm and/or the nucleus. The mechanism of cellular translocation by CPPs is still not fully understood, although macropinocytosis seems to be the commonly assumed route. It is likely that CPPs from the different groups act by distinct transport mechanisms. For many CPPs, the cargoes must be covalently conjugated, but in some cases (MPG) a mixture is sufficient. Independent of the binding of the cargo, an excess of CPP is necessary. Examples of cargoes internalized by CPPs include the transport of a fibroblast growth factor (FGF) receptor phosphopeptide by penetratin to inhibit FGF receptor signaling in living neurons, and internalization of the 21-mer galanin receptor antisense by penetratin or transportan in order to regulate galanin receptor levels and modify pain transmission *in vivo*. A broad range of therapeutics, such as proteins, DNA,

antibodies, oligonucleotides, PNAs and imaging agents, are translocated by CPPs into target cells. Until now, CPP-based technologies have served as useful tools in biomedical research, especially due to their non-invasive and efficient delivery of bioactive molecules into cells, both *in vitro* and *in vivo*. CPPs with an affinity towards actively proliferating cells are of special importance, as they open new vistas in cancer and developmental biology research.

Radiolabelled tumor-specific peptides have found application for diagnostic purposes. For example, they may be used *in vitro* on tumor sections to obtain information on the so-called receptor status of cells; alternatively, they may be injected into the body in order to locate tumors. ^{125}I is a useful radioligand for *in vitro* application, whereas the short-lived ^{123}I is more suitable for *in vivo* administration. Nowadays, peptides labelled with radioactive iodine isotopes are increasingly replaced by peptide derivatives equipped with chelators for ^{111}In or ^{99}Tc . Radiolabelled peptides appear to be very useful both for tumor diagnosis in cancer patients, and for tumor therapy.

An interesting approach to cancer chemotherapy is based on the targeting of cytotoxic peptide conjugates to their receptors on tumors. Cytotoxic conjugates are hybrid molecules consisting of a peptide carrier (which binds to the receptors that are up-regulated on tumors) and a suitable cytotoxic moiety. An early example of hormone drug conjugates was that of the DNA intercalator daunomycin linked to the N-terminal amino group of Asp, and also to the ϵ -amino groups of Lys residues of the β -melanocyte stimulating hormone [26]. Furthermore, cytotoxic compounds such as doxorubicin linked to LH-RH, bombesin, and somatostatin could be targeted to certain tumors that expressed specific peptide-receptors in higher numbers than normal cells. Consequently, these conjugates were seen to be especially lethal for cancer cells [27]. Novel chemically modified analogues of neuropeptide Y for tumor targeting have been described by Beck-Sickinger and coworkers [28]. Especially, the Y_1 -receptor selective $[\text{Lys}(\text{DOTA})^4, \text{Phe}^7, \text{Pro}^{34}]\text{NPY}$ (DOTA: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid, a chelate ligand for metal ions) labelled with ^{111}In has been proven to be a very promising analogue. From *in vitro* and *in vivo* studies it has been suggested that receptor-selective NPY analogues have promising properties for future applications in nuclear medicine for breast tumor diagnosis and therapy. Peptides that target tumor blood vessels have been identified by phage display and coupled to anticancer drugs [29–31]. Tumor-targeting peptide oligonucleotide conjugates have been described for the application of antisense oligodeoxynucleotides as therapeutic agents inhibiting gene expression [32].

The blood–brain barrier (BBB) limits the transfer of soluble peptides and proteins via passive diffusion through the brain capillary endothelial wall. However, the BBB permeability of potent pharmaceuticals is required for the treatment of many CNS-derived diseases. In order to target peptides to the CNS consideration must be given to both increasing bioavailability and enhancing brain uptake. To date multiple strategies have been studied, but each strategy is associated with its own set of complications and considerations [33]. The capability of peptides to cross the BBB and enter the brain depends on several factors such as size, conformation, flexibility, as well as amino acid composition and arrangement. For these reasons, special methods for modification of peptide drugs are required

which are more complex than those discussed in Section 9.2.1. Especially, glycosylation has proven to be a useful tool for enhancing biodistribution to the brain. Glycosylated opioid peptides show improved analgesia and higher metabolic stability which might be due to the increased bioavailability. Structural modifications are very important to enhance stability. In the case of Met-enkephalin, the conversion into the cyclic analog DPDPE, H-Tyr-D-Pen-Gly-Phe-D-Pen-OH (disulfide bond: D-Pen²-D-Pen⁵), resulted in a δ -opioid specific peptide analog with a saturable mode of transport at the BBB [34]. Further possibilities for enhancing BBB permeability are lipidization, cationization, vector-based strategies, and the use of prodrugs and nutrient transporters.

9.3

Protein Pharmaceuticals

9.3.1

Importance and Sources

Proteins constitute a major fraction of the biopolymers present in all organisms with respect to diversity and mass. A huge proportion of these biomolecules have regulatory functions in maintaining biochemical or cellular equilibria in healthy organisms, though they may also be involved in both pathophysiological events and healing processes. Until the late 1970s, the human body was the only source of endogenous proteins such as growth hormone or coagulation factor VIII used for replacement therapy. The selection of a suitable protein source for its isolation was based on such criteria as the ease of obtaining sufficient quantities of the appropriate tissue, the amount of the chosen protein in this tissue, and any properties that would aid in their stabilization and isolation. Preferentially, tissues or organs from domesticated animals, easily obtainable microorganisms and plants were chosen as sources for the isolation. Isolated proteins, sometimes in the form of poorly defined mixtures, have been used in many traditional medicines and in so-called alternative medicine. However, purified active proteins are of great importance for both causal and symptomatic treatment, and for prophylaxis.

Since its introduction in 1977, the use of recombinant DNA technology (Section 4.6.1) has provided a new and highly efficient means of producing large amounts of rare and/or novel proteins. The development of molecular cloning techniques offers a new production method for proteins, and has consequently exerted an enormous medical, industrial and agricultural impact. Once a protein-encoding gene has been isolated from its parent organism, it may be genetically engineered if desired, and overexpressed in either bacteria, yeast, or mammalian cell cultures. The biotechnological isolation of a recombinant protein is much easier as it may constitute up to about 35% of the overproducer's total cell protein. The recombinant polypeptide- and protein-based drugs present in the market include a very wide range of compounds such as hormones, enzymes, monoclonal antibodies, vaccines, vaccines, radio-immuno conjugates and various cellular factors.

9.3.2

Endogenous Pharmaceutical Proteins

The developments in molecular biology have led to a therapeutic concept based on the pharmaceutical application of endogenous proteins which includes

- the discovery and synthesis of proteins with therapeutic potential by gene technology,
- the elucidation of their biological actions *in vitro* and *in vivo*,
- the development of drugs based on the primary protein lead molecule.

Very important indications for pharmaceutical proteins include cancer, infectious diseases, AIDS-related diseases, heart disease, respiratory diseases, autoimmune disorders, transplantations, skin disorders, diabetes, genetic disorders, digestive disorders, blood disorders, infertility, growth disorders, and eye conditions. In the last decade cancer was by far the most prevalent target, accounting for more than 40% of the total number of new medicines according to disease area. The top five drug types were vaccines, monoclonal antibodies, gene therapeutics, growth factors, and interferons [35].

A list of proteins of general pharmaceutical interest is provided in Table 9.2. Most of these proteins were cloned during the 1980s [36], at which time examples of approved protein-based products included epidermal growth factor (EGF), Factor VIII, tissue plasminogen activator (tPA), insulin, hepatitis B vaccine, various interferons, monoclonal antibodies, and growth hormone. Many of these proteins are in the meantime produced by recombinant synthesis.

Many of the early protein drug candidates failed in clinical trials due to their immunogenicity, short half-life, or low specificity. It has been estimated that up to end of the last century, about 100 drugs produced by biotechnology had been approved, but that approximately 350 biotechnology drugs are currently under development. Initially, many pharmaceutical proteins were of nonhuman origin, and caused immune responses against the drug itself. Others suffered from suboptimal affinity or poor half-life, resulting in poor efficacy.

9.3.3

Engineered Protein Pharmaceuticals**9.3.3.1 Selected Recombinant Proteins**

Therapeutic proteins with various biological actions, including growth factors, interferons, interleukins, tissue plasminogen activators, clotting factors, colony stimulating factors, erythropoietin and others, have also been engineered to improve the effectiveness as protein therapeutics in the clinic or clinical development pipeline.

Recombinant DNA technology [37] is the preferred method of production (for more details see Section 4.6.1). Peptides and proteins that do not require post-translational modification, for example insulin, can be synthesized in produced in prokaryotes. Shorter peptides are often expressed as fusion proteins to protect them from proteolysis and to increase process efficiency. Such fusion proteins contain a cleavage

Table 9.2 Selected pharmaceutical proteins.^a

Protein (abbreviation)	aa	Isolated from	Indication/mode of action
Albumin (HSA)	585	Liver (1975)	Plasma expander
Angiogenin (TAF)	123	Bowel cancer cells (1985)	Wound healing; tumors
α_1 -Antitrypsin (AAT)	394	Blood (1978)	Anticoagulant
Antithrombin III (AT3)	432	Liver (1979)	Anticoagulant
Erythrocyte differentiation factor (EDF)	110	Leukemia cells (1987)	Tumors
Erythropoietin (EPO)	165	Urine (1977)	Aplastic anemia
Factor VII	406	Plasma (1980)	Blood clotting
Factor VIII	2332	Liver (1983)	Hemophilia A
Factor IX	416	Plasma (1975)	Hemophilia B
Factor XIII	1372	Plasma (1971)	Surgical adhesive
Fibroblast growth factor (basic) (bFGF)	146	(1986)	Wound healing, tumors
Fibronectin (FN)	96	(1970)	Wound healing
Granulocyte colony-stimulating factor (G-CSF)	174–177	Tumor cells (1986)	Leukemia, other tumors
Granulocyte macrophage colony stimulating factor (GM-CSF, CSF-2)	127	T cell (1984)	Anemia, tumors
Hepatitis B surface antigen (HBS, HbsAg)	226	Virions (1977)	Hepatitis vaccine
Human collagenase inhibitor (HCI, TIMP)	184	Fibroblasts (1983)	Arthritis
Interferon- α (IFN- α)	166	Leucocytes (1979)	Hairy cell leukemia, tumors
Interferon- β (IFN- β)	166	Fibroblasts (1979)	Keratitis, hepatitis B
Interferon- γ (IFN- γ , MAF)	146	Lymphocytes (1981)	Tumors, arthritis
Interleukin-1 (IL-1, ETAF, LAF)	152	Neutrophils (1984)	Tumors
Interleukin-2 (IL-2, TCGF)	133	T cells (1980)	Tumors
Interleukin-3 (IL-3, Multi-CSF, BPA, MCGF)	133		Leukemia, other tumors
Interleukin-4 (IL-4, BSF-1, BCGF-1)	129		Leukemia, infections
Interleukin-5 (IL-5, TRP, BCGF-II)	112		Autoimmune diseases
Interleukin-6 (IL-6, BSF2, IFN- β 2, BCDF)	184	T cells (1985)	Leukemia
Lipase	135	microorganisms	Digestive disturbances
Lipomodulin, lipocortin (AIP)	346		Arthritis, allergies
Lung surfactant protein (LSP, PSF)	248	Sputum (1986)	Emphysema
Lymphotoxin (LT, TNF- β)	171	Lymphocytes (1984)	Tumors
Macrophase inhibitory factor (MIF)	Het.	Lymphocytes (1981)	
Macrophage colony stimulating factor (CSF-1, M-CSF)	224	Urine (1982)	Leukemia, tumors
Monoclonal antibody OKT3, Orthoclone OKT3		Hybridoma (1979)	Transplantation
Nerve growth factor (NGF- β)	118		Injuries
Platelet-derived growth factor (PDGF)	241	Platelets (1983)	Wound healing

Table 9.2 (Continued)

Protein (abbreviation)	aa	Isolated from	Indication/mode of action
Plasminogen activator (PAI I)	376–379	Lymphosarcoma (1984)	Blood clotting
Protein C (PC)	262	Plasma (1979)	Anticoagulant
Protein S	635		Anticoagulant
Streptokinase	416	Streptococcus	Myocardial infarct, thrombosis
Superoxide dismutase (SOD)	153	Placenta (1972)	After-treatment of myocardial infarct
Tissue plasminogen activator (tPA)	527	Uterus (1979)	Myocardial infarct, embolism
Transforming growth factor- α (TGF- α)	50	Tumor cells (1982)	Wound healing
Transforming growth factor- β (TGF- β)	112	Kidney tumor (1983)	Wound healing, tumors
Tumour necrosis factor (TNF- α , cachectin, DIF)	157	Tumor (1985)	Tumors
Urokinase (UK)	366	Urine (1982)	Thromboses, embolism
Uromodulin, Tamm-Horsfall protein	616	Urine (1985)	Inflammations

^aBased on data published by Blohm *et al.* [36].

site, e.g., for a protease or cyanogen bromide (cleaves Met) located N-terminally in order to liberate the target peptide during work-up. When proper folding, assembly, and post-translational modification of the target protein is a prerequisite, the quality and efficacy can be largely improved by utilizing cultivated mammalian cells for production. About 60–70% of all recombinant protein pharmaceuticals are produced in mammalian cells. For this purpose, Chinese hamster ovary (CHO) cells or human embryo kidney (HEK-293) cell lines are used preferentially. More than 200 peptide and protein pharmaceuticals have been approved by the FDA in US.

Acute myocardial infarction and other thrombotic obstructions of blood vessels are indications for the therapy with recombinant *tissue plasminogen activator* (tPA). This compound belongs to the ‘big’ pharmaceutical market products. It might be demonstrated that, in the case of tPA, the removal of natural domains may improve the pharmacokinetics and specificity of the protein drug. Retaplast (Boehringer Mannheim) is an extremely truncated tPA molecule lacking the N-terminal finger domain, the epidermal growth factor domain, and the kringle 1 domains. The resultant drug (Rapilysin), which has an improved half-life compared with Retaplast, is used in the treatment of myocardial infarction. *Human serum albumin* maintains plasma colloid osmotic pressure and serves as a carrier of both intermediate metabolites and various therapeutics, as discussed in Section 9.2.1. It is applied for symptomatic relief and supportive treatment in the management of, e.g., shock and burns.

Pharmaceutical proteins, which act on immunological functions include, particularly, the interferons and interleukins as well as the growth and differentiation factors, the latter being highly specific triggers of the differentiation steps in

hematopoiesis. Angiogenesis factors such as angiogenin and fibroblast growth factors are normally not distributed throughout the bloodstream, but are formed and act locally, for example in the surroundings of inflammations, injuries or tumors. These compounds are not only involved in wound healing, but are also indirectly associated with tumor therapy, as anti-angiogenic compounds may prevent the vascularization of solid tumors.

Interleukins (IL) and interferons (IFN) are members of the cytokine family [38]. They are natural peptides produced by the cells of most animals' immune systems in response to challenges by foreign agents, e.g., viruses, bacteria, parasites, and tumor cells. These peptides and their recombinant analogues [39] or derivatives [40] are suited to bolstering immune responses for the treatment of neoplastic diseases, viral infection, and immunodeficiencies. In order to overcome the disadvantages associated with the therapeutic application of these proteins, engineering efforts have been (and are being) directed towards the design and expression of variants with low toxicity and suitable binding profiles. IL-2, which is an approved therapeutic for advanced metastatic cancer, is a representative example. Despite the great potential initially promised, IL-2 has found limited use due to its systemic toxicity. Proleukin (Chiron) is a mutant of IL-2 in which one (Cys¹²⁵) of the three Cys residues has been converted to Ser, without affecting the biological activity. However, this minimal alteration safeguards that a greater portion of the recombinant product is produced in the correctly folded form.

Recombinant *growth hormone* is used for the treatment of children suffering from dwarfism, a disorder brought about by deficient endogenous synthesis of this hormone. Interestingly, until 1985 growth hormone was obtainable only from human pituitaries removed at autopsy, as the growth hormone of other species is not active in humans. However, this early drug was heavily criticized because of suspected viral contamination. The recombinant hormone analogue Protropin (Genentech), which has an additional Met at the N-terminal end, was approved in 1985, and this was followed 2 years later by Humatrop (Eli Lilly), which had an identical sequence as the native hormone.

Epidermal growth factor (EGF) combined with poly(acrylic acid) gels has been shown to be successful in the treatment of corneal epithelial wounds [41].

Finally, it should be mentioned that the cost to develop a successful drug and to bring it onto the market is, on average, US\$ 600 million, this being associated with an average development time of about 10 years [42]. By contrast, improved engineered drugs have been successful both in medical and financial terms. For example, the humanized mAb Herceptin (see Section 9.3.3.3) generated US\$ 188 million during its first year of sales, and is undoubtedly one of the most successful anticancer drugs launched to date.

Despite an almost 80-year history of the use of proteins as therapy – starting with the commercial introduction of insulin in 1923, and followed by the approval of recombinant insulin as the first biotechnological drug in 1982 – interest has increased most significantly during the past two decades. The enormous advances in molecular biology (genetic engineering), cell biology and modern techniques in protein chemistry have promoted this rapid development. At present, most

efforts are still directed towards the discovery of new proteins with pharmaceutical potential, and the engineering of therapeutic proteins to provide the clinical benefits as discussed above. It is likely that drug developments in the near future will be characterized as a marriage of selection-based and knowledge-based approaches. Mutagenesis, selection, and high-throughput screening (HTS) techniques (cf. Section 9.4.6) will be guided both by structural knowledge, obtained by systematic determination of protein structures, and a better understanding of the biological and molecular mechanisms (molecular medicine).

9.3.3.2 Peptide-Based Vaccines

Peptide-based vaccines are peptides used in immunotherapeutic strategies for vaccination against, e.g., tumors (e.g., adenocarcinoma, glioma, melanoma, etc.), Alzheimer's disease, pathogenic microorganisms (e.g., *Pseudomonas aeruginosa*), and malaria [43, 44]. Peptide vaccines may be designed based on the subunit of a pathogen, either with naturally occurring immunogenic peptides or synthetic peptides corresponding to highly conserved regions required for the pathogen's function. The aim of this strategy is vaccination with a minimal structure that consists of a well-defined antigen and elicits effectively a specific immune response, without potentially hazardous risks.

In some countries, especially in south-east Asia and Africa, a relatively high percentage of the population has been infected with the highly infectious hepatitis B virus. This causes jaundice and, as a late consequence of chronic infection, even gives rise to liver tumors. A hepatitis B vaccine, isolated from the blood of virus carriers, has been available since 1982. Some years later, the first vaccine based on a recombinant protein containing the pure viral surface antigen was described.

A completely synthetic vaccine was first described about two decades ago [45, 46]. A CD8⁺ cytotoxic T-cell (CTL) epitope of influenza virus NP was conjugated to the general immune enhancer Pam₃Cys-Ser-Ser (cf. Section 6.5). This relatively simple construct resulted in efficient priming of virus-specific cytotoxic T cells when injected into mice, without any additional adjuvant.

During the past few years, a number of approaches have been identified to develop vaccines against viruses, harmful bacteria, and tumors for humans and cattle based on genetic vaccination, recombinant viruses, attenuated mycobacteria, or vaccination of protein subunits [47–49].

Synthetic peptides may even take into account the immunological diversity of cytotoxic T lymphocyte responses among patients in the frame of a personalized therapy. Tumors express many different antigens that distinguish them from normal healthy tissue. The microenvironment of the tumor tissue supports tolerance and limits T-cell immunity. Tumor vaccines aim at reversing tumor-induced immunosuppression by eliciting high-avidity T cells against subdominant tumor antigen epitopes. In the case of vaccines against Alzheimer's disease, circulating antibodies are directed towards the CNS and prevent β -amyloid formation or even dissolve the aggregates. Even peptides with post-translational modifications (glycosylation, lipidation) can be obtained synthetically. Modified peptides resist proteolytic cleavage and display improved metabolic stability *in vivo*.

Knowledge of the antigenicity of peptides has improved significantly during the past few years as a result of X-ray crystallographic analyses of complexes between peptides and monoclonal antibodies. However, this has not yet been achieved for immunogenicity. The development of a peptide-based vaccine requires the induced antibodies not only to recognize but also to neutralize the infectious agent. However, there are no chemical rules for designing peptide immunogens that elicit neutralizing antibodies.

Further progress will include the design of vaccines based on artificial proteins, e.g., multiantigen peptides, branched polypeptides, fusion and recombinant peptides, as well as T cell epitopes and tumor antigen peptides. For immunization purposes, peptides are required to exceed a certain molecular mass, and hence they are either conjugated to a protein, or single peptide antigens are incorporated into an antigenic peptide dendrimer, also called a multiple antigen peptide, MAP (M_r 3–100 kDa). This approach has been reported to increase the immunogenicity of weakly immunogenic monomeric peptides, presumably because of the multivalency and improved half-life *in vivo* (cf. Section 7.5.2).

An interesting approach to the therapeutic management of autoimmune diseases involves the design and application of peptide analogues of disease-associated epitopes to be used as immunomodulatory drugs [50].

9.3.3.3 Monoclonal Antibodies [51, 52]

Monoclonal antibodies (mAb) can be considered as a group of natural drugs as they mimic their natural function in an organism, but without inherent toxicity [53, 54]. The therapeutic application of mAb has become a major part of treatments in various diseases such as transplantation, oncology, cardiovascular, autoimmune, and infectious diseases. Antibody engineering technologies are advancing to enable further tuning of the effector function and serum half-life. More than 20 mAb are on the market and have received authorization to be applied for the treatment of the mentioned severe diseases, and over 150 are currently being evaluated in clinical trials. Antibodies exert their action either by (i) blocking cell–cell interactions, (ii) simulating cell membrane receptors, (iii) blocking lymphokine–cell interactions, or (iv) destroying their target cell by activating complement or mediating antibody-dependent cell-mediated cytotoxicity (ADCC).

The murine anti-human CD3 mAb (Orthoclone, OKT3) was the first monoclonal antibody marked for therapeutic purposes. It was launched in 1986 by Ortho Biotech for acute kidney transplant rejection, but first-dose reactions and anti-murine antibodies remain drawbacks in its clinical application. The application of OKT3 is associated with increasing susceptibility to infections and the cytokine release syndrome. The latter is characterized by shaking chills, fever, hypotension, diarrhea and vomiting, arthralgia, and even the development of respiratory distress.

Generally, the use of rodent mAb as therapeutic agents is hampered because the human organism recognizes them as foreign. An entirely antigenic nonhuman protein, e.g., a murine mAb, becomes human-friendly when small parts of the initial murine mAb are engrafted or inserted onto human IgG molecules,

creating either chimeric or humanized mAb. In chimeric mAb only the Fc part of the Ig molecule is human, whereas in humanized mAb only the complementarity-determining regions of the novel IgG molecule are murine and 90–95% of the molecule is human. Those near-human clinical mAb have been created by fusing murine variable domains to human constant domains in order to retain binding specificity while simultaneously reducing the portion of the mouse sequence [55]. The first example of an approved chimeric antibody was ReoPro (Abciximab) from Centocor, an anticoagulant, which was registered at the end of 1994 in the USA. Zenapax is a complementary determining region (CDR) grafted mAb targeted to the interleukin-2 (IL-2) receptor on T cells for application in preventing transplant rejection. The above discussed side effects of the xenogenic protein OKT3 led to the development of HuM291, a humanized OKT3, with only mild-to-moderate symptoms related to cytokine release [56]. Further examples of therapeutics are listed in Table 9.3. Trastuzumab is the first mAb approved by FAD for the treatment of solid tumors. Alemtuzumab kills malignant and normal hematopoietic cells which bear the cell-surface marker CD52. The chimeric anti-TNF- α mAb infliximab was first approved for the treatment of Crohn's disease.

All currently available mAb are produced in mammalian cell cultures, but this is an expensive process. Economic alternatives may be the development of transgenic animals (goats or cows) that have been genetically engineered to produce mAb in their milk [57]. Besides recent developments to reduce murine components, fully human antibodies will be the next-generation therapeutics [58]. Indeed, various techniques already exist for the development of 100% human antibodies, such as the direct isolation of human antibodies from phage display libraries [59] and transgenic mice containing human antibody genes and disrupted endogenous immunoglobulin loci [60]. A human anti-TNF- α mAb, designated D2E7 (BASF/CAT), was undergoing Phase III clinical trials for the treatment of rheumatoid arthritis [61]. Antibodies and

Table 9.3 Selected therapeutic monoclonal antibodies.

Name	Target antigen	Therapeutic use
Orthoclone (OKT3)	CD3	Renal transplants
Mabthera/Rituxan (Rituximab)	CD 20	Non-Hodgkin's lymphoma
Zenapex (Daclizumab)	CD 25	Renal transplants
Herceptin (Trastuzumab)	HER-2	Cancer
Simulect (Basilimab)	CD 25	Renal transplant
Remicade (Infliximab)	TNF- α	Crohn's disease; rheumatoid arthritis
Alemtuzumab (Campath-1H)	CD25	B-cell chronic lymphocytic leukemia
Simulect (Basilixmab)	CD 25	Organ transplants
Daclizumab	IL-2 receptor α -chain	Organ transplants; noninfec. Uveitis
Trastuzumab	ERBB2	Cancer
Cetuximab	EGFR	Cancer
Rituximab	CD20	B-cell lymphomas, autoimmunity
Efalizumab	CD11a	Psoriasis
Omalizumab	IgE	Asthma

antibody derivatives constitute about 25% of pharmaceutical proteins currently under development, and it is very clear that the immune system is an excellent target for new therapeutic efforts.

Fusion proteins (FP) are constructed by fusion of the Fc part of human IgG1 with a natural soluble receptor or ligand of a target molecule. Although artificial, fusion proteins are, however, completely human molecules retaining the function of the soluble receptor or ligand. The additional Fc part of IgG1 is responsible for a sufficient half-life which makes them clinically useful.

The use of proteomics and genomics combined with phage display allows at present the rapid selection of mAb directed against new targets. Current research is focussed on the selection of mAb with improved pharmaco-kinetics and bio-distribution, as well as on a better control of side-effects generated by some antibody treatments.

9.3.3.4 Future Perspectives

Drug research and pharmaceutical treatment stand at the dawn of an entirely new scientific era. In mid-2001 the human genome sequence had (mostly) been completed, indicating a total of 30 000 to 40 000 unique human genes [62, 63]. Moreover, according to the number of splice variants and functional variants resulting from post-translational modifications, the number of proteins will most likely exceed the number of genes. The *Human Genome Browser at UCSC*, a mature web tool [64] for the rapid display of any requested portion of the genome at any scale, together with several dozen aligned annotation tracks, is provided at <http://genome.ucsc.edu>.

The next major challenge is directed towards the human proteome. Proteomics represents a formidable task, and may result ultimately in the characterization of every protein encoded by the human genome. The proteome is defined as the entirety of proteins expressed in a cell, in a tissue, in a body fluid, or an organism at a certain time and under certain conditions. As different stages of development or pathological events are reflected in changes in the proteome, proteome analysis is usually carried out as a differential approach, by detecting changes in the protein expression profile [65, 66]. More information about proteomics and the role of peptides in proteomics is given in Chapter 10.

An understanding of the structure, function, molecular interactions, and regulation of every protein in various cell types is a future goal of the highest importance. However, due to the magnitude of this task, powerful tools in biochemistry, molecular biology, and bioinformatics – combined with massive automation – will be required to reach this goal. Indeed, knowledge gained of the molecular basis of many human diseases such as diabetes, cancer, arthritis, and Alzheimer's disease might eventually lead to the introduction of new therapeutic strategies.

Microarrays might represent the backbone of medical diagnostics during the 21st century. They consist of immobilized biomolecules spatially addressed on substrates, e.g., planar surfaces (typically coated microscope glass slides), microwells, or arrays on beads. In a typical protein microarray (for a recent review see [67])

proteins or peptides are arrayed on a solid support. After washing and blocking surface unreacted sites, the array is probed with a sample containing the counterparts of the molecular recognition events under investigation. In the case of interaction, a signal is revealed by a variety of detection techniques, either by direct detection, e.g., mass spectrometry, atomic force microscopy, surface plasmon resonance, quartz crystal microbalance, or by a labelled probe on the surface. The simplest variant for protein binding is performed via surface absorption, as has been demonstrated in standard ELISA and Western blot for a long time. This principle is based on adsorption of proteins and other macromolecules either by electrostatic forces on charged surfaces or by hydrophobic interactions. To eliminate several drawbacks, other attachment mechanisms for capture ligands such as physical entrapment, covalent binding and oriented biorecognition have been developed.

Three categories of protein arrays are discussed: (i) protein function arrays, (ii) protein detection arrays, also termed analytical arrays, and (iii) reverse phase arrays [68]. The *protein function arrays* comprise thousands of native proteins that are immobilized in a defined pattern so that each protein present in a cell at a certain time occupies a specific x/y -coordinate on the chip. Such devices are used for parallel screening of a variety of biochemical interactions such as the investigation of effects of substrates or inhibitors on the activity of enzymes, protein–drug or peptide hormone effector interactions, or the study of epitope mapping. These arrays will find useful application for studies of activities and binding profiles of native proteins, and will also be useful in addressing the specificity of small, protein-binding molecules, including drug candidates. A *protein detection microarray* consists of large numbers of arrayed protein-binding agents (antigens or antibodies). This chip allows the recognition of target proteins and polypeptides in cell extracts or other complex biological solutions. This approach seems to be useful for monitoring the levels and chemical states of native proteins, and can be considered as the proteomics version of DNA microarrays. Analytical arrays have been used to assay antibodies for diagnosis of allergy, autoimmunity diseases or for monitoring large scale protein expression. In the so-called *reverse phase microarrays* cell lysates, tissues or serum probes are spotted on the surface and probed with one antibody per analyte for a multiplex readout.

Without doubt, protein microarray technology is not so easy to perform as DNA technology due to the complex physical and chemical structure of proteins, including the enhancement of protein molecular variability by posttranslational modification. A limiting aspect in the protein microarray approach is the difficulty in maintaining the native state of the protein upon surface immobilization. An attempt to overcome this problem was published Ramachandran *et al.* [69] in 2004. This group has spotted protein expression plasmids instead of purified proteins on the microarray surface, thereby generating a nucleic acid programmable protein array. The latter reduces the process of building a protein array to a single step. Finally, it can be concluded that immobilization strategies and the design of an ideal local environment on the solid surface are both essential for the success of protein microarray technology [70, 71].

9.4

Peptide Pharmaceuticals

9.4.1

Large-Scale Peptide Synthesis

Procedures for the industrial production of peptide-derived active pharmaceutical ingredients (APIs) [2] differ significantly from the well-known laboratory-scale peptide synthesis methods. The term “large-scale” refers to batches ranging from kilograms to metric tons, according to the definition given in a review on this subject by Andersson *et al.* [72]. The chemistry does not differ markedly between large-scale manufacturing processes and those used under laboratory-scale conditions. However, the development of an economic, efficient and safe procedure which fulfils the requirements imposed by regulatory authorities generally comprises the final goal of large-scale peptide production [73]. Reaction conditions must be worked out in much more detail than for small-scale synthesis and the absolute minimum of starting materials must be carefully explored to reduce costs associated with raw material and waste after completing the reaction. Without doubt, the purification process is a very crucial step, or in other words, the bottleneck, in large-scale manufacturing processes [74, 75]. The overall purification strategy is determined on the basis of parameters such as size, polarity, solubility and, especially, the impurity profile of the peptide under investigation.

In the course of the development of the anti-HIV drug Fuzeon™ by Roche and Trimeris (see below and Chapter 5.3.4), suppliers had to produce starting materials and reagents at the metric ton scale with high requirements for purity.

The scale-up development strategy must take into account various technical, economic, and safety aspects. Extreme reaction conditions such as high pressure, or temperature, long reaction times, highly anhydrous conditions, and very specialized equipment must be avoided. Reaction temperatures generally range from -20°C to $+100^{\circ}\text{C}$.

The reactors used for large-scale solution phase synthesis (Figure 9.1) include steel reaction vessels, systems for heating and cooling, and a heterogeneous group of units for filtration, concentration under reduced pressure, and hydrogenation. Intermediates isolated during the course of synthesis should be obtained as solids rather than as oils, and the method of choice for this is either precipitation (crystallization if possible) or chromatography. Environmental and economic aspects also determine the selection of reagents and solvents used in industrial processes. For example, it is necessary to eliminate diethyl ether as a precipitation agent due to the high risk of explosion, and also to substitute the ozone-destroying dichloromethane with other solvents. Corrosive cleavage agents such as trifluoroacetic acid (TFA) and HF, or the toxic hydrazoic acid HN_3 which occurs as the by-product of azide couplings, are highly hazardous and must be avoided. The coupling agent BOP should also be substituted, as the by-product hexamethylphosphortriamide (HMPA), which is formed in coupling reactions, is known to be a carcinogen.

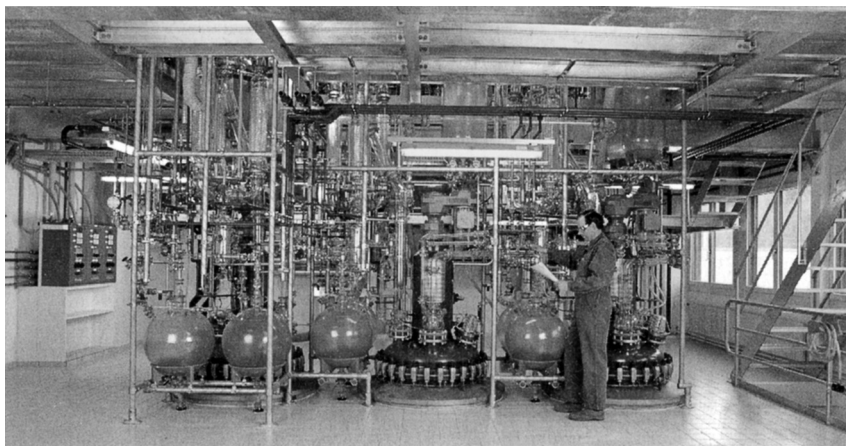


Figure 9.1 Modern production plant for solution phase synthesis (Photo: Bachem AG).

From the industrial point of view a good coupling agent must fulfil several requirements [76]. It must be a cost-effective reagent working at high efficiency, and be safe for producer, user, and environment. Furthermore, it must be produced in large quantities. The highly efficient (but very expensive) coupling additive HOAt has been normally substituted by the less expensive HOBt. At present, 6-chloro-1-hydroxybenzotriazole (Cl-HOBT) is the preferred additive since its active esters are more reactive than HOBt esters and, additionally, the chloro substituent stabilizes the structure, making Cl-HOBT a less hazardous reagent. In summary, HCTU/TCTU together with Cl-HOBT (see Section 4.3.7) are the preferred coupling agents in large scale API synthesis.

For economic reasons, it is detrimental to use more than two equivalents of activated amino acids. Reagents and reactants should be used in amounts close to stoichiometry. The decision as to whether the more expensive preactivated amino acids instead of nonactivated amino acids are used is usually made when the necessary development studies have been completed. On occasion, *in situ* activation protocols may be more time-consuming and accompanied by lower yields and higher amounts of impurities compared to protocols using the more expensive preactivated starting materials.

