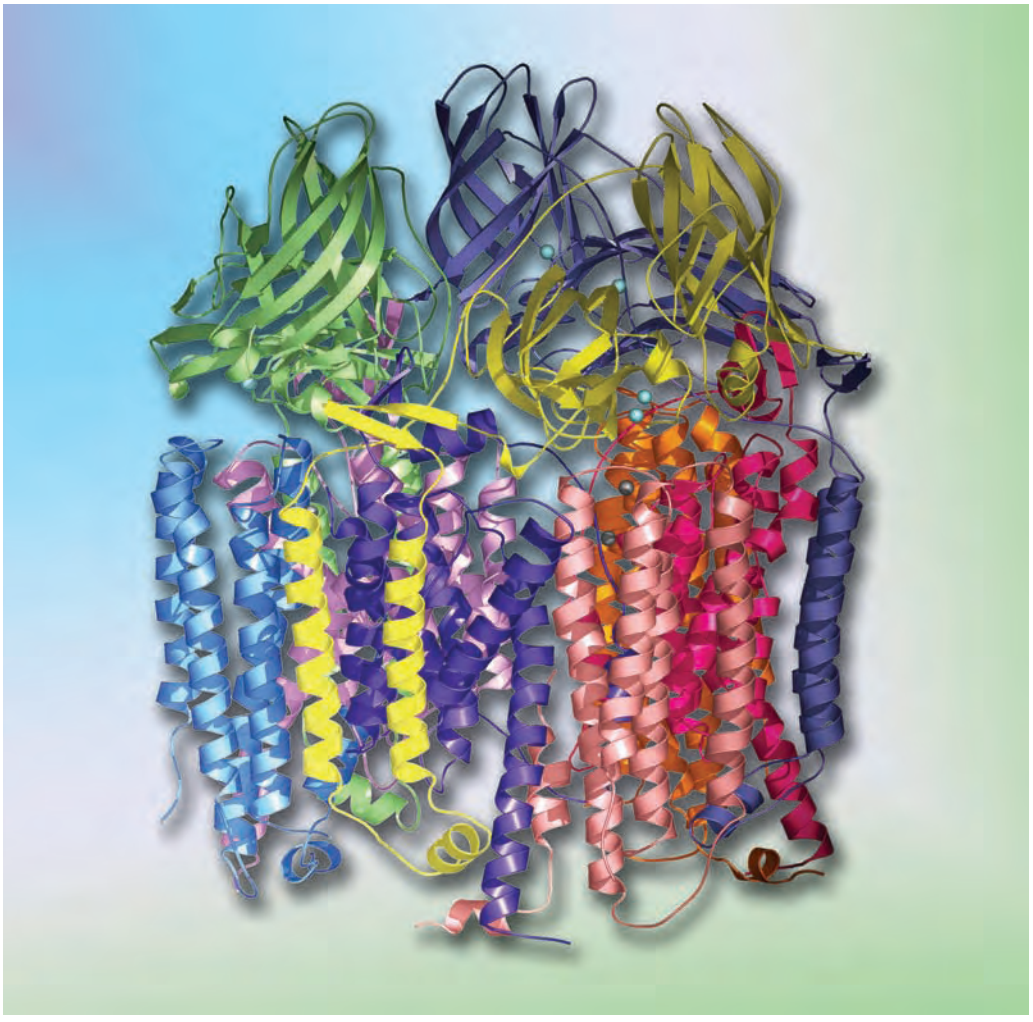
Specialist Periodical Reports

Edited by E Farkas and M Ryadnov

Amino Acids, Peptides and Proteins

Volume 37



RSC Publishing

Amino Acids, Peptides and Proteins

Volume 37

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

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

Preface

Etelka Farkas[a] and Maxim Ryadnov[b]
DOI: 10.1039/9781849734677-FP005

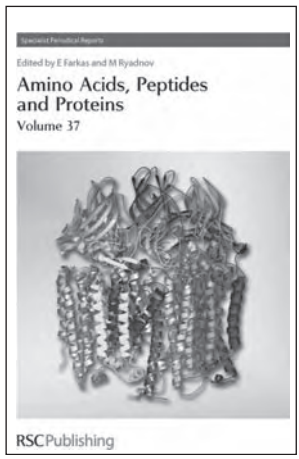
Since the last publication of the series in 2007 there has been a considerable progress in elucidating and exploiting protein structure at all length scales ranging from individual amino acids to nanostructured protein-based materials. Traditionally, the series provides a compilation of the most recent findings and developments in the field and reviewes 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-lauch 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 bioconjugation including synthetic, semi-synthetic and chemoselective approaches (Mihala and Hudecz). Following this, peptide research is reviewed over three subsequent chapters starting from a presciptive or de novo peptide design setting out from basic principles of protein folding and topology (Ryadnov), through rational exploitaion 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 empahsis 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 applicaiton 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 (Heddle 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

[a]*University of Debrecen, Department of Inorganic and Analytical Chemistry, Debrecen, Hungary. E-mail: efarkas@science.unideb.hu*
[b]*National 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.



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.

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 diaminoproprionic 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

Other Amino Acids

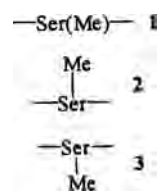
Aad	α-Aminoadipic acid
βAad	β-Aminoadipic acid
Abu	α-Aminobutyric acid
Aib	α-Aminoisobutyric acid; α-methylalanine
βAla	β-Alanine; 3-aminopropionic acid (avoid Bal)
Asu	α-Aminosuberic acid
Aze	Azetidine-2-carboxylic acid
Cha	β-cyclohexylalanine
Cit	Citrulline; 2-amino-5-ureidovaleric acid
Dha	Dehydroalanine (also ΔAla)
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 CH_2CH_2OH, homoarginine $CH_2CH_2CH_2NHC(=NH)NH_2$, 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', $NH_2CH(CH_2CH_2Ph)CO_2H$, abbreviated Hph, but also 'homophenylalanine', $NH_2CH(CH_2Ph)CH_2CO_2H$ 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 $NH_2CH(CH_2Ph)CH_2CO_2H$.
Hyl	δ-Hydroxylysine
Hyp	4-Hydroxyproline
αIle	*allo*-Isoleucine; 2*S*, 3*R* 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 αMeTyr 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; *N*-methylglycine
Sta	Statine; (3*S*, 4*S*)-4-amino-3-hydroxy-6-methyl-heptanoic acid
Thi	β-Thienylalanine
Tic	1,2,3,4-Tetrahydroisoquinoline-3-carboxylic acid
αThr	*allo*-Threonine; 2*S*, 3*S* 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

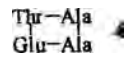
It would be superfluous to attempt to repeat all the detail which can be found at the above address, and the ramifications are extensive, but a few remarks focussing on common misuses and confusions may assist. The three-letter symbol standing alone represents the unmodified intact amino acid, of the 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; $NH_2AlaAlaCO_2H$ 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**.



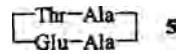
Note that the oxygen atom is not shown: it is contained in the three-letter symbol – showing it, as in Ser(OMe), would imply that a peroxy group was

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.



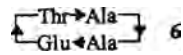
4

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**.



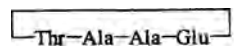
5

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**.



6

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



7

Substituents and Protecting Groups

Ac	Acetyl
Acm	Acetamidomethyl
Adoc	1-Adamantyloxycarbonyl
Alloc	Allyloxycarbonyl
Boc	*t*-Butoxycarbonyl
Bom	π-Benzyloxymethyl
Bpoc	2-(4-Biphenylyl)isopropoxycarbonyl
Btm	Benzylthiomethyl
Bum	π-*t*-Butoxymethyl
Bu^i	*i*-Butyl
Bu^n	*n*-Butyl
Bu^t	*t*-Butyl
Bz	Benzoyl
Bzl	Benzyl (also Bn); Bzl(OMe) = 4-methoxybenzyl and so on
Cha	Cyclohexylammonium salt
Clt	2-Chlorotrityl
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*-Propyl
Pr^n	*n*-Propyl
Tfa	Trifluoroacetyl
Tos	4-Toluenesulphonyl (also Ts)
Troc	2,2,2-Trichloroethoxycarbonyl
Trt	Trityl, triphenylmethyl
Xan	9-Xanthydryl
Z	Benzyloxycarbonyl (also Cbz). Z(2C1) = 2-chlorobenzyl-oxycarbonyl and so on

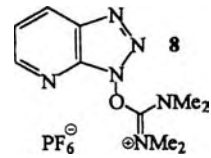
Amino Acid Derivatives

DKP	Diketopiperazine
NCA	*N*-Carboxyanhydride
PTH	Phenylthiohydantoin
UNCA	Urethane *N*-carboxyanhydride

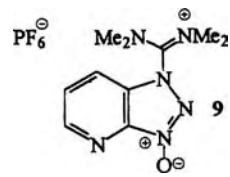
Reagents and Solvents

BOP	1-Benzotriazolyloxy-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 ***H***exafluorophosphate salt of the *O*-(7-***A***zabenzotriazol-lyl)-***T***etramethyl ***U***ronium 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	*N*,*N*′-Dimethylbarbituric acid
NMM	*N*-Methylmorpholine
PAM	Phenylacetamidomethyl resin
PEG	Polyethylene glycol
PtBOP	1-Benzotriazolyloxy-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
WSCI	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	Adrenocorticotropic 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

Amino acid and peptide bioconjugates

Nikolett Mihala and Ferenc Hudecz*
DOI: 10.1039/9781849734677-00001

Introduction

Bioconjugate research is a dynamic and trans-disciplinary field with fast development. During the last two decades the main focus of bioconjugate chemistry has been the chemical synthesis and functional characterization of two- or three, sometimes even multi-component systems in which the partner molecules are attached by covalent bond and preserve relevant functional properties (like biological activity or "reporter properties") after conjugation.

To achieve this, from the viewpoint of organic chemistry there are two main "restrictions": No.1. Only that part of the partner molecule could be chosen (or structurally modified/derivatized, if needed) for conjugation, which has no or negligible effect on the desired functional property of the product. No.2. There is a need to develop conjugation strategy (*e.g.* development a new site of reaction, inclusion a spacer entity), which will lead to a conjugate compound possessing the selected characteristics of both parent molecules (*e.g.* after coupling a fluorophore to a peptide hormone, the conjugate must be able to act as a hormone as well as a fluorophore).

Considering these "limitations" several new and exciting organic synthetic approaches and strategies, like "click chemistry",[1–3] bioorthogonal chemistry[4–7] and also novel analytical methodologies were developed.[8] The intellectual challenge and the practical importance of this research field could be demonstrated by the emerging appreciation of the journal of Bioconjugate Chemistry, established by the American Chemical Society in 1990 as documented by the IF > 5.0 value.[9]

Several comprehensive books and reviews published in the last years have summarized and discussed the state-of-the-art in the area of synthesis and application of bioconjugate techniques.[10–13] However, many aspects of the synthesis of peptide-bioconjugate that can not be encompassed by this chapter. These include, for instance the synthesis of glycopeptide/glycoprotein,[14,15] peptide-oligonucleotide bioconjugates[16] or the application of peptide bioconjugates in nanoscience.

We have made an attempt to provide an adequate coverage on the more widely used and available methods. The references and suggestions for further reading will hopefully provide good starting points for the inquisitive reader.

Current trends

Synthesis

The state-of-art in the repertoire of organic synthetic transformations enables the access of highly complex molecular structures. Obviously the chemical

Research Group of Peptide Chemistry, Hungarian Academy of Sciences, Eötvös L.University, P.O.Box 32, H-1518, Budapest 112t, Hungary. E-mail: fhudecz@elte.hu

toolbox for the conjugation of amino acids/peptides is much more restricted because should proceed in mild conditions. *i.e.* in aqueous solution and at near-ambient temperature. Moreover, such modifications must be highly chemoselective in the sense that only a single target moeity is to be modified in the presence of a myriad of other functionalities.

Bioconjugation methods of peptides rely heavily on chemoselective modification of functionalities of the side chains of amino acids. Lysine (tipically by acylation or alkylation) and cysteine (by alkylation, acylation and redox reactions) side chains are the most commonly functionalized amino acids. Alternatively, the carboxylic functionality of aspartate and glutamate residues can be activated by formation of active ester.

Recently, significant advancements have opened new avenues in bioconjugation chemistry. Several strategies have emerged that allow specific tailoring of oligo- and polypeptides through either endogenous residues or introduced functionality.

At the same time existing techniques have been expanded to enable the synthesis of a wider range of bioconjugates. Classical organic reactions have been explored in the context of bioconjugation as well. Furthermore a number of potentially useful bioconjugation reactions have been described, but have not yet been used in diverse conjugation applications.

Few methods – if any – are designed to cope with all reaction conditions, but all known factors and practical consideration should be taken into account when selecting a strategy.

1 Amide bond forming reactions

The peptide/amide bond formation is one of the most fundamental and widespread chemical tool in nature, underlying the properties of a huge array of organic biomolecules, synthetic polymers and materials, including peptides and proteins. The outstanding stability of amides is attractive especially in the synthesis of peptide-bioconjugates.

For the experimental (bio)chemist the number of methods available for the synthesis of amide linkage are nearly beyond counting. However this very high number of the chemical reactions emphasizes the need for improved strategies for synthesis of amide functionality. V. R. Pattabiraman and J. W. Bode has recently given a critical overview of examples of ground breaking, alternative amide bond forming strategies.[17]

1.1 Coupling reagents

Traditional approaches for formation of amide bond rely on coupling reagents. These agents convert the unreactive carboxylic function into an activated form for the reaction with a suitable amine to produce the desired amide. The most recently developed coupling reagents have been thoroughly reviewed and their potentials and limitations assessed by E. Valeur and M. Bradley[18] and by A. El-Faham and F. Albericio.[19]

1.2 Ligation methods

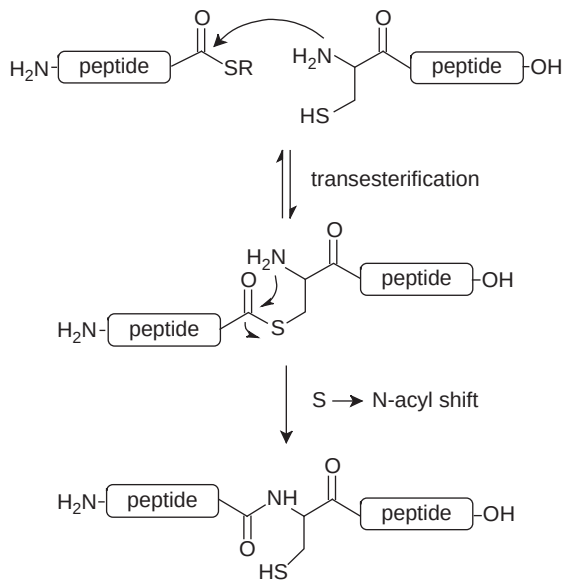
Chemoselective ligation methods have been lately summarized and thoroughly reviewed by T.K. Tiefenbrunn and P.E. Dawson[20] and also by C.P.R.Hackenberger and D. Schwarzer.[21]

1.2.1 Native chemical ligation. The native chemical ligation (NCL) method described by S.B. Kent and co-workers in 1994[22] exploits a chemoselective reaction between two unprotected fragments, a C-terminal thioester and an *N*-terminal cysteine, in aqueous solution at neutral pH thus forming a ''native'' amide bond at the ligation site (Scheme 1). First an equilibration between the nucleophilic thiol of the *N*-terminal cysteine and the electophilic thioester takes place then the newly generated thioester undergoes spontaneously an irreversible intramolecular *S*- to *N*- acyl transfer yielding a native bond between the two fragments. The requirement for cysteine at the ligation juncture is an intrinsic restriction of the NCL strategy.[23]

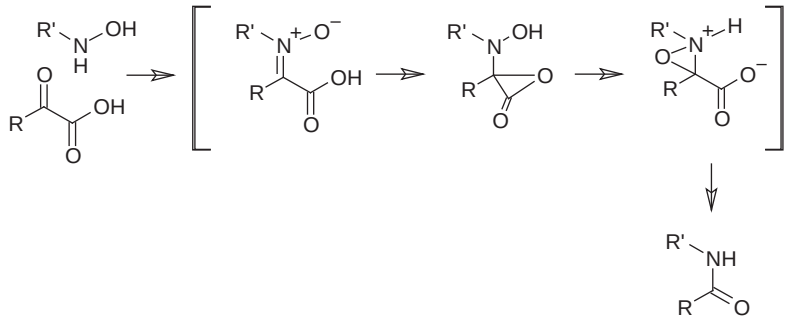
To overcome this limitation a number of different techniques (for reviews see[24–26]) has been developed. Although NCL and related technologies has became a powerful tool for peptide hence protein synthesis surprisingly, so far have only found limited application in the synthesis of peptide-bioconjugates.[27–31]

1.2.2 Chemoselective ligation. In 2006 J.W. Bode and co-workers reported on a mechanistically unique, highly chemoselective amide forming reaction.[32] The ligation of unprotected fragments is accomplished by an unusual decarboxylative condensation of α-ketoacids with *N*-alkylhydroxylamines as depicted in Scheme 2.[33] No reagents are required for the ligation and no by-products, other than carbon dioxide and water are produced and more importantly there is no need for a special amino acid at the ligation site. Lately, α-ketoacid–hydroxylamine amide ligation (KAHA) was applied for the synthesis of human glucagon-like peptide, GLP-1 (7–36).[34] The 30-amino-acid residue peptide was prepared in good yields and purity.

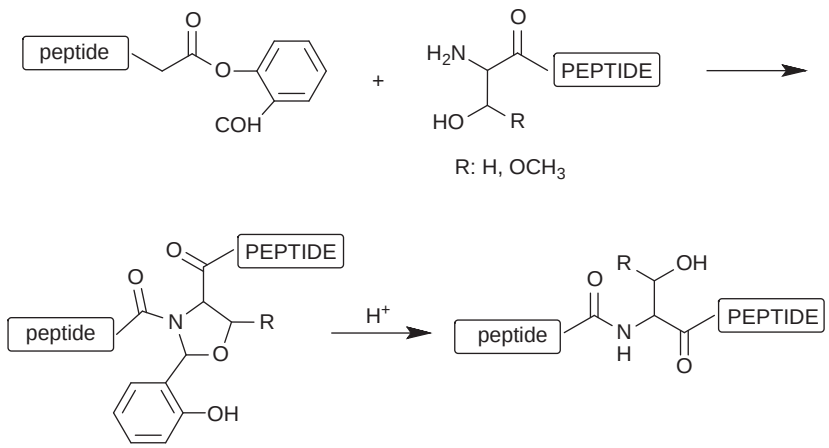
Recently X. Li *et al.* has published their preliminary investigations on a new chemoselective ligation method.[35] The reaction which involves a



Scheme 1 The principle of native chemical ligation.



Scheme 2 The α-ketoacid–hydroxylamine amide ligation.



Scheme 3 The salicylaldehyde ester-induced chemoselective ligation.

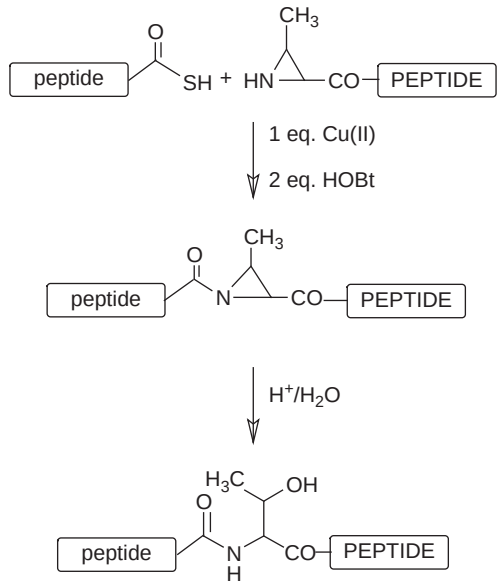
peptide having an O-salicylaldehyde ester at the *C*-terminus and Ser/Thr peptide possessing a protecting group at the *C*-terminus proceeds rapidly in Py/AcOH (1:1). The forming N,O-benzylidene acetal intermediate in a separate step is treated with a mixture of TFA/H_2O/i-Pr_3SiH to give a native peptide bond (Xx- Ser/Thr) at the ligation site (Scheme 3).

Another study on a novel chemical ligation resulting in Xxx-Thr bond was reported by F. B. Dyer and co-workers.[36] The reaction between an unprotected peptide thioacid and an aziridine-2-carbonyl containing peptide proceeds in the presence of stoichiometric amount of Cu(II) ion (Scheme 4). To suppress epimerization 1-hydroxybenzotriazole (HOBt) was added to the mixture of DMF-buffer. The forming aziridine - without isolation - was converted using water as nucleophile into the ligation product containing a Thr residue at the ligation site. The amino, carboxyl and aliphatic/aromatic hydroxyl groups were found compatible with this ligation method.

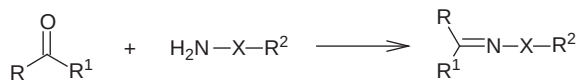
2 Reactions yielding a non-amide bond

2.1 Imine ligation

The condensation reaction of a nitrogen base with aldehydes or ketones is accomplished in aqueous solutions at neutral pH forming a carbon-nitrogen



Scheme 4 The aziridine-mediated peptide ligation at Xaa-Thr sites.



Scheme 5 The hydrazone/oxime ligation.

bond is a highly chemoselective reaction and as such has found wide application in the conjugation of biomolecules even though it shows sluggish reaction kinetics.

When the nitrogen base is a hydrazine the forming products are hydrazones (C=N–N)[37] while oximes[38,39] (C=N–O) are generated when the nitrogen base is an alkoxyamine (Scheme 5). Both hydrazones and oximes are significantly more stable than those of simple imines (C=N), the products of condensation reaction of amines with aldehydes or ketones. Furthermore hydrazone can be reduced with sodium borohydride to produce the more stable hydrazide.

Recently, anilines were suggested to be effective catalysts of these reactions.[40,41] This could greatly broadened their utility by allowing the use of a significantly lower concentration of reagents and reaction conditions at a pH value closer to neutral.

2.1.1 Hydrazone ligation. Hydrazone ligation chemistry has became central to the synthesis of a wide range of bioinspired compounds such as hydrazide reactive peptide tags for protein labelling.[42] Further utility of the aniline catalyzed hydrazone ligation has been shown for attaching protein capture reagents, aldehyde-modified antibodies to model silicon dioxide surface demonstrating that the method can be exploited for biosensor

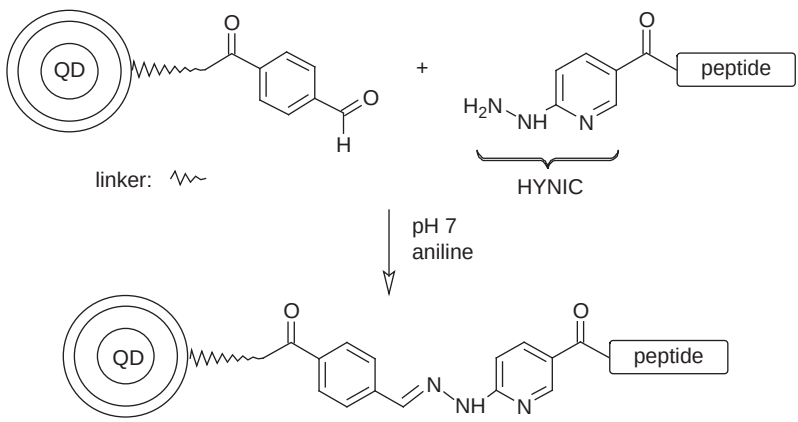
applications in the biomedical field.[43] Hydrazone ligation resulted an efficient and versatile method also for the covalent modification of quantum dots (QD).

J. B. Blanco-Canosa and co-workers reported on a straightforward aniline catalyzed arylhydrazone ligation strategy with chemoselective covalent modification of CdSe ZnS QDs.[44] The reaction between the benzaldehyde functionalized QDs and 2-hydrazinonicotinoyl (HYNIC) protease substrate oligopeptide with a fluorescent label was found compatible with neutral aqueous buffer. The reaction proceeded quantitatively at micromolar concentrations without the use of an excess of the peptide target. Average peptide/QD ratios from 2:1 to 11:1 (mol/mol) were achieved with excellent control over the desired valency.

G. Iyer and co-workers developed a very similar method for the synthesis of modified QDs. It is based on bis-aryl hydrazone bond formation mediated by aromatic aldehyde and hydrazinonicotinate acetone hydrazone (HYNIC) activated peptide coated quantum dots. The versatility of the approach was demonstrated by controlled preparation of antibody–QD bioconjugates (Scheme 6).[45]

In an elegant study, F.M. Brunel and co-workers described a sequential hydrazone ligation strategy to assemble multifunctional viral nanoparticles.[46] The Cowpea mosaic virus was functionalized by benzaldehyde moiety and than conjugated to a tumour targeting peptide bearing arylhydrazone group. The kinetics of covalent modification can be monitored spectroscopically at $\lambda \sim 350$ nm due to the strong absorbance of the forming bisarylhydrazone product. In a second step, the conjugate was incubated with an arylhydrazido group modified and fluorescent PEGylated peptide. This sequential arylhydrazone ligation strategy enables the introduction of different labels in a ratio that is controlled by the reaction conditions. It is important to note that the modular nature of the method can be tailored for specific applications.

A two-step modular linkage strategy was described by D.E. Prasuhn and co-workers[47] for the multivalent display of oligopeptides and DNA on



Scheme 6

CdSe/ZnS core/shell QDs. The approach exploits the chemoselective, aniline-catalyzed hydrazone coupling chemistry to append hexahistidine sequences onto oligopeptides and DNA for subsequent self-assembly to QDs. This specifically facilitates the ratiometrical self-assembly to hydrophilic QDs. The versatility of these QD-biomolecular conjugates were demonstrated in a variety of targeted biosensing, hybridization, and cellular uptake assays.

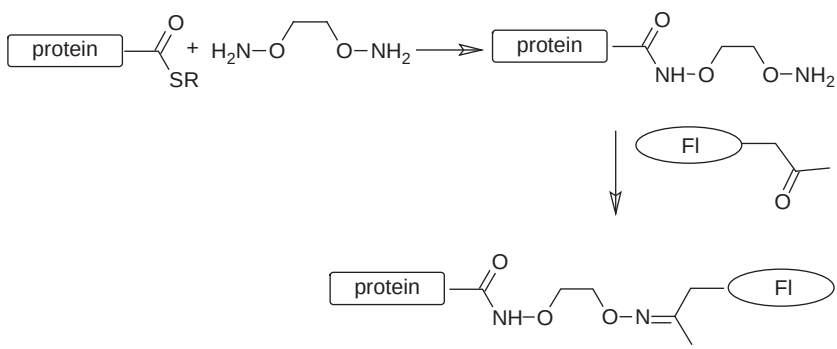
A.R. Blanden *et al*. studied the reaction between 3-formyltyrosine and hydrazine-containing fluorophores using 4-aminophenylalanine as catalyst. The authors found that 4-aminophenylalanine preserved $\sim$70% of the catalytic efficacy of aniline with the advantage of being much more compatible with biomolecules.[48]

2.1.2 Oxime ligaton. Oxime ligation is a simple, yet powerful tool suited for application in many areas of bioconjugate chemistry due its high chemoselectivity and the stability of forming oxime bond under physiological conditions. Of late, a number of synthetic strategies for the introduction of functionalities allowing oxime ligation was described.

M. R. Carrasco *et al*. showed a highly efficient synthetic strategy for Boc-protected aminooxy and *N*-alkylaminooxy amines starting from two- and three-carbon Cbz-protected amino alcohols.[49] The amino functionality enables the incorporation of the aminooxy moiety by standard amide formation protocol into biomolecules and after removal of the protecting group the forming aminooxy and *N*-alkylaminooxy groups enables conjugation with desired target molecules establishing various distances between the reaction partners.

R. S. Goody and co-workers established a simple and general method to incorporate oxyamino functionality into proteins. The homofunctional 1,2-*bis*(oxyamino)ethane linker could be introduced into protein thioester. The oxyamino - modified proteins react rapidly with keto fluorescent dyes under mild condition (Scheme 7).[50]

M. B. Francis reported[51,52] screening studies that have identified the most reactive *N*-terminal residues in the previously described pyridoxal 5′-phosphate (PLP)-mediated reaction. This site-specifically oxidizes the *N*-terminal amine of a proteins to afford a ketone functionality. A generalizable



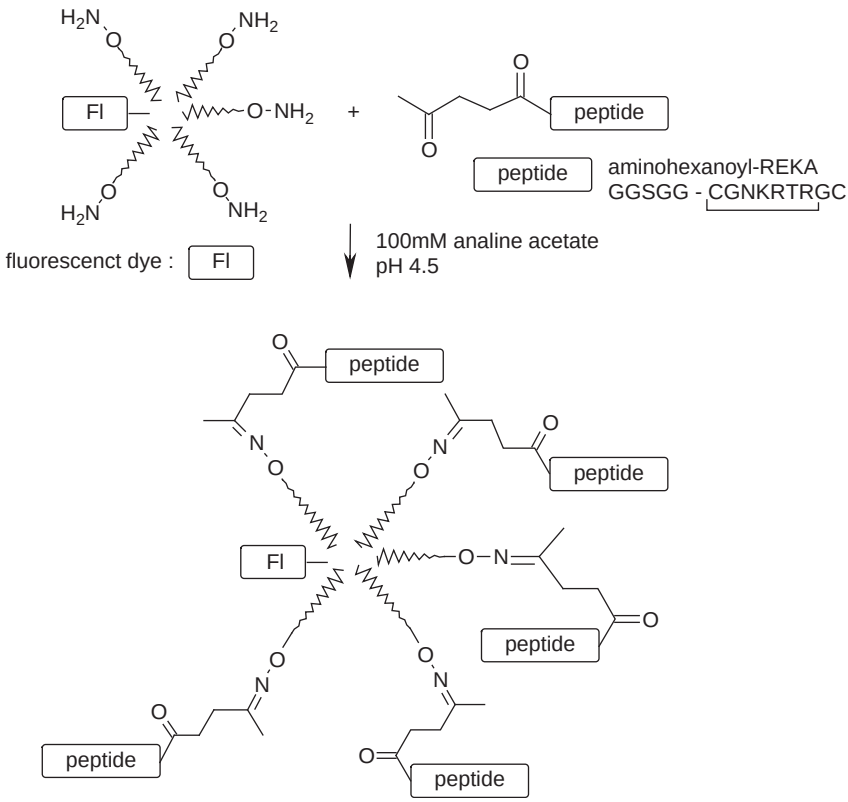
Scheme 7

combinatorial peptide library screening platform was developed, which enables the identification of highly reactive motifs toward a desired bioconjugation reaction.

E. H. M. Lempens *et al.* developed a surface immobilization technique based on aniline-catalyzed oxime ligation chemistry.[53] A set of aldehyde- or glyoxyl-functionalized peptide and protein was conjugated to biosensor chip modified with alkoxyamines.

In another study of the same research group, the controlled conjugation of tumour homing oligopeptides to an AB_5-type dendritic scaffold via aniline-catalyzed oxime chemistry was reported.[54] A short linear and a cyclic tumour homing peptide modified on solid phase by levulinic acid were incubated using 6:1 peptide/dendron ratio (mol/mol) to the aminooxy-functionalized scaffold obtaining complete substitution. The resulting pentavalent compounds were investigated for tumour homing and found that the peptide properties and overall size greatly influenced their biological behaviour (Scheme 8).

O. Renaudet *et al.* presented an iterative oxime ligation strategy for the assembly of multivalent bioconjugates. The method was illustrated by the assembly of structurally diverse tetravalent and hexadecavalent glycoclusters on a cyclodecapeptide template.[55]



Scheme 8

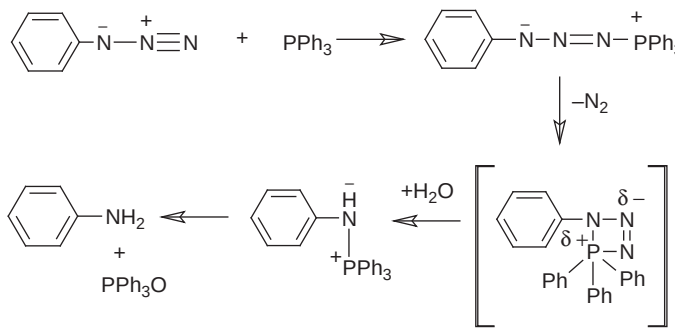
2.2 Staudinger ligation

The roots of this ligation method go back to the investigations of Hermann Staudinger, who in 1919 described the reaction of azides with triarylphosphines through an aza-ylide intermediate, which in aqueous medium rapidly hydrolyzes to the corresponding amine (Scheme 9).[56]

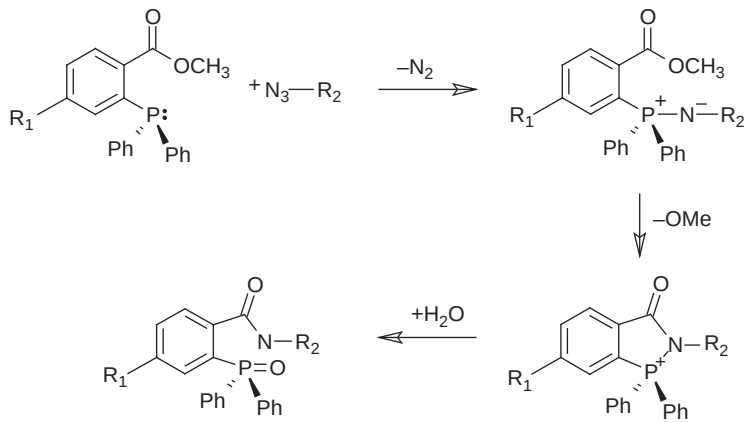
The Staudinger ligation developed by C. Bertozzi and co-workers[57] is a variant of this classical Staudinger reduction in which the triaryl phosphine substituted with electophilic group (*i.e.* methylester) captures the nucleophilic aza-ylide intermediate by an intermolecular cyclization and creates a stable amide (Scheme 10).

Of particular note is the high compatibility of the Staudinger ligation with biological systems thus this method can offer a very attractive alternative or complement to the more traditional ligation techniques. However, it can not be applied to visualize low-abundance entities or to follow rapid biological processes due to its slow kinetics.[58]

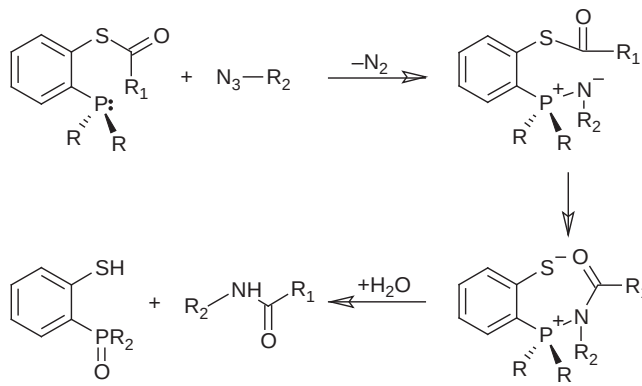
It is not surprising that - even considering its serious drawbacks, like the aforementioned sluggish kinetics or the sensitivity of phosphines to air oxidation - the Staudinger ligation, as a conjugation strategy, has been widely



Scheme 9 The Staudinger reduction.



Scheme 10 The Staudinger ligation.



Scheme 11 The traceless Staudinger ligation.

applied for the synthesis of oligopeptides and proteins through coupling chemical probes to biomolecules as well as to cell surface engineering *in vivo*.

It is worth to mention that the very first bioorthogonal reaction ever performed not in cultured cells, but in living animals was a Staudinger ligation.[59] Azido-tagged carbohydrates were detected by a phospine probe in laboratory mice.

Two methods have been developed simultaneously to modify the Staudinger ligation into a ''traceless'' ligation approach. Both of them are based on trapping the nucleophilic nitrogen of the aza-ylide through intramolecular reaction an electrophilic moiety.

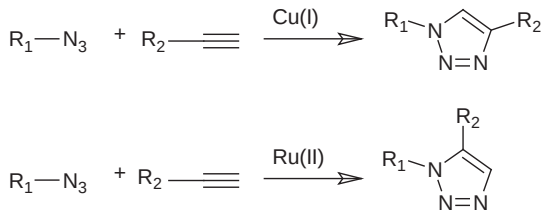
In the traceless ligation described by Raines and co-workers[60] for the ligation of two peptide fragments a phosphinothioester traps the aza-ylide. The forming intermediate after rearrangement and hydrolysis establish the amide bond. The mechanistically very similar method investigated by Saxon *et al*. exploits acylated phosphine derivatives (Scheme 11).[61]

The Staudinger ligation as chemoselective and bioorthogonal conjugation method was recently reviewed by E. M. Sletten and C. R. Bertozzi[58] C. I. Schilling *et al*.[62] and by S. S. van Berkel.[63]

2.3 Cycloadditions

Since the introduction of cycloadditions for performing bioconjugation there has been a steady growing steam of publications describing the application of these methods. In particular, azide–alkyne cycloadditions proved to be highly relevant for biological applications.[1] The breadth and depth of this research area clearly demonstrates the increasing need of the further development of novel bioconjugation methods.

2.3.1 Cu(I)-catalyzed azide-alkyne cycloaddition. In 2002 B. Sharpless[64] and M. Meldal[65] concurrently and independently discovered that metal catalysis dramatically accelerate the classical [3 + 2] Huisgen cycloaddition.[66] This variant of the Huisgen 1,3-dipolar azide–alkyne cycloaddition was termed CuAAC for Cu-catalyzed Azide-Alkyne Cycloaddition and often referred to as click reaction. CuAAC is a prototype of ''click chemistry'' reaction a term coined in 2001 by B. Sharpless and co-workers in their



Scheme 12 The metal catalyzed 1,3-dipolar azide-alkyne cycloaddition.

conceptual and futuristic paper to distinguish a set of chemical reactions that efficiently link two components.[67] A click chemistry reaction by definition should be modular, stereospecific and give very high yields. Moreover these reactions are based on readily available starting materials, simple reaction conditions (ideally, the process should be insensitive to oxygen and water) and generate inoffensive by-products.

Copper(I) and ruthenium(I) are the most commonly used catalysts in the reaction. The use of Cu(I) results in exclusively the formation of 1,4-disubstituted 1,2,3-triazole whereas Ru(I) results in the formation of the 1,5-regioisomer as it is depicted in Scheme 12.

The copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition (CuAAC) has recently enjoyed tremendous interest as a synthetic route for the preparation of complex materials due to its high conversion and mild experimental conditions. The application of Cu-catalyzed azide-alkyne cycloaddition has increased exponentially in the last years not only in biochemistry, but also in organic synthesis, inorganic chemistry and polymer chemistry. The forming triazole ring is similar to the amide bond, but not susceptible to neither enzymatic degradation nor oxidation/ reduction. Recently, triazol was reported as disulphide bond mimetic as well.[68]

Hallmark feature of the Cu-catalyzed azide-alkyne cycloaddition is its high chemoselectivity between the two partners. The reaction proceeds rapidly at ambient temperature in aqueous solvents and requires no protecting groups.[3,69] Moreover, the small and stable azide functional group can be readily incorporated into diverse set of biomolecules either chemically, or by bioengineering/metabolic methods without dramatically changing their biological functions. A straightforward strategy on this was reported by T. Fricke and co-workers for the introduction of alkyne functionality into expressed peptides.[70]

The Cu-catalyzed azide-alkyne click reaction was also exploited in the synthesis of sophisticated bioconjugates encompassing a variety of biomolecules, such as oligopeptides,[71] proteins,[72] carbohydrates, lipids, nucleic acids and macromolecular assemblies like viral[73,74] and other nanoparticles.[75] This chemistry was widely utilized in chemical proteomics,[76] activity-based protein profiling[77] and target guided synthesis of enzyme inhibitors.[78] The CuAAC allowed *in vivo* and *in vitro* protein modification, nucleic acid immobilization and tracking, *in vivo* lipid labelling and virus surface remodelling.[79,80]

It should be noted, however, while this strategy offers exquisite specificity, it does have some limitations. One major shortcoming of the Cu-promoted

click reaction is that it works exclusively on terminal alkynes. In addition, copper catalysts might be incompatible with living systems due to its toxicity.[81]

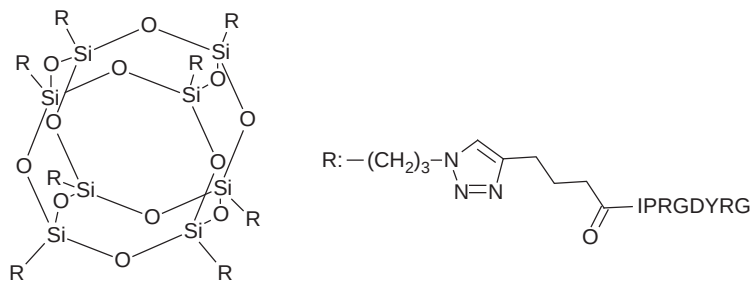
Even in more complex and challenging environment, *in vivo* cell culture, click chemistry has found its successful application. The versatility of the CuAAC in living cells was demonstrated by V. Hong *et al*. in experiments on mammalian cells (CHO, Jurkat or HeLa) labelled efficiently in 5 min. with no loss in cell viability.[82]

The ultimate test of the concept was the non-invasive imaging of glycans in transparent live zebrafish embryos reported by D. Soriano del Amo *et al*.[83]

Of late, J. R. Heath and co-authors described the use of an iterative *in situ* click chemistry approach to design an Akt-specific branched peptide triligand composed of 5, 6 and 7 amino acid residue modules and conjugated via their additional *N*-terminal unnatural amino acid, bearing azido and/or alkyne group. The conjugate was employed as a drop-in replacement for monoclonal antibodies in multiple biochemical assays. The results pointed to the potential for iterative *in situ* click chemistry for the preparation of effective antibody replacements with novel inhibitory properties.[84]

S. Fabritz *et al*. described and ''exotic'' cube-octameric silsesquioxane (POSS) based conjugation scaffolds for copper catalyzed azide-alkyne cycloaddition. The octaazido/octaalkyno functionalised templates were conjugated with an unprotected peptide forming a POSS assembled octamer (Scheme 13).[85]

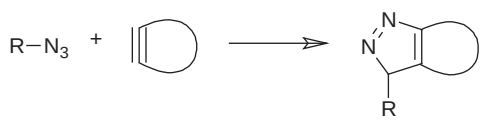
Lately, a number of reviews discussed the impact of Cu(I)-catalyzed azide-alkyne cycloaddition on various fields of chemical biology. M. Glaser and E. G. Robins summarized the recent applications of CuAAC in radiosynthesis of tracers for positron emission tomography.[86] C. Le Droumaguet *et al*. outlined the design and recent developments of fluorogenic Cu(I)-catalyzed alkyne–azide cycloaddition reactions as a powerful tool for bioconjugation, materials science, organic synthesis and drug discovery.[87] E. Lallana and co-workers discussed the impact of Cu(I)-catalyzed alkyne–azide cycloaddition in the synthesis of various nanosized drug delivery systems.[88] E. Fernandez-Megia *et al*. summarized the key aspects in the development of the efficient and bio-benign functionalization of biomacromolecules through CuAAC.[69]



Scheme 13

2.3.2 Strain-promoted azide/alkyne cycloaddition. The potential toxicity of copper has been addressed by the development of an alternative, copper-free bioconjugation method, which enables the *in vivo* application. To lower the activation barrier without a catalyst the [3 + 2] strain-promoted azide/alkyne cycloaddition (SPAAC)[89] employs highly strained cyclic alkynes (Scheme 14). The [3 + 2] strain-promoted azide/alkyne cycloaddition (SPAAC)[89] or the copper free click chemistry proceeds well at ambient temperature and pressure and exhibit an exceptional tolerance towards a wide range of functional groups and reaction conditions**.** Moreover, cycloalkynes are abionic, biologically stable and inert thus allowing biological imaging in living cells or even in animals in reaction with an azide-tagged biomolecule.

The first cyclic alkyne experimented in strain-promoted azide/alkyne cycloaddition was the smallest isolable cycloalkyne, the cyclooctyne. The first generation of cyclooctynes suffered from relatively low reactivity in reaction with azides. Since then, different classes of analogues (Fig. 1**)** have been designed achieving more than two orders of magnitude faster reaction rates. However, these cycloalkynes frequently require tedious multistep synthesis and the increased reactivity often goes hand in hand with increased sensitivity to moisture or air. Latest examples can be reviewed in relevant references.[90–94] The relatively limited commercial availability of cycloalkyne derivatives suitable for bioconjugation is an additional factor that may hamper the widespread application of the copper free click reactions.



