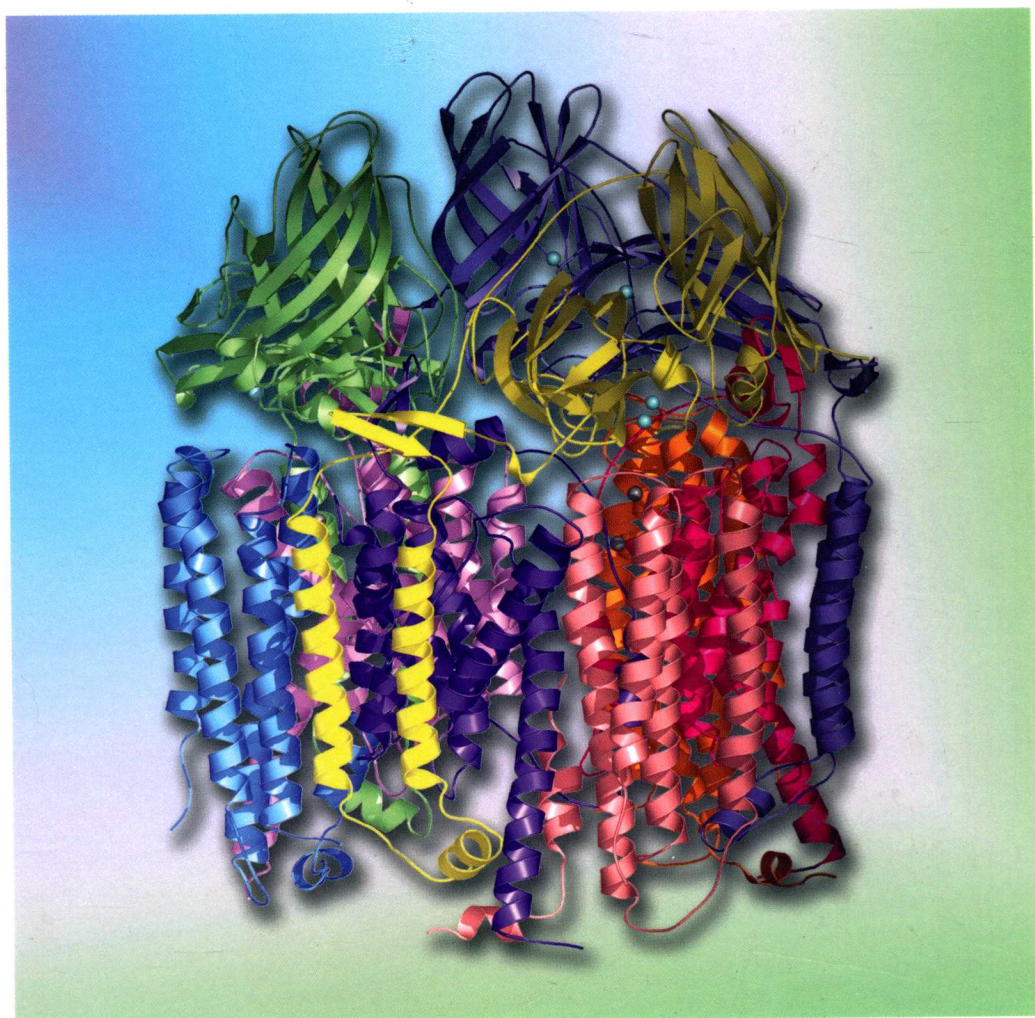


Specialist Periodical Reports

Edited by E Farkas and M Ryadnov

Amino Acids, Peptides and Proteins

Volume 37



RSCPublishing

A Specialist Periodical Report

Amino Acids, Peptides and Proteins

Volume 37

Editors

Etelka Farkas, *University of Debrecen, Hungary*

Maxim Ryadnov, *National Physical Laboratory and University of Edinburgh, UK*

Authors

Luc Brunsveld, *Technische Universiteit Eindhoven, The Netherlands*

Jason Crain, *University of Edinburgh, UK*

Jonathan G. Heddle, *RIKEN Advanced Science Institute, Japan*

Ferenc Hudecz, *Hungarian Academy of Sciences, Hungary*

Nikolett Mihala, *Hungarian Academy of Sciences, Hungary*

Lech-Gustav Milroy, *Technische Universiteit Eindhoven, The Netherlands*

Lidia Nieto, *Technische Universiteit Eindhoven, The Netherlands*

Imre Sóvágó, *University of Debrecen, Hungary*

Jeremy R. H. Tame, *Yokohama City University, Japan*

Shuguang Zhang, *Massachusetts Institute of Technology, USA*



RSC Publishing

If you buy this title on standing order, you will be given FREE access to the chapters online. Please contact sales@rsc.org with proof of purchase to arrange access to be set up.

Thank you

ISBN: 978-1-84973-406-6

ISSN: 1361-5904

DOI: 10.1039/9781849734677

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

© The Royal Society of Chemistry 2012

All rights reserved

Apart from fair dealing for the purposes of research for non-commercial purposes or for private study, criticism or review, as permitted under the Copyright, Designs and Patents Act 1988 and the Copyright and Related Rights Regulations 2003, this publication may not be reproduced, stored or transmitted, in any form or by any means, without the prior permission in writing of The Royal Society of Chemistry, or in the case of reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of the licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to The Royal Society of Chemistry at the address printed on this page.

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Printed and bound in Great Britain by CPI Group (UK) Ltd,
Croydon, CR0 4YY

Amino Acids, Peptides and Proteins

Volume 37

Preface

Etelka Farkas^a and Maxim Ryadnov^b

DOI: 10.1039/9781849734677-**FP005**