Although many protected amino acid derivatives are available commercially, minimum protection schemes are preferred for large-scale synthesis (see Section 5.2.2). In particular, the side-chain protection of arginine is minimized to the inexpensive HCl salts, as shown by the first industrial solution-phase synthesis of ACTH(1–24) [77] (Figure 9.2).

The large-scale solution-phase manufacture of [1-desamino,D-Arg⁸]vasopressin, DDAVP (Desmopressin), Mpa-Tyr-Phe-Gln-Asn-Cys-Pro-D-Arg-Gly-NH₂ (Mpa, 3-mercaptopropionic acid; disulfide bond: Mpa¹-Cys⁶), an antidiuretic used to treat diuresis associated with diabetes insipidus, nocturnal enuresis, and urinary incontinence [78], is performed via a [(3 + 4) + 2] segment coupling strategy. Interestingly, the N-terminal segment Mpa(Acm)-Tyr-Phe-NH-NH₂ is synthesized by chymotryp-

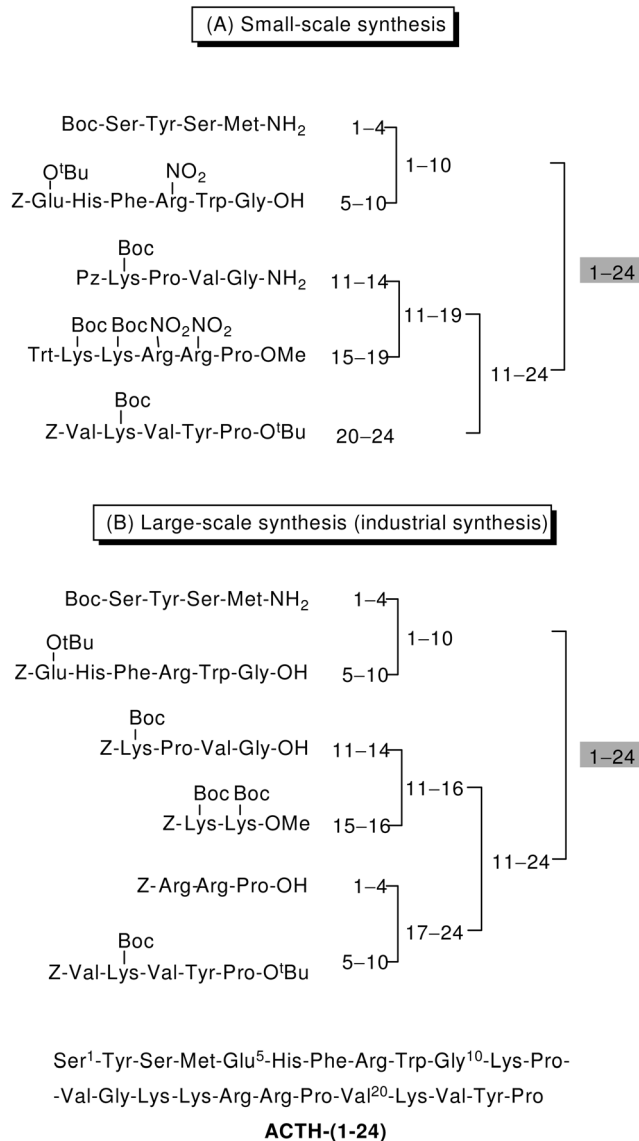


Figure 9.2 Strategy and tactics of the industrial synthesis of ACTH-(1-24) [77].

sin-catalyzed coupling of Mpa(Acm)-Tyr-OEt with H-Phe-NH-NH₂, underlining that enzyme-mediated coupling (cf. Section 4.6.2) is also highly efficient in industrial processes.

As shown in Section 4.5, solid-phase peptide synthesis has many advantages over the classical solution procedure, such as shorter production cycle times and often higher yields and purity. Thus this approach is also attractive for large-

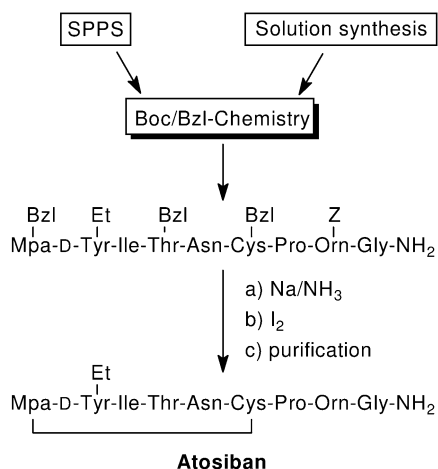


Figure 9.3 Schematic illustration of a drug development approach to a common intermediate resulting from solid-phase and solution-phase strategy demonstrated for the oxytocin antagonist Atosiban [80].

scale manufacture of selected peptides and, especially, for peptide fragments (cf. Section 5.3) [79]. Through the refinement of the SPPS it has been possible to produce relatively complex peptide APIs economically, on a scale up to hundreds of kilograms or even metric tons. It has been reported that approximately half of peptide-based APIs are manufactured using SPPS techniques. For mid-scale SPPS special equipment has been developed which differs significantly from the commercially available lab-scale synthesizer. For example, starting from 2 kg resin in a commercially available 60 L Labortech solid-phase reactor unit the procedure results in about 9 kg peptidyl resin which corresponds to about 1 to 1.5 kg peptide. Such equipment fulfils the standard of *current Good Manufacturing Practice* (cGMP) [2] according to the Federal Regulations of the Food and Drug Administration in the same way as the plant for solution-phase peptide production shown in Figure 9.1. Furthermore, reactors with much higher capacity have been developed, culminating in the 10 000 L reactor developed by Roche Colorado which is shown below in Figure 9.4.

A mid-scale industrial synthesis (Figure 9.3) with an estimated future annual production scale in the range of 50–100 kg has been described for an oxytocin antagonist, named Atosiban [80], which is used to treat preterm labor and delivery [81]. The synthesis strategy for Atosiban, Mpa-D-Tyr(Et)-Ile-Thr-Asn-Cys-Pro-Orn-Gly-NH₂ (disulfide bond: Mpa¹-Cys⁶) is based on a common intermediate for solution-phase and solid-phase syntheses. First, the required quantities of Atosiban for toxicology and early phase clinical studies during drug development were synthesized using the rapid solid-phase method (see Section 4.5). An increasing demand for the peptide in clinical Phase II trials, where defined doses for studies in humans and the determination of a safety profile are required according to the regulations, led to a change in the synthesis protocol and the introduction of a solution-phase scale-up (2 + 5) + 2 strategy. Thus, it was desirable to direct both manufacturing methods



Figure 9.4 The World's largest solid phase peptide synthesizer with a volume of 10 000 L (Photo: Roche Colorado).

to a common intermediate with an identical side-chain protecting group pattern (Figure 9.3).

Under these conditions, the following steps such as deprotection, oxidation, purification, and final isolation are associated with a similar profile of impurities. This combined strategy, leading to a common intermediate, is of general importance in the industrial-scale production of peptides.

The SPS/SPPS-hybrid approach (see Section 5.3.4) has been used in several industrial processes, with synthesis of the 36-peptide enfuvirtide (T20; Fuzeon™) being the most exciting example at present [82]. Enfuvirtide (sequence see Figure 5.11) is derived from the ectodomain of HIV-1 gp41, and is the first representative of a novel family of anti-retroviral agents that inhibit membrane fusion. The major importance of enfuvirtide in the development of a drug to treat HIV initiated, in the first instance, a solid-phase manufacturing process based on Fmoc chemistry [83]. However, due to a need for production in the region of metric tons, the strategy was changed at an early stage to a phase change process involving three segments, synthesized on 2-chlorotriyl resin (Section 4.5.1). The World's largest solid-phase reactor (10 000 L), shown in Figure 9.4, is used at Roche Colorado for the synthesis of the segments. Initially, Fmoc-enfuvirtide-(27–35)-OH is coupled after cleavage from the resin to H-Phe-NH₂ (enfuvirtide synthesis scheme, Figure 5.12). The N-terminal Fmoc group is cleaved, resulting in H-enfuvirtide-(27–36)-NH₂ which is elongated with Fmoc-enfuvirtide-(17–26)-OH, yielding Fmoc-enfuvirtide-(17–36)-NH₂. After removal of the N-terminal

protecting group, this fragment is coupled to Ac-enfuvirtide-(1–16)-OH, providing the fully protected Ac-enfuvirtide-(1–36)-NH₂. Deprotection with TFA/dithiothreitol/H₂O gives the crude 36-peptide derivative with a relatively high HPLC purity (>70%). This is further purified by preparative reversed-phase HPLC, and subsequently lyophilized. Enfuvirtide can be synthesized economically on a metric ton scale and has also provided a tremendous boost for large-scale peptide production [84]. The annual production capacity for this peptide drug at Roche amounts to 6000 kg per year, which is 60 to 300 times the annual production of other synthetic peptides such as calcitonin or leuprolide.

A method, termed Diosynth Rapid Solution Synthesis of Peptides (DioRaSSP) has been developed for large-scale manufacturing of peptides in solution [85]. This procedure combines the advantages of the homogenous character of classical SPS with the generic character and the amenability of automation inherent to SPPS. This approach is characterized by repetitive cycles of coupling by water-soluble carbodiimides and deprotection in a permanent organic phase. Intermediates do not need to be isolated, the process is easy to scale up yielding products of reproducible high purity. Furthermore, the first fully automated solution-phase peptide synthesizer for application in the DioRaSSP process has been developed. Leuprolide, buserelin, deslorelin, goserelin, histrelin, and triptorelin are examples of large-scale synthesis according to the DioRaSSP approach. It has been reported that only the size of the available reaction vessels should prove to be the limiting factor during scale-up towards multi-100 kg batches.

9.4.2

Peptide Drugs and Drug Candidates [86–90]

As shown in Table 9.4, peptides have now found widespread use as active pharmaceutical ingredients (APIs), and more than 40 peptides are on the market worldwide. Most peptides address specifically one receptor or one family of receptors, exerting a well-defined spectrum of biological answers upon binding. However, many peptides are usually regarded not to be useful as drugs because they lack metabolic stability *in vivo* and are not orally available. Moreover, many body barriers cannot be crossed by peptidic compounds. Often, peptides are more expensive to produce and hence need to be more potent than other alternatives. However, the past decade has witnessed a renaissance of peptides to be applied as drug molecules. This coincides with advancements in chemical modification of peptides, administration, and formulation, as discussed above. Several of the top best-selling drugs approved by the FDA are relatively unmodified peptides, and many of these stem from natural sequences. *Gonadoliberein* (GnRH) (cf. Section 3.3.3.2) agonists and antagonists are used for the treatment of prostate cancer. For example, *leuprolide* acetate (LupronTM), [D-Trp⁶]GnRH and [D-Leu⁶, desGly¹⁰NH₂]GnRH-NHEt, belongs to the so-called blockbuster drugs with annual worldwide sales exceeding US \$2 billion. Leuprolide shows relative activities of 3600 and 5000%, respectively, compared to the native hormone, and is indicated for treating advanced prostate cancer due to its capability of decreasing testosterone levels. Leuprolide has also been used, under court order,

Table 9.4 Selected approved peptide drugs and their manufacturing methods.

Peptide	aa	Synth./ Strategy ^a	Quantity ^b kg p.a.
Abarelix (Plenaxis TM ; GnRH antagonist)	10		
ACTH-(1–24) (Synacthen TM)	24	SPS	50–100
Atosiban (Tractocile TM , Antocin TM ; Oxytocin analogue)	9	SPS	50–100
Bacitracin (mixture of related cyclic peptide antibiotics from <i>Bacillus subtilis</i> (Tracy)	10	F	
Bleomycin (Blenoxane TM , cytostatic glycopeptide antibiotic from <i>S. verticillus</i>)		F	
Buserelin (Profact TM , Suprefact TM ; GnRH agonist)	9	SPPS	
Calcitonin (human) [Cibacalcin TM]	32	SPS	
Calcitonin (salmon) [Miacalcin TM]	32	SPS, SPPS	
Calcitonin (eel) [Thyrocalcitonin Eel TM]	32	SPS, SPPS	
Caspofungin (Cancidas TM ; antifungal lipopeptide)	6	F, SS	
Cetorelix (Cetrotide TM ; GnRH antagonist)	10	SPS	10–100
Cholecystokinin (CCK-33)	33	SPS	
Cyclosporin (Sandimmune TM , Neoral TM)	11	F, E	
Daptomycin (Cubicin TM ; cyclic lipopeptide)	13	F, SS	
Darbepoetin α (Aranesp TM ; erythropoietin analogue)	165	R	
Deslorelin (Suprelorin TM ; GnRH agonist)	9	SPPS	
Desmopressin (Minirin TM ; Vasopressin analogue)	9	SPS, SPPS	50–100
Elcatonin (Dicarba-eel-calcitonin)	31	SPS, SPPS	
Eledoisin	11	SPS	
Enfuvirtide (Fuzeon TM ; T20)	36	SPS/SPPS-Hybrid	~6000
Eptifibatide (Integrilin TM ; cyclic RGD peptide)	7	SPS	>200
Exenatide (Byetta TM ; synth. Exendin-4)	39	SPPS	
Glucagon	29	SPS, R, E	
GnRH (LH-RH)	10	SPS, SPPS	150–200
Goserelin (Zoladex TM ; GnRH analogue)	10	SPPS	
Icatibant (Firazyr TM ; Bradykinin antagonist)	10	SPPS	
Insulin and insulin analogs	51	SPS, SS, R, E	
Lanreotide (Somatuline TM ; SST analogue)	8	SPPS	100–200
Leuprolide (Lupron TM , Viadur TM , Eligard TM ; GnRH agonist)	9	SPS, SPPS	25–50
Lypressin (Diapid TM , Vasopressin analogue)	9	SPS	50–100
Nesiritide (Natrecor TM ; hBNP)	32	R	
Octreotide (Sandostatin TM , SST analogue)	8	SPS	100–200
Pitressin TM (Vasopressin analogue)	9	SPS	50–100
Polymyxin B and E (peptide antibiotics)	10	F	

Table 9.4 (Continued)

Peptide	aa	Synth./ Strategy ^a	Quantity ^b kg p.a.
Pramlintide (Symlin™; hAmylin analogue)	37	SPS, SPPS	>10
Preotact [Preos™; PTH-(1–84)]	84	R	
Secretin (human, porcine)	27	SPS, E	
Sincalide (Kinevac™; CCK-8)	8	SPPS	
Somatostatin (SST)	14	SPS, SPPS	
Teriparatide [Forteo™; PTH-(1–34)]	34	SPS, R	
Terlipressin (Glypressin™; Vasopressin analogue)	12	SPS, SPPS	50–100
Thymopentin	5	SPS	
Thymosin α ₁ (Zadaxin™; Thymalfasin)	28	SPPS	200–400
Thyroliberin	3	SPS	
Triptorelin (Decapeptyl™, Trelstar™; GnRH agonist)	10	SPPS	
Vancomycin (Vancocin™, glycopeptide antibiotic)	7	F	
Ziconotide (Prialt™, ω-conotoxin M-VII-A)	25	SPPS	1–5

^aSPS (Solution-Phase Synthesis); SPPS (Solid-Phase Peptide Synthesis); SS (Semisynthesis); R (Recombinant Synthesis); E (Extraction); F (Fermentation).

^bData with the exception of enfuvirtide are taken from Bruckdorfer *et al.* [88].

to cause male chemical castration. *Goserelin* acetate (Zoladex™), [D-Ser(tBu)⁶, AzaGly¹⁰]GnRH (AzaGly: azaglycine, hydrazinecarboxylic acid), is also a superagonist and has been marketed by AstraZeneca. Goserelin is an effective hormonal treatment for prostate cancer as it reduces testosterone production, thereby removing the growth stimulus for cancer cells within the prostate [91]. Instead of a therapy by receptor down-regulation, which is accompanied by strong pain in the initial phase, antagonists can also be used as this application does not show such effect. The GnRH antagonist *abarelix*, manufactured by Praecis Pharmaceuticals in the US and sold under the brand name Plenaxis™ reduces the amount of testosterone in patients with advanced symptomatic prostate cancer for which no other treatment options are available. It does not cure prostate cancer but can relieve symptoms. The long-acting nonapeptide superagonist *buserelin*, [D-Ser(tBu)⁶]GnRH-(1–9)-NH₂, has found clinical application for disorders such as estrogen-dependent tumors (carcinoma of the prostate and the breast) or endometriosis, and is undergoing evaluation as a contraceptive agent. Antagonists are also being tested as male and female contraceptive agents. Antagonists of the “third generation” are in clinical trials for induced hormone suppression, e.g., in sex steroid-dependent benign and malignant diseases, and for premature LH surges in assisted reproduction. Members include *cetrorelix*, Ac-[D-Nal¹, D-p-Cl-Phe², D-Pal³, D-Cit⁶, D-Ala¹⁰]GnRH (SB-75) and *detirelix*, Ac-[D-Nal¹, D-p-Cl-Phe², D-Trp³, D-Harg(Et)₂⁶, D-Ala¹⁰]GnRH (Cit: citrulline; p-Cl-Phe: 4-chlorophenylalanine; Harg(Et)₂: N,N'-diethylhomoarginine; Nal: 3-(2'-naphthyl)-alanine; Pal: 3-(3'-pyridyl)-alanine).

Icatibant (Firazyr) is a potent and highly specific competitive *bradykinin* B2 receptor antagonist with the sequence H-D-Arg-Arg-Pro-Hyp-Gly-Thi-Ser-D-Tic-Oic-Arg-OH (Thi: 3-(2-thienyl)alanine; Tic: 3,4-dihydro-1*H*-isoquinoline-3-carboxylate; Oic: octahydroindole-2-carboxylate). It has been approved for application against hereditary angioedema, and is under investigation for a number of other conditions in which bradykinin is assumed to play a significant role.

Another class of peptide drugs that is related to peptide hormones has been used traditionally for the treatment of several diseases. In particular, *insulin* (see Section 3.3.1.3) should be mentioned in this context. While insulin is used in diabetic patients to lower the blood glucose level, its antagonist *glucagon* (see Section 3.3.1.2) increases blood glucose concentration. Glucagon is used in the treatment of hypoglycemia; for this it can be applied parenterally, by using a portable pump, nasally or as eye drops. Today, insulin is produced by recombinant technology (for more information cf. Section 4.6.1 and Figures 4.48 and 4.49). In addition to native human insulin, both fast-acting and slow-acting insulin derivatives have been obtained by amino acid replacements. Insulin is one of the oldest biopharmaceuticals approved, and currently more than 2000 kg are marketed each year. Recombinant human insulin was first launched by Eli Lilly in 1982, and over the past few decades insulin analogues have been designed with the aim of improving therapy [92]. In order to improve on the pharmacokinetics of insulin, an engineered form of human insulin, termed insulin Lispro (HumalogTM, LiprologTM) was produced by Eli Lilly. In this variant, only the partial sequence -Pro²⁸-Lys²⁹- has been reversed [93]. Because of this manipulation, the insulin exist as a monomer at physiological concentrations and, consequently, has a faster onset, but shorter duration of action due to enhanced absorption after subcutaneous administration. Humalog is a block buster drug. For example, in 2004 Humalog accounted for \$ 1.1 billion of Lilly's worldwide revenues from diabetes care of \$ 2.6 billion. Insulin glargine (Lantus, formerly known as HOE901), 21^A-Gly-^{30B}a-L-Arg-^{30B}b-L-Arg-human insulin, is a long-acting recombinant human insulin analogue produced by DNA technology [94]. The substitution of Asn²¹ of the A chain by Gly, and the N-terminal extension of the B chain by two Arg residues, resulted in a change in the isoelectric point from 5.4 of the native insulin to 6.7 of insulin glargine. As a result, it is soluble in slightly acidic conditions (pH 4.0) and precipitates at the neutral pH of subcutaneous tissue. In this way, the absorption of insulin glargine is delayed, thereby providing a fairly constant basal insulin supply for about 24 h.

Exenatide (ByettaTM) is a synthetic 39-peptide with the same sequence as exendin-4, a peptide from the saliva of the lizards *Heloderma suspectum* and *H. horridum*. It mimics the function of glucagon-like peptide-1 (GLP-1), and strongly activates the pathway to improve glycemic control in patients with type-2 diabetes [95]. ByettaTM is supplied as a sterile solution for subcutaneous injection.

Pramlintide (SymlinTM) is a synthetic analogue of human amylin (cf. Section 3.3.5.3), and is used by injection for antihyperglycemic therapy of diabetic patients treated with insulin. The application of SmylinTM contributes to glucose control after meals.

Teduglutide (ALX-0600) is a dipeptidyl peptidase IV-resistant GLP-2 analogue improving intestinal function in short bowel syndrome patients [96].

Oxytocin (cf. Section 3.3.4.1) and its analogues are applied either intranasally or by injection, and cause uterine contractions. They are used to induce labor, control bleeding after childbirth, and support milk secretion during breastfeeding. The oxytocin analogue atosiban (Tractocile™, Antocin™) (cf. Section 9.2; Figure 9.4) acts as an oxytocin receptor antagonist and is used clinically to suppress premature labor between weeks 24 and 33 of gestation.

Vasopressin (cf. Section 3.3.4.2) and its analogues, e.g., desmopressin (Minirin™) (cf. Section 9.2), administered by injection, support the kidneys in reabsorbing water in the body. They also raise the blood pressure by constricting the blood vessels.

Secretin (cf. Section 3.3.1.2) with the brand name Human Secretin™ is used, by intravenous injection, to stimulate pancreatic and gastric secretion.

Calcitonin (CT) from different origin (cf. Section 3.3.5.3) is administered nasally or by injection to treat osteoporosis and high blood calcium levels. Elcatonin, [Asu¹⁻⁷]eel calcitonin and second-generation analogues of CT with reduced side effects and new dosage forms (nasal, and potentially oral) will enhance the usefulness of calcitonin therapy.

Parathyroid hormone (cf. Section 3.3.5.1) regulates the metabolism of calcium and phosphate in the body [97]. At present, beside the recombinant native hormone [rhPTH-(1-84); Preos™] [98], the fragment [rhPTH-(1-34); teriparatide, Forteo™] [99] is available, while the analogue [Leu²⁷]cyclo(Glu²²-Lys²⁶)hPTH-(1-31)-NH₂ (Ostabolin-C) [100] awaits approval. The PTHs treat osteoporosis by strongly stimulating bone formation and strengthening bone microarchitecture in humans, rodents, and monkeys, with few or no side effects. Teriparatide is the first anabolic drug stimulating new bone formation approved by the FDA. Furthermore, studies have been started using these PTHs in cancer patients as a novel tool to treat bone marrow depletion caused by chemotherapeutic drugs and ionizing radiation.

Erythropoietin (EPO) stimulates the production of red blood cells, and is used in the treatment of anemia caused by kidney disease. Darbepoetin alfa (Aranesp™) is produced by recombinant DNA technology in modified Chinese hamster ovary cells (CHO cells) and differs from the endogenous EPO (165 aa) by containing two additional N-linked oligosaccharide moieties. In 2001 it was approved by the FDA for treatment of anemia in patients with chronic renal failure by intravenous or subcutaneous injections. Like EPO its application increases the risk of cardiovascular problems.

Hematide is a novel synthetic PEGylated erythropoietin mimicking peptide that acts as an erythropoiesis-stimulating agent (ESA) with a prolonged half-life and slow clearance times. It was designed to bind and activate the EPO receptor in order to stimulate erythropoiesis and to treat anemia associated with chronic kidney disease. The amino acid sequence of Hematide is unrelated to EPO and, for this reason, it is not likely to induce a cross-reactive immune response against either endogenous or recombinant EPO. The sequence of Hematide was originally derived from phage display [101].

Tissue factor pathway inhibitor (TFPI) exerts important role(s) as a natural anticoagulant. Novel peptides that mimic fragments of TFPI are in development with

the aim to stop tumor growth by tapping into TFPI's innate capability to inhibit blood vessel growth.

The human *brain natriuretic peptide* (hBNP) (cf. Section 3.3.6.3) has been approved as a vasodilatory cardiovascular drug for intravenous administration. Nesiritide (Natreacor™) is the recombinant form of hBNP, which is normally produced by the ventricular myocardium. Nesiritide is a drug used to treat acutely decompensated congestive heart failure with dyspnea at rest or minimal exertion [102]. It promotes vasodilation, natriuresis, and diuresis. Furthermore, BNP may be a useful addition for disease monitoring in heart failure patients.

Somatostatin (SST) (cf. Section 3.3.1.4) analogues have been synthesized and tested to identify some with higher selectivity and longer half-life. For example, *octreotide* (Sandostatin™), H-D-Phe-c-(Cys-Phe-D-Trp-Lys-Thr-Cys)-Thr-ol (disulfide bond: Cys²-Cys⁷), is 70-fold more potent than SST in inhibiting somatotropin release *in vivo* 15 min after administration, and it is characterized by a long duration of action after intramuscular administration. This analogue is used in the treatment of somatotropin- and thyrotropin-secreting pituitary tumors, carcinoid tumors, and in further indications. More recently, two additional analogues, namely *lanreotide*, H-D-βNal-c-(Cys-Tyr-D-Trp-Lys-Val-Cys)-Thr-NH₂ (disulfide bond: Cys²-Cys⁷), and *vapreotide*, H-D-Phe-c-(Cys-Tyr-D-Trp-Lys-Val-Cys)-Trp-NH₂ (disulfide bond: Cys²-Cys⁷) have become available for clinical use. Radiolabelled somatostatin analogues such as ⁹⁰Y-DOTA-Tyr³-octreotide (⁹⁰Y-DOTATOC) with the β-emitter ⁹⁰Y have been developed for radiotherapy [103].

Ziconotide, CKGKGAKCSR¹⁰LMYDCCTGSC²⁰RSGKCa (disulfide bonds: C¹/C¹⁶, C⁸/C²⁰, C¹⁵/C²⁵), is a novel non-opioid, non-local anesthetic, developed for the treatment of severe chronic pain. It was also previously referred to as Prialt, CI 1009, or SNX-111, and is the synthetic form of the cone snail peptide ω-conotoxin M-VII-A, a neuron-specific N-type calcium channel blocker with an analgesic activity about 800-fold stronger than that of morphine [104]. Ziconotide is currently licensed for continuous intrathecal infusion (into the spinal canal) in the treatment of chronic intractable pain, and its analgesic efficacy has been demonstrated in both animal and human studies. Ziconotide-induced analgesia is not associated with the development of tolerance, respiratory depression or endocrine side effects, as is common in opioids [105].

Cyclosporin A (CsA) (cf. Section 3.3.8.1) is an immunosuppressant indicated for the prophylaxis of organ rejection in kidney, liver, and heart allogeneic transplants. In principle, the clinical use of CsA is limited because of poor water solubility, associated with very important adverse side effects. Since oral or parenteral formulation forms result in CsA being distributed widely throughout the body, the development of alternative dosage forms which deliver the drug specifically to the target site is needed. Water-soluble prodrugs of CsA with tailored conversion rates have been developed [106].

Antithrombotic therapeutics like *Abciximab*, a human-murine chimeric Fab fragment of a monoclonal antibody against the GP IIb/IIIa receptor, have demonstrated their clinical effectiveness. However, due to some disadvantages alternative GP IIb/IIIa receptor inhibitors have been developed. *Eptifibatide* is

a small cyclic 7-peptide containing an RGD-sequence mimicking the receptor blocker barbourin, found in the venom of the southeastern pigmy rattlesnake, and is applied as a therapeutic for coronary thrombosis. Further integrin-specific peptidomimetic antagonists are, e.g., *tirofiban* , which is used as an adjunct to angioplasty; *WO9736858 A1* , which is potentially useful for the treatment of tumor metastasis, solid tumor growth, osteoporosis, angiogenesis, humoral hypercalcemia of malignancy, restenosis, and smooth muscle cell migration; and last – but not least – *BIO-1211* , a small-molecule, tight-binding inhibitor of the integrin $\alpha_4\beta_1$ which is an adhesion receptor that plays an important role in allergic inflammation and contributes to antigen-induced late responses (LAR) and airway hyperresponsiveness (AHR). *BIO-1211* is in pre-clinical studies for asthma and inflammatory bowel disease.

The vitamin K-dependent human *protein C* [107] concentrate is employed for therapy of patients with life-threatening blood clotting complications.

Thymalfasin (thymosin α_1 , $T\alpha_1$, *Zadaxin* TM) is a synthetic 28-peptide with multiple biochemical activities primarily directed towards immune response enhancement. $T\alpha_1$ was originally isolated from thymosin fraction 5, a bovine thymus extract. Chronic hepatitis B infection is a serious disease because of its worldwide distribution. This peptide was effective in treatment of chronic hepatitis B, both as monotherapy and combined with interferon- α (INF- α). Further clinical trials are necessary since few side effects have been observed. $T\alpha_1$ should have the potential for the enhancement of the activity of antivirals such as INF- α , lamivudine, and ribavirin as viral hepatitis therapy.

Daptomycin (*Cubicin* TM, **3** in Section 6.1, cf. Section 6.5) is a branched cyclic 13-peptide linked by an ester bond between the terminal kynurenine and the hydroxy group of Thr bearing a lipophilic tripeptide tail. This lipopeptide antibiotic, originally discovered at Elli Lilly, is now licensed to Cubist Pharmaceuticals and used in the treatment of certain infections caused by Gram-positive organisms, and was approved in the US in 2003 for the treatment of skin infections. From 2003 to 2004 revenues of Cubist Pharmaceuticals increased from \$ 3.7 million to \$ 68.1 million, while \$ 58.6 million of the whole amount was generated by *Cubicin* TM alone.

Glycopeptide antibiotics inhibiting peptidoglycan biosynthesis, with *vancomycin* as one representative, are indicated for serious infections where other antibiotics are not effective. Vancomycin was discovered by Eli Lilly as early as 1956. It consists of seven amino acids containing in total five aromatic rings. The sugar components are L-vancosamine and D-glucose. Despite recent incidences of bacterial resistance to vancomycin, it became almost legendary because of its performance against methicillin-resistant *S. aureus* (MRSA).

Telavancin (TD-6424) is a second-generation semisynthetic lipoglycopeptide antibacterial agent based on the vancomycin scaffold. It exhibits potent antibacterial action *in vitro* against a broad array of important Gram-positive pathogens [108]. As observed with vancomycin, telavancin inhibits late-stage peptidoglycan biosynthesis in a substrate-dependent fashion, and also perturbs bacterial cell membrane potential and permeability.

The glycopeptide antibiotic *bleomycin* , produced by *S. verticillus* , inhibits DNA synthesis and is used in cancer chemotherapy, including testicular cancer, non

Hodgkin's lymphoma, and Hodgkin's lymphoma. Generally, *anticancer peptides* [109] display antitumor activity based on different modes of action. They may be derived from sites of protein interaction, phosphorylation, or cleavage, and, e.g., interfere with apoptotic pathways. Peptide-based approaches are reported to target, e.g., MDM2, p53, NF- κ B, ErbB2, MAPK, Smac/DIABLO, IAP BIR domains, and Bcl-2 interaction domains. In addition, therapeutic cancer targeting peptides have been developed and have shown clinical promise because they can be conjugated with cytotoxic agents and hence be delivered to tumor tissues. Monoclonal antibodies or peptides recognizing cell-surface receptors that are up-regulated on tumor cells can be used as homing devices for tumor-targeting strategies.

Active specific immunotherapy (ASI) is an approach to induce cellular immunity in the tumor-bearing host and is more promising than passive immunotherapy techniques. ASI is a promising approach to treating cancer. Cells taken from the host are reintroduced to the host after use of *ex vivo* techniques, e.g. irradiation, hapten conjugation, neuraminidase treatment. Genetic modulation of the tumor cells to produce immunostimulatory molecules can also be performed. For example, clinical trials with granulocyte-macrophage colony-stimulating factor (GM-CSF)-modified tumor cells have produced encouraging results. Furthermore, it could be demonstrated that various cancer vaccines can stimulate antibody and cell-mediated immune responses against tumor-associated antigens. For example, sialyl-T_n (ST_n) is an ideal candidate for ASI therapy. Theratope vaccine is a cancer vaccine that was designed by Biomira, Inc. (Edmonton, Alberta, Canada). It is composed of a synthetic 43-peptide glycosylated with sialyl-T_n antigen that emulates the carbohydrate seen on human tumors. The glycopeptide is conjugated to keyhole limpet hemocyanin (KLH) to elicit the immune response. Theratope vaccine is being well-tolerated with minimal toxicity.

The RGD peptide cyclo(-Arg-Gly-Asp-D-Phe-N^{Me}Val-) (*cilengitide*) is a highly selective ligand for α_v integrins, which are important in angiogenesis. Hence, it has been studied for treating cancer by inhibiting angiogenesis [110, 111], and has reached clinical Phase III trials for the treatment of glioblastoma (brain tumors).

There is no doubt, that at present an exciting time for peptide drug development has been initiated, as demonstrated by the large-scale production of enfuvirtide (T20, FuzeonTM) which represents a landmark in industrial peptide synthesis [82] (for more information cf. Sections 5.3.4 and 9.4.1). Enfuvirtide is a 36-peptide therapeutic derived from a protein subdomain. Especially, the natural sequence is taken from the ectodomain gp41 moiety of the HIV-1 precursor protein gp 160. Enfuvirtide inhibits viral fusion by interaction with the transient conformational forms of both the gp41 and gp120 target proteins, which themselves are derived from proteolytic cleavage of the gp 160 precursor [112]. First, it was assumed that enfuvirtide was only a competitive inhibitor of the complex formation [113]. However, further studies show that enfuvirtide resistance mutations from patients map to both the gp41 and the interacting gp 120 proteins. This led to the suggestion that enfuvirtide may bind to multiple sites in both proteins, potentially acting by an allosteric mechanism [112].

Before enfuvirtide received approval in the US by the FDA in March 2003, peptides comprised only about 0.0025% (by mass) of the worldwide annual production of

drugs. Peptide drugs such as the immune suppressor cyclosporin, which nowadays is indispensable in modern organ transplantation, are highly beneficial in therapy, and the current sales of cyclosporin exceed US \$ 1 billion each year. The sales values of calcitonin, which has become an important drug in the treatment of hypercalcemia, Paget's disease, osteoporosis and pain (preferentially of patients suffering from bone cancer) are in the same sales range.

Without doubt, the therapeutic application of peptides has great potential in various indications such as blood pressure, neurotransmission, growth, digestion, reproduction, and metabolic regulation. The control of almost all biological processes in living cells is exerted by proteins, and involves various types of molecular recognition. Much of this activity is mediated by enzymes, but many regulatory processes are initiated by specific protein–protein interactions that still constitute a largely unexploited area of targets in drug discovery and drug design. Unfortunately, the interacting surfaces very often lack the classical features necessary for inhibition with small molecules [114]. The development of peptide and nonpeptide integrin antagonists [115], the design of a platelet-derived growth factor (PDGF), a binding molecule with anti-angiogenic and tumor regression properties [116], and the “death receptor antagonist” Bcl-2 [117] have proven that the inhibition of protein–protein interactions is a viable therapeutic strategy. The current focal point of research in this field is directed towards the development of high-affinity protein–protein interaction antagonists and agonists that mimic the binding interface at selected interaction hot spots, based on the “hot spot” concept [118].