Scheme 14 The [3 + 2] strain-promoted azide-alkyne cycloaddition.

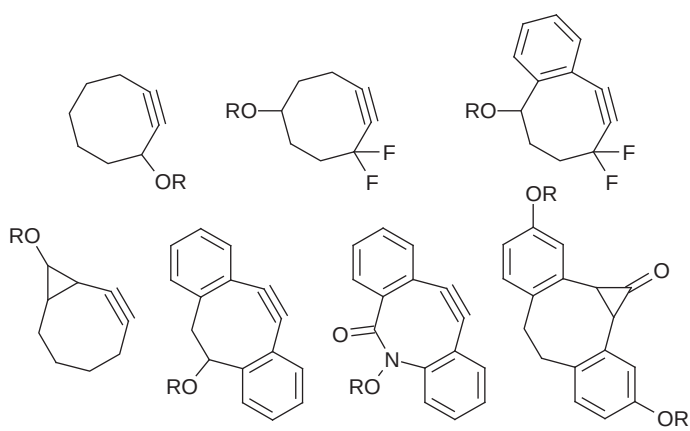


Fig. 1

The main drawbacks of this reaction is the production of a mixture of regioisomeric products, unlike the Cu(I)-catalyzed cycloaddition, which produces only one regioisomer with high yield. In addition, the strain-promoted azide/alkyne cycloaddition possess relatively sluggish reaction kinetics. Another shortcoming is the hydrophobicity of the alkyne compounds that may result in limited solubility in aqueous solutions and more importantly, in non specific bonding to biomolecules (although covalent attachment to cysteine residues is not completely excluded either[7,95]).

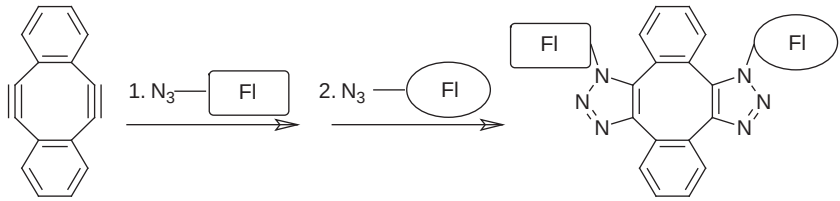
The first application of Cu-free click chemistry in animal model involved the reaction of a fluorescent cycloalkyne with a carbohydrate-azide and allowed the visualization of spatiotemporal changes in cell-surface glycosylation in *Caenorhabditis elegans*[96] and later in developing zebrafish.[97] Recently, a set of cyclooctynes bearing FLAG epitope tag was evaluated in glycoconjugate labelling - in comparison to the Staudinger ligation – both *in vitro* and in mice.[7]

In a comparative study by P. Wu and co-workers, the efficacy of Cu(I)-catalyzed and strain-promoted azide-alkyne biarylazacyclooctynone cycloadditions were thoroughly evaluated in different biological applications: detection of purified recombinant glycoproteins in crude cell lysates, labelling of glycans on the surface of live cells and in zebrafish embryos. The optimized Cu(I) catalyzed cycloaddition was found superior in all studied biological settings.[98]

Besides glycans, proteins were visualized by strain promoted azide/alkyne cycloaddition in live cells. D. A. Tirrell *et al*. described the use of SPAAC to label newly synthesized proteins in living cells with three different cell-permeable coumarin–cyclooctyne conjugates,[95] After 10 minute exposure of cyclooctyne derivative the cells previously pulse-labeled with azidohomoalanine were characterized by substantial, eight or higher fold enhancements in fluorescence relative to cells treated with Met.

The viability of copper free chemistry for the immobilization of a variety of molecules onto a solid surface and microbeads was demonstrated by A. Kuzmin and co-workers.[93] 18F labelled cyclooctynes was utilized in the synthesis of radiotracers for PET imaging.[99–101] Copper-free click chemistry was employed in proteomics for capturing azido-peptides by a resin-immobilized cyclooctyne.[102] M. D.Burkart *et al*. studied protein–protein interactions in non-ribosomal peptide biosynthesis via selective cross-linking.[103] Azido-lipids presented on liposomes were also conjugated to cyclooctyne-tagged reagents derivatizing membrane bilayers.[104]

L. A. Canalle *et al*. explored click chemistry as a potential cost-effective and selective immobilization method for the production of an enzyme-linked immunosorbent assay (ELISA).[105] Coatings were formulated containing either a terminal alkyne or cyclooctyne handle, and a diagnostic oligopeptide was subsequently immobilized onto these coatings by CuAAC or copper-free strain-promoted azide-alkyne cycloaddition, respectively. While the coating, synthesized via CuAAc, showed high background in the ELISA experiments due to the copper catalyst used in the immobilization step, the coating prepared via copper free reaction was successfully employed in ELISA for monitoring rheumatoid arthritis. This strategy



Scheme 15 The strain-promoted "double-click" reaction.

could be a cost-effective alternative to existing (strept)avidin-biotin immobilization methods.

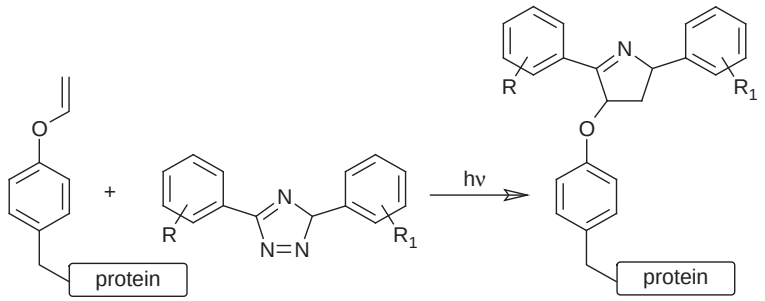
The strain-promoted "double-click" (SPDC) reaction was developed by I. Kii[106] and co-workers. The method takes advantage of the Sonderheim diyne as cycloalkyne as outlined (Scheme 15) allowing the consecutive conjugation of two molecules possessing azido functionality. The viability of this concept in bioorthogonal chemistry was demonstrated in labelling of cells pretreated with azido- sugar by a fluorescent azido-dye.

The strain promoted cycloaddition has been proved orthogonal with the Cu-catalyzed azido-alkyne cycloaddition. P. Kele and co-workers reported[107] the dual labelling of peptides, proteins (*e.g.* bovine serum albumin) and silica nanoparticles using consecutive strain promoted and Cu-catalyzed azido-alkyne reaction. For example an *N*-terminal cycooctyne derivative and a propargylglycine were incorporated into a matrix metalloproteinase substrate peptide than subsequent one-pot labelling with different fluorophores forming a fluorescence resonance energy transfer (FRET) system afforded a sensitive probe of enzyme activity.[108]

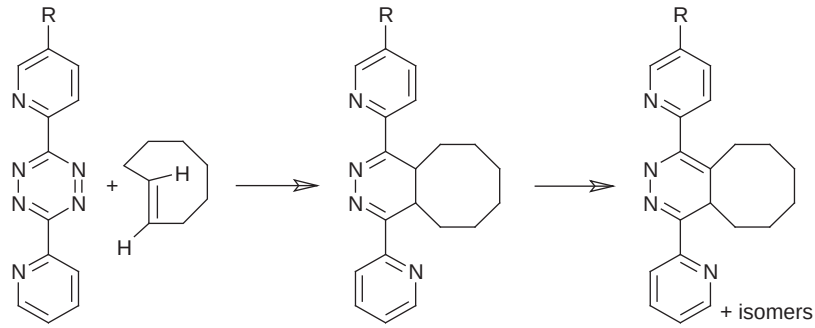
Copper free click reactions have been reviewed by various authors recently: J. M. Baskin and C. R. Bertozzi outlined the development of cycloalkyne reagents and provided a brief outlook of their applications as well.[109] The relevance of SPAAC in chemical biology in more details was summarized in different papers by C. R. Bertozzi and co-authors.[58,110] M. D. Best,[2] E. Sletten *et al.*,[5] M. Boyce *et al.*[6] discussed the strain-promoted cycloaddition in comparison of other bioorthogonal reactions.

2.3.3 Other cycloaddition reactions. Besides the well-established copper catalysed and strain-promoted cycloadditions novel methods have been explored for bioconjugation. M. F. Debets and co-authors summarized bioconjugation strategies centered around strain-promoted cycloadditions[111] and around the azides[112] (in particular, their potential to undergo metal-free reactions with strained unsaturated systems) with special emphasis on reaction kinetics.

Photoinducible 1,3-dipolar cycloaddition. The photoinducible 1,3-dipolar cycloaddition reported by Q. Lin *et al.*[113] is another ligation technique that takes advantage of the investigations pioneered by Huisgen. The versatility of this transformation involving small-ring heterocycles and simple alkenes (Scheme 16) has been demonstrated for both *in vitro* and *in vivo* modification of proteins. The photoclick reaction proceeds under mild



Scheme 16



Scheme 17 The tetrazine ligation.

conditions in high yield with excellent selectivity. The kinetics of this cyclization is comparable to that of the copper-catalyzed azide-alkyne cycloaddition. The photoclick chemistry has been used to label proteins rapidly (within $\sim$1 min) both *in vitro* and in *E. coli*.[114] The main advantage of using an alkene reporter along with photoclick chemistry for protein imaging is the potential of spatiotemporal control.[115] Of late, R. K. V. Lim and Q. Ling[116] reviewed the photoinducible, bioorthogonal tetrazole–alkene cycloaddition reaction and its application for probing protein dynamics and function in living cells. Additionally, the synthesis, structure-reactivity studies of tetrazoles, including their optimization for applications in biology are discussed.

Tetrazine ligation. The tetrazine ligation means the reaction of a *trans*-cyclooctene and an *s*-tetrazine in an inverse-demand Diels-Alder reaction followed by a retro-Diels-Alder reaction.[117] The reaction (Scheme 17) is extremely rapid thus allowing modifications of biomolecules at extremely low concentrations. Selectivity, fast reaction rate and aqueous compatibility make the reaction suitable for various *in vivo* and *in vitro* applications. However, tetrazines has some background activity towards amines, thiols, and water. To date, this inverse-demand Diels-Alder reaction between s-tetrazines and *trans*-cyclooctenes is the fastest known bioorthogonal reaction. This exceptionally fast kinetics allows real-time imaging of covalent protein modification.

For example near-infrared-emitting fluorophore labelled tetrazine were conjugated to dienophile-modified extracellular proteins in living cancer cells.[118]

The technique has been further exploited for the design of fluorogenic probes that increase in fluorescence intensity upon ligation and for labelling biomarkers on cells with magneto-fluorescent nanoparticles.[119,120]

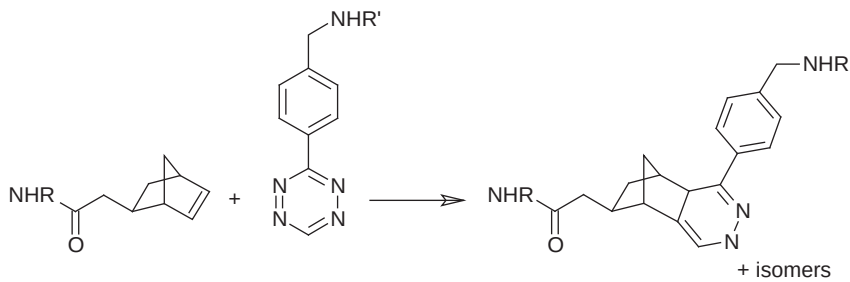
This ligation found several application in the synthesis of radiolabelled probes for PET imaging.[121–123] The application of tetraazine ligation in biomedicinal chemistry has been recently reviewed by N. K. Devaraj and R. Weissleder.[124]

The tetraazide ligation was found mutually orthogonal with strain-promoted azide–cyclooctyne cycloaddition reactions. The orthogonality was demonstrated with simultaneous labelling of two different live cell populations in the same culture.[125]

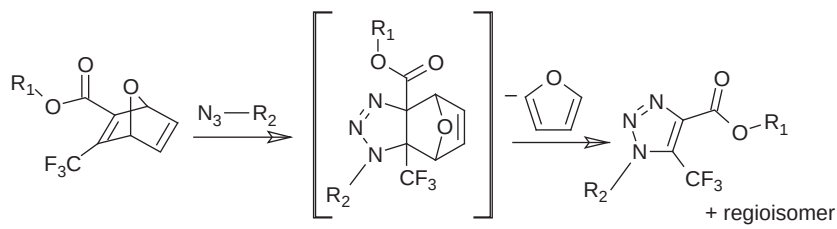
Norbornene cycloaddition. A very similar concept was introduced by S.A. Hilderbrandt and co-workers. In this bioorthogonal reaction the mono-aryl tetrazine reacts with norbornene as depicted in Scheme 18.[126] This reaction is slower than the above mentioned one, but still faster than strain-promoted azide-alkyne cycloaddition.

The bioorthogonality of this reaction was tested by the conjugation of norbornene and rhodamine to a monoclonal antibody. The method was exploited in targeting of norbornene-coated quantum dots to living cancer cells labelled with tetrazine-modified proteins[127] as well as in the synthesis of radiometallated antibodies.[128]

Cyclobutenenorbornene ligation. Another tetrazine-base ligation (Scheme 19) was developed by Pipkorn *et al.* as a method to ligate a transporter



Scheme 18 The norbornene cycloaddition.



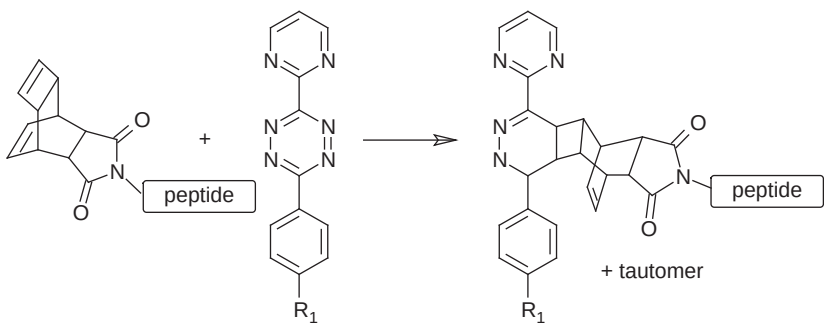
Scheme 19 The cyclobutenenorbornene ligation.

oligopeptide to the anticancer drug temolozomide (TMZ). The ligation involves inverse electron demand Diels-Alder reaction of cyclobutene and biaryltetrazine. The conjugation dramatically increased the antitumour activity of the drug in prostate cancer cells *in vitro*.[129,130] Similar TMZ containing BioShuttle constructs labelled with a fluorescent dye were evaluated as teranostics in human breast cancer cells.[131]

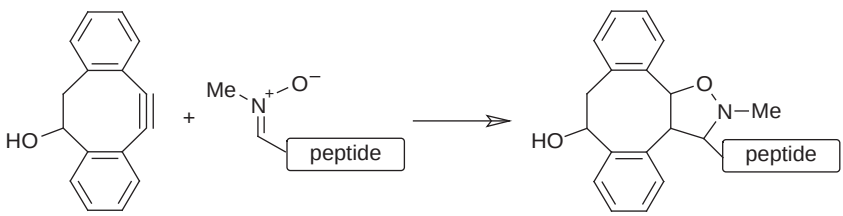
Oxanonbornadiene cycloaddition. The oxanorbornadiene cycloaddition is a 1,3-dipolar cycloaddition followed by a retro-Diels Alder reaction to generate a triazole-linked conjugate with the elimination of a furan molecule [132] (Scheme 20). The versatility of this ligation for protein modification was demonstrated by labelling hen egg white lysozyme with an azido-coumarine dye.

Using this metal-free triazole formation, F. Rutjes *et al*. were able to conjugate a chelated ^{111}In-radioisotope to an azide-functionalized cyclic peptide for SPEC imaging[133] and to attach tat peptide to PS-PEG for the preparation of polymersome nanoreactors.[134] The oxanorbornadiene reagents are relatively easy to prepare, but some undesired side-reactions was observed in the presence of electron-rich double bond, particularly with unsubstituted oxanorbornadienes.

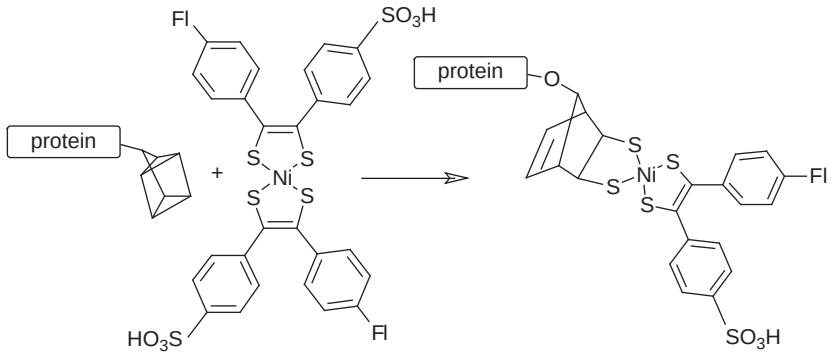
Strain-promoted alkyne–nitrone cycloaddition. X. Ning and co-workers envisioned a novel bioorthogonal reaction pair, namely a strain-promoted alkynes and a nitrone, which gives *N*-alkylated isoxazolines in high yield[135] (Scheme 21). The Strain-Promoted Alkyne–Nitrone cycloaddition (SPANC) was successfully employed for site-specific modification of a pentapeptide and IL8 in one-pot synthesis. The potential advantage of this



Scheme 20 The oxanonbornadiene cycloaddition.



Scheme 21 The strain-promoted alkyne–nitrone cycloaddition.



Scheme 22 The quadricyclane ligation.

method is the fast reaction kinetics and the relatively straightforward synthesis of nitrones.

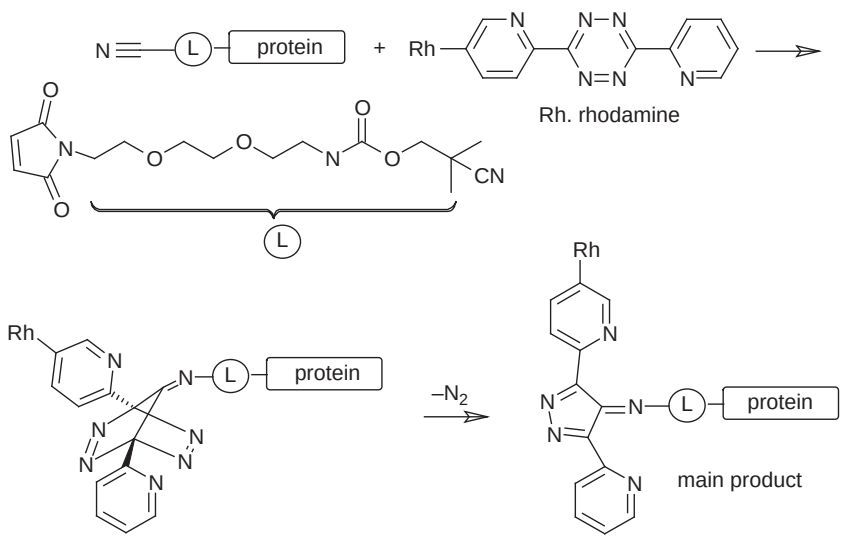
A very similar strategy – the chemoselective and bioorthogonal reaction between dibenzocyclooctyne and cyclic nitrone - was applied by C. S. McKay and co-workers for labelling human tumour cells.[136]

Quadricyclane ligation. Recently E. M. Sletten and C. R. Bertozzi reported the [2 + 2 + 2]-cycloaddition reaction of quadricyclane (a highly strained hydrocarbon) and a nickel *bis*(dithiolene) complex as bioorthogonal ligation reaction for chemoselective protein labelling (Scheme 22). Quadricyclane is relatively small, abiotic, unreactive with biomolecules, but highly strained. It reacts readily with electron-poor π systems but not with alkenes, alkynes, or cyclooctynes. The second order rate constant of quadricyclane ligation was found comparable to the rates reported for bioorthogonal strain-promoted azide/alkyne cycloaddition reactions.

The ligation proceeds readily in aqueous solutions at physiological pH, is highly selective and can be applied alongside other bioorthogonal reactions like the 1,3-dipolar cycloadditions or Diels-Alder reactions ([4 + 2]-cycloaddition) or oxime formation. A shortcoming of this method is the red-ox sensitivity of the Ni complex, which might preclude some applications of the quadricyclane ligation in living systems.

[4 + 1] cycloaddition. H. Stöckmann and co-workers reported the first bioorthogonal [4 + 1] cycloaddition.[137] This variant of click chemistry involves an isonitrile-tetrazine pair. In their study the selective labelling of a tertiary isonitrile-tagged protein by a fluorophore was investigated (Scheme 23).

2.3.4 Sequential bioconjugation strategies involving cycloaddition. M. Galibert and co-workers described (also as an extension of their former technique[138]) a one-pot triple orthogonal chemoselective ligation through the combination of the CuAAC reaction, oxime ligation, and thioether bond formation enabling the preparation of sophisticated biomolecular structures without the requirement of special protection schemes.[139] To illustrate this strategy a multifunctional, four c(RGDfK)-containing cyclodecapeptide scaffold was designed encompassing a carbohydrate residue along with a fluorescent probe or a nucleic acid fragment.



Scheme 23 The [4 + 1] cycloaddition.

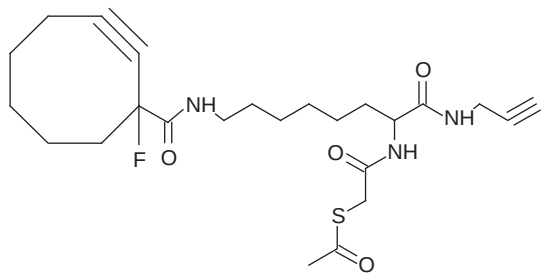


Fig. 2

In the same vein D. M. Beal *et al*. envisioned a heterotrifunctional template (Fig. 2) that utilizes thiolate-maleimide and click chemistries (both copper-free and copper-mediated) to effect sequential biomolecule conjugations in a one-pot process.[140] This template was proposed for the design of chemically-enhanced/enabled biotherapeutics, especially through the expression of discontinuous (and heterogeneous) epitopes.

P. A. Ledin *et al*. reported a strategy for efficient dendrimer assembly and heterobifunctionalization employed three sequential click reactions (SPAAC + CuAAC).[141]

R. M. J. Liskamp and co-workers applied CuAAC and thio acid/sulphonyl azide ligation in the preparation of chelator conjugated mono-, di- and tetrameric somatostatin analogues.[142]

2.4 Chemoselective reactions

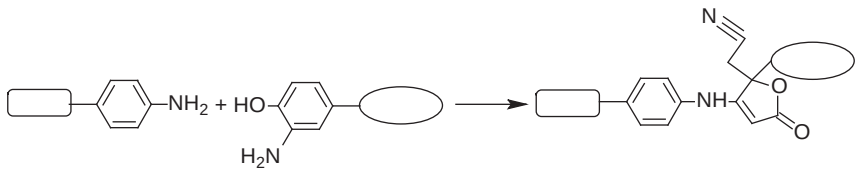
The studies outlined below represent more proof-of-concept than widespread applicability.

M. B. Francis and co-workes reported on a new protein bioconjugation method which exploits the addition of anilines to *o*-aminophenols in the

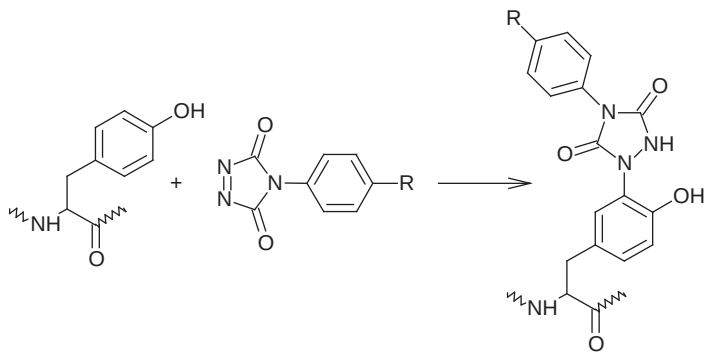
presence of sodium periodate (Scheme 24). The reaction is carried out in aqueous buffer at pH 6.5 and can reach high conversion in 2–5 minutes.[143]

Another promising study by C. F. Barbas and co-workers describes a new class of cyclic diazodicarboxamides that reacts chemoselectively with phenols therefore the phenolic side chain of tyrosine (Scheme 25). The efficiency of the ene-like reaction was investigated under mild aqueous milieu at pH range 2–10. The effectiveness of tyrosine ligation reaction was demonstrated in the synthesis of small molecule, peptide, enzyme, and antibody conjugates.[144]

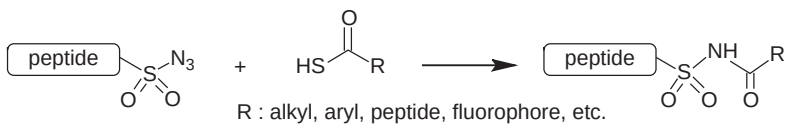
The chemoselective amidation reaction involving the reaction of an aminoethane sulfonyl azide with a thioacid developed by Williams and co-workers[145] was explored by D. T. S. Rijkers *et al.*[146] for peptide (Scheme 26) and Zhang *et al.*[147] for recombinant protein bioconjugation. The 'sulfo-click' reaction can be performed in nearly all organic and aqueous solvents/buffers and is not affected by unprotected functional groups. A major drawback of the method is that the free sulfhydryl groups in a competitive reaction can reduce the sulfonyl azide into the corresponding amide. Detailed protocols are provided for conjugate peptides or dendrimers with biophysical tags, fluorescent probes, metal chelators, and small peptides.



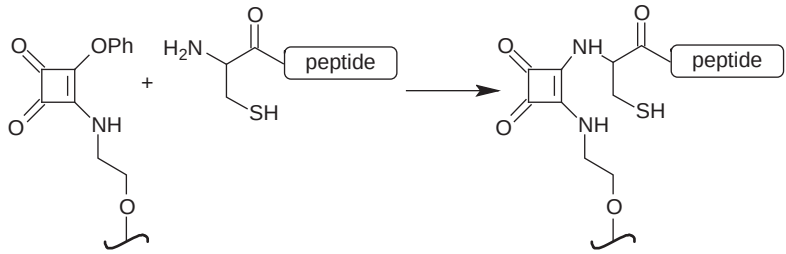
Scheme 24



Scheme 25



Scheme 26



Scheme 27

Immobilizing peptides or proteins enables the elucidation of ligand – receptor interaction in complex biological systems. A highly chemoselective reaction that captures peptides exclusively via *N*-terminus cysteine residue in a peptide is described by P. Sejwal *et al.*[148]. At pH 5.5, only *N*-terminal cysteines reacts with phenoxy amino squarate moieties presented on self-assembled monolayers (SAMs) of alkanethiols on gold films (Scheme 27). The selectivity of this surface reaction tolerate the presence of internal cysteines in close proximity to basic residues such as histidines.

3 PEGylation

The conjugation of different length of poly(ethyleneglycol) polymer chains (PEG) or oligo(ethyleneglycol) with defined number of units[149] to another molecule most typically to a drug or a therapeutic protein is commonly referred as PEGylation.

This technique using a wide variety of stable and releasable linkers can help to meet the challenges of improving the physico-chemical properties like solubility, therapeutic value of many drugs, peptides. The PEGylated compound may have favourably altered pharmacokinetic properties (*e.g.* blood clearance, biodistribution) over the unmodified biologics such as improved uptake, enhanced protection from proteolytic degradation, reduced toxicity, and enhanced physicochemical stability. Recently, this technology has been reviewed by R. R. Somani *et al.*[150] and G. Pasut and F. M. Veronese,[151] and will not be discussed in detail here.

4 Crosslinking agents

The increasing arsenal of commercially available crosslinking agents allows an appropriate selection for a specific application on the basis of the chemical reactivities of the bioconjugate components as well as of other relevant properties. The characteristics to be considered are for example the chemical specificity (homo- or heterobifunctional linker), the length of the spacer arm if needs any, the cleavability of spacer, the solubility of the linker in aqueous solution, cell membrane permeability etc. The reagent can be tailored in very specific and customized ways to optimize a wide range of applications. To find the best reagent Pierce has even developed an online interactive cross-linker selection guide.[152] However, it may happen that the

most suitable cross-linker for a specific conjugation reaction must be developed empirically.

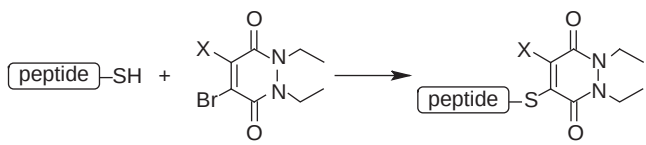
Recently, G. Leriche *et al.* presented an overview of the cleavable linkers used in chemical biology. The linkers were classified according to the cleavage mechanism for example by enzymes, nucleophilic/basic reagents, reducing agents, photo-irradiation, electrophilic/acidic reagents, organometallic and metal reagents, oxidizing reagents.[153]

Bromopyridazinedione-mediated protein and peptide bioconjugation was reported by V. Chudasama *et al.*[154] Bromopyridazinedione-mediated reversible bioconjugation to a single cysteine residue containing protein and somatostatin was demonstrated (Scheme 28, X = H, Br.) The conjugates are cleavable in an excess of thiol, including cytoplasmically-relevant concentrations of glutathione, and show a high level of hydrolytic stability. Both mono- and di-bromopyridazinediones can be used to mediate reversible protein and peptide bioconjugation.

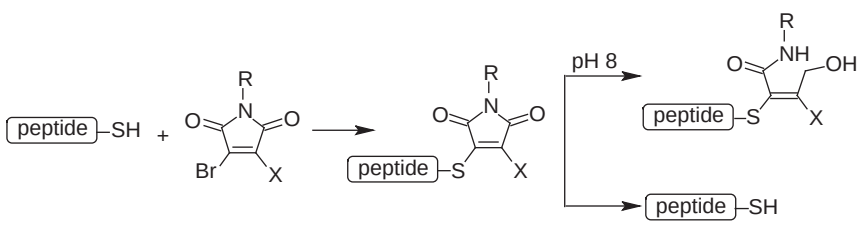
C. P. Ryan and co-workers reported on the development of a bromomaleimide reagent (Scheme 29). This compound, though controlled maleimide hydrolysis, allows the construction of bromomaleimide-mediated bioconjugates, which are either stable or cleavable in an aqueous, thiol-mediated reducing environment.[155]

This thiol-mediated cleavage can be considered for the design and application of a novel biotinylated affinity tag, which enables the successful pull-down of a single cysteine containing protein onto neutravidin beads. This strategy was further employed for the synthesis of both fluorophore-labelled and unlabelled, thiol-stable glycoprotein mimics and thiol-stable fluorescent labelling of the disulphide bond of somatostatin.

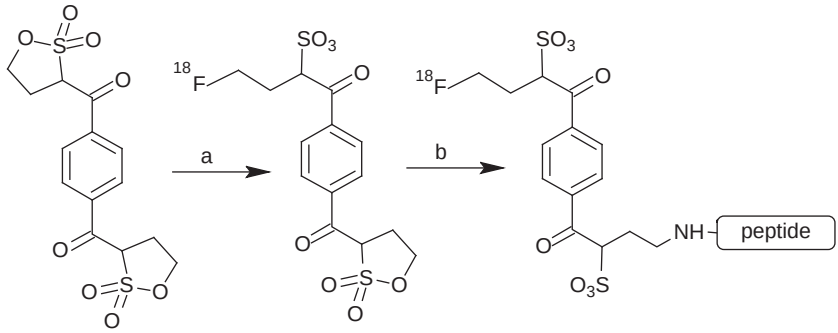
In parallel A. D. Baldwin and K. L. Kiick described that the succinimide thioether formed by the reaction of thiol with *N*-ethylmaleimide in the presence of other thiol compounds at physiological pH undergoes retro and exchange reactions.[156] The kinetics of this reaction could be modulated by the Michael donor's reactivity; therefore, the degradation of maleimide-thiol adducts could be tuned for controlled release.



Scheme 28



Scheme 29



Scheme 30

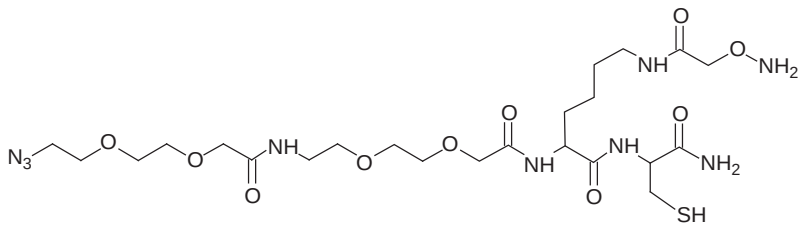


Fig. 3

T. Priem *et al*. reported on a novel homobifunctional cross-linker derived from a *bis*-sultone benzenic scaffold.[157] The utility of this bioconjugation reagent was demonstrated by the synthesis of a ^{18}F-labelled peptides Scheme 30. In the first step a nucleophilic fluorination *via* the ring-opening of a first sultone moiety takes place (a) followed by the nucleophilic ring-opening of the second sultone by the amine group of the peptide (b). Beyond the one-step radiolabelling of the peptide, the second main advantage of this approach is the release of free sulphonic acid moieties making the separation of the targeted [^{18}F]-tagged sulfonated compound from its non-sulfonated precursor easier and thus faster.

G. Clave and co-workers prepared a heterotrifunctional cross-linking reagent, whose full-orthogonality between its three bioconjugable functions was obtained without using temporary protecting groups[158] (Fig. 3). The chemoselective reactions were accomplished in the order of Michael addition, oxime ligation and the CuAAC reaction. To illustrate efficacy of this novel bioconjugation reagent, the construction of three different FRET cassettes based on cyanine and/or xanthene dye pairs and their subsequent grafting to biomolecules (*i.e*. aflatoxin B2, AFB2) and biopolymers (*i.e*., DNA fragments) were explored.

5 Selected applications of peptide-bioconjugates

Bioconjugates are widely used in various branches of biomedicinal research. Four main area could be identified: a) design and synthesis of drug-carrier conjugates to improve their efficacy/selectivity for potential therapeutic

use,[159,160] b) research related to synthetic antigens/immunogens for better understanding immunorecognition and to develop efficient synthetic vaccines and or target antigens for immunodiagnostics,[161,162] c) biosensor research including conjugates of radioisotope/fluorophore/spin label for intracellular analysis and proteomics,[163] d) nanomaterial based bioconjugates.[75] A novel need could be also noted: the construction of suitable bioconjugates to identify intracellular and cell surface membrane related structures, molecules to be considered as biomarkers even for targeting purposes.

During the last decades novel tendencies in drug discovery/development lead to the establishment of protein substances as drug candidates, even marketed drugs, such as monoclonal antibodies and their conjugates with chemotherapeutic agent or radioligand.[164] Also, mainly for the treatment of solid tumours, investigations on macromolecular drug-conjugates with synthetic polymer resulted in promising data.[165–167] Recently inspiring results with bioconjugates of porhpyrin for photodynamic studies were also published.[168]

On the other hand increasing interest could be observed in relation to the preparation, characterization and structure-function analysis of small or medium size oligopeptide conjugates with cell membrane targeting and/or cell penetrating properties.[169–172] This development could be due to certain concerns expressed in connection with protein drugs and protein-drug conjugates (*e.g.* difficulties in their analytical characterization, size related biodistribution, development of immune response during treatment, production cost).[173,174]

In the literature increasing number of papers are available on identification of membrane (biomarker) proteins with potential targeting ligand binding capacity. However, only limited number of studies was published on chemotherapeutic drug cargo-conjugates with novel ''cell targeting'' oligopeptides.[175]

This short and necessarily very sketchy outline could be indicative of the importance of bioconjugate chemistry both in basic biomolecular and also in the more applied drug research and documents its originality and international recognition.

5.1 Targeting with peptide bioconjugates

Imaging and therapeutic agents, although used for different purposes, both highly benefit from precise targeting. Advances in biology over the last years have profoundly impacted the elucidation of biological processes that occur at molecular or cellular level. This deeper insight into an increasing number of cellular pathways allows the identification of new potential peptides that may deliver their ''cargo'' to a diversity of cellular targets.

An ideal delivery vehicle would exhibit high binding affinity for the target, specific uptake and retention in the target and would be flexible for chemical modification. Additional criteria for application in living organisms are the rapid clearance from non-target tissue, adequate capillar permeability, and high stability *in vivo*, low immunogenicity and toxicity. Considering the roles peptides play in living systems they hold great promise for efficient targeting.

For example regulatory peptides that exert their action though their specific high-affinity binding receptors seem to be particularly good candidate for targeting. Many of these receptors in certain pathological conditions like cancer, inflammation, vascular and infectious diseases were shown to be massively overexpressed and thus target in diagnosis and therapy.

In cancer research - among many others - somatostatin (SST),[142,176–179] integrin,[180–183] gastrin-releasing peptide (GRP)[184–186] cholecystokinin (CCK),[187,188] gonadotropin-releasing hormone,[189–191] α-melanocyte stimulating hormone (α-MSH),[192] and glucagon-like peptide-1 (GLP-1)[193] receptors, just to mention a few, have been successfully applied in tumour receptor targeting. Recently, a daunomycin conjugate of an ErbB-2-binding oligopeptide was utilized to perform proteomic analysis for comparing the expression profiles of HL-60 cells after administration of cells with Dau-conjugate vs free drug.[194]

Successful techniques for target-specific peptide discovery involve the screening of phage display and combinatorial libraries.[195] These approaches were applied for the development of the delivery of drugs and tumour-imaging agents deep into the tumour parenchyma using tumour-homing peptides (iRGD).[196–198]

Incorporation of an appropriate linker between the targeting peptide and the labelling moiety may result in improved pharmacokinetic properties.

To improve peptide binding affinity, the concept of multivalency has been proposed. Originally, this approach involves the use of peptide homodimers or homomultimers in which peptide ligands of the same type are dimerized or ''multiplized'' with suitable linkers. Recently, a novel trend using peptide heterodimers has emerged as a promising method for targeting multi-receptor over-expressed cells.[199,200]

The combination of modern nanotechnology and molecular biology has yielded the emerging field of nanoplatform-based bioconjugates. Besides their unique physical properties, nanoparticles also afford a large surface area that allows the presentation of one, or even multiple types of functional groups.[75] Prospectively, this constructions may provide a platform that combines targeting ligands, imaging probes, and therapeutic drugs thus allowing simultaneous diagnosis, drug delivery, and monitoring of therapeutic response.

5.2 Molecular imaging by peptide bioconjugates

Molecular imaging by peptide bioconjugates, an emerging research field, enables the non-invasive visualization and quantification of biological processes at cellular even at molecular levels within living systems. In some applications real-time imagining of interaction between physiological targets and ligands can be achieved. Thus, this techniques hold great potential in basic research, early diagnosis, monitoring treatment efficacy and thus in drug/diagnostics discovery and development.

Upon labelling with an imaging moiety, peptides might suffer a significant loss in functionality (*e.g.* binding affinity, lowered quantum yield), thus the incorporation of an appropriate spacer between the active site and the reporter molecule or the conjugation at a position not involved in binding can be of utmost importance.

Recently, very comprehensive reviews on peptide-based probes for molecular imaging were published by research group of X. Chen.[201,202] The current state of development of peptide heterodimers as dual-targeted molecular probes is discussed by Y. Yan *et al*.[203] and their utilization for *in vivo* cancer imaging was discussed by Z. Liu and F. Wang.[204]

L. Zhu and co-workers reported a new type of *in vivo* fluorogenic probe that enables simultaneous and active targeting of overexpressed receptors, $\alpha_V\beta_3$ integrins, extracellular proteases, matrix metalloproteinases (MMPs), in the tumour.[200] This c(RGDyK)-conjugated MMP fluorogenic probe efficiently targeted both $\alpha_V\beta_3$ integrin- and MMP-overexpressing U87MG tumour-bearing mouse model.

The reviews by S. Lee and X. Chen,[205] by L. Jennings and N. J. Long[206] and also by J. Kuil *et al*.[207] evaluate the advantages and comprehensively list the challenges observed in the development dual-modality imaging probes and summarize the current state-of-the-art systems with emphasis on their unique design strategies and applications.

Based on the philosophy mentioned above a new term approach was born. ''Theranostics'' - portmanteau for therapeutics and diagnostics – in principle one concept for delivering therapy and monitoring the pharmaceutical effect. These type of conjugates constructed by a combination of diagnostics and therapeutics approach, offer promising tools that are capable of multiple functions *in vivo*, including targeting, image contrast, biosensing, and drug delivery. The review by S.S. Kelkar and T.M. Reineke overviewed the different theranostic strategies developed for the diagnosis and treatment of disease, with an emphasis on cancer.[208] In these conjugates therapeutic strategies such as nucleic acid delivery, chemotherapy, hyperthermia (photothermal ablation), photodynamic, and radiation therapy are combined with one or more imaging functionalities for both *in vitro* and *in vivo* studies.

5.2.1 Peptide-bioconjuagte in nuclear imaging. Nuclear imaging is the premier clinical imaging method due to the exceptional sensitivity. The most sensitive molecular imaging techniques are the radionuclide-based PET and SPECT. These nuclear imaging modalities are able to visualize biomolecules even at picomolar concentration in human body. Driven by the success of the first peptide-based imaging agent OctreoScan® - an ^{111}In labelled somatostatin analogue approved by FDA in 1994 for detection of somatostatin-producing tumours in the nervous and endocrine systems - peptide-based probes have become established methods in nuclear medicine. To develop radiolabelled peptide probes, the targeting peptide should be radiolabelled efficiently with high specific radioactivity and be stable under physiological conditions. Various techniques that enable the efficient radiolabelling of peptides via a chelating moiety or a prosthetic group have been developed. Several radionuclides, including ^{99m}Tc,^{123}I, ^{111}In, ,^{64}Cu, ^{68}Ga and ^{18}F for diagnostic use or ^{90}Y and ^{177}Lu for therapeutic use, have been employed for radiolabelling. Special attention has to be paid to restricted synthesis time window and specific labelling techniques due to the short physical half life of the radionuclides. Thus, the implementation of emerging synthesis methods like click reactions offers the possibility to design a wide variety of novel radiotracers.[101,209]

In their review M. Schottelius and H. J. Wester[210] summarized the peptide radiopharmaceuticals that are currently under clinical and preclinical investigation. Moreover critical steps and issues in the development and optimization of radiolabelled peptide-probes are discussed. R.D.G. Correia and co-worker overviewed the chemical strategies explored to improve the biological performance of different families of radiometallated peptides for nuclear molecular imaging and/or targeted radionuclide tumour therapy.[211] A practical guide to the construction of radiometallated bioconjugates for PET is provided by B. M. Zeglis and J. S. Lewis.[212] Recently, K. Chen and P. S. Conti[213] published an overview of the current status and trends in the development of target-specific peptide-based probes for PET imaging.

5.2.2 Peptide - bioconjugates in optical imaging, visualization. Optical imaging methods provide many advantages over other imaging modalities that include high sensitivity, use of nonradioactive agent, experimental simplicity and flexibility. Novel specific probes has been developed to allow visualization, characterization, and quantification of biologic processes both at the cellular and subcellular levels.

A wide selection of fluorophores with different excitation and emission ranges can be chosen on the basis of experimental conditions. The selection differs for *in vitro* and *in vivo* applications. Fluorophores in the visible range with emission between $\lambda = 400$ and 600 nm such as 7-amino-4-methylcoumarin (AMC), fluorescein isothiocyanate (FITC), and 5-carboxytetramethylrhodamine (TAMRA) are generally used for *in vitro* cellular imaging. Fluorophores which emit at the near-infrared (NIR) region ($\lambda =$ 650–900nm) are used for *in vivo* applications because autofluorescence of cells is less significant at longer wavelengths. The recent development of optical imaging enables the detection of fluorophores from picomolar to nanomolar concentrations offering an alternative of nuclear techniques.