Since the last publication of the series in 2007 there has been a considerable progress in elucidating and exploiting protein structure at all length scales ranging from individual amino acids to nanostructured protein-based materials. Traditionally, the series provides a compilation of the most recent findings and developments in the field and reviews literature predominantly published over the last two years. This volume is not meant to be an exception. However, given the five-year break in the publishing of the series it was deemed appropriate to revisit traditional concepts in the light of recent discoveries and breakthroughs. In a sense, this volume can be viewed by our reader as a re-launch of the series in response to increasing dynamics, volume and frequency of publishing. Partly for this reason, the volume does not aim to provide an exhaustive coverage of all research trends in the area of amino acids, peptides and proteins. Instead, the presented selection of chapters is designed to touch key and contemporary points, within which scientific advances appear to be most apparent.

The volume opens with an overview of amino-acid and peptide bio-conjugation including synthetic, semi-synthetic and chemoselective approaches (Mihala and Hudecz). Following this, peptide research is reviewed over three subsequent chapters starting from a prescriptive or de novo peptide design setting out from basic principles of protein folding and topology (Ryadnov), through rational exploitation of local protein-protein interactions with a pharmaceutically relevant case study of targeting nuclear receptors (Milroy, Nieto and Brunsveld), to the role of biophysical model systems in the elucidation of tertiary contacts in peptides and proteins at the atomistic level (Crain). One chapter provides an overview of metal complexation by amino acids, peptides and their derivatives. A particular emphasis is made on the role of metal complexation in biochemical reactions and biologically relevant ligand binding (Farkas and Sóvágó). The volume closes at the protein side of the spectrum with two chapters highlighting the application of peptide self-assembly for the fabrication of novel nanostructured materials (Zhang) and the construction principles of protein self-assembled architectures including protein nanotubes, channels and cages and their use and impact in nanotechnology (Hedde and Tame).