Because of their enormous potential for diversity, it is possible that peptides may be uniquely suited for influencing biological control processes based on molecular recognition. As shown in Chapter 8, it is now possible to construct very large libraries of peptides. When discussing peptides as potential pharmaceutical agents, although the final goal is efficacy *in vivo*, the ultimate need for high potency towards the target protein must be linked with few side effects and good (preferentially oral) bioavailability. Unfortunately, one major disadvantage of peptide pharmaceuticals is their putative metabolic instability.

Nowadays nearly 300 new peptide-based drugs are at different stages of development [119] and about 400 are in the pipeline (<http://www.biopharma.com>). Peptide drugs represent 1% of total API with an annual market of US \$ 300–500 millions and an annual growth rate of 15–25%. More than 10 known bulk peptide producers and over 20 companies are offering custom peptide synthesis. These numbers and capacities are growing [120] and the same is true for specialized biotech companies. A huge step for further synthetic and artificial sequences to biological application results from the sequencing of the human genome that will provide potential new drugs for current medical and pharmaceutical needs.

9.4.3

Peptides as Tools in Drug Discovery

Peptide research on drug discovery and design is an important field in the development of peptide mimetics (see Chapter 7), with the potential to generate

important new drugs. Peptides control numerous body processes and, as such, represent an untapped wellspring of new drugs for treating a variety of diseases. Therefore, the current challenge is to produce small molecules which mimic peptides and proteins, in order to overcome the ineffectiveness of peptides as drugs when administered orally.

The use of peptides for affinity labelling of receptors is important in attempts to identify, characterize, and isolate hormone or neurotransmitter receptors. The general approach is to establish a covalent bond between a ligand and its receptor; this can be achieved by chemical affinity labelling and photoaffinity labelling, with the latter technique being the preferred method for receptor identification and isolation. For this purpose, a chemically stable but photolabile moiety is conjugated to a potent ligand. When the modified ligand has bound to its receptor site, photolysis generates highly reactive nitrenes or carbenes that react with chemical functionalities on the receptor molecule, thereby forming a covalent bond between the ligand and the receptor [121].

Synthetic peptides are also used for the delineation of receptor types and subtypes. Receptors for almost all bioactive peptides are expressed by different target cells linking the hormone signal to slightly varying biological effects. Multiple types and subtypes of receptors exist, which complicates receptor pharmacology, notably as each subtype plays a particular functional role *in vivo*. Consequently, the design and synthesis of peptides directed toward receptor subtype binding and the determination of the appropriate kinetics are essential aims of current peptide research.

Target-based screening to identify compounds for development is a prerequisite to a powerful methodology in drug discovery research. Conventionally, drugs have been discovered by screening either natural compound collections or small chemical compound libraries. An alternate approach would be the chemical synthesis of compounds based on structural data available for a given target. Unfortunately, all these methods are generally cumbersome and time consuming, and drug companies now routinely assay several hundred thousand compounds against each new drug target by the use of modern HTS techniques (see below). In connection with this, several complementary methods now exist by which large combinatorial peptide libraries may be made available (see Chapter 8).

The importance of peptides as tools in drug discovery has been reviewed by Grøn and Hyde-DeRuyscher [122]. The initial step in drug discovery is the selection of a suitable target molecule, and the number of proteins seen as potential targets for drug intervention in order to control human disease or injury has been estimated to be in the range 2000 to 5000 [123]. Despite this, the drugs that are currently on the market, together with those which have been discovered during the past 100 years, have been calculated to be directed against not more than 500 target proteins [124]. Interestingly, the term *chemogenomics* has been coined as relating to the discovery and description of all possible drugs to all possible drug targets [125]. As mentioned above, the terms genomics and proteomics (cf. Chapter 10) define the process of identifying and classifying all genes in a genome, as well as the correlation between a gene expression pattern and the phenotype at different stages. Protein modeling forms an integral part of the drug discovery effort [126]. A functional understanding

of novel gene products will increase the number of suitable drug targets based on clear synergies of the combination of target structural information with combinatorial chemistry.

9.4.4

Peptides Targeted to Functional Sites of Proteins

A functional site of a target protein is characterized as an area where binding of a ligand – a small molecule or a protein – modulates activity. As shown above, most proteins interact with other proteins, but the number of residues critical for binding sometimes is rather low, comprising three to ten amino acids [127]. Thus peptides, for example from combinatorial libraries, may act as “surrogate” ligands. Functional sites are mostly located at grooves in the protein surface [128], and comprise flexible areas where favorable interactions with the ligand support formation of the protein–ligand complex. Often, interactions with single water molecules stabilize the empty functional site of the native protein. One of the driving forces for peptide binding to a target protein is the displacement of water molecules from recesses or cavities in the protein, mainly because of entropic reasons. Target-specific peptides can be used to understand the nature of functional sites and to identify potential binding partners; moreover, they serve as valuable tools in structure-based and HTS drug discovery, as will be shown below.

As mentioned, many peptides have poor pharmacological properties. Consequently, the question remains as to how a peptide ligand that binds to an active site of a target protein can be converted into a drug. Peptides may act immediately as agonists or antagonists under special circumstances, as is the case of cell-surface receptors. Because most small peptides are easily proteolyzed, rapidly excreted and poorly bioavailable, special short-lived peptides are only used for the treatment of acute health problems by intravenous or subcutaneous injection. These limitations have thus necessitated the development of techniques to replace portions of peptides with nonpeptide structures, and this has resulted in nonpeptide therapeutics. Additionally, it is possible to design peptidomimetics (cf. Chapter 7 and Section 9.4.7) that are protease-resistant, readily cross the plasma membrane, and also show desirable pharmacokinetic properties [129].

9.4.5

Peptides Used in Target Validation

Target validation is necessary to clarify the function of a protein in a specific biochemical pathway. Peptides may also find application for target validation in the drug discovery process. The increasing amount of genome data, both of the human cell and of selected human pathogens, has provided a rich source of interesting targets. The best candidates for pharmacological interventions can be elucidated by the usual target validation tools such as gene knockouts and targeted mutations, in combination with bioinformatics. Unfortunately, genetic knockouts and mutations may result in the complete loss of all functions of the target protein, and the

deletion of the protein can therefore be misleading. Target validation with peptides will be faster and can be achieved much more selectively. A peptide normally interferes with only one of several specific functional sites of a protein target, and this resembles the action of a drug. Suitable peptide ligands can either be introduced into a cell or expressed inside cells. They may bind to the target protein, and the physiological effects of binding can be monitored in order to predict the response of a drug binding to the same site.

Several means for validating a target with a peptide have been developed. Upon injection of target-specific peptides, for example, for the Src homology 3 (SH3) domain into *Xenopus laevis* oocytes, an acceleration of progesterone-stimulated maturation was observed [130]. This effect might be caused by peptide-induced modulation of the protein, or by a signal transduction pathway. Furthermore, peptide ligands that are specific for the tyrosine kinase Lyn SH3 domain have been transported into mast cells by electroporation, resulting in an inhibition of mast cell activation [131]. The activity of peptidic ligands might be of short duration because of intracellular proteolysis, though this limitation can be overcome using peptides composed of either D-amino acids or β -amino acids and γ -peptides [132, 133].

Another delivery route, already discussed in Section 9.2.2, is based on linking peptides to other peptides or protein domains. The resulting peptide conjugates have the capacity to cross the plasma membrane in order to modulate target activity inside cells. In principle, peptides can also be expressed inside cells, either alone or fused to an innocuous reporter protein, for example, to the green fluorescent protein [134] using recombinant DNA [135].

9.4.6

High-throughput Screening (HTS) Using Peptides as Surrogate Ligands

A third possibility to utilize peptides in the drug discovery process is the design of *in vitro* modular assays suitable for HTS technology systems of small molecule libraries. Peptides directed to special protein functional sites are used to format an assay where molecules are tested for their capability to displace bound peptide ligands, or to prevent binding. Several competitive binding assays are currently in use to detect inhibitors of peptide binding, and for many targets compounds have been identified to inhibit the activity of the target protein. Various detection formats, including scintillation proximity assays, time-resolved fluorescence (TRF), fluorescence polarization (FP) and fluorescence resonance energy transfer (FRET), have found application for the detection of inhibitors using peptide surrogate ligands. The assays can be performed automatically in a high throughput mode, and it is possible to collect up to 200 000 or more data points per day with appropriate robotic workstations. Any target protein for which a peptide surrogate ligand has been elucidated can be used in the inhibitor screening of large compound libraries.

A universal assay technology called Transcreener™ HTS Assay Platform (Bell-Brook Labs, Madison, WI, USA) has been developed which relies on a proprietary fluorescence polarization detection method for group transfer enzymes that enables

an entire family of enzymes to be screened using the same detection reagents [136]. Group transfer reactions, such as phosphorylation and glycosylation involving peptidic substrates, serve as important on/off switches for signaling proteins in diverse disease pathways. The Transcreener™ platform relies on detection of the product of donor molecule cleavage; e.g., ADP for kinases, Coenzyme A for acetyltransferases, etc. There is only one “donor product” for each type of group transfer reaction, so a single set of Transcreener™ detection reagents can be used for all family members, regardless of the acceptor substrate.

The homogeneous time-resolved fluorescence (HTRF) assay [137] eliminates many disadvantages associated with some conventional screening assay methodologies like in-plate binding assays and radiometric assays. HTRF is performed in completely homogeneous solution without the need for coating plates, solid supports or time-consuming separation steps. Furthermore, background fluorescence is eliminated and there is no requirement for special handling, monitoring or disposal of reagents. Each microplate is measured in less than one second. HTRF is based on fluorescence resonance energy transfer between the donor fluorophore europium cryptate (Euk) and the acceptor fluorophore XL665 which is a modified allophycocyanin (Figure 9.5).

A slow signal decay is observed at 665 nm when two biomolecules labelled with the fluorophores bind to each other. The energy of the laser at 337 nm is absorbed by EuK which transfers its energy to XL665, that emits the fluorescence signal at 665 nm with a slow decay time. HTRF has been proven to be feasible for the detection of protein–protein interaction and receptor binding with a variety of targets like tyrosine kinases, viral proteases and antibodies.

In this respect the AlphaScreen™ technology is an ideal tool that allows screening for a broad range of targets. The technology provides an easy and reliable means to determine the effect of compounds on biomolecular interactions and activities in particular protein–protein interactions [138]. AlphaScreen™ relies on the use of donor and acceptor beads that are coated with a layer of hydrogel providing functional groups for bioconjugation. When a biological interaction between molecules brings the beads into proximity, a cascade of chemical reactions

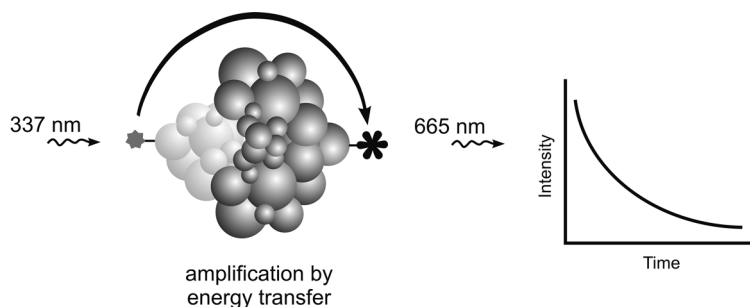


Figure 9.5 Simplified principle of the homogenous time-resolved fluorescence (HTRF) screening assay according to Kolb *et al.* [137].

is initiated to produce a greatly amplified signal. Upon laser excitation, a photosensitizer in the donor bead converts ambient oxygen to a more excited singlet state. The singlet state oxygen molecules diffuse across to react with a chemiluminescer in the acceptor bead that further activates fluorophores contained within the same bead. The fluorophores subsequently emit light at 520–620 nm. In the absence of a specific biological interaction, the singlet state oxygen molecules produced by the donor bead go undetected without the close proximity of the acceptor bead. AlphaScreen has successfully been developed for enzyme assays (kinase, helicase, protease and others), interaction assays (ligand/receptor, protein/protein, protein/DNA), immunoassays, and GPCR functional assays (cAMP, IP₃) [139].

A further interesting tool for HTS is based on a conformationally dependent binding of peptides to receptors. Traditional assays searching for agonists and antagonists of hormone receptors are based on a binding assay where a labelled natural ligand competes with library constituents. Unfortunately, these assays are not capable of differentiating between an agonist or antagonist; additional cell-based model systems, or even animal models, are then required for the elucidation of the biological effect. In contrast, HTS assays can be formatted for the search for compounds with specific effects on receptor conformation that will contribute to knowledge on biological effects.

The emerging field of epigenetics, in particular involving histone modification, demonstrates the power of peptides as surrogate substrates in drug discovery. Histones are known to carry plenty of different posttranslational modifications like acetyl and methyl groups which are added by many enzyme families. Distinct modification patterns of histones comprising the histone code are read by many transcription factors and enzymes with specific binding motifs for distinct modifications triggering cellular events like gene transcription or chromatin condensation [140, 141]. Methylation of distinct lysine residues of the histone tails is carried out by at least 50 SET domains containing histone methyltransferases (HMT) identified in the human genome so far [142]. Lysine specific histone methyltransferases (KMT) use S-adenosyl methionine (SAM) as a cosubstrate and catalyze the transfer of the methyl group from AdoMet to these lysine residues, thereby producing S-adenosyl homocysteine (AdoHcy). The differences in accepted substrates have crucial impact on the biological roles of the enzymes. Different methylation marks correlate with distinct processes from transcriptional activation or repression, respectively, to stem cell maintenance and differentiation, X-inactivation, and DNA damage response [143, 144]. KMT are intimately linked to tumorigenesis and histone methyltransferase inhibitors are thought to be of value for a therapeutic intervention [145, 146]. Despite this fact, only a very few small molecule inhibitors for histone methyltransferases are known besides generic analogs of SAM. In 2005 the first specific inhibitor for a KMT (SU(VAR)3–9) was identified by screening a small compound library [147]. Only recently, a specific G9a histone methyltransferase inhibitor was identified employing a HTS approach [148]. Selective inhibitors are strongly demanded as KMT have distinct genomic targets and biological roles, e.g. in differentiation and development and usage of pan-inhibitors will likely cause side-effects. In particular, unspecific SAM analogs will interfere with

the many biological functions executed by a huge number of unrelated SAM-dependent methyltransferases. As a preferred assay format, the methylation of substrate peptides representing truncated versions of the natural substrate proteins (histones) is determined by a FRET-based assay approach.

HTS technologies underwent a revolution during the late 1990s, with the result that most pharmaceutical companies now use HTS as the primary tool for lead discovery [149–151]. New HTS techniques have significantly increased throughput, and have also reduced assay volumes in offering a new technology for the 21st century [152]. The transition from slow, manual, low-throughput screening to robotic *ultrahigh-throughput screening* (*uHTS*) will soon allow screening of more than 200 000 samples per day. In addition, new fluorescence methods [153, 154], photoactivatable ligands [155], and miniaturized HTS technologies [156–159], together with key advances in both detection platforms and liquid handling technologies have contributed to the rapid development of *uHTS*. Modern detection platforms demonstrate significant improvements in sensitivity and throughput, whilst new liquid handling methods allow the dispensing of compounds and reagents in volumes consistent with miniaturized assay formats. The development from 96-well screening on the microscale towards higher density (for example, 1536-well) nanoscale formats and the advent of homogeneous fluorescence detection technologies serve as benchmarks in HTS development. Ullmann *et al.* [160] described both new applications and instrumentation for confocal fluctuation fluorescence-based HTS, and new two-dimensional applications of this methodology in which molecular brightness analysis (FIDA) is combined with molecular anisotropy measurements and other reactant principles.

In summary, peptide surrogate ligands play a fundamental role in modern drug discovery programs. Besides other applications, they have been used to format sensitive HTS systems in order to identify compounds that modulate the function of the target protein. Most importantly, they are the starting points for drug leads [161].

9.4.7

Artificial Peptide Analogs in Drug Discovery

The phage-display approach [162] consists of generating peptide phage-display libraries (cf. Section 8.2.5), screening them against targets of interest, and identifying high affinity and specificity target binding compounds. Several companies use this approach successfully for peptide drug discovery. Linear and constrained loop peptide libraries as tools for peptide drug discovery are commercially available with greater than 10 billion in each [163]. The phage-display technology often results in novel peptide leads with little to no sequence similarity to any known human sequences. This can be an advantage over natural peptides, which can be highly unstable and have multiple binding partners *in vivo*. Naturally derived peptides or peptide derivatives are often covered by existing intellectual property or in the public domain. Since phage display derived peptides are often completely novel and large motifs can be generated, it is possible to get strong patent protection around a family of peptide binders.

Peptide aptamers are artificial peptides and proteins selected from combinatorial libraries that display conformationally constrained variable regions [135, 164]. They bind to target proteins with a high specificity and a strong affinity [165]. Peptide aptamers are selected from randomized expression libraries based on their *in vivo* binding capacity to the appropriate target protein. Inserted peptides are expressed as part of the primary sequence of a structurally stable protein, termed “scaffold”. This is achieved by the insertion of oligonucleotides encoding the peptide into existing or engineered restriction sites in the open reading frame encoding the scaffold. An ideal scaffold should not interact with any cellular molecule or organelle and should not show enzymatic activity. Peptide aptamers are capable of disrupting specific protein interactions. Peptide aptamer technology has the advantage over existing techniques that the reagents identified are designed for expression in eukaryotic cells. Analogous to intracellular antibodies, peptide aptamers are capable of binding specifically to a given target protein, both *in vitro* and *in vivo*, with the potential to block selectively the function of their target protein.

Peptide aptamers can modulate the function of their cognate targets. Because peptide aptamers introduce perturbations that are similar to those caused by therapeutic molecules, their use identifies and/or validates therapeutic targets with a higher confidence level than is typically provided by methods that act upon protein expression levels. The unbiased combinatorial nature of peptide aptamers enables them to ‘decorate’ numerous polymorphic protein surfaces, whose biological relevance can be inferred through characterization of the peptide aptamers. Bioactive aptamers that bind druggable surfaces can be used in displacement screening assays to identify small-molecule hits to the surfaces. The peptide aptamer technology has a positive impact on drug discovery by addressing major causes of failure and by offering a seamless, cost-effective process from target validation to hit identification. In summary, peptide aptamers are powerful new tools for molecular medicine. Blocking the intracellular function of a target protein by peptide aptamers allows the investigation of distinct physiological and pathological processes within living cells. Furthermore, peptide aptamers meet the requirements for the development of novel diagnostic and therapeutic strategies with potential importance for a broad variety of various disease entities such as metabolic disorders, infections, and cancer.

9.5

Review Questions

- Q9.1. Name five different sources/routes to therapeutic peptides and proteins.
- Q9.2. How can pharmacokinetics and bioavailability of peptide and protein drugs be improved?
- Q9.3. Which routes for administration of peptide and protein drugs do you know?
- Q9.4. What is the difference between endogenous protein pharmaceuticals and engineered protein pharmaceuticals?
- Q9.5. How can monoclonal antibodies be employed for therapeutic purposes?

- Q9.6. Enfuvirtide (Fuzeon™; T20) is the longest synthetic peptide drug with the largest annual production amount. On what annual scale is it produced? Summarize the synthetic approach.
- Q9.7. Give at least five examples of approved peptide drugs.
- Q9.8. How can synthetic peptides be employed in drug discovery?
- Q9.9. What are surrogate ligands in high-throughput screening.
- Q9.10. Explain peptide aptamer technology.

References

- 1 P.M. Watt, *Nat. Biotechnol.* **2006**, 24, 177.
- 2 T. Vorherr, *Chim. Oggi* **2003**, 21, 14.
- 3 R.C. Ladner, A.K. Sato, J. Gorzelany, M. de Souza, *Drug Discov. Today* **2004**, 9, 529.
- 4 L.A. Landon, J. Zou, S.L. Deutscher, *Curr. Drug Discov. Technol.* **2004**, 1, 113.
- 5 G. Wright, A. Carver, D. Cottom, D. Reeves, A. Scott, P. Simons, I. Wilmut, I. Garner, A. Colman, *Biotechnology* **1991**, 9, 830.
- 6 C. Cunningham, A.J.R. Porter, *Methods in Biotechnology*, Vol. 3 Humana Press, Totowa, **1997**.
- 7 J. McCafferty, D.R. Glover, *Curr. Opin. Structural. Biol.* **2000**, 10, 417.
- 8 F.M. Veronese, G. Pasut, *Drug Discov. Today* **2005**, 10, 1451.
- 9 R. Duncan, *Anticancer Drugs* **1992**, 3, 175.
- 10 M. Werle, A. Bernkop-Schnürch, *Amino Acids* **2006**, 30, 351.
- 11 E. Lipp, *Genetic Eng. Biotechnol. News* **2008**, 28 (7) 44.
- 12 J.A. Dumont, S.C. Low, R.T. Peters, A.J. Bitonti, *BioDrugs* **2006**, 20, 151.
- 13 S.J. Rich, C.E. Bello-Quintero, *J. Manag. Care Pharm.* **2004**, 10, 318.
- 14 T. Doan, E. Massarotti, *J. Clin. Pharmacol.* **2005**, 45, 751.
- 15 J.L. Nichol, *Pediatr. Blood Cancer* **2006**, 47, 723.
- 16 S. Sutherland, *Drug Discov. Today* **2004**, 9, 683.
- 17 V. Balan, D.R. Nelson, M.S. Sulkowski, G.T. Everson, L.R. Lambiase, R.H. Wiesner, R.C. Dickson, A.B. Post, R.R. Redfield, G.L. Davis, A.U. Neumann, B.L. Osborn, W.W. Freimuth, G.M. Subramanian, *Antivir. Ther.* **2006**, 11, 35.
- 18 W. Wang, Y. Ou, Y. Shi, *Pharm. Res.* **2004**, 21, 2105.
- 19 L.L. Baggio, Q. Huang, D.J. Drucker, *Diabetes* **2004**, 53, 2492.
- 20 J. Silverman, Q. Liu, A. Bakker, W. To, A. Duguay, B.M. Alba, R. Smith, A. Rivas, P. Li, H. Le, E. Whitehorn, K.W. Moore, C. Swimmer, V. Perloth, M. Vogt, J. Kolkman, W.P. Stemmer, *Nat. Biotechnol.* **2005**, 23, 1556.
- 21 D.K. Pettit, W.R. Gombotz, *Trends Biotechnol.* **1998**, 16, 343.
- 22 J.A. Dumont, A.J. Bitonti, D. Clark, S. Evans, M. Pickford, S.P. Newman, *J. Aerosol. Med.* **2005**, 18, 294.
- 23 M. Zorko, Ü. Langel, *Adv. Drug Deliv. Rev.* **2005**, 57, 529.
- 24 K.M. Wagstaff, D.A. Jans, *Curr. Med. Chem.* **2006**, 13, 1371.
- 25 Ü. Langel (Ed.), *Handbook of Cell-Penetrating Peptides*, CRC Press, Boca Raton, USA, **2007**.
- 26 J.M. Varga, *Methods Enzymol.* **1985**, 112, 259.
- 27 A.V. Schally, A. Nagy, *Eur. J. Endocrinol.* **1999**, 141, 1.
- 28 D. Zwanziger, I.U. Khan, I. Neundorff, S. Sieger, L. Lehmann, M. Friebe, L. Dinkelborg, A.G. Beck-Sickinger, *Bioconj. Chem.* **2008**, 19, 1430.
- 29 C. Rousselle, P. Clair, J.M. Lefauconnier, M.M. Kaczorek, J.M. Scherrmann, J. Temsamani, *Mol. Pharmacol.* **2000**, 57, 679.

- 30 D. Derossi, G. Chassaing, A. Prochiantz, *Trends Cell Biol.* **1998**, 8, 84.
- 31 W. Arap, R. Pasqualini, E. Ruoslahti, *Science* **1998**, 279, 377.
- 32 W. Mier, R. Eritja, A. Mohammed, U. Haberkorn, M. Eisenhut, *Bioconj. Chem.* **2000**, 11, 855.
- 33 K.A. Witt, T.J. Gillespie, J.D. Huber, R.D. Egleton, T.P. Davis, *Peptides* **2001**, 22, 2329.
- 34 R.G. Egleton, T.P. Davis, *J. Pharm. Sci.* **1999**, 88, 392.
- 35 A. Persidis, *Nature Biotechnol.* **1998**, 16, 1378.
- 36 D. Blohm, C. Bollschweiler, H. Hillen, *Angew. Chem. Int. Ed.* **1988**, 27, 207.
- 37 G. Gellissen (Ed.), *Production of Recombinant Proteins*, Wiley-VCH, Weinheim, **2004**.
- 38 N. Dafny, P.B. Yang, *Eur. J. Pharmacol.* **2005**, 523, 1.
- 39 D.J. Cassel, S. Choudhri, R. Humphrey, R.E. Martell, T. Reynolds, A.B. Shanafelt, *Curr. Pharm. Des.* **2002**, 8, 2171.
- 40 F.F. Little, W.W. Cruikshank, *Exp. Opin. Ther.* **2004**, 4, 837.
- 41 W.C. Zhou, C. Chen, B. Buckland, J. Aunins, *Biotechnol. Bioeng.* **1997**, 55, 783.
- 42 M. Hall, *An A to Z of British Medicine Research*, Association of British Pharmaceutical Industry, London, **1998**.
- 43 M.H.V. Van Regenmortel, *Biologicals* **2001**, 29, 209.
- 44 P.J. Cachia, R.S. Hodges, *Biopolymers (Pept. Sci.)* **2003**, 71, 141.
- 45 K. Deres, H. Schild, K.-H. Wiesmüller, G. Jung, H.D. Rammensee, *Nature* **1989**, 342, 561.
- 46 H. Schild, K. Deres, K.-H. Wiesmüller, G. Jung, H.G. Rammensee, *Eur. J. Immunol.* **1991**, 21, 2649.
- 47 K.-H. Wiesmüller, B. Fleckenstein, G. Jung, *Biol. Chem. Hoppe-Seyler* **2001**, 382, 571.
- 48 T. Ben-Yedidia, R. Arnon, *Curr. Opin. Biotechnol.* **1997**, 8, 442.
- 49 J.H. Fritz, S. Brunner, M.L. Birnstiel, M. Buschle, A. von Gabain, F. Mattner, W. Zauner, *Vaccine* **2004**, 22, 3274.
- 50 A. Mouzaki, S. Deraos, K. Chatzantoni, *Curr. Med. Chem.* **2005**, 12, 1537.
- 51 A. Nissim, Y. Chernajowsky, *Handb. Exp. Pharmacol.* **2008**, (181), 3.
- 52 L.G. Presta, *Curr. Pharm. Biotechnol.* **2002**, 3, 237.
- 53 G. Galfre, C. Milstein, *Methods Enzymol.* **1981**, 73, 3.
- 54 R.M. Sharkey, D.M. Goldenberg, *CA Cancer J. Clin.* **2006**, 56, 226.
- 55 T.J. Vaughan, J.K. Osbourn, P.R. Tempest, *Nature Biotechnol.* **1998**, 16, 535.
- 56 D.J. Norman, F. Vincenti, A.M. de Mattos, J.M. Barry, D.J. Lewitt, N.I. Wedel, M. Maia, S.E. Light, *Transplantation* **2000**, 70, 1707.
- 57 D.P. Pollock, J.P. Kutzko, E. Birck-Wilson, J.L. Williams, Y. Echelard, H.M. Meade, *J. Immunol. Methods* **1999**, 231, 147.
- 58 M.A. van Dijk, J.G.J. van den Winkel, *Curr. Opin. Chem. Biol.* **2001**, 5, 368.
- 59 J. McCafferty, A.D. Griffiths, G. Winter, D.J. Chiswell, *Nature* **1990**, 348, 552.
- 60 M. Bruggemann, M.J. Taussig, *Curr. Opin. Biotechnol.* **1997**, 8, 455.
- 61 J. Kempeni, *Ann. Rheum. Dis.* **1999**, 58, 170.
- 62 E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al. *Nature* **2001**, 409, 860.
- 63 J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, et al. *Science* **2001**, 291, 1304.
- 64 W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, D. Haussler, *Genome Res.* **2002**, 12, 996.
- 65 F. Lottspeich, *Angew. Chem. Int. Ed.* **1999**, 38, 2476.
- 66 M. Walker (Ed.), *The Proteomics Protocols Handbook*, Humana Press, Totowa, N.J., **2005**.
- 67 M. Cretich, F. Damin, G. Pirri, M. Chiari, *Biomol. Eng.* **2006**, 23, 77.
- 68 H. Zhu, M. Snyder, *Curr. Opin. Chem. Biol.* **2003**, 7, 55.
- 69 N. Ramachandran, E. Hainsworth, B. Bhullar, S. Eisenstein, B. Rosen,

- A.Y. Lau, J.C. Walter, L. LaBaer, *Science* **2004**, 305, 86.
- 70 T. Cha, A. Guo, X.Y. Zhu, *Proteomics* **2005**, 5, 414.
- 71 B. Guilleaume, A. Bunes, C. Schmidt, F. Klimek, G. Moldenhauer, W. Huber, D. Arlt, U. Korf, S. Wiemann, A. Poustka, *Proteomics* **2005**, 5, 4705.
- 72 L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, B. Nilsson, M. Verlander, *Biopolymers* **2000**, 55, 227.
- 73 T. Vorherr, F. Dick, B. Sax, *Chimia* **2005**, 59, 25.
- 74 T. Vorherr, *Chim. Oggi* **2007**, 25, 22.
- 75 T. Vorherr, *Spec. Chem. Mag.* **2007**, 27.
- 76 O. Marder, F. Albericio, *Chim. Oggi* **2003**, 21, 6.
- 77 M. Feurer in: *Peptides: Proceedings of the 5th American Peptide Symposium*, M. Goodman, J. Meienhofer (Eds.), Wiley, New York, **1977** p. 487.
- 78 K.-E. Andersson, B. Bengtsson, O. Paulsen, *Drugs Today* **1988**, 24, 509.
- 79 M. Verlander, *Int. J. Peptide Res. Ther.* **2007**, 13, 75.
- 80 C. Johansson, L. Blomberg, E. Hlebowicz, H. Nicklasson, B. Nilsson, L. Andersson, in: *Peptides 1994*, H.L.S. Maia (Ed.), Escom, Leiden, **1995**, p. 34.
- 81 P. Melin, *Baillieres Clin. Obstet. Gynaecol.* **1993**, 7, 577.
- 82 B.L. Bray, *Nat. Rev. Drug. Discov.* **2003**, 2, 587.
- 83 M.C. Kang, B.L. Bray, M. Lichty, C. Mader, G. Merutka, *US 6015881*, **2000**.
- 84 V. Marx, *Chem. Eng. News* **2005**, 83, 16.
- 85 I.F. Eggen, T. Balelaar, A. Petersen, P.B.W. Ten Kortenaar, *Org. Proc. Res. Dev.* **2005**, 5, 98.
- 86 D.J. Ward, *Peptide Pharmaceuticals: approaches to the design of novel drugs*, Open University Press, Philadelphia, **1991**.
- 87 P.W. Lathan, *Nature Biotechnol.* **1999**, 17, 755.
- 88 T. Bruckdorfer, O. Marder, F. Albericio, *Curr. Pharm. Biotechnol.* **2004**, 5, 29.
- 89 L. Gentilucci, A. Tolomelli, F. Squassabia, *Curr. Med. Chem.* **2006**, 13, 2449.
- 90 A.K. Sato, M. Viswanathan, R.B. Kent, C.R. Wood, *Curr. Opin. Biotechnol.* **2006**, 17, 638.
- 91 M. Roach, A. Izaguirre, *Expert. Opin. Pharmacother.* **2007**, 8, 257.
- 92 A.H. Barnett, D.R. Owens, *Lancet* **1997**, 349, 47.
- 93 E. Ciszak, J.M. Beals, B.H. Frank, J.C. Baker, N.D. Carter, G.D. Smith, *Structure* **1995**, 3, 615.
- 94 R. Hilgenfeld, M. Dörschug, K. Geisen, H. Neubauer, R. Obermeier, G. Seipke, H. Berchtold, *Diabetologia* **1992**, 35 (Suppl. 1), A193.
- 95 L.L. Nielsen, A.A. Young, D.G. Parkes, *Regul. Peptide* **2004**, 117, 77.
- 96 P.B. Jeppesen, E.L. Sanguinetti, A. Buchman, L. Howard, J.S. Scolapio, T.R. Ziegler, J. Gregory, K.A. Tappenden, J. Holst, P.B. Mortensen, *Gut* **2005**, 54, 1224.
- 97 J.P. Potts, *J. Endocrinol.* **2005**, 187, 311.
- 98 H. White, A. Ahmad, *Curr. Opin. Investig. Drugs* **2005**, 10, 1057.
- 99 C. Deal, J. Gedeon, *Clev. Clin. J. Med.* **2003**, 70, 585.
- 100 L.A. Sorbera, *Drugs Fut.* **2006**, 31, 670.
- 101 D.L. Johnson, F.X. Farrell, F.P. Barbone, F.J. McMahon, J. Tullai, K. Hoey, O. Livnah, N.C. Wrighton, S.A. Middleton, D.A. Loughney, E.A. Stura, W.J. Dower, L.S. Mulcahy, I.A. Wilson, L.K. Jolliffe, *Biochemistry* **1998**, 37, 3699.
- 102 A.S. Kesselheim, M.A. Fischer, J. Avorn, *Health Affairs* **2006**, 25, 1095.
- 103 A.M. Comaru-Schally, A.V. Schally, *Int. J. Oncol.* **2005**, 26, 301.
- 104 Z. Xia, Y. Chen, Y. Zhu, F. Wang, X. Xu, J. Zhan, *BioDrugs* **2006**, 20, 275.
- 105 S. Eldabe, *Fut. Neurol.* **2007**, 2, 11.
- 106 A.R. Hamel, F. Hubler, A. Carrupt, R.M. Wenger, M. Mutter, *J. Pept. Sci.* **2004**, 63, 147.
- 107 C.T. Esmon, *Chest* **2003**, 124, 26S.
- 108 J.L. Pace, G. Yang, *Biochem. Pharmacol.* **2006**, 71, 968.
- 109 Y.L. Janin, *Amino Acids* **2003**, 25, 1.
- 110 P. Burke, S. DeNardo, L. Miers, K. Lamborn, S. Matzku, G. DeNardo, *Cancer Res* **2002**, 62, 4263.