In general, fluorophore molecules are sensitive to the environment, and their emissions may vary with changes of pH, solvent, and buffer components but more subtle alternations in their microenviroment such as changes in conformation, electronic structures can affect their photophysical properties. This potential drawback is exploited in the design of peptide-based fluorescent biosensors. In their recent review by E. Pazos *et al*. the different strategies employed for the preparation of fluorescent peptide based sensors were summarized and selected applications were discussed.[214]

Peptide-based chemosensors that rely on fluorescence changes upon phosphorylation are well-established tools in spatiotemporal, real-time monitoring protein kinase activity. The review by Juan A. Gonzalez-Vera summarizes the different fluorescent peptide-based strategies, introduces a few novel and attractive emerging assays, discusses their advantages and limitations, and highlights possible future directions.[215]

W. P. Heal *et al*.[216] reviewed the activity-based protein profiling (ABPP) one of the most powerful chemical proteomic technology with special emphasis on its application from drug target elucidation and *in vivo* visualisation of enzyme activity to comprehensive profiling of the catalytic content of living systems, and the discovery of new biological pathways.

A comprehensive review on bioorthogonal reactions in activity based profiling was provided by L. I. Willems and co-workers.[77] Another review by K. A. Kalesh surveys the contribution of Cu(I)-catalyzed 1,3-dipolar cycloaddition reaction in catalomics with special emphasis on activity-based protein profiling, posttranslational modifications and enzyme inhibitor developments.[76]

An interesting application of activable cell-penetrating peptides is the visualization tumours during surgery.[217] R. Y. Tsien *et al*. employed fluorescently labelled, polycationic cell-penetrating peptide coupled via a cleavable linker to a neutralizing peptide. Upon exposure to proteases characteristic of tumour tissue, the linker was cleaved, dissociating the inhibitory peptide and allowing the CPP to bind to and enter tumour cells. Surgery guided by this method resulted in fewer residual cancer cells, resulting in improved precision of tumour resection.

Q. T. Nguyen and co-workers using phage display techniques identified an oligopeptide, which binds selectively to peripheral nerves. The fluorescently labelled peptide derivative was able to delineate all peripheral nerves in mice without obvious toxicity. These results suggested that this conjugate could facilitate the surgical repair of injured nerves without accidental transection.[218]

An alternative approach was suggested for the introduction of fluorescent or chromophoric reporter molecules into peptides/protein by F. Diederich *et al*. Non-fluorescent/chromophoric precursors were conjugated to obtain fluorescence/chromophoric properties in the final product. The [2 + 2] cycloaddition–cycloreversion reaction afforded 1,1-dicyanobuta-1,3-diene-scaffolded peptides as new imaging chromophores.[219]

5.3 A few further applications

5.3.1 Biosensors. The field of biosensor research is constantly evolving to develop devices that have higher sensitivity and specificity, and are smaller, portable, and cost-effective. J. E. Dover summarized the recent advances in peptide probe–based biosensors for detection of infectious agents and evaluated the relative advantages and disadvantages of each technology.[220]

5.3.2 Peptide arrays. Screening of arrays and libraries of compounds is well-established as a high-throughput method for detecting and analysing interactions in both biological and chemical systems. Relatively short peptides enable the mapping the precise binding sites in such biocomplexes. Peptide array libraries usually contain partly overlapping fragments from the studied protein. The peptide library is attached to a solid support using various techniques such as SPOT-synthesis and photolithography. After incubation with the partner biomolecule the usually semi-quantitative detection of interacting peptides is carried out by using immunodetection assays. A recent review by Katz *et al*. discuss the potential and real advantages of the use of peptide arrays as a tool to study protein–protein interactions.[221] S.S.Li and C.Wu reviewed methods for identification of binding motifs and interaction networks in modular domains.[222] The very high-density peptide arrays and their fast development are reviewed by

F. Breitling *et al*. with an emphasis on methodologies.[223] M. Kohn suggested immobilization methods for small molecule, peptide and protein microarrays.[224]

5.3.3 Delivery systems. The growing number of potent therapeutic agents that can not enter into the clinic due to poor delivery and/or low bioavailability have made of delivery a vital element in therapeutic development . Several technologies have been designed to improve cellular uptake of therapeutic molecules, including cell-penetrating peptide bioconjugates. F.Heintz and his co-authors outlined the structure/function and cellular uptake mechanism of cell-penetrating peptides in the general context of drug delivery highlighted the application of peptide carriers and provided an update of their clinical evaluation.[225]

Aluri *et al*. summarized the peptide-based environmentally responsive delivery systems, which harnesses tumour pathology to trigger release of therapeutic agents at the target site. The review surveyed the different mechanisms of peptide-mediated tumour targeting : the enhanced permeability and retention effect, the mechanism of redox-, pH- ligand- and temperature mediated targeting.[226]

5.3.4 Peptide-bioconjugates as drug. There has been a rapid expansion in the use of peptides as drugs over the last years, and this is likely to continue. Most peptides cannot be administered orally as they are rapidly inactivated by gastrointestinal enzymes, but other forms of administration may be equally problematic. Bioconjugation can open an alternative route of delivery. Recent developments and a variety of examples has been recently summarized.[171,227] Other articles are providing valuable views and useful summary on advances synthetic peptide therapeutics[228] and pharmaceuticals.[229]

5.3.5 Dendric structures based peptide bioconjugates. J. Sebestik *et al*. in their review covered the broad area of peptide and glycopeptide dendrimers. The impact of cluster effect to biological, chemical, and physical properties is discussed in appropriate details. Synthesis of dendrimers by convergent and divergent approaches, ''Lego'' chemistry, ligation strategies, and click chemistry is given with many examples. Purification and characterization and their application is also surveyed. Biocompatibility and toxicity of dendrimers, as well as their applications in nanoscience, nanotechnology, drug delivery, and gene delivery are discussed critically.[230]

D. J. Welsh and D. K. Smith compared the covalent, dendric structure and non-covalent (micellar self-assembly) RGD ligand presentation strategies in an integrin bindig assay. Comparable efficiency in terms of achieving multivalent organisation of a ligand array was observed.[231]

H. F. Gaertner *et al*. developed a strategy to construct a two-chain covalent heterodendrimer (a ''diblock'') using cystamine core PAMAM dendrimers. In these structure two functionalities were orthogonally introduced. Finally, by incorporating a linker into the diblock architecture a third functional element were conjugated.[232]

N.A.Ulrich and co-workers synthesized peptide dendrimers of different generations *via* thioether ligation for studying multivalency in catalysis ester

hydrolysis or aldol reactions in aqueous medium. The efficiency of the esterase catalysts of G4, G5, G6 dendrimers was comparable to that of their lower generation analogues, a remarkable reactivity increase was observed in G5 and G6 aldolase dendrimers.[233]

Of late, L. Röglin and co-authors outlined several synthetic strategies and a few applications of monodispers, multivalent and multimodal dendrimers.[234]

5.4 Future perspectives of bioconjugates

Synthetic conjugation strategies of a molecule, not only a peptide, with useful functional properties (*e.g.* radionuclide, fluorophore, toxin proteins, chemotherapeutic drug) with a biological molecule (*e.g.* transport protein, monoclonal antibody, enzyme, hormone, receptor ligands, nucleic acid, polysaccharides) or with relevant bio-modified nanostructures (particles, liposomes, quantum dots, etc.) were developed during the last two decades and even most quickly during the last five years. These conjugates are mainly, but not exclusively for biomedical applications, like *in vitro* analysis, monitoring, diagnostics, *in vivo* diagnosis or even for therapy. The advantage of conjugation of a small molecule with solid surface by covalent linkage led also to the development of different array technologies, high throughput screening and many more discoveries. In this review we aimed to provide a summary on the classical organic chemistry derived approaches and their penetration into the field of peptide bioconjugation. There is hope that the reader by studying this chapter have received appropriate stimuli to enter this domain of bioscience with enthusiasm.

References

1 J. Lahann, ed., *Click Chemistry for Biotechnology and Materials Science* Wiley, 2009.
2 M. D. Best, *Biochemistry*, 2009, **48**, 6571.
3 S. I. Presolski, V. P. Hong and M. G. Finn, in *Current Protocols in Chemical Biology*, Wiley, 2011, vol. 3.
4 R. K. V. Lim and Q. Lin, *Chem Commun*, 2010, **46**, 1589.
5 E. M. Sletten and C. R. Bertozzi, *Angew Chem Int Ed*, 2009, **48**, 6974.
6 M. Boyce and C. R. Bertozzi, *Nat. Methods*, 2011, **8**, 638.
7 P. V. Chang, J. A. Prescher, E. M. Sletten, J. M. Baskin, I. A. Miller, N. J. Agard, A. Lo and C. R. Bertozzi, *Proc Natl Acad Sci U S A*, 2010, **107**, 1821.
8 K. E. Sapsford, K. M. Tyner, B. J. Dair, J. R. Deschamps and I. L. Medintz, *Anal Chem*, 2011, **83**, 4453.
9 S. Aggarwal, M. W. Ndinguri, R. Solipuram, N. Wakamatsu, R. P. Hammer, D. Ingram and W. Hansel, *Int J Cancer*, 2011, **129**, 1611.
10 G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, 2008.
11 S. S. Mark, ed., *Bioconjugation Protocols Strategies and Methods*, 2nd edn., Humana Press, 2011.
12 D. M. J. Shan S. Wong, *Chemistry of Protein and Nucleic Acid Cross-Linking and Conjugation*, 2nd edn., CRC Press, 2011.
13 J. Kalia and R. T. Raines, *Current Organic Chemistry*, 2010, **14**, 138.
14 R. J. Payne and C. H. Wong, *Chem Commun*, 2010, **46**, 21.
15 Y. Yuan, J. Chen, Q. Wan, R. M. Wilson and S. J. Danishefsky, *Peptide Science*, 2010, **94**, 373.

16 K. Lu, Q.-P. Duan, L. Ma and D.-X. Zhao, *Bioconjug Chem*, 2009, **21**, 187.
17 V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471.
18 E. Valeur and M. Bradley, *Chem Soc Rev*, 2009, **38**, 606.
19 A. El-Faham and F. Albericio, *Chem Rev*, 2011, **111**, 6557.
20 T. K. Tiefenbrunn and P. E. Dawson, *Biopolymers*, 2010, **94**, 95.
21 C. P. R. Hackenberger and D. Schwarzer, *Angew Chem Int Ed*, 2008, **47**, 10030.
22 P. Dawson, T. Muir, I. Clark-Lewis and S. Kent, *Science*, 1994, **266**, 776.
23 S. B. H. Kent, *Chem Soc Rev*, 2009, **38**, 338.
24 R. L. Yang, K. K. Pasunooti, F. P. Li, X. W. Liu and C. F. Liu, *J Am Chem Soc*, 2009, **131**, 13592.
25 H. Rohde and O. Seitz, *Biopolymers*, 2010, **94**, 551.
26 J. Offer, *Peptide Science*, 2010, **94**, 530.
27 E. Byun, J. Kim, S. M. Kang, H. Lee and D. Bang, *Bioconjug Chem*, 2011, **22**, 4.
28 A. M. Angeles-Boza, A. Erazo-Oliveras, Y.-J. Lee and J.-P. Pellois, *Bioconjug Chem*, 2010, **21**, 2164.
29 Z. Tan, S. Shang and S. J. Danishefsky, *Angew Chem Int Ed*, 2010, **49**, 9500.
30 J. Xiao and T. J. Tolbert, *Organic Letters*, 2009, **11**, 4144.
31 K. S. A. Kumar, L. Spasser, S. Ohayon, L. A. Erlich and A. Brik, *Bioconjug Chem*, 2011, **22**, 137.
32 J. W. Bode, R. M. Fox and K. D. Baucom, *Angew Chem Int E*, 2006, **45**, 1248.
33 I. Pusterla and J. W. Bode, *Angew Chem Int Ed*, 2012, **51**, 513.
34 J. Wu, J. Ruiz-Rodriguez, J. M. Comstock, J. Z. Dong and J. W. Bode, *Chem. Sci*, 2011, **2**, 1976.
35 X. Li, H. Y. Lam, Y. Zhang and C. K. Chan, *Organic Letters*, 2010, **12**, 1724.
36 F. B. Dyer, C.-M. Park, R. Joseph and P. Garner, *J Am Chem Soc*, 2011, **133**, 20033.
37 T. P. King, S. W. Zhao and T. Lam, *Biochemistry*, 1986, **25**, 5774.
38 K. Rose, *J Am Chem Soc*, 1994, **116**, 30.
39 H. F. Gaertner, K. Rose, R. Cotton, D. Timms, R. Camble and R. E. Offord, *Bioconjug Chem*, 1992, **3**, 262.
40 A. Dirksen, T. Hackeng and P. Dawson, *Angew Chem Int Ed*, 2006, **45**, 7581.
41 A. Dirksen, S. Dirksen, T. M. Hackeng and P. E. Dawson, *J Am Chem Soc*, 2006, **128**, 15602.
42 G. M. Eldridge and G. A. Weiss, *Bioconjug Chem*, 2011, **22**, 2143.
43 J.-Y. Byeon, F. T. Limpoco and R. C. Bailey, *Langmuir*, 2010, **26**, 15430.
44 J. B. Blanco-Canosa, I. L. Medintz, D. Farrell, H. Mattoussi and P. E. Dawson, *J Am Chem Soc*, 2010, **132**, 10027.
45 G. Iyer, F. Pinaud, J. Xu, Y. Ebenstein, J. Li, J. Chang, M. Dahan and S. Weiss, *Bioconjug Chem*, 2011, **22**, 1006.
46 F. M. Brunel, J. D. Lewis, G. Destito, N. F. Steinmetz, M. Manchester, H. Stuhlmann and P. E. Dawson, *Nano Letters*, 2010, **10**, 1093.
47 D. E. Prasuhn, J. B. Blanco-Canosa, G. J. Vora, J. B. Delehanty, K. Susumu, B. C. Mei, P. E. Dawson and I. L. Medintz, *ACS Nano*, 2010, **4**, 267.
48 A. R. Blanden, K. Mukherjee, O. Dilek, M. Loew and S. L. Bane, *Bioconjug Chem*, 2011, **22**, 1954.
49 M. R. Carrasco, C. I. Alvarado, S. T. Dashner, A. J. Wong and M. A. Wong, *J. Org. Chem.*, 2010, **75**, 5757.
50 L. Yi, H. Sun, Y.-W. Wu, G. Triola, H. Waldmann and R. S. Goody, *Angew Chem Int Ed*, 2010, **49**, 9417.
51 L. S. Witus, T. Moore, B. W. Thuronyi, A. P. Esser-Kahn, R. A. Scheck, A. T. Iavarone and M. B. Francis, *J Am Chem Soc*, 2010, **132**, 16812.

52 L. S. Witus and M. Francis, in *Current Protocols in Chemical Biology*, John Wiley & Sons, Inc., 2009, pp. 125.
53 E. H. M. Lempens, B. A. Helms, M. Merkx and E. W. Meijer, *ChemBioChem*, 2009, **10**, 658.
54 E. H. M. Lempens, M. Merkx, M. Tirrell and E. W. Meijer, *Bioconjug Chem*, 2011, **22**, 397.
55 O. Renaudet, D. Boturyn and P. Dumy, *Bioorg Med Chem Lett*, 2009, **19**, 3880.
56 H. Staudinger and J. Meyer, *Helvetica Chimica Acta*, 1919, **2**, 619.
57 E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007.
58 E. M. Sletten and C. R. Bertozzi, *Acc Chem Res*, 2011, **44**, 666.
59 J. A. Prescher, D. H. Dube and C. R. Bertozzi, *Nature*, 2004, **430**, 873.
60 B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Organic Letters*, 2000, **2**, 1939.
61 E. Saxon, J. I. Armstrong and C. R. Bertozzi, *Organic Letters*, 2000, **2**, 2141.
62 C. I. Schilling, N. Jung, M. Biskup, U. Schepers and S. Brase, *Chem Soc Rev*, 2011, **40**, 4840.
63 S. S. van Berkel, M. B. van Eldijk and J. C. M. van Hest, *Angew Chem Int Ed*, 2011, **50**, 8806.
64 V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew Chem Int Ed*, 2002, **41**, 2596.
65 C. W. Tornøe, C. Christensen and M. Meldal, *J Org Chem*, 2002, **67**, 3057.
66 R. Huisgen, *Angew. Chem Int Ed Eng*, 1963, **2**, 633.
67 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew Chem Int Ed*, 2001, **40**, 2004.
68 K. Holland-Nell and M. Meldal, *Angew Chem Int Ed*, 2011, **50**, 5204.
69 E. Lallana, R. Riguera and E. Fernandez-Megia, *Angew Chem Int Ed*, 2011, **50**, 8794.
70 T. Fricke, R. J. Mart, C. L. Watkins, M. Wiltshire, R. J. Errington, P. J. Smith, A. T. Jones and R. K. Allemann, *Bioconjug Chem*, 2011, **22**, 1763.
71 O. Avrutina, M. Empting, S. Fabritz, M. Daneschdar, H. Frauendorf, U. Diederichsen and H. Kolmar, *Organic & Biomolecular Chemistry*, 2009, **7**.
72 M. Sainlos, C. Tigaret, C. Poujol, N. B. Olivier, L. Bard, C. Breillat, K. Thiolon, D. Choquet and B. Imperiali, *Nat Chem Biol*, 2011, **7**, 81.
73 Q. Zeng, S. Saha, L. A. Lee, H. Barnhill, J. Oxsher, T. Dreher and Q. Wang, *Bioconjug Chem*, 2010, **22**, 58.
74 D. Banerjee, A. P. Liu, N. R. Voss, S. L. Schmid and M. G. Finn, *ChemBioChem*, 2010, **11**, 1273.
75 W. R. Algar, D. E. Prasuhn, M. H. Stewart, T. L. Jennings, J. B. Blanco-Canosa, P. E. Dawson and I. L. Medintz, *Bioconjug Chem*, 2011, **22**, 825.
76 K. A. Kalesh, H. Shi, J. Ge and S. Q. Yao, *Org Biomol Chem*, 2010, **8**.
77 L. I. Willems, W. A. van der Linden, N. Li, K.-Y. Li, N. Liu, S. Hoogendoorn, G. A. van der Marel, B. I. Florea and H. S. Overkleeft, *Acc Chem Res*, 2011, **44**, 718.
78 S. K. Mamidyala and M. G. Finn, *Chem. Soc. Rev.*, 2010, **39**, 1252.
79 P. S. Banerjee, P. Ostapchuk, P. Hearing and I. S. Carrico, *J Virol*, 2011, **85**, 7546.
80 K. G. Patel and J. R. Swartz, *Bioconjug Chem*, 2011, **22**, 376.
81 D. C. Kennedy, C. S. McKay, M. C. B. Legault, D. C. Danielson, J. A. Blake, A. F. Pegoraro, A. Stolow, Z. Mester and J. P. Pezacki, *J Am Chem Soc*, 2011, **133**, 17993.
82 V. Hong, N. F. Steinmetz, M. Manchester and M. G. Finn, *Bioconjug Chem*, 2010, **21**, 1912.

83 D. Soriano del Amo, W. Wang, H. Jiang, C. Besanceney, A. C. Yan, M. Levy, Y. Liu, F. L. Marlow and P. Wu, *J Am Chem Soc*, 2010, **132**, 16893.
84 S. W. Millward, R. K. Henning, G. A. Kwong, S. Pitram, H. D. Agnew, K. M. Deyle, A. Nag, J. Hein, S. S. Lee, J. Lim, J. A. Pfeilsticker, K. B. Sharpless and J. R. Heath, *J Am Chem Soc*, 2011, **133**, 18280.
85 S. Fabritz, D. Heyl, V. Bagutski, M. Empting, E. Rikowski, H. Frauendorf, I. Balog, W.-D. Fessner, J. J. Schneider, O. Avrutina and H. Kolmar, *Org Biomol Chem*, 2010, **8**.
86 M. Glaser and E. G. Robins, *J Labelled Comp Radiopharm*, 2009, **52**, 407.
87 C. Le Droumaguet, C. Wang and Q. Wang, *Chem Soc Rev*, 2010, **39**, 1233.
88 E. Lallana, A. Sousa-Herves, F. Fernandez-Trillo, R. Riguera and E. Fernandez-Megia, *Pharmaceutical Research*, 2012, **29**, 1.
89 N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J Am Chem Soc*, 2004, **126**, 15046.
90 J. C. Jewett, E. M. Sletten and C. R. Bertozzi, *J Am Chem Soc*, 2010, **132**, 3688.
91 J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. van Hest, D. J. Lefeber, P. Friedl and F. L. van Delft, *Angew Chem Int Ed*, 2010, **49**, 9422.
92 J. C. Jewett and C. R. Bertozzi, *Organic Letters*, 2011, **13**, 5937.
93 A. Kuzmin, A. Poloukhtine, M. A. Wolfert and V. V. Popik, *Bioconjug Chem*, 2010, **21**, 2076.
94 E. M. Sletten, H. Nakamura, J. C. Jewett and C. R. Bertozzi, *J Am Chem Soc*, 2010, **132**, 11799.
95 K. E. Beatty, J. D. Fisk, B. P. Smart, Y. Y. Lu, J. Szychowski, M. J. Hangauer, J. M. Baskin, C. R. Bertozzi and D. A. Tirrell, *ChemBioChem*, 2010, **11**, 2092.
96 S. T. Laughlin and C. R. Bertozzi, *ACS Chemical Biology*, 2009, **4**, 1068.
97 J. M. Baskin, K. W. Dehnert, S. T. Laughlin, S. L. Amacher and C. R. Bertozzi, *Proc Natl Acad Sci U S A*., 2010, **107**, 10360.
98 C. Besanceney-Webler, H. Jiang, T. Q. Zheng, L. Feng, D. S. del Amo, W. Wang, L. M. Klivansky, F. L. Marlow, Y. Liu and P. Wu, *Angew Chem Int Ed*, 2011, **50**, 8051.
99 R. D. Carpenter, S. H. Hausner and J. L. Sutcliffe, *ACS Medicinal Chemistry Letters*, 2011, **2**, 885.
100 H. L. Evans, R. L. Slade, L. Carroll, G. Smith, Q.-D. Nguyen, L. Iddon, N. Kamaly, H. Stockmann, F. J. Leeper, E. O. Aboagye and A. C. Spivey, *Chem Commun*, 2012, **48**, 991.
101 M. E. Martin, S. G. Parameswarappa, M. S. O'Dorisio, F. C. Pigge and M. K. Schultz, *Bioorg Med Chem Lett*, 2010, **20**, 4805.
102 M. A. Nessen, G. Kramer, J. Back, J. M. Baskin, L. E. J. Smeenk, L. J. de Koning, J. H. van Maarseveen, L. de Jong, C. R. Bertozzi, H. Hiemstra and C. G. de Koster, *J Proteome Res*, 2009, **8**, 3702.
103 G. H. Hur, J. L. Meier, J. Baskin, J. A. Codelli, C. R. Bertozzi, M. A. Marahiel and M. D. Burkart, *Chemistry & Biology*, 2009, **16**, 372.
104 H. E. Bostic, M. D. Smith, A. A. Poloukhtine, V. V. Popik and M. D. Best, *Chem Commun*, 2012, **48**, 1431.
105 L. A. Canalle, T. Vong, P. H. H. M. Adams, F. L. van Delft, J. M. H. Raats, R. G. S. Chirivi and J. C. M. van Hest, *Biomacromolecules*, 2011, **12**, 3692.
106 I. Kii, A. Shiraishi, T. Hiramatsu, T. Matsushita, H. Uekusa, S. Yoshida, M. Yamamoto, A. Kudo, M. Hagiwara and T. Hosoya, *Org Biomol Chem*, 2010, **8**, 4051.
107 P. Kele, G. Mezö, D. Achatz and O. S. Wolfbeis, *Angew Chem Int Ed*, 2009, **48**, 344.

108 D. E. Achatz, G. Mező, P. Kele and O. S. Wolfbeis, *ChemBioChem*, 2009, **10**, 2316.
109 J. M. Baskin and C. R. Bertozzi, *Aldrichimica Acta*, 2010, **43**, 15.
110 J. C. Jewett and C. R. Bertozzi, *Chem Soc Rev*, 2010, **39**, 1272.
111 M. F. Debets, S. S. van Berkel, J. Dommerholt, A. J. Dirks, F. P. J. T. Rutjes and F. L. van Delft, *Acc Chem Res*, 2011, **44**, 805.
112 M. F. Debets, C. W. J. van der Doelen, F. P. J. T. Rutjes and F. L. van Delft, *ChemBioChem*, 2010, **11**, 1168.
113 Y. Wang, W. J. Hu, W. Song, R. K. V. Lim and Q. Lin, *Organic Letters*, 2008, **10**, 3725.
114 Y. Wang, W. Song, W. J. Hu and Q. Lin, *Angew Chem Int Ed*, 2009, **48**, 5330.
115 W. Song, Y. Wang, Z. Yu, C. I. R. Vera, J. Qu and Q. Lin, *ACS Chemical Biology*, 2010, **5**, 875.
116 R. K. V. Lim and Q. Lin, *Acc Chem Res*, 2011, **44**, 828.
117 M. L. Blackman, M. Royzen and J. M. Fox, *J Am Chem Soc*, 2008, **130**, 13518.
118 D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox and A. Y. Ting, *J Am Chem Soc*, 2011, **134**, 792.
119 J. B. Haun, N. K. Devaraj, S. A. Hilderbrand, H. Lee and R. Weissleder, *Nat Nano*, 2010, **5**, 660.
120 J. B. Haun, N. K. Devaraj, B. S. Marinelli, H. Lee and R. Weissleder, *ACS Nano*, 2011, **5**, 3204.
121 Z. Li, H. Cai, M. Hassink, M. L. Blackman, R. C. D. Brown, P. S. Conti and J. M. Fox, *Chem Commun*, 2010, **46**, 8043.
122 R. Selvaraj, S. Liu, M. Hassink, C.-w. Huang, L.-p. Yap, R. Park, J. M. Fox, Z. Li and P. S. Conti, *Bioorg Med Chem Lett*, 2011, **21**, 5011.
123 R. Rossin, P. Renart Verkerk, S. M. van den Bosch, R. C. M. Vulders, I. Verel, J. Lub and M. S. Robillard, *Angew Chem Int Ed*, 2010, **49**, 3375.
124 N. K. Devaraj and R. Weissleder, *Acc Chem Res*, 2011, **44**, 816.
125 M. R. Karver, R. Weissleder and S. A. Hilderbrand, *Angew Chem Int Ed*, 2012, **51**, 920.
126 N. K. Devaraj, R. Weissleder and S. A. Hilderbrand, *Bioconjug Chem*, 2008, **19**, 2297.
127 H.-S. Han, N. K. Devaraj, J. Lee, S. A. Hilderbrand, R. Weissleder and M. G. Bawendi, *J Am Chem Soc*, 2010, **132**, 7838.
128 B. M. Zeglis, P. Mohindra, G. I. Weissmann, V. Divilov, S. A. Hilderbrand, R. Weissleder and J. S. Lewis, *Bioconjug Chem*, 2011, **22**, 2048.
129 R. Pipkorn, W. Waldeck, B. Didinger, M. Koch, G. Mueller, M. Wiessler and K. Braun, *J Peptide Sci*, 2009, **15**, 235.
130 M. Wiessler, W. Waldeck, C. Kliem, R. Pipkorn and K. Braun, *Int J Med Sci*, 2010, **7**, 19.
131 M. Wiessler, U. Hennrich, R. Pipkorn, W. Waldeck, L. Cao, J. Peter, V. Ehemann, W. Semmler, T. Lammers and K. Braun, *Theranostics*, 2011, **1**, 381.
132 S. S. van Berkel, A. J. Dirks, M. F. Debets, F. L. van Delft, J. J. L. M. Cornelissen, R. J. M. Nolte and F. P. J. T. Rutjes, *ChemBioChem*, 2007, **8**, 1504.
133 P. Laverman, S. A. Meeuwissen, S. S. van Berkel, W. J. G. Oyen, F. L. van Delft, F. P. J. T. Rutjes and O. C. Boerman, *Nuclear Medicine and Biology*, 2009, **36**, 749.
134 S. F. M. van Dongen, W. P. R. Verdurmen, R. J. R. W. Peters, R. J. M. Nolte, R. Brock and J. C. M. van Hest, *Angew Chem Int Ed*, 2010, **49**, 7213.
135 X. Ning, R. P. Temming, J. Dommerholt, J. Guo, D. B. Ania, M. F. Debets, M. A. Wolfert, G.-J. Boons and F. L. van Delft, *Angew Chem Int Ed*, 2010, **49**, 3065.

136 C. S. McKay, J. A. Blake, J. Cheng, D. C. Danielson and J. P. Pezacki, *Chem Commun*, 2011, **47**, 10040.
137 H. Stockmann, A. A. Neves and S. Stairs, K. M. Brindle and F. J. Leeper, *Org Biomol Chem*, 2011, **9**, 7303.
138 M. Galibert, P. Dumy and D. Boturyn, *Angew Chem Int Ed*, 2009, **48**, 2576.
139 M. Galibert, O. Renaudet, P. Dumy and D. Boturyn, *Angew Chem Int Ed*, 2011, **50**, 1901.
140 D. M. Beal, V. E. Albrow, G. Burslem, L. Hitchen, C. Fernandes, C. Lapthorn, L. R. Roberts, M. D. Selby and L. H. Jones, *Org Biomol Chem*, 2012, **10**, 548.
141 P. A. Ledin, F. Friscourt, J. Guo and G.-J. Boons, *Chemistry – A European Journal*, 2011, **17**, 839.
142 C.-B. Yim, I. Dijkgraaf, R. Merkx, C. Versluis, A. Eek, G. E. Mulder, D. T. S. Rijkers, O. C. Boerman and R. M. J. Liskamp, *J Med Chem*, 2010, **53**, 3944.
143 C. R. Behrens, J. M. Hooker, A. C. Obermeyer, D. W. Romanini, E. M. Katz and M. B. Francis, *J Am Chem Soc*, 2011, **133**, 16398.
144 H. Ban, J. Gavrilyuk and C. F. Barbas, *J. Am. Chem. Soc.*, 2010, **132**, 1523.
145 N. Shangguan, S. Katukojvala, R. Greenberg and L. J. Williams, *J Am Chem Soc*, 2003, **125**, 7754.
146 D. T. S. Rijkers, R. Merkx, C. B. Yim, A. J. Brouwer and R. M. J. Liskamp, *J Peptide Sci*, 2010, **16**, 1.
147 X. Zhang, F. Li, X.-W. Lu and C.-F. Liu, *Bioconjug Chem*, 2009, **20**, 197.
148 P. Sejwal, S. K. Narasimhan, D. Prashar, D. Bandyopadhyay and Y. Y. Luk, *J Org Chem*, 2009, **74**, 6843.
149 Á. Bartos, F. Hudecz and K. Uray, *Tetrahedron Letters*, 2009, **50**, 2661.
150 R. R. Somani, G. K. Shinde, P. Y. Shirodkar, T. Pal and P. Naik, *Pharma Rev*, 2009, **7**, 68.
151 G. Pasut and F. M. Veronese, *J Control Release* http://dx.doi.org/10.1016/j.jconrel.2011.10.037
152 www.piercenet.com
153 G. Leriche, L. Chisholm and A. Wagner, *Bioorg Med Chem*, 2012, **20**, 571.
154 V. Chudasama, M. E. B. Smith, F. F. Schumacher, D. Papaioannou, G. Waksman, J. R. Baker and S. Caddick, *Chem. Commun. (Camb)*, 2011 **47**, 8781.
155 C. P. Ryan, M. E. B. Smith, F. F. Schumacher, D. Grohmann, D. Papaioannou, G. Waksman, F. Werner, J. R. Baker and S. Caddick, *Chem. Commun. (Camb)*, 2011, **47**, 5452.
156 A. D. Baldwin and K. L. Kiick, *Bioconjug Chem*, 2011, **22**, 1946.
157 T. Priem, C. Bouteiller, D. Camporese, A. Romieu and P.-Y. Renard, *Org Biomol Chem*, 2012, **10**.
158 G. Clave, H. Volland, M. Flaender, D. Gasparutto, A. Romieu and P.-Y. Renard, *Org Biomol Chem*, 2010, **8**, 4329.
159 P. Khare, A. Jain, A. Gulbake, V. Soni, N. K. Jain and S. K. Jain, *Critical Reviews in Therapeutic Drug Carrier Systems*, 2009, **26**, 119.
160 C. Van Der Walle, ed., *Peptide and Protein Delivery*, Academic Press 2011.
161 N. L. Dudek, P. Perlmutter, M. I. Aguilar, N. P. Croft and A. W. Purcell, *Current Pharmaceutical Design*, 2010, **16**, 3149.
162 A. Jakab, G. Schlosser, M. Feijlbrief, S. Welling-Wester, M. Manea, M. Vila-Perello, D. Andreu, F. Hudecz and G. Mező, *Bioconjug Chem*, 2009, **20**, 683.
163 S. Pavan and F. Berti, *Anal Bioanal Chem*, 2012, **402**, 3055.

164 P. Sapra, A. T. Hooper, C. J. O'Donnell and H.-P. Gerber, *Expert Opinion on Investigational Drugs*, 2011, **20**, 1131.
165 R. Duncan, *Current Opinion in Biotechnology*, 2011, **22**, 492.
166 L. A. Canalle, D. W. P. M. Loewik and J. C. M. van Hest, *Chem. Soc. Rev.*, 2010, **39**, 329.
167 A. Vaidya, A. Agarwal, A. Jain, R. K. Agrawal and S. K. Jain, *Curr Pharm Des*, 2011, **17**, 1108.
168 F. Giuntini, C. M. A. Alonso and R. W. Boyle, *Photochem Photobiol Sci*, 2011, **10**, 759.
169 S. Majumdar and T. J. Siahaan, *Medicinal Research Reviews*, 2010, **32**, 637.
170 L. Jorgensen, H. M. Nielson and I. 978-0-470-72338-8, eds., *Delivery Technologies for Biopharmaceuticals: Peptides, Proteins, Nucleic Acids and Vaccines*, Wiley, 2009
171 G. Brendt, ed., *Peptides as Drugs*, Wiley, 2009.
172 Z. Bánóczi, A. Gorka-Kereskényi, J. Reményi, E. Orbán, L. Hazai, N. Tőkési, J. Oláh, J. Ovádi, Z. Béni, V. Háda, Cs. Szántay Jr., F. Hudecz, G. Kalaus and Cs. Szántay, *Bioconjug Chem*, 2010, **21**, 1948.
173 J. P. Stephan, K. R. Kozak and W. L. Wong, *Bioanalysis*, 2011, **3**, 677.
174 G. Schlosser, A. Jakab, G. Pocsfalvi, K. Vékey, F. Hudecz and G. Mező, *Rapid Commun Mass Spectrom*, 2009, **23**, 1249.
175 V. J. Venditto and E. E. Simanek, *Mol Pharm*, 2010, **7**, 307.
176 V. K. A. Sreenivasan, O. A. Stremovskiy, T. A. Kelf, M. Heblinski, A. K. Goodchild, M. Connor, S. M. Deyev and A. V. Zvyagin, *Bioconjug Chem*, 2011, **22**, 1768.
177 C. Wängler, B. Waser, A. Alke, L. Iovkova, H.-G. Buchholz, S. Niedermoser, K. Jurkschat, C. Fottner, P. Bartenstein, R. Schirrmacher, J.- C. Reubi, H.- J. Wester and B. Wängler, *Bioconjug Chem*, 2010, **21**, 2289.
178 H. Cai, Z. Li, C.-W. Huang, A. H. Shahinian, H. Wang, R. Park and P. S. Conti, *Bioconjug Chem*, 2010, **21**, 1417.
179 G. Mező, I. Szabó, I. Kertesz, R. Hegedus, E. Orbán, U. Leurs, S. Bősze, G. Halmos and M. Manea, *J Peptide Sci*, 2011, **17**, 39.
180 M. Schottelius, B. Laufer, H. Kessler and H.-J. r. Wester, *Acc Chem Res*, 2009, **42**, 969.
181 K. Chen and X. Chen, *Theranostics*, 2011, **1**, 189.
182 Z. Liu, F. Wang and X. Chen, *Theranostics*, 2011, **1**, 201.
183 R. H. Kimura, Z. Miao, Z. Cheng, S. S. Gambhir and J. R. Cochran, *Bioconjug Chem*, 2010, **21**, 436.
184 L. Mu, M. Honer, J. Becaud, M. Martic, P. A. Schubiger, S. M. Ametamey, T. Stellfeld, K. Graham, S. Borkowski, L. Lehmann, L. Dinkelborg and A. Srinivasan, *Bioconjug Chem*, 2010, **21**, 1864.
185 V. Sancho, A. Di Florio, T. W. Moody and R. T. Jensen, *Current Drug Delivery*, 2011, **8**, 79.
186 L. Hosta-Rigau, I. Olmedo, J. Arbiol, L. J. Cruz, M. J. Kogan and F. Albericio, *Bioconjug Chem*, 2010, **21**, 1070.
187 S. Roosenburg, P. Laverman, F. van Delft and O. Boerman, *Amino Acids*, 2011, **41**, 1049.
188 S. Roosenburg, P. Laverman, L. Joosten, A. Eek, W. J. G. Oyen, M. de Jong, F. P. J. T. Rutjes, F. L. van Delft and O. C. Boerman, *Bioconjug Chem*, 2010, **21**, 663.
189 M. Manea, U. Leurs, E. Orbán, Z. Baranyai, P. Ohlschläger, A. Marquardt, A. Schulcz, M. Tejeda, B. Kapuvári, J. Tóvári and G. Mező, *Bioconjug Chem*, 2011, **22**, 1320.

190 P. Schlage, G. Mező, E. Orbán, S. Bősze and M. Manea, *J Controlled Release*, 2011, **156**, 170.
191 I. Szabó, M. Manea, E. Orbán, A. Csámpai, S. Bősze, R. Szabó, M. Tejeda, D. Gaál, B. Kapuvári, M. Przybylski, F. Hudecz and G. Mező, *Bioconjug Chem*, 2009, **20**, 656.
192 G. Ren, S. Liu, H. Liu, Z. Miao and Z. Cheng, *Bioconjug Chem*, 2010, **21**, 2355.
193 Z. Wu, I. Todorov, L. Li, J. R. Bading, Z. Li, I. Nair, K. Ishiyama, D. Colcher, P. E. Conti, S. E. Fraser, J. E. Shively and F. Kandeel, *Bioconjug Chem*, 2011, **22**, 1587.
194 E. Orbán, M. Manea, A. Marquadt, Z. Bánóczi, G. Csík, E. Fellinger, S. Bősze and F. Hudecz, *Bioconjug Chem*, 2011, **22**, 2154.
195 K. C. Brown, *Current Pharmaceutical Design*, 2010, **16**, 1040.
196 K. N. Sugahara, T. Teesalu, P. P. Karmali, V. R. Kotamraju, L. Agemy, O. M. Girard, D. Hanahan, R. F. Mattrey and E. Ruoslahti, *Cancer Cell*, 2009, **16**, 510.
197 E. Ruoslahti, S. N. Bhatia and M. J. Sailor, *J Cell Biology*, 2010, **188**, 759.
198 P. Laakkonen and K. Vuorinen, *Integrative Biology*, 2010, **2**.
199 J. S. Josan, H. L. Handl, R. Sankaranarayanan, L. Xu, R. M. Lynch, J. Vagner, E. A. Mash, V. J. Hruby and R. J. Gillies, *Bioconjug Chem*, 2011, **22**, 1270.
200 L. Zhu, J. Xie, M. Swierczewska, F. Zhang, X. Lin, X. Fang, G. Niu, S. Lee and X. Chen, *Bioconjug Chem*, 2011, **22**, 1001.
201 S. Lee, J. Xie and X. Chen, *Chem Rev*, 2010, **110**, 3087.
202 S. Lee, J. Xie and X. Chen, *Biochemistry*, 2010, **49**, 1364.
203 Y. Yan and X. Chen, *Amino Acids*, 2011, **41**, 1081.
204 Z. Liu and F. Wang, *Current Pharmaceutical Biotechnology*, 2010, **11**, 610.
205 S. Lee and X. Chen, *Molecular Imaging*, 2009, **8**, 87.
206 L. E. Jennings and N. J. Long, *Chem Commun*, 2009, 3511.
207 J. Kuil, A. H. Velders and F. W. B. van Leeuwen, *Bioconjug Chem*, 2010, **21**, 1709.
208 S. S. Kelkar and T. M. Reineke, *Bioconjug Chem*, 2011, **22**, 1879.
209 C. Waengler, R. Schirrmacher, P. Bartenstein and B. Waengler, *Curr. Med. Chem.*, 2010, **17**, 1092.
210 M. Schottelius and H.-J. Wester, *Methods*, 2009, **48**, 161.
211 J. D. G. Correia, A. Paulo, P. D. Raposinho and I. Santos, *Dalton Transactions*, 2011, **40**.
212 B. M. Zeglis and J. S. Lewis, *Dalton Transactions*, 2011, **40**, 6168.
213 K. Chen and P. S. Conti, *Advanced Drug Delivery Reviews*, 2010, **62**, 1005.
214 E. Pazos, O. Vazquez, J. L. Mascarenas and M. Eugenio Vazquez, *Chem Soc Rev*, 2009, **38**, 3348.
215 J. A. Gonzalez-Vera, *Chem Soc Rev*, 2012.
216 W. P. Heal, T. H. T. Dang and E. W. Tate, *Chem Soc Rev*, 2011, **40**, 246.
217 Q. T. Nguyen, E. S. Olson, T. A. Aguilera, T. Jiang, M. Scadeng, L. G. Ellies and R. Y. Tsien, *Proc Natl Acad Sci U S A.*, 2010.
218 J. L. Whitney Ma Fau - Crisp, L. T. Crisp Jl Fau - Nguyen, B. Nguyen Lt Fau - Friedman, L. A. Friedman, B Fau - Gross, P. Gross La Fau, - Steinbach, R. Y. Steinbach, P Fau - Tsien, Q. T. Tsien Ry Fau - Nguyen and Q. T. Nguyen, *Nat Biotechnol*, 2011, **29**, 352.
219 D. T. S. Rijkers, F. de Prada López, R. M. J. Liskamp and F. Diederich, *Tetrahedron Letters*, 2011, **52**, 6963.
220 J. E. Dover, G. M. Hwang, E. H. Mullen, B. C. Prorok and S.-J. Suh, *J Microbiol Methods*, 2009, **78**, 10.

221 C. Katz, L. Levy-Beladev, S. Rotem-Bamberger, T. Rito, S. G. D. Rudiger and A. Friedler, *Chem Soc Rev*, 2011, **40**.
222 S. S. Li and C. Wu, *Methods Mol Biol*, 2009, **570**, 67.
223 F. Breitling, A. Nesterov, V. Stadler, T. Felgenhauer and F. R. Bischoff, *Molecular BioSystems*, 2009, **5**.
224 M. Kohn, *J Peptide Sci*, 2009, **15**, 393.
225 F. Heitz, M. C. Morris and G. Divita, *Br J Pharmacol*, 2009, **157**, 195.
226 S. Aluri, S. M. Janib and J. A. Mackay, *Advanced Drug Delivery Reviews*, 2009, **61**, 940.
227 M. Castanho and N. Santos, eds., *Peptide Drug Discovery and Development*, Wiley, 2011.
228 P. Vlieghe, V. Lisowski, J. Martinez and M. Khrestchatisky, *Drug Discovery Today*, 2010, **15**, 40.
229 C. L. Stevenson, *Current Pharmaceutical Biotechnology*, 2009, **10**, 122.
230 J. Sebestik, P. Niederhafner and J. Jezek, *Amino Acids*, 2011, **40**, 301.
231 D. J. Welsh and D. K. Smith, *Org Biomol Chem*, 2011, **9**.
232 H. F. Gaertner, F. Cerini, A. Kamath, A.-F. o. Rochat, C.-A. Siegrist, L. Menin and O. Hartley, *Bioconjug Chem*, 2011, **22**, 1103.
233 N. A. Uhlich, T. Darbre and J.-L. Reymond, *Org Biomol Chem*, 2011, **9**.
234 L. Roglin, E. H. M. Lempens and E. W. Meijer, *Angew Chem Int Ed*, 2011, **50**, 102.

Self-assembling peptide materials

Shuguang Zhang
DOI: 10.1039/9781849734677-00040

1 Introduction

1.1 Discovery of the first self-assembling peptide

While working on yeast genetics and protein chemistry and trying to understand a left-handed Z-DNA structure in the laboratory of Alexander Rich at Massachusetts Institute of Technology in 1989, I identified a protein *Zuotin* (Zuo in Chinese means left and there is a Z in it) for its ability to bind to left-handed Z-DNA in the presence of 400-fold excess of sheared salmon DNA that contains ubiquitous right-handed B-DNA and other form DNA structures, in it, there is a repeat segment with the sequence AEAEAKA-KAEAEAKAK, thus I named EAK16 for its amino acid composition and peptide length (Fig. 1). In the yeast zuotin, the sequence repeats 3 times.[1]

Initial computer modelling of this EAK16 sequence showed the structure to be a α-helix: its lysines and glutamic acids on the side-chains with *i, i + 3* and *i, i + 4* arrangements could form potential ionic bonds. As a curiosity, I wondered if this peptide could be synthesized and studied to satisfy my scientific curiosity. At the time, I could not at all predict that my curiosity led me into an entirely unexpected field of peptide materials, now thriving and many products and applications have been realized in diverse fields.