Each chapter incorporates current trends of the reviewed topic and the authors' outlook of future perspectives. This is to facilitate the monitoring of the covered areas and their potential expansion with the inclusion of

^aUniversity of Debrecen, Department of Inorganic and Analytical Chemistry, Debrecen, Hungary. E-mail: efarkas@science.unideb.hu

^bNational Physical Laboratory, Teddington, TW11 0LW, UK, and School of Physics and Astronomy, University of Edinburgh, Mayfield Road, EH9 3JZ, UK.
E-mail: max.ryadnov@npl.co.uk

other specialist reports in subsequent volumes. All chapters are compiled by leading researchers in their subject areas which offers this series as an appealing source of information for the research community in both academia and industry.

A short guide to abbreviations and their use in peptide science

Abbreviations, acronyms and symbolic representations are very much part of the language of peptide science – in conversational communication as much as in its literature. They are not only a convenience, either – they enable the necessary but distracting complexities of long chemical names and technical terms to be pushed into the background so the wood can be seen among the trees. Many of the abbreviations in use are so much in currency that they need no explanation. The main purpose of this editorial is to identify them and free authors from the hitherto tiresome requirement to define them in every paper. Those in the tables that follow – which will be updated from time to time – may in future be used in this Journal without explanation.

All other abbreviations should be defined. Previously published usage should be followed unless it is manifestly clumsy or inappropriate. Where it is necessary to devise new abbreviations and symbols, the general principles behind established examples should be followed. Thus, new amino-acid symbols should be of form *Abc*, with due thought for possible ambiguities (Dap might be obvious for diaminopropionic acid, for example, but what about diaminopimelic acid?).

Where alternatives are indicated below, the first is preferred.

Amino Acids

Proteinogenic Amino Acids

Ala	Alanine	A
Arg	Arginine	R
Asn	Asparagine	N
Asp	Aspartic acid	D
Asx	Asn or Asp	
Cys	Cysteine	C
Gln	Glutamine	Q
Glu	Glutamic acid	E
Glx	Gln or Glu	
Gly	Glycine	G
His	Histidine	H
Ile	Isoleucine	I
Leu	Leucine	L
Lys	Lysine	K
Met	Methionine	M
Phe	Phenylalanine	F
Pro	Proline	P
Ser	Serine	S
Thr	Threonine	T
Trp	Tryptophan	W
Tyr	Tyrosine	Y
Val	Valine	V

Copyright © 1999 European Peptide Society and John Wiley & Sons, Ltd. Reproduced with permission from *J. Peptide Sci.*, 1999, **5**, 465–471.