- 111 S.L. Goodman, G. Hoelzemann, G.A.G. Sulyok, H. Kessler, *J. Med. Chem.* **2002**, 45, 1045.
- 112 S. Liu, H. Lu, J. Niu, Y. Xu, S. Wu, S. Jiang, *J. Biol. Chem.* **2005**, 280, 11259.
- 113 T. Matthews, M. Salgo, M. Greenberg, J. Chung, R. DeMasi, D. Bolognesi, *Nat. Rev. Drug Discov.* **2004**, 3, 215.
- 114 A.G. Cochran, *Chem. Biol.* **2000**, 7, 85.
- 115 H.W. Lark, G.W. Stroup, S.M. Hwang, I.E. James, D.J. Rierman, F.H. Drake, S.N. Bradbeen, A. Mathur, K.F. Erhand, K.A. Newlander Stephen, T. Ross, K.L. Salyers, B.R. Smith, W.H. Miller, W.F. Huffman, M. Gowen, *J. Pharmacol. Exp. Ther.* **1999**, 291, 612.
- 116 M.A. Blaskovich, Q. Lin, F.L. Delarue, J. Sun, H.S. Park, D. Coppola, A.D. Hamilton, S.M. Sebti, *Nature Biotechnol.* **2000**, 18, 1065.
- 117 J.-L. Wang, D. Liu, Z.-J. Zhan, S. Shan, X. Han, S.M. Srinivasula, C.M. Croce, E.S. Alnemri, Z. Huang, *Proc. Natl. Acad. Sci. USA* **2000**, 97, 7124.
- 118 T. Clackson, J.A. Wells, *Science* **1995**, 267, 383.
- 119 G. Welsh, *Trends Biotechnol.* **2005**, 23, 553.
- 120 V. Glaser, *Genetic Eng. Biotechnol. News* **2002**, 22, 25.
- 121 G. Dorman, G.D. Prestwich, *Trends Biotechnol.* **2000**, 18, 64.
- 122 H. Grøn, R. Hyde-DeRuyscher, *Curr. Opin. Drug Discov. Dev.* **2000**, 3, 636.
- 123 J. Drews in: *Human Disease: From Genetic Causes to Biochemical Effects*, J. Drews, S. Ryser (Eds.), Blackwell, Berlin, **1997**.
- 124 J. Drews, S. Ryser, *Nature Biotechnol.* **1997**, 15, 1318.
- 125 P.R. Caron, M.D. Mullican, R.D. Mashal, K.P. Wilson, M.S. Su, M.A. Murcko, *Curr. Opin. Chem. Biol.* **2001**, 5, 464.
- 126 M. Norin, M. Sundström, *Curr. Opin. Drug Discov. Dev.* **2001**, 4, 284.
- 127 J.A. Wells, *Proc. Natl. Acad. Sci. USA* **1996**, 93, 1.
- 128 R.A. Laskowski, N.M. Luscombe, M.B. Swindells, J.M. Thornton, *Protein Sci.* **1996**, 5, 2438.
- 129 A.S. Ripka, D.H. Rich, *Curr. Opin. Chem. Biol.* **1998**, 2, 441.
- 130 A.B. Sparks, L.A. Quilliam, J.M. Thom, C.J. Der, B.K. Kay, *J. Biol. Chem.* **1994**, 269, 23853.
- 131 T.P. Stauffer, C.H. Martenson, J.E. Rider, B.K. Kay, T. Meyer, *Biochemistry* **1997**, 36, 9388.
- 132 D.H. Appella, L.A. Christianson, I.L. Karle, D.R. Powell, S.H. Gellman, *J. Am. Chem. Soc.* **1996**, 118, 13071.
- 133 D. Seebach, A.K. Beck, D.J. Bierbaum, *Chem. Biodivers.* **2004**, 1, 1111.
- 134 M. Chalfie, Y. Tu, G. Euskirchen, W.W. Ward, D.C. Prasher, *Science* **1994**, 263, 802.
- 135 P. Colas, B. Cohen, T. Jessen, I. Grishina, J. McCoy, R. Brent, *Nature* **1996**, 380, 548.
- 136 Y. Liu, L. Zalameda, K.W. Kim, M. Wang, J.D. McCarter, *Assay and Drug Dev. Tech.* **2007**, 5, 225.
- 137 A.J. Kolb, P.V. Kaplita, D.J. Hayes, Y.-W. Park, C. Pernell, J.S. Major, G. Mathis, *Drug Discov. Today* **1998**, 3, 333.
- 138 R.J. Kordal, A.M. Usmani, W.T. Law, *Microfabricated Sensors: Application of Optical Technology for DNA Analysis*, American Chemical Society, Washington D.C. **2002**.
- 139 Y. Hou, D.E. McGuinness, A.J. Prongay, B. Feld, P. Ingravallo, R.A. Ogert, C.A. Lunn, J.A. Howe, *J. Biomol. Screen.* **2008**, 13, 406.
- 140 B.D. Strahl, C.D. Allis, *Nature* **2000**, 403, 41.
- 141 T. Jenuwein, C.D. Allis, *Science* **2001**, 293, 1074.
- 142 J.F. Couture, R.C. Trievel, *Curr Opin Struct Biol.* **2006**, 16, 753.
- 143 K. Plath, J. Fang, S.K. Mlynarczyk-Evans, R. Cao, K.A. Worringer, H. Wang, C.C. de la Cruz, A.P. Otte, B. Panning, Y. Zhang, *Science* **2003**, 300, 13.
- 144 C. Martin, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2005**, 6, 838.
- 145 R. Schneider, A.J. Bannister, T. Kouzarides, *Trends Biochem. Sci.* **2002**, 27, 396.

- 146 P.A. Jones, S. Baylin, *Cell* **2007**, 128, 683.
- 147 D. Greiner, T. Bonaldi, R. Eskeland, E. Roemer, A. Imhof, *Nature Chem. Biol.* **2005**, 1, 143.
- 148 S. Kubicek, R.J. O'Sullivan, E.M. August, E.R. Hickey, Q. Zhang, M.L. Teodoro, S. Rea, K. Mechtler, J.A. Kowalski, C.A. Homon, T.A. Kelly, T. Jenuwein, *Mol. Cell* **2007**, 25, 473.
- 149 A.W. Lloyd, *Drug Discov. Today* **1997**, 2, 397.
- 150 J.J. Burbaum, *Drug Discov. Today* **1998**, 3, 313.
- 151 S.J. Rhodes, R.C. Smith, *Drug Discov. Today* **1998**, 3, 361.
- 152 R.P. Hertzberg, A.J. Pope, *Curr. Opin. Chem. Biol.* **2000**, 4, 445.
- 153 M. Auer, K.J. Moore, F.J. Meyer-Almes, R. Guenther, A.J. Pope, K.A. Stoeckli, *Drug Discov. Today* **1998**, 3, 457.
- 154 P. Kask, K. Palo, D. Ullmann, K. Gall, *Proc. Natl. Acad. Sci. USA* **1999**, 96, 13756.
- 155 G. Dorman, G.D. Prestwich, *Trends Biotechnol.* **2000**, 18, 64.
- 156 K.R. Oldenburg, J. Zhang, T. Chen, A. Maffia, III, K.F. Blom, A.P. Combs, D.Y. Chung, *J. Biomol. Screen.* **1998**, 3, 55.
- 157 U. Haupts, M. Ruediger, A.J. Pope, *Drug Discov. Today* **2000**, 1, 3.
- 158 R. Turner, D. Ullmann, S. Sterrer in: *Handbook of Drug Screening*, R. Seethala, P.B. Fernandes (Eds.), Marcel Dekker, New York, **2001**.
- 159 J. Wölcke, D. Ullmann, *Drug Discov. Today* **2001**, 6, 637.
- 160 D. Ullmann, M. Busch, T. Mander, *J. Pharm. Technol.* **1999**, 99, 30.
- 161 B.A. Kenny, M. Bushfield, D.J. Parry-Smith, S. Forgarty, J.M. Treheme, *Prog. Drug Res.* **1998**, 51, 245.
- 162 B.K. Kay, J. Winter, J. McCafferty, *Phage Display of Peptides and Proteins: A Laboratory Manual*, Academic Press, San Diego, **1996**.
- 163 M.A. Spear, X.O. Breakefield, J. Beltzer, D. Schuback, R. Weissleder, F.S. Pardo, R. Ladner, *Cancer Gene Ther.* **2001**, 8, 506.
- 164 F. Hoppe-Seyler, I. Crnkovic-Mertens, E. Tomai, K. Butz, *Curr. Mol. Med.* **2004**, 4, 529.
- 165 I.C. Baines, P. Colas, *Drug Discov. Today* **2006**, 11, 334.

10

Peptides in Proteomics

10.1

Genome and Proteome

The technology for sequencing the whole genome of an organism has experienced tremendous advances during the past decades. Many complete genomes of organisms with different complexity have been sequenced. This huge amount of data on the genome (DNA) and the transcriptome (RNA) level remains now to be correlated with molecular and cellular functions. However, global analysis of changes in gene transcription does not necessarily provide valid information on the proteins encoded by DNA and RNA. Proteins are the main effector molecules in a living organism. The term proteome was originally coined in 1994 as the “PROTEin complement of the genOME” [1]. A striking example is found when comparing different development states of insects. A caterpillar experiences significant changes in phenotype during metamorphosis into a butterfly. Both have the same genome but significant differences in the proteome. Accordingly the proteome is supposed to be a sensitive monitor for physiological changes and pathological disorders. In order to meet the challenge of correlating the genome of an organism with the proteome, the entirety of proteins present in the cell, tissue or organism at a certain time under certain conditions, the discipline proteomics as a parallel to genomics has emerged.

Proteome analysis represents a technical challenge because the number of different proteins being expressed by a cell under certain conditions ranges from several thousand for simple prokaryotes to more than 10 000 for eukaryotes. These proteins may largely differ with respect to physical and chemical properties, like hydrophilicity, hydrophobicity, pI, post-translational modifications, etc. In addition, the concentrations of proteins in a cell, tissue, or organ encompasses a tremendous dynamic range. It has been estimated that the concentration range of a proteome spans six to eight orders of magnitude in a cell and even ten to twelve in the human body, with blood plasma as an example [2]. There are highly abundant proteins, in most cases thoroughly characterized, besides proteins in low abundance that often exert regulatory functions and for which the physiological roles often remain to be elucidated. Moreover, the complexity of a proteome is increased further by gene and protein splicing as well as by the introduction of a large variety of hitherto described or yet undescribed types of post-translational modification [3] (cf. Section 3.2.2).

Unlike for DNA/mRNA there is no amplification method comparable to PCR for low abundant proteins and there are no general protein binding partners. As a consequence proteome analysis, the parallel analysis of the entirety of all proteins produced by a cell, a tissue or an organism at a certain time, has to deal with the separation of a large number of different proteins present at different concentrations with different physical and chemical properties. Therefore, proteome analysis has to rely on highly efficient separation strategies combined with sensitive analytical methods for protein identification. Last but not least, it is also necessary to distinguish proteins with respect to biological activity, because proteins, and especially enzymes, may either exist in an inactive form, e.g. as a “silent” pro-form or as an inhibited enzyme, or as active proteins. In this context peptides play important roles on different levels: in proteome analysis, proteins are characterized after proteolytic degradation in the form of peptides (see below). In addition peptides, but also other small-molecule protein ligands and inhibitors, may be used in the context of functional proteomics for enrichment, pulldown, or labelling of proteins.

Moreover, peptides have potential as disease markers. Different peptidomes might be present in the serum of cancer patients and healthy individuals as a result of cancer-associated changes in proteolytic processing [4]. Based on this assumption proteomic peptide patterns (as opposed to identified peptide sequences), recorded by high resolution mass spectrometers and interpreted by suitable pattern analysis software, could in the future serve as diagnostic tools [5].

10.2

Separation Methods

10.2.1

Depletion Strategies

About 50–100 highly abundant proteins, which are usually not in the focus of the investigation, are often removed prior to the separation step. This is especially important for proteome analysis for example of blood plasma, as the classical, abundant plasma proteins are of relatively low interest in plasma proteomics where primarily biomarkers present in the plasma are to be identified [6]. Depletion strategies may comprise precipitation steps with organic solvents and ultrafiltration. Besides such unspecific methods where the risk of eliminating interesting low-abundant proteins in the depletion step is very high, specific depletion approaches using affinity chromatography have been established [7, 8].

10.2.2

Two-Dimensional Polyacrylamide Gel-Electrophoresis

Two-dimensional polyacrylamide gel-electrophoresis (2D-GE) remains the most widely used technique for protein separation in proteomics. 2D-GE usually combines isoelectric focussing (IEF) with SDS-polyacrylamide gel-electrophoresis (SDS-PAGE).

This allows the simultaneous detection of more than 5000 proteins which are separated according to their pI in the isoelectric focussing step and to their molecular mass in the SDS-polyacrylamide gelelectrophoresis step [9]. The advent of immobilized pH gradient (IPG) strips has been particularly important for IEF. However, 2D-GE still lacks the ability to resolve important protein classes like membrane-associated proteins [10], while low abundance proteins often fall below the detection limit [11].

Besides problems with sensitive detection and quantification using conventional staining methods, the analysis of proteins occurring with low abundance is especially problematic. Detection methods using fluorescent dyes are highly sensitive, but quantification is problematic because the dye is covalently linked to the proteins in an unselective manner prior to the separation step. Uniform labelling cannot be achieved and extensive labelling may lead to fluorescence quenching and solubility problems [12].

Proteome analysis is usually a differential approach where cells corresponding to different states are analyzed with respect to the protein profile and these two proteomes are subsequently compared to each other. However, separations by 2D-GE are often difficult to reproduce. Perfect superimposition cannot always be realized because of slightly varying experimental conditions. Hence, proteome analysis and comparison relying exclusively on image analysis is not straightforward.

Fluorescent differential 2D gel electrophoresis (DIGE) allows the comparison of two different samples on the same 2D-gel: Two protein samples are labelled with two spectrally distinct fluorophors which have the same charge and similar size. The differently labelled samples are combined and separated in the same gel which ensures that all samples will be subject to exactly the same 2D-GE conditions. The fluorescence images at the two different wavelengths are overlaid and subtracted, visualizing only differences in protein abundance [13, 14].

10.2.3

Gel-Free Methods – Two-Dimensional Liquid Chromatography (2D-LC, MudPIT)

“Gel-free” methods such as the *multidimensional protein identification technology* (MudPIT), represent an alternative to the 2D-GE-based strategies. MudPIT is based on liquid chromatography (LC) for separation and tandem mass spectrometry (MSⁿ) for peptide sequencing and is often combined with biochemical or chemical tagging steps.

Shotgun proteomics with MudPIT proceeds via the separation and identification of complex peptide mixtures by two-dimensional (or multi-dimensional) liquid chromatography (2D-LC), followed by MSⁿ. The proteins are digested with a protease and bound to a strong cation exchange column (first dimension), from where they are successively eluted with increasing salt concentrations onto a reversed-phase (RP) column (second dimension), which separates the peptides according to hydrophobicity. The LC system is coupled on-line to a mass spectrometer. The eluted peptides are ionized and subjected to gas-phase fragmentation sequencing (MS/MS) [15, 16].

More than 100 000 MS/MS spectra can nowadays be recorded within 24 h in a fully automated manner. Database search algorithms are subsequently employed to match the acquired spectra to peptide sequences from a protein database, e.g. SEQUEST, Mascot, Spectrum Mill, ProteinLynx, XTandem, or OMSSA [17]. The technique is suitable for the detection and identification of low abundance proteins, proteins with extreme pI, as well as membrane proteins. There is no limitation with respect to protein molecular mass, because the proteolytically generated peptide fragments are separated and identified. Membrane proteins are usually dissolved in formic acid and the loops between the hydrophobic transmembrane domains are cleaved e.g. with cyanogen bromide. The resulting buffer-soluble fragments are then subjected to proteolysis with trypsin [18].

10.3

Peptide and Protein Analysis in Proteomics

10.3.1

Mass Spectrometry

Mass spectrometry techniques of different sophistication levels are being employed in proteomics. In a chemometric approach, proteomic peptide patterns of healthy and pathologic tissue or serum are compared without extracting sequence information from the complex high resolution mass spectra.

Different mass spectrometric techniques are appropriate to identify proteins and peptides sensitively with high throughput. Once the proteins have been separated by 2D-GE every single spot has to be excized and digested by a protease (e.g. trypsin). This digestion step by the protease yields peptide fragments that can be identified by mass spectrometry (peptide mass fingerprint, PMF). In this case, genuine mass spectrometric peptide sequencing is usually not necessary, provided the genome of the organism has been sequenced and annotated. Protein sequence databases have been generated from genomic information. Sequencing and annotation of the genome provides protein sequences on the basis of the DNA sequence, where the peptides arising from tryptic digestion can be predicted. Matching of the molecular masses observed experimentally with the theoretically predicted peptide fragments within some mass error tolerance results in a hit list that eventually leads to the identification of the protein under investigation, provided the tryptic peptide fragments can be detected with sufficient sensitivity. Further sequencing of the peptides is often not necessary.

Tandem or multidimensional mass spectrometry (MS/MS or MSⁿ) is a technique that provides real sequence data by fragmentation of a peptide in the gas phase and, hence, allows peptide identification at a high level of confidence. For example, in an MS/MS experiment, a precursor ion (peptide) is selected and isolated for further collision, e.g. with an inert gas, that will produce daughter ions with distinctive signature (collision-induced dissociation, CID). Peptide mass sequencing (PMS) by MS/MS provides a partial primary structure of a peptide, and, consequently, of the

parent protein, which is then used for protein identification. Tandem mass spectrometry is usually combined with 2D-LC in the MudPIT approach [19].

10.3.2

Quantitative Proteomics

Quantitative proteomics aims at pairwise comparison of the abundance of proteins in two or more proteomes. Although several strategies have been published, “gel-based” quantitative proteomics has nowadays been largely superseded by “gel-free” MS-based approaches [19]. In the latter case quantification is performed with the mass spectrometric data. Basically three different approaches can be distinguished:

- metabolic stable-isotope labelling
- isotope tagging by chemical reaction
- stable-isotope incorporation via enzyme reaction

10.3.2.1 Metabolic Stable-Isotope Labelling

Differential labelling of two proteomes with stable isotopes in the SILAC (*Stable isotope labelling by amino acids in cell culture*) approach can be achieved by adding isotope-labelled amino acids, e.g. arginine with six ^{13}C atoms, to the cell culture [20]. While the cultivated cells in one state (proteome 1) are being fed with unlabelled Arg, the culture medium for the cells corresponding to the second state (proteome 2) contains ^{13}C -Arg. Similar approaches have proven successful with other compounds labelled with ^{13}C or other isotope combinations ($^{14}\text{N}/^{15}\text{N}$, $^{16}\text{O}/^{18}\text{O}$).

10.3.2.2 Tagging Methods

Chemical approaches have the advantage that a broad variety of isotopically labelled chemicals targeting reactive functional groups in proteins is available, e.g. thiols, amines (N-terminus, side chain), carboxylates, indole moieties [19]. The diverse functional groups in proteins allows one to design tailored tags for quantification. An enrichment step to reduce sample complexity is also possible. However, the conjugation reaction is required to be specific, complete and easy to perform.

One powerful LC-MS based strategy for proteomics involves application of *isotope-coded affinity tags* (ICAT) [21–23]. This approach is suitable for comparison of protein expression in proteomes by treating two proteome samples with non-isobaric forms of a chemical labelling reagent containing different isotopes like $^1\text{H}/^2\text{H}$, $^{12}\text{C}/^{13}\text{C}$, or $^{14}\text{N}/^{15}\text{N}$ (“light and heavy tags”). The tags comprise a thiol-specific reactive group with multiple incorporation of either light or heavy isotopes, conjugated to biotin, thus providing tag–protein or tag–peptide conjugates for the two proteome states that have different masses but are chemically identical.

Typically, the proteins of the two different proteome samples are denatured, solubilized and the disulfide bridges are reduced. The two proteome samples are then derivatized with the light and heavy ICAT tags, respectively (Figure 10.1). The higher nucleophilicity of thiol and thiolate groups at $\text{pH} < 8$ safeguards selective reaction with Cys residues in the presence of N-terminal and lysine side-chain amino

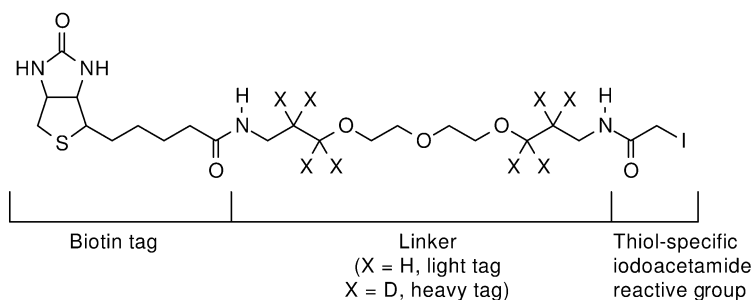


Figure 10.1 Molecular composition of a light and heavy tag for ICAT labelling.

groups, the imidazole ring, and the hydroxy groups of serine, threonine, and tyrosine. The differently tagged samples are pooled and digested with trypsin. The peptides containing Cys residues are enriched by avidin affinity chromatography, because they are conjugated to the biotinylated tag. This step reduces the complexity by a factor of 10 because only Cys-containing peptides are analyzed [22]. The labelled peptides of the two proteomes are combined and then subjected to multidimensional LC. The simultaneously eluted peptides labelled with the light and heavy tags are identified by on-line sequencing with MS/MS and database search. Pairwise comparison of two peptide peaks that carry the light and the heavy tag allows quantification of the ratio by comparing the relative signal intensities [23].

A closely related technique is the cleavable isotope-coded affinity tagging (cICAT), where the tags contain a cleavage site in the form of a photocleavable linker or an acid-labile linker to remove the biotin moiety, which results in improved MS/MS spectra [24, 25]. Although ICAT and cICAT reduce the complexity of the peptide mixtures to be separated and sequenced, this may lead to insufficient sequence coverage for protein identification. Moreover, ICAT methods are not appropriate for the detection of post-translational modifications (PTM), as the PTM and the Cys residue required for ICAT conjugation do not necessarily reside on the same peptide fragment.

The *isotope-coded protein labelling* (ICPL) approach addresses all amino groups (N-terminus and side chain) at the protein level of two different cellular states (proteomes) with light or heavy ICPL tags. It is compatible with all commonly used protein and peptide separation techniques to reduce complexity and provides high protein sequence information, including PTMs and isoforms [26].

The iTRAQ (*isobaric tag for relative and absolute quantification*) technique also does not suffer from the shortcoming of insufficient sequence coverage. Like ICAT and ICPL it is based on chemically tagging functional groups of the proteins (ICAT: thiols; ICPL: N-terminus or lysine). However, with iTRAQ, not the proteins, but the peptides generated by proteolysis are tagged. Moreover, isobaric tags having the same atomic mass but different arrangements are used. There is no mass difference detectable in the conventional mass spectra, but the tags release characteristic low mass reporter ions upon fragmentation in the MS/MS experiment. This allows comparison and quantification of up to four samples in one experiment. Since each peptide is labelled, proteome coverage is expanded and tracking of PTM is possible [27].

Synthetic peptides have large problem-solving potential in absolute proteomics, as for example shown for the strategy named AQUA (*absolute quantification*), where isotopically labelled synthetic peptides corresponding to the native counterparts formed by proteolysis are employed as an internal standard in the mass spectra. The availability of a pure, well-defined peptide in accurately known amounts is a precondition for AQUA absolute proteomics, as opposed to the relative quantification methods described above.

10.3.2.3 Enzymatic Stable-Isotope Labeling

As the generation of peptides in gel-free proteome analysis requires digestion by enzymes, stable isotopes can also be introduced into the peptide by performing the digestion in H_2^{18}O , which incorporates one or two ^{18}O at the C-terminal carboxyl group of the generated peptides, leading to variability in the MS pattern of the peptides. The mass offset of 2 Da or 4 Da may, however, impose problems of isotopic overlap of the peptide pairs [19, 28].

10.4

Activity-Based Proteomics

While all strategies mentioned so far in this chapter are appropriate to detect changes in protein abundance (*abundance-based proteomics*), it is not possible to analyze the function and activity of the enzymes in a proteome with these methods. Changes in the phenotype of a cell or a tissue or even the transition from a healthy to a pathogenic state may, however, not be due to mere changes in the overall amount of one or more proteins, but may also depend on the fraction of active protein. In order to meet this challenge, chemical methods (*chemical proteomics*, *functional proteomics*) have been developed during the past decade [29–31]. Methods that are able to provide information on the activity state of proteins are now known as *activity-based protein profiling* (ABPP) or *activity-based proteomics* (ABP). The basic concept of chemical proteomics in general is the application of synthetic peptides or small molecules that selectively address the active site of enzymes or enzyme families. Selective molecules that can be tuned in their affinity to different target enzymes are used to generate subproteomes prior to further analysis, which significantly reduces the amount of data. This method is also suitable for the detection of low-abundance or membrane-associated proteins, via affinity enrichment based on biotin tags. Since only active proteins are detected, while inactive proforms and inhibitor bound enzymes remain unlabelled [31], direct assignment of effector proteins responsible for specific biological events is possible.

There are different approaches to meet the requirements of chemical proteomics, depending on the target protein family and the nature of the peptide/small molecule ligand. Protein families, where irreversibly binding ligands are known, can be covalently tagged in a straightforward manner with conjugates containing family-specific irreversible inhibitors (Section 10.4.1). Enzymes where no irreversible inhibitors are available may be covalently tagged by using tailored suicide

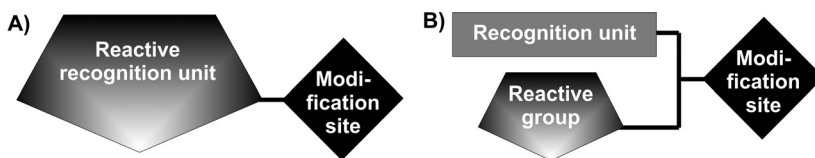


Figure 10.2 Basic concepts for the design of an affinity-based probe in functional proteomics. (A) An irreversible inhibitor (reactive recognition unit) is equipped with a modification site for attachment to a solid surface or a reporter group. (B) A reversibly binding protein ligand, is equipped with a modification site and, if required, with a reactive group to trigger the formation of a covalent bond to the target protein.

substrate conjugates (Section 10.4.1). Protein families, where only reversibly binding ligands are known, may be enriched by using inhibitor affinity chromatography (IAC), or by employing conjugates with photoreactive groups (Section 10.4.2).

Activity-based probes (ABP) basically comprise a *recognition unit* which may be an irreversible inhibitor (Figure 10.2, *reactive recognition unit*) or a reversible inhibitor coupled across a linker to a *modification site* that allows attachment to a solid surface (for affinity purification) or of a *reporter group* (e.g. fluorophor, biotin, radiolabel), which serves for identification.

10.4.1

Irreversibly Binding Affinity-Based Probes

Conjugates of irreversibly binding ligands (inhibitors) with reporter groups have been proven to be valuable molecular tools for functional proteomics. They are suitable to retrieve the members of a class of proteins in a family-specific manner. Hence, they are also called *directed affinity based probes*. Such tools contain a reactive recognition unit (Figure 10.2A), which means that the irreversible inhibitor simultaneously serves as recognition unit and reactive group. Two sub-classes encompassing either reactive mechanism based probes or suicide substrate probes may be distinguished. Reactive mechanism based probes intrinsically display high reactivity towards nucleophiles. Suicide substrate probes are transformed into a highly reactive species by the target enzyme in the course of the enzyme reaction, which then reacts with active site residues. Consequently, a covalent bond to the active site of the enzyme is formed. In both cases the molecular probes exclusively target enzymes and only enzymatically active representatives are being covalently labeled. Such irreversible affinity based probes are usually conjugated across a linker to an appropriate reporter group (fluorophor, biotin, radiolabel) and are suitable for gel-based *activity-based protein profiling*.

Fluorophosphonates **1** (Figure 10.3) address serine proteases [31–33] because they readily react with the catalytic serine residue in the active site. Active site serine residues are present in all types of serine hydrolases that form one of the largest enzyme families present in eukaryotes and include the serine proteases together with

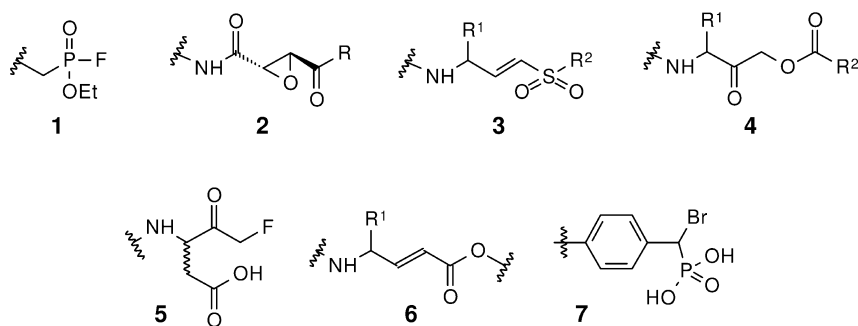


Figure 10.3 Irreversible enzyme inhibitors used in affinity-based protein profiling (ABPP).

lipid hydrolases, esterases and amidases. Serine hydrolases have been estimated to represent approximately 1% of the predicted protein products encoded by the human genome. Consequently, such affinity based probes are of high relevance to functional proteomics, both in plants and animals.

The thiol group of cysteine is a potent nucleophile even under physiological conditions at near-neutral pH. Cysteine proteases are therefore amenable for targeting with electrophilic groups, like epoxy succinyl moieties **2** [34–36], vinyl sulfones **3** [37–40], acyloxymethyl ketones **4** [41–44], fluoromethyl ketones **5** [45], and vinyllogous amino acids **6** (Figure 10.3).

Additional recognition elements present in the affinity-based probe, like for example peptides or peptide-like linkers, confer on the probes improved selectivity towards certain classes of proteases. The epoxy succinyl moiety of type **2** also occurs in the natural products E-64 (**8**) and CA074 (**9**). It is considered a general inhibitor for cysteine proteases. Modified epoxy succinyl-derived conjugates have been employed as affinity labels for the papain class of cysteine proteases [46]. For compound DCG-04 (**10**), a general probe for lysosomal cathepsins, broad structure–activity relationship studies with respect to the S^2/P^2 position (leucine residue) are reported. Moreover, a concept for the specific targeting of single members of the cysteine protease subclass of cathepsins has been developed. The structural and binding characteristics of E-64 (**8**) and CA074 (**9**) were combined in the design of new highly potent and selective chimeric inhibitor tools **11** (Figure 10.4) [34].

Solid phase synthesis of peptidyl vinyl sulfones **3** permits the rapid optimization of linker length and properties [47]. In most strategies for solid phase synthesis of vinyl sulfone type molecular tools the vinyl sulfone part is introduced at a late stage of the synthesis. Attachment of a vinyl sulfone-containing aspartic acid across the side-chain carboxylic acid to Rink resin allowed the generation of a positional scanning library (cf. Section 8.2.4) of peptide vinyl sulfones **3** that subsequently was used to profile substrate and inhibitor selectivity of the catalytic subunits of the proteasome [38]. Side-chain immobilization of a phenolic vinyl sulfone derivative was shown to be possible on 2-chlorotrityl resin. Subsequent chain elongation and finally conjugation with a reporter group, e.g. a fluorescent dye, provides the functional vinyl sulfone after cleavage from the resin [39].

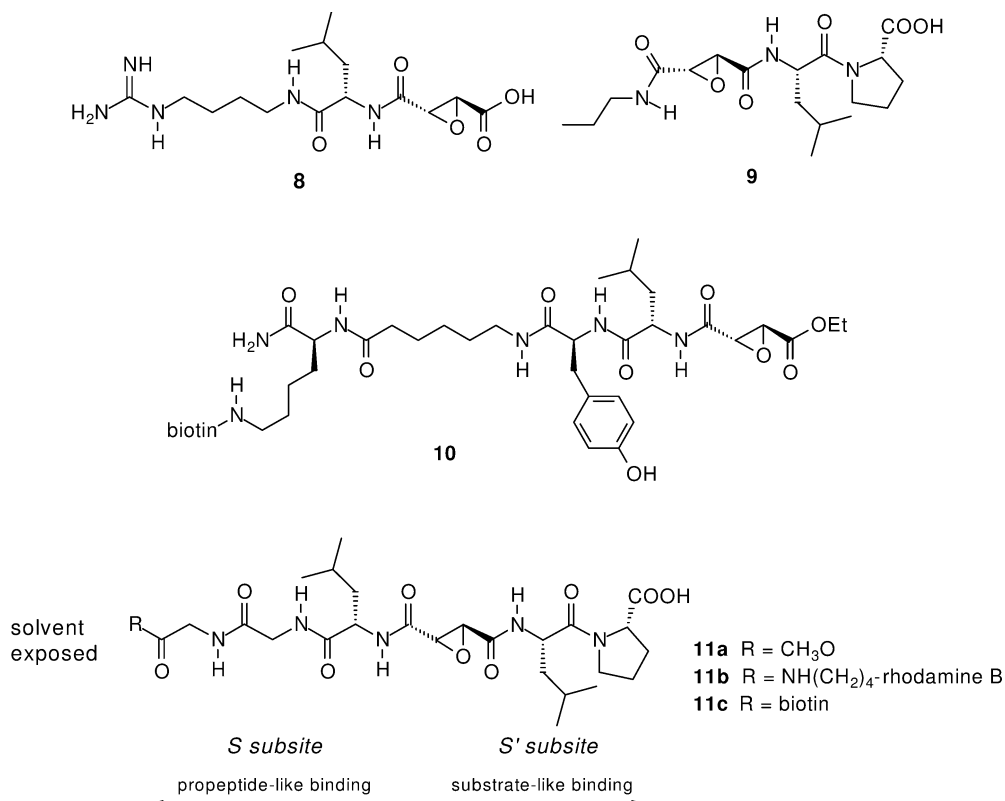


Figure 10.4 Epoxysuccinyl-type affinity-based probes (ABPs) derived from naturally occurring protease inhibitors E-64 and CA074.

Acyloxymethyl ketones **4** [41–44] intrinsically display low reactivity towards weak nucleophiles but readily react with thiol residues in the active site of enzymes. Acyloxymethyl ketones **4** that comprise an aspartate residue close in the sequence and fluoromethyl ketones **5** are reported to display some selectivity towards caspases [41, 45]. Acrylates, and especially substituted acrylates like vinylogous amino acids **6**, are appropriate for targeting cysteine proteases [48]. A number of probes containing either a single amino acid or an extended peptide sequence have been shown to target caspases, legumains, gingipains and cathepsins. However, experiments directed towards caspase subproteome generation suffered from shortcomings: Several currently available activity-based probes have the limitation of a high level of background labeling when applied to crude proteomes. Moreover, not all caspases can be addressed in a similar manner because they differ significantly with respect to the subsite requirements [44].

In a similar fashion as shown for cysteine protease targeting, affinity-based probes directed towards protein tyrosine phosphatases (PTPs) rely on the reaction between an electrophilic probe (e.g. **7**, Figure 10.3) and an active site Cys residue [49, 50].

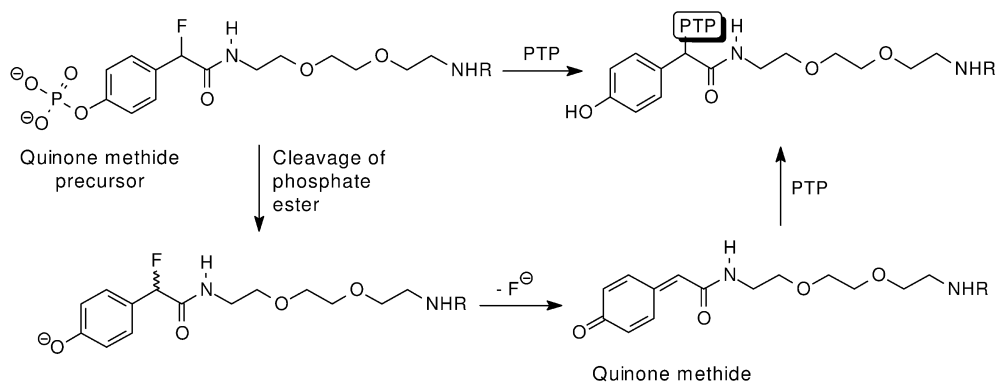


Figure 10.5 Mechanism of quinone methide generation by the action of a protein tyrosine phosphatase on an O-phosphorylated quinone methide precursor.