Alexander Rich who always has an open mind then strongly supported me to order the actual EAK16 peptide. For economic reasons, we ordered a custom synthesis through the Biopolymers Laboratory at Massachusetts Institute of Technology.

When the EAK16 peptide was first studied following the reported method using circular dichroism spectroscopy, an unexpected result occurred, instead of showing the expected computered modelled α-helix, the peptide formed an exceedingly stable β-sheet structure. Upon adding salt, a thin layer membrane-like substance and transparent materials occurred in the petri dish.[2,3]

I later met the late legendary Francis Crick and told him about the discovery. Crick suggested me to look under scanning electron microscope (SEM), as I did. It took me more than one year to understand how the seemingly soluble short peptides underwent self-assembly to form well-ordered nanofibers and scaffolds and to form naked-eye visible materials. I and my colleagues published the yeast *Zuotin* where the first self-assembling peptide that formed visible material was discovered.[1] Since then self-assembling peptide has been expanded in a number of directions in past two decades.[4–13]

1.2 Self-assembling peptide materials

Designer materials that are self-assembled molecule by molecule or atom by atom to produce novel supramolecular architectures belong to

Laboratory of Molecular Design, Canter for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA, 02139-4307, USA. E-mail: shuguang@MIT.EDU

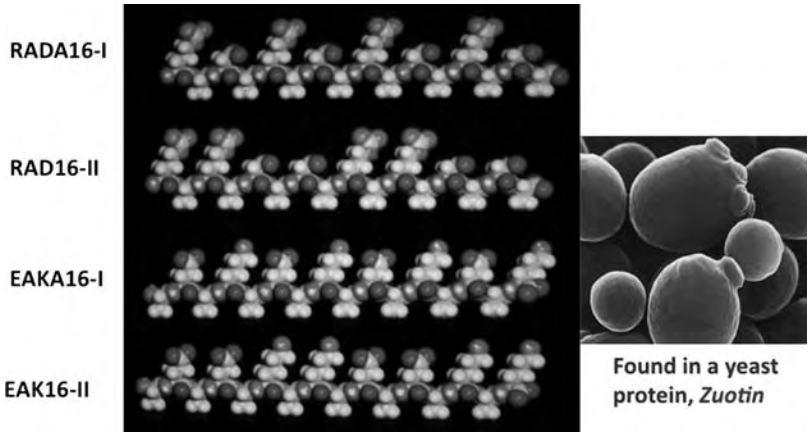


Fig. 1 The simple and molecular models of the designer amphiphilic self-assembling peptides that form well-ordered nanofibers. These peptides have two distinctive sides, one hydrophobic and the other hydrophilic. The hydrophobic side forms a double sheet inside of the fiber and the hydrophilic side forms the outside of the nanofibers that interacts with water molecules, forming an extremely high water content hydrogel that contains as high as 99.9% water. At least three types of molecules can be made, with $-$, $+$, $-/+$ on the hydrophilic side. The individual self-assembling peptide molecules are $\sim$5 nm long. The first such peptide, EAK16-II, was discovered from a yeast protein, zuotin.[1,2] This peptide inspired us to design a large class of self-assembling peptide construction motifs. When dissolved in water in the presence of salt, they spontaneously assemble into well-ordered nanofibers and then further into scaffolds.

"bottom-up" instead of "top-down" approach, and likely become an integral part of materials manufacture. This approach requires a deep understanding of individual molecular building blocks, their structures and dynamically assembly properties.[4,5]

These self-assembling peptides have alternating hydrophobic, namely, alanine, valine, leucine, isoleucine, and phenylalanine, and hydrophilic sides, which include positively charged lysine, arginine, histidine and negatively charged aspartic acids and glutamic acids.[4,5]

The complementary ionic sides have been classified into modulus I, II, III, IV and mixed moduli. This classification is based on the hydrophilic surface of the molecules that have alternating $+$ and $-$ charged amino acid residues, either alternating by 1, 2, 3, 4 and so on. For example, charge arrangements for the different moduli are as follows: modulus I, $-+-+-+-+$; modulus II, $--++\ \ --++$; modulus III, $---+++$; and modulus IV, $----++++$ (Fig. 2). The charge orientation can also be designed in reverse orientations that yield entirely different molecules with distinct molecular behaviors. These well-defined sequences allow them to undergo ordered self-assembly, resembling some situations found in well-studied polymer assemblies. This simple idea is the basis of the self-assembling peptide building blocks.

1.3 Basic chemical properties of the self-assembling peptide systems

The peptide synthesis has become more and more affordable that uses conventional mature solid phase or solution peptide synthesis chemistry. The peptide production cost directly correlates with the motifs length, purity of peptides, skill of the manufactures and the chirality of amino

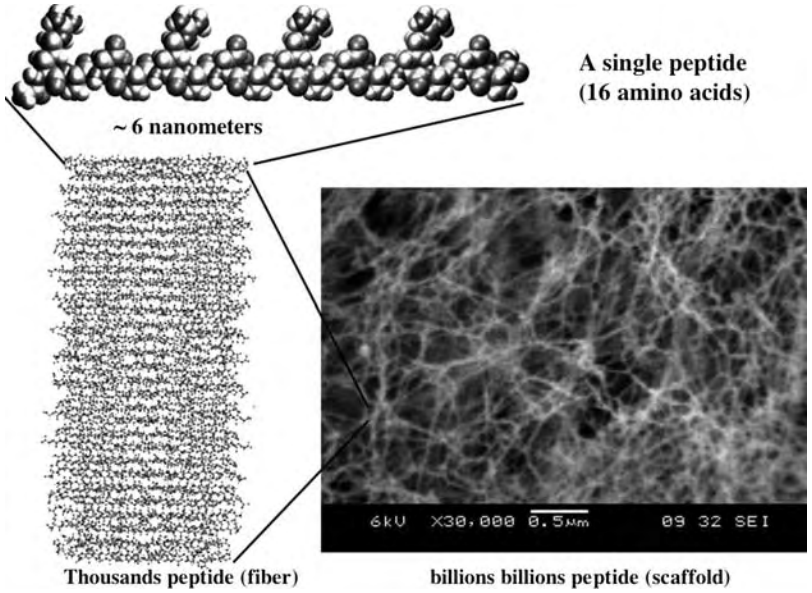


Fig. 2 Self-assembling peptide RADA16-I nanofiber scaffold hydrogel. Upper panel) Amino acid sequence of RADA16-I, molecular model of a single RADA16-I nanofiber, the calculated peptide dimensions are ~6 nm long depending on end capping, 1.3 nm wide and 0.8 nm thick; tens and hundreds of thousands of individual peptides self-assemble into a nanofiber, SEM images of RADA16-I nanofiber scaffold. Note the scale bar, 0.5 μm or 500 nm (SEM image courtesy of Fabrizio Gelain). RADA16-I Peptide form nanofibers in aqueous solution that further form hydrogel with extremely high water content (99.5–99.9% w/v water).

acids. Most of self-assembling peptides are readily soluble in water since their amino acid molecules consist of alternating hydrophilic and hydrophobic that contain 50% charged residues with distinct polar and non-polar surfaces and periodic repeats of 2–4 times. The self-assembly is accelerated by millimolar salt concentration under the physiological pH conditions to form ordered-nanostructure such as nanofiber, nanotube and nanovesicle.[2–8]

For example, RADA16-I and RADA16-II with arginine and aspartate residues replacing lysine and glutamate have been designed. The alanines form overlapping hydrophobic interactions in water, both positive Arg and negative Asp charges are packed together through intermolecular ionic interactions in a checkerboard-like manner. They self-assemble to form nanofibers ~10 nm in diameter, and theses nanofibers interwove into scaffolds that retain extremely high hydration, >99% in water (1 mg–10 mg/ml, w/v) (Fig. 3).[14]

The formation of the scaffold and its mechanical properties are influenced by several factors 1) amino acid sequence, 2) the level of hydrophobicity 3) length of the peptides, and 4) self-assembling time. For example to the extent of the hydrophobic residues, Ala, Val, Ile, Leu, Tyr, Phe, Trp (or single letter code, A, V, I, L, Y, P, W) can significantly influence the mechanical properties of the scaffolds and the speed of their self-assembly. The higher the content of hydrophobicity, the easier it is for scaffold formation and the better for their mechanical properties.[15]

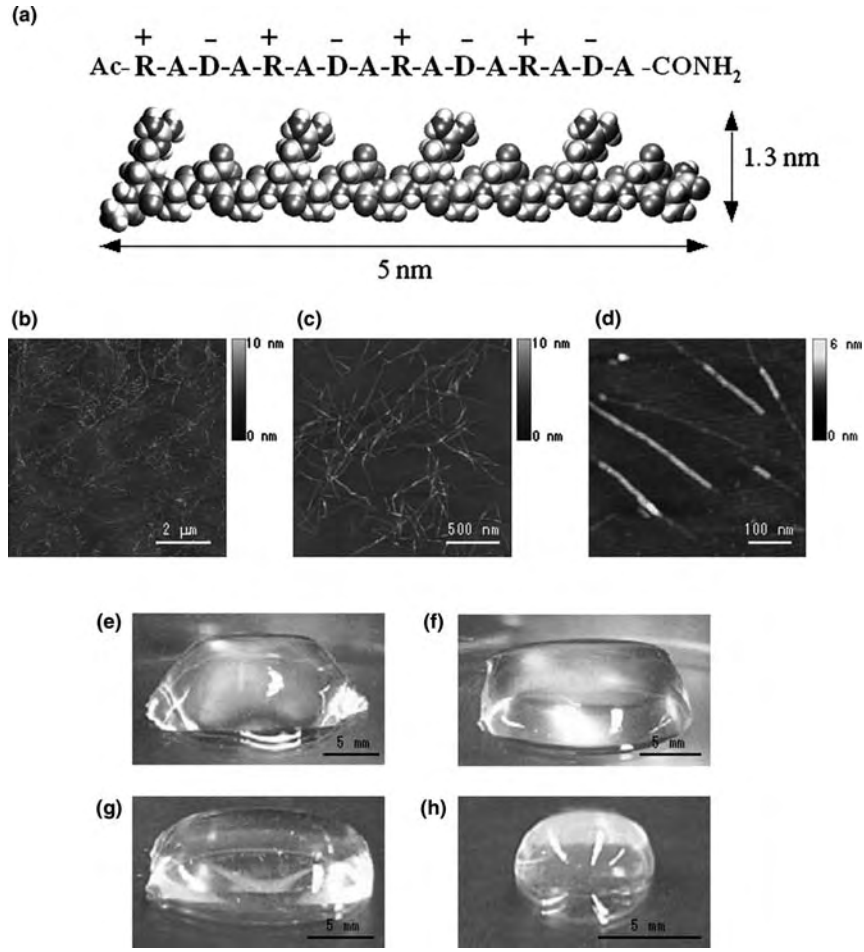


Fig. 3 Peptide RADA16-I. **a**) Amino acid sequence and molecular model of RADA16-I, the dimensions are ~5 nm long, 1.3 nm wide and 0.8 nm thick; **b**) AFM images of RADA16-I nanofiber scaffold, 8 μm^2, **c)** 2 μm^2 **d)** 0.5 μm^2. Note the different height of the nanofiber, ~1.3 nm, in **d** suggesting a double layer structure; Photographs of RADA16-I hydrogel at various condition, **e**) 0.5 wt% (pH 7.5), **f**) 0.1 wt% (pH 7.5, Tris.HCl), **g**) 0.1 wt% (pH 7.5, PBS) before sonication, **h**) re-assembled RADA16-I hydrogel after 4 times of sonication, respectively (image courtesy of Hidenori Yokoi).[14]

1.4 The proposed general assembly of nanofiber formations

The detailed mechanism how self-assembling peptides self-organize themselves in water at such low concentration still remains inadequately understood at present, nevertheless various peptides as biological materials have found diverse applications.

Several plausible ideas for different nanostructures have been proposed. First, for nanofibers, a molecular model to interpret the formation of EAK16 and RADA16 was proposed. These two peptides are representative of a class of peptides that undergo self-assembly into ordered-nanofibers: 1) numerous intermolecular hydrogen bonds between the peptides -C=O·····H-N- in the conventional β-sheets on the peptide backbones, 2) the side chains of positively and negatively charged residues form

intermolecular ionic bonds in a checkerboard manner, 3) hydrophobic interactions between the peptides from the amino acid residues pused by water molecules, 4) alternating polar and nonpolar surface interaction. It is known that salt ions facilitate the self-assembly, however it is not yet clear where the monovalent ions may coordinate the charged residues in a higher order of geometry.[14]

1.5 Dynamic reassembly of self-assembling peptides

The self-assembling peptides form stable β-sheet structure in water. The interactions between the peptides and β-sheets are 1) non-covalent hydrogen bonds along the backbones, 2) the arrays of ionic + and - charge interactions, 3) alanine hydrophobic interactions and van der Waals interactions, and 4) water-mediated hydrogen bond formations. Thus the nanofibers can be disrupted mechanically using sonication.[14] The self-assembling process is reversible and dynamic since these peptides are short and simple. Numerous individual peptides can be readily self-organized through the weak interactions. However, they can undergo dynamic reassembly repeatedly (Fig. 4), similar to the material self-healing process. Since the driving energy of the assembly in water is not only through hydrophobic van der Waals interactions, but also the arrays of ionic interactions as well as the peptide backbone hydrogen bonds, this phenomenon can be further exploited for production and fabrication of many self-assembling peptide materials.[14]

Unlike processed polymer microfibers in which the fragments of polymers cannot undergo reassembly without addition of catalysts or through material processing, the supramolecular self-assembly and reassembly event is likely to be wide spread in many unrelated fibrous biological materials where numerous weak interactions are involved. Self-assembly and reassembly are a very important property for fabricating novel materials, and it is necessary to fully understand its detailed process in order to design and to improve biological materials.

AFM images reveal that the nanofibers range from several hundred nanometers to a few microns in length before sonication. After sonication, the fragments were broken into ~20–100 nanometers. The kinetics of the nanofiber reassembly is closely followed at 1, 2, 4, 8, 16, 32 and 64 minutes as well as 2, 4, and 24 hours (Fig. 4). The nanofiber length reassembly is as a function of time: by 2 hours, the peptide nanofibers have essentially reassembled to their original length. The β-sheet structure had little change since the β-sheets at the molecular level remain unchanged despite the nanofiber length change.[14]

1.6 Molecular modeling of the self-assembly process

We carried out molecular modeling of the self-assembly process. For molecular modeling clarity, the RADA16-I β-sheet is presented as a non-twisted strand. It is known that these peptides form stable β-sheet structure in water, thus they not only form the intermolecular hydrogen bonding on the peptide backbones, but they also have two distinctive sides, one hydrophobic with array of overlapping alanines (Fig. 5, green color sandwiched inside), similar to what is found in silk fibroin or spider silk assemblies. The other side of the backbones has negatively charged (−)

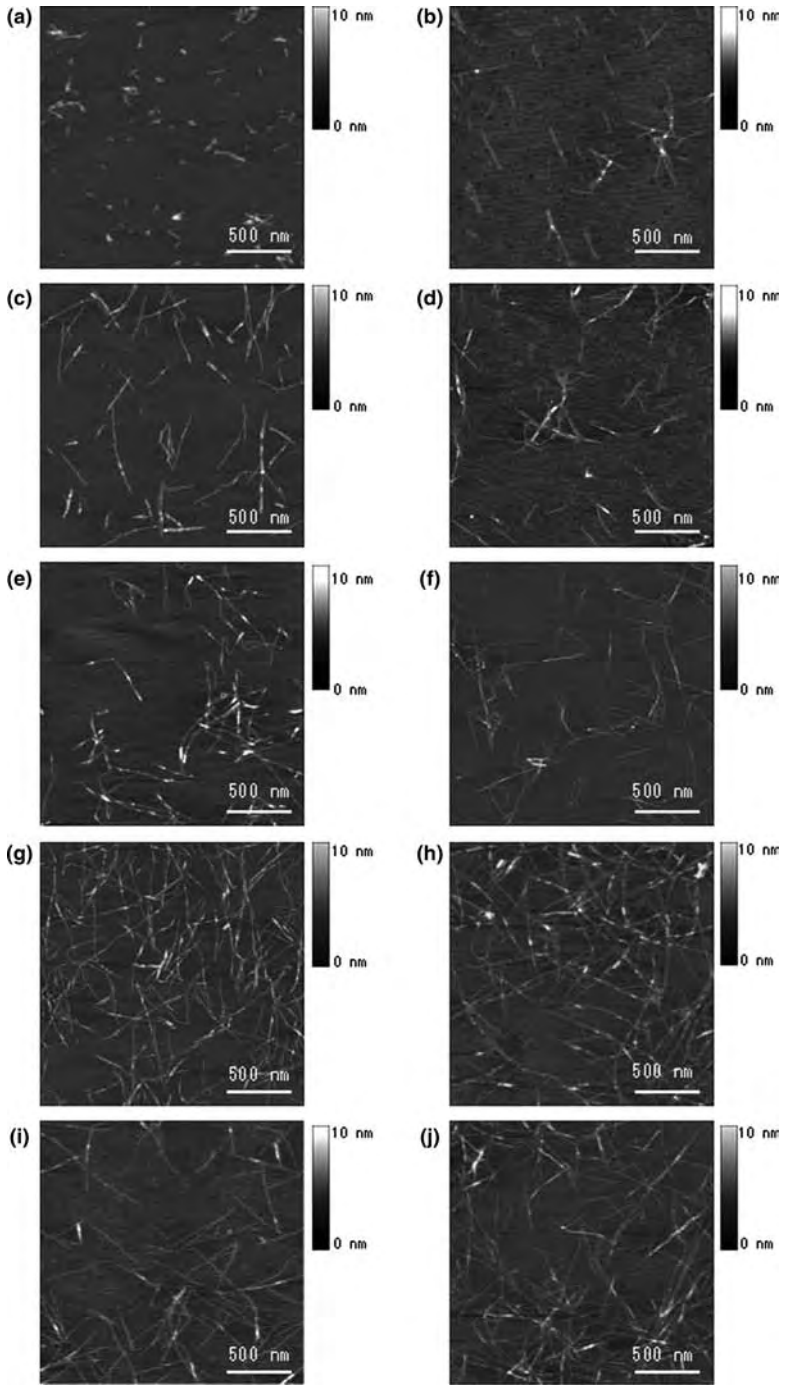


Fig. 4 AFM images of RADA16-I nanofiber at various time points after sonication. The observations were made using AFM immediately after sample preparation. **a)** 1 min after sonication; **b**) 2 min; **c**) 4 min; **d**) 8 min; **e**) 16 min; **f**) 32 min; **g**) 64 min; **h**) 2 hours; **i**) 4 hours; **j**) 24 hours. Note the elongation and reassembly of the peptide nanofibers over time. By ~1–2 hours, these self-assembling peptide nanofibers have nearly fully re-assembled (image courtesy of Hidenori Yokoi, PNAS).[14]

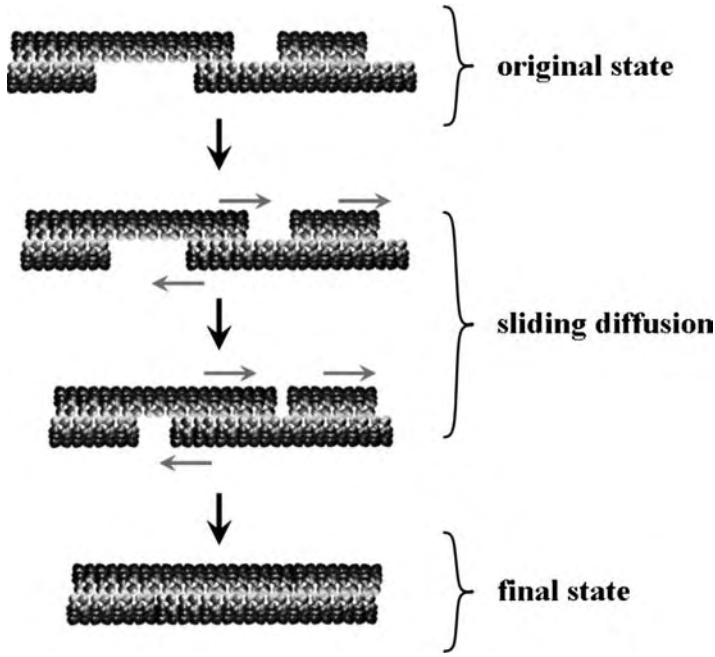


Fig. 5 A proposed molecular sliding diffusion model for dynamic reassembly of a single peptide nanofiber consisting thousands of individual peptides. When the peptides self-assemble into stable β-sheets in water, they form intermolecular hydrogen bonds along the peptide backbones. The β-sheet structure has two distinctive sides, one hydrophobic with an array of alanines and the other with negatively charged and positively charged amino acids. These peptides form anti-parallel β-sheet structures. The alanines form overlap packed hydrophobic interactions in water, a structure that is found in silk fibroin from silkworm and spiders. On the charged sides, both positive and negative charges are packed together through intermolecular ionic interactions in a checkerboard-like manner. When the fragments of nanofiber first meet, the hydrophobic sides may not fit perfectly but with gaps. However, the non-specific hydrophobic interactions permit the nanofiber to slide diffusion along the fiber in either direction that minimizes the exposure of hydrophobic alanines and eventually fill the gaps. The sliding diffusion phenomenon was also proposed for nucleic acids of polyA and polyU in 1956.[15–16] For clarity, these β-sheets are not presented as twisted strands. Color code: green, alanines; red, negatively charged amino acids; blue, positively charged amino acids [Colour image available on-line] (image courtesy of Hidenori Yokoi).[14]

amino acids, represented as red, and positively charged (+) amino acids, represented as blue.

The alanines form packed hydrophobic interactions in water, which can be disrupted mechanically during sonication. However, these hydrophobic cohesive ends could find each other quickly in water since the exposure of hydrophobic alanine arrays to water is energetically unfavorable. Since the hydrophobic alanines' interaction is non-specific, they can slide diffuse along the nanofiber, like trains sliding along train tracks. The same sliding diffusion phenomenon was also observed in early studies of nucleic acids where polyA and polyU form complementary base pairings that can slide diffuse along the double helical chains.[15,16] If however, the bases are heterogonous, containing G, A, T, C, then the bases cannot undergo sliding diffusion. Likewise, if the hydrophobic side of the peptides does not always contain alanine, containing residues such as valine and isoleucine, it would become more difficult for sliding diffusion to occur due to their structural constraint.

On the charged side, both positive and negative charges are packed together through intermolecular ionic interactions in a checkerboard-like manner. Likewise, the collectively complementary + and − ionic interactions may also facilitate reassembly. Similar to restriction-digested DNA fragments, these nanofiber fragments could form various assemblies like blunt and protruding ends. The fragments with various protruding ends as well as blunt ends can reassemble readily through hydrophobic and ionic interactions (Fig. 5).

2 General self-assembling peptide materials

There are the general peptide materials that are generic with no specific biological active motifs, in other words, not tailor-made for specific purpose. In an analogy, the general peptide materials are like pieces of bread without butter, cream spread or jam; or a bowl of pure rice without any added flavors.

There are also several classes of peptide mateirals, 1) like Legos, they self-assemble to form nanofibers; 2) like lipids or surfactants, they self-assemble into nanotubes and nanoviescles; 3) like biological paint, they self-assemble on surfaces to modify the surfaces.

2.1 Self-assembling peptide nanofibers

The first self-assembling peptide EAK16 and RADA16 are general peptides without active motifs. Here, the RADA16 made first milestone for new economic development. RADA16 has become commercial products for various 3D tissue cell cutlures (BD Biosciences, http://www.bdbiosciences.com/). It has also been successfully developed for various applications as medical devices.

The peptides form the stable second structures including α-helix and β-sheet. But some peptides secondary structures are more dynamic under various environmental factors: i) the amino acid sequence arrangements (even with the identical composition), ii) the molecular size of the peptide, iii) peptide concentration, iv) pH of the solution, v) temperature, vi) the medium composition, such as solvent or substrate, vii) ionic strength, and viii) the presence of denaturation agents, such as sodium dodecyl sulfate (SDS), urea and guanidium hydrochloride. These factors can significantly influence the dynamic behaviours of peptide secondary structures and also affect the process of self-assembly.[2–4]

2.2 Specific self-assembling peptide materials

Although self-assembling peptides are promising materials, they show no specific cell interaction because their sequences are not naturally found in living systems. In order to introduce the specific needs for individuals, the next logical step is to tailor-make the materials by directly coupling biologically active and functional peptide motifs onto the generic peptide (Fig. 6). Accordingly, the second generation of designer scaffolds will have significantly enhanced interactions with cells and tissues as well as other molecules.

We have made a wide range of designer self-assembling peptide nanofiber scaffolds for both 3D tissue cell cultures and sustained molecular releases.[16–23]

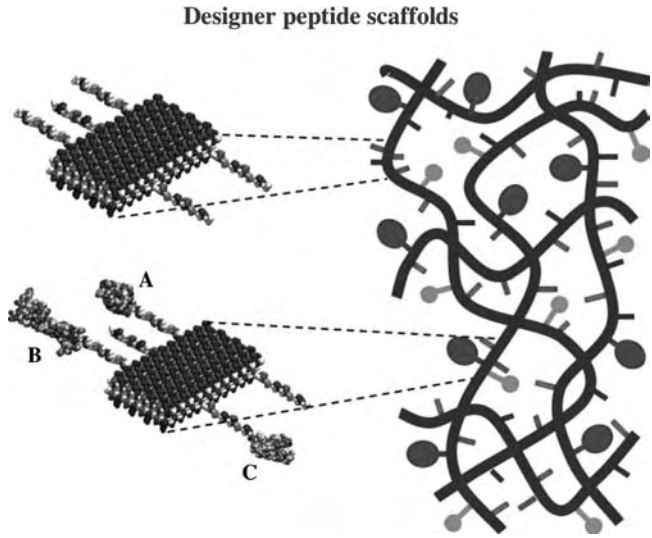


Fig. 6 Molecular and schematic models of the designer peptides and of the scaffolds. Direct extension of the self-assembling peptide sequence by adding different functional motifs. Light turquoise cylinders represent the self-assembling backbone and the yellow, pink, and tan lines represent various functional peptide motifs. Molecular model of a self-assembling peptide nanofiber with functional motifs flagging from both sides of the double b-sheet nanofibers. Either few or more functionalized and active peptide can be mixed at the same time. The density of these functionalized peptides can be easily adjusted by simply mixing them in various ratios, 1:1-1,000,000 or more before the assembling step. They then will be part of the self-assembled scaffold.[17] [Colour image available on-line]

The designer self-assembling peptides with additional active motifs play much more important and biological role and attractive biological materials with added functionalities and added values. Importantly, although additional active motifs were appended to the generic peptide backbone, the nanofiber structures remain quite similar as shown by SEM (Fig. 7).[17] Interestingly, the nanofiber structure is similar to the widely used Matrigel except the Matrigel is not pure with many other substances and growth factors in it. In some cases, when the active sequences become longer, the self-assembling speed and mechanical property are affected.[18–19,21–22] The 3D nanofiber materials are superb scaffolds for diverse tissue cells, not only for differentiation, proliferation, migration, but also for long-term maintenance.[17–19,21–22]

In order to fully understand how cells behave such as the 3-dimensional (3-D) microenvironment, 3-D gradient diffusion, 3-D cell migration and 3-D cell-cell contact interactions and in tissue engineering and regenerative medicine, it is important to develop a well-controlled 3-D tissue culture system where every single ingredient is known.

For over 100 years since the Petri dish was invented and used for tissue culture studies, almost all tissue cells have been studied on the 2D Petri dish and various formats of coated 2D surfaces. However, this 2D surface is rather unlike 3D tissue and the body's microenvironment. Thus it is important to develop a true 3D microenvironment to mimic the real tissue and body situation. The commonly used biomaterials are inadequate due to

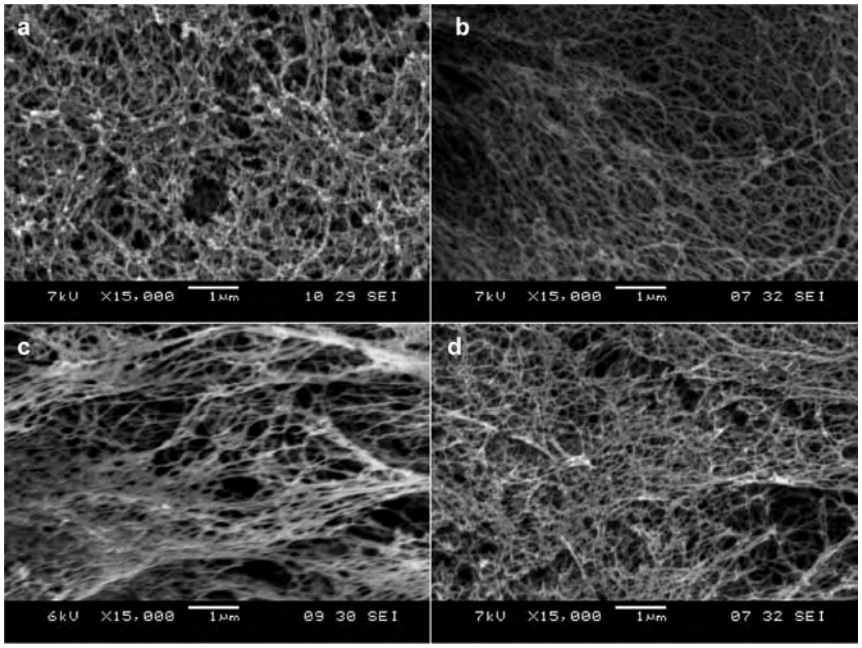


Fig. 7 SEM images of Matrigel and various designer peptide nanofiber scaffolds. a) Matrigel, b) RADA16, c) RADA16-BMHP1, d) RADA16-BMHP2 nanofiber scaffolds assembled in PBS solutions. Matrigel nanostructures are comparable in size to nanofibers found after self-assembly of the designer peptides. Clusters and aggregates of the unidentified naturally derived proteins in Matrigel (a) are absent in the pure peptide scaffolds shown in (b), (c) and (d). The interwoven nanofibers are $\sim$10 nm in diameter in each of the peptide scaffolds with $\sim$5 nm–200 nm pores. The appended functional motifs did not prevent peptide self-assembly.[17]

their microfiber and micropore size. Animal derived collagen gel and Matrigel contain other residue materials that are not always adequate for finely controlled studies. Thus, designer scaffold becomes more desirable (Fig. 8).

In order to achieve the fine-tuning and control, we have designed tailor-made peptides to suit for specific individual needs of studies or applications through appending specific active motifs onto the basic peptide, such as RADA16-I or EAK16-II.

From synthetic organic chemistry aspect, both of peptides C- or -N termini could be attached to the modified motifs. However, the functional motifs are always located on the C-termini because solid phase peptide synthesis initiates synthesis from the C-termini and proceeds toward N-terminus. The longer the peptide sequence is made, the more probable the coupling error would occur. Thus in order to avoid peptide synthesis errors, the active sequence motifs should always be at the C-terminus without exception (Fig. 6).[16–23]

Usually a spacer comprising 2-glycine residues is added to guarantee flexible and correct exposure of the motifs to cell surface receptors. If one combines a few designer peptides with different active motifs, these different functional motifs in various ratios can be incorporated in the same scaffold (Fig. 6). Upon exposure to solution at neutral pH, the functionalized sequences self-assemble, leaving the added motifs on both sides of each

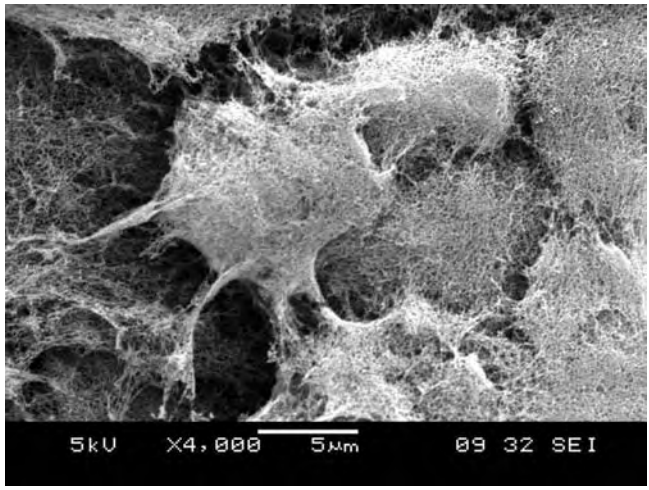


Fig. 8 SEM images of adult mouse neural stem cells (NSC) embedded in designer peptide nanofiber scaffold RADA16-BMHP1 (1% v/w) after 14 day *in vitro* cultures. Cluster of three visible mouse neural stem cells embedded in 3-D self-assembling RADA16-BMHP1. It is important to point out that the nanoscales of peptide scaffold and the extracellular matrix made by cells are indistinguishable. This is totally unlike the many processed biopolymer microfibers that are often in 10–50 μm in diameter, which is 1000–5000 bigger.[17]

nanofiber (Fig. 6). Nanofibers take part in the overall scaffold, thus providing functionalized microenvironments with specific biological stimuli (Fig. 6).

Self-assembling peptide scaffolds with functional motifs can be commercially produced at a reasonable cost. Thus, this method can be readily adopted for wide spread uses including the study of cell interactions with their local- and micro-environments, cell migrations in 3D, tumor and cancer cells interactions with normal cells, cell processes and neurite extensions, cell based drug screen assays and other diverse applications.

We have produced different designer peptides from a variety of functional motifs with different lengths. We showed that the addition of motifs in some cases to the self-assembling peptide RADA16-I did not significantly inhibit self-assembling properties. Furthermore, one can combine the RADA16-I nanofiber with the active designer self-assembling peptides by mixing the modified peptides. Although their nanofiber structures are indistinguishable from the RADA16-I scaffold, the appended functional motifs significantly influence cell behaviors.

Using the designer self-assembling peptide nanofiber system, every ingredient of the scaffold can be defined. Furthermore, it can be combined with multiple functionalities including soluble factors. Cells reside in a 3-D environment where the extracellular matrix receptors on cell membranes can bind to the functional ligands appended to the peptide scaffolds. It is likely that higher tissue architectures with multiple cell types, rather than monolayers, could be constructed using these designer 3-D self-assembling peptide nanofiber scaffolds.

Even if only a fraction of functionalized motifs on the 3-D scaffold are available for cell receptor binding, cells may likely receive more external stimuli than when in contact with coated 2-D Petri dishes or RGD-coated (or other motifs) polymer microfibers, which is substantially larger than the

cell surface receptors and in most cases, larger than the cells themselves. There, cells are not in a real 3-D environment, but rather, they are on a 2-D surface wrapping around the microfiber polymers with a curvature that depends on the diameter of the polymers. It is plausible in a 2-D environment, where only one side of the cell body is in direct contact with the surface, that receptor clustering at the attachment site may be induced; on the other hand, the receptors for growth factors, cytokines, nutrients and signals may be on the other sides that directly expose to the culture media. Perhaps cells may become partially polarized. In the 3-D environment, the functional motifs on the nanofiber scaffold surround the whole cell body in all dimensions. Thus growth factors may form a gradient in 3-D nanoporous microenvironment.

In our search for additional functional motifs, we found that a class of bone marrow homing peptides BMHP is one of the most promising active motifs for stimulating adult mouse neural stem cell (NSC) adhesion and differentiation. This observation suggests a new class of designer self-assembling peptides for 3-D cell biology studies.[17]

We attached several functional motifs including cell adhesion, differentiation and bone marrow homing motifs on the C-termini. We used them to study neural stem cell, and these functionalized peptides underwent self-assembly into nanofibers structure as well. More functionalized self-assembling peptides have been shown to promote specific cellular responses and long-term cell survival, proliferation, migration, and morphological differentiation, but these peptides are expensive for most applications.[17–23]

Designer self-assembling peptide scaffolds are also shown interesting interactions with functional proteins for study of molecular releases since the release kinetics results suggested that protein diffusion through nanofiber scaffolds primarily depended on the size of proteins.[24–27]

For lipid-like peptide surfactants, some expanded to choose 2-4 amino-acid to have useful applictions. Another way is from primary sequence geometry angle to design cone-shaped, Ac-GAVILRR-NH_2, has a hydrophilic head with two positive charges and a relatively large head size and a hydrophobic tail with decreasing hydrophobicity and side-chain size with a cone shape, can self-assembly of interesting nanodonut structure.

3 Diverse uses of self-assembling peptide nanofibers

A wide range of diverse uses in various areas have been developed from these self-assemlbing peptide materials. They include 1) reparative and regenerative medicine, 2) accelerated wound healing in human clinical studies, 3) sustained molecular releases (small molecules, protein growth factors and monoclonal antibodies), 4) stabilization of deiverse membrane proteins in solution and dry surface for nanobiotechnological device fabrications, and 5) production of G protein-coupled receptors. Some of they are highlighted below.

3.1 Reparative and regenerative medicine

The self-assembling peptides are easy to use in the tissue engineering, wound healing, addressing chronic wound problems, and regenerative medicine

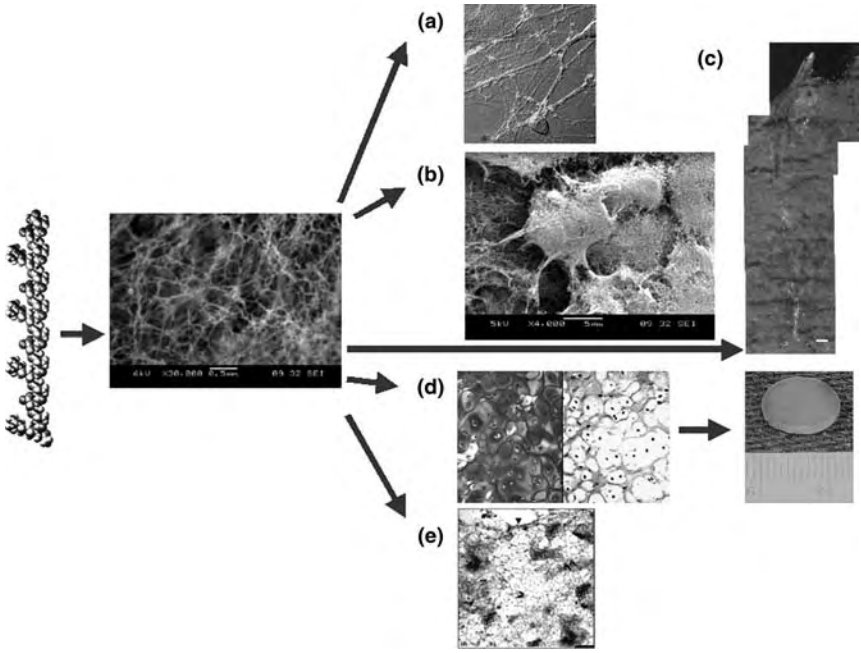


Fig. 9 From designer peptide to scaffold to tissues. **a)** Active synapses on the peptide surface. Primary rat hippocampal neurons form active synapses on peptide scaffolds. The confocal images shown bright discrete green dot labeling indicative of synaptically active membranes after incubation of neurons with the fluorescent lipophilic probe FM-143. FM-143 can selectively trace synaptic vesicle turnover during the process of synaptic transmission. The active synapses on the peptide scaffold are fully functional, indicating that the peptide scaffold is a permissible material for neurite outgrowth and active synapse formation. **b)** Adult mouse neural stem cells embedded in 3D scaffold (image courtesy of Fabrizio Gelain). **c)** Brain damage repair in hamster. The peptide scaffold was injected into the optical nerve area of brain that was first severed with a knife (image courtesy of Rutledge Ellis-Behnke). The gap was sealed by the migrating cells after a few days. A great number of neurons form synapses. **d)** Peptide KLD12 (KLDLKLDLKLDL), chondrocytes in the peptide scaffold and cartilage. The chondrocytes stained with TB showing abundant GAG production (left panel) and antibody to type II collagen demonstrating abundant Type II collagen production (right panel). A piece of premolded cartilage with encapsulated chondrocytes in the peptide nanofiber scaffold. The cartilage formed over a 3–4 week period after the initial seeding of the chondrocytes (image courtesy of John Kisiday). **e)** Von Kossa staining showing transverse sections of primary osteoblast cells on HA-PHP-RADA16-I self-assembling peptide nanofiber scaffold. Scale bar = 0.1 mm. The intensely stained black areas represent bone nodules forming (image courtesy of Maria Bokhari).[7]

(Fig. 9). Ellis-Behnke and colleagues used RADA16-I to repair injured rat and mouse brain structures, and their result showed the peptide scaffold hydrogel was excellent not only for axons to regenerate through the site of an acute brain injury but also to knit the injured brain tissue together seamlessly. This work represents an enabling nanobiomedical technology for tissue brain repair and restoration (Fig. 9).

Ellis-Behnke and colleagues found that the self-assembling peptide scaffolds are useful in repairing the damaged brain.[25] They also found that the peptide scaffold hydrogel instantly stopped bleeding in a few seconds during the procedure of repairing injured brain. They then expanded to study haemostasis of brain, spinal cord, femoral artery, and liver of rat.[26–28]

The early self-assembling peptide scaffold work has inspired others to expand to different kinds of tissues, organs and animals.[29–33] RADA16 functionalized with biologically active motifs also induced favourable reparative injured spinal cords.[34–35]

In order to better understand the individual molecular and material building blocks, their structures, assembly properties, dynamic behaviours and application for rapid haemostasis, we also used D-amino acids, the chiral self-assembling peptide d-EAK16 also forms 3-dimensional nanofiber scaffold. This study not only provided insights for understanding the chiral assembly properties for rapid haemostasis, but also to aid in further design of self-assembling D-form peptide scaffolds for clinical trauma emergency.[29–30]

3.2 Sustained molecular deliveries

Since the self-assembling peptides form the nanofiber scaffolds and its process is dynamic over the time, it is possible to use it for control drug delivery of molecular medicine, for small molecules and large proteins.[35–44]

Using EAK16 II, RAD16-II and RAD16-I as a model to slow release hydrophobic molecules showed that these types of self-assembling peptide scaffolds encapsulated hydrophobic drug. Various dye molecules including phenol red, bromophenol blue, 8-hydroxypyrene-1, 3, 6-trisulfonic acid trisodium salt, 1, 3, 6, 8-pyrenetetrasulfonic acid tetrasodium salt, and Coomassie Brilliant Blue G-250 through RADA16 hydrogels, providing an alternate route of controlled release of small molecules.[36] Furthermore, nanofiber encapsulated Camptothecin or ellipticine have confirmed to inhibit tumour growth.

The scaffolds have also used for sustained release of proteins including lysozyme, trypsin inhibitor, BSA, MMP-13 and monoclonal antibody IgG (Fig. 10),[37] active cytokines βFGF, TGF, VEGF and BDNF.[38] Furthermore, It can be used for sustained release from a few days to over 100 days (>3 months) when the experiments were terminated.[39] It is likely for sustained-release much longer since the content has not decrease

Fig. 10 Molecular representation of lysozyme, trypsin inhibitor, BSA, and IgG as well as of the ac-$(RADA)_4$-$CONH_2$ peptide monomer and of the peptide nanofiber. Color scheme for proteins and peptides: positively charged (blue), negatively charged (red), hydrophobic (light blue). Protein models were based on known crystal structures.[37] [Colour image available on-line]

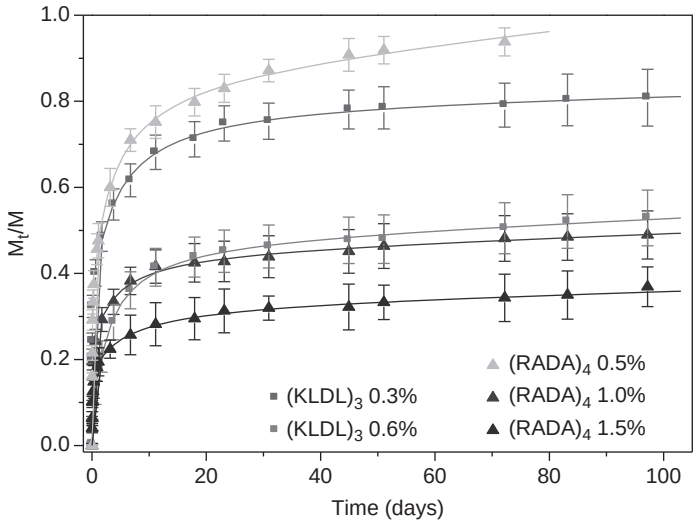


Fig. 11 The release profiles during the entire 3-month period for IgG through hydrogels of different peptides and different peptide nanofiber densities. Hydrogels consisted of the self assembling peptides (i) ac-(RADA)4-CONH2 with concentration 0.5% w/v (light blue, ▲), 1.0% w/v (blue, ▲), and 1.5% w/v (dark blue, ▲) and of (ii) ac-(KLDL)3-CONH2 with concentration 0.3% w/v (red, ■) and 0.6% w/v (magenta, ■). Release experiments were performed in PBS, pH 7.4 at room temperature. Data points represent the average of 5 samples.[39]

signficantly when the experiment results were collected (Fig. 11).[39] Future experiments will be carried out to test these ideas.