Other Amino Acids

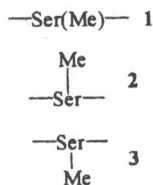
Aad	α -Aminoadipic acid
β Aad	β -Aminoadipic acid
Abu	α -Aminobutyric acid
Aib	α -Aminoisobutyric acid; α -methylalanine
β Ala	β -Alanine; 3-aminopropionic acid (avoid Bal)
Asu	α -Aminosuberic acid
Aze	Azetidine-2-carboxylic acid
Cha	β -cyclohexylalanine
Cit	Citrulline; 2-amino-5-ureidovaleric acid
Dha	Dehydroalanine (also Δ Ala)
Gla	γ -Carboxyglutamic acid
Glp	pyroglutamic acid; 5-oxoproline (also pGlu)
Hph	Homophenylalanine (Hse = homoserine, and so on). Caution is necessary over the use of the prefix homo in relation to α -amino-acid names and the symbols for homo-analogues. When the term first became current, it was applied to analogues in which a side-chain CH_2 extension had been introduced. Thus homoserine has a side-chain $\text{CH}_2\text{CH}_2\text{OH}$, homoarginine $\text{CH}_2\text{CH}_2\text{CH}_2\text{NHC}(=\text{NH})\text{NH}_2$, and so on. In such cases, the convention is that a new three-letter symbol for the analogue is derived from the parent, by taking H for homo and combining it with the first two characters of the parental symbol – hence, Hse, Har and so on. Now, however, there is a considerable literature on β -amino acids which are analogues of α -amino acids in which a CH_2 group has been inserted between the α -carbon and carboxyl group. These analogues have also been called homo-analogues, and there are instances for example not only of ‘homophenylalanine’, $\text{NH}_2\text{CH}(\text{CH}_2\text{CH}_2\text{Ph})\text{CO}_2\text{H}$, abbreviated Hph, but also ‘homophenylalanine’, $\text{NH}_2\text{CH}(\text{CH}_2\text{Ph})\text{CH}_2\text{CO}_2\text{H}$ abbreviated Hph.
	Further, members of the analogue class with CH_2 interpolated between the α -carbon and the carboxyl group of the parent α -amino acid structure have been called both ‘ α -homo’- and ‘ β -homo’. Clearly great care is essential, and abbreviations for ‘homo’ analogues ought to be fully defined on every occasion. The term ‘ β -homo’ seems preferable for backbone extension (emphasizing as it does that the residue has become a β -amino acid residue), with abbreviated symbolism as illustrated by βHph for $\text{NH}_2\text{CH}(\text{CH}_2\text{Ph})\text{CH}_2\text{CO}_2\text{H}$.
Hyl	δ -Hydroxylysine
Hyp	4-Hydroxyproline
α Ile	<i>allo</i> -Isoleucine; 2S, 3R in the L-series
Lan	Lanthionine; S-(2-amino-2-carboxyethyl)cysteine
MeAla	N-Methylalanine (MeVal = N-methylvaline, and so on). This style should not be used for α -methyl residues, for which either a separate unique symbol (such as Aib for α -methylalanine) should be used, or the position of the methyl

	group should be made explicit as in α MeTyr for α -methyltyrosine.
Nle	Norleucine; α -aminocaproic acid
Orn	Ornithine; 2,5-diaminopentanoic acid
Phg	Phenylglycine; 2-aminophenylacetic acid
Pip	Pipecolic acid; piperidine-s-carboxylic acid
Sar	Sarcosine; <i>N</i> -methylglycine
Sta	Statine; (3 <i>S</i> , 4 <i>S</i>)-4-amino-3-hydroxy-6-methyl-heptanoic acid
Thi	β -Thienylalanine
Tic	1,2,3,4-Tetrahydroisoquinoline-3-carboxylic acid
α Thr	<i>allo</i> -Threonine; 2 <i>S</i> , 3 <i>S</i> in the L-series
Thz	Thiazolidine-4-carboxylic acid, thiaproline
Xaa	Unknown or unspecified (also Aaa)

The three-letter symbols should be used in accord with the IUPAC-IUB conventions, which have been published in many places (*e.g.* *European J. Biochem.* 1984; **138**: 9–37), and which are (May 1999) also available with other relevant documents at: <http://www.chem.qnw.ac.uk/iubmb/iubmb.html#03>

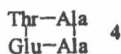
It would be superfluous to attempt to repeat all the detail which can be found at the above address, and the ramifications are extensive, but a few remarks focussing on common misuses and confusions may assist. The three-letter symbol standing alone represents the unmodified intact amino acid, of the L-configuration unless otherwise stated (but the L-configuration may be indicated if desired for emphasis: *e.g.* L-Ala). The same three-letter symbol, however, also stands for the corresponding amino acid *residue*. The symbols can thus be used to represent peptides (*e.g.* AlaAla or Ala-Ala = alanylalanine). When nothing is shown attached to either side of the three-letter symbol it is meant to be understood that the amino group (always understood to be on the left) or carboxyl group is unmodified, but this can be emphasized, so AlaAla = H-AlaAla-OH. Note however that indicating free termini by presenting the terminal group in full is wrong; NH₂AlaAlaCO₂H implies a hydrazino group at one end and an α -keto acid derivative at the other. Representation of a free terminal carboxyl group by writing H on the right is also wrong because that implies a 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. Thus an *O*-methylserine residue could be shown as **1**, **2**, or **3**.