A relatively broad spectrum of enzymes may be addressed with affinity-based probes of the suicide substrate type. The name suicide substrate originates from the fact that the reactive species that eventually forms a covalent bond to the enzyme is generated in the course of the enzyme reaction itself. In this context, precursors for electrophilic quinone methide intermediates have found widespread application in the irreversible tagging of protein families.

Affinity-based probes directed towards protein tyrosine phosphatases (PTPs) rely on the reaction of an electrophilic probe with incorporation of a quinone methide based phosphatase suicide inhibitor (Figure 10.5). The protein tyrosine phosphatase hydrolyzes the phenyl phosphate, releasing a phenoxide ion that contains an additional leaving group (e.g. fluoride) for example in the benzyl position. Fluoride elimination results in the formation of a quinone methide, which is an electrophilic species and reacts with nucleophiles present in the active site [51, 52]. However, the reaction of the quinone methide with nucleophiles is rather slow, which permits diffusion of the reactive species out of the active site. Consequently, these affinity-based probes must be used in a rather high concentration and their application in complex proteomes remains to be proven.

The general principle of cleavage of a specific group (phosphate ester, sulfate ester, O-phenyl glucoside, or peptide) by enzymatic hydrolysis followed by elimination to create a quinone methide has been proven amenable not only for phosphatases (Figures 10.5 and 10.6 12). Molecular tools for targeting sulfatases (with the aryl O-sulfate **13** [53], glycosidases (with the aryl O-glucoside **14**) [54] and proteases (with the peptidylamide **15**) [55] (Figure 10.6) have been developed.

10.4.2

Reversibly Binding Affinity-Based Probes

Irreversible inhibitors for many classes of enzymes are not known and many other proteins of interest in a physiological or pathological context do not display enzymatic

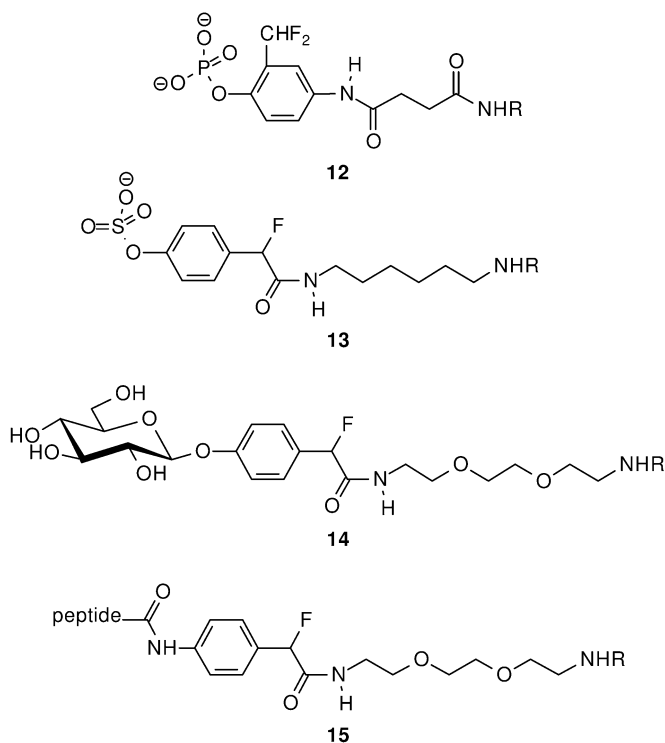


Figure 10.6 Suicide substrates of phosphatases, sulfatases, glycosidases, and proteases based on quinone methide intermediates.

activity. Hence, reversibly binding ligands (peptides or small molecules) are valuable tools in proteome analysis. They may be used in enrichment strategies like *inhibitor affinity chromatography* (IAC) or fishing with inhibitors immobilized on magnetic beads. Alternatively, for application in gel-based approaches, the attachment of a reporter group (e.g. fluorophor, biotin or radiolabel) is required.

10.4.2.1 Inhibitor Affinity Chromatography (IAC)

Subsets of the proteome can be enriched by affinity purifications on special columns or on magnetic beads and subsequently analyzed using 2D-GE and mass spectrometry. This approach is suited for the pulldown of all proteins that bind to a certain ligand (Figure 10.7).

Enrichment of the target proteins by inhibitor affinity chromatography (IAC) is a viable alternative for the majority of proteins, whenever inhibitors or irreversibly binding ligands are not known. For that purpose the inhibitor, or generally the protein ligand, is chemically modified in such a way that it can be immobilized to a solid surface (affinity chromatography matrix, magnetic beads, or surface plasmon resonance sensor chips).

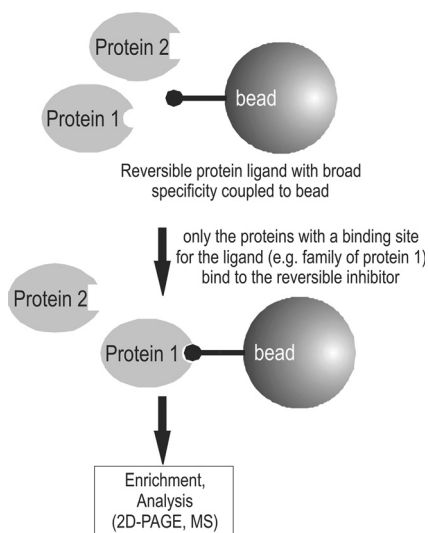


Figure 10.7 Enrichment of protein families on a mechanistic basis with reversibly binding protein ligands attached to affinity chromatography material or magnetic beads.

This has been successfully proven by the creation of metalloprotease [56–58] or kinase sub-proteomes [59–63]. For optimization of the elution conditions in the IAC process, surface plasmon resonance (SPR) represents an excellent method [64].

10.4.2.2 Labelling Strategies with Reversibly Binding Protein Ligands

Gel-based approaches for subproteome analysis with reversibly binding protein ligands require covalent crosslinking between the probe and the target proteins, because a non-covalently associated protein–ligand complex would dissociate under the denaturing conditions of 2D-GE. Consequently, an additional chemical step is necessary to obtain such a covalent bond. Novel engineered synthetic molecular tools are required that are able both to address the target protein family on a common mechanistic basis and to undergo a cross-linking reaction upon external triggering. Such a cross-linking reaction can be brought about photochemically. Photoaffinity labeling is a well-established method e.g. for elucidation of ligand binding sites in biochemistry [65]. There are different photoreactive groups (benzophenones [65, 66], 2*H*-azirines [67], aryl azides) available that can be incorporated into affinity-based probes.

The incubation of a proteome with an affinity-based probe comprising a reversible protein ligand (recognition unit), a photoreactive moiety, and a reporter group, followed by photochemical cross-linking has proven amenable for retrieving proteins that bind to the recognition unit of the probe. This concept is also suitable for the discovery of new members of a protein family. Hence, a subproteome is labelled with high sensitivity and can then be subjected to further analysis.

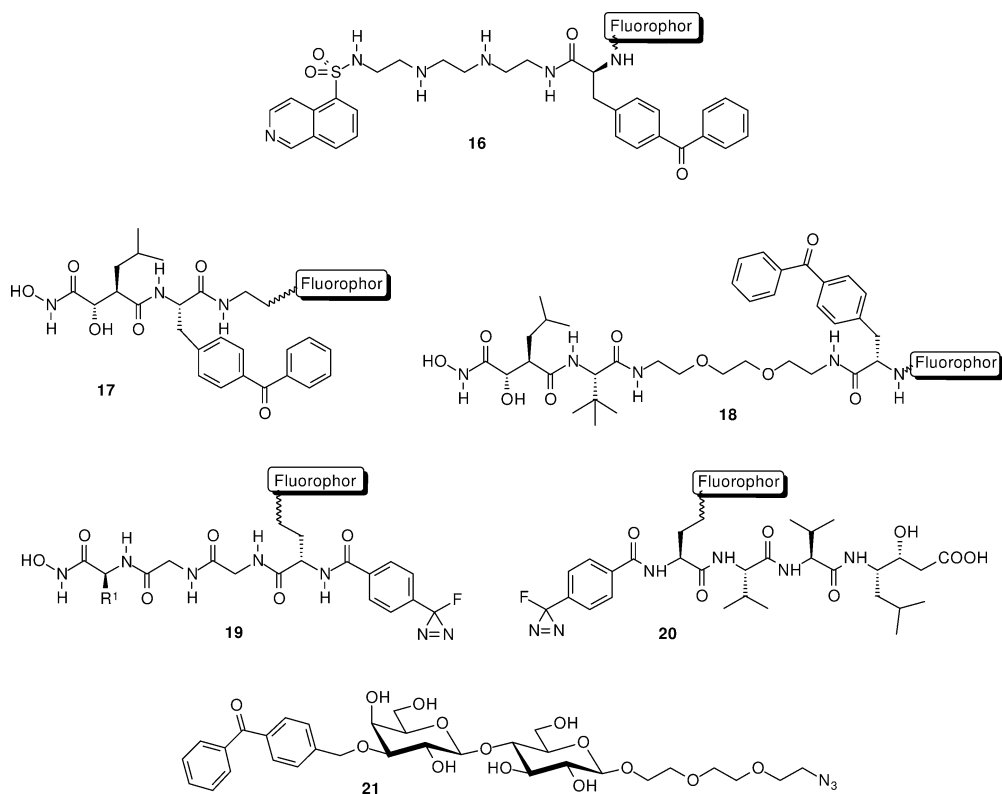


Figure 10.8 Affinity-based probes (ABPs) based on reversible inhibitors/ligands.

Although false positive results cannot be excluded, this method is appropriate to significantly reduce the number of proteins in a complex proteome. The basic concept was pioneered by Hagenstein et al. [68] for kinase subproteome tagging with the conjugate **16** (Figure 10.8) composed of a broad spectrum kinase inhibitor (isoquinoline sulfonamide), *p*-benzoylphenylalanine (Bpa) as the photoreactive group and carboxyfluorescein as the fluorophore. It has subsequently been adapted and validated for metalloproteases with **17–19** [58, 66, 67] by using zinc-chelating hydroxamate inhibitors that were conjugated to a photoreactive group and different fluorophors.

In addition, the approach was also successfully used for tagging aspartic proteases with **20** [69] and carbohydrate binding proteins (lectins) with **21** [70]. In the latter case a sugar moiety employed as the reversible ligand was not attached directly to a reporter group, instead, an azide moiety was incorporated. This chemical functionality subsequently allows in a completely orthogonal way the attachment of reporter groups in the frame of a Cu^I catalyzed 1,3-dipolar cycloaddition (*click chemistry*). Such an approach was also used for subsequent fluorophore attachment after metalloprotease subproteome generation [66].

10.5

Review Questions

- Q10.1. What is the proteome?
- Q10.2. Can an organism or a cell have two or more different proteomes?
- Q10.3. How many different proteins may be produced by a eukaryotic cell at a certain point of time?
- Q10.4. What problem is imposed on proteome analysis by the dynamic range of protein concentrations?
- Q10.5. What is meant by depletion? Why is it sometimes necessary in proteome analysis?
- Q10.6. Explain the differences between the 2D-GE/peptide mass fingerprint approach and the MudPIT approach.
- Q10.7. Why is it necessary to have the genome of the organism under investigation sequenced and annotated for proteome analysis using peptide mass fingerprint?
- Q10.8. What information is available from MSⁿ experiments?
- Q10.9. How is a quantitative comparison between two proteomic states possible?
- Q10.10. Explain the difference between the ICAT and the iTRAQ approach.
- Q10.11. Why is activity-based proteomics important?
- Q10.12. Compare the different molecular tools that may be used in activity-based proteomics.
- Q10.13. Give some examples of molecular tools used in activity-based proteomics, which involve either irreversible or reversible inhibitors.
- Q10.14. How can suicide substrates be employed in activity-based proteomics?
- Q10.15. Why is it necessary to incorporate photoreactive groups into probes for some families of proteins?

References

- 1 V. Wasinger, S. Cordwell, A. Cerpa-Poljak, J. Yan, A. Gooley, M. Wilkins, M. Duncan, R. Harris, K. Williams, I. Humphery-Smith, *Electrophoresis* **1995**, *16*, 1090.
- 2 N.L. Anderson, N.G. Anderson, *Mol. Cell. Proteomics* **2002**, *1*, 845.
- 3 C. O'Donovan, R. Apweiler, A. Bairoch, *Trends Biotechnol.* **2001**, *19*, 178.
- 4 U. Kruse, M. Bantscheff, G. Drewes, C. Hopf, *Mol. Cell. Proteomics* **2008**, *7*, 1887.
- 5 J.D. Wulfschuhle, L.A. Liotta, E.F. Petricoin, *Nat. Rev. Cancer* **2003**, *3*, 267.
- 6 J.M. Jacobs, J.N. Adkins, W.-J. Qian, T. Liu, Y. Shen, D.G. CampII, R.D. Smith, J. Proteome Res. **2005**, *4*, 1073.
- 7 W.-C. Lee, K.H. Lee, *Anal. Biochem.* **2004**, *324*, 1.
- 8 P.G. Righetti, E. Boschetti, L. Lomas, A. Citterio, *Proteomics* **2006**, *6*, 3980.
- 9 A. Görg, W. Weiss, M.J. Dunn, *Proteomics* **2004**, *4*, 3665.
- 10 V. Santoni, M. Molloy, T. Rabilloud, *Electrophoresis* **2000**, *21*, 1054.
- 11 S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9390.
- 12 T. Rabilloud (Ed.), *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods*, Springer, Berlin, Heidelberg, 2000.

- 13 M. Unlu, M.E. Morgan, J.S. Minden, *Electrophoresis* **1997**, 18, 2071.
- 14 S. Gharbi, P. Gaffney, A. Yang, M.J. Zvelebil, R. Cramer, M.D. Waterfield, J.F. Timms, *Mol. Cell. Proteomics* **2002**, 1, 91.
- 15 M.P. Washburn, D. Wolters, J.R. Yates III, *Nat. Biotechnol.* **2001**, 19, 242.
- 16 R. Aebersold, M. Mann, *Nature* **2003**, 422, 198.
- 17 J.G. Rohrbough, L. Brechi, N. Merchant, S. Miller, P.A. Haynes, *J. Biomol. Techniques* **2006**, 17, 327.
- 18 M.P. Washburn, D. Wolters, J.R. Yates III, *Nat. Biotechnol.* **2001**, 19, 242.
- 19 A. Panchaud, M. Affolter, P. Moreillon, M. Kussmann, *J. Proteomics* **2008**, 71, 19.
- 20 S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, *Mol. Cell. Proteomics* **2002**, 1, 376.
- 21 S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, *Nat. Biotechnol.* **1999**, 17, 994.
- 22 S.P. Gygi, B. Rist, T.J. Griffin, J. Eng, R. Aebersold, *J. Proteome Res.* **2002**, 1, 47.
- 23 F. Tureček, *J. Mass Spectrom.* **2002**, 37, 1.
- 24 K.C. Hansen, G. Schmitt-Ulms, R.J. Chalkley, J. Hirsch, M.A. Baldwin, A.L. Burlingame, *Mol. Cell. Proteomics* **2003**, 2, 299.
- 25 J. Li, H. Steen, S.P. Gygi, *Mol. Cell. Proteomics* **2003**, 2, 1198.
- 26 A. Schmidt, J. Kellermann, F. Lottspeich, *Proteomics* **2005**, 5, 4.
- 27 P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlett-Jones, F. He, A. Jacobson, D.J. Pappin, *Mol. Cell. Proteomics* **2004**, 3, 1154.
- 28 O.A. Mirgorodskaya, Y.P. Kozmin, M.I. Titov, R. Korner, C.P. Sonksen, P. Roepstorff, *Rapid Commun. Mass Spectrom.* **2000**, 14, 1226.
- 29 D.A. Jeffrey, M. Bogyo, *Curr. Opin. Biotech.* **2003**, 14, 87.
- 30 M.C. Hagenstein, N. Sewald, *J. Biotechnol.* **2006**, 124, 56.
- 31 M.J. Evans, B.F. Cravatt, *Chem. Rev.* **2006**, 106, 3279.
- 32 Y. Liu, M.P. Patricelli, B.F. Cravatt, *Proc. Natl. Acad. Sci. USA* **1999**, 96, 14694.
- 33 D. Kidd, Y. Liu, B.F. Cravatt, *Biochemistry* **2001**, 40, 4005.
- 34 N. Schaschke, I. Assfalg-Machleidt, W. Machleidt, L. Moroder, *FEBS Lett.* **1998**, 421, 80.
- 35 N. Schaschke, I. Assfalg-Machleidt, T. Laßleben, C.P. Sommerhoff, L. Moroder, W. Machleidt, *FEBS Lett.* **2000**, 482, 91.
- 36 D. Greenbaum, A. Baruch, L. Hayrapetian, Z. Darula, A. Burlingame, K.F. Medzihradsky, M. Bogyo, *Mol. Cell. Proteomics* **2002**, 1, 60.
- 37 M. Bogyo, S. Verhelst, V. Bellingard-Dubouchaud, S. Toba, D. Greenbaum, *Chem. Biol.* **2000**, 7, 27.
- 38 T. Nazif, M. Bogyo, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 2967.
- 39 G. Wang, U. Mahesh, G.Y.J. Chen, S.Q. Yao, *Org. Lett.* **2003**, 5, 737.
- 40 H. Ovaa, P.F. van Swieten, B.M. Kessler, M.A. Leeuwenburgh, E. Fiebigler, A.M.C.H. van der Nieuwendijk, P.J. Galaray, G.A.V.D. Marel, H.L. Ploegh, H.S. Overkleeft, *Angew. Chem. Int. Ed.* **2003**, 42, 3626.
- 41 N.A. Thornberry, E.P. Peterson, J.J. Zhao, A.D. Howard, P.R. Griffin, K.T. Chapman, *Biochemistry* **1994**, 33, 3934.
- 42 L. Faleiro, R. Kobayashi, H. Fearnhead, Y. Lazebnik, *EMBO J.* **1997**, 16, 2271.
- 43 L.M. Martins, T. Kottke, P.W. Mesner, G.S. Basi, S. Sinha, N. Frigon, E. Tatar, J.S. Tung, K. Bryant, A. Takahashi, P.A. Svingen, B.J. Madden, D.J. McCormick, W.C. Earnshaw, S.H. Kaufmann, *J. Biol. Chem.* **1997**, 272, 7421.
- 44 D. Kato, K.M. Boatright, A.B. Berger, T. Nazif, G. Blum, C. Ryan, K.A.H. Chehade, G.S. Salvesen, M. Bogyo, *Nature Chem. Biol.* **2005**, 1, 33.
- 45 M.-L. Liao, R.C. Panicker, S.Q. Yao, *Tetrahedron Lett.* **2003**, 44, 1043.
- 46 D. Greenbaum, K.F. Medzihradsky, A. Burlingame, M. Bogyo, *Chem. Biol.* **2000**, 7, 569.
- 47 H.S. Overkleeft, P.R. Bos, B.G. Hekking, E.J. Gordon, H.L. Ploegh, B.M. Kessler, *Tetrahedron Lett.* **2000**, 41, 6005.

- 48 N. Winssinger, S. Ficarro, P.G. Schultz, J.L. Harris, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 11139.
- 49 S. Kumar, B. Zhou, F. Liang, W.-Q. Wang, Z. Huang, Z.-Y. Zhang, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 7943.
- 50 S. Kumar, B. Zhou, F. Liang, H. Yang, W.-Q. Wang, Z.-Y. Zhang, *J. Proteome Res.* **2006**, 5, 1898.
- 51 Q. Zhu, X. Huang, G.Y.J. Chen, S.Q. Yao, *Tetrahedron Lett.* **2003**, 44, 2669.
- 52 L.-C. Lo, T.-L. Pang, C.-H. Kuo, Y.-L. Chiang, H.-Y. Wang, J.-J. Lin, *J. Proteome Res.* **2002**, 1, 35.
- 53 C.-P. Lu, C.-T. Ren, S.-H. Wu, C.-Y. Chu, L.-C. Lo, *ChemBioChem* **2007**, 8, 2187.
- 54 C.-S. Tsai, Y.-K. Li, L.-C. Lo, *Org. Lett.* **2002**, 4, 3607.
- 55 Q. Zhu, A. Girish, S. Chattopadhyaya, S.Q. Yao, *Chem. Commun.* **2004**, 13, 1512.
- 56 J.R. Freije, R. Bischoff, *J. Chromatogr. A* **2003**, 1009, 155.
- 57 J.R. Freije, T. Klein, J.A. Ooms, J.P. Franke, R. Bischoff, *J. Proteome Res.* **2006**, 5, 1186.
- 58 M.I. Collet, J. Lenger, K. Jenssen, H.P. Plattner, N. Sewald, *J. Biotechnol.* **2007**, 129, 316.
- 59 G. Lolli, F. Thaler, B. Valsasina, F. Roletto, S. Knapp, M. Uggeri, A. Bachi, V. Matafora, P. Storici, A. Stewart, H.M. Kalisz, A. Isacchi, *Proteomics* **2003**, 3, 1287.
- 60 H. Daub, K. Godl, B. Klebl, G. Müller, *Assay Drug Dev. Technol.* **2004**, 2, 215.
- 61 J. Wissing, K. Godl, D. Brehmer, S. Blencke, M. Weber, P. Habenberger, M. Stein-Gerlach, A. Missio, M. Cotten, S. Müller, H. Daub, *Mol. Cell. Proteomics* **2004**, 3, 1181.
- 62 D. Brehmer, K. Godl, B. Zech, J. Wissing, H. Daub, *Mol. Cell. Proteomics* **2004**, 3, 490.
- 63 D. Brehmer, Z. Greff, K. Godl, S. Blencke, A. Kurtenbach, M. Weber, S. Müller, B. Klebl, M. Cotten, G. Kéri, J. Wissing, H. Daub, *Cancer Res.* **2005**, 65, 379.
- 64 K. Jenssen, K. Sewald, N. Sewald, *Bioconj. Chem.* **2004**, 15, 594.
- 65 G. Dorman, G.D. Prestwich, *Trends Biotechnol.* **2000**, 18, 64.
- 66 A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey, B.F. Cravatt, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 10000.
- 67 E.W.S. Chan, S. Chattopadhyaya, R.C. Panicker, X. Huang, S.Q. Yao, *J. Am. Chem. Soc.* **2004**, 126, 14435.
- 68 M.C. Hagenstein, J.H. Mussgnug, K. Lotte, R. Plessow, A. Brockhinke, O. Kruse, N. Sewald, *Angew. Chem. Int. Ed.* **2003**, 42, 5635.
- 69 S. Chattopadhyaya, E.W.S. Chan, S.Q. Yao, *Tetrahedron Lett.* **2005**, 46, 4053.
- 70 L. Ballell, K.J. Alink, M. Slijper, C. Versluis, R.M.J. Liskamp, R.J. Pieters, *ChemBioChem* **2005**, 6, 291.

Glossary

- A β** amyloid- β .
A₂bu 2,4-diaminobutyric acid.
aa amino acid.
AA antamanide (anti-amanita peptide).
Aad 2-aminoadipic acid.
 β -Aad 3-aminoadipic acid.
AAP antimicrobial animal peptides.
aatRS amino acyl tRNA synthetase.
Ab antibody.
ABPP activity-based protein profiling.
ABP activity-based proteomics or activity-based probes.
Abu α -aminobutyric acid.
Abz aminobenzoic acid.
Abzyme *catalytic antibody*, a monoclonal antibody with catalytic activity.
Ac acetyl.
ACE 2 angiotensin-converting enzyme 2.
ACE angiotensin-converting enzyme.
ACHR acetylcholine receptor.
Acm acetamidomethyl.
ACP acyl carrier protein.
ACTH corticotropin.
Active pharmaceutical ingredients
(API) a term for pharmaceutical market products considering peptides not only as hormones, as in the past, but as a major class of pharmaceutical products in antibiotics, antiviral and other therapeutic areas.
AD Alzheimer's disease.
Ada adamantyl.
ADME abbreviation of the pharmacological parameters absorption, delivery, metabolism, excretion.
Adenohypophysis the anterior glandular lobe of the pituitary gland that constitutes of functional unit for the control of the secretion of many hormones, including ACTH, prolactin and growth hormone.
Adoc adamantyloxycarbonyl.
ADP adenosine diphosphate.
Aet aminoethyl.
Ag antigen.
AGaloc tetra-*O*-acetyl- β -D-galactopyranosyloxycarbonyl.
AGE advanced glycation end products.
AGloc tetra-*O*-acetyl-D-glucopyranosyloxycarbonyl.
Agonist a ligand (e.g. hormone) of a receptor that elicits the same response as the native ligand.
AGRP agouti-related protein.
Ahx 2-aminohexanoic acid (norleucine).
 ϵ -Ahx 6-aminohexanoic acid.
AHZ β -alanyl-histidinato zinc.
Aib α -aminoisobutyric acid.
AIDS acquired immunodeficiency syndrome.
Alanine scan systematic substitution of each amino acid residue of a native peptide by a simple amino acid such as alanine. A first step in structure–activity relationship studies.
alle allo-isoleucine (2*S*,3*R* in the L-series).
Alzheimer's disease (AD) the most prominent severe dementia in the elderly population, first described by Alzheimer in 1907. AD is a widespread, neurodegenerative, dementia-inducing disorder characterized mainly by amyloid deposits surrounding dying neurons (senile plaques), neurofibrillar degeneration with tangles, and cerebrovascular angiopathies. AD is clinically characterized by a progressive

- loss of cognitive abilities, progressive memory and intellectual deficits. Amyloid- β and tau protein are responsible for the formation of the plaques and tangles of AD. The mechanism of neurodegeneration caused by amyloid- β in AD is controversial.
- AKH** adipokinetic hormone.
- All** allyl (used only in 3-letter code names).
- Ala** alanine.
- β -Ala** β -alanine.
- Alloc** allyloxycarbonyl.
- Allocam** allyloxycarbonylaminomethyl.
- Allom** allyloxymethyl.
- Aminoxy peptides** correlates of β -peptides composed of α -aminoxy acids as analogues of β -amino acids with replacement of C^β by an oxygen atom.
- AMP** acronym for (1) antimicrobial peptides, or (2) adenosine monophosphate.
- AM-PS** aminomethyl polystyrene.
- ANF** atrial natriuretic factor.
- Anchoring group** *linker*, a moiety bound to the polymeric support for the attachment of the first amino acid in SPPS.
- Angiogenesis** the sprouting of new blood vessels from pre-existing ones. It is a necessary process in normal physiology and can be found both during development and in the adult.
- ANP** atrial natriuretic peptide.
- Ans** anthracene-9-sulfonyl.
- Antagonist** an analogue of a biologically active peptide acting as a competitive inhibitor. It occupies the appropriate receptor and displaces the $\rightarrow$ agonist from the receptor, but does not transmit the biological signal.
- Antibody** a protein produced by B lymphocytes or B cells responsible for humoral immunity. An enormously diverse collection of related proteins mediates humoral immunity that is most effective against bacterial infections and the extracellular phases of viral infections.
- Antigen** a foreign macromolecule, predominantly a protein, carbohydrate or nucleic acid, that triggers the immune response, usually performed by production of defense proteins, known as antibodies.
- Antifreeze proteins (AFPs)** proteins synthesized from teleost fish that encounter extreme cold seawater conditions for protection against freezing.
- Antisense peptide** *complementary peptide*, a peptide sequence hypothetically deduced from the nucleotide sequence that is complementary to the nucleotide sequence coding for a naturally occurring peptide (sense sequence).
- Anxiety peptide** a peptide interacting with the benzodiazepine receptor causing increased anxiety.
- Aoc** 1-azabicyclo[3.3.0]octane-2-carboxylic acid.
- Aoe** (S)-2-amino-8-oxo-(S)-9,10-epoxidecanoic acid.
- AOP** 7-azabenzotriazol-1-yloxytris(dimethylamino)phosphonium hexafluorophosphate
- APA** Australian Peptide Association.
- Apa** 6-aminopenicillanic acid.
- API** active pharmaceutical ingredient.
- Apm** 2-aminopimelic acid.
- APM** aspartame.
- APS** American Peptide Society.
- AQUA** absolute quantification
- AQP** aquaporin.
- Ar** aryl.
- Arg** arginine.
- Asn** asparagine.
- Asp** aspartic acid.
- Asu** aminosuberic acid.
- AT** angiotensin.
- At** azabenzotriazolyl.
- ATP** adenosine triphosphate.
- AVP** [8-arginine]vasopressin.
- Azadepsipeptide** a pseudopeptide in which the α -carbon atom in a depsipeptide is replaced isoelectronically by a trivalent nitrogen.
- Azapeptide** a class of backbone-modified peptides in which the α -CH of one or more amino acid residues in the peptide chain is isoelectronically replaced by a trivalent nitrogen atom.
- Azatide** a biopolymer mimetic consisting of α -aza-amino acids (hydrazine carboxylates).
- Azoc** (4-phenylazophenyl) isopropoxycarbonyl.
- B cells** *B lymphocytes*, cells playing an important role in the humoral immune system being specialized to secrete large amounts of antigen-specific antibodies.
- Bac5** bactericin.
- BAL** backbone amide linker.
- BBB** blood-brain barrier.

- Bet** α -betainyl.
- BGloc** tetrabenzylglucosyloxycarbonyl.
- BGP** bone Gla protein.
- BHA** benzhydrylamine.
- Bic** benzisoxazol-5-yloxycarbonyl.
- Bip** biphenyl-4-sulfonyl.
- BK** bradykinin.
- Blood-brain barrier (BBB)** one barrier separating the central nervous system (CNS) from the periphery, located at the endothelial cells of the brain tissue capillaries besides the blood–cerebrospinal fluid barrier (B–CSF–B) at the choroid plexus and the circumventricular organs.
- BLP** bombinin-like peptide.
- BMP** brain morphogenetic protein.
- BNP** brain natriuretic peptide.
- Boc** *tert*-butoxycarbonyl.
- BOI** 2-[(1*H*-benzotriazol-1-yl)oxy]-1,3-dimethylimidazolidinium hexafluorophosphate.
- Bom** benzylloxymethyl.
- BOP** benzotriazol-1-yloxytris(dimethylamino)phosphonium hexafluorophosphate.
- Bpa** *p*-benzoylphenylalanine.
- Bpoc** 2-(4-biphenyl)isopropoxycarbonyl.
- BPTI** basic pancreatic trypsin inhibitor.
- BroP** bromotris(dimethylamino)phosphonium hexafluorophosphate.
- BrPhF** 9-(4-bromophenyl)-9-fluorenyl.
- BSA** bovine serum albumin.
- Bsmoc** 1,1-dioxobenzo[b]thiophen-2-ylmethoxycarbonyl.
- Bspoc** 2-(*tert*-butylsulfonyl)-2-propenyloxycarbonyl.
- Bt** benzotriazolyl.
- Btb** 1-*tert*-butoxycarbonyl-2,3,4,5-tetrachlorobenzoyl.
- Btm** benzylthiomethyl.
- BTU** *O*-benzotriazolyl-*N,N,N',N'*-tetramethyluronium hexafluorophosphate.
- iBu** isobutyl.
- Bum** *tert*-butoxymethyl.
- tBu** *tert*-butyl.
- Bz** benzoyl.
- Bzl(4-Me)** 4-methylbenzyl.
- Bzl** benzyl (Bn in contemporary organic synthesis).
- CADD** computer-aided drug design.
- CaM** calmodulin.
- Cam** carboxamidomethyl.
- CAMD** computer-aided molecular design.
- CAMM** computer-assisted molecular modeling.
- cAMP** cyclic adenosine 3',5'-monophosphate.
- CBD** chitin binding domain.
- CCAP** crustacean cardioactive peptide.
- CCK** cholecystokinin.
- CD** circular dichroism.
- CDI** carbonyldiimidazole.
- cDNA** complementary DNA.
- CE** capillary electrophoresis.
- Central nervous system (CNS)** part of the vertebrate nervous system consisting of the brain (in vertebrates that have a brain), and the spinal cord. It contains the majority of the nervous system and plays a fundamental role, together with the peripheral nervous system, in the control of behavior.
- CF₃-BOP** 6-(trifluoromethyl)benzotriazol-1-yloxytris(dimethylamino)phosphonium hexafluorophosphate.
- CF₃-HBTU** 2-[6-(trifluoromethyl)benzotriazol-1-yl]-1,1,3,3-tetramethyluronium hexafluorophosphate.
- CF₃-PyBOP** 6-(trifluoromethyl)benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate.
- CG** chorionic gonadotropin.
- Cg** chromogranin.
- cCMP** acronym for (1) current Good Manufacturing Practice, or (2) cyclic guanosine 3', 5'-monophosphate.
- CGRP** calcitonin gene related peptide.
- Cha** β -cyclohexylalanine or cyclohexylammonium salt.
- cHp** cycloheptyl.
- CID** collision-induced dissociation.
- CIEF** capillary isoelectric focusing.
- Cit** citrulline.
- CJD** Creutzfeldt–Jakob disease.
- Cl-HOBt** 1-hydroxy-5-chloro-benzotriazol.
- CLIP** corticotropin-like intermediate lobe peptide.
- Clt** 2-chlorotriethylchloride.
- Cm** carboxymethyl.
- CM** casomorphin or chorionic mammotropin.
- CN** calcineurin.
- CNBr** cyanogen bromide.
- Cne** 2-cyanoethyl.
- CNP** C-type natriuretic peptide.
- CNS** central nervous system.
- cOc** cyclooctyl.

Colony-stimulating factors

(CSF) *hematopoietic growth factors*, glycoprotein growth factors that are involved in proliferation, differentiation and survival of hematopoietic progenitor cells.

COSY correlated NMR spectroscopy.

CP carboxypeptidase.

Cpa 4-chlorophenylalanine.

cPe cyclopentyl.

CPP cell-penetrating peptide.

CPS acronym for (1) convergent peptide synthesis, or (2) Chinese Peptide Society.

Creutzfeldt-Jakob disease (CJD) a sporadically occurring human transmissible spongiform encephalopathy (TSE) characterized by fatal cerebral disorder and apparently caused by prion protein.

CRF corticotropin releasing factor.

CRIF corticotropin release-inhibiting factor.

CRL cerulein.

CsA cyclosporin A.

CSF colony stimulating factor.

CSPPS convergent solid-phase peptide synthesis.

CST cortistatin.

CT calcitonin.

Cy cyclohexyl.

Cya cysteic acid.

Cyoc 2-cyano-*tert*.-butoxycarbonyl.

Cyp cyclophilin.

Cys cysteine.

CZE capillary zone electrophoresis.