Our results not only provide evidence for long-term sustained molecular release from self-assembling peptide scaffolds, but also inspire others to design more self-assembling peptides to control molecular release for clinical applications.

3.3 Self-assembling peptides materials for human clinical studies of accelerated-wound healing

Human wound healing clinical process still seem to be similar to procedures used many years ago, use in suture during operations, try to prevent infection and other treatment measures. Accelerated-wound healing will benefit both patients and clinical workers. However, accelerated surgical and trauma wound healing especially chronical diabetic ulcer wound healing, is still problematic, the self-assembling medical technology may allieviate the problem and put to good practice.

The self-assembling peptide scaffolds have been found to be permissible for all tissue cell cultures including human cells tested, and the degradation products of the scaffolds are natural amino acids that pose no harm to human body. Furthermore a variety of stringent animal tests with rigorous controls have showed that the peptides did not elicit noticeable immune responses, nor caused measurable inflammatory reactions through injections and surgical procedures at various tissue sites. It thus encouraged people to carry out human clinical trial for accelerated-wound healing indication.

Of all subjects treated for various surgical wounds including cardiovascular surgeries for coronary artery bypass and synthetic blood vessel replacement, gastrointestinal surgery for partial hepatectomy and

gastrointestinal treatment for endoscopic excision of mucosa. Those clinical results are very encouraging and beneficial for patients. There were no observed adverse and undesirable side effects so far. This is not surprising since the designer self-assembling peptide scaffolds are totally pure synthetic amino acid based materials; there are no animal-derived impurities, no chemical and biological contaminations, no organic solvents, and no toxic compounds.

Since the successful human clinical trials, additional trials for several indications have been planned or launched for human tooth wound healing, skin wound healing from other diseases and injuries, in various parts of the world. Based on the previous successful clinical trials, it is anticipated that these clinical trials will likely be successful. It is hoped that the self-assembling peptide scaffolds will become an enabling medical technology that will truly benefit the society.

4 Surface modification self-assembling peptides

I also designed another class of self-assembling peptides that specifically assemble on surfaces to instantly change the physical, chemical and biological characteristics of surfaces.[45] In an analogy, it is like a grass lawn to cover the soil or a layer of paints to change the color and texture of surfaces. Except in this case, the coating is a nanometer layer of biological active peptides that can specifically interact with other molecules or cells. Not only we can pattern the cell for arbituary shapes, but cells can be confined in the region where they are seeded. An example of cells on the patterned surface is showing here (Fig. 12).

5 Lipid-like self-assembling peptide surfactant materials

The third classes of peptides are the lipid-like peptide surfactants, these peptides are similar to that of natural phospholipids, with tuneable hydrophobic tails to various degrees of hydrophobic amino acids such as alanine, valine or leucine, a hydrophilic head with either negatively charged aspartic and glutamic acids or positively charged histidine, lysine or arginine.[46–55] They undergo self-assembly to form well-ordered nanostructures including nanotubes, nanovesicles and micelles (Fig. 13).

5.1 A class of lipid-like self-assembling peptides

The lipid-like peptide is a new class of short peptides. Cationic, anionic, and zwitterionic peptide detergents were designed.[46–55] These lipid-like peptides are a class of molecules with properties similar to surfactants. They have hydrophilic heads comprised of 1–2 residues, and hydrophobic tails 3–6 residues long. They are about 2 nm-3 nm in length, and their ionic character and strength can be controlled by selecting appropriate amino acids or by capping the termini. Lysine or aspartic acid was used for the hydrophilic head. To control the detergent ionic nature, each peptide was capped by acetylation at the N-terminus, or selectively amidation at the C-terminus when required. Alanine, valine, leucine, and isoleucine were used for the hydrophobic tails (Figs. 13 and 14).

These lipid-like peptides behave comparably to traditional detergents, but offer several advantages over other novel detergents. Their chemical

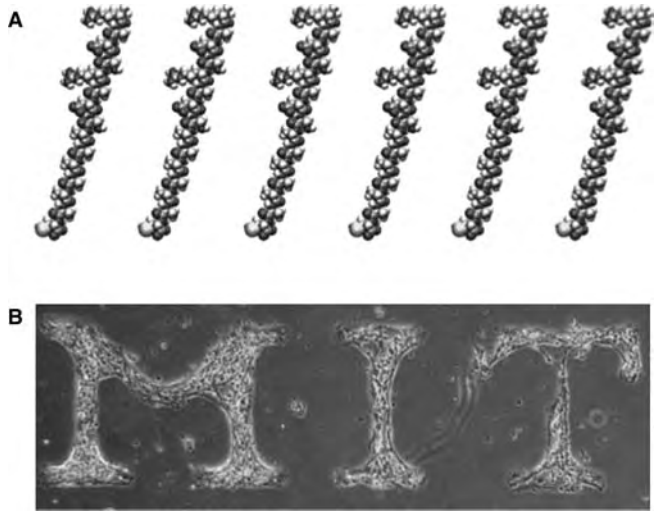


Fig. 12 Cells are patterned on the peptide coated surface. A) Molecular models of the surface self-assembling peptides. This type of peptide has three distinct segments: a biologically active segment where it interacts with other proteins and cells; a linker segment that can not only be flexible or stiff, but also sets the distance from the surface, and an anchor for covalent attachment to the surface.[45] These peptides can be used as ink for an inkjet printer to directly print on a surface, instantly creating any arbitrary pattern, as shown here. B) Mouse neural cells are seeded on the coated surface. The cells only attach where the adhesion surfaces and not the areas with non-adhesive substrate, such as oligoPEG.

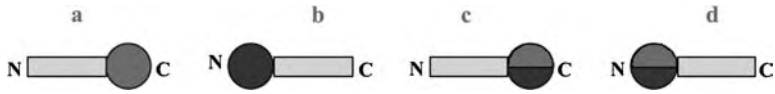


Fig. 13 The lipid-like peptides. These peptides have a hydrophilic head and a hydrophobic tail, much like lipids or detergents. They sequester their hydrophobic tail inside of micelle, vesicles or nanotube structures and their hydrophilic heads expose to water. At least three kinds molecules can be made, with $-$, $+$, $-/+$ heads and in 2 orientations.

properties are similar to commonly used detergents, they can be systematically designed and produced at high purity, and they remain stable for long periods of time.

5.2 Structure and of nanotube and nanovesicle formations

The lipid-like self-assembling peptides can form the ordered-nanotubes and nanovesicles (Figs. 15 and 16),[46–57] but the mechanism is not clear. A plausible path from the monomer state to the assemblies, two peptides form dimmer tail-to-tail packing to form a bilayer, monomeric peptides form small segments of the bilayer ring, with hydrophobic tails packing together to avoid water and hydrophilic heads exposed to water on the inner and outer portion of the tube, subsequently stack through non-covalent interactions to form longer nanotubes (Fig. 17).

I also designed with cone-shaped peptide Ac-GAVILRR-$CONH_2$ with large size amino acid arginin at the C-terminus and progressly reduce the size to glycine at the N-terminus.[51–52]

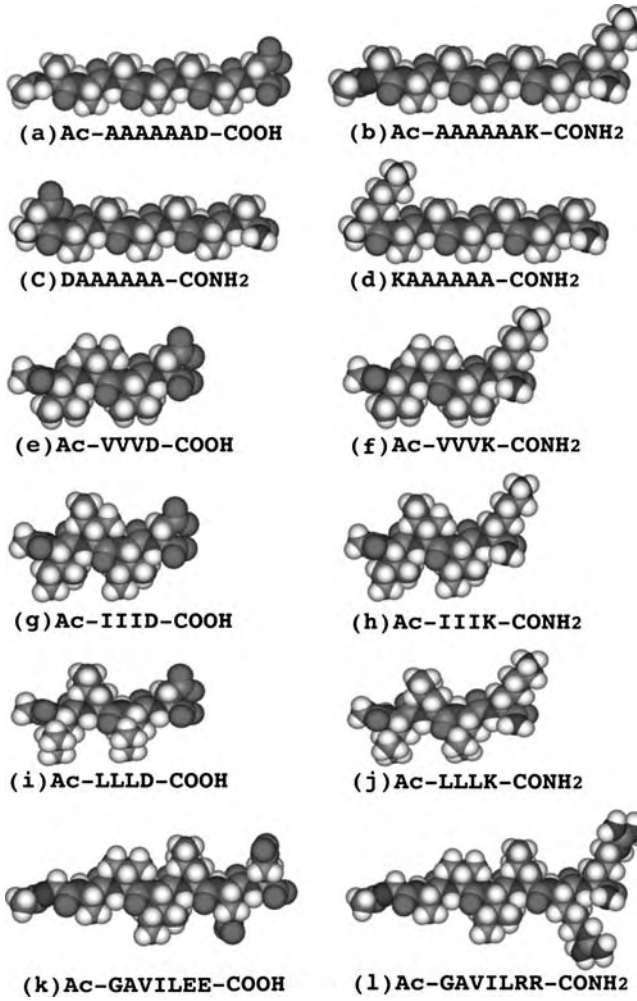


Fig. 14 Molecular models of peptide detergents at neutral pH. a) Ac-AAAAAAD-COOH. b) Ac-AAAAAAK-$CONH_2$. c) DAAAAAA-$CONH_2$. d) KAAAAAA-$CONH_2$. e) Ac-VVVD-COOH. f) Ac-VVVK-$CONH_2$. g) Ac-IIID-COOH. h) Ac-IIIK-$CONH_2$. i) Ac-LLLD-COOH. j) Ac-LLLK-$CONH_2$. k) Ac-GAVILEE, l) Ac-GAVILRR. Aspartic acid (D) is negatively charged and lysine (K) is positively charged. The hydrophobic tails of the peptide detergents consist of alanine (A), valine (V), isoleucine (I) and leucine (L). Each peptide is $\sim$2 nm–2.5 nm long, similar size to biological phospholipids. Color code: teal, carbon; red, oxygen; blue, nitrogen and white, hydrogen. [Colour image available on-line]

Charlotte A.E. Hauser and colleagues, have also discovered some ultra-small amphilic peptides as short as 3 to 7 residues that form strong hydrogels with a wide range of applications including tissue cell cultures, cosmetics, spine repair and more.[58–61] One of the peptides form very stable α-helical strucutre, one of the shortest peptide that could form helical structure.[56]

Further study of the mechanism is great importance because these lipid-like peptides have successfully stabilized diverse membrane proteins including GPCRs and used for molecular deliveries.

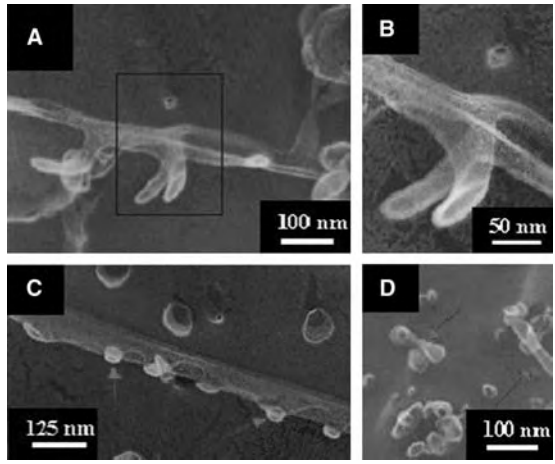


Fig. 15 High-resolution of TEM images of Ac-G_6D_2 showing different structures and dynamic behaviors of these structures. A) A pair of finger-like structure branching off from the stem. B) Enlargement of the box in (A), the detail opening structures are clearly visible. C) The openings (arrows) from the nanotube where may resulted in the growth of the finger-like structures. Some nanovesicles are also visible. D) The nanovesicles may undergo fission (arrows). (image courtesy of Dr. Steve Yang).[47]

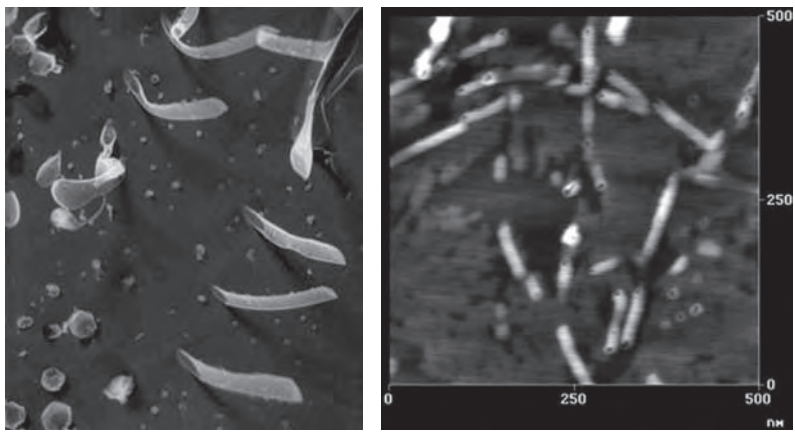


Fig. 16 (Left panel) TEM of nanotubes and nanovesicles of lipid-like peptide ac-VVVVVVD-OH in water. Micelles are also present (image courtesy of Dr. Steve Yang). (Right panel) The AFM image of nanotubes of A_6K lipid-like self-assembling peptides.[46] When the solution pH is less than the Lysine pKa 10, the peptide bears a positive charge. The openings of peptide nanotubes are clearly visible.[50] These nanotube structures can also undergo structural changes depending on various conditions, particularly pH changes, ionic strength of salts, temperature and incubation time. The other sheets like materials are likely the un-assembled peptides at the time of the image collected.

5.3 Stabilize membrane proteins

Membrane proteins play vital roles in all living systems. They involve in energy conversions, cell-cell and cell-environmental communications and sensing, specific ion channels and pumps, transporters, and all sorts of transports. Membrane proteins are also essential for our 5 senses: sight,

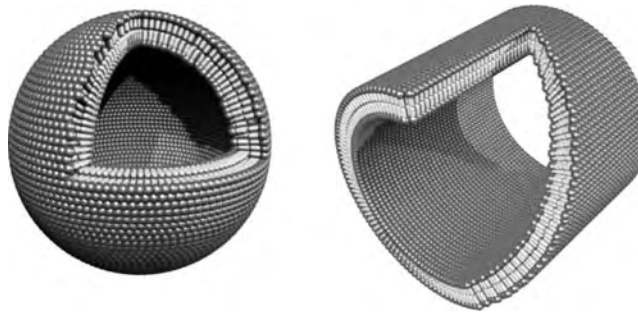


Fig. 17 Molecular modeling of cut-away structures formed from the peptides with negatively charged heads and non-polar tail. Peptide nanotube with an area sliced away. Peptide nanovesicle. Color code: red, negatively charged aspartic acid heads; green, non-polar tail. The non-polar tails are packed inside of the bilayer away from water and the aspartic acids are exposed to water, much like other surfactants. The modeled dimension is 50 nm–100 nm in diameter. [Colour image available on-line]

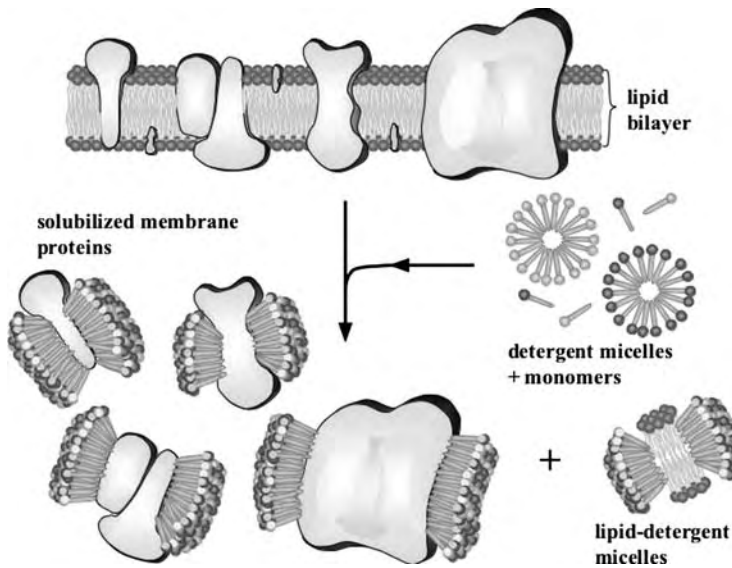


Fig. 18 A proposed scheme for how the designer lipid-like peptides stabilize membrane proteins. These simple designer self-assembling lipid-like peptides have been used to solubilize, stabilize and crystallize membrane proteins. These peptides have a hydrophilic head and a hydrophobic tail, much like other biological lipids. They use their tail to sequester the hydrophobic part of membrane proteins, and the hydrophilic heads exposed to water. Thus, they make membrane proteins soluble and stable outside of their native cellular lipid milieu. These lipid-like peptides are very important for overcoming the barrier of high resolutions of molecular structure for challenging membrane proteins.

hearing, smell, taste, touch and temperature sensing; and G-protein coupled receptors (GPCRs) are crucial in learning, memory, stem cell renewal and differentiation, body-plan development, the immune system, aging and more. However, our understanding of their structures and function falls far behind than that of soluble proteins (Fig. 18).

For past a few years, our teams gradually overcome some of the obstacles of the membrane proteins production and purification. We solubilized and

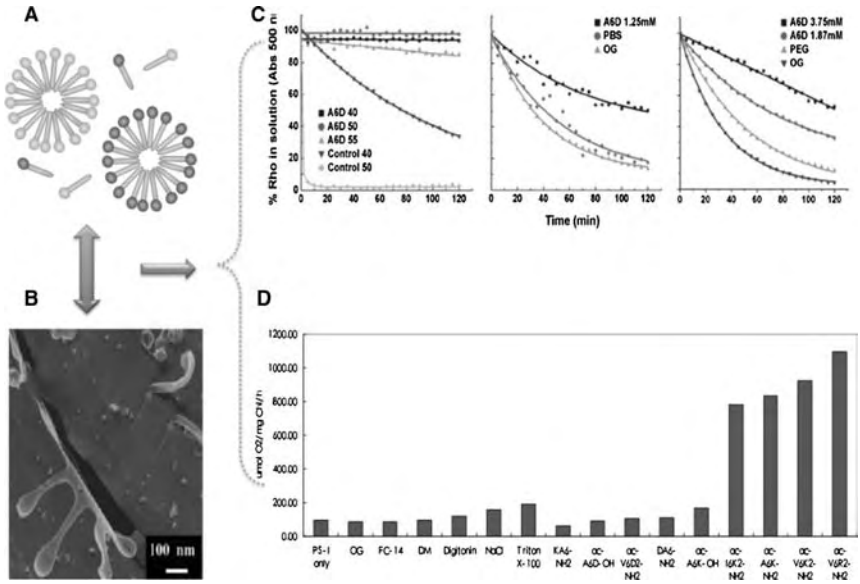


Fig. 19 Designer self-assembling peptides stabilize membrane proteins. (A) A proposed model peptide surfactants; (B) Quick-freeze/deep-etch TEM image of peptide surfactant (V_6D) dissolved in water; Stability kinetics of rhodopsin (Rho) in different surfactants; (C) Kinetics of bovine rhodopsin under different conditions. Stability of rhodopsin in the absence of OG at different temperatures. Half-life of rhodopsin was as follows: not available in 2.5 mM A_6D at 40 °C, 50 °C, and 55 °C (Left); (D) Decay of A500 in delipidated rhodopsin in the absence of OG. Half-life of rhodopsin was as follows: 122 min in 1.25 mM A_6D, 47 min in PBS, and 27 min 1% OG (Middle); Stability of delipidated rhodopsin at 40 °C (Right);[65] Lipid-like peptides stabilize functional photosystem I.[64]

stabilized several classes of membrane proteins using the lipid-like peptides, including *E.coli* Glycerol-3-phosphate dehydrogenase,[62] the multi-domain protein complex Photosystem-I (PSI) on surface in dry form[63] and in aqueous solution,[64] G protein-coupled receptor bovine rhodopsin (Fig. 19)[65] and many olfactory receptors.[66–68]

5.4 Lipid-like peptides used in cell-free productions of high yield membrane proteins

Cell-free production of proteins is a robust method and capable for molecular fabrications quickly, often in a few hours. It can easily be industrially standardized with rigorous quality controls. The technology is simple and versatile, unskilled person can be trained in a few hours to perform the task. Various functional proteins including membrane proteins, particularly difficult G-protein coupled receptors and other 7-transmembrane proteins (7TM), have been made for a variety of studies and for a wide range of applications.

Recently, cell-free protein productions have become cost effective and competitive. Grams amount of functional proteins and materials can be made in a few hours instead of weeks. Cell-free system can be integrated into other devices because it's simplicity and robustness.

Selecting the right surfactant is thus crucial because bottlenecks in elucidating the structure and function of membrane proteins are the difficulty

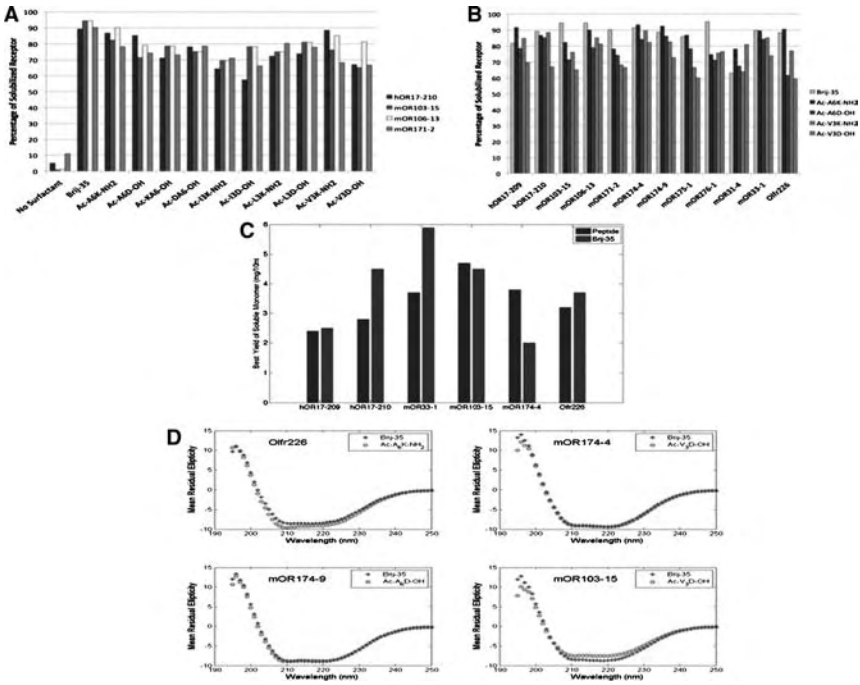


Fig. 20 Designer lipid-like peptides are used in the cell-free systems. Tens of functional olfactory receptors purification and secondary structure have been achieved. Olfactory receptors are solubility in Brij-35 and peptide detergents. Each receptor was expressed in the presence of Brij-35 or a peptide detergent using a commercial *E. coli* cell-free expression system (Qiagen, RiNA and Invitrogen). (A) The presence of a detergent was necessary to solubilize the olfactory receptors, and all of the peptide detergents were able to solubilize four unique receptors; (B) The detergent peptides and Brij-35 were able to solubilize similar fractions of protein. Peptides that were positively charged or had longer tails tended to solubilize higher fractions of receptors; (C) Detergent peptides can yield milligram quantities of solubilized olfactory receptors, and the maximum yield of the monomeric form of all tested olfactory receptors expected in a 10 ml reaction. Only results from the most effective detergent peptide are shown; (D) CD spectra of Brij-35 and peptide detergent-produced olfactory receptors to Olfr226, mOR174-4, mOR174-9 and mOR103-15 (Images courtesy of Karolina Corin).[66–67]

of producing large quantities of functional receptors. Various self-assembling peptide surfactants in commercial *E.coli* cell-free systems can rapidly produce milligram quantities of soluble G-protein coupled receptors (GPCRs) that include the human formyl peptide receptor (FPR), human trace amine-associated receptor (TAAR), vomeronasal type 1 receptor 1 (VNR), and other olfactory receptors.[66–67]

Furthermore, using short, designer lipid-like peptides as surfactants, we not only produced 12 unique mammalian olfactory receptors, but also solubilized and stabilized to maintain their structure and function (Fig. 20).[67–68] These simple and inexpensive lipid-like peptide surfactants will likely make significant contributions to facilitate the rapid production of GPCRs, and perhaps other membrane.

6 Summary

Since the unexpected discovery of the self-assembling peptide EAK16-II in yeast Zuotin, we have come a long way, from initial surprises, puzzlement,[1–3]

no understanding at all to, in outline, not only gradually understand the design principals at the molecular level, the molecular and fine material structures, interactions of the peptides, the dynamic self-assembly behaviours, but also how we can further improve their designs. From there, we not only subsequently expanded designer materials using 20 natural L-amino acids or some non-natural D-amino acids, but also we proceeded to optimize their sequence for 3D tissue cell cutlures,[18–24] delivering bioactive therapeutics such as drugs and growth factors and antibodies.[36–40] Recent advances in functionalization have also led to the development of better synthetic tissue culture bioactive scaffolds that promote cell proliferation, migration and differentiation for regenerative medicine. Furthermore, these self-assembling peptides have become an enabling medical technology[24–30,69–80] that will find a wide ranges of uses in surgery, emergency care, accelerated wound healing and more.

In science there are numerous examples of curiosity-driven research and unintentional discoveries that eventually lead to technological breakthroughs and new economic development. It is extremely difficult to imagine not to pursue the curiosity-driven research and explorations. Curiosity-driven research not only provide new insights into mysteries of nature and generate new knowledge, but it also can translate the knowledge into new enabling technology that will eventually be developed knowledge-based economy.

There are few examples: X-ray, the structure of DNA double helix, DNA-RNA and RNA-RNA hybridizations, reverse transcription, RNA splicing, RNA as enzymes, telomeres, Natural Killer cells (NK cells), programmed cell death, microRNA, RNA interference, S-layer proteins, Kuru disease, carbon 60, carbon nanotubes, carbon graphene and the indispensible worldwide web -- www. The discovery of the self-assembling peptide is another good example that an unexpected curiosity-driven discovery led to the development of an enabling medical technology that will benefit the society. The recent example of successful clinical trials of the self-assembling peptide scaffolds for accelerating wound healings and regenerative medicine provide a glimpse of what is coming for wide spread uses of chiral self-assembling peptides. Therefore curiosity-driven research must be strongly encouraged and fully supported, despite the current emphasis of application-driven research.

References

1 S. Zhang, C. Lockshin, A. Herbert, E. Winter and A. Rich, *EMBO. J*, 1992, **11**, 3787.
2 S. Zhang, T. Holmes, C. Lockshin and A. Rich, *Proc. Natl. Acad. Sci.*, 1993, **90**, 3334.
3 S. Zhang, C. Lockshin, R. Cook and A. Rich, *Biopolymers*, 1994, **34**, 663.
4 S. Zhang, *Biotechnology Advances*, 2002, **20**, 321.
5 S. Zhang, *Nat Biotechnol*, 2003, **21**, 1171.
6 X. Zhao and S. Zhang, *Chem Soc Rev*, 2006, **35**, 1110.
7 C. A. E. Hauser and S. Zhang, *Chem Soc Rev*, 2010, **39**, 2780.
8 C. A. E. Hauser and S. Zhang, *Nature*, 2010, **468**, 516.

9 S. Koutsopoulos, L. Kaiser, H. M. Eriksson and S. Zhang, *Chem Soc Rev*, 2011, **41**, 1721.
10 E. Loo, S. Zhang and C. A. E. Hauser, *Biotechnology Advances*, 2012, **30**, 593.
11 A. Lakshmanan, S. Zhang and C. A. E. Hauser, *Trends in Biotechnology*, 2012, **30**, 165.
12 E. Wu, S. Zhang and C. A. E. Hauser, *Advanced Function Materials*, 2012, **22**, 456.
13 Zhang, S. *Accounts of Chemical Research* 2012, **45,** (in press).
14 H. Yokoi, T. Kinoshita and S. Zhang, *Proc Natl Acad Sci USA*, 2005, **102**, 8414.
15 A. Rich and D. R. Davies, *J. Amer. Chem. Soc*., 1956, **78**, 3548.
16 G. Felsenfeld, D. R. Davies and A. Rich., *J. Amer. Chem. Soc*., 1957, **79**, 2023.
17 M. R. Caplan, E. M. Schwartzfarb, S. Zhang, R. D. Kamm and D. A. Lauffenburger, *Biomaterials*, 2002, **23**, 219.
18 S. Zhang, F. Gelain and X. Zhao, *Seminars in Cancer Biology*, 2005, **15**, 413.
19 F. Gelain, D. Bottai, A. Vescovi and S. Zhang, *PLoS One*, 2006, **1**, e119.
20 A. Horii, X. M. Wang, F. Gelain and S. Zhang, *PLoS One*, 2007, **2**, e190.
21 X. Wang, A. Horii and S. Zhang, *Soft Matter*, 2008, **4**, 2388.
22 Y. Chau, Y. Luo, A. C. Y. Cheung, Y. Nagai, S. Zhang, J. B. Kobler, S. M. Zeitels and R. Langer, *Biomaterials*, 2008, **29**, 1713.
23 Y. Kumada and S. Zhang, *PLoS One*, 2010, **5**, e10305.
24 Y. Kumada, N. A. Hammond and S. Zhang, *Soft Matter*, 2010, **6**, 5073.
25 R. G. Ellis-Behnke, Y. X. Liang and S. W. You, *et al*, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 5054.
26 R. G. Ellis-Behnke, Y. X. Liang and D. K. C. Tay, *et al*., *Nanomedicine, Nanotechnology, Biology and Medicine*, 2007, **2**, 207.
27 R. G. Ellis-Behnke and G. E. Schneider, *Methods Mol Biol*., 2011, **726**, 259.
28 J. Guo, K. K. G. Leung, H. Su, Q. Yuan, L. Wang, T.-H. Chu, W. Zhang, J. K. S. Pu, G. K. P. Ng, W. M. Wong, X. Dai and W. Wu, *Nanomedicine*, 2009, **5**, 345.
29 Z. Luo, X. Zhao and S. Zhang, *Macromol Biosci*, 2008, **8**, 785.
30 Z. Luo, S. Wang and S. Zhang, *Biomaterials*, 2011, **32**, 2013.
31 H. Song, L. Zhang and X. Zhao, *Macromole. Biosci*, 2009, **10**, 33.
32 H. Meng, L. Chen, Z. Ye, S. Wang and X. Zhao, *J Biomed Materials Research Part B: Applied Biomaterials*, 2008, **89B**, 379–391.
33 P. W. Kopesky, E. J. Vanderploeg, J. D. Kisiday, D. D. Frisbie, J. D. Sandy and A. J. Grodzinsky, *Tissue Eng Part A*, 2010, **17**, 83.
34 F. Gelain, S. Panseri, S. Antonini, C. Cunha, M. Donega, J. Lowery, F. Taraballi, G. Cerri, M. Montagna, F. Baldissera and A. Vescovi, *ACS Nano*, 2011, **5**, 227.
35 D. Cigognini, A. Satta, B. Colleoni, D. Silva, M. Donegà, S. Antonini and F. Gelain, *PLoS One*, 2011, **6**, e19782.
36 Y. Nagai, L. D. Unsworth, S. Koutsopoulos and S. Zhang, *J. Control Release*, 2006, **115**, 18.
37 S. Koutsopolous, L. D. Unsworth, Y. Nagai and S. Zhang, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 4623.
38 F. Gelain, L. D. Unsworth and S. Zhang, *J Control Release*, 2010, **145**, 231.
39 S. Koutsopoulos and S. Zhang *J. Control Release*, 2012, doi: 10.1016/j.jconrel.2012.03.014
40 L. Ruan, H. Zhang, H. Luo, J. Liu, F. Tang, Y. K. Shi and X. Zhao, *Proc Natl Acad Sci USA*., 2009, **106**, 5105.
41 F Tang and X. Zhao, *J Biomater Sci Polym Ed*., 2010, **21**, 677.
42 M. Wu, Z. Ye, Y. Liu, B. Liu and X. Zhao, *Mol Biosyst*., 2011, **7**, 2040.

43 J. Liu, L. Zhang, Z. Yang and X. Zhao, *Int J Nanomedicine*, 2011, **6**, 2143.
44 F. Li, J. Wang, F. Tang, J. Lin, Y. Zhang, E. Y. Zhang, C. Wei, Y. K. Shi and X. J. Zhao, *Nanosci Nanotechnol*, 2009, **9**, 1611.
45 S. Zhang, L. Yan, M. Altman, M. Lässle, H. Nugent, F. Frankel, D. Lauffenburger, G. M. Whitesides and A. Rich, *Biomaterials*, 1999, **20**, 1213–1220.
46 S. Vauthey, S. Santoso, H. Gong, N. Watson and S. Zhang, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 5355.
47 S. Santoso, W. Hwang, H. Hartman and S. Zhang, *Nano Letters*, 2002, **2**, 687.
48 G. von Maltzahn, S. Vauthey, S. Santoso and S. Zhang, *Langmuir*, 2003, **19**, 4332.
49 S. J. Yang and S. Zhang, *Supramolecular Chemistry*, 2006, **18**, 389–396.
50 A. Nagai, Y. Nagai, H. Qu and S. Zhang, *J. Nanosci Nanotechnol*, 2007, **7**, 2246.
51 U. Khoe, Y. L. Yang and S. Zhang, *Macromol. Biosci.*, 2008, **8**, 1060.
52 U. Khoe, Y. L. Yang and S. Zhang, *Langmuir*, 2009, **25**, 4111.
53 C. Chen, S. Zhang, M. Cao, J. Wang, H. Xu, X. Zhao and J. R. Lu, *Biomacromolecules*, 2010, **11**, 402.
54 X. Zhao, F. Pan and J. R. Lu, *Progress in Natural Science*, 2008, **18**, 653.
55 Z. Luo, B. Åkerman, S. Zhang and B. Nordén, *Soft Matter*, 2010, **6**, 2260.
56 S. Bucak, C. Cenker, I. Nasir, U. Olsson and M. Zackrisson, *Langmuir*, 2009, **25**, 4262.
57 V. Castelletto, D. R. Nutt, I. W. Hamley, S. Bucak, Ç. Cenker and U. Olsson, *Chemical Communications*, 2010, **46**, 6270.
58 C. A. E. Hauser, R. Deng, A. Mishra, Y. Loo, U. Khoe, F. Zhuang, D. W. Cheong, A. Accardo, M. B. Sullivan, C. Riekel, J. Y. Ying and U. A. Hauser, *Proc Natl Acad Sci USA*, 2011, **108**, 1361.
59 A. Mishra, Y. Loo, R. Deng, Y. J. Chuah, H. T. Hee, J. Y. Ying and C. A. E. Hauser, *NanoToday*, 2011, **6**, 232.
60 F. Zhuang, K. Oglęcka and C. A. E. Hauser, *Membranes*, 2011, **1**, 314.
61 A. Lakshmanan and C. A. Hauser, *Int J Mol Sci*, 2011, **12**, 5736.
62 J. Yeh, S. Du, A. Tordajada, J. Paulo and S. Zhang, *Biochemistry*, 2005, **44**, 16912–16919.
63 P. Kiley, X. Zhao, M. Vaughn, M. A. Baldo, B. D. Bruce and S. Zhang, *PLoS Biol*, 2005, **3**, 1180.
64 K. Matsumoto, M. Vaughn, B. D. Bruce, S. Koutsopoulos and S. Zhang, *J Phys Chem B*, 2009, **113**, 75.
65 X. Zhao, Y. Nagai, P. J. Reeves, P. Kiley, H. G. Khorana and S. Zhang, *Proc Natl Acad Sci USA*, 2006, **103**, 17707.
66 X. Wang, K. Corin, P. Baaske, C. J. Wienken, M. Jerabek-Willemsen, D. B. Stefan Duhrc and S. Zhang, *Proc Natl Acad Sci USA*, 2011, **108**, 9049.
67 K. Corin, P. Baaske, D. B. Ravel, J. Song, E. Brown, X. Wang, C. J. Wienken, M. Jerabek-Willemsen, S. Duhr, Y. Luo, D. Braun and S. Zhang, *PLoS One*, 2011, **6**, e25067.
68 S. Koutsopoulos, L. Kaiser, H. M. Eriksson and S. Zhang, Designer peptide surfactants stabilize diverse functional membrane proteins, *Chemical Society Reviews*, 2011, **41**, 1721.
69 S. Zhang, *Nature Biotechnology*, 2004, **22**, 151.
70 S. Zhang, T. Holmes, M. DiPersio, R. O. Hynes, X. Su and A. Rich, *Biomaterials*, 1995, **16**, 1385.
71 T. Holmes, S. Delacalle, X. Su, A. Rich and S. Zhang, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 6728.
72 J. Kisiday, M. Jin and B. Kurz, *et al.*, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 9996.

73 A. Schneider, J. A. Garlick and C. Egles, *PLoS ONE*, 2008, **3**, e1410.
74 S. Y. Fung, H. Yang and P. Chen, *PLoS ONE*, 2008, **3**, e1956.
75 D. Marini, W. Hwang, D. Lauffenburger, S. Zhang and R. D. Kamm, *Nano Lett*., 2002, **2**, 295.
76 M. A. Bokhari, G. Akay, S. Zhang and M. A. Birch, *Biomaterials*, 2005, **26**, 5198.
77 D. A. Narmoneva, O. Oni, A. L. Sieminski, S. Zhang, J. P. Gertler, R. D. Kamm and R. T. Lee., *Biomaterials*, 2005, **26**, 4837.
78 M. E. Davis, J. P. M. Motion, D. A. Narmoneva, T. Takahashi, D. Hakuno, R. D. Kamm, S. Zhang and R. T. Lee, *Circulation*, 2005, **111**, 442.
79 M. E. Davis, P. C. H. Hsieh, T. Takahashi, Q. Song, S. Zhang, R. D. Kamm, A. J. Grodzinsky, P. Anversa and R. T. Lee, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 8155.
80 Z. Luo and S. Zhang, *Chemical Society Reviews* 2012, **42**, DOI:10.1039/C2CS15360B.

Metal complexes of amino acids and peptides

Etelka Farkas* and Imre Sóvágó
DOI: 10.1039/9781849734677-00066

1 Introduction

This chapter deals with the most important results and observations published on various aspects of the metal complex formation with amino acids, peptides and related ligands during the past two-three years. The major sources of the references collected here are the Abstracts reported by the Web of Science Databases on the Internet but the title pages of the most common journals of inorganic, bioinorganic and coordination chemistry have also been surveyed.

The previous volume (Vol. 36) of this series covered the papers published in 2003 and 2004 and the five year gap between the preparation of the last and recent reviews resulted in a tremendous change in the number and especially in the major subjects of the related publications. In the previous volumes, the traditional structural, kinetic and thermodynamic characterization of the complex formation processes of these ligands were in the focus of the studies and the contents and structure of the Chapters were in correspondence with it. The basic information on the coordination chemistry of all amino acids and the most common peptides are now available for almost all important metal ions and the number of publications on the simple kinetic and structural characterization of these complexes is significantly reduced. On the other hand, the number of publications increased in the fields of analytical, industrial, synthetic and biomedical applications of the complexes of amino acids and peptides. The understanding of these applications requires the exact structural and thermodynamic characterization of the complexes too, which is reflected in the comprehensive nature of the most valuable publications. Taking into account these changes in the major directions of the corresponding research the sub-headings of the Chapter were changed accordingly. The metal complexes of amino acids (Section 2) and peptides (Section 3) are treated separately, although there are many publications dealing with the coordination chemistry of both types of ligands.

The basic coordination chemistry of amino acids and related ligands have already been reviewed. Nowadays, a decisive majority of the papers connected to the subject of metal ion - amino acid complexes summarize results of bioinduced works. Developement of new metal complex based anticancer agents with high selectivity and low toxicity is a hot area and initiates numerous investigations. This topic is very diverse, involves synthesis and characterization of new compounds, the development and tests of new synthetic routes and new analytical methods, as well as investigation of

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

biological effects. In this subject the results obtained for the interaction of lanthanide-amino acid complexes with target nucleic acids have been reviewed.[1] In another review the current results for the potential antitumor agent palladium(II)-amino acid complexes have been summarized.[2] As it is well-known, there are numerous potential metalloenzyme inhibitors among the amino acids/derivatives and a review about the aminopeptidase inhibitors has summarized the current achievements.[3] Non-covalent interactions play crucial roles in biological systems and are extensively studied. The results obtained for metal ion-nucleotides (or nucleosides) -amino acid (amine) ternary systems as models for evaluation of factors affecting the formation and stability of stacks have been collected and analyzed.[4] A short review discusses the possible role of Cys and iron in the progression of Parkinson's desease.[5] Metallosupramolecular compounds (with involvement of amino acids or derivatives) have been fabricated with different purposes and for various applications as can be seen in some of the most recent reviews.[6–9] Reviews to show the development of fluorescent and colorimetric chemosensors for detection of amino acids,[10] and also a new technique, infrared consequence spectroscopy to investigate protonated and metal ion cationized complexes of different ligands, among others amino acids in gas phase have been published, too.[11]

Manganese-amino acid/derivative complexes as models for Photosystem II have been studied for a long period and the results have been summarized in numerous reviews previously. The most recent summarization focuses on the results obtained by computational chemistry, Density Functional Theory computational techniques to study the structure and mechanism of the manganese catalytic site in Photosystem II.[12]

In the case of peptides the significant development in peptide synthesis and the widespread applications of the various automatic, solid phase peptide synthesizers made it possible to synthesize relatively large peptide molecules in such quantities which are suitable even for the use of the classical experimental techniques of coordination chemistry. The structure of metalloproteins revealed that two amino acids, histidine and cysteine, are much more common in the active centres of these molecules than the other ones. As a consequence, peptides containing histidyl and/or cysteinyl residues in different number and location in the sequence have been prepared in high number and their complex formation was studied with a series of metallic elements having biological significance. In accordance, the subsections are dealing with (i) the coordination chemistry of small model peptides, (ii) metal complexes and biological role of peptides containing histidyl residues (iii) metal complexes and biological role of peptides containing cysteinyl residues and (iv) their applications in analytical, industrial and medicinal chemistry.