Note that the oxygen atom is not shown: it is contained in the three-letter symbol – showing it, as in Ser(OMe), would imply that a peroxy group was

present. Bonds up or down should be used only for indicating side-chain substitution. Confusions may creep in if the three-letter symbols are used thoughtlessly in representations of cyclic peptides. Consider by way of example the hypothetical cyclopeptide threonylalanylalanylglutamic acid. It might be thought that this compound could be economically represented 4.



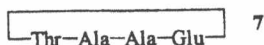
But this is wrong because the left hand vertical bond implies an ester link between the two side chains, and strictly speaking if the right hand vertical bond means anything it means that the two Ala α -carbons are linked by a CH_2CH_2 bridge. This objection could be circumvented by writing the structure as in 5.



But this is now ambiguous because the convention that the symbols are to be read as having the amino nitrogen to the left cannot be imposed on both lines. The direction of the peptide bond needs to be shown with an arrow pointing from CO to N, as in 6.



Actually the simplest representation is on one line, as in 7.



Substituents and Protecting Groups

Ac	Acetyl
Acm	Acetamidomethyl
Adoc	1-Adamantyloxycarbonyl
Alloc	Allyloxycarbonyl
Boc	<i>t</i> -Butoxycarbonyl
Bom	π -Benzyloxymethyl
Bpoc	2-(4-Biphenyl)isopropoxycarbonyl
Btm	Benzylthiomethyl
Bum	π - <i>t</i> -Butoxymethyl
Bu ⁱ	<i>i</i> -Butyl
Bu ⁿ	<i>n</i> -Butyl
Bu ^t	<i>t</i> -Butyl
Bz	Benzoyl
Bzl	Benzyl (also Bn); Bzl(OMe) = 4-methoxybenzyl and so on
Cha	Cyclohexylammonium salt
Clt	2-Chlorotriyl
Dcha	Dicyclohexylammonium salt
Dde	1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl
Ddz	2-(3,5-Dimethoxyphenyl)-isopropoxycarbonyl

Dnp	2,4-Dinitrophenyl
Dpp	Diphenylphosphinyl
Et	Ethyl
Fmoc	9-Fluorenylmethoxycarbonyl
For	Formyl
Mbh	4,4'-Dimethoxydiphenylmethyl, 4,4'-Dimethoxybenzhydryl
Mbs	4-Methoxybenzenesulphonyl
Me	Methyl
Mob	4-Methoxybenzyl
Mtr	2,3,6-Trimethyl,4-methoxybenzenesulphonyl
Nps	2-Nitrophenylsulphenyl
OA11	Allyl ester
OBt	1-Benzotriazolyl ester
OcHx	Cyclohexyl ester
ONp	4-Nitrophenyl ester
OPcp	Pentachlorophenyl ester
OPfp	Pentafluorophenyl ester
OSu	Succinimido ester
OTce	2,2,2-Trichloroethyl ester
OTcp	2,4,5-Trichlorophenyl ester
Tmob	2,4,5-Trimethoxybenzyl
Mtt	4-Methyltrityl
Pac	Phenacyl, PhCOCH_2 (care! Pac also = PhCH_2CO)
Ph	Phenyl
Pht	Phthaloyl
Scm	Methoxycarbonylsulphenyl
Pmc	2,2,5,7,8-Pentamethylchroman-6-sulphonyl
Pr ⁱ	<i>i</i> -Propyl
Pr ⁿ	<i>n</i> -Propyl
Tfa	Trifluoroacetyl
Tos	4-Toluenesulphonyl (also Ts)
Troc	2,2,2-Trichloroethoxycarbonyl
Trt	Trityl, triphenylmethyl
Xan	9-Xanthrydryl
Z	Benzoyloxycarbonyl (also Cbz). Z(2C1) = 2-chlorobenzyl-oxycarbonyl and so on