Dab α,γ -diaminobutyric acid.

DAG diacylglycerol.

DAST *N,N*-diethylaminosulfur trifluoride.

DBIP diazepam-binding inhibitor peptide.

DBU 1,8-diazabicyclo[5.4.0]undec-7-ene.

Dcb 2,6-dichlorobenzyl.

DCC *N,N'*-dicyclohexylcarbodiimide (also DCCI).

DCHA dicyclohexylamine.

Dcha dicyclohexylammonium salt.

DCM dichloromethane.

Dcp acronym for (1) α,α -dicyclopropylglycine, or (2) dipeptidyl carboxypeptidase.

Dcpm dicyclopropylmethyl.

DCU *N,N'*-dicyclohexylurea (also DCHU).

DDAVP [1-desamino,D-Arg⁸]vasopressin.

Dde 1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl.

Ddz 2-(3,5-dimethoxyphenyl) isopropoxycarbonyl.

DEA diethylamine.

DEAE diethylaminoethanol.

Deg C ^{α,α} -diethylglycine.

DEPBT 3-(diethoxyphosphoryloxy)-1,2,3-benzotriazine-4(3*H*)-one.

DEPC diethyl pyrocarbonate.

DFIH 2-fluoro-4,5-dihydro-1,3-dimethyl-1*H*-imidazolium hexafluorophosphate.

Dha α,β -didehydroalanine (more commonly: α,β -dehydroalanine).

Dhbt 3,4-dihydro-4-oxobenzotriazin-3-yl.

Diabetes-associated peptide (DAP) an alternative term for amylin.

Diabetes mellitus a chronic metabolic disease caused by insulin deficiency. This disease results from either insufficient insulin secretion or decreased sensitivity of the insulin receptor in the target cells. Two types of diabetes mellitus are known, *insulin-dependent (type 1) diabetes mellitus (IDDM)*, also termed juvenile-onset diabetes mellitus, caused by a complete deficiency of pancreatic β cells, often strikes suddenly in childhood and requires insulin to sustain life, and *non-insulin-dependent (type II) diabetes mellitus (NIDDM)*, that may be associated with loss of fully active insulin receptors on normally insulin-responsive cells and is strongly correlated with obesity.

DIC *N,N'*-diisopropylcarbodiimide (also DIPCI).

Dio-Fmoc diisooctyl-Fmoc.

DioRASP Diosynth rapid solution synthesis of peptides.

DIPEA diisopropylethylamine (also DIEA).

DKP diketopiperazine.

DLP defensin-like peptides.

Dmab 4-[[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl]amino] benzyl.

DMF *N,N*-dimethylformamide.

Dmh 2,6-dimethylhept-4-yl.

Dmp 2,4-dimethoxyphenyl.

DMSO dimethylsulfoxide.

DNA deoxyribonucleic acid.

Dnp dinitrophenyl.

Dns 5-(dimethylamino)naphthalene-1-sulfonyl (dansyl).

DOPA 3,4-dihydroxyphenylalanine.

DP IV dipeptidyl peptidase IV.

Dpg C ^{α,α} -diphenylglycine.

Dpm diphenylmethyl (also Bzh, benzhydryl).

- DPPA** diphenyl phosphorazidate.
Dpr 2,3-diaminopropionic acid.
DPTU *N,N'*-diphenylthiourea.
DSC di(*N*-succinimidyl)carbonate.
DSIP delta sleep-inducing peptide.
DSK drosulfakinin.
Dsu (2*S*,7*S*)-2,7-diaminosuberic acid.
Dts dithiasuccinoyl.
DTT dithiothreitol.
DVB divinylbenzene.
Dyn dynorphin
- <**E** single-letter code for pGlu.
EC enzyme commission (enzyme nomenclature).
ECE endothelin converting enzyme.
ECGF endothelial cell growth factor.
ED₅₀ median effective dose.
EDC *N*-ethyl-*N'*-(3-dimethylaminopropyl) carbodiimide.
EDF acronym for (1) epidermal growth factor, or (2) erythrocyte differentiation factor.
EDFR epidermal growth factor receptor.
EDT ethanedithiol.
EDTA ethylenediamine tetraacetic acid.
ee enantiomeric excess.
EEKQ ethyl 2-ethoxy-1,2-dihydroquinoline-1-carboxylate.
EF elongation factor.
EGF epidermal growth factor.
EH eclosion hormone.
ELAM endothelial leukocyte adhesion molecule.
ELH egg-laying hormone.
ELISA enzyme-linked immunosorbent assay.
EM electron microscopy.
EMEA European Agencies for the Evaluation of Medicinal Products.
Endoplasmic reticulum (ER) an organelle occurring in all eukaryotic cells that is responsible for synthesis of lipids and proteins and additionally for transport of proteins and carbohydrates to the Golgi apparatus.
ENK enkephalin.
Epigenetics in biology a term that refers to changes in gene expression that are stable between cell divisions, and sometimes between generations, but do not involve changes in the underlying DNA sequence.
EPL expressed protein ligation.
EPO erythropoietin.
Epoc 2-ethynyl-2-propyloxycarbonyl.
- EPS** European Peptide Society.
ER endoplasmic reticulum.
ESI-MS electrospray ionisation mass spectrometry.
ES-MS electrospray mass spectrometry.
ESR electron spin resonance.
ET endothelin.
Et ethyl.
ETH ecdysis-triggering hormone.
Etm ethoxymethyl.
Ets ethylsulfanyl.
- Fa** 3-(2-furyl)acryloyl.
Fab antigen-binding Ig fragment.
FAB-MS fast atom bombardment mass spectrometry.
FACS fluorescence-activated cell sorter.
Farn farnesyl.
FaRP FMRFamide-related peptide.
Fbg fibrinogen.
Fc ferrocenyl.
Fd ferredoxin.
FGF fibroblast growth factor.
FKBP FK506 binding protein.
Fm 9-fluorenylmethyl.
FMDV foot-and-mouth disease virus.
Fmoc 9-fluorenylmethoxycarbonyl.
FN fibronectin.
For formyl.
FPLC fast protein liquid chromatography.
FPP farnesyl pyrophosphate.
FRET fluorescence resonance energy transfer.
FRL formin-related peptide.
FS follistatin.
FSF fibrin-stabilizing factor.
FSH follicle-stimulating hormone.
FTase farnesyltransferase.
- GABA** γ -aminobutyric acid.
Gal galanin.
GC gas chromatography.
gCSF granulocyte colony stimulating factor.
GFC gel filtration chromatography.
GFP green fluorescent protein.
GGTase geranylgeranyltransferase.
GH growth hormone.
GHRH growth hormone releasing hormone.
GHRP growth hormone releasing peptide.
GHS growth hormone secretagogue.
GHS-R growth hormone secretagogue receptor.

- GIP** glucose-dependent insulinotropic polypeptide.
- Gla** 4-carboxyglutamic acid.
- GLC** gas liquid chromatography.
- Glc** glucose, glycosyl.
- GlcNAc** *N*-acetyl- β -glucosamine.
- Gln** glutamine.
- GLP** glucagon-like peptide.
- Glp** pyroglutamic acid (also pGlu and <E).
- Glu** glutamic acid.
- Gly** glycine.
- Glycome** a term coined for the glycan repertoire of an organism.
- Glycomics** a term for the characterization by function and structure of glycans in the studied system.
- Glycopeptide dendrimers** regularly branched structures containing both carbohydrates and peptides.
- GMP** guanosine monophosphate.
- GnIH** gonadotropin-inhibitory hormone.
- GnRH** gonadotropin releasing hormone.
- Golgi apparatus** a cell organel consisting of a stack of membrane-bound cisternae located between the endoplasmic reticulum (ER) and the cell surface mainly devoted to posttranslational processing of proteins.
- GPC** gel permeation chromatography.
- GPCR** G-protein coupled receptor.
- GPh** guanidinophenyl.
- GRF** growth hormone releasing factor.
- Growth-hormone secretagogues (GHS)** molecules having lost opiate activity but postulated as a growth hormone (GH) releaser.
- GRP** gastrin-releasing peptide.
- GRPP** glicentin-related pancreatic peptide.
- GSF** glutathion-S-transferase.
- GSH** glutathione reduced.
- GSSG** glutathione oxidized.
- GT** gastrin.
- GTP** guanosine triphosphate.
- Gva** δ -guanidinovaleric acid.
- h** human.
- HA** head activator.
- HAPipU** *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-bis(pentamethylene)uronium hexafluorophosphate.
- HAPyU** *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-bis(tetramethylene)uronium hexafluorophosphate.
- Harg** homoarginine.
- HATU** *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; correct IUPAC name: 1-[bis-(dimethylamino)methylumyl]-1*H*-1,2,3-triazolo[4,5-*b*]pyridin-3-oxide hexafluorophosphate or 1-[(dimethylamino)-(dimethyliminium)methyl]-1*H*-1,2,3-triazolo[4,5-*b*]pyridin-3-oxide hexafluorophosphate.
- HAV** hepatitis A virus.
- Hb** hemoglobin.
- HBUTU** 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; correct IUPAC name: 3-[Bis (dimethylamino)methylumyl]- 3*H*-benzotriazol-1-oxide hexafluorophosphate.
- HBV** hepatitis B virus.
- Hbz** 2-hydroxybenzyl.
- HCRT** hypocretin.
- HCV** hepatitis C virus.
- Hcys** homocysteine.
- HDL** high density lipoprotein.
- Hep** heptyl.
- Hepes** *N*-(2-hydroxyethyl)piperazine-*N'*-2-ethanesulfonic acid.
- HF** hydrogen fluoride.
- HG** human little-gastrin.
- hGH** human growth hormone.
- Hip** hippuric acid.
- His** histidine.
- Hiv** α -hydroxyisovaleric acid.
- HIV** human immunodeficiency virus.
- HK-1** hemokinin-1.
- Hmb** 2-hydroxy-4-methoxybenzyl.
- HMBA** hydroxymethylbenzoic acid.
- HMPA** 4-hydroxymethyl-3-methoxyphenoxyacetic acid.
- HMPAA** 4-(hydroxymethyl)phenoxyacetic acid.
- HMPT** hexamethylphosphorous triamide (auch HMP or HMPA).
- HN** humanin.
- Hnb** 2-hydroxy-6-nitrobenzyl.
- HOAt** 1-hydroxy-7-aza-1*H*-benzotriazole.
- HOBt** 1-hydroxy-1*H*-benzotriazole.
- Hoc** cyclohexyloxycarbonyl.
- HODhbt** 3,4-dihydroxy-3-hydroxy-4-oxo-1,2,3-benzotriazine.
- HONdc** *N*-hydroxy-5-norbornene-2,3-dicarboximide.
- HOPip** *N*-hydroxypiperidine.
- HOSu** *N*-hydroxysuccinimide.
- HPCE** high-performance capillary electrophoresis.

- HPLC** high-performance liquid chromatography.
- HPSEC** high-performance size exclusion chromatography.
- HSA** human serum albumin.
- Hse** homoserine.
- Hsp** heat shock protein.
- HSPS** high speed peptide synthesis.
- HSV** herpes simplex virus.
- HTRF** homogeneous time-resolved fluorescence.
- HTS** high-throughput screening.
- Hyp** hydroxyproline.
- Hypophysis** *pituitary gland*, a vertebrate endocrine gland located at the base of the brain, and connected to the midbrain by the hypophyseal stalk. The hypophysis consists of the anterior pituitary (AP, adenohypophysis), the middle part, and the posterior pituitary (neurohypophysis).
- Hyv** α -hydroxyisovaleric acid.
- IAC** inhibitor affinity chromatography.
- IAD** isoaspartyl dipeptidase.
- Iboc** isobornyloxycarbonyl.
- IC** inhibitory concentration.
- ICAM** intracellular adhesion molecule.
- ICAT** isotope-coded affinity tag.
- cICAT** cleavable isotope-coded affinity tagging.
- ICPL** isotope-coded protein labeling.
- IEC** ion-exchange chromatography.
- IEF** isoelectric focussing
- IF** initiation factor.
- IFN** interferon.
- Ig** immunoglobulin.
- IGF** insulin-like growth factor.
- IHB** inhibin.
- IIDQ** 1-isobutoxycarbonyl-2-isobutoxy-1,2-dihydroquinoline.
- IL** interleukin.
- Ile** isoleucine.
- im** imidazole.
- in** indole.
- Incretin effect** the augmentation of glucose-stimulated insulin secretion by intestinal derived peptides that are released in the presence of glucose or nutrients in the gut.
- iNoc** isonicotinylloxycarbonyl.
- INSLP** insulin-like peptides.
- IP** inositol phosphate.
- IPC** immobilized pH gradient
- IPL** intein-mediated protein ligation.
- IPNS** isopenicillin N synthase.
- iPr** isopropyl.
- IR** infrared.
- IRaa** internal reference amino acid.
- IS-MS** ion spray mass spectrometry.
- iTRAQ** isobaric tag for relative and absolute quantification
- IU** international unit.
- Iva** isovaline.
- IvDde** 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl.
- JAK proteins (JAKs)** proteins of the Janus kinase (JAK) family which are non-receptor tyrosine kinases.
- JAKs** JAK proteins.
- J chain** joining chain.
- JHBP** juvenile hormone binding protein.
- JPS** Japanese Peptide Society.
- kB** kilo base pair.
- KC** katacalcin.
- KCL** kinetically controlled ligation.
- kDa** kilodalton.
- KGF** keratinocyte growth factor.
- Killer peptide (KP)** an engineered 10-peptide acting as functional mimotope of the microbial yeast killer toxin.
- KKS** kallikrein-kinin system.
- K_M** Michaelis constant.
- KP** killer peptide.
- KTX** kalitoxin.
- LA** α -lactalbumin.
- Lac** lactic acid.
- Lan** lanthionine.
- LAP** leucine aminopeptidase.
- LD₅₀** lethal dose 50%.
- LDL** low density lipoprotein.
- LD-MS** laser desorption mass spectrometry.
- LDTof** laser desorption time-of-flight.
- Leu** leucine.
- LH** luteinizing hormone.
- LHRH** luteinizing hormone releasing hormone.
- Limited proteolysis** proteolytic cleavage of a scissile peptide bond directed and limited to the specificity of the acting peptidase.
- LF** lactoferrin.
- Lfcin** lactoferricin.
- LPH** lipotropic hormone.
- LPPS** liquid-phase peptide synthesis.
- LPS** lipopolysaccharide.
- LSF** lung surfactant factor.

- LVP** lysine vasopressin.
Lys lysine.
- MA** mixed anhydride.
mAb monoclonal antibody.
Mal maleoyl.
MALDI-ToF MS matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.
MARS multiple automatic robotic synthesizer.
MBHA methoxybenzhydramine.
Mbom 4-methoxybenzyloxymethyl.
MBP acronym for (1) maltose binding protein, or (2) myelin basic protein.
Mbs 4-methoxybenzenesulfonyl.
Mcc microcins.
MCDP mast cell degranulating peptide.
MCH melanin-concentrating hormone.
MD molecular dynamics.
MDP maduropeptin.
Me methyl.
'me-too' drug (MTD) a drug with identical indication and similar formulation as another drug previously approved by FDA.
Mee methoxyethoxymethyl.
MeO methoxy.
Mes mesityl.
Met methionine.
MGP matrix Gla protein.
MHC major histocompatibility complex.
MIC minimum inhibitory concentration.
MIF melanotropin release-inhibiting factor.
MIH melanotropin release-inhibiting hormone.
Mio-Fmoc monoisooctyl-Fmoc.
MK midkine.
MMA metamorphisin A.
Mmt 4-methoxytrityl.
Moa 6-methyloctanoic acid.
Mob 4-methoxybenzyl.
MoEt 2-(*N*-morpholino)ethyl.
Mot motilin.
MP myelopeptides.
MPGF major proglucagon fragment.
MPS multiple peptide synthesis.
MRF melanotropin-releasing factor.
MRH melanotropin-releasing hormone.
MRIH melanotropin release-inhibiting hormone.
mRNA messenger ribonucleic acid.
MS mass spectrometry.
MS/MS tandem mass spectrometry.
Msc 2-(methylsulfonyl)ethoxycarbonyl.
MSH melanocyte stimulating hormone (melanotropin).
Mspoc 2-(methylsulfonyl)-3-phenyl-2-propenyloxycarbonyl.
MT metallothioneins.
Mtb 2,4,6-trimethoxybenzenesulfonyl.
MTD 'me-too' drug.
MTLRP motilin-related peptide.
Mtm methylthiomethyl.
Mtr 4-methoxy-2,3,6-trimethylbenzenesulfonyl.
Mts 2,4,6-trimethylbenzenesulfonyl (mesitylsulfonyl).
MudPIT multidimensional protein identification technology.
MVD mouse vas deferens.
Mz 4-(methoxyphenylazo)benzyloxycarbonyl.
NADH nicotinamide adenine dinucleotide (reduced).
NADPH nicotinamide adenine dinucleotide phosphate (reduced).
Nbb 3-nitro-4-bromomethylbenzhydramido.
Nboc 2-nitrobenzyloxycarbonyl.
NBS *N*-bromosuccinimide.
Nbs nitrobenzenesulfonyl.
2-Nbz 2-nitrobenzyl.
Nbz 4-nitrobenzyl.
NC nociceptin.
NCA α -amino acid *N*-carboxy anhydride.
NCL native chemical ligation.
Ncy norcysteine.
Nde 1-(4-nitro-1,3-dioxindan-2-ylidene)ethyl.
Neu neuraminic acid.
NeuNAc *N*-acetylneuraminic acid.
Neurohypophysis the posterior lobe of the pituitary which is anatomically distinct from the adenohypophysis.
Neuropeptidomics a technological approach for detailed analyses of endogenous peptides from the brain. The neuropeptidomics approach is an excellent example of how to further reveal the way neuropeptides are diversified from a single gene, which releases a variety of regulated, biologically active peptides.
Neurotransmitter a substance transmitting nerve impulses across the synapses between certain types of nerve cells.
NGF nerve growth factor.

NIDDM noninsulin-dependent (type II) diabetes mellitus.
NK neurokinin.
Nle norleucine.
NM neuromedin.
NMB neuromedin B.
NMC neuromedin C.
NMDA *N*-methyl-D-aspartate.
NMM *N*-methylmorpholin.
NMR nuclear magnetic resonance.
NMS neuromedin S.
NMU neuromedin U.
NN neuromedin N.
NOE nuclear Overhauser effect.
NOESY nuclear Overhauser enhanced spectroscopy.
NP_γ neuropeptide γ .
Np 4-nitrophenyl.
NP acronym for (1) neurophysin, (2) natriuretic peptides or (3) nitrophenols.
NP_γ neuropeptide γ .
NPF neuropeptide F.
NPFF neuropeptide FF.
NPK neuropeptide K.
Nps 2-nitrophenylsulfenyl.
NPS neuropeptide S.
NPW neuropeptide W.
NPY neuropeptide Y.
Npys 3-nitro-2-pyridinesulfanyl.
NRPS nonribosomal peptide synthesis.
Nsc 2-[(4-nitrophenyl)sulfonyl]ethoxycarbonyl.
NT neurotensin.
NTA *N*-thiocarboxy anhydride.
Nu/Nuc⁻ nucleophile.
Nva norvaline.
Nvoc 6-nitroveratryloxycarbonyl (4,5-dimethoxy-2-nitrobenzyloxycarbonyl).
O2Nbz 2-nitrobenzyl ester.
O2Np 2-nitrophenyl ester.
OAl allyl ester.
OAt 1-hydroxy-7-azabenzotriazyl ester.
OBt 1-hydroxybenzotriazyl ester.
OtBu *tert*-butyl ester.
OBz benzyl ester.
OcHx cyclohexyl ester.
OCM oncostatin M.
OD optical density.
OEt ethyl ester.
OGp 4-guanidinophenyl ester.
OGP osteogenic growth peptide.
OJP ovarian jelly-peptides.
OMe methyl ester.

ON osteonectin.
ONbz 4-nitrobenzyl ester.
ONC onconase.
Oncopeptidomics a term describing the comprehensive multiplexed analysis of endogenous peptides from a biological sample, under defined conditions, to discover a probable valid peptide tumor biomarker.
ONdc 5-norbornene-2,3-dicarboximido ester.
ONp 4-nitrophenyl ester.
OPA 2-phthaldialdehyde.
OPcp pentachlorophenyl ester.
OPfp pentafluorophenyl ester.
OPN osteopontin.
ORD optical rotatory dispersion.
Orn ornithine.
OSM oncostatin M.
OSu *N*-hydroxysuccinimide ester.
OT oxytocin.
OTce 2,2,2-trichloroethyl ester.
OTcp 2,4,5-trichlorophenyl ester.
Ox 1,3-oxazolidine.
Pac phenacyl.
PACAP pituitary adenylate cyclase activating polypeptide.
PAGE polyacrylamide gel electrophoresis.
PAM acronym for (1) 4-(hydroxymethyl)phenylacetamidomethyl or (2) peptidylglycine α -amidating monooxygenase.
PAMP proadrenomedullin N-terminal 20-peptide.
Pancreas a glandular organ of which the bulk is an exocrine gland producing digestive enzymes, whereas a smaller part (1–2%) consists of the islets of Langerhans, an endocrine gland responsible for maintaining energy metabolite homeostasis.
PAOB 4-phenylacetoxymethyl.
PAPS 3'-phosphoadenosine-5'-phosphosulfate.
Pbf 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl.
PCR polymerase chain reaction.
PD-ECGF platelet-derived endothelial cell growth factor.
PDGF platelet-derived growth factor.
PDH pigment dispersing hormone.
PDI protein disulfide isomerase.
PD-MS plasma desorption mass spectrometry.

- Peptaibiome** the entire expression of fungal peptides with the characteristic amino acid Aib comprising peptides with 5 to 21 residues including a C-terminal amino alcohol.
- Peptaibiotics** fungal *peptides* containing Aib exhibiting *antibiotic* or other bioactivities.
- Peptaibols** *peptides* containing Aib and a C-terminal 1,2-amino alcohol comprising the most abundant subgroup of the → peptaibiotics.
- Peptibody** a hybrid of a peptide and an antibody.
- Peptidome** all peptides with their post-translational modifications expressed in a cell, tissue, body fluid, organ or organism. A term derived by analogy with proteome.
- Peptidomics** a technology for comprehensive analysis of the whole peptidome, aimed at thorough visualization and analysis of small endogenous polypeptides in the molecular mass range 1–20 kDa.
- Pdpm** pyridyldiphenylmethyl.
- PEG** polyethylene glycol.
- Pen** penicillamine (β -mercaptovaline. β , β -dimethylcysteine).
- Pf** 9-phenylfluoren-9-yl.
- Pfp** pentafluorophenyl.
- PFPyU** O-pentafluorophenyl-1,1,3,3-bis (tetramethylene)uronium hexafluorophosphate.
- PFTU** O-pentafluorophenyl-1,1,3,3-tetramethyluronium hexafluorophosphate.
- PG** protecting group.
- pGlu** pyroglutamic acid.
- Ph** phenyl.
- Phac** phenylacetyl.
- Phacm** phenylacetamidomethyl.
- Phe** phenylalanine.
- Phg** phenylglycine.
- PHI** peptide histidine isoleucine amide.
- Phlac** phenyllactic acid.
- Phth** phthaloyl.
- pl** isoelectric point.
- Pic** 4-picolyl (pyridyl-4-methyl).
- PIH** prolactin-release-inhibiting hormone.
- Pip** pipercolic acid (piperidine-2-carboxylic acid).
- Pipoc** piperidinyloxycarbonyl.
- PITC** phenyl isothiocyanate.
- Piv** pivaloyl.
- PKA** protein kinase A.
- PKS** phyto-sulfokinin.
- PL** placenta lactogen.
- PLC** phospholipase C.
- pMBzl** 4-methylbenzyl.
- Pmc** 2,2,5,7,8-pentamethylchroman-6-sulfonyl.
- Pme** 2,3,4,5,6-pentamethylbenzenesulfonyl.
- pNA** 4-nitroaniline.
- PNA** peptide nucleic acid.
- Poc** cyclopentyloxycarbonyl.
- POMC** proopiomelanocortin.
- PP** pancreatic polypeptide.
- PPA** Propylphosphonic acid anhydride.
- PPlase** peptidyl prolyl *cis/trans* isomerase.
- PPST** tyrosyl protein sulfotransferase.
- Pr** propyl.
- PRH** prolactin-releasing hormone.
- iPr** *iso*-propyl.
- PRL** prolactin.
- nPr** *n*-propyl.
- Pro** proline.
- Proteome** the entirety of proteins present in the same organism at a certain time under certain conditions. The 'PROTEin complement of the genome'.
- PrRP** prolactin-releasing peptide.
- PSA** preformed symmetrical anhydride.
- PS-SCL** positional scanning synthetic combinatorial libraries.
- PST** pancrastatin.
- Psty** polystyrene.
- PTC** phenylthiocarbamyl.
- PTH** acronym for (1) phenylthiohydantoin, or (2) parathyroid hormone.
- PTHrP** parathyroid hormone-related hormone.
- PTK** protein-tyrosine kinase.
- PTM** post-translational modification.
- Ptmse** 2-phenyl-2-trimethyl-silylethyl.
- PTnm** 3-nitro-1,5-dioxaspiro[5.5]undec-3-ylmethoxycarbonyl.
- PTP** protein-tyrosine phosphatase.
- PyAOP** 7-azabenzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate.
- PyBOP** benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate.
- PyBroP** bromotripyrrolidinophosphonium hexafluorophosphate.
- Pyl** pyrrollysine.
- PYY** peptide tyrosine tyrosine.
- Pz** 4-(phenyldiazenyl)benzyloxycarbonyl.
- QC** glutaminyl cyclase.
- QCl** 5-chloro-8-quinolyl.

- QpTOF** quadrupole/time-of-flight.
- QSAR** quantitative structure–activity relationship.
- RAFT** regioselectively addressable functionalized template.
- RAMP** receptor activity modifying protein.
- Rc** rusticyanin.
- Rd** rubredoxin.
- RELM** resistin-like molecules.
- RER** rough endoplasmic reticulum.
- RET** resonance energy transfer.
- R_f** retention factor (TLC).
- RFaP** RFamide peptides.
- RGD** fibrinogen binding sequence (-Arg-Gly-Asp-).
- RIA** radioimmunoassay.
- RIP** acronym for (1) ras inhibitory peptide or (2) ribosome-inactivating protein.
- RNA** ribonucleic acid.
- RNase** ribonuclease.
- ROE** rotating frame nuclear Overhauser effect.
- ROESY** rotating frame nuclear Overhauser enhanced spectroscopy.
- Rough endoplasmic reticulum (RER)** a large portion of the ER studded with ribosomes and involved in synthesis and transport of proteins.
- RP** reversed phase.
- RPCH** red pigment concentrating hormone.
- RP-HPLC** reversed phase high performance liquid chromatography.
- RTK** ranatachikinin.
- SA** symmetrical anhydride.
- Saa** sugar amino acid.
- SABR** structure-activity bioavailability relationships.
- SAM** S-adenosylmethionine.
- Sar** sarcosine.
- SAR** structure-activity relationship.
- Sarcoplasmic reticulum (SR)** a special type of smooth ER occurring in smooth and striated muscle cells, containing large stores of Ca²⁺ and involved in biosynthesis.
- SASRIN** Super Acid Sensitive ResIN.
- SAXS** small angle X-ray scattering.
- StBu** *tert*-butylsulfanyl.
- SBzl** thiobenzyl (benzylsulfanyl).
- SCAL** safety catch acid-labile linker or safety catch amide linker.
- Scg-MT** schistomyotropin.
- SCL** synthetic combinatorial libraries.
- Scy I** scyliorhinin I.
- SDB** styrene divinylbenzene.
- SDS** sodium dodecylsulfate.
- Sec** selenocysteine.
- SEC** size-exclusion chromatography.
- SER** smooth endoplasmic reticulum.
- Ser** serine.
- Shk** stichodactyla toxin.
- SIH** somatotropin release-inhibiting hormone.
- SILAC** stable isotope labeling by amino acids in cell culture.
- SLC** sublethal concentration.
- Sle^x** sialyl-Lewis^x.
- SM** somatomedin.
- Smooth endoplasmic reticulum (SER)** a multipurpose organelle involved in several metabolic processes such as the synthesis of lipids and steroids, and the metabolism of carbohydrates.
- SMPS** simultaneous multiple peptide synthesis.
- SN** secretoneurin.
- SP** substance P.
- SP5** splenopentin.
- SPCL** synthetic peptide combinatorial library.
- SPPS** solid-phase peptide synthesis.
- SPR** surface plasmon resonance.
- SR** sarcoplasmic reticulum.
- SRH** somatotropin releasing hormone.
- Srt** sortase.
- SRTX** sarafotoxin.
- ssDNA** single-stranded DNA.
- SST** somatostatin.
- Sta** statine, (3S,4S)-4-amino-3-hydroxy-6-methylhexanoic acid.
- STH** somatotropin.
- Suc** succinoyl.
- SVG** sauvagine.
- SZ** Suzukacillin.
- T lymphocytes (T cells)** cells developed in the thymus mediating cellular immunity.
- T20** enfuvirtide (FuzeonTM).
- Tacm** trimethylacetamidomethyl.
- TASP** template-assembled synthetic protein.
- TBAF** tetrabutylammonium fluoride.
- TBTU** 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate.
- Tcboc** 2,2,2-trichloro-*tert*-butoxycarbonyl.
- Tce** 2,2,2-trichloroethyl.

- TCL** thin layer chromatography.
- Tcp** trichlorophenyl.
- TEA** triethylamine.
- Teoc** 2-(trimethylsilyl)ethoxycarbonyl.
- TFA** trifluoroacetic acid.
- Tfa** trifluoroacetyl.
- TFAA** trifluoroacetic anhydride.
- TFE** trifluoroethanol.
- TFFH** tetramethyl fluoroformamidinium hexafluorophosphate.
- Tfm** trifluoromethyl.
- TFMSA** trifluoromethanesulfonic acid.
- TGF** transforming growth factor.
- TH** thymic hormone.
- THF** tetrahydrofuran.
- Thi** β -thienylalanine.
- Thr** threonine.
- Thx** thyroxine.
- Thz** thiazolidine-4-carboxylic acid.
- TIC/Tic** 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid.
- Tip** 2,4,6-triisopropylbenzenesulfonyl.
- Tipseoc** triisopropylsilylethoxycarbonyl.
- TLC** thin layer chromatography.
- TLE** thin layer electrophoresis.
- TM** thrombomodulin.
- Tmb** 2,4,6-trimethoxybenzyl
- TMS** trimethylsilyl.
- TMSBr** trimethylsilyl bromide.
- Tmse** 2-(trimethylsilyl)ethyl.
- Tmz** α -methyl-2,4,5-trimethylbenzoyloxycarbonyl.
- Tn** troponin.
- TNBS** 2,4,6-trinitrobenzenesulfonic acid.
- TNF** tumor necrosis factor.
- TOCSY** total correlation spectroscopy.
- TOF** time-of-flight.
- Tos** tosyl (4-toluenesulfonyl).
- TP** thymopoietin.
- TP5** thymopentin.
- tPA** tissue plasminogen activator.
- TPST** tyrosine protein sulfotransferase.
- t_R** retention time.
- TROSY** transferred rotational correlated NMR spectroscopy.
- TRF** time-resolved fluorescence.
- TRH** thyrotropin releasing hormone.
- Tris** tris(hydroxymethyl)aminomethane.
- tRNA** transfer RNA.
- TRNOE** transferred nuclear Overhauser effect.
- Troc** 2,2,2-trichloroethoxycarbonyl.
- Trp** tryptophan.
- Trt** triphenylmethyl (trityl).
- TS** thrombospondin.
- Tse** 2-(4-toluenesulfonyl)ethyl.
- TSH** thyroid stimulating hormone (thyrotropin).
- Tsoc** 2-(4-toluenesulfonyl)ethoxycarbonyl.
- Tyr** tyrosine.
- Ucn** urocortin.
- UF** ultrafiltration.
- uHTS** ultra-high throughput screening.
- UK** urokinase.
- UNCA** urethane protected α -amino acid *N*-carboxy anhydride.
- uPA** urokinase-type plasminogen activator.
- UT** urotensin.
- UV** ultraviolet.
- Val** valine.
- VIC** vasoactive intestinal contractor.
- VIP** vasoactive intestinal peptide.
- VLDL** very low density lipoprotein.
- VP** vasopressin.
- Vpu** viral protein U.
- VT** vasotocin.
- WSCI** water-soluble carbodiimide.
- Xaa** unknown or unspecified amino acid (also *Aaa*).
- XAL** 5-(9-aminoxanthen-2-oxy)valeric acid
- Xan** 9*H*-xanthen-9-yl, 9-xanthidryl.
- XRP** xenopsin-related peptide.
- Z** benzyloxycarbonyl.
- Zte** 1-benzyloxycarbonylamino-2,2,2-trifluoroethyl.