The readers of metallopeptide chemistry are in a good position becuase a large number of different reviews has been published in the last few years. Some of them are devoted to specific aspects of this subject, *e.g.* copper(II) chelators based upon tetrazole derivatives of peptides,[13] the role of cysteinyl and histidyl peptides in nickel(II) homeostasis,[14] the formation of peptide bonds in the neighborhood of metal ions[15] or the increasing application of mass spectrometry in this field.[16] Most of these rewiews were, however,

promoted by the research on the possible role of metal ions in neurodegeneration[17–23] including a book on the role of metallostasis in the neurological processes.[24]

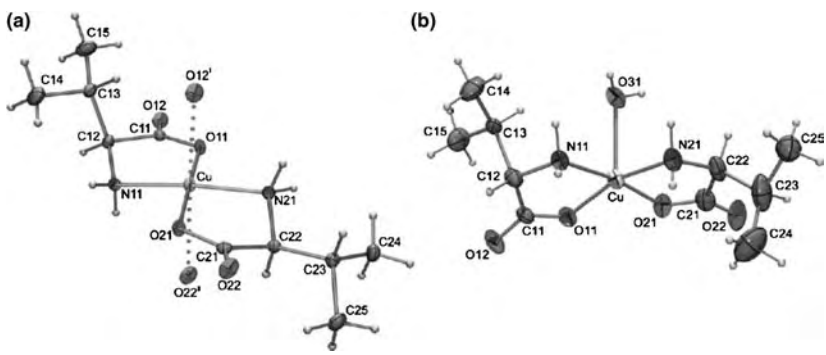
2 Amino acid complexes

2.1 Synthesis and characterization of metal ion-amino acid complexes

Binary complexes. Although many questions related *e.g.* to the stability, bonding mode and structure of metal ion-amino acid binary complexes were clarified prior to this reviewed period, still numerous papers have been published in this subject during the past *ca.* two years. Some new complexes have been prepared and characterized,[25–27] solution studies have also been performed for determination of stoichiometry, thermodynamic and sometimes also redox stability of the complexes,[38–49] or distribution of metal ions among amino acids have been determined.[50] Several results dealing with various kinetic aspects of metal ion assisted oxidation of amino acids have been published, too.[49,51–54] Interesting new results have been provided by the numerous theoretical calculations (made in some cases together, in other cases not with experimental studies),[37,55–64] and the number of papers appeared with results for the metal complexes of amino acids in gas-phase is astonishingly high.[61–72]

The X-ray crystal and molecular structures of trans bis(L-valinato)-copper(II) and *cis* aquabis(L-valinato)copper(II) have been determined in a recent work. Molecular modelling calculations were attempted to resolve factors that influenced the isomerisation and crystallization of either the aqua *cis*- or the anhydrous *trans*-isomer. The conformers with *trans*-configuration were found to be the most stable in vacuum, but those with *cis*-configuration formed more favourable intermolecular interactions. Consequently, both *cis*- and *trans*-isomers were predicted to exist in aqueous solution. According to the crystal structure simulations and predictions, *cis*-isomer (Scheme 1b) was found to require water molecules to form energetically more stable crystal packing than *trans*-isomer (Scheme 1a).[25]

CO as one of the fundamental signalling molecules is known to have numerous toxical, but also a range of beneficial effects. To develop an



Scheme 1 (Reprinted with permission from M. Markovic, N. Judas and J. Sabolovic, *Inorg. Chem.*, 2011, **50**, 3632. Copyright © 2011 by American Chemical Society.)

efficient form of CO-therapy, in which issues such as dosage and delivery may be carefully controlled a range of CO-releasing molecules have been prepared previously. In a recent work, new complexes *via* incorporation of amino acids and amino esters into the coordination sphere of Group 6 metal (Cr, Mo, W) carbonyl complexes have been synthesized and their structures have been characterized by X-ray diffraction. The incorporation of biologically compatible leaving groups such as amino acids or amino esters into the coordination sphere was found to offer a considerable potential for tuning and modulating the rates of the CO-release, which have occurred more readily when sterically demanding amino esters have been involved.[30]

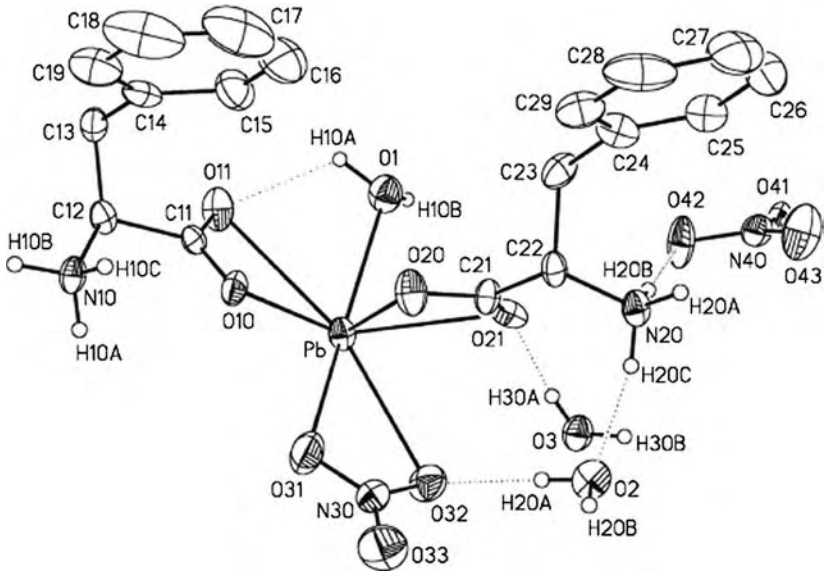
Investigation of the interrelation between the structure, stability, and reactivity of transition metal complexes formed with amino acids and oligopeptides is vitally important for modelling metalloenzymes. As imidazole-N and thiol-S donor atoms are frequently present in the catalytic centres of metalloenzymes, studying the interaction of metal ions with His and Cys and their derivatives takes on particular significance and several papers have been published recently in this subject. A combination of pH-potentiometry, spectrophotometry, EPR, and NMR relaxation methods as well as quantum chemical computations were used to characterize the complex formation, chemical exchange, structure of the complexes formed in the copper(II)-L/DL-histidine systems. Hydrogen bonding between carboxyl and imidazolyl groups through the water molecule was resulted in considerable stereoselectivity of the mono-protonated complex and preferential formation of the *cis*- $[Cu(L\text{-}His)(L\text{-}HisH)]^+$ isomer. At the same time, small stereoselectivity in the $[Cu(His)_2]$ formation with preferential accumulation of the *meso*-form was resulted from favourable axial coordination of the imidazolyl group in the *cis*-[Cu(L-His)(D-His)] isomer.[38] Not only the copper(II), but also manganese(II), cobalt(II), nickel(II) and zinc(II) complexes of L-His have been investigated by extended X-ray absorption fine structure (EXAFS).[26] In another work, dinuclear copper(II)-zinc(II) complexes as models for superoxide dismutase (SOD) enzyme were tried to be developed by using His or Gly as chelating agent. However, the formation of mixed metal complex was found only with His, but not with Gly.[28] There are His residues also in the classic binding motif of the zinc finger, where, in most cases, two His and two Cys residues are directly coordinated to the zinc(II) ion. However, the zinc finger is a sensitive target for toxic heavy metals like cadmium(II) and the replacement of the zinc ion by another metal ion results in the inhibition of the functionality of the zinc-dependent complex. To get new information in this subject infrared multiple photon dissociation (IRMPD) spectroscopy was used to study the gas-phase structures of the zinc(II) and cadmium(II) complexes and the role of these metal ions in the conformational dependence of His was evaluated.[61] Again, IRMPD spectroscopic measurements and theoretical calculations were made for studying the effects of halide anions on the zwitterion stability of His (as well as Glu and Arg) and a comparison of these effects to those realised in the presence of cesium(I) was also made.[62] Theoretical study has been performed to determine the effect of various monovalent (lithium(I), sodium(I), potassium(I)) and bivalent (magnesium(II), calcium(II),

nickel(II), copper(II), zinc(II)) metal ions as well as water coordination on the structure and complex forming properties of the neutral and zwitterionic forms of L-His.[63]

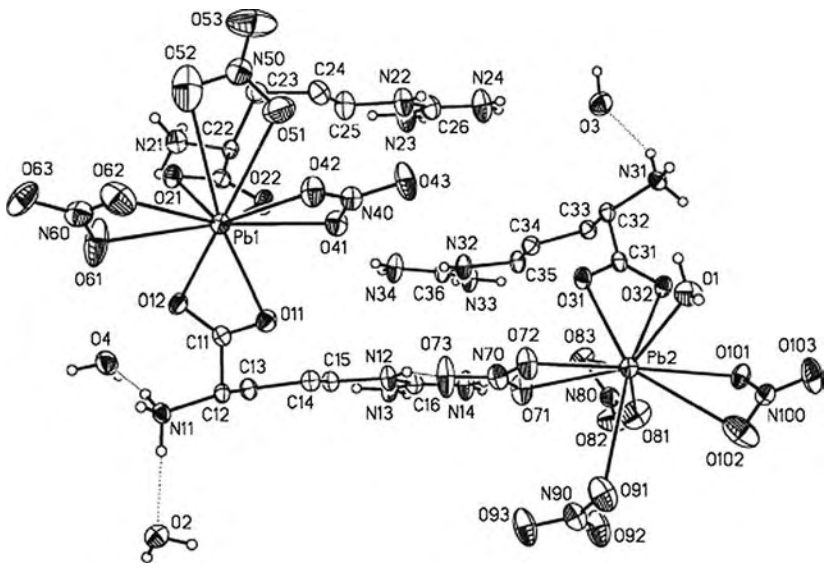
As the thiol group can readily coordinate with many metal ions in biological systems, the complexation of various metal ions with Cys and its derivatives, Met, N-acetyl-cysteine and cystine, has been of great interest since decades. This fact is one of the reasons why numerous papers published during the past few years are discussing results for characterization of metal complexes with thiol-containing amino acids.[29,32–34,41,52,55–57,64,79,81] For example, a new gold(I) complex with N-acetyl-cysteine was synthesized and characterized by different techniques and antibacterial activity of the complex against Staphylococcus aureus and Escherinchia coli bacterial cells was also tested.[29] In another work, a density functional theory based method together with a continuum solvation method was used to study the complexation of Cys with zinc(II), magnesium(II), calcium(II), iron(III) and manganese(II), and metal ion dependent coordination mode was found. Bidentate (N,S) type coordination was preferred by zinc(II) and iron(III), while sulfur-free bonding mode was predominated by magnesium(II) and calcium(II). With manganese(II), the existence of several equally stable bonding isomers was identified.[55] Significant alteration by metal ion substitution was not found in the tridentate coordination mode, when density functional theory and second order Moller-Plesset Perturbation theory methods were used to study the coordination geometries, electronic features, entropies and energetics of complexes formed between lithium(I), sodium(I), potassium(I), berillium(II), magnesium(II) or calcium(II) and Cys.[56] On the other hand, according to some results obtained recently by IRMPD spectroscopy, the size of the alkaline metal cation has significant effects on gas-phase conformation of Cys. The tridentate coordination mode, in which the amine and carbonyl groups of the amino acid backbone and the sulfur atom of the side chain predominates in the lithium(I) and sodium(I) complexes, but with potassium(I), rubidium(I) and cesium(I), in addition to the above mentioned tridentate mode, also a bidentate way *via* both oxygens of the carboxylic acid group occurs.[64]

Although heavy metal ions are expected to interact with various bioligands and the interaction might lead to dramatic physiological consequences, limited number of results is available for amino acid – toxic metal ion systems. However, some papers display new results for complexes of amino acids with toxic heavy metals like lead(II),[31,36,40,50,65] cadmium(II),[32,33,40,41,46,50,57,61] mercury(II),[43,57,58] aluminium(III).[59] X-Ray characterization of lead(II) complexes formed with L-Phe, L-Ile, L-Val or L-Arg was performed. According to the results, the solid state structures of the binary Pb^{2+}- L-Phe, Pb^{2+}- L-Ile, Pb^{2+}- L-Val and the ternary Pb^{2+}- L-Val-L-*i*Leu complexes show a lead center coordinated by two amino acid ligands, while the Pb^{2+}-L-Arg complex is a cluster involving two lead centers and three Arg molecules. As representative examples, the crystallographic views of the asymmetric units of $[Pb(OH_2)_2(Phe)_2(NO_3)]$ (NO_3) and $[Pb_2(OH_2)(HArg)_3(NO_3)_7]3H_2O$ are shown on Schemes 2 and 3 respectively.[31]

Lead(II), cadmium(II) and zinc(II) metal ions and Gly, Ala and Val ligands have been involved into the study, in which the molar heat capacities



Scheme 2 (Reprinted with permission from C.D.L Saunders, L.E. Longobardi, N. Burford, M.D. Lumsden, U. Werner-Zwanziger, B. Chen and R. McDonald, *Inorg. Chem.*, 2011, **50**, 2799. Copyright © 2011 by American Chemical Society.)



Scheme 3 (Reprinted with permission from C.D.L. Saunders, L.E. Longobardi, N. Burford, M.D. Lumsden, U. Werner-Zwanziger, B. Chen and R. McDonald, *Inorg. Chem.*, 2011, **50**, 2799. Copyright © 2011 by American Chemical Society.)

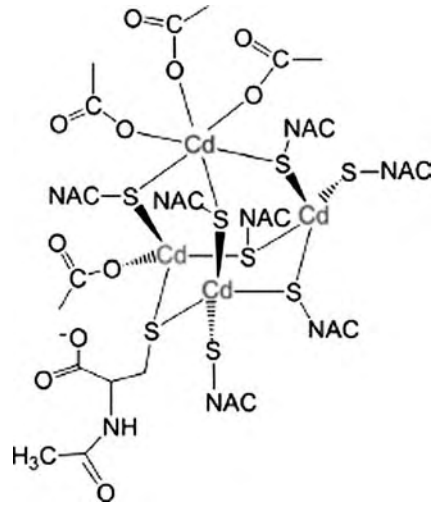
and standard molar formation enthalpies for the bis-complexes have been determined.[41] In another work, the high affinity of mercury(II) and cadmium(II) ions to bind to thiolate group initiated the quantum chemical

study on complexation of neutral and deprotonated Cys with the above mentioned two toxic metals and zinc(II).[57] As it is known, methylmercury (MeHg) biomagnifies through the food-chain and is a neurotoxin to biological organisms including humans. The toxic effect is in direct correlation with the high affinity of MeHg to bind "soft" Lewis bases including sulfur-containing amino acids and residues. Because, sulfur and selenium belong to the same group in the periodical table and show chemical similarity, selenium can substitute sulfur in the biological molecules and the selenium analogues can be obtained. In a recent work, synthesis and characterization of four new MeHg-selenoamino acid complexes and quantum chemical calculations on their structural, electronic, spectroscopic, and thermodynamic properties have been performed. In the same work the results obtained with the seleno derivatives were compared to the results for the corresponding complexes of the sulfur containing molecules (cysteine, penicillamine, glutathione and methionine). This comparison showed the very similar structural properties, but the different electronic and thermodynamic properties. The theoretical backgrounds of similarities and differences, as well as some biological consequences have been also evaluated in the paper.[58]

The complex formation between cadmium(II) ion and sulfur-containing amino acids or derivatives have been investigated in several laboratories.[32,33,41] The interaction of this metal ion with *N*-acetylcysteine (H_2NAC) in aqueous solution was investigated by Cd K- and L_3-edge X-ray absorption and ^{113}Cd NMR spectroscopic techniques. Different conditions were necessary to use at pH = 11 and pH = 7.5 to achieve the predominant existence of the cadmium-tetrathiolate complex. In alkaline solution, the ^{113}Cd NMR spectrum supported the existence of an oligomeric Cd(II)-NAC species with single thiolate bridges between the cadmium ions and with presence of CdS_3O_3, CdS_3O and CdS_4 coordination sites (Scheme 4).[41]

IRMPD spectroscopy was used to determine the gas-phase structure of lead(II)-amino acid (Ala, Val, Leu, Ile, Pro and Lys) complexes with and without a solvent molecule present.[65] The fluorescence spectroscopic study, in which the fluorescence of three aromatic amino acids, Trp, Tyr and Phe could be quenched by Pd(II), resulted not only in determination of the coordination modes, but also in development of a method for determination of this metal.[36] The same three aromatic amino acids have been involved in another study, in which their complexation with dimethyltin(IV) in aqueous solution was investigated.[42] Thermodynamics of mixed ligand complexation of mercury(II) with EDTA and His or Lys have been also investigated.[43] Numerous previous speculations have been already made for various possible toxicity mechanisms of Al(III). In connection with this problem, in a recent work, density functional theory has been applied and the obtained results supported the real possibility of the spontaneous substitution of magnesium(II) by aluminium(III) in the magnesium(II)-Asp (Hasp) complexes.[59]

Gas-phase studies (using either adequate experimental method or theoretical calculations) can characterize inherent amino acid interactions with metal cations in the absence of solution effects. The results, which can



Scheme 4

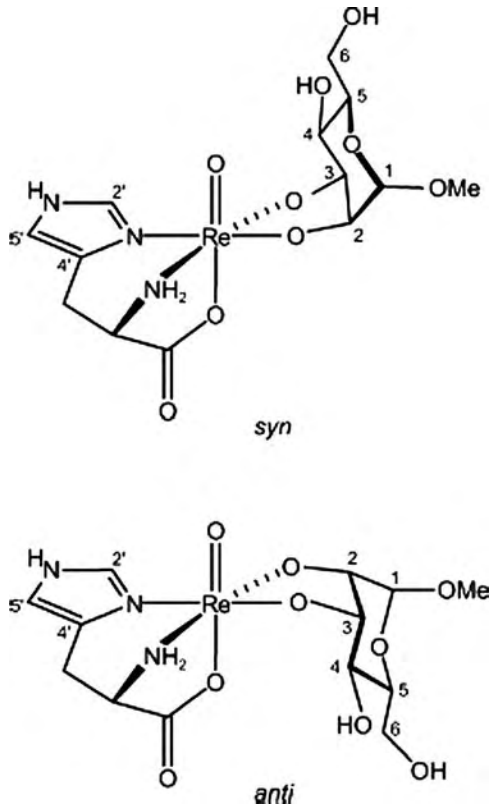
be collected by the gas-phase study, provide useful information, *e.g.* for interpreting the fragmentation mechanisms in mass spectrometric measurements. In addition to the already mentioned papers,[61–65] several others have been published with results of gas-phase studies during the *ca.* past two years.[66–72] For example, IRMPD (infrared multiple-photon spectroscopy) was used to characterize the dimeric complexes formed in the interaction of Trp with alkaline earth cations.[66] The ever increasing amount of ionizing radiation to which people are exposed to in the modern society is believed to be a major cause of damage to living cells and investigation of the mechanisms of damage and repair of biologically essential molecules, *e.g.* amino acids might provide useful information for this problem. This idea has initiated investigation of electron induced fragmentation of the complexes formed between cesium(I), sodium(I), potassium(I) or silver(I) and Trp.[67] Reaction mechanisms for the fragmentation processes, collisionally activated complexes of alkali metal cations with Pro and *N*-methyl alanine,[68] or with Cys,[69] and also lithium(I) complex with Cys,[69] rubidium(I) with Gly, Ser, Thr, Pro[71] have been also investigated. In a theoretical study, the bonding isomer, in which the zwitterionic Gly is coordinated to the manganese(II) *via* its two carboxylate oxygens was found to present the most stable mode of binding.[72]

Ternary complexes. Investigations on ternary complexes with amino acids have been in the focus of interest since several decades and a large number of papers has come out in this subject anually. During the past *ca.* two years, numerous bioinspired investigations on metal ion – amino acid – ''second ligand'' ternary complexes have been performed and the obtained results have been summarized in many publications.[43,73–106] Several systems have been studied to understand modified pharmacological and toxicological properties of drugs, when those are administered in the form of metal complexes (metallodrugs) and the ternary complexes served as models for gaining new information about the mechanism of the action. In other

cases, the authors have aimed to collect novel details about the interactions of metal ions with various biomolecules. In such types of works, in addition to the chemical characterization of the complexes, various biological tests have been also performed.[73–80] For instance, following the detailed characterization (X-ray, IR, FIR, NIR-Vis-UV, HF EPR, magnetic susceptibility) of the synthesized $[Ni(imidazole)_2(L\text{-}Tyr)_2]\cdot 4H_2O$ complex, investigation on its antibacterial and antifungal activity was also made. The results showed octahedral geometry with very small distortion, but this small distortion was sufficient to affect the electronic d-d bands. Hydrogen-bond chain was found to exist in the structure, but analysis of the magnetic properties confirmed that the hydrogen-bond chain did not provide a magnetic exchange pathway because of the rather long intermolecular Ni-Ni separation. Compared to the Tyr-containing binary complex, more significant antibacterial, antifungal effects of this new ternary complex have been detected against some microorganisms.[78]

As ''second ligands'' various biologically important compounds, such as DNA, DNA purine bases, ATP, pyrophosphates, amines have been frequently used.[81–90] As chirality is known as a key factor determining the pharmacological behaviour of metal complexes, works are frequently directed to study this problem. Cobalt(II), copper(II) and zinc(II) complexes formed with 1,2-diaminobenzene and L-, D- or L,D-Trp have been synthesized and characterized. The complex formed with the L-Trp was found to show the highest ability for binding DNA *via* non-covalent interactions.[83] Chiral recognition of α-amino acids by an optically active cobalt(II)-cyclen complex was observed in another work, in which the preference towards D-Phenylglycine and D-Ile compared to their counterparts was detected. In the postulated mechanism, the optically active cyclen skeleton allows chiral recognition of amino acids through a twist of amino acid backbone that is caused by the repulsion between the bulky amino acid residue and cyclen.[87] Formation of ternary complexes is involved in the ''kinetic and fixed ligand methods'', that was successfully used for chiral differentiation, discrimination of some cyclic-β-amino acids and β-3-homo amino acids.[91,92]

Reactions of cisplatin and its hydrated products with the amino acids Cys and Met play an important role in the metabolism of the widely used antineoplastic drug cisplatin, therefore computational study has been performed to clarify the interaction of *cis*-$[Pt(NH_3)_2Cl(H_2O)]^+$ and *cis*-$[Pt(NH_3)_2(OH)(H_2O)]^+$ with Cys or Met in a wide pH-range.[83,84] Currently, significant cytotoxicity of ruthenium(II) complexes with azpy (azpy = 2-(phenylazo)pyridine) against a series of tumor-cell lines was found. The binding of hydrolyzed α-$[Ru(azpy)_2Cl_2]$ (α refers to the isomer in which the coordinating pairs Cl, N(py), N(azo) are *cis*, *trans* and *cis*, respectively) to DNA purine bases and amino acid residues (guanine, adenine, methionine and histidine) was investigated by using density functional theory in a recent work. The results demonstrated that the ruthenium(II) containing hydrolyzed complex preferentially bound to the amino acid residues.[95] Complexes of oxidorhenium(V) with L-His and pyranosides (methyl α-D-mannopyranoside and methyl β-D-ribopyranoside) have been synthesized and characterized by X-ray diffraction, NMR spectroscopy,



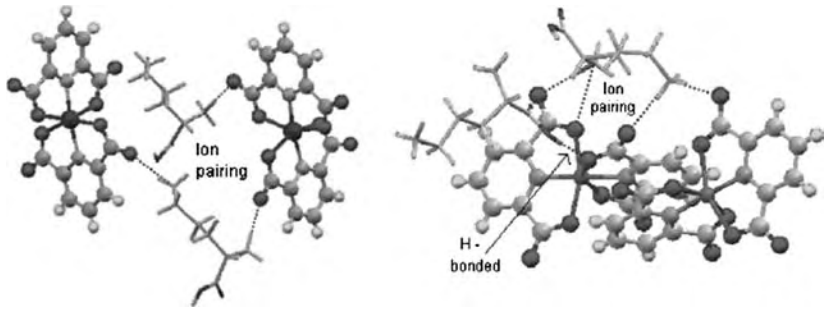
Scheme 5

elemental analysis and mass spectrometry. NMR study, made following the dissolution of the crystallised ternary complex formed with methyl α-D-mannopyranoside, provided evidence for the existence of two isomers in solution (Scheme 5). The *syn* isomer that formed the crystals was the tentative major species also in solution.[96]

Two unusual ternary dinuclear vanadium(V)-tetraperoxo-glycine complexes were obtained *via* the reaction of V_2O_5, Gly and H_2O_2. The complexes were fully characterized and are thought to provide some useful information to clarify the chemical reactivity of V(V)-diperoxo species in key cellular events.[97]

Interestingly, when L-His and L-Orn in their fully protonated (dicationic) forms have been reacted with bis-copper(II) dipicolinato and bis-cobalt(II)-dipicolinato dianions and the new complexes have been synthesized, the amino acids have been found not to get inside the coordination sphere, but have been strongly hydrogen bonded and intercalated in the layered structures of the dipicolinato complex anions. 1:1 cation to anion ratio was shown by the X-ray result with L-His, bit it was 2:2 with L-Orn. (see Scheme 6 for the assembly of L-Orn dication with bis-cobalt(II)-dipicolinato (left) and bis-copper(II) dipicolinato (right) anions).[99]

Similarly, ternary complexes were not formed, but substitution reaction happened, when the reactivity of ruthenium(II)-arene complexes and



Scheme 6 (Reprinted from Polyhedron, 30/1, B. Das, J.B. Baruah, Dications ofl-histidine andl-ornithine in layers of copper(II) and cobalt(II) dipicolinato anions, 22–26, Copyright (2011), with permission from Elsevier.)

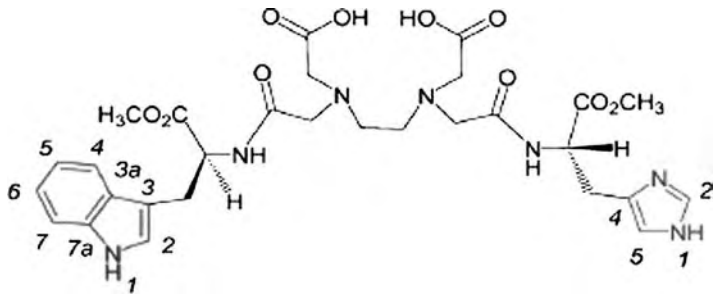
osmium(II)-arene complexes of 3-hydroxy-2(1*H*)-pyridones towards various biomolecules has been studied in a recent work. In this work, to get answer for the question whether the reactivity of these complexes toward biomolecules is the determining parameter of their anticancer activity, reactions with the involvement of different amino acids have been performed and clear evidence for the displacement of the pyridone ligands from the metal centers by amino acids has been found.[27]

2.2 Synthesis and structural studies on metal complexes of amino acid derivatives

Modification of amino acids can significantly alter their metal binding ability or selectivity towards various metal ions. This fact is one of the reasons why, numerous different derivatives of amino acids have been prepared during the reviewed period and, in many cases, their metal binding ability have been investigated. It is evident that the number of possibilities to create new amino acid derivatives is ''endless'' and the aims of studies can be very different, but quite frequently, the biochemical and pharmacological properties of the newly synthesized derivatives and their complexes initiate works in this subject. Also, metal complexes of amino acid derivatives are often prepared as structural or functional models for biomolecules.

Amino acid esters,[30,107] amides,[108] aminohydroxamic acids,[109,110] oximes,[111,112] phosphinic/phosphonic derivatives,[113–119] amines[120] are among the permanently investigated molecules and some new results *e.g* for thermodynamic and redox stability, bonding mode and biological activity of their metal complexes have been published within the period covered in this chapter. Some new functionalysed amino acids such as N-phthaloyl[121] and N-p-tolylsulfonyl[122] or 4-toluenesulfonyl[123] derivatives have been also synthesized and their metal complexes with selected metal ions have been characterized.

Trp/His side chains around metal sites are crucial to diverse metalloprotein function. To get answer for the question that how aromatic rings around metal-sites have unique capabilities through the control of the spatiotemporal distribution of noncovalent interaction elements to achieve

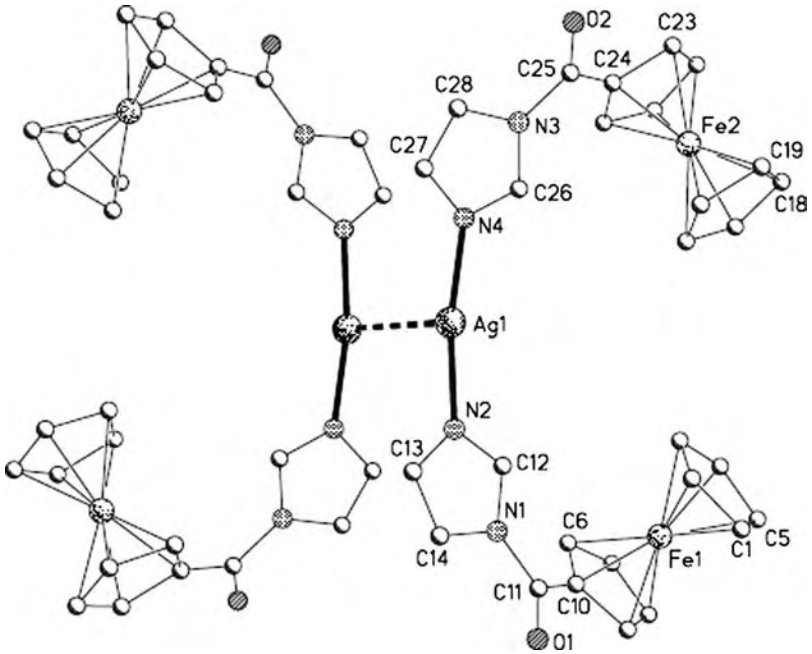


Scheme 7

diverse chemical functionality, hetero-bifunctionally modified EDTA-bisamide and DTPA-bisamide ligands featuring Trp and His side chains have been prepared and their complexes with calcium(II), cobalt(II), copper(II), nickel(II), manganese(II), zinc(II), cadmium(II) and iron(II) have been studied. As an example, the formula of the EDTA derivative is presented in Scheme 7.[124]

For the development of potential manganese-based MRI probes, three new chelators based on the 6-amino-6-methylperhydro-1,4-diazepine scaffold and possessing three pendant *N*-acetic or *N*-α-methylacetic groups have been synthesized. Although thermodynamic and relaxometric study with these new hexadentate ligands showed the lower stability of these manganese(II) complexes compared to the stability of the corresponding complex with the parent molecule, but interestingly, in solution, existence of a mixture of two isomeric species differing in coordination number (7 and 6) and in the number (1 and 0) of bound water molecules was found with both ligands.[125] C-branched new macrocyclic amino acids have been synthesized and the structure of their nickel(II), zinc(II) and copper(II) complexes have been determined. The X-ray diffraction results demonstrated formation of diastereometrically pure complexes with different geometries depending on the metal ion.[126] The unique properties of ferrocene have initiated the insertion of this building block into various compounds including amino acids. Recently, click reaction has been employed to synthesize a wide range of ferrocene amino acids.[117] In another work, silver(I) and copper(I) complexes of five ferrocenyl containing ligands including ferrocenyl imidazole have been investigated. When the counter anions were triflate, the crystal structure showed a dimerization of the complex, with the ferrocenyl moieties occupying *cis* positions, by means of a weak Ag - - Ag interaction as this is shown by the molecular diagram related to the overall structure of the dimer (Scheme 8).[128]

Valuable results have been obtained when chiral phosphinoferrocene carboxamides wearing amino acid pendants as efficient ligands have been tested in palladium-catalysed asymmetric allylic alkylation reaction,[129] and also, when redox-responsive oligoferrocene peptid foldamers from Fmoc-protected 1-amino-1′-ferrocene carboxylic acid (Fmoc = 9-fluorenylmethyloxycarbonyl) *via* solid-phase synthesis have been prepared and characterized both in solution and solid states.[130]



Scheme 8 (Reprinted from Journal of Organometallic Chemistry, 695/4, S.Quintal, M. C. Gimeno, A. Laguna, M. J. Calhorda, Silver(I) and copper(I) complexes with ferrocenyl ligands bearing imidazole or pyridyl substituents, 558–566, Copyright (2011), with permission from Elsevier.)

Covalent linkers can be established, for example, *via* side-chain functionalization of amino acids. The imidazole-containing His is remarkable in this respect. Alkylation of histidine side chain and the subsequent metallation of the histidinium salt constitute an attractive approach to bioorganometallic chemistry. In a recent work, *C,N*-protected His has been successfully alkylated at both side-chain nitrogens. Following this, the corresponding histidinium salts have been metallated with ruthenium(II). The obtained complexes showed appreciable activity in the catalytic transfer hydrogenation of ketones.[131] Ruthenium amino acids have been used in a few laboratories to synthesize new bis(terpyridine)ruthenium(II) complexes with different purposes.[132,133] For instance, investigation has been performed to clarify the possibility of the metal-metal communication across an amide linkage in a dinuclear bis(terpyridine)ruthenium complex.[132]

Captopril is a pseudo-cysteinyl-proline and, because it is a potent and specific inhibitor of the zinc-containing angiotensin converting enzyme (ACE), can be used as an antihypertensive agent. However, side-effects can arise during the captopril treatment, what can be caused by the interaction of this molecule with other metal ions present in the plasma. This problem has initiated the investigation on complexation of captopril with copper(II), nickel(II), iron(III) and tin(II).[134]

The biological roles of the natural amino acid derivative β-citryl-L-glutamate, which was isolated at the first time from the brains of newborn rats are unclear. In a recent work, the main goal was the evaluation of the

interaction of this bioligand with some metal ions like iron(II), iron(III), copper(II), zinc(II), calcium(II), magnesium(II), when solution equilibrium work on these systems have been performed.[135]

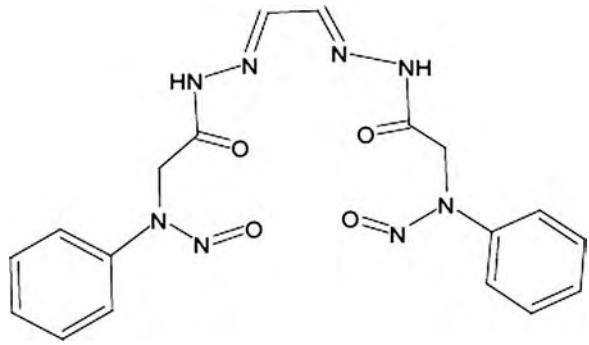
Diversity and structural variability make Schiff bases very attractive compounds. Amino acid Schiff bases are an important class of metal binding ligands and their complexes have effective applications in different fields. This fact is one of the reasons why numerous investigations in this subject have been made during the past few years.[136–154] To develop new cleavage agents (potential metallodrugs) for DNA, ternay copper(II) complexes containing Trp-based or Tyr-based Schiff base as ''ligand A'' and 1,10-phenantroline or 2,2′-bipyridine as ''ligand B'' have been synthesized and characterized by various techniques. Due to the presence of the more extended aromatic moiety in 1,10-phenantroline, the two ternary complexes formed with this ligand showed the higher DNA binding and cleavage activity.[136,137] 1,10-Phenantroline was also the ''ligand B'' in another work, in which copper(II) and zinc(II) complexes of *N*-salicylidene-Ala and *N*-salicylidene-His have been prepared and tested against some pathogens.[128] *N*-salicylidene-Arg, hydroxynaphthylidene-Arg and *N*-salicylidene-Lys were used to synthesize new iron(III) complexes as photocleavers of DNA in visible light. The synthesized complexes exhibited photo-induced DNA cleavage activity in UV-A light and visible light *via* a pathway that involves the formation of reactive hydroxyl radical species.[139] The high importance of the organotin(IV) complexes of Schiff bases is clearly demonstrated by the recently published perspective type paper,[140] but some new results also for tin(II) complexes of amino acid based fluorinated Schiff bases have come out.[141] As it is well-known, metal complexes of amino acid based Schiff bases are frequently employed as models for more complicated biological systems. In this subject, new dinuclear complexes containing Zn-O-Zn motif as metalloenzyme mimics and also some copper(II) analogues have been prepared and characterized. A template strategy, which led to the one-pot synthesis of the complexes, was used in this work.[142] Chiral tridentate amino acid Schiff bases have been used to synthesize effective zinc complexes as catalysts for the ring-opening polymerization of lactides at ambient temperature producing polymers with controlled and narrow molecular weight distribution. The complexes developed have been fully characterized.[143] In a recent experiment, two nitrosyl (NO) moieties have been inserted into an amino acid Schiff base and a fascinating new compound, glyoxal bis(*N*-nitroso phenylglycine has been developed (see Scheme 9 for the formula).[144]

Manganese(II), cobalt(II), nickel(II), copper(II) and zinc(II) complexes of this new hexadentate ligand have been prepared and characterized. Among the new species, the copper complex has shown the highest antimicrobal activity against all the microorganisms tested.[144]

Previously, little attention has been given to Schiff bases derived from β-amino acids. Quite recently some new results have been published also on this field.[145,146]

2.3 Application of metal ion-amino acid/derivative complexes

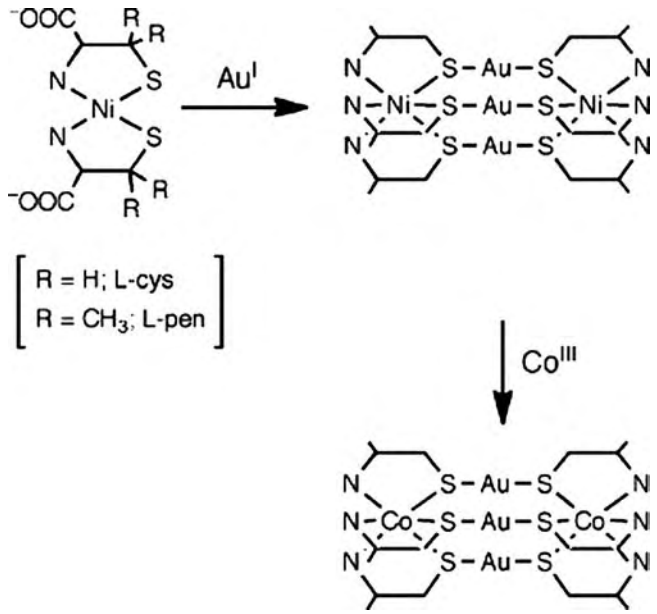
The interest to coordination polymers with well-regulated network structures is caused by their potential applications for creation of various



Scheme 9 (Reprinted from European Journal of Medicinal Chemistry, 45/7, M. Patil, R. Hunoor, K. Gudasi, Transition metal complexes of a new hexadentate macroacyclic N2O4-donor Schiff base: Inhibitory activity against bacteria and fungi, 2981–2986, Copyright (2010), with permission from Elsevier.)

functional materials. New, well-regulated networks have been developed with the aim of possibility of various potential applications in the fields of catalysis, non-linear optics, ion-exchange, luminescence, magnetism, and so on. However, a lot of factors leading to certain structure of transition metal complexes with α-aminocarboxylates or with their derivatives are still not clear and initiate a lot of investigations. As a consequence, a huge number of papers that came out during the period of coverage can be associated with these subjects.[122,155–176] For example, two new 1D coordination polymers have been isolated and characterized in a recent work. One was built from $[Cu(L\text{-}Pro)]^+$ units, and the second was made from trinuclear Cu_3 blocks $Cu_3(Gly)_4(H_2O)_2(NO_3)_2$. $Cu(L\text{-}Pro)(ClO_4)(H_2O)_2$ rotated polarized light and relative rotation of polarized light for the bands associated with d–d transitions was higher in solid state than in solution. This result along with changes of the positions of bands in CD spectra at dissolving of $Cu(L\text{-}Pro)(ClO_4)(H_2O)_2$ provides evidence for changes of the donor set around the metal(II) ion. In both compounds copper(II) ions were linked through carboxy-groups in syn-anti configuration, donor atoms of which occupied equatorial positions in coordination spheres of copper(II) ions. Magnetic measurements revealed ferromagnetic exchange interaction between copper(II) ions in the Pro-containing complex.[155] In another work, on heating cobalt(II), aspartate and N-donor ligands (4,4′-bipyridine, *trans*-bis(4-pyridyl)-ethylene and 1,2-bis(4-pyridyl)ethane) in an aqueous – methanol solvent mixture various new coordination polymers have been synthesized. The structures of the three new compounds have been determined by X-ray diffraction method.[156] Fascinating different motifs can be obtained, when mononuclear metal complexes with tiol-containing amino acids such as D-penicillamine and L-cysteine are used as chiral metalloligands toward various metal ions with different coordination geometries. As an example, Scheme 10 shows the synthetic routes used to prepare S-bridged $Au^{I}Ni^{II}$ pentanuclear chiral metalloaggregates and $Au^{I}Co^{III}$ analogues.[157]

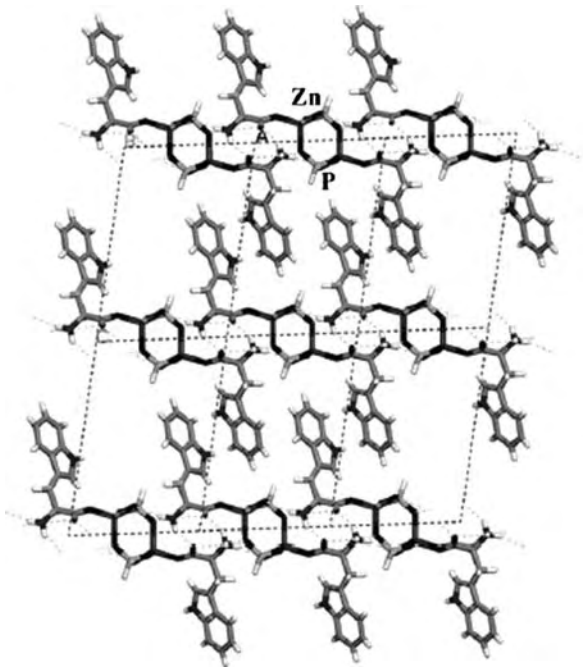
Two new homochiral inorganic-organic hybrid frameworks $[Zn(HPO_3)(C_{11}N_2O_2H_{12})]$ and $Zn_3(H_2O)(PO_4)(HPO_4)(C_6H_9N_3O_2)_2(C_6H_8N_3O_2)]$ have



Scheme 10

been synthesized in presence of chiral amino acids L-Trp and L-His. The formation of such one-dimensional structures might be potentially applied in chiral catalysis, biochemistry processes or as functional materials. For demonstration, the structure of the former comples viewed along the [010] direction is presented in Scheme 11.[158]

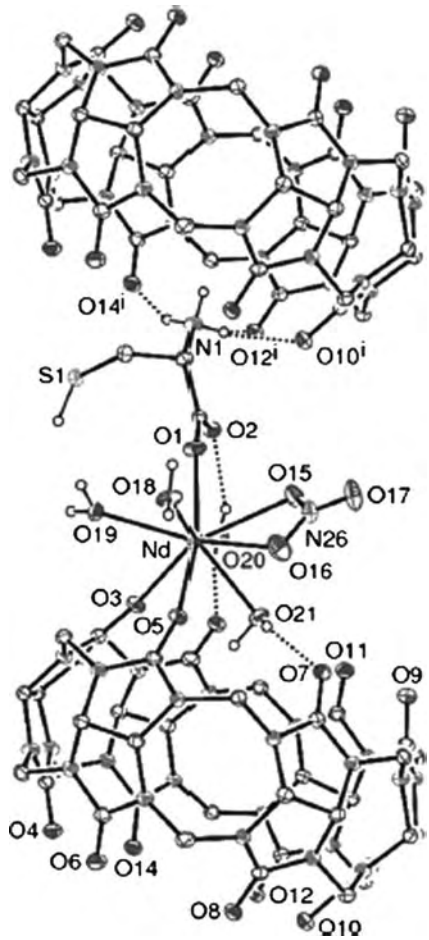
In another work two functionalized hybrid copper(II)- and cobalt(II)-containing polyoxometallates, $NaK_4[(A\text{-}\beta\text{-}SiW_9O_{34})Cu_4(OH)_3(H_2O)(H_3N(CH_2)_3COO)_2]\cdot 18H_2O$, $K_3\{Co(H_2O)_5\}_{0.5}[(B\text{-}\beta\text{-}SiW_9O_{34})Co_4(OH)_2(H_2O)_3(H_3N(CH_2)_3COO)_2]\cdot 18H_2O$, (trivacant Keggin anions) were built under mild conditions. These complexes represent the first polyoxowolframate compounds directly bound to amino acids.[159] More recently, lanthanides have started playing a major role in the field of molecular magnetism. Two three-dimensional 2p-3d-4f heterometallic frameworks featuring a nanosized $Ln_6Cu_{24}Na_{12}$ (Ln = Gd, Dy) cluster as a node have been obtained under microwave irradiation conditions through the reaction of H_2ANMA (H_2ANMA = L-alanine-N-monoacetic acid), $Cu(NO_3)_2$, and $Ln(NO_3)_3$ (Ln = Gd or Dy) with NaOH in deionized water. The gadolynium-containing compound exhibits ferrimagnetic behavior.[160] 3d-4f "hybrid" Co^{III}/Ln^{III} (Ln = Eu or Dy) molecular materials have been successfully synthesized by the use of 2-amino-isobutyric acid and characterization of their magnetic and optical properties has been performed.[161] Homochiral L-cysteate and 4,4′-bipyridine as mixed ligands were reacted with nickel(II), copper(II), manganese(II), zinc(II) or cadmium(II) ions and five new chiral coordination polymers were obtained. Depending on the metal ion used, three different types of architectures were synthesized.[162] Under a variety of solvent conditions, strong $\pi - \pi$ stacking interactions of the 1,8-naphthalimide groups organized the



Scheme 11

copper(II) complexes into supramolecular structures, when the copper(II) complexes were formed with the tri- and tetrafunctional enantiopure new ligands prepared from 1,8-naphthalic anhydride and the amino acids, L-Ala, L-Asn and D-phenylglycine.[163] When, the strong $\pi - \pi$ stacking 1,8-naphthalimide was linked to Gly, Ala or Ser and each of these synthesized new ligands was mixed with an appropriate metal salt such as salts of iron(II), cobalt(II), nickel(II), copper(II) or zinc(II), and 4,4′- bipyridine homochiral, helical supramolecular frameworks were fabricated.[164] Zwitterionic form of L-Cys as a chiral linker was successfully used to create one-dimensional assemblies *via* the reaction of this amino acid with neodymium, europium, or terbium nitrate and cucurbit[6]uril (Scheme 12).[165]

Due to its d^{10} electron configuration and softness, also the cadmium(II) ion is a preferred metal ion to create coordination frameworks.[166–168] A new example for ligand chirality-controlled *in situ* supramolecular hydrogel formation was found. The chirality of the Phe was found to control the formation of Phe-Cu(II) supramolecular metallogels. Intriguingly, the gelation capability was enhanced with increasing enantiomeric excesses (ee%). As decreasing both enantiomeric excesses of Phe towards its D- and L-forms, the gelation ability of Phe-Cu(II) based supramolecular metallogelator was found to be weakened and eventually disappeared, when ee% was lower than 30%. This behaviour was suggested to originate from the stereoselectivity of inter-ligand interactions, because efficient intermolecular hydrogen bonding, hydrophobic and/or $\pi - \pi$ stacking interactions might be weakened as ee% is lowered.[169] In another work, new gels based on



Scheme 12 (Reprinted with permission from P. Thuery, *Inorg. Chem.*, 2011, **50**, 10558. Copyright © 2011 by American Chemical Society.)