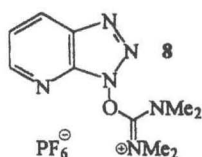
Amino Acid Derivatives

DKP	Diketopiperazine
NCA	<i>N</i> -Carboxyanhydride
PTH	Phenylthiohydantoin
UNCA	Urethane <i>N</i> -carboxyanhydride

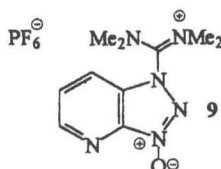
Reagents and Solvents

BOP	1-Benzotriazolylloxy-tris-dimethylamino-phosphonium hexafluorophosphate
CDI	Carbonyldiimidazole
DBU	Diazabicyclo[5.4.0]-undec-7-ene
DCCI	Dicyclohexylcarbodiimide (also DCC)
DCHU	Dicyclohexylurea (also DCU)

DCM	Dichloromethane
DEAD	Diethyl azodicarboxylate (DMAD = the dimethyl analogue)
DIPCI	Diisopropylcarbodiimide (also DIC)
DIPEA	Diisopropylethylamine (also DIEA)
DMA	Dimethylacetamide
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMS	Dimethylsulphide
DMSO	Dimethylsulphoxide
DPAA	Diphenylphosphoryl azide
EEDQ	2-Ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline
HATU	This is the acronym for the 'uronium' coupling reagent derived from HOAt, which was originally thought to have the structure 8 , the <i>Hexafluorophosphate</i> salt of the <i>O</i> -(7- <i>Azabenzotriazol-yl</i>)- <i>Tetramethyl Uronium</i> cation.



In fact this reagent has the isomeric *N*-oxide structure **9** in the crystalline state, the unwieldy correct name of which does not conform logically with the acronym, but the acronym continues in use.



Similarly, the corresponding reagent derived from HOBt has the firmly attached label HBTU (the tetrafluoroborate salt is also used: TBTU), despite the fact that it is not actually a uronium salt.

HMP	Hexamethylphosphoric triamide (also HMPA, HMPTA)
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	1-Hydroxybenzotriazole
HOCT	1-Hydroxy-4-ethoxycarbonyl-1,2,3-triazole
NDMBA	<i>N,N'</i> -Dimethylbarbituric acid
NMM	<i>N</i> -Methylmorpholine
PAM	Phenylacetamidomethyl resin
PEG	Polyethylene glycol
PtBOP	1-Benzotriazolyl-oxy-tris-pyrrolidinophosphonium hexafluorophosphate
SDS	Sodium dodecyl sulphate
TBAF	Tetrabutylammonium fluoride

TBTU	See remarks under HATU above
TEA	Triethylamine
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TFMSA	Trifluoromethanesulphonic acid
THF	Tetrahydrofuran
WSCl	Water soluble carbodiimide: 1-ethyl-3-(3'-dimethylamino-propyl)-carbodiimide hydrochloride (also EDC)

Techniques

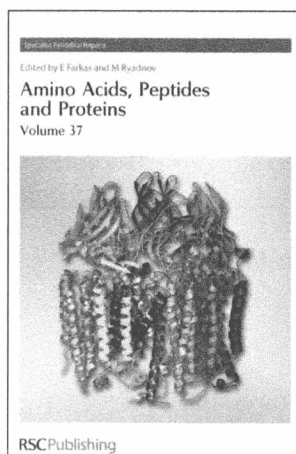
CD	Circular dichroism
COSY	Correlated spectroscopy
CZE	Capillary zone electrophoresis
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
ESR	Electron spin resonance
FAB	Fast atom bombardment
FT	Fourier transform
GLC	Gas liquid chromatography
hplc	High performance liquid chromatography
IR	Infra red
MALDI	Matrix-assisted laser desorption ionization
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
nOe	Nuclear Overhauser effect
NOESY	Nuclear Overhauser enhanced spectroscopy
ORD	Optical rotatory dispersion
PAGE	Polyacrylamide gel electrophoresis
RIA	Radioimmunoassay
ROESY	Rotating frame nuclear Overhauser enhanced spectroscopy
RP	Reversed phase
SPPS	Solid phase peptide synthesis
TLC	Thin layer chromatography
TOCSY	Total correlation spectroscopy
TOF	Time of flight
UV	Ultraviolet