Index

a

- abarelix 508–509
- Abciximab 499, 512
- abundance-based proteomics 535
 - absolute quantification 534
 - AQUA 535
- abzyme 299–300
- ACE 124, 125, 143, 430
 - inhibitors 125, 430
- ACE 2, *see* angiotensin-converting enzyme 2
- acetaldehyde/chloroanil test 274
- N^α-acetylcarnosine 64
- achatin 74, 144
- Acm 208, 209–210, 323, 334, 336, 348, 354
- acromegaly 115, 424, 425
- ACTH 110, 113–117, 130, 133–134, 177, 503, 504, 508
- actin 162
- active ester 194, 214, 234–237, 239, 241, 258, 269, 279, 300, 320, 371
 - leaving group capacity 235
 - oxazolone formation 236
 - racemization 236, 237, 257, 258, 300
 - tetrahedral intermediate 235, 237
- active specific immunotherapy (ASI) 514
- activity-based protein profiling (ABPP) 535–536
- activity-based proteomics (ABP) 535
- acyl azide 214, 217, 225–226, 245
- acyl enzyme 293, 295, 296
- acyl halides 225, 237–238
 - racemization 237–238
- O-acyl isopeptide method 273
- O-acyl isourea 232–234
- acyl migration 210, 214, 232, 233, 273, 344, 347, 378, 379
- acyltransferase 90, 293
- N-acyl urea 233, 234
- adenosine triphosphate (ATP) 68, 76, 78, 94, 102, 288, 289
- adhesion molecules 82
- adipokinetic hormones (AKH) 140
- ADME 368, 414
- ADP-ribosylation 79, 93
- adrenocorticotrophic hormone, *see* ACTH
- adrenocorticotropin, *see* ACTH
- adrenomedullin 121, 123
- advanced glycation end-products 34
- affinity enrichment 535
- agonism 120, 413, 428
- agonist 65, 66, 106, 123, 125, 126, 138, 139, 177, 418, 430, 459, 520
- agouti protein 118
- Aimoto thioester approach 329
- Akabori method 23
- alamethicin 149
- alanine scan 413
- alaphosphin 148
- albumins 486, 488, 494, 495
 - S-alkylation 192, 211, 213, 216
- allatostatin families 141, 142
- allostatin-like peptides 119, 514
- allyl transfer 191, 261, 332, 392
- allyl-type protecting group 392, 400
- alloc protecting group 392
- alphascreen technology 519–520
- Alzheimer's disease 273, 280, 497, 500
- Amadori rearrangement 19, 93
- amanitins 160
- amatoxins 73, 160, 161, 162
- amidase 293, 296, 331
- amidation 86–87, 113, 129, 357, 487
- amide, primary 218, 262
- ω-amide protection 218–221
- amine-capture method 354

- amino acid 24–26, 38, 466
 - N^α-alkylated 418
 - biosynthesis 365
 - chloride 192, 230, 237, 267
 - composition 24–26, 460
 - conformationally constrained 416
 - dehydro 365, 396, 418
 - deletion 417
 - C^α-dialkyl 417, 418, 420
 - exchange 12, 176
 - glycosylated 280, 387, 390
 - helix compatibility 38
 - β-homo 435
 - isofunctional exchange 176
 - modification 79, 417
 - N-methyl 243, 247, 377, 378, 419, 420
 - nonproteinogenic 2, 38, 147, 290, 417, 418, 422
 - preactivation 249
 - phosphorylated 395
- β-amino acid 365, 368, 376, 422, 426, 435–437, 487, 518
- D-amino acid 3, 10, 39–42, 73, 94, 96, 137, 138, 144, 297, 365, 368, 372, 376, 377, 418, 419, 478, 487, 518
- amino acid activation 76, 94, 247
- amino acid analysis 20, 24–26, 34, 35
- amino acid anhydride 232
 - N-carboxy 75, 226, 229–231, 245, 248, 324, 447
 - mixed 76, 194, 199, 226–229, 236, 245, 246, 271, 320, 324, 378
 - symmetrical 232, 233, 234–236, 240, 275, 330
 - N-thiocarboxy 231
 - urethane protected N-carboxy 231
- aminoacyl-tRNA 76–78, 94, 297
- α-aminoisobutyric acid 149, 280, 420
- 6-aminopenicillanic acid 148
- aminopeptidase 22, 125
- amphiphilic structure 42–44, 155
- amphiregulin 109
- amphomycin 399
- amylin 121–123, 510
- amyloid-β 273, 341
- anchimeric assistance 23
- anchoring group 199, 251, 256, 258, 259, 265, 275, 336, 464 *See also* linker
- ancovenin 153
- androctonus 158
- angiotensin converting enzyme 124, 143, 325, 430
- angiotensin-converting enzyme 2 125
- angiotensin kinin system 123–125
- angiotensinogen 124
- angiotensins 123–125
- anhydride 229–231
- 2-anilino-5(4H)-thiazolone 28, 29
- anorexigenic peptide 112
- antagonism 413, 428
- antagonist 3, 113, 119, 124, 125, 414, 418, 419, 424, 427, 429, 430, 459, 470, 483, 507, 510, 511, 513, 517, 520
- antamanide 73, 162, 367
- antho-Kamide 144
- anthopleurin-A 159
- antho-Rlamide I 144
- antho-RNamide 144
- antibody
 - catalytic 300
 - monoclonal 498, 512
- antibody-catalyzed synthesis 299
- antibody library 461
- antifreeze protein type III 128
- antigen 1, 16, 17, 40, 68, 70, 82, 90, 444, 445, 446, 462, 497, 498, 513, 514
- antihemophilic factor 72
- antisense nucleotides 434
- antisense peptide 491
- antithrombin 485, 494
- α₁-Antitrypsin 494
- anxiety peptide 144
- apamin 159, 160
- apidaecin 155
- apidaecin Ia 155
- Aranesp[®] 508, 511
- arene sulfonyl protecting group 203
- arginine 183, 203, 503
 - guanidino protecting group 203, 217
 - lactam formation 202, 203
 - methylation 92
 - nitro 202
- argiotoxin 159
- array 156, 414, 463–466, 501, 513
- asialoglycoprotein receptor 81
- asparagine
 - β-amide protecting group 218
 - α-aspartyl peptide 206, 218
 - cyclization reaction 218, 371
 - ω-nitrile formation 218, 220
- aspartame 2, 245, 294, 325
- aspartic acid, β-carboxy protecting group 207
- aspartimide 93, 190, 199, 206, 387
- α-aspartyl peptides 206, 218
- β-aspartyl peptides 206, 218
- atosiban 119, 505, 508, 511
- atrial natriuretic peptide 123, 127
- atriopeptide 127, 278

atriopeptigen 127
 atriopentin 278
 autocrine hormone 66
 auxiliary-mediated ligation 346
 avimers 488
 azapeptide 421
 azatide 431
 azide coupling 199, 226, 324, 328, 502

b

Bac5 155
 bacilysin 148
 bacitracin 74, 146, 147, 150
 backbone amide linker 373, 375
 backbone amide protecting group 272, 273
 backbone-engineered ligation 355–356
 backbone cyclization 11, 257, 262, 367, 371, 379
 backbone modification 421–423
 backing-off strategy 226, 256
 bactenectin 155
 bacteriocins 152
 Barlos resin 260–261, 336
 barnase 329
 β -barrel 68, 442
 Bcl-2 514, 515
 bee venom 159, 160
 benzodiazepine receptor 145
 benzyl ester 187, 198, 206, 214, 260, 342, 387
 benzyloxycarbonyl group 63, 186, 187, 202, 205, 215, 221, 387
 betacellulin (BTC) 109
 binary encoding 471, 473
 – nonsequential 471
 bioassays, solid-phase 256
 biochemical communication 65, 66, 96
 biochemical protein ligation 281
 biochemical peptide/protein synthesis 77, 125, 175, 252, 288, 317, 344, 348, 355, 358, 394
 biological membranes 51, 67, 490
 biological target 414, 477
 biologically active peptides 11, 96, 97, 216, 277, 434
 bioreactor, continuous-flow 289
 bleomycin 146, 508
 blood brain barrier 67, 132, 144, 411, 437, 491, 492
 blood clotting 70, 71, 81, 513
 blood coagulation 70, 71, 72, 80, 91
 blood pressure regulating peptide families 123
 blunt ends 282

Boc 184, 187–193, 203–207, 210, 212, 214, 217, 218, 220, 224, 228, 229, 231, 236, 245, 246, 257–260, 264, 329, 333, 342, 390, 391, 398, 400, 461, 464
 Boc/Bzl/Pac tactics 323, 324
 Boc/Bzl tactics 214, 224, 267
 bombesin 130, 143, 417, 491
 bombesin family 143, 417, 491
 bombinin 74
 bombolitin 160
 BOP 218, 239, 245, 269, 373, 502
 borapeptide 421
 BQ-123 126
 bradykinin 124, 125, 430, 510
 bradykinin-potentiating peptides (BPP) 124
 brain natriuretic peptide 127, 512
 brevenin-1 155
 bromophenolblue test 274
N-bromosuccinimide 27, 28
 bufokinin 107
 α -bungarotoxin 157
 buserelin 113, 488, 507, 508
tert-butyl 189, 190, 192, 194, 196, 198, 207, 210, 214, 216, 217, 224, 229, 238, 258, 267, 275, 322, 380, 402
tert-butoxycarbonyl group (Boc) 188, 189
 butyryl choline esterase 223
 byetta 508, 510

c

caerulein peptides 98
 calcitonin 120–123, 130, 278, 289, 333, 338, 488, 507, 511, 515
 calcitonin gene-related peptide 120–122, 130
 calcium ions 72
 calmodulin 15, 108
 campath 499
 capillary electrophoresis 14, 16, 31, 249, 341
 capillary isoelectric focusing 16
 capillary zone electrophoresis 16
 capping 269, 271, 332, 464
 capreomycin 146
 captopril 430
 carassin 107
 carbene insertion 472
 carbodiimide 231–235
 – DCC 198, 231–235, 258, 269, 279
 – polymeric 233
 carbohydrate 81, 82, 83, 85, 92, 140, 221, 285, 347, 348, 386–391, 432, 446, 514
 N^{ϵ} -(carboxymethyl)lysine 93
 carboxy protecting group 180, 193–195, 199, 200, 206, 216, 223, 253, 318, 374, 375, 400
 C^{α} -carboxy protecting group 193–199, 318

- γ -carboxyglutamic acid 34
- carboxylation 79, 81, 485
- carboxypeptidase 23, 31, 33, 125, 430
- cardiac peptide hormones 127–128
- cardionatrin IV 127
- carnosine 63, 64
- α -casein exorphin 139
- β -casomorphins 139, 373
- β -casorphins 139
- casoxins 139
- catalytic antibody 300
- CCK receptor 98
- cDNA 18, 64, 74, 98, 282, 283, 286, 477
- cecropin 44, 75, 154–156
- cell adhesion 82, 85, 246, 374, 413
- cell-free transcription/translation 290
- cell-free translation approach 288–290
- cell-penetrating peptides (CPP) 490
 - protein transduction domains 490
- cellulose 15, 64, 256, 350, 461, 464, 465
- cephalosporin c 147
- cerulein 97, 98, 108
- ceruletide 98
- cetorelix 113, 508, 509
- chain elongation, linear 78, 181, 229, 255, 269, 318, 320, 340, 342, 354, 433, 537
- chaotropic salt 272
- chemical evolution 75
- chemical ligation 264, 339, 343–359, 394, 395, 400, 488
- chemical proteomics 535
 - activity-based protein profiling 535
 - activity-based proteomics 535
 - irreversible affinity based probes 536
 - reversibly binding affinity-based probes 539, 540
- chemical shift 53, 416
- chemogenomics 516
- chemokines 91, 350, 369
- chemoselective ligation 320, 344, 345, 351, 355, 443
- chitin 18, 256, 351, 352
- chlamydocin 372
- chloromethylation 257
- 6-chloro-1-hydroxybenzotriazole (Cl-HOBt) 503
- 2-chlorotrityl group 199, 332, 335, 341, 402, 506
- chlorotoxin 159
- cholecystokinin/pancreozymin (CCK-PZ) 91, 97, 98, 104, 108, 402, 429, 508
- choline ester 223, 400, 401
- choline esterase 223, 400
- α -choriogonadotropin 91
- christmas factor 72
- chromatography
 - affinity 14–16, 18, 530, 534, 536, 540, 541
 - gel filtration 15
 - gel permeation 15, 18
 - hydrophobic interaction 15, 18
 - ion-exchange 14, 15, 18, 25
 - metal affinity 18
 - reverse-phase high performance liquid 15
 - size-exclusion 14, 15, 64
 - thin-layer 14
- chymotrypsin 26, 34, 220, 330, 394
- cilengitide 514
- cinnamycin 153
- cionin 97, 98
- circular dichroism 43, 49–50
 - vibrational 50
- cis* peptide bonds 7, 371, 376, 419
- CLIP 117, 130, 133
- cloning 30, 283, 478, 492
- clostripain 26
- α -cobratoxins 157
- coeruloplasmin 81
- cofactor 63, 81, 120
- coiled coil 42, 71, 439, 443, 444
- colicins 152
- colistin 146, 147
- collagen 38, 71, 73, 80, 117, 120, 385, 386, 485
- collision-induced dissociation 31, 532
- colony stimulating factors 494, 514
- combinatorial chemistry 176, 256, 277, 280, 350, 413, 457, 458, 460, 476, 478, 517
- combinatorial organic synthesis 457
- combinatorial peptide synthesis 457–478
 - mixture-based 459–460
 - solid-phase 460, 461, 468
- combinatorial synthesis 280, 338, 375, 429, 458, 459, 460, 466, 467, 470, 474
- complement system 70, 80
- compound libraries 429, 457, 461, 467, 471, 472, 474, 478, 516, 518, 520
- computer-aided molecular design (CAMD) 415
- conantokins 157, 158
- conformation 3, 6, 24, 36, 39–54, 71, 77, 81, 129, 176, 200, 255, 272, 299, 367, 368, 371, 373, 411, 413, 415, 420, 423, 424
- conformational analysis 3, 53, 176, 417
- conformational homogeneity 176, 381
- conopressins 157, 158
- conotoxins 157, 158

- controlled pore glass 29, 256
 - contryphans 157
 - convergent synthesis 320–321, 326–327, 342, 349, 444
 - core sequences 269, 270
 - corticoliberin 110, 113–115
 - corticostatin 104, 115
 - corticotropin 64, 112, 115–117, 130, 175, 278
 - corticotropin-like intermediate lobe peptide 117, 130
 - corticotropin release-inhibiting factor 112, 117
 - corticotropin-releasing hormone 114
 - cortistatin 105
 - countercurrent distribution 14, 18, 64
 - coupling, oxidative 210
 - coupling reaction 179, 194, 195, 218, 224, 226, 228, 229, 231, 233, 238, 241, 244, 247, 249, 250, 269, 273, 279, 280, 300, 319, 326, 334, 341, 342, 445, 463, 502
 - C-peptide 102, 103
 - crabolin 159, 160
 - cripto 109
 - cross-linked enzyme crystals 294
 - crustacean cardioactive peptide 140, 141
 - crystallization method 278, 467
 - C-type natriuretic peptide 127, 128
 - curtatocins 159
 - Curtius rearrangement 214, 226, 326, 377
 - cyanogen bromide 27, 495, 532
 - cyanuric fluoride 238
 - cyclization 368–381
 - chain-to-tail 370
 - head-to-side chain 370
 - head-to-tail 199, 369, 371, 374, 378
 - high dilution 371, 373, 382
 - pseudo-dilution 373, 382
 - side chain-to-head 380
 - side chain-to-side chain 369, 380
 - side chain-to-tail 370
 - tail-to-side chain 380
 - turn-inducing elements 368, 371
 - cyclo-(His-Pro-) 112
 - cyclodepsipeptide 149, 365, 367, 377
 - cyclodimerization 371, 376, 383
 - cyclohexyl 196, 206, 215, 232, 257
 - cyclo-oligomerization 373
 - cyclopeptide 365–381, 440
 - β -amino acid 365, 368, 376, 426
 - conformational design 368
 - conformational flexibility 368
 - β -hairpin mimetics 369
 - heteromeric 146
 - homomeric 146
 - N-methylation 368
 - pentapeptide 371–373, 376, 377
 - protein epitope mimetics 369
 - tetrapeptides 372, 376, 377
 - tripeptides 376
 - turn 365, 367–369, 371
 - cyclopeptide synthesis
 - backbone anchoring 374
 - cyclization auxiliary 377
 - cyclization in solution 373
 - cyclodimerization 371, 376, 383
 - cyclo-oligomerization 373
 - diphenylphosphoryl azide (DPPA) 226, 373
 - 1,3-dipolar cycloaddition 380, 381, 542
 - disulfide 366, 367, 369, 380–388
 - epimerization 372, 373, 377
 - metathesis macrocyclization 379, 380
 - on-resin cyclization 373, 374, 375, 380
 - pentafluorophenylester ring closure 371
 - side chain anchoring 374
 - Staudinger ligation 357, 379
 - triphosgene 377
 - cyclophilin (Cyp) 7, 151
 - cycloscan 367
 - cyclosporin A 7, 96, 151, 512
 - cyclosporin O 365
 - cyclosporin synthetase 96, 151
 - cyclotheonamide 422
 - cyclotide 370
 - Cyl-1 372
 - Cyl-2 372
 - cysteine protease 293, 294, 353, 537, 538
 - cystine-peptide 381
 - condensation of fragments 381
 - oxidative refolding 381
 - cysteine, thiol protecting group 208, 210, 383
 - cytokines 82, 485, 496, 498, 499
- d**
- dactinomycin 146
 - dalfopristin 147
 - dansyl chloride 21, 22
 - dansyl method 21, 22
 - daptomycin 365, 368, 399, 508, 513
 - darbepoetin alfa 508, 511
 - DAST 238
 - DCC 198, 219, 231–234, 237, 258, 269, 279, 341, 435
 - DDAVP 119, 418, 503
 - Dde 192, 205
 - deamidation 19, 36, 218
 - deconvolution
 - recursive 467, 475

- defensin 74, 75, 155, 156
 - dehydroalanine 153, 211
 - delta-sleep inducing peptide 130, 144
 - deltorphins 137, 138
 - de-novo design 413, 439, 442, 444
 - de-novo sequencing 32
 - dephosphorylation 87, 88, 395, 398
 - depletion 120, 511, 530
 - depsipeptide 147, 149, 245, 273, 421
 - dermorphins 74, 137–138
 - des-Gln¹⁴-ghrelin 108, 109
 - deslorelin 507, 508
 - desmopressin 325, 488, 503, 511
 - detection 14–16, 250, 500, 518–519, 521, 531, 532, 535
 - detirelix 113, 509
 - deuteration 52
 - diabetes-associated peptide 122
 - diabetes insipidus 119, 418, 503
 - diabetes mellitus 100, 102, 122, 283, 488, 493, 501, 510, 511
 - insulin-dependent 102
 - noninsulin-dependent 122
 - N,N'*-dialkyl urea 233
 - diazepam-binding inhibitor 156
 - diazepam-binding inhibitor peptide 144–145
 - dibenzofulvene 189, 191, 267, 323
 - didemnin 372
 - diethylglycine 426
 - differential 2D gel electrophoresis (DIGE) 531
 - difficult sequence 271–273, 339
 - incomplete acylation 271
 - dihedral angle 36, 41, 420, 438
 - diketopiperazine 19, 36, 256, 261, 264, 271, 277, 332, 375
 - β -2,4-dimethyl-3-pentyl ester 207
 - 4-(dimethylamino)pyridine 258
 - 2,4-dinitrofluorobenzene 21
 - dinitrophenyl method 21
 - Diosynth rapid solution peptide synthesis (DioRaSSP) 507
 - diphenylphosphoryl azide, *see* DPPA
 - diptericin 394
 - directed assembly 320
 - discontinuous epitopes 178, 367, 412
 - disulfide bond (disulfide bridge) 10, 44, 50, 79, 116, 152, 155–158, 217, 269, 341, 366, 360, 381, 383, 384, 386, 492, 503, 505, 512
 - cleavage 24
 - collagen peptides 386
 - insulin 80, 101, 382, 383, 384
 - interchain 10, 34, 104, 381, 382, 386
 - intrachain 10, 103, 158, 381, 424
 - minicollagen-1 385
 - diversity 460, 468, 473, 483, 484, 492, 497, 515
 - diversomer 458, 473
 - Dmab 197, 199, 208
 - 2,4-Dmb ester 195
 - DNA ligase 282
 - DNA sequence analysis 30, 281
 - DNA synthesis 283, 513
 - DNA template-controlled ligation 351
 - dolichol pyrophosphate 84
 - DPDPE 380, 424, 492
 - DPPA 226, 245, 373, 376
 - drug candidate 411, 414, 457–460, 501, 507
 - drug delivery systems 488–492
 - drug design 3, 49, 153, 414–416, 457, 515
 - Dts 191, 268, 269
 - duramycins 153
 - dwarfism 115, 286, 496
 - dynamic combinatorial library 476–477
 - dynamic range 529
 - dynorphins 132, 135, 136
- e**
- early placenta insulin-like peptide 104
 - Edman degradation 23, 28–30, 32, 270, 474
 - Edman microsequencing 471
 - EGF family 109–110
 - eisenin 90
 - elastin 73
 - elcatonin 121, 508, 511
 - electrophilic aromatic substitution 214, 217
 - electrostatic interaction 39, 45, 92, 439
 - eledoisin 106, 107, 508
 - β -elimination 19, 189, 211, 220, 336, 387, 391, 396, 400
 - elongation factors 78
 - enalapril 430
 - encoding methods 470–474
 - nonchemical 474
 - end group analysis 21–24
 - C-terminal 21
 - N-terminal 21, 34
 - endocrine hormone 66, 97, 111, 115
 - endokinins 105, 106
 - endomorphins 141
 - endopeptidase 26, 124, 129
 - endorphins 131–136
 - α -endorphin 134, 135
 - β -endorphin 113, 117, 130, 133, 134, 278, 430
 - γ -endorphin 134, 135
 - δ -endorphin 134, 135
 - α -neo 132, 134–136

- endothelin 123, 125–127
 endothelium-derived relaxing factor 126
 endothiopeptide 421
 enduracidin 147, 148
 enfuvirtide 4, 235, 278, 336, 337, 484, 506, 507, 508, 509, 514
 enkephalin (EK) 67, 132–135, 424
 – Leu- 132, 319
 – Met- 132, 137, 492
 enniatins 63
 enolization 246
 enzymatic degradation 31, 81, 367, 486
 enzymatic ligation 297, 330, 358, 359
 enzymatic synthesis 283, 291, 294, 318, 319, 331, 388, 389, 391, 394
 enzyme 220, 221, 223, 245, 277, 281, 295, 397, 400, 411–416, 536
 enzyme assay 17, 473, 520
 enzyme inhibitor 1, 15, 414, 428, 429, 434, 467, 537
 epidermal growth factor 109, 493, 496
 epidermin 153
 epimerization 94, 96, 181, 206, 209, 213, 233, 246–247, 249, 270, 280, 297, 339–341, 372, 373, 377
 epiregulin 109
 epitope
 – continuous 177, 178, 367, 412
 – discontinuous 178, 367, 412
 epitope mapping 177, 463, 467, 470, 501
 eptifibatide 508, 512
 equilibrium-controlled enzymatic synthesis 292–294, 359
 erythrocyte differentiation factor 494
 erythropoietin 487, 490, 493, 511
 ESI 20, 31
 esterase 221, 293, 294, 400
 ethyl 195, 228
 eumenine mastoparan-AF 159, 160
 excluded protecting group method (EPG) 280, 359
 exenatide 508, 510
 exendins 99, 510
 exorphins 139–140
 expansion of the genetic code 290
 expressed protein ligation 297, 351–352, 368
 extended native chemical ligation 346
- f**
- F₁F₀-ATPase 68
 Fab fragment 68, 512
 FACS 470
 factor II 71
 S-farnesylation 89, 400
 farnesylation 89
 farnesyltransferase (F Tase) 89
 Fc fragment 68, 489
 fentanyl 131, 132, 134
 fibrin 71, 72
 fibrinogen 72, 485
 fibrinopeptides 72
 fibrin-stabilizing factor 72
 fibroblast growth factors 490, 494, 496
 fibroin 39, 71
 fibronectin 494
 fibrous proteins 71
 fingerprint 31, 532
 FK506 binding proteins 7
 fluorescence correlation spectroscopy 49, 56
 fluorescence resonance energy transfer 55, 518, 519
 9-Fluorenylmethyl (Fm) 192, 197, 209, 238
 Fmoc 188, 204–206, 210, 213, 217, 220
 Fmoc-protected amino acid halides 189, 228, 238, 261, 464
 Fmoc/tBu tactics 27, 213, 224, 276, 322
 foldamer 431, 436
 follicle-stimulating hormone releasing hormone 113
 folliliberin 110
 follitropin 110, 116
 forced peptide synthesis 246
 force-field calculations 415
 N^{fm}-formylation 217
 N-formyl group 149
 Forteo® 120, 509, 511
 fourier transform infrared spectroscopy 51
 fragmentation 26, 31, 32, 221, 223, 264, 531, 534, 540
 fragment condensation 181, 235, 320, 330, 331, 383, 384, 433
 fragment growing 416
 fulicin 74
 functional proteomics 530, 535–537
 fusion proteins 15, 16, 18, 287, 483, 487, 489, 493, 500
 Fuzeon® *see* enfuvirtide
- g**
- G protein-coupled receptors 65, 104–105, 107, 118, 119, 121, 135, 146, 147, 400
 G proteins 121, 400
 galanin 130, 142, 490
 gas chromatography, electron capture detection 473
 gas phase sequencer 29, 30
 gastric inhibitory polypeptide 100, 156
 gastrin 66, 91, 97–99, 104, 130, 402, 417, 483

- gastrin family 97–99
 - gastrin-releasing peptide 130, 417, 483
 - gastroenteropancreatic peptide families 96, 97
 - gastrointestinal hormone 97
 - gene 120, 121, 281
 - gene transfer 447
 - genetic algorithms 421, 422
 - genetic engineering 55, 281, 285–287, 496
 - genome 2, 49, 175, 282, 457, 477, 500–501, 515–517, 529, 532, 537
 - S-geranylgeranylation 89
 - geranylgeranyltransferase 89
 - ghrelin 108, 109, 115
 - GIP(7–42) 156
 - glucicentin 99
 - glucicentin-related pancreatic peptide (GRPP) 99
 - global restriction approach 423
 - globin 140
 - globulins 124
 - glucagon 64, 67, 88, 99–102, 104, 112, 278, 424, 425, 483, 508, 510
 - glucagon-like peptides 99, 100, 101, 483
 - glucose-dependent insulinotropic polypeptide 99, 100
 - α -glucosidase 221
 - β -glucosidase 221
 - glutamic acid 34, 90, 98, 112, 205, 206, 207, 219
 - glutaminy cyclase 90
 - γ -carboxy protecting group 374, 375
 - glutamine
 - nitrile formation 218
 - pyroglutamyl peptide 218
 - ω -amide protecting group 219
 - glutathione 8, 9, 63, 93
 - gluten 140
 - gluten-exorphin A5 139, 140
 - gluten-exorphins 140
 - glycans 81, 84–86, 90, 394
 - glycoconjugate 82
 - glycoforms 81
 - glycopeptide remodeling 386
 - N-glycopeptide 390
 - O-glycopeptide 391
 - glycopeptide synthesis 86, 346, 386, 387, 391, 392, 394
 - hydrophilic resins 393
 - native chemical ligation 344–346, 350, 352, 353, 394–395
 - safety-catch 378, 395
 - solid-phase peptide synthesis 3, 4, 175, 190, 251–279, 393, 446
 - sugar-assisted ligation 347
 - glycoproteins
 - N-linked 395
 - O-linked 84
 - trimming 84
 - endo-glycosidase 390
 - exo-glycosidase 390
 - C-glycoside 83, 85
 - N-glycoside 83, 387
 - O-glycoside 83
 - β -N-glycosidic bond 387
 - α -O-glycosidic bond 388
 - glycosidic bond 82, 83, 85, 93, 221, 387, 388, 391, 394
 - glycosulfopeptide 402
 - glycosylation 81–86
 - block 387
 - stepwise 84
 - glycosyltransferases 386, 390
 - gonadoliberin 90, 110, 113, 419, 507
 - gonadotropin-releasing hormone 110, 419
 - gonadotropins 110, 113, 419
 - goserelin 507, 508, 509
 - GPI-linked proteins 90
 - gramicidins 11, 39, 63, 64, 67, 95, 96, 146, 147, 149, 365
 - gramicidin A 39, 67, 149
 - gramicidin S 11, 63, 95, 96, 146, 147, 365
 - granulocyte colony stimulating factor 494
 - granulocyte macrophage colony stimulating factor 514
 - graph theoretical methods 415
 - greek key motif 46
 - green fluorescent protein 55, 278, 335, 518
 - growth hormone 100, 103, 104, 108, 109, 114, 115, 286, 386, 424, 425, 484, 492, 496
 - growth hormone release-inhibiting hormone 110
 - growth hormone secretagogue (GHS-R) 108, 115
 - receptor 108, 115
 - growth hormone-releasing hormone 98–99, 100, 110, 114, 115
 - receptor 110
 - GRPP 99
 - GTP-binding proteins 89
 - 4-guanidinophenyl ester 295, 330
 - guanidino protection 202–204
 - guanine nucleotide-binding proteins 15
 - guanylation 204, 211
- h**
- Hageman factor 72
 - β -hairpin 42, 43, 46, 369

- β -hairpin mimetics 369
- handles 256, 280, 332, 341–343
- HAPipU 241, 242
- HAPyU 377
- HATU 240–245, 373, 376
- HBTU 218, 235, 240, 243, 269, 280, 373
- HC Toxin 372
- head activator 145
- helix
 - bundles 42, 46, 439, 442, 443
 - compatibility 38
- α -helix
 - initiators 426
- $\beta\alpha\beta$ -helix 442
- 3_{10} -helix 38, 41, 42
- helix-turn-helix motif 46
- helospectine I 99
- helospectine II 99
- hematide 511
- hemocyanins 514
- hemoglobin 68, 140
- hemokinin-1 105, 106
- hemopressin 139, 140
- hemorphins 140
- heparin binding growth factor 91, 109
- hepatitis B surface antigen 494
- heptyl ester 223, 396, 397
- herceptin 496
- heregulin 109
- heterodetic peptide 11, 146
- heteromeric peptide 10, 74, 146
- hexafluoroacetone 245, 246
- HFMS resin 336
- high-frequency signal 474
- high-molecular weight kininogen 71
- high-throughput screening 290, 433, 458, 473, 497, 518–521
- hirudin 341, 485
 - imidazole group protection 211–215
- histidine 18, 45, 64, 211
- histone methyltransferases (HMT) 520
- histones 92, 520, 521
- HIV protease 278
- HIV-1 gp 41, 490, 506
- Hmb 200, 201, 206, 272
- HMBA resin 260
- Hnb 200, 201, 206, 272, 377
- HOAt 218, 225, 234, 237, 239–242, 272, 324, 340, 372, 373, 503
- HOBT 206, 213, 218, 225, 234–239, 242, 258, 280, 319, 320, 324, 326, 329, 334, 336, 340, 503
- HOdhbt 326, 327, 329
- homocysteine 216, 520
- homodetic peptide 11
- homogeneity characterization 17, 20, 249, 326, 381
- homogeneous time-resolved fluorescence (HTRF) screening assay 519
- homology design 48
- homology modeling 48, 416
- homomeric peptide 10, 11, 146
- HONB 237
- hormone
 - autocrine 66
 - endocrine 66, 97, 111, 115
 - paracrine 66
 - pituitary 115–118, 145
- HOSu 234, 237, 328, 329, 340
- HPLC 15, 16, 25, 26, 31, 249, 250, 277, 318, 342, 507
- ^{15}N HSQC 416
- HTRF assay 519
- Humalog 510
- human genome 2, 7, 175, 281, 457, 500, 515, 520, 537
- human immunodeficiency virus (HIV) 2, 349, 490
- Humatrop 287, 496
- hybrid approach 325, 332, 335, 484, 506
- HYCRON linker 393
- hydrazide 23, 182, 198, 199, 200, 217, 218, 225, 226, 256, 260, 262, 276, 356
- hydrazinolysis 192, 193, 198, 208, 212, 214, 225, 226, 262, 268, 333
- hydrazone-forming ligation 356
- hydrins 25, 273, 343
- hydrogen bond 36–40, 44, 45, 50–53, 200, 219, 236, 237, 271, 411, 428, 432–440
- hydrogen fluoride 187, 198, 323
- hydrophilic resins 393
- hydrophobic collapse 48, 429
- hydrophobic interaction 15, 45, 46, 428, 429, 501
- hydrophobic protein-ligand interactions 429
- hydroxy group protection 214–216
- α -hydroxylating monooxygenase 87
- hydroxylation 36, 79, 80, 87
- N-hydroxypiperidinyl ester 236
- hylambatins 107
- hypertensin 125
- hypocretins 136–137
- hypophysis 64, 111, 117, 136
- hypothalamic releasing hormones 110, 111
- hypothalamus 90, 100, 101, 104, 108–119, 136, 143–146

i

icatibant 508, 510
 IDSP sequence 82
 IgE pentapeptide (human) 68, 499
 IgF receptor 101, 103, 115
 imidazole protection 211–214
 imine capture ligation 351
 immobilized pH gradient 531
 immune system 68, 70, 90, 105, 106, 114, 128, 483, 496, 500
 immunoglobulins 15, 68, 70, 487–489, 499
 incretins 100, 101
 indole protecting group 217–218
 indolicidin 155
 inflammatory compounds 82
 inhibin 328
 inhibiting factors 110, 112, 117
 inhibitor affinity chromatography (IAC) 536, 540
 insect diuretic peptides 127
in-situ neutralization 272
 insulin 2, 20, 21, 34, 35, 66, 67, 80, 100–104, 108, 145, 175, 209, 284, 285, 287, 288, 294, 425, 493, 496, 510
 – B chain 34, 35, 70, 80, 102–104, 140, 287, 385, 510
 – C-peptide 102, 103
 – glargine (HOE901) 287, 510
 – Lispro 287, 510
 insulin-like growth factor(s) 101, 103, 115
 – growth factor 1 115
 insulin-like peptides 103, 104
 insulin receptor 102
 insulin superfamily 101–104
 integrilin 508
 integrins 82, 100, 368, 413, 514
 intein-mediated protein ligation 351–353
 interaction– dipole-dipole 45
 – ion-dipole 45
 intercellular adhesion molecule 82
 interferons 284, 286, 493, 495, 496
 interleukin 66, 278, 345, 493–496, 499
 intermediate filaments 71
 intermedin/adrenomedullin-2 121, 123
 internal reference amino acid 256
 introns 76, 123, 282, 286
 inverse peptide synthesis 318
 iodolysis 210
 islet amyloid polypeptide 122
 isoaspartyl peptide formation 205, 206, 276
 isokinetic mixture 468
 isopenicillin N synthase 148
 isopeptide bond 8
 isotope-labeled hydrolytic agents 249

isovaline 420
 isoxazolium method 236
 iteration method 474

j

jak proteins 535

k

Kaiser oxime resin 332
 Kaiser test 273, 341
 KALA amphipathic peptide 159, 490
 kaliuretic peptide 128
 kallidin 124
 kallikrein–kinin system 124
 kallikreins 72, 124
 kassinin 106, 107
 keratin 38, 39, 71
 α -ketoacid–hydroxylamine amide ligation 357, 358
 ketone bodies 102
 kinase 65, 87, 88, 102, 330, 331, 519, 541
 kinetically controlled enzymatic synthesis 292–294
 kinetically controlled ligation 348–350
 kininase II 124
 kinins 123, 124
 kyotorphin 143

l

α -lactalbumin 140
 β -lactam antibiotics 148
 lactoferrin 139
 lactogenic hormone 116
 β -lactoglobulin 139, 140
 lactorphins 139, 140
 ladder sequencing 32, 33
 lanreotide 104, 508, 512
 lantibiotics 74, 153, 366
 Lantus 287, 510
 large-scale peptide synthesis 278, 324, 484, 502–507
 LDV sequence 82
 lead compound 126, 176, 177, 414, 457, 458, 460
 leptin 108, 115
 Leu-enkephalin 132, 134
 leukocyte interferon 82, 83, 85
 leuprolide 113, 507, 508
 Levinthal paradox 48
 Leydig cell insulin-like peptide 104
 LF-transferase 297
 LH-RH 110, 113, 325, 338, 488, 491, 508
 liberins 90, 110
 library 177, 430, 458, 459, 472–477