N-isopropylacrylamide and L-Orn were prepared. After the polymerization process the alpha-amino acid groups remained unbound. The presence of free alpha-amino groups attached to the polymeric network of the gels enabled the binding of these groups to metal ions. Interesting results relating the influence of the complexation with copper(II) ions on the swelling behaviour of the gels were obtained.[170] A thermolysis method was used to fabricate porous lanthanide oxides *via* thermolysis of precursors, which were synthesized hydrothermally with lanthanide (La, Ce, Pr, Nd) salts and sodium oxalate in the presence of glutamine or asparagine, respectively. Surprisingly, the asparagine and glutamine exhibited greatly different complexation abilities with lanthanide cations under hydrothermal conditions. Nanofibers of the crystalline oxalate precursor was produced by using glutamine as additive, but an amorphous compound precursor with a 3D macroporous structure was obtained in the presence of asparagine.[171]

To develop a tumor targeting nano-sized delivery system of *cis*-dichlorodiammine platinum(II), polymer metal complex micelles have been fabricated successfully from folate-conjugated PEG-*graft*-α,β-poly[(N-amino acidyl)-aspartamide].[162] With dicarboxylate and sulfate bridges a series of novel, *in vitro* cytotoxic dinuclear platinum(II) complexes of a new chiral amino acid derivative, 2-(((1*R*,2*R*)-2-aminocyclohexyl)amino)acetic acid, have been prepared.[173]

Molecular recognition phenomena are in the focus of interest by many reasons. Metallacrowns are known as a class of anion recognition agents, supramolecular catalysts or chiral separators. Chriral Ln(III)[15–metallacrown–5] complexes with phenyl side chains have been prepared and these complexes are able to encapsulate aromatic carboxylates reversibly in their hydrophobic cavities. Interestingly, the X-ray results illustrated differential binding of phenylalanine to the metallacrown based upon the chirality of the guest. Preferential binding of the *S*-isomer was found, because it forms an additional hydrogen bond to the metallacrown ring with the pendent ammonium group. On the contrary, the amine of the *R*-isomer orients away from the metallacrown face.[174] Magnetic and sorption properties of pentanuclear Cu(II)[12–metallacrown–4] have been evaluated *via* investigation of sorption of MeOH and EtOH.[175] Phenylalanine-based new double-stranded chelators have been synthesized and their metal binding abilities towards hard metal ions, such as lithium(I) and potassium(I), and also towards alkaline metal ions, such as lithium(I) and potassium(I), and towards various lanthanide ions have been examined. The results are thought to provide useful information *e.g.* to the development of chelators for the selective separation of lanthanides.[176]

To collect additional information for the preparation of new inorganic materials that incorporate proteins, model protein surfaces have been created from porous hen egg white lyzozyme crystals containing rhodium(III) ions and, as a result of this work, the cooperative effects of amino acids and hydrogen bonds promoting metal accumulation reactions have been clarified.[177] Development of metal complex based catalysts disposing high selectivity and efficiency is always a challenge. In a recent work, as bioinspired selective oxidation catalysts, transition metal complexes of amino acids encapsulated within solid supports have been created.[178]

As chiral sensors are important *e.g.* from the viewpoint of medical care and pharmaceutical technologies, numerous studies have been performed in this subject during the period reviewed[179–192] For example, chiral sensing system based on the formation of diastereomeric metal complex on a homocysteine monolayer has been developed for differentiate enantiomers of various amino acids,[179] while a copper(II)ion mediated, imprint polymer-based new sensor has been proposed for enantioselective sensing of L-His at trace level.[180] Selective coordination of Cys to mercury(II) was used to create a rapid, sensitive, selective fluorescence method for detection of Cys,[171] but, in another analytical method for the determination of thiols, the protocol is based on the oxidation of thiols by Ce(IV) and the formation of fluorescent Ce(III) disulfide complex.[182] By conjugation of two dansyl fluorofores with a Lys residue, a new fluorescent sensor has been synthesized and successfully used for detection of mercury(II).[183] Among the

examined metal ions, nickel was found to provide the best selectivity toward the detection of Trp, when a metal ion-assisted simple infrared optical sensor was developed for the selective detection of Trp in biological fluids.[184] The results obtained for the electrochemical behaviour of cobalt nanoparticles attached to graphene modified glassy carbon electrode are promising to develop a nonenzymatic electrochemical sensor *e.g.* for detection of Cys and *N*-acetyl-cysteine.[185] In this subject, in another work, a cobalt hydroxide nanoparticles modified glassy carbon electrode was used to study the electrocatalytic oxidation of Cys, Met, Cystine and *N*-acetyl-cysteine[186], and the electrochemical behaviour of Cys and cystine on carbon-paste electrodes modified with iron(II), cobalt(II), nickel(II) and copper(II) phthalocyanines has been also investigated.[187]

L-phenylalaninamide, L-Lys and L-Thr were used as chiral selectors when enantioseparation of dansylated amino acids by ligand-exhange capillary electrophoresis method was carried out. The central metal ions involved in the investigation were copper(II), cobalt(II), cadmium(II), nickel(II) and zinc(II).[188] The recent progress achieved in capillary electrophoresis related to analysis of amino acid enantiomers has been also summarized in another paper.[189] A palladium-based extraction system for enantioselective liquid-liquid extraction of underivatized amino acids was also developed during the period reviewed.[190]

A protocol was developed for the synthesis of a new technetium complex, which is stable at neutral pH and is suitable *via* a new labelling strategy for the labelling of pharmacophores, amino acids and carbohydrates.[193] It is a serious problem if pollution of water and soils by heavy metals happens and this problem has to be solved. EDTA-like surfactants or hydroxamate-based ones are among the known efficient agents for extraction of several metal cations. Based on a recent work, biodegradable, non-toxic and good foaming sugar-based agents have been also suggested as candidates for this purpose. The sugar-based surfactants have been functionalysed with chelating moieties like amino acids.[194]

3 Peptide Complexes

3.1 Thermodynamic and structural studies on the metal complexes of small model peptides. The range of experimental techniques

The stoichiometry, stability constants and structure of metal complexes formed with glycylglycine and related small peptides have already been well characterized. On the other hand, taking into account the large number of possible metal ions and especially the various small peptides many equilibrium and structural data are still missing from the literature. Several papers have been published recently on the determination of protonation equilibria[195,196] and for the stability constants of the cerium(IV),[197] and zinc(II)[198] complexes of GlyGly and related peptides. The stability constants of amino acid and peptide complexes are generally determined by pH-potentiometric measurements. It has been shown recently that comparable results can be obtained with the use of copper selective electrodes.[199] The major advantage of the latter technique that it enables conditional stability constants to be determined for ligands for which acid dissociation constants are not available.

The application of various forms of mass spectroscopic technique seems to be a rapidly growing field in the coordination chemistry of peptides. A recent review[16] based upon 253 references provides a general overview on this subject with a special emphasis on the applicability of electrospray ionization mass spectrometry (ESI-MS). It is obvious from this compilation that ESI-MS is a very efficient technique to study the noncovalent interactions of peptides with metal ions and to elucidate the stoichiometry of complexes formed under different conditions. The major advantage of the application of mass spectrometry comes from its widespread applicability for either small or large peptide molecules and for the interaction with very different metal ions. Some selected examples of these applications include the studies on the interactions of alkaline and alkaline earth metals with the peptides containing Ala and Arg residues[200,201] and with the dipeptide carnosine,[202] the silver(I) complexes of substance P[203] and the copper(II) and nickel(II) complexes of various multihistidine peptides.[204] The metal ion interaction with the peptide backbone in polyalanine peptides was also explored by means of mass spectrometric measurements.[205] Moreover, mass spectrometry proved to be an efficient tool to study the metal ion promoted cleavage of peptide bond in various palladium(II) or copper(II) complexes of peptides containing Met, His or Cys residues.[206] The experimental difficulties coming from the possibility of metal deposition during the ESI-MS analysis of peptide complexes have also been reported.[207]

The use of density functional theory (or DFT calculations) is also a rapidly increasing field for the characterization of the interaction of peptides with metal ions. Cobalt(I) and cobalt(II) complexes $(Gly)_nGly$ peptides[208] have been evaluated *via* DFT calculations. In another study the calculations were performed for the oxovanadium(IV) complexes of various tripeptides containing His or Cys residues.[209] It was found that DFT calculations enable various structural and spectroscopic features, like electron paramagnetic resonance parameters to be calculated. The acidity of free and metal ion coordinated imidazole residues was evaluated in another publication,[210] and the major coordination sites and bonding energies of peptides in silver(I) complexes was also reported.[211]

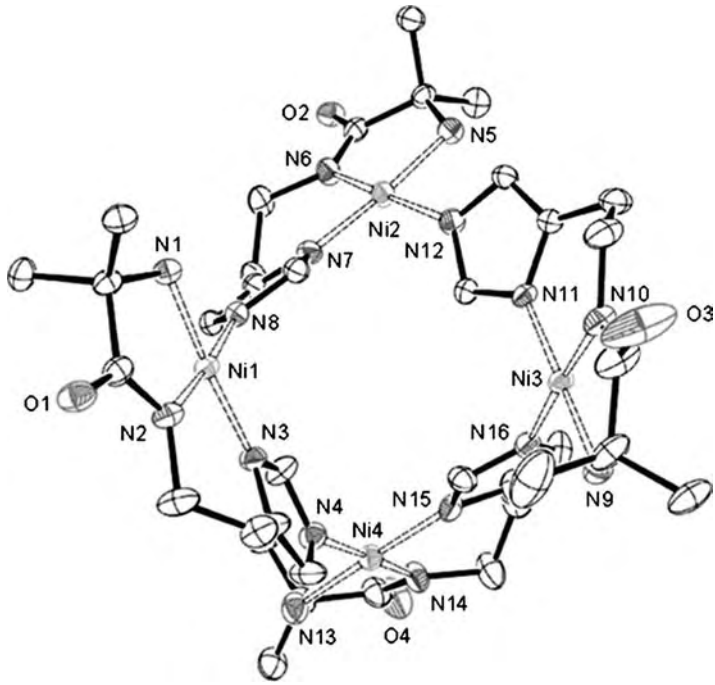
Relatively low number of papers have been published on the role of weakly coordinating side chains in peptide complexes. A very distorted trigonal bipyramidal coordination of zinc(II) was described in the peptides of α-aminoisobutyric acid,[212] while the aromatic side chains of PhePhe were found to be able to encapsulate various metal ions.[213] A systematic study has been performed to understand the influence of the extra carboxylate groups of Asp and Glu residues in their palladium(II) complexes.[214] In previous publications it was reported that aspartyl residues present in N-terminal or internal positions of a peptide molecule may significantly enhance the thermodynamic stability of corresponding copper(II) and nickel(II) complexes, while glutamyl moieties do not affect significantly the coordination behaviour of peptide molecules. Palladium(II) is a soft metal ion and generally it has a low affinity towards the oxygen donor ligands. Interestingly, it was found that β-carboxylate groups of aspartyl residues can significantly increase the stability and modify the coordination geometry of palladium(II) complexes, too.

3.2 Metal complexes of peptides of histidine

Imidazole-N donor atoms of histidyl residues are the most common metal binding sites in metalloenzymes and in a series of other metalloproteins. Various peptides of histidine are the most frequently used and simplest models to understand the structural features and catalytic activity of these molecules. The discovery of the molecular background of the various forms of neurodegenerative diseases gave a big impetus for further studies on the metal complexes of histidyl peptides. It is now widely accepted that the development of neurodegeneration is linked to the conformational changes of natural proteins. These proteins are generally rich in histidyl residues providing a good chance for stable interaction with metal ions. As a consequence, the number of studies on the various complexes of peptides containing histidyl residues increased tremendously in the last few years. The most important results of these studies will be discussed in the next few subsections.

3.2.1 Complexes of small model peptides containing histidyl residues. A large number of metal complexes of simple dipeptides containing one histidyl residue have already been well characterized because the N- and C-terminal histidyl residues have significantly different effect on the complex formation processes. More recently, the gold(III) complexes of GlyHis and AlaHis have been prepared from acidic solutions and their structures characterized by X-ray crystallography.[215] The same 3N-coordination mode $[NH_2,N^-(amide),N(3)_{im}]$ was found in these complexes as reported previously for the corresponding copper(II), nickel(II) and palladium(II) systems. An interesting feature of the peptides with NH_2-X-His sequence is that they can easily form tetranuclear complexes in slightly alkaline samples with the involvement of the above mentioned donor atoms and the deprotonated pyrrole-type N(1) donor of imidazole. The nickel(II) complex of a pseudopeptide α-methyl-alanyl-histamine has been studied recently and the species with $[Ni_4H_{-8}L_4]$ stoichiometry was prepared in the solid state and its structure determined by X-ray crystallography.[216] This tetrameric complex is shown by Scheme 13 demonstrating the square planar coordination geometry of the nickel(II) ions and the bridging imidazolato residues.

L-carnosine (β-alanyl-L-histidine) is an endogenous dipeptide widely distributed in several animal species. The presence of a β-amino acid in the sequence results in a difference between the structures of complexes formed with carnosine and other dipeptides. The importance of this dipeptide is further enhanced by its possible therapeutic applications as an antioxidant and antiinflammatory agent. These applications are, however, limited by the breakdown of the peptide bond by specific peptidases. More recent studies prove that this problem can be overcome by the conjugation of carnosine with various carbohydrate derivatives including cyclodextrin,[217] D-trehalose[218] and other small sugar moieties.[219] F-moc-L-carnosine was prepaped in another laboratory and its ability to form amyloid fibrils was studied in the presence and absence of zinc(II) ions.[220] The copper(II) complexes of several tetrapeptides containing N-terminal β-amino acids (Ala and Asp) and two histidyl residues were also investigated.[221] It was concluded that the



Scheme 13 (Reprinted with permission from K. Selmeczi, P. Gizzi, D. Champmartin, P. Rubini, E. Aubert, S. Dahoui and D. Henry, *Inorg. Chem.*, 2010, **49**, 8222. Copyright © 2010 by American Chemical Society.)

insertion of the β-carboxylate groups in peptide amide bonds results in a significant structural modification of the peptide chain and in the unususal coordination abilities of the ligands.

Copper(II) complexes of low molecular weight histidyl peptides are often used as superoxide dismutase (SOD) mimics. The copper(II) complex of the tetrapeptide HisGlyGlyTrp has been studied by ESI-MS, UV-vis and potentiometric measurements. Significant pH-dependence of the speciation was obtained and the SOD activity of the major species existing under physiological conditions was evaluted.[222] Both terminal amino and side chain imidazole nitrogen donors are considered as primary metal binding sites of peptides. As a consequence, the complex formation processes of oligopeptides containing these residues in well separated locations are especially complicated. Copper(II) and nickel(II) complexes of a pseudo tetrapeptide GlyGlyGlyHistamine and its N-protected counterpart were studied by means of potentiometric and various spectroscopic techniques.[223] Substantially higher stabilities were determined for the ML complexes of the terminally free peptide supporting the metal binding of both crucial moieties. The increase of pH resulted in the formation of various amide bonded species for both metal ions. Luteinizing hormone releasing hormone (LHRH: pEHWSYGLRPG) is a decapeptide containing one histidyl residue. Its biological activity is connected to its interaction with divalent metal ions. NMR spectroscopy was used to follow the interaction of LHRH with copper(II), nickel(II) and zinc(II) ions. The formation of a

specific helical structure with the involvement of His(2) and Leu(7) residues was concluded.[224]

Oxytocin and vasopressin are well-known peptide hormones containing disulfide bridges. The histidine analogs of these hormons have been prepared, in which cysteinyl residues are replaced by histidine. Copper(II)[225] and zinc(II)[226] complexes of these peptides were studied by potentiometric and NMR measurements. It was found that the peptide can efficiently bind both metal ions at physiological pH and the major species contain the His-M-His structural motif instead of the Cys-S-S-Cys bridge. Three new peptide hydroxamic acids containing histidyl residues were synthesized and their copper(II), nickel(II) and zinc(II) complexes studied by potentiometric and various spectroscopic techniques.[227] A competition between the hydroxamate and imidazole functions was found in all systems depending on the metal, the ligand and especially the pH of solutions. The imidazole-N donor was found to play the most determinant role in the copper(II) containing systems, while its role was negligible in the corresponding zinc(II) complexes. The complexes with the coordination of deprotonated amide nitrogens predominated in slightly alkaline samples.[227] The histamine and histidine conjugates of 2′-deoxyriboadenosine have been prepared in another study and copper(II) complexes were investigated by means of potentiometry, UV-vis and EPR spectroscopies. It was found that these conjugates bind copper(II) ions very efficiently and the conclusions were supported by theoretical calculations, too.[228,229]

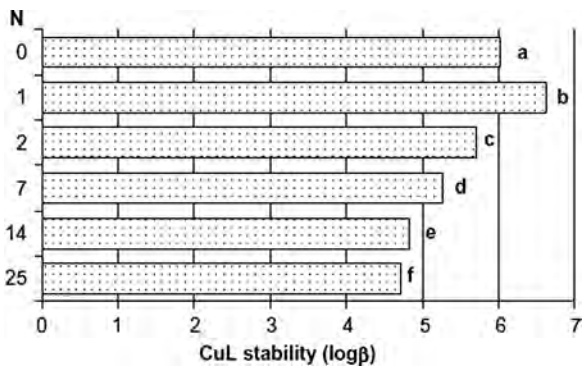
3.2.2 Complexes of multihistidine peptides. The studies on the metal complexes of peptides containing two or more histidyl residues became one of the most rapidly growing research areas in the field of metallopeptide chemistry. Some representatives of these ligands have already been reported in the previous subsection,[221,225,226] while many other examples will be discussed with the fragments of prion protein and related substances in the next subsection. The simplest models of these peptides and their most important coordination features are presented here.

Copper(II), nickel(II) and in some cases the zinc(II), cobalt(II) and cadmium(II) complexes of N-terminally protected oligopeptides containing two to four histidyl residues were studied by several groups using potentiometric and various spectroscopic techniques.[230–235] The metal binding ability of all these peptides is increased by the involvement of the imidazole residues in coordination. The stoichiometries and the structures of the major species are very much dependent on the sequence of the peptides and the nature of metal ions and no chance to recall all information in this review. There are, however, several general conclusions on the complex formation with these ligands as shown below:

(i) Imidazole-N donor atoms are the primary metal binding sites of these ligands and the simultaneous coordination of these nitrogen atoms results in the formation of 2N, 3N or 4N macrochelates. The peptides containing the histidyl and sarcosyl (= N-methylglycyl) residues in alternating positions, Ac-$(HisSar)_n$His, are the best suited ligands to follow macrochelation bacause sarcosyl residue prevents amide deprotonation.[230,231] It was found that the thermodynamic stability of these macrochelates follows the

trend: Cu(II) > Ni(II) > Zn(II) > Co(II) ~ Cd(II). Of course, the increase in the number of histidyl residues also enhanced the metal binding affinity of the ligands. Moreover, the stability of peptide complexes was very much influenced by the distance of histidyl residues. This is demonstrated by Scheme 14 where the stability constants of 2N macrochelates of copper(II) are plotted as the function of the number of separating amino acid residues. It is clear from this Figure that both adjacent and the very distant locations reduce the stability constants of complexes. The comparison of these data with those reported for peptides containing three or four histidyl residues leads to similar conclusions: the highest stability is obtained with peptides where the histidyl sites are separated by one or two amino acid residues.[231]

(ii) The macrochelation can increase the thermodynamic stability of peptide complexes, but cannot prevent the deprotonation and metal ion coordination of amide groups. The imidazole-N(3) donors are the anchors for these reactions and the deprotonation of the first two amide groups generally takes place in a cooperative process resulting in the tridentate (N_{im},$2N^-$) coordination mode. The pH range of this reaction is very close to the physiological one for copper(II) ($5 < pH < 7.5$) and occurs in the slighly alkaline samples for nickel(II), but the accurate values depend on the sequence of peptides. The deprotonation and metal ion coordination of the third amide residue can occur by increasing pH, around pH 8.0 for copper(II) and 8–10 for nickel(II). It is important to note that the metal ion promoted amide coordination was observed only in the copper(II) and nickel(II) containing systems[232–235] and these reactions were definitely ruled out for the corresponding cobalt(II) and cadmium(II) complexes.[230] Some examples have already been reported for the zinc(II) induced amide binding but it was generally observed in the presence of side chain carboxylates as it will be shown for the amyloid fragments in the next subsection. Coordination geometry of copper(II) complexes corresponds to the usual distorted octahedral environment, while the corresponding nickel(II) complexes are square planar, diamagnetic species.[230,233,235]



Scheme 14 (Reprinted from Inorganica Chimica Acta, 362, C. Kállay, K. Várnagy, G. Malandrinos, N. Hadjiliadis, D. Sanna and I. Sóvágó, Thermodynamic and structural characterization of the macrochelates formed in the reactions of copper(II) and zinc(II) ions with peptides of histidine, 935–945, Copyright (2009), with permission from Elsevier.)

(iii) The presence of well-separated histidyl residues makes possible the formation of polynuclear copper(II) and nickel(II) complexes with these peptides. The most interesting examples of these complexes will be discussed in the next paragraphs in connection with the prion peptide fragments.

In a recent study it was also reported that the cyclopeptides with multi-His motifs are also effective ligands to bind copper(II) ions. The formation of $4N_{im}$ macrochelates was suggested in sightly acidic samples, but the amide coordination of these ligands is suppressed as compared to those of the linear counterparts.[236]

3.2.3 Complexes of peptide fragments of natural proteins involved in neurodegeneration. Various forms of neurodegenerative diseases are among the major risks for human health. The development of neurodegeneration is generally linked to the misfolding or abnormal conformational changes and subsequent aggregation of otherwise natural proteins. Alzheimer's disease and Parkinson's disease are the most frequent and best studied forms of neurodegeneration, but the infectious character of prion diseases makes these disorders also widely studied. In the last 10 to 15 years a great number of publications seem to support the role of metal ions in the development of neurodegenerative diseases. Neither the treatment of the diseases nor the exact mechanism of the aggregation processes of proteins are completely understood promoting further studies for the clarification of neurodegeneration. Several reviews[17–23] and some books are dealing with the involvement of metal ions in these processes and the most recent was published in 2011.[24]

There are two major approaches of the studies on the role of metal ions in neurodegeneration. Lots of papers are published every year on the metal ion-protein interactions and also on their peptide fragments, separately. This Chapter is devoted to the coordination chemistry of peptides, therefore we will focus on the studies of peptide complexes, but more and more publications report the results from both aspects. The complexation with copper(II) is the best studied in all cases, but the iron(II/III), zinc(II), nickel(II) and manganese(II) containing systems are also widely studied.

The interaction of metal ions with prion protein received increasing attention in the last decade. The whole protein contains ten histidyl residues of which at least six are available for metal binding. Previous studies reveal that this protein has an outstanding copper(II) binding affinity and the results obtained for this interaction are reviewed recently.[18] The most recent studies are still dealing with the copper(II) complexes with a special emphasis on the selectivity of binding sites. On the other hand, the role of other transition elements and the effects of structural modifications of the peptides are in the centre of measurements. A new technique nanopore analysis can be used to study the conformational changes of individual peptides and protein molecules. It was found that the conformation of prion protein and its peptide fragments can be modulated by divalent metal ions including copper(II) and zinc(II), but not magnesium and manganese(II).[237] The pH dependence and concomitant redox activity of the copper(II)-prion protein interaction was observed in another study.[238]

The octarepeat domain consisting of four repeated octapeptide segments -$(PHGGGWGQ)_4$- is considered as the major copper(II) binding site of prion protein. The binding modes of the corresponding species were first described by X-ray, potentiometric and various spectroscopic studies and now supported by computer simulations.[239] In addition to the usual spectroscopic measurements voltammetry was also used to characterize the complexes of copper(II) with the PEG-ylated tetra-octarepeat fragment.[240] In the presence of excess copper(II) ions the formation of $[Cu_3H_{-6}L]$ and $[Cu_4H_{-8}L]$ species was detected with the involvement of imidazole and peptide nitrogens and a carbonyl oxygen in a square pyramidal coordination geometry. It was also observed that these species easily react with NO molecule but this interaction was ruled out in the presence of ligand excess when the $[Cu(N_{im})_4]$ coordinated species predominate. The copper(II) complex of the N-terminal repeat region of the opossum prion protein was also studied. Similar coordination geometry but slightly lower stability was reported than for the corresponding human counterparts.[241] The copper binding properties of the mutants with extended octarepat domain led to interesting conclusions. It was found that the domains with eight or nine total repeats arrested in the multihistidine coordination mode dictated by the increased binding affinity. It was also found that this behaviour correlates with the early onset of prion disease.[242] Former studies have already demonstrated that the zinc(II) binding affinity of prion fragments is much weaker than that of copper(II). On the other hand, recent X-ray absorption spectroscopic measurements revealed the possible competition of zinc(II) with copper(II) for binding with histidine.[243] Similar conclusions were drawn from the potentiometric and spectroscopic measurement performed on the copper(II) and zinc(II) complexes of zebra-fish prion-like protein. The binding affinity of copper(II) was found to be much higher than that of zinc(II) but the existence of heteronuclear species was also identified.[244,245]

Histidines of prion protein outside the octarepeat domain were also considered as metal binding sites of the protein. Copper(II) complexes of HuPrP(76-114) containing two histidyl residues (H77 and H85) inside and two histidyl residues (H96 and H111) outside the octarepeat domain were studied by potentiometric and spectroscopic techniques.[246] The coordination chemistry of these peptides correlates well to those described for the multihistidine peptides in the previous subsection. The macrochelates predominate in the slightly acidic pH range and it is followed by the formation of $[N_{im},2N(amide)]$ and $[N_{im},3N(amide)]$ coordinated species by increasing pH. All these peptide fragments can bind as much copper(II) ions as the number of histidyl residues in the sequence supporting that all histidyl residues are independent metal binding sites. The relative binding affinities of the histidyl sites were obtained from the evaluation of circular dichroism spectroscopic measurement, and the following stability order was obtained: $His111 > His96 \gg His77 \sim H85$. Another study on similar fragments came to similar conclusions.[247] The enhanced metal binding affinity of the peptide fragments outside the octarepeat domain can be easily explained by the presence of proline in the octarepeat. Proline works as a breakpoint for amide coordination and rules out the amide deprotonation towards the N-termini of peptides. As a consequence, the size of the first $[N_{im},N(amide)]$

chelate is seven- and six-membered for the octarepeat and outside this domain, respectively.The explanation of the outstanding binding affinity of the H111 residue is not evident, but further studies on small fragments revealed that point mutations in position 109 with amino acids containing side chain alcoholic-OH or thioether functions may contribute to the enhancement of stability.[248] The weak axial interaction of Met109 residue and its role in the redox activity of the copper binding site was concluded in another study.[249] The strong palladium(II) binding affinity of this region of the protein was also reported.[250]

The nickel(II) complexes of several fragments of prion protein have also been studied by the combined application of potentiometric and spectroscopic techniques.[251] All data supported that complex formation processes of nickel(II) are very similar to those of copper(II) but with a significantly reduced stability, which shifts the complex formation into the slightly alkaline pH range. The major difference in the corresponding copper(II) and nickel(II) complexes was obtained for the preferred nickel(II) binding sites of peptides. H96 was found as the predominating nickel(II) binding site of all peptides, while H111 was reported for the corresponding copper(II) complexes. The square planar coordination geometry was characteristic of these nickel(II) complexes providing a good chance for the application of NMR measurements. By means of NMR experiments nickel(II) was used as a diamagnetic probe to study the differences in the coordination behaviour of peptide fragments of chicken and human prion proteins.[252] It was found that the two prion proteins exhibited different nickel(II) and copper(II) preferences.

Most of the studies on the prion fragments are dealing with the characterization of human prion protein but the structural similarity with the prions of other species promoted studies in this field, too. Chicken prion protein contains a tandem repeat region of hexapeptides with the sequence -$(PHNPGY)_4$- with additional histidyl residues outside this domain. It was found that the hexarepeat domain can bind only two copper(II) ions and the tyrosyl residue is also involved in metal binding. Superoxide dismutase (SOD) activity of the avian and human tandem repeat domains were also compared and no activity was found for the avian fragments.[253] The copper binding affinity of the avian fragments outside the tandem repeat regions was investigated in another study.[254] From the thermodynamic data, it was concluded that chicken peptide fragments are distinctly better ligands for coordination of copper(II) ions with respect to the human fragments.

Doppel protein (Dpl) is a paralog of the cellular form of prion protein (PrP^C) exhibiting common structural and biochemical properties. Its sequence is missing the octarepeat and toxic regions but it also contains histidyl residues for metal binding. Copper(II) complexes of the terminally protected peptide fragment Dpl(122-130) (=KPDNKLHQQ) of human doppel protein and its Asp-Asn mutated form was studied by potentiometric and spectroscopic measurements.[255] The comparison of the binding constants with those of prion peptide fragments (PrP106-114) showed that doppel peptide had a higher metal binding affinity at acidic pH, whereas the prion fragment binds the metal tightly at physiological pH. Similar

results were obtained for the copper(II) complexes of another fragment Dpl(122-139) in comparison with the aspargine-mutated peptide.[256]

The polypeptide amyloid-β (Aβ) consists of 39 to 43 amino acids and is considered to be responsible for the development of Alzheimer's disease (AD). In the last ten years more and more studies proved the involvement of metal ions, especially iron(II/III), copper(II) and zinc(II) in the onset of this disease. Several reviews and books[19–24] have been published recently to summarize the previous results on the metal-amyloid-β interaction, while the most important findings reported in the last three years are collected in this Chapter. There are two major directions of the studies on the metal-amyloid interactions. One group of scientists, mainly the biochemists and biologists try to understand the role of metal ions in the aggregation processes. The whole polypeptides, generally Aβ(1-40) or Aβ(1-42) are used in these experiments and a very wide range of metal ions are involved in these studies. The determination of the binding affinity and binding sites of Aβ is in the center of the research of the other groups of scientists. It is evident from the amino acid sequence of the peptide that the possible metal binding sites are located in the amino terminus of Aβ, therefore the fragments Aβ(1-16) and Aβ(1-28) or even shorter peptides are also studied in these experiments. The essential elements, copper(II), zinc(II) and iron(II/III) are the most frequently used metal ions in these studies. At that point, it is also worthwhile to mention that two major approaches are used to quantify the metal binding ability of these polypeptides. The dissocation constant (K_d) is used in one approach. This value can be easily determined, but it is a conditional constant valid only for certain conditions and pH values. The stability constants (log β) values are used in the other approach. The determination of these thermodynamic constants are sometimes rather difficult but they can provide information on the coexistence of various species and binding modes. The other advantages and disadvantages of these approaches are discussed in two recent reviews.[18,22]

The major difficulties in the determination of thermodynamic stability constants of Aβ complexes comes from the low solubility of these complexes but conjugation with polyethyleneglycol (PEG) helps to overcome this problem. Complex formation of zinc(II) ions with Aβ(1-16)PEG and its fragments were studied by potentiometric and NMR techniques.[257] It was found that the peptide can efficiently bind zinc(II) ions but the major metal binding site is different for copper(II) and zinc(II). In the case of copper(II) the amino terminus was described as the primary binding site,[258] while the preference for the coordination at the internal histidyl residues (His13 and His14) was suggested for zinc(II). As a consequence, polynuclear mixed metal species can be easily formed with this peptide as it was reported for the copper(II)-zinc(II)-Aβ system.[259] Similar studies have been performed for the elucidation of the complex formation with nickel(II).[260] It was found that the major nickel(II) binding sites are the same as those of copper(II) and the existence of polynuclear mixed metal species was also suggested.

In addition to the classical potentiometric and spectroscopic techniques, theoretical calculations and mass spectrometric measurements are also frequently used to understand the metal ion-Aβ interaction. In the case of the copper(II)-Aβ system, the theoretical calculations supported the

previous findings for the preferred metal binding sites.[261] The precise affinity for iron binding of amyloid peptides cannot be determined by the classical techniques but it was reported on the basis of computer calculations.[262] On the basis of NMR and ESI-MS measurements, however, no interaction of aluminium(III) and iron(III) ions with Aβ(1-28) was found around physiological pH.[263] On the contrary, the binding of iron(III) by Aβ(1-40) was reported from another mass spectrometric study.[264] The efficient copper(II), zinc(II) and nickel(II) binding of Aβ(1-40) was also reported from mass spectrometric measurements and it was concluded that pH-dependent metal binding may induce conformational changes, which affect the affinity towards other metal ions.[265] The important role of 11-14 residues of Aβ in zinc(II) binding was proved by mass spectrometry and their role in the zinc(II) induced dimerization was also concluded.[266]

Huge number of studies have been performed on the factors influencing the aggregation processes of Aβ(1-40) and related peptides. These studies are extremely important from biological and especially from a medical point of view but generally do not give a comprehensive characterization of the role of metal ions in these processes. Therefore we do not go into details in this subject just give a list of the most important studies involving metal ions. A number of biophysical techniques were used to study the interaction of aluminum(III), copper(II), iron(II) and zinc(II) with Aβ(1-42).[267] It was found that these metal ions alter differently the aggregation and toxicity of the peptide and the specific role of aluminum(III) was emphasized.[268] The effect of silver ions on the conformational changes of peptides was studied by CD spectroscopy and a protective role of this metal ion was concluded.[269] A series of other studies, however, revealed the outstanding role of copper(II) and zinc(II) ions in the conformational changes or aggreagation proceses of the peptides.[270–273]

Metal ion catalyzed oxidative damage of peptides and proteins is one the most popular hypothesis on the possible role of metal ions in neurodegeneration. The less common oxidation states of metal ions are also involved in these oxidation processes of peptides. In one of the corresponding studies it was concluded that copper(I) can bind more tightly Aβ peptides than copper(II) does.[274] As a consequence, the low oxidation state complexes of copper and iron are generally considered as critical reactants in the oxidative transformations of peptides or in the formation of reactive oxygen species (ROS).[275–280] The interaction of amyloid peptides with cell membranes is also considered as crucial point in the pathology of AD. The important role of copper(II) and zinc(II) is widely accepted in this interaction[281] but the enhanced toxicity from aluminum(III) was also concluded.[282]

The experiments for the development of efficient therapeutic agents for the treatment of AD received also increasing attention. Various chelating agents are considered as promising candidates in this field. The efficiancy of clioquinol in amyloid aggregation was evaluated experimentally,[283] while theoretical calculations have also been performed to design the most appropriate chelators.[284] The design of multi-target directed ligands against plaque formation with additional antioxidant and metal complexing properties is a new and promising strategy in the treatment.[285] Insulin-degrading enzyme is the main metalloprotease involved in Aβ degradation.

Its effectiveness is modulated by the presence of various metal ions, but the inhibitory effect of copper(II) was reversed by adding zinc(II).[286] Various metal based drugs including complexes of rhodium(III) and iridium(III) were also tried as potential inhibitors of amyloid fibrillogenesis.[287]

α-Synuclein is a 140 amino acid residue protein and plays a key role in the pathogenesis of Parkinson's disease (PD). It is also a metal binding protein but its metal complexes are much less studied and the role of metal ions in the pathological processes is poorly understood. On the basis of a recent study three major copper(II) binding domains of the protein have been proposed: (i) the (1-9) N-terminal residue, (ii) His-50 and (iii) Asp and Glu in the C-terminal domain.[288] In another study, two independent and non-interacting copper-binding sites were identified in the N-terminal region of the protein.[289] The interactions of di- and trivalent cations copper(II), lead(II), iron(II) and iron(III) with the C-terminal domains of the protein were studied by ESI-MS and fluorescence spectroscopy.[290]

3.2.4 Complexes of other biologically relevant ligands with histidyl residues. Albumin is one of the most effective metal binding proteins. It has been well known for a long time that the high copper(II) and nickel(II) binding ability of albumin and related ligands is connected to the ivolvement of nitrogen atoms of the terminal amino, two deprotonated amide and His(3) residues in coordination (ATCUN motif). Two papers have been published on the exchange reaction of copper(II) ions between amyloid-β peptide and human serum albumin (HSA) or its N-terminal fragment (DAHK).[291,292] The results unambiguously prove that both HSA and DAHK are able to remove copper(II) from its Aβ complexes and to reduce the copper(II) induced aggregation. In another study it was demonstrated that even small molecules including common buffers and heterocyclic ligands can modulate the binding properties of the multimetal binding site of HSA.[293] Bovine serum albumin (BSA) was also used to study the metal ions induced protein aggregation processes.[294] A significant difference was observed in the copper(II) and zinc(II) promoted aggregation reactions.

Histones are also histidyl-rich proteins with enhanced metal binding affinity. The copper(II) and nickel(II) complexes of the peptide fragments of histone H2B was studied by potentiometric and various spectroscopic techniques.[295–298] Histidyl residues were identified as the primary metal binding sites but the specific roles of other side chain residues were also emphasized. It was concluded that these findings may contribute to the better understanding of the metal ion induced toxicity and carcinogenesis.

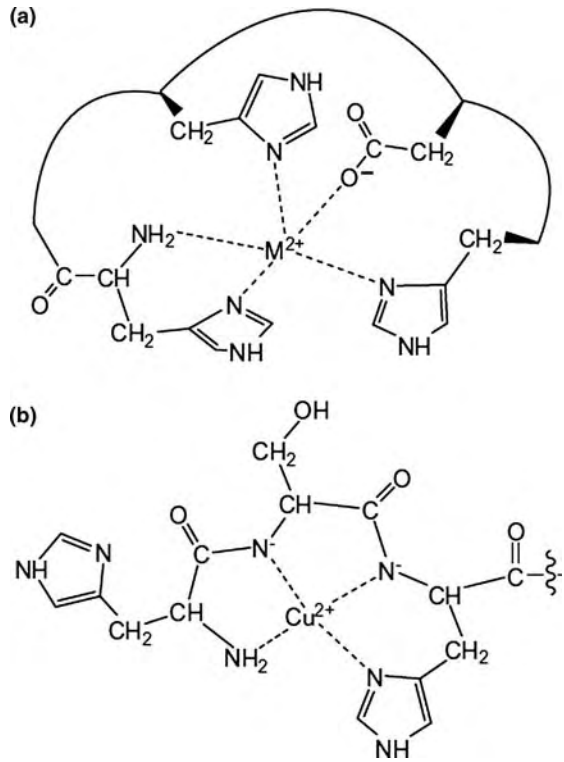
Histatins are a family of human salivary antimicrobial peptides. It was found that histatine-5, which has an albumin-like N-terminal sequence can bind nickel(II) five times stronger than HSA. On the basis of these results the involvement of these peptides in nickel(II) allergy and toxicity was conluded.[299] Two specicific copper(II) binding sites of the same peptide were identified by spectroscopic measurements and the significant oxidative activity of the complexes were considered to be responsible for the biological activity of these molecules.[300] The zinc(II) complex of a 12 amino acid fragment of histatine-5 containing three histidyl residues were investigated by means of NMR and CD techniques. The imidazole-N donor of His

residues were identified as metal binding sites but the three dimensional structures of the corresponding zinc(II) and copper(II) complexes were significantly different.[301] Hpn is another histidyl-rich bacterial protein related to Helicobacter pylori and have been suggested to be involved in nickel(II) homeostatis. Several fragments of Hpn containing two or three histidyl residues were synthesized and their copper(II) and nickel(II) complexes studied by potentiometric and various spectroscopic techniques. The albumin-like N-terminus of the peptides resulted in the enhanced metal binding affinity of the peptides and a specific effect from the hexaglutamyl sequence was also observed.[302]

The copper(II) and zinc(II) binding ability of the terminally protected penta-and decapeptide fragments of histidine-rich glycoprotein (HRG) was investigated by potentiometric and spectroscopic methods. The presence of six histidyl residues in the decapeptide offered a high affinity metal binding site for both metal ions. the formation of dinuclear copper(II) complexes was also suggested.[303] The N-terminal dodecapeptide fragment of human endostatin have also been synthesized and its copper(II) and zinc(II) complexes were studied. Endostatin is a recently identified broad spectrum angiogenesis inhibitor and rich in histidyl residues. The albumin-like N-terminus of the peptide resulted in the enhanced copper(II) binding affinity, but relatively stable zinc(II) complexes were also formed. The structures of the major zinc(II) and copper(II) containing species were, however, completely different as shown by Schemes 15.a and 15.b, respectively.[304]

Alloferons are multihistidine peptides and their coordination chemistry resembles those reported for model peptides in Section 3.2.2. The results obtained for the copper(II) complexes of various fragments revealed the formation of high stability macrochelates in slightly acidic samples followed by the existence of amide bonded species by increasing pH.[305] In the case of a decapeptide fragment containing albumine-like N-terminus, the formation of dinuclear copper(II) complexes was also suggested.[306] Neuropeptide Gamma (NPG) is another natural multihistidine peptide and its ability for complexation with copper(II) and the products of metal catalysed oxidation were also studied.[307] In addition to the metal binding *via* the histidyl residues and amide nitrogens, the stabilizing role of the N-terminal Asp residue was observed. The copper(II) catalysed oxidation was found to be site-specific including the oxidation of two His residues and the cleavage of the G(3)-H(4) and R(11)-H(12) peptide bonds. Neurokinin A is a decapeptide found in mammalian neuronal tissue and contains an N-terminal His residue in the sequence. Its complex formation processes were found to be quite similar to those of simple model peptides with N-terminal His residues. Copper(II) catalysed oxidation of the peptide was also followed by LC-MS technique and Met and His residues were susceptible for oxidation.[308]

Angiogenin is one of the most potent angiogenesis factors, whose activity may be affected by copper(II) ions. The complex formation of the Ac-KNGNPHREN-NH_2 and Ac-PHREN-NH_2 fragments with copper(II) was found to be very similar to those of other terminally protected monohistidine peptides, but the involvement of side chain glutamyl carboxylate residue in copper(II) binding was also concluded.[309] The interactions of copper(II) and zinc(II) ions with the N-terminal tetradecapeptide fragment



Scheme 15 (Reprinted from Journal of Inorganic Biochemistry, 103/7, A Kolozsi, A. Jancsó, N. V. Nagy and T. Gajda, N-terminal fragment of the anti-angiogenic human endostatin binds copper(II) with very high affinity, 940–947, Copyright (2009), with permission from Elsevier.)

of human nerve growth factor (NFG) were characterized by means of potentiometric and spectroscopic techniques. The results demonstrate that the amino group is the first anchoring site for copper(II) and is involved in zinc(II) coordination at physiological pH.[310]

ESI-MS measurements were used to study the zinc(II) binding of the histidine-rich peptide H5WYG. H(11), H(15) and H(19) residues were found to be involved in metal binding.[311] This conclusion has been confirmed later with the studies using alanine mutated peptides.[312] It was concluded that this binding mode can promote the membran fusion at physiological pH. The toxicities of free peptides built up from glycine and histidine and their nickel(II) complexes were compared in another study and it was found that the complexation enhances the toxicity.[313] This observation was explained by the structural changes of peptides upon nickel(II) binding which allows a free penetration across the cell membrane. Nickel(II) complexes of the peptide fragments of Cap43 protein were studied by NMR spectroscopy. The sequences of the peptides correspond to a one to three times repeated motif: Ac-$(TRSRSHTSEG)_n$-NH_2 ($n = 1–3$). It was found that the peptides can bind as much nickel(II) ions as the number of histidyl residues in the sequence.[314] Angiotensin is an octapeptide hormone (DRVYIHPF) containing a single histidyl residue. Electrospray tandem mass spectrometry was used to determine the copper(II), nickel(II),

magnesium(II) and calcium(II) binding sites of the peptide. Completely different binding modes for the copper(II), nickel(II) and magnesium(II), calcium(II) species were suggested, respectively.[315] Hepcidine a 25-residue peptide hormone is considered as the master regulator of iron metabolism. Moreover, the peptide contains a histidyl residue at position 3 providing a good chance for the interaction of divalent metal ions. In agreement with this expectation, the high copper(II) binding affinity of the peptide was concluded from ESI-MS measurements. Interestingly, no interaction with iron(II) or iron(III) ions was detected under the same conditions. The mutation of the peptide with alanine or asparagine at position 3 reduced the copper(II) binding affinity.[316]

3.3 Metal complexes of peptides of cysteine

L-cysteine and its peptides have an outstanding ability to form stable complexes with a series of metal ions. The formation of the metal-thiolate bond is, however, rather selective for the soft transition and posttransition elements. Moreover, peptides of cysteine can be easily oxidised to disulfides providing a great variety for the parallel complex formation and redox reactions. L-cysteine is a common constituent of some natural proteins with enhanced metal binding affinity. They include metallothioneins, phytochelatins, iron sulfur proteins, blue copper proteins and zinc finger proteins. A huge number of papers have been published on the chemistry of these substances, but they are not the subject of this Chapter. Here, we will focus on the complexes of small, natural and model peptides containing cysteinyl residues in different numbers and locations.