Miscellaneous

Ab	Antibody
ACE	Angiotensin-converting enzyme
ACTH	Adrenocorticotrophic hormone
Ag	Antigen
AIDS	Acquired immunodeficiency syndrome
ANP	Atrial natriuretic polypeptide
ATP	Adenosine triphosphate
BK	Bradykinin
BSA	Bovine serum albumin
CCK	Cholecystokinin
DNA	Deoxyribonucleic acid
FSH	Follicle stimulating hormone

GH	Growth hormone
HIV	Human immunodeficiency virus
LHRH	Luteinizing hormone releasing hormone
MAP	Multiple antigen peptide
NPY	Neuropeptide Y
OT	Oxytocin
PTH	Parathyroid hormone
QSAR	Quantitative structure–activity relationship
RNA	Ribonucleic acid
TASP	Template-assembled synthetic protein
TRH	Thyrotropin releasing hormone
VIP	Vasoactive intestinal peptide
VP	Vasopressin

J. H. Jones



Cover

The crystal structure of particulate methane monooxygenase (pMMO) reveals many unexpected features including, a trimeric oligomerization state and three distinct metal centers. Image reproduced by permission of Amy Rosenzweig from *Dalton Transactions*, 2005.

Preface

v

Etelka Farkas and Maxim Ryadnov

Amino acid and peptide bioconjugates

1

Nikolett Mihala and Ferenc Hudecz

Introduction	1
Current trends	1
1 Amide bond forming reactions	2
2 Reactions yielding a non-amide bond	4
3 PEGylation	22
4 Crosslinking agents	22
5 Selected applications of peptide-bioconjugates	24
References	31

Self-assembling peptide materials

40

Shuguang Zhang

1 Introduction	40
2 General self-assembling peptide materials	47
3 Diverse uses of self-assembling peptide nanofibers	51
4 Surface modification self-assembling peptides	55
5 Lipid-like self-assembling peptide surfactant materials	55
6 Summary	61
References	62

Metal complexes of amino acids and peptides	66
--	-----------

Etelka Farkas and Imre Sóvágó

1	Introduction	66
2	Amino acid complexes	68
3	Peptide Complexes	85
	References	105

Model systems for folding and tertiary contacts in peptides: A perspective from the physical sciences	119
--	------------

Jason Crain

1	Introduction	119
2	Thermodynamic basics of liquid mixtures and hydrophobicity	120
3	Current trends – in what state is the art?	124
4	Model systems for self-organization	131
5	Future perspectives	144
	References	146

Protein nanotubes, channels and cages	151
--	------------

Jonathan G. Heddle and Jeremy R. H. Tame

1	Introduction	151
2	Current trends	154
3	Nanofibres (Amyloid-like fibrils)	155
4	Nanotubes	157
5	Cage proteins	168
6	Channel proteins	178
7	Examples of useful principles and methods	179
8	<i>Ab initio</i> design	182
9	Future perspective and applications	182
	References	183

Prescriptive peptide design	190
------------------------------------	------------

Maxim G Ryadnov

1	Introduction	190
2	Generic considerations for prescriptive peptide design	190
3	Structural space of de novo design	194
4	Current trends: structural rationales	199
5	Current trends: de novo motifs	211
6	Future perspectives	229
	References	230

**Targeting alpha-helix based protein interactions; nuclear receptors
as a case study** **238**

Lech-Gustav Milroy, Lidia Nieto and Luc Brunsveld

1	Introduction	238
2	Nuclear receptors	243
3	α -Helix-inspired drug discovery	254
	Conclusions	265
	References	266