- ligand-receptor interaction 415
- ligase C-N 291, 296, 332, 353
- ligation 343–359
 - expressed protein 351–353, 358
 - extended native chemical ligation 346–348
 - hydrazone-forming 356
 - kinetically controlled ligation 348–350
 - native chemical 344–345, 350–353, 394, 395, 400
- ligation strategies
 - desulfurization 348
- linker 222, 252–263
 - orthogonal 472, 473
- lipase 221, 223, 494
- lipidation 79, 88–90, 400, 401, 497
- Lipinski rules 411
- lipomodulin 494
- lipopeptides 147, 151, 220, 365, 398–401, 508, 513
 - antimicrobial 398
 - Diels–Alder ligation 400
 - Pam₃Cys 399
 - S-farnesylated 400
 - S-palmitoylated 400
 - synthetic adjuvants 399
 - vaccine 399
- lipophilic segment coupling 333, 334
- lipoproteins 399
- lipotropic hormone 116
- lipotropin 116
- β-lipotropin 132, 133, 134, 135, 278, 328
- Liprolog/Humalog 510
- liquid phase sequencer 29
- liquid-phase peptide synthesis 245, 251, 278–280, 466
- lisinopril 430
- liver cell growth factor 71
- locustatachykinin I 106, 107
- locustatachykinin II 106, 107
- long-acting natriuretic peptide 127
- loop mimetics 426
- losartan 430
- luliberin 110
- luteinizing hormone 113, 116, 325, 447
- luteinizing hormone-releasing hormone 113, 447
- luteotropin 116
- lutropin 116
- lymphokines 498
- lypressin 508
- lysine, ε-amino protection 205
- lysine side chain N-acetylation 92
- lysine specific histone methyltransferases (KMT) 520
- lysozyme 277, 349, 443
- lysyl hydroxylase 80
- m**
- mabthera/rituxan 499
- machine learning 415
- macrocyclization 6, 245, 373, 377, 379, 380
- macrophage colony stimulating factor 494, 514
- macrophage inhibitory factor 494
- magainins 44, 74, 154, 155
- mahagony 118
- maillard reaction 19, 93
- major histocompatibility complex 69, 70, 436
- MALDI-ToF MS 274
- margatoxin 158
- mass spectrometry 16, 20, 31–32, 35, 249, 341, 342, 350, 501, 530–532, 540
- mast cell degranulating peptide 160
- mastoparan 159, 160, 490
- Mbs 203, 204, 213
- melanin-concentrating hormone 131
- melanocortin 117, 118
- melanocortin system 118
- melanocyte-stimulating hormone (MSH) 89, 116–118, 130–134, 418, 425, 491
 - α-MSH 89, 116–118, 130–134, 381, 418
 - β-MSH 116–118
 - γ-MSH 117, 118, 133, 134
- melanoliberin 117
- melanostatin 110, 117
- melanotans 118
- melanotropin 64, 89, 110, 113, 116–118
- α-melanotropin *see* melanocyte-stimulating hormone
- α-MSH *see* melanocyte-stimulating hormone
- β-melanotropin 64
 - β-MSH *see* melanocyte-stimulating hormone
 - γ-MSH *see* melanocyte-stimulating hormone
- mellitin 43, 159
- membrane 16, 18, 44, 51, 52, 67, 68, 70, 71, 81, 102, 152, 153, 155, 289, 336, 400, 411
- Merrifield synthesis 181, 252, 253
- Merrifield tactics 265
- messenger RNA 75, 282
- metabolic stability 366
- metal affinity chromatography 18
- metal chelation 44
- metalloprotease 430, 541, 542
- metalloproteins 440
 - miniaturized 440
- metathesis, ring-closing 379
- Met-enkephalin 132, 134, 135, 137, 498
- methionine

- oxidation 24, 36, 93, 217,
 - sulfoxide formation 24, 34, 36, 217
 - thioether group protection 190, 216, 217
 - methyl ester 89, 198, 257, 320, 330, 434
 - α -methylphenacyl ester resin 333
 - MHC proteins 69, 70, 436
 - microcalorimetry 49
 - microglobulins 70
 - microheterogeneity 81
 - microscopic reversibility 291
 - microwave-enhanced peptide synthesis 280
 - midkine 326
 - minicollagen-1 385
 - mismatch sequence 269–271
 - milk protein-derived opioid peptides 139
 - model peptide 51, 177, 178, 250, 295, 299, 440, 490
 - molecular chaperones 7
 - molecular modeling 3, 176, 366, 415, 429
 - monitoring on-resin 273–274
 - monoclonal antibodies 177, 299, 467, 483–485, 493, 498, 499
 - morphiceptin 139
 - morphine 131–134, 137, 139, 430, 512
 - morphine modulating neuropeptides 430
 - motilin 108, 109, 130, 131, 175
 - motilin-related peptide (MAP) 108
 - MS peptide sequencing 16
 - MSI-78 155
 - Mtr 204
 - Mts 204
 - MudPIT 531, 533
 - mucins 85
 - Mukaiyama's reagent 243
 - multidimensional protein identification technology 531
 - multidimensional separations 14
 - multipin synthesis 461, 462, 463
 - multiple antigen peptides 444, 445, 446
 - multiple peptide synthesis 175, 177, 459–465
- n**
- nafarelin 113
 - nalorphine 134
 - naloxone 134, 140
 - naltrexone 134
 - nanoscale polymer particles 447
 - nanospray 31
 - native chemical ligation 344–346, 350–355, 394, 395, 400
 - Natrecor[®] 128, 508, 512
 - natriuretic peptides 127
 - NBB resin 261, 333
 - neighboring group participation 264, 388
 - α -neoendorphin 132, 134–136
 - β -neoendorphins 132, 134–136
 - neokytorphin 143
 - nerve growth factor 487, 494
 - nesiritide 128, 508, 512
 - neurohormones 67, 130
 - neurohypophyseal hormones 110, 118–120
 - neurokinins 105, 106, 143
 - neuromedin B 142
 - neuromedin C 143
 - neuromedin N 143, 145
 - neuromedin U 143
 - neuromedins 142, 143, 145
 - neuronal network 415
 - neuropeptide 97, 101, 103, 106, 107, 123, 128–130, 136–138, 142–146
 - neuropeptide F 145
 - neuropeptide FF 130, 145
 - neuropeptide γ 106
 - neuropeptide K 106
 - neuropeptide receptor 137, 145
 - neuropeptide 26Rfa 145
 - neuropeptide S 146
 - neuropeptide W 146
 - neuropeptide Y 106, 107, 123, 137, 146, 333, 491
 - neuropeptide Y family 106–108, 146
 - neurophysins 119
 - neuoregulins 109
 - neurotensin 90, 130, 131, 143, 145, 370
 - neurotensin-like peptide 143, 145
 - neurotransmitter 1, 65, 67, 88, 96–98, 100, 105, 106, 112, 126, 130, 132, 144, 516
 - neutralization, *in-situ* 272
 - ninhydrin 25, 273
 - ninhydrin reaction 25
 - ninhydrin test 273
 - nisin 74, 153, 154
 - nitrile formation 218, 220
 - 4-nitrophenyl ester 236
 - 3-nitrotyrosine 34, 93
 - NMDA receptor 158
 - NMR spectroscopy 20, 49, 52–54
 - gel-phase 274
 - solid-state 52
 - nociceptin 130, 135, 138, 139
 - nociceptin/orphanin 130, 138, 139
 - nocistatin 138, 139
 - nomenclature 7, 8, 11, 12, 38, 344
 - nonmammalian tachykinins 106
 - non peptide drug 3, 412, 415, 433, 473
 - non-native chemoselective ligation 349
 - nonribosomal peptide synthesis (NRPS) 94–96

NOP receptor 138
 Npys 205, 210
 Nsc 192, 193
 S1-nuclease 282
 nuclear overhauser effect 53
 nucleoproteins 30
 nucleotidylation 79
 Nvoc 183, 465, 466

o

obestatin 109
 octadecaneuropeptide (ODN) 144
 octreotide 105, 425, 508, 512
 oligocarbamate 431, 432, 437
 oligopyrrolinone 431, 438,
 oligosulfonamide 437
 – vinylogous 431
 oligosulfone 431
 oligourea 431
 omphalotin A 365, 377, 378
 ophthalmic acid 64
 opioid peptides 131–136
 – endogenous 131, 132, 135
 – milk protein-derived 139
 opioid receptor 131–135, 138–142, 430
 optical rotation index 49
 optical rotatory dispersion 49
 orexins 130, 135, 136
 ORL1 receptor 138
 ornithine, ω -amino group protection 204–205
 orphanin FQ 138
 orthoclone (OKT3) 494, 498, 499
 orthogonal ligation 350
 orthogonality 182, 223, 224, 262, 267,
 268–269, 322, 338, 380
 – hidden 262
 ostabolin-C 120, 511
 osteoclast 120
 ovalbumin 88
 overlapping fragments 34
 5(4H)-oxazolone
 – 2-alkoxy 228, 233, 234, 248, 249
 – formation 247, 248
 oxidative degradation 276
 oxime-forming ligation 356
 oxyntomodulin 99
 oxytocin 64, 116, 118–120, 123, 130, 176, 177,
 209, 320, 325, 418, 424, 488, 505, 511
 – antagonist 505
 – receptor 511

p

PACAP-38 99, 101
 palmitoylation 400

panbo-RPCH 140, 141
 pancreatic islet hormones 109
 pancreatic polypeptide 106, 107, 108
 pancreozymin 98
 PAOB 223
 PAPS 91
 paracrine hormone 66
 parallel synthesis 460–467
 parathormone 120
 parathyrin 120, 278
 parathyroid hormone 120, 121, 284, 511
 parathyroid hormone-related peptide 121
 parvulins 7
 Pbf 204
 PEG-protein conjugates 488
 PEGylation techniques 486–488
 pelvetin 90
 penetratin 490
 penicillin G acylase 221
 penicillin N 147, 148
 pentafluorophenyl ester 235, 246, 258, 343
 Pep5 153
 peplomycin 146
 pepstatin 428
 peptaibol 149, 398
 peptaibols 149
 peptibodies 487
 peptiCLECs 294
 peptide
 – aminoxy 435, 436, 437
 – analysis 14
 – antibiotics 74, 96, 146–156, 282, 398,
 508
 – bioavailability 3, 176, 368, 412, 420, 485,
 488, 492, 515
 – carbo 432
 – cystine bridge 381–382
 – defense 152, 154, 156
 – heterodetic 11, 146
 – heterodetic cyclic 104, 118, 424
 – heteromeric 10, 74, 146
 – homodetic 11
 – homogeneity 20
 – homomeric 10, 11, 146
 – hydrazino 431
 – hydrolysis 21, 23, 24, 188, 198, 274, 291,
 293, 294, 390, 427
 – mass map 31
 – pharmacophoric groups 3, 176, 368, 413,
 414, 429
 – proteolytic cleavage 79, 98, 99, 104, 124,
 293, 295, 497
 – therapeutic 2, 285, 287, 483–489, 493–497,
 498

- thioester 346
- toxin 96, 156–162
- vinylogous 422, 431, 438
- β-peptide 4, 431, 432, 435, 436, 437
- γ-peptide 8, 431, 432, 437, 486, 518, 537
- peptide aldehyde 276, 356
- peptide amide 87, 100, 107, 108, 122, 123, 138, 143, 145, 160, 260, 332
- peptide analysis 14
 - amino acid composition 24, 460, 491
 - C-terminal end group 21–24
 - N-terminal end group 21–24, 34
- peptide antibiotic 74, 96, 146–156, 282
- nonribosomally synthesized 74, 147–152
- ribosomally synthesized 74, 152–156
- peptide aptamers 522
- peptide bond
 - *cis* 7, 41, 371, 376, 419
 - double, bond character 6, 36
 - formation 75, 78, 94, 96, 178–181, 188, 224–246, 281, 291, 292, 296–299, 321, 328, 351, 376
 - antibody-catalyzed formation 297–300
 - sensitivity 16
 - *trans* 419
- peptide cleavage 34, 275–277, 428
- peptide dendrimer 444–446, 498
- peptide dendrimer synthesis 444–446
- peptide-derived active pharmaceutical ingredients 502
- peptide drug 484, 488, 492, 502–515
 - ADME 414
- peptide drug delivery systems 488–492
- peptide ester 260, 276, 374, 375
- peptide folding 417
- peptide histidine isoleucine 98, 100
- peptide histidine isoleucine amide 99, 100
- peptide hormone 64–66, 80, 91, 96, 97, 101, 108, 110, 115, 117, 119–121, 124, 127, 128, 133, 140, 176, 413, 417, 429, 485, 501, 510
- peptide hydrazide 225, 260, 276, 356
- peptide leucine arginine 45
- peptide library 470, 474, 477
- peptide ligase 291, 444
- peptide mass fingerprint 532
- peptide modification 412, 418, 423, 429
- peptide nanotubes 440, 441
- peptide nucleic acids 431, 434–435
- peptide pharmaceuticals 325, 502–522
- peptide polymer 447
- peptide production plant 484, 502, 505, 507
- peptide synthesis
 - concepts 317–359
 - large-scale 278, 324, 484, 502–507
 - LF-transferase 297, 298
 - microwave-enhanced 280
 - non-ribosomal 74, 297
 - polymer reagent 279
 - protease-catalyzed 291–295, 297, 358
 - ribosomal 75–79, 295
 - segment condensation 181, 277, 296
 - stepwise 323, 337
 - strategy 3
 - tactics 317–324
- peptide synthesizer 267, 268, 274, 275, 462, 506, 507
- peptide synthetases 94
- peptide thioester synthesis 346
- peptide toxin 156–162
- peptide YY 106–108
- peptide-based vaccines 497–498
- peptidome 530
- peptidomimetic 411–413, 418–439
 - alkene 422
 - amide analog 421
 - aminoxy acid 422
 - carbapeptide 424
 - hydrazino acid 422
 - hydroxyethylene 422, 428
 - ketomethylene isostere 421
 - phosphonamide 299, 421
 - reduced amide 421
 - retro- 421
 - retro-inverso 422, 423
 - sulfinamide 421
 - sulfonamide 421
 - thioester 421
- peptidyl-α-hydroxyglycine α-amidating lyase 87
- peptidyl prolyl *cis/trans* isomerases 7
- peptidyl transferase 78, 281, 291, 295
- peptidyl transferase center (PTC) 78, 281
- peptidylglycine α-amidating monooxygenase 86, 113
- peptoids 432–434
- peptoid libraries 433
- peptoid synthesis 432
- β-peptoids 433
- Pfp 434
- pGlu 90, 112, 140
- phage 477, 483, 488, 491, 499, 500, 521
- phage coat protein 477, 478
- phage display 477, 478, 488, 491, 499, 500, 511, 521
- phallotoxins 73, 160, 161, 162

- pharmacophoric groups 3, 176, 368, 412–414, 416, 429
- pharmacokinetics 414, 486, 495, 510
 - ADME 368, 414
- phase change synthesis 335
- phenyl ester, halogen-substituted 236
- 3-phenyl-2-thiohydantoin 28
- phenylalanine, constrained derivatives 420
- phenyl isothiocyanate 28, 29
- 3-phenylproline 420
- phenylthiocarbamoyl 28, 32
- N-phenylthiomethyl 272
- 3'-phosphoadenosine-5'-phosphosulfate (PAPS) 92
- phosphonium reagent 239, 241, 373
- phosphopeptides 395–398
- phosphorylation 35, 79, 87–88, 91–93, 346, 395, 396, 398, 514, 519
- phosphoserine 26, 395, 398
- phosphothreonine 395, 398
- phosphotyrosine 88, 395, 396, 398
- phosphotyrosine binding domain 88
- phosphotyrosine mimetics 398
- photoaffinity labelling 291, 516, 541
- photolabile linker 261, 472
- phyllomedusin 107
- physalaemin 90, 106, 107
- picolyl ester method 342
- pituitary hormones 115–118
- pitressin 508
- pituitary adenylate cyclase activating polypeptide (PACAP) 99, 101, 115
- placentin 104
- plant defensins 75
- plasma kinins 123, 124
- plasma thromboplastin antecedent 72
- plasmid 283, 285, 289
- plasmin 72
- plasminogen 386, 493, 495
- plasminogen activators 493
- platelet-derived growth factor 494, 515
- pleiotrophin 326
- Pmc 206, 324, 334
- polarized light 49
- poly(dimethylacrylamide) 255, 393
- polyarginine 447
- polyethylene 255, 256, 278, 461–465, 474
- polylysine 446, 447
- polymerase I 282
- polymeric support 251
 - solid 181, 198, 251–262, 373
 - soluble 198, 278, 285, 466–467
- polymyxins 74, 146, 151
- polyproteins 80, 113, 128, 129, 134
 - precursor 80, 128, 134
- polystyrene 15, 253–257, 259–262, 272, 278, 337, 339, 342, 393, 461, 465
- POMC 80, 113, 117, 132, 133, 134
- porin 67, 68
- positional scanning 474–476, 537
- post-source decay 31
- post-translational modification 21, 34–36, 64, 79–94, 153
 - carboxylation 81
 - glycosylation 81–86
 - hydroxylation 80
 - lipidation 88–90
 - phosphorylation 87–88
 - prenylation 89
 - sulfatation 91–92
- pox virus growth factors 109
- PP-fold family 107
- pramlintide 509, 510
- prekallikrein 72
- prenylation 88, 89
- Preos 120, 509, 511
- preprodynorphin 136
- preprotein 79
- prepro-protein 80
- pre-sequence 80, 273, 286
- preview analysis 270
- preview synthesis 277
- primary structure 10, 11, 20, 21, 26, 35, 36, 53, 84, 97, 99, 105, 107, 108, 112, 117, 124–127, 133, 135–140, 142, 146, 151, 157–159, 175, 336, 413
- primer 282
- prior capture-mediated ligation 353–355
- proaccelerin 72
- procollagen 80
- proconvertin 71, 72
- proctolin 144
- prodynorphin 132, 134, 136
- proenkephalin 132–134
- proglucagon 99, 100
- prohormone converting enzyme 129
- proinsulin 80, 102, 287
- prolactin 110, 112, 116, 123, 146
- prolactin release-inhibiting hormone or factor 110
- prolactin-releasing hormone or factor 110
- prolactin-releasing peptide 146
- prolactoliberin 110
- prolactostatin 110
- prolyl hydroxylase 80
- proopiomelanocortin (POMC) 80, 117, 134

- proprotein 80
 - protease-catalyzed peptide synthesis 291–295, 297
 - proteasome 93, 537
 - protecting group 179–224, 317–324
 - alkyl type 192–193
 - backbone amide 199, 272, 273
 - C α carboxy 193–199, 223, 318
 - enzyme-labile 220–221, 296, 400
 - intermediary 180
 - orthogonality 182, 223, 268, 322
 - photolabile 261, 465
 - real 195
 - scheme 211, 214, 224, 268, 321–324
 - semipermanent 180–182, 190, 201, 224, 266
 - tactical 181, 321, 323, 324
 - temporary 181, 182, 204, 212, 220, 224, 251, 252, 257, 265, 269, 272, 320, 322, 387, 391, 402
 - protecting schemes 182, 202, 204, 211, 214, 224, 268, 321–324, 396, 495
 - protection
 - guanidino group 202–204
 - hydroxy group 214–216
 - imidazole group 211–214
 - indole group 217–218
 - lysine ϵ -amino group 204, 205
 - ornithine δ -amino group 203, 204
 - side-chain 201
 - methionine thioether group 216, 217
 - ω -amide group 218–220
 - ω -carboxy group 205–208
 - protein
 - analysis 532–535
 - arrays 501
 - ^{15}N -labeled 416
 - membrane-bound 18, 69, 89, 400
 - protein C 495, 513
 - protein S 90, 495
 - therapeutic 485, 487–489, 493, 497
 - protein adsorption 501
 - protein carbonyl 34, 35
 - protein detecting array 501
 - protein domain 85, 518
 - SH2 domain 88
 - SH3 domain 470, 518
 - protein epitope mimetics 369
 - protein folding 2, 48, 55, 81, 82, 178, 325, 413, 417, 439, 442, 443
 - protein function array 501
 - protein inactivation 19
 - protein kinases 87, 99, 102, 121, 330
 - protein-ligand interaction 412, 416, 429
 - protein microarrays 500, 501
 - protein modeling 516
 - protein pharmaceuticals 489, 492–501
 - protein phosphatases 7, 87, 395
 - protein phosphoglycosylation 85
 - protein phosphorylation 87, 88, 395
 - protein–protein interaction 2, 85, 88, 91, 94, 177, 367, 379, 412, 515, 519
 - protein splicing 18, 352, 353, 529
 - protein structure database 54
 - protein-tyrosine kinases 87
 - protein tyrosine phosphatases (PTPase) 88, 538, 539
 - protein tyrosine sulfation 91
 - proteolysis, limited 79, 80, 112, 291
 - proteome 500, 529–540
 - proteome analysis 16, 31, 32, 500, 529–531
 - database search 532, 534
 - differential approach 500, 531
 - proteomics 2, 31, 79, 175, 414, 500, 501, 516, 529–533, 535–537, 543
 - prothrombin 72, 81
 - prothymosin 335, 340, 341
 - protropin 286, 496
 - pseudobiopolymers 431–438
 - pseudoproline 200, 272, 351, 376, 419
 - PTH/PTHrP receptor (PTHR1) 120, 121
 - purification 13–20, 33, 64, 228, 229, 251, 256, 270, 277, 323, 326, 333, 337–339, 341–346, 374, 414, 445, 460, 462, 487, 502, 506, 536, 540
 - purification techniques 17–18, 338
 - PyBOP 218, 239, 269, 280, 396, 433
 - pyroglutamic acid 90, 98, 112
 - pyroglutamyl formation 79, 90–91
 - pyroglutamyl peptidase II 112
 - pyroglutamyl peptides 90, 218
 - pyrrolysine 5, 76
- q**
- quaternary structure 12, 436
 - quantitative proteomics
 - gel-based 533, 540, 541
 - gel-free 533
 - ICAT 533, 543
 - ICPL 543
 - isobaric tag for relative and absolute quantification 543
 - isotope-coded affinity tags 543
 - isotope-coded protein labelling 543
 - iTRAQ 543
 - metabolic stable-isotope labelling 533
 - SILAC 533
 - 8-quinolyl ester 236, 237

R

- racemization 19, 180, 181, 188, 192, 194, 198, 209–211, 213, 214, 216, 225, 226, 228, 233–242, 246–251, 256–258, 261–263, 269, 271, 272, 280, 281, 294, 300, 318–321, 324, 325, 346, 372, 373, 378, 445
- radiolabelled tumor-specific peptides 491
- Ramachandran plot 36, 37
- ranakinin 107
- ranamargarin 107
- ranatachykinin A 107
- random coil 43, 50, 51, 255, 272
- random sampling 416
- random screening 429
- ras proteins 89
- reagent mixture method 468
- receptor 1, 3, 7, 15, 44, 65, 66, 68, 70, 74, 81, 82, 87, 88, 90, 91, 93, 98, 99, 102–109, 111–115, 118–126, 128, 130–146, 152, 156–159, 176, 366–369, 380–399, 400, 402, 411–420, 424, 425, 428–430, 443, 470, 472, 473, 477, 486–491, 498–500, 507, 509, 510, 511–517, 519, 520
- receptor activity modifying protein 123
- receptor down-regulation 65, 419, 509
- receptor mapping 416
- receptor tyrosine kinases 65, 102
- recombinant DNA techniques 281, 285–286, 442
- recombinant growth hormone 496
- recombinant protein 16, 18, 30, 283–285, 351, 352, 394, 400, 485, 492, 493, 495, 497
- recombinant tissue plasminogen activator 495
- red pigment-concentrating hormone 140
- regioselectively addressable functionalized templates 442, 443
- relaxin 102–104, 382
- relaxin-like factor (RLF) 104
- relaxin-like peptide family 102, 103
- release factors 78, 79
- release inhibiting hormones 110, 114, 115
- releasing hormones 110, 112, 129
- remicade 499
- renin 124
- reporter group 536, 537, 540–542
- resin 251–265
 - 2-chlorotriyl 332, 335, 341, 346, 402, 506, 537
 - Barlos 260, 261, 336
 - BHA 260
 - chloromethyl 257
 - HMBA 260
 - HMPB 258
 - HYCRAM 261
 - MBHA 260, 265, 329
 - Merrifield 255, 257, 258
 - oxime 280, 332, 374, 400
 - PAM 258–260, 266
 - Rink amide 260, 261, 432
 - super acid-sensitive 258
 - thioester 346, 374
 - Wang 257, 258, 354, 376
- resin loading 255, 256, 271, 273, 373, 445
- restriction endonuclease 282–285
- retaplast 495
- reverse phase microarrays 501
- reverse thioester ligation 350, 351
- reverse transcriptase 282
- reversed-phase HPLC 14, 15, 18, 29, 34, 333, 338, 341
- RGD 82, 246, 367, 368, 413, 470, 513, 514
- Rh^I catalysis 392
- ribonuclease A 295, 325, 326, 329–331
- ribonuclease S 231, 327, 328
- ribosomal peptide synthesis 75–79, 94–96, 295
- ribosomes 73, 75, 77–79, 288, 291, 297
- ribozyme 281, 291, 295
- ristocetin A 146
- RNA
 - mRNA 5, 75–78, 102, 109, 282–283, 288, 289, 530
 - suppressor tRNA 290
 - tRNA 5, 75–79, 288, 297
- rough endoplasmic reticulum 80
- ruhemann's violet 25

S

- safety-catch linker 262–265, 395
 - alkanesulfonamide 262
 - arenesulfonamide 262
 - DSA 263
 - SCAL 264
- Sakakibara approach 326, 327
- salt coupling 194
- salt-induced peptide formation 75
- SAR by NMR 54, 416
- sarafotoxins 125–126
- saralasin[®] 125
- sarenin 125
- sauvagine 114
- scalar coupling 53
- scatchard plot 65
- scavenger 188, 191, 197, 204, 215–217, 219, 266, 275, 337, 402, 460
- Schlack-Kumpf method 30, 33

- scissile bond 291, 296
- scorpion peptide toxins 158
- scylorhinus I and II 107
- SDS-polyacrylamide gelelectrophoresis (SDS-PAGE) 16, 530, 531
- sea anemone toxins 159
- second messenger 65, 120
- secondary structure 11, 36–37, 42–53, 272–273, 412, 413, 418, 419, 423, 425–426, 442, 444
 - mimetics 425–426
- secretin 66, 97–101, 108, 136, 177, 320, 509, 511
- segment condensation 149, 181, 226, 233, 237, 240, 277, 295, 296, 320–322, 325–330, 332–334, 336, 339–341
- segment coupling, solid support-mediated 332–333
- E-selectin 82–83
- selenocysteine 5, 76, 383, 386
- separation methods 12–16, 530–532
- sequence analysis 20–21, 24–26, 28–35, 54, 281
- serine hydrolase 536, 537
- serine protease 70–72, 297, 536
- serine protecting group 273
- serum albumin 488, 495
- Shaker peptide 158
- β -sheet 39–40, 42–46, 48, 50, 51, 53, 75, 155–157, 176, 272, 277, 367, 423, 438, 439, 442
- Sheppard tactics 267, 323
- sialyl-Lewis X 82, 83, 390
- side-chain modification 346, 418–421
- signal deconvolution 51
- signal hypothesis 79
- signal peptidase 80, 102, 129, 286
- signal peptide 79, 80, 98, 129
 - signal sequence-based peptides 102, 103
- silica 255, 256
- silk fibroin 39, 71
- silyl derivatives 473
- SiMb 200, 272
- simulated annealing 416
- simulect 499
- site-directed mutagenesis 415–416
- site-specific drug delivery 489
- sleeper peptide 158
- solid-liquid-phase peptide synthesis 279
- solid-phase chemical ligation 350
- solid-phase peptide synthesis 3, 4, 175, 181, 190, 204, 251–253, 266–268, 270, 280, 339–341, 393, 399, 445, 460, 504
 - batchwise synthesis 274
 - continuous flow mode 274
 - convergent 339–341
 - linear 277, 278, 319, 332, 339–341
 - stepwise 181, 338, 339, 445
- solid phase sequencer 29
- solid-to-solid conversion 294
- soluble handle approach 280, 341
- solution phase/solid phase hybrid
 - approach 332–334
 - lipophilic 333–334
- solution phase synthesis 199, 252, 322, 323, 325, 326, 327, 337, 339, 341, 402, 460, 507
- somatoliberein 100, 110, 114
- somatomedins 116
- somatostatin 104–105, 110, 115, 142, 285, 286, 338, 368, 382, 420, 424, 425, 436, 488, 491, 512
- somatostatin family 104–105
- somatotropin 110, 114–116, 285, 518
 - release inhibiting peptide 110, 112, 114, 115, 117
- releasing peptides 130, 146, 417, 483
- sortase-mediated ligation 359
- spatial screening 367–369, 425
- spatially addressable parallel synthesis 465–466
- spider peptide toxins 159
- spinning-cup sequencer 29
- split and combine method 468–471
- split intein 378–379
- split intein-mediated circular ligation 378–379
- spot synthesis 461, 464
- SPS/SPPS-hybrid approach 332–336, 506
- Src homology 2 (SH2) domain 88
- Src homology 3 (SH3) domain 470, 518
- stability problems 19–20
- statine 428
- statins 110
- statistical method 415
- Staudinger ligation 357, 379
- stereochemical product analysis 249–251
- streptogranin 147
- streptokinase 495
- structural diversity 95, 148, 457, 458
- structure prediction 48
- structure-based molecular design 460
- Stuart factor 71, 72
- subproteome 535, 538, 541, 542
- substance P 105–106, 130
- substrate mimetic approach 295–297, 330, 331, 358
- subtiligase 295, 329, 330
- subtilin 153

- subtilisin 26, 294, 295, 329, 390, 391
- succinimide formation 276
- sugar-assisted ligation 347
- sulfation 91–92
- sulfated peptide 220, 402
- sumoylation 93, 94
- superoxide dismutase 495
- supersecondary structure 46
- surface complementary matching 416
- surface plasmon resonance 49, 501, 540, 541
- surfactin 151, 365
- switch peptide 273
- Symlin 509, 510
- syndecan-3 118
- synthetic adjuvants 399
- synthesis strategy 343, 505
- synthesis tactics 201
- t**
- T20, *see* enfuvirtide
- tachykinins 105–107, 123, 129, 131, 417
- tachykinin family 105–106
- tachykinin receptors 106
- tachyplesins 156
- tagging 471, 472, 473, 531, 533, 534, 539, 542
 - cICAT 534
 - cleavable isotope-coded affinity tagging 534
- tandem mass spectrometry 31, 531–532
- target validation 517–518, 522
- targeted diversity approach 460
- tat fragment 490
- TBTU 210, 240, 241, 269, 320, 373
- teabag synthesis 461–462
- teduglutide 510
- teicoplanin 398
- telavancin 513
- template-associated synthetic proteins (TASP) 441–443
- template-mediated ligation 353–355
- teprotide 430
- teriparatide 509, 511
- terlipressin 509
- tertiary fold 46, 47, 439
- tertiary structure 11, 13, 36, 40, 44–48, 178, 439, 442
- therapeutic peptide engineering 483–486
- thermolysin 33, 294
- thiocarboxy segment condensation 328
- thioester-forming ligation 355
- thiol capture ligation 354
- thiol protecting group 208, 210, 211, 383
- thiol protection 208–211
- thionin 75
- thiopeptin 147
- thiophenyl ester 235
- thiostrepton 147, 148
- thiotemplate mechanism 94, 365
- three-dimensional structure 11, 36–49
- threonine hydroxy group protection 214–216
- thrombin 26, 71, 72, 81, 422, 470, 485
- thrombin receptor activator peptide 71
- thrombogen 72
- thromboplasmin 72
- thymalfasin 509, 513
- thymopentin 509
- thymopoietin 367
- thymosins 278, 509, 513
- thyroxine 112
- thyro-liberase 112, 115
- thyroliberin 90, 110, 112, 115, 509
- thyrotropin 104, 110, 112, 113, 116, 117, 512
- thyrotropin-releasing hormone 110
- tifludom 430
- tirofiban 513
- tissue factor 71, 72, 511
- tissue factor pathway inhibitor (TFPI) 511
- tissue plasminogen activator 386, 493, 495
- Tmb 208, 210, 220, 346
- TNBS test 274
- 4-toluolsulfonyl group 192
- torsion angle 6, 36–39, 42, 44, 53, 415, 419–421
- toxic peptides 73, 156, 161, 491
- toxin 157–159, 336, 372
- toxin II 158
- toxin V 158
- toxin Sh-I 159
- transcreeper HTS assay platform 518, 519
- transcription 76, 88, 92, 282, 283, 434, 520, 529
- transcriptome 529
- transfection 282, 283, 285
- transfer RNA 5, 75–79, 94, 290, 297
- transferred NOE 54
- transferrin 489
- transforming growth factor α 109, 495
- transforming growth factor β 495
- transition state 236, 238, 298, 299, 427
- transition-state analogue 298, 299
- transition state inhibitor 427, 428
- translation 76, 79, 175, 288, 289
- translation system 288, 289
- transportan 490
- trapoxin A 372
- trapoxin B 372
- triflavin 34

- trifluoroacetyl 192, 205, 247
 - 2,2,2-trifluoroethanol 50, 272, 326
 - trigger factor 65
 - triiodothyronine 112
 - 2,4,6-trimethylbenzenesulfonyl 203
 - triphenylphosphine 205
 - trityl (Trit) 202, 208, 210, 213, 220, 238, 261, 320, 323, 332, 346, 400
 - N-tritylhistidine 237
 - Trojan horse peptides, *see* cell-penetrating peptides
 - tropomyosin 142
 - TROSY 52
 - trypsin 26, 31, 34, 35, 129, 294, 295, 338, 348, 370, 532, 534
 - tryptophan protecting group 187, 188, 204, 210, 217–218
 - tumor-associated antigens 521
 - tumor necrosis factor- α 495, 499
 - tumor necrosis factor- β 494
 - turn 37, 40–42
 - α -turn 41
 - β -turn 41, 42, 53, 157, 367, 377, 419, 421, 423, 425, 426, 439
 - γ -turn 41, 42, 367, 426
 - π -turn 41
 - turn mimetics 425, 426, 430
 - two-dimensional polyacrylamide gel-electrophoresis 530–531
 - tyrocidine A 365
 - tyrocidins 96, 146, 365, 374
 - tyrosine hydroxy group protection 214–216
 - tyrosyl protein sulfotransferase 91, 92, 98
- u**
- ubiquitin 80, 94
 - ubiquitinylation 94
 - ultrafiltration 14, 16, 278, 289, 530
 - ultrahigh-throughput screening 521
 - UNCA 231
 - uperin 107
 - uperolein 107
 - urethane-type protecting group 182, 221, 223, 248
 - urocortin/urotensin 1 114
 - urokinase 495
 - uromodulin 495
 - uronium reagent 210, 240–243
 - urotensin 114
- v**
- V8 protease 26, 358
 - valinomycin 149
 - vancomycin 146, 153, 509, 513
 - vapreotide 104, 512
 - vasoactive intestinal contractor 126
 - vasoactive intestinal peptide (VIP) 99, 100
 - vasopressin 64, 116, 118–120, 123, 125, 130, 131, 158, 418, 424, 488, 571
 - vasotocin 280
 - vector 282–283, 351, 492
 - vessel dilator 128
 - viomycin 146
- w**
- wobble hypothesis 77
 - W(X)₆wamides 142
- x**
- xanthenyl resin 332, 335
 - X-ray crystallography 6, 49, 51, 54–55, 414, 438
 - cyclotron beam 54
- z**
- Zadaxin 509, 513
 - zein 341
 - γ -zein 341
 - zenapax 499
 - ziconotide 509, 512
 - zinc finger 44
 - zymogen 72, 80