The number of publications using the classical techniques for studying the metal complexes of simple di- or tripeptides of cysteine is relatively low. This can be easily explained by the experimental difficulties arising in the studies with thiol compounds. However, the rapidly growing applications of various mass spectrometric techniques opened a new perspect for the measurement of metal-peptide complexes. Some of the most recent publications and reviews in this field have already been discussed in Section 3.1, while another review is specifically devoted to metal-thiolate ligands in real biological samples.[317] Some other examples for the application of mass spectroscopy include the studies on copper(I/II) complexes of thiolate peptides,[318–320] because the fast redox reactions between copper(II) and peptides of cysteine rule out the use of the classical potentiometric and spectroscopic measurements. Moreover, mass spectrometry can be successfully applied to follow simultaneously the complex formation and metal induced cleavage of disulfide ligands. Oxytocin is a well-known peptide hormone containing a disulfide bridge and its copper(II) and nickel(II) complexes were studied by the combined application of mass spectrometric and other spectroscopic measurements[321,322] or DFT calculations.[323] The results obtained for the binding mode of the ligand in the corresponding complexes confirmed the previous results obtained by classical techniques, but MS provided further insight into the fragmentation and ionization processes. Isothermal titration calorimetry is also a relatively new and efficient experimental technique to study complicated metal-ligand interactions. This method was used to study the interaction of mercury(II) ions

with the di- and tripeptides of cysteine, histidine and tryptophan.[324] It is not surprising that the highest stability complexes were formed with the peptides containing the thiolate binding sites.

The discovery of the essential role of Helicobacter pylori in nickel(II) metabolism and metal ion homeostasis in gastric adaptation promoted the studies on the metal complexes of peptides containing cysteinyl and histidyl residues. The Hpn and HspA proteins are rich in histidyl and cysteinyl residues and are responsible for the above-mentioned processes. Nickel(II), zinc(II), cadmium(II) and bismuth(III) complexes of terminally protected decapeptides containing adjacent Cys residues were studied by potentiometric and various spectroscopic techniques.[325] Thiol sulfur atoms were the anchoring sites for all metal ions and the highest stability were reported for the complexation with bismuth(III), while the stability of nickel(II) complexes was higher than that of zinc(II). The complexes of the same metal ions with peptides containing separated cysteinyl residues have also been studied by the same group.[326,327] The following stability order of the metal ions was obtained: $Bi(III) \gg Cd(II) > Zn(II) > Ni(II)$ reflecting a reversed tendency in the stabilities of zinc(II) and nickel(II) complexes. Bismuth(III) complexes of peptides containing the Cys-X-Cys motif have also been studied and the data revealed the specific role of thiolate residues in metal binding.[328,329] The understanding of the structural and catalytic properties of nickel containing superoxide dismutase enzyme made further studies also necessary on the nickel(II) complexes of cysteinyl peptides.[330,331] It was found that the tripeptide AsnCysCys binds nickel(II) in a square planar coordination geometry with (2N,2S) binding sites and behaves as a structural and functional mimic of Ni-SOD.

The characterization of zinc finger proteins is another field which promotes the studies on the metal complexes of peptides with Cys and His residues. Using a family of six peptides it was possible to assess the effect of the hydrophobic side chains on the thermodynamic and kinetic properties of the corresponding zinc(II) complexes.[332] It was found that the packing of hydrophobic amino acids into a well-defined hydrophobic core can have a drastic influence on both the binding constant and the kinetics of metal exchange. Structural characterization of these zinc(II) and cadmium(II) complexes is generally a difficult task because of the spectroscopic silence of the metal ions. It has been shown recently, that correlation of ^{113}Cd NMR and ^{111m}Cd PAC (Perturbed Angular Correlation) spectroscopies provides a powerful approach for the elucidation of the Cd(II)-substitued Zn(II) proteins.[333]

The tripeptide glutathione (γ-glutamyl-L-cysteinylglycine) is the most abundant low-molecular-weight thiol-containing molecule in the cell. Phytochelatins [PC(n)] with the general sequence $(\gamma\text{-Glu-Cys})_n\text{Gly}$ ($n = 2$–11) are cysteine rich peptides synthesized by plants and are involved in metal ion bioregulation and phytoremediation. It is evident that the amino acid sequence of phytochelatins is closely related to that of glutathione. In the last few years huge number of papers were published on various aspects of the interactions of metal ions with glutathione, phytochelatins and related ligands. Most of these studies are focusing on the biological aspects of these

interactions and they are not discussed in this Chapter. We focus only on the basic coordination chemistry of these substances.

The possible application of glutathione and related ligands in heavy metal detoxification promoted the studies on the characterization of metal complexes of thiolate ligands. Cadmium(II) complexes of glutathione was studied by NMR measurements. The results revealed the significant pH-dedendence in the stoichiometry and binding modes of the major species, but the thiolate function was described as the predominating metal binding site.[334] DFT calculations have also been performed for the same system providing further evidence for the governing role of Cd-S bond.[335] ^{207}Pb NMR spectroscopy was used to determine the metal binding site of glutathione in its lead(II) complexes. The results indicate that glutathione binds lead(II) in a trigonal pyramidal geometry (PbS_3 centrum) and it was also proved that ^{207}Pb NMR is a promising technique to study lead(II)-thiolate interactions.[336] The zinc(II) complexes of oxidized glutathione and its nine analogs with C-terminal modifications have been studied by potentiometric and spectroscopic measurements.[337] In agreement with previous findings the reduced metal binding ability of the disulfides was obtained as compared to those of the thiolate ligands. The role of glycyl residue in the stabilization and structuring of the complexes was also reported.

In the case of various forms of phytochelatins the complex formation with toxic elements was in the focus of studies. It was confirmed that electrochemical techniques can be efficiently used for the clarification of the interaction of cadmium(II) with short phytochelatins.[338] ESI-MS measurements in combination with the results of voltammetric and circular dichroism experiments were used to describe the interaction of PC(5) with cadmium(II) and lead(II) ions.[339] Some other aspects of the biological role and coordination chemistry of phytochelatins are available in a review published recently.[340]

3.4 Metal complexes of other biologically relevant peptides and derivatives

A great number of peptide hormones and other peptides do not contain histidyl and/or cysteinyl residues but they have outstanding biological significance with interesting coordination behaviour. This is especially true for the complexes of peptides conjugated with other ligands. Several examples of these ligands and their complexes were studied recently and they are summarized shortly in this subsection.

The coordination ability of insect kinin analogs has been studied by potentiometric and spectroscopic techniques and by theoretical calculations.[341] It was found that the insertion of 4-aminopyroglutamate moiety instead of Pro residue leads to an effective ligand towards copper(II) ions. The enhanced stability of the complexes was explained by the specific conformation of the peptides caused by the structural modification. Human islet amyloid polypeptide (hIAPP) is a peptide hormone believed to be involved in the pathogenesis of type 2 diabetes. The rat analog of this peptide is highly homologous to human amylin but does not contain any histidyl residue. Copper(II) complexes of rat amylin fragments and several

mutants have been studied recently with the combined applicaton of several experimental techniques and DFT calculations.[342] It was found that even the terminally protected small hexa- and tetrapeptide fragments (Ac-VRSSNN-NH_2 and Ac-SSNN-NH_2) can bind copper(II) ions effectively in the physiological pH range. Two or three deprotonated amide nitrogens of the peptide backbone and a side chain asparagine were suggested to be involved in metal binding. The curiosity of this result comes from the fact that the involvement of amide nitrogens in metal binding generally requires the presence of an anchoring group (terminal amino, histidyl or cysteinyl residues) in the sequence. These data, however, demonstrate that peptides can be effective ligands even in the absence of anchoring groups, if more polar side chains are present in a specific sequence. Capreomycin is a polypeptide antitubercular antibiotic and an important therapeutic agent. Its structure is based upon a cyclic polypeptide built up from amino acid derivatives.[343] The high number of nitrogen donor atoms makes this molecule an effective complexing agent for copper(II) ions. One of the amino functions of the α,β-diaminopropionic acid residue was suggested as the anchoring site, with the additional coordination of amide nitrogens and amino group of β-lysine.

Peptide nucleic acids (PNA) are synthetic analogs of DNA containing the same nucleobases on a peptide backbone. Studies on their metal complexes started several years ago and a most recent observation suggests that the preferred handedness of a ligand modified PNA can be switched as a consequence of metal ion coordination to the ligand.[344] Nucleopeptides represent another group of molecules obtained by the conjugation of amino acids and nucleobases. A recent study on a novel dithymine L-serine tetrapeptide based supramolecular assembly revealed that the structure of this assembly can be modified by the interactions with metal ions.[345] Patellamides are natural cyclic peptides with important biological applications. Copper(II) complexes of derivatives obtained by the substitution of the oxazoline and tiazole heterocycles by dimethyl-imidazoles were studied by various spectroscopic techniques and DFT calculations.[346] It was found that the rigidity of the dimethyl-imidazole rings significantly changes the properties of both the free ligands and their copper(II) complexes. Pseudopeptides derived from valine have been synthesized, in which the metal binding sites are separated by different length aliphatic chains.[347] The properties of copper(II) and zinc(II) complexes of these ligands differ not only in the thermodynamic stability but also in stoichiometry suggesting the different coordination modes for the two metal ions. Schiff base ligands derived from aldehydes and various peptides can represent another big group of peptide derivatives. They are generally good complexing agents as it was demonstrated by a recent publication.[348] Lanthanide(III) ions generally form low stability complexes with peptides. A recent study on the La(III), Ho(III) and Eu(III complexes of various di- to pentapeptides came to the same conclusion indicating only the involvement of oxygen donors in metal binding.[349] The insertion of the unnatural aminodiacetate side chains into the sequence of hexapeptides, however, significantly enhanced the affinity of the peptides towards lanthanide(III) ions.[350]

3.5 Metal ion promoted hydrolytic and oxidative reactions of peptides

In the previous paragraphs the thermodynamic and structural characterization of metal-peptide interactions has been summarized. Metal ions are, however, able to promote various irreversible transformations of peptide molecules. Among these processess the metal ion catalyzed hydrolysis of peptide bond, the oxidation or oxidative cleavage of peptides and the oxidation of thiol residues or the cleavage of the disulfide moieties represent the most common reactions. The most important findings obtained in these fields are collected in the next few paragraphs.

Various metal ions or their metal complexes are able to catalyze either the formation or the hydrolysis of peptide bonds. An outstanding review on the formation of peptides within the coordination sphere of metal ions of classical and organometallic complexes have been published very recently.[15] This compilation contains 748 references and information also on the prebiotic aspects of this subject, therefore the individual papers are not discussed here.

The metal ion promoted hydrolytic cleavage of peptides and the design of artificial metallopeptidases are rapidly increasing fields in the coordination chemistry of peptides. Various palladium(II) and platinum(II) complexes are probably the most studied catalysts in these reactions. A common feature of these complexes that they can promote the selective cleavage of the peptide bonds. The most common residues include the methionyl and histidyl side chains which can easily interact with these residues with the subsequent hydrolysis of the neighbouring peptide bonds. The most recent publications provide further examples for these reactions and provide insight into the mechanism of the hydrolytic reactions with theoretical calculations.[351–354] The possible biochemical and industrial applications of these stubstances was also reported.[355] Sequence specific hydrolytic reactions of the peptides at the Xaa-Ser/Thr bond were also presented.[356,357] It was established that a square planar nickel(II) complex with His imidazole and three preceeding peptide amide coordination is required for this reaction. Former studies revealed that the insertion of azaGly residue into a peptide enhances the thermodynamic stability of the copper(II) complexes. At the same time, it was found in a recent study that copper(II) ions are able to induce the cleavage of these peptides. The rate of hydrolysis was also influenced by the presence of histidyl residues in the sequence.[358] Effective artificial hydrolases can be based on zinc(II) complexes, too. These reactions were studied *via* the zinc(II) complexes of new artificial peptides containing histidine side chains.[359] The hydrolysis of serine containing peptides catalyzed by molybdate anion was also observed.[360] No reduction of molybdenum was detected in the hydrolytic reaction. The cleavage of the disulfide bond of peptides is also a biologically significant process. The gold(I) assisted cleavage was investigated by DFT calculations and the reaction was found to be very much site specific.[361] In contrast with the previous reactions metal complexes can be promising enzyme inhibitors, too. The inhibition of the activity of the α-amylase enzyme was presented by means of an organotin-peptide complex.[362]

Metal ion induced oxidative damage of peptides and proteins and the simultaneous formation of reactive oxygen species (ROS) is one of the most

popular hypothesis on the possible role of metal ions in various human diseases. This idea promoted a huge number of studies on the redox chemistry of peptides in the presence of metal ions. Some of these studies have already been cited in Sections 3.2.3 and 3.2.4 in connection with the peptides of histidine.[274–280] Here we give a brief account of other studies demonstrating the great variety of metal ions and ligands involved in these studies.

Simple di- and tripeptides were used to study the influence of the amino acid relative position on the oxidative modification. It was found that oxidative behaviour of peptides is influenced by both the reactivity and specific location of amino acid residues.[363] The easy oxidation of cysteine and other thiols to disulfides has been well-known for a long time but more and more data prove the involvement of methionyl[364] and histidyl[365–367] residues in oxidative transformations. In agreement with the findings reported for the redox active metalloenzymes, copper[365–369] and iron[364] are the most common metal ions in the redox reactions of peptides, too. Moreover, peptide complexes may behave as effective inhibitors of various oxido-reductase enzymes providing a possibility for the application of these substances in food and cosmetic industry[370] or in the treatment of neuro-degenerative disorders.[371–374] The involvement of both oxidation states of metal ions in these reactions promotes the systematic investigations of the redox properties of copper(I/II) complexes of simple peptides. A recent study on the copper(II) complexes of various imidazole containing ligands using cyclic voltammetry provided further insight into the relationship between the structural and electrochemical parameters of copper(II) complexes.[375]

3.6 Applications of metal ion peptide complexes and derivatives

The favourable structural, thermodynamic, kinetic and transport properties of peptide complexes together with the simplification of peptide synthesis gave a big impetus to the various applications of peptide complexes. The medical applications are probably the best known but the analytical, synthetic and catalytic use of the complexes of peptides and their conjugates also represent a rapidly growing area in peptide chemistry.

Medical applications of peptide complexes include both therapeutic and diagnostic applications. In a recent review the use of small peptides to create biochemically specific molecular imaging agents and radiotherapeutic pharmaceuticals has been collected.[376] Ruthenium complexes are promising anticancer agents and recently their application in the treatment of Alzheimer's disease was also suggested. An octaarginine conjugate has been found to improve the uptake and DNA localization efficiency of ruthenium complexes.[377] A potential strategy for the treatment of Alzheimer's disease is based upon the inhibition of amyloid aggregation. The results obtained for the use of a mixed binuclear platinum(II)-ruthenium(II) complex[378] and for the organometallic derivatives of ruthenium(II) peptide complexes were reported in this field.[379,380]

^{99m}Tc complexes are probably the most frequently applied radio-pharmaceutical agents. The conjugation of these complexes with peptide molecules may help the transport and targeting of these agents.[381–383]

Various conjugates of peptides are used in many other fields of chemistry. A recent review on the peptide-nucleic acid conjugates[384] gives an overview on the synthesis, properties and applications of these substances. The synthetic applications of metal cyclam-peptide complexes and zinc(II) cyclen-peptide conjugates have also been reported.[385,386] Lanthanide complexes of various peptide conjugates were found to enhance the paramagnetic anisotropic effect which is valuable in the structural analysis of proteins.[387] A peptide-based reagent was described to selectively bind trivalent actinides over trivalent lanthanides and it is a very important step in the recycling process of nuclear waste.[388] Finally, it is intersting both from pratical and theoretical aspects that a new lanthanide binding motif, lanthanide fingers have been developed. These peptides utilize and Asp(2)Glu(2) metal-coordination environment to bind lanthanides in a tetracarboxylate mode.[389] Another structural modification of peptides can be obtained by the binding of separated side chains to metal ions which induces unstructured peptides to be alpha-helical in aqueous solutions.[390] Among the most valuable analytical applications of peptide complexes the design of new fluorescence probes[391,392] and GlyHisGlyHis based metal detectors fixed on silica surfaces can be mentioned.[393]

The development of metal complexes that cleave nucleic acids hydrolytically at physiological conditions is of great interest in the field of artificial metallonucleases. Peptide complexes represent an important group of these substances. The recently studied examples include the copper(II) and nickel(II) complexes of peptides containing histidyl and lysyl residues,[394–396] and the gold(III), palladium(II) and platinum(II) complexes of methionyl peptides.[397] The structure of peptide complexes and the role of various species in the hydrolytic or oxidative damage of nucleic acids are also involved in these papers.

References

1 G. Jie, Yu. Haijia, Z. Haiyuan, Xu. Haixia and Qu. Xiaogang, *Prog. Chem.*, 2009, **21**, 866.
2 E. Gao, C. Liu, M. Zhu, H. Lin, Q. Wu and L. Liu, *Anti-Cancer Agents Med. Chem.*, 2009, **9**, 356.
3 A. Mucha, M. Drag, J. P. Dalton and P. Kafarski, *Biochimie*, 2010, **92**, 1509.
4 N. A. Corfú, A. Sigel, B. P. Operschall and H. Sigel, *J. Indian Chem. Soc.*, 2011, **88**, 1093.
5 G. N. L. Jameson, *Monatsh. Chem.*, 2011, **142**, 325.
6 F. Mancin, P. Scrimin, P. Tecilla and U. Tonellato, *Coord. Chem. Rev.*, 2009, **253**, 2150.
7 A. Igashira-Kamiyama and T. Konno, *Dalton Trans.*, 2011, **40**, 7249.
8 H. J. Schneider and R. M. Strongin, *Acc. Chem. Res.*, 2009, **42**, 1489.
9 P. S. Dieng and C. Sirlin, *Int. J. Mol. Sci.*, 2010, **11**, 3334.
10 T. D. Fridgen, *Mass Spectrom. Rev.*, 2009, **28**, 586.
11 Y. Zhou and J. Yoon, *Chem. Soc. Rev.*, 2012, **41**, 52.
12 P. Gatt, R. Stranger and R. J. Pace, *J. Photochem. Photobiol. B*, 2011, **104**, 80.
13 E. L-Chrusczinska, *Coord. Chem. Rev.*, 2011, **255**, 1824.
14 D. Witkowska, M. Rowinska-Zyrek, G. Valensin and H. Kozlowski, *Coord. Chem. Rev.*, 2012, **256**, 133.
15 W. Beck, *Z. Anorg. Allg. Chem.*, 2011, **637**, 1647.

16 D. D. Carlton and K. A. Schug, *Anal. Chim. Acta*, 2011, **686**, 19.
17 H. Kozlowski, A. Janicka Klos, J. Brasun, E. Gagelli, D. Valensin and G. Valensin, *Coord. Chem. Rev*., 2009, **253**, 2665.
18 H. Kozlowski, M. Luczkowski and M. Remelli, *Dalton Trans*., 2010, **39**, 6371.
19 J. A. Duce and A. I. Bush, *Progress in Neurobiology*, 2010, **92**, 1.
20 V. B. Kenche and K. J. Barnham, *British J. Pharmacol*., 2011, **163**, 211.
21 C. Migliorini, E. Porciatti, M. Luczkowski and D. Valensin, *Coord. Chem. Rev*., 2012, **256**, 352.
22 G. Arena, G. Pappalardo, I. Sóvágó and E. Rizzarelli, *Coord. Chem. Rev*., 2012, **256**, 3.
23 P. Faller and C. Hureau, *Dalton Trans*., 2009, 1080.
24 Neurodegeneration, Metallostasis and Proteostasis, ed. D. Milardi and E. Rizzarelli, RSC Publishing, 2011.
25 M. Markovic, N. Judas and J. Sabolovic, *Inorg. Chem*., 2011, **50**, 3632.
26 M. Yu, W. Chu and Z. Wu, *Nucl. Instrum. Methods Phys. Res. A*, 2010, **619**, 408.
27 M. Hanif, H. Henke, S. M. Meier, S. Martic, M. Labib, W. Kandioller, M. A. Jakupec, V. B. Arion, H. B. Kraatz, B. K. Keppler and C. G. Hartinger, *Inorg. Chem*., 2010, **49**, 7953.
28 N. C. S. Selvam, R. T. Kumar, L. J. Kennedy and J. J. Vijaya, *Asian J. Chem*., 2011, **23**, 4328.
29 P. P. Corbi, F. A. Quintao, D. K. D. Ferraresi, W. R. Lustri, A. C. Amaral and A. C. Massabni, *J. Coord. Chem*., 2010, **63**, 1390.
30 W. Q. Zhang, A. C. Whitwood, I.J.S Fairlamb and J. M. Lynam, *Inorg. Chem*., 2010, **49**, 8941.
31 C. D. L. Saunders, L. E. Longobardi, N. Burford, M. D. Lumsden, U. Werner-Zwanziger, B. Chen and R. McDonald, *Inorg. Chem*., 2011, **50**, 2799.
32 F. Jalilehvand, V. Mah, B. O. Leung, J. Mink, G. M. Bernard and L. Hajba, *Inorg. Chem*., 2009, **48**, 4219.
33 F. Jalilehvand, B. O. Leung and V. Mah, *Inorg. Chem*., 2009, **48**, 5758.
34 M. Nazir and I. I. Naqvi, *J. Saudi Chem. Soc*., 2010, **14**, 101.
35 J. Han and Y. S. Chi, *J. Korean Soc. Appl. Biol. Chem*., 2010, **53**, 821.
36 Q. Y. Xu, Z. F. Liu, X. L. Hu, L. Kong and S. P. Liu, *Chem. J. Chinese Univ*., 2011, **32**, 1492.
37 H. D. Devi and N. R. Singh, *Asian J. Chem*., 2011, **23**, 3587.
38 V. G. Shtyrlin, Y. I. Zyavkina, E. M. Gilyazetdinov, M. S. Bukharov, A. A. Krutikov, R. R. Garipov, A. S. Mukhtarov and A. V. Zakharov, *Dalton Trans*., 2012, **41**, 1216.
39 J. Singh and O. D. Gupta, *Asian J. Chem*., 2010, **22**, 4311.
40 M. A. V. R. da Silva, L. M. N. B. F. Santos, A.C.P Faria and F.S.A Sistelo, *J. Therm. Anal. Calorim*., 2010, **100**, 475.
41 F. Jalilehvand, Z. Amini, K. Parmar and E. Y. Kang, *Dalton Trans*., 2011, **40**, 12771.
42 F. Gharib, F. Jaberi, A. Shamel, P. Farzaneh, S. Babashpour and M. Moazzami, *J. Chem. Eng. Data*, 2011, **56**, 3128.
43 A. M. Ryzhakov, M. S. Gruzdev, D. F. Pyreu, E. V. Kozlovskii and R. S. Kumeev, *Russ. J. Coord. Chem*., 2010, **36**, 565.
44 A. Heller, O. Ronitz, A. Barkleit, G. Bernhard and J. U. Ackermann, *Appl. Spectrosc*., 2010, **64**, 930.
45 M. Taha, B. S. Gupta and M. J. Lee, *J. Chem. Eng. Data*, 2011, **56**, 3541.
46 A. A. Mohamed, F. I. El-Dossoki and H. A. Gumaa, *J. Chem. Eng. Data*, 2010, **55**, 673.
47 A. Milicevic, G. Branica and N. Raos, *Molecules*, 2011, **16**, 1103.
48 A. Manceau and A. Matynia, *Geochim. Cosmochim. Acta*, 2010, **74**, 2556.

49 V. Sharma and K. D. Gupta, *Monatsh. Chem.*, 2011, **142**, 481.
50 P. C. do Nascimento, M. S. da Marques, D. Bohrer, L. M. de Carvalho and C. W. Carvalho, *Biol. Trace Elem. Res.*, 2011, **141**, 76.
51 S. Skounas, C. Methenitis, G. Pneumatikakis and M. Morcellet, *Bioinorg. Chem. Appl.*, 2010, **2010**, 1.
52 H. M. Abdel-Halim, A. S. Abu-Surrah and H. M. Baker, *Z. Anorg. Allg. Chem.*, 2010, **636**, 872.
53 K. Klacanová, P. Fodran and M. Rosenberg, *Monatsh. Chem.*, 2010, **141**, 823.
54 M. N. Kumara, K. Mantelingu, D. G. Bhadregowda and K. S. Rangappa, *Int. J. Chem. Kin.*, 2011, **43**, 599.
55 H. Pesonen, R. Aksela and K. Laasonen, *J. Phys. Chem. A*, 2010, **114**, 466.
56 R. Shankar, P. Kolandaivel and L. Senthilkumar, *J. Phys. Org. Chem.*, 2011, **24**, 553.
57 S. Mori, T. Endoh, Y. Yaguchi, Y. Shimizu, T. Kishi and T. K. Yanai, *Theor. Chem. Acc.*, 2011, **130**, 279.
58 A. M. Asaduzzaman, M. A. K. Khan, G. Schreckenbach and F. Y. Wang, *Inorg. Chem.*, 2010, **49**, 870.
59 J. F. Fan, L. J. He, J. Liu and M. Tang, *J. Mol. Model*, 2010, **16**, 1639.
60 A. Lamsabhi, O. Mo and M. Yanez, *Can. J. Chem./Rev. Can. Chim.*, 2010, **88**, 759.
61 T. E. Hofstetter, C. Howder, G. Berden, J. Oomens and P. B. Armentrout, *J. Phys. Chem. B*, 2011, **115**, 12648.
62 J. T. O'Brien, J. S. Prell, G. Berden, J. Oomens and E. R. Williams, *Int. J. Mass spectrom.*, 2010, **297**, 116.
63 M. Remko, D. Fitz and B. M. Rode, *Amino Acids*, 2010, **39**, 1309.
64 M. Citir, E.M.S Stennett, J. Oomens, J. D. Steill, M. T. Rodgers and P. B. Armentrout, *Int. J. Mass spectrom.*, 2010, **297**, 9.
65 M. B. Burt, S. G. A. Decker, C. G. Atkins, M. Rowsell, A. Peremans and T. D. Fridgen, *J. Phys. Chem. B*, 2011, **115**, 11506.
66 W. K. Mino Jr., J. Szczepanski, W. L. Pearson, D. H. Powell, R. C. Dunbar, J. R. Eyler and N. C. Polfer, *Int. J. Mass spectrom.*, 2010, **297**, 131.
67 L. Feketeová, M. W. Wong and R. A. J. O'Hair, *Eur. Phys. J. D*, 2010, **60**, 11.
68 M. K. Drayß, P. B. Armentrout, J. Oomens and M. Schäfer, *Int. J. Mass spectrom*, 2010, **297**, 18.
69 P. B. Armentrout, E. I. Armentrout, A. A. Clark, T. E. Cooper, E. M. S Stennett and D. R. Carl, *J. Phys. Chem. B*, 2010, **114**, 3927.
70 P. B. Armentrout, S. J. Ye, A. Gabriel and R. M. Moision, *J. Phys. Chem. B*, 2010, **114**, 3938.
71 V. N. Bowman, A. L. Heaton and P. B. Armentrout, *J. Phys. Chem. B*, 2010, **114**, 4107.
72 M. H. Khodabandeh, M. D. Davari, M. Zahedi and G. Ohanessian, *Int. J. Mass spectrom.*, 2010, **291**, 73.
73 S. Regupathy and M. S. Nair, *Spectrochim. Acta, Part A*, 2010, **75**, 656.
74 N. Pal, M. Kumar and G. Seth, *E-J. Chem.*, 2011, **8**, 1174.
75 G. G. Mohamed, H. F. A. El-Halim, M. M. I. El-Dessouky and W. H. Mahmoud, *J. Mol. Struct.*, 2011, **999**, 29.
76 M. A. Neelakantan, M. Sundaram and M. S. Nair, *Spectrochim. Acta, Part A*, 2011, **79**, 1693.
77 A. A. Shoukry, *J. Solution Chem.*, 2011, **40**, 1796.
78 A. Wojciechowska, M. Daszkiewicz, Z. Staszak, A. Trusz-Zdybek, A. Bienko and A. Ozarowski, *Inorg. Chem.*, 2011, **50**, 11532.
79 X. C. Wang, Q. K. Gao, K. J. Wang and J. R. Wang, *Chem. Pap.*, 2012, **66**, 188.

80 N. S. Youssef, E. A. El Zahany and M. M. Ali, *Phosphorus, Sulfur, Silicon Relat. Elem.*, 2010, **185**, 2171.
81 B. Jiang and L. Zhou, *Struct. Chem.*, 2011, **22**, 1353.
82 A. Adam, S. Verma and G. Seth, *E-J. Chem.*, 2011, **8**, S404.
83 F. Arjmand, M. Muddassir and R. H. Khan, *Eur. J. Med. Chem.*, 2010, **45**, 3549.
84 A. A. El-Sherif, M. M. Shoukry, R. M. El-Bahnasawy and D. M. Ahmed, *Cent. Eur. J. Chem.*, 2010, **8**, 919.
85 J. Zhou and G. Lu, *Spectrochim. Acta, Part A*, 2011, **78**, 1305.
86 R Bregier-Jarzebowska and L. Lomozik, *Polyhedron*, 2010, **29**, 3294.
87 S. Tashiro, Y. Ogura, S. Tsuboyama, K. Tsuboyama and M. Shionoya, *Inorg. Chem.*, 2011, **50**, 4.
88 R. Aydin and A. Yirikogullari, *J. Chem. Eng. Data*, 2010, **55**, 4794.
89 H. A. Azab, S. A. El-Korashy, Z. M. Anwar, B.H.M Hussein and G. M. Khairy, *J. Chem. Eng. Data*, 2010, **55**, 3130.
90 H. A. Azab, Z. M. Anwar and R. G. Ahmed, *J. Chem. Eng. Data*, 2010, **55**, 459.
91 A. R. M. Hyyrylainen, J. M. H. Pakarinen, E. Forró, F. Fülöp and P. Vainiotalo, *J. Mass Spectrom.*, 2010, **45**, 198.
92 R. Berkecz, A. R. M. Hyyrylainen, F. Fülöp, A. Péter, T. Janáky, P. Vainiotalo and J. M. H. Pakarinen, *J. Mass Spectrom.*, 2010, **45**, 1312.
93 T. Zimmermann and J. V. Burda, *Dalton Trans.*, 2010, **39**, 1295.
94 T. Zimmermann, Z. Chval and J. V. Burda, *J. Phys Chem. B*, 2009, **113**, 3139.
95 Y. Wu and L. Zhou, *Inorg. Chim. Acta*, 2010, **363**, 3274.
96 P. Grimminger and P. Klüfers, *Dalton Trans.*, 2010, **39**, 715.
97 C. Gabriel, M. Kaliva, J. Venetis, P. Baran, I. Rodriguez-Escudero, G. Voylatzis, M. Zervou and A. Salifoglu, *Inorg. Chem.*, 2009, **48**, 476.
98 S. Verma and G. Seth, *J. Indian Chem. Soc.*, 2011, **88**, 1435.
99 B. Das and J. B. Baruah, *Polyhedron*, 2011, **30**, 22.
100 S. Sharma, D. Dalwadi and M. Neog, *J. Serb. Chem. Soc.*, 2010, **75**, 75.
101 M. M. H. Khalil, E. H. Ismail, S. A. Azim and E. R. Souaya, *J. Therm. Anal. Calorim.*, 2010, **101**, 129.
102 N. S. Youssef, E. A. El Zahany and M. M. Ali, *Phosphorus, Sulfur Silicon Relat. Elem.*, 2010, **185**, 2171.
103 S. Sobel and G. Theophall, *Chem. Speciation Bioavailability*, 2010, **22**, 201.
104 S. Sobel, A. Haigney, M. Kim, D. Kim, G. Theophall, J. Nunez, D. Williams, B. Hickling and J. Sinacori, *Chem. Spec. Bioavailab.*, 2010, **22**, 109.
105 B. B. Tewari, *J. Chem. Eng. Data*, 2010, **55**, 1779.
106 A. A. Shoukry and W. M. Hosny, *Cent. Eur. J. Chem.*, 2012, **10**, 59.
107 A. Hammershoi, M. Schau-Magnussen, J. Bendix and A. Molgaard, *Acta Cryst.*, 2010, **66**, 319.
108 N. Tajdini and A. Moghimi, *Asian J. Chem.*, 2010, **22**, 5737.
109 M. M. Khalil and R. K. Mahmoud, *J. Chem. Eng. Data*, 2010, **55**, 789.
110 C. T. Miranda, S. Carvalho, R. T. Yamaki, E. B. Paniago, R. H. U. Borges and V. M. De Bellis, *Inorg. Chim. Acta*, 2010, **363**, 3776.
111 R. Sharma and M. Nagar, *J. Indian Chem. Soc.*, 2010, **87**, 1021.
112 A. Kufelnicki, S. V. Tomyn, R. V. Nedelkov, M. Haukka, J Jaciubek-Rosinska, M. Pajak, J. Jaszczak, M. Swiatek and I. O. Fritsky, *Inorg. Chim. Acta*, 2010, **363**, 2996.
113 R. Starosta, M. Florek, J. Krol, M. Puchalska and A. Kochel, *New J. Chem.*, 2010, **34**, 1441.
114 B. S. Murray, D. Parker, C. M. G. dos Santos and R. D. Peacock, *Eur. J. Inorg. Chem.*, 2010, **18**, 2663.

115 B. V. Rao, *J. Therm. Anal. Calorim.*, 2010, **100**, 577.
116 A. Kamecka, B. Kurzak, J. Jezierska and A. Wozna, *J. Solution Chem.*, 2011, **40**, 1041.
117 M. Menelaou, M. Daskalakis, A. Mateescu, C. P. Raptopoulou, A. Terzis, C. Mateescu, V. Tangoulis, T. Jakusch, T. Kiss and A. Salifoglou, *Polyhedron*, 2011, **30**, 427.
118 A. D. Naik, J. Beck, M. M. Dirtu, C. Bebrone, B. Tinant, K. Robeyns, J. Marchand-Brynaert and Y. Garcia, *Inorg. Chim. Acta*, 2011, **368**, 21.
119 A. N. Kozachkova, A. V. Dudko, N. V. Tsaryk, V. V. Trachevskii, A. B. Rozhenko and V. I. Pekhn'o, *Russ. J. Inorg. Chem.*, 2011, **56**, 1494.
120 S. Hashemian and M. H. Moslemine, *Asian J. Chem.*, 2010, **22**, 4641.
121 R. Sharma and M. Nagar, *Phosphorus, Sulfur, Silicon Relat. Elem.*, 2010, **185**, 1526.
122 R. He, H. H. Song, Z. Wei, J. J. Zhang and Y. Z. Gao, *J. Solid State Chem.*, 2010, **183**, 2021.
123 J. Zhang, L. Li, L. Ma, F. Zhang, Z. Zhang and S. Wang, *Eur. J. Med. Chem.*, 2011, **46**, 5711.
124 C. M. Yang, *Dalton Trans.*, 2011, **40**, 3008.
125 L. Tei, G. Gugliotta, M. Fekete, F. K. Kalman and M. Botta, *Dalton Trans.*, 2011, **40**, 2025.
126 D. Pellico, M. Gomez-Gallego, R. Escudero, P. Ramirez-Lopez, M. Olivan and M. A. Sierra, *Dalton Trans.*, 2011, **40**, 9145.
127 V. S. Sudhir, N. Y. P. Kumar and S. Chandrasekaran, *Tetrahedron*, 2010, **66**, 1327.
128 S. Quintal, M. C. Gimeno, A. Laguna and M. J. Calhorda, *J. Organomet. Chem.*, 2010, **695**, 558.
129 J. Tauchmann, I. Cisarova and P. Stepnicka, *Dalton Trans.*, 2011, **40**, 11748.
130 D. Siebler, C. Forster and K. Heinze, *Dalton Trans.*, 2011, **40**, 3558.
131 A. Monney, G. Venkatachalam and M. Albrecht, *Dalton Trans.*, 2011, **40**, 2716.
132 A. Breivogel, K. Hempel and K. Heinze, *Inorg. Chim. Acta*, 2011, **374**, 152.
133 K. Heinze, K. Hempel and A. Breivogel, *Z. Anorg. Allg. Chem.*, 2009, **635**, 2541.
134 M. S. Refat, S. Alghool and R. F. de Farias, *Synth. React. Inorg. Met.-Org. Nano-Metal Chem.*, 2010, **40**, 585.
135 M. Hamada-Kanazawa, M. Kouda, A. Odani, K. Matsuyama, K. Kanazawa, T. Hasegawa, M. Narahara and M. Miyake, *Biol. Pharm. Bull.*, 2010, **33**, 729.
136 P. R. Reddy, A. Shilpa, N. Raju and P. Raghavaiah, *J. Inorg. Biochem.*, 2011, **105**, 1603.
137 P. R. Reddy and A. Shilpa, *Polyhedron*, 2011, **30**, 565.
138 T. P. Ndifon, O. A. Moise, N. N. Julius, D. Y. Mbom, G. P. Awawou and D. N. Lynda, *Res. J. Chem. Environ.*, 2010, **14**, 50.
139 M. S. A. Begum, S. Saha, M. Nethaji and A. R. Chakravarty, *J. Inorg. Biochem.*, 2010, **104**, 477.
140 M. Nath and P. K. Saini, *Dalton Trans.*, 2011, **40**, 7077.
141 H. L. Singh, *Spectrochim. Acta, Part A*, 2010, **76**, 253.
142 A. Arbaoui, C. Redshaw, N. M. Sanchez-Ballester, M. R. J. Elsegood and D. L. Hughes, *Inorg. Chim. Acta*, 2011, **365**, 96.
143 D. J. Darensbourg and O. Karroonnirun, *Inorg. Chem.*, 2010, **49**, 2360.
144 M. Patil, R. Hunoor and K. Gudasi, *Eur. J. Med. Chem.*, 2010, **45**, 2981.
145 G. Qiu, Y. Li, W. Yang and Y. Zou, *J. Chem. Crystallogr.*, 2011, **41**, 898.
146 C. R. Bhattacharjee, P. Goswami and M. Sengupta, *J. Coord. Chem.*, 2010, **63**, 3969.

147 E. Zamanifar and F. Farzaneh, *Reac. Kinet. Mech. Cat.*, 2011, **104**, 197.
148 Z. L. Fang and Q. X. Nie, *J. Coord. Chem.*, 2010, **63**, 2328.
149 N. M. Hosny, N. Nawar, S. I. Mostafa and M. M. Mostafa, *J. Mol. Struct.*, 2011, **1001**, 62.
150 B. M. Draskovic, G. A. Bogdanovic, M. A. Neelakantan, A. C. Chamayou, S. Thalamuthu, Y. S. Avadhut, J.S.A.D Gunne, S. Banerjee and C. Janiak, *Cryst. Growth Des.*, 2010, **10**, 1665.
151 S. Handa, V. Gnanadesikan, S. Matsunaga and M. Shibasaki, *J. Amer. Chem. Soc.*, 2010, **132**, 4925.
152 D. Arish and M. S. Nair, *J. Coord. Chem.*, 2010, **63**, 1619.
153 R. K. Dubey and M. D. Pandey, *J. Indian Chem. Soc.*, 2009, **86**, 1262.
154 J. Albert, M. Crespo, J. Granell, J. Rodriguez, J. Zafrilla, T. Calvet, M. Font-Bardia and X. Solans, *Organometallics.*, 2010, **29**, 214.
155 E. A. Mikhalyova, S. V. Kolotilov, O. Cador, F. Pointillart, S. Golhen, L. Ouahab and V. V. Pavlishchuk, *Inorg. Chim. Acta*, 2010, **363**, 3453.
156 M. P. Yutkin, M. S. Zavakhina, D. G. Samsonenko, D. N. Dybtsev and V. P. Fedin, *Russ. Chem. Bull., Int. Ed.*, 2010, **59**, 733.
157 A. Igashira-Kamiyama and T. Konno, *Dalton Trans.*, 2011, **40**, 7249.
158 Z. Dong, L. Zhao, Z. Liang, P. Chen, Y. Yan, J. Li, J. Yu and R. Xu, *Dalton Trans.*, 2010, **39**, 5439.
159 G. Rousseau, O. Oms, A. Dolbecq, J. Marrot and P. Mialane, *Inorg. Chem.*, 2011, **50**, 7376.
160 G. L. Zhuang, W. X. Chen, H. X. Zhao, X. J. Kong, L. S. Long, R. B. Huang and L. S. Zheng, *Inorg. Chem.*, 2011, **50**, 3843.
161 M. Orfanoudaki, I. Tamiolakis, M. Siczek, T. Lis, G. S. Armatas, S. A. Pergantis and C. J. Milios, *Dalton Trans.*, 2011, **40**, 4793.
162 H. Y. Li, F. P. Huang and Y. M. Jiang, *Inorg. Chim. Acta*, 2011, **377**, 91.
163 D. L. Reger, J. J. Horger, A. Debreczeni and M. D. Smith, *Inorg. Chem.*, 2011, **50**, 10225.
164 D. L. Reger, J. J. Horger, M. D. Smith, G. J. Long and F. Grandjean, *Inorg. Chem.*, 2011, **50**, 686.
165 P. Thuery, *Inorg. Chem.*, 2011, **50**, 10558.
166 M. Bartholomä, B. Ploier, H. Cheung, W. Ouellette and J. Zubieta, *Inorg. Chim. Acta*, 2010, **363**, 1659.
167 K. A. Siddiqui, G. K. Mehrotra, S. S. Narvi and R. J. Butcher, *Inorg. Chem. Commun.*, 2011, **14**, 814.
168 R. Pandey, M. Yadav, P. Kumar, P. Z. Li, S. K. Singh, Q. Xu and D. S. Pandey, *Inorg. Chim. Acta*, 2011, **376**, 195.
169 J. S. Shen, G. J. Mao, Y. H. Zhou, Y. B. Jiang and H. W. Zhang, *Dalton Trans.*, 2010, **39**, 7054.
170 M. Karbarz, J. Romanski, K. Michniewicz, J. Jurczak and Z. Stojek, *Soft Matter*, 2010, **6**, 1336.
171 Z. R. Shen, J. G. Wang, P. C. Sun, D. T. Ding and T. H. Chen, *Dalton Trans.*, 2010, **39**, 6112.
172 Y. Xue, X. Tang, J. Huang, X. Zhang, J. Yu, Y. Zhang and S. Gui, *Colloids Surf., B*, 2011, **85**, 280.
173 C. Gao, G. Xu and S. Gou, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 6386.
174 C. S. Lim, J. Jankolovits, P. Zhao, J. W. Kampf and V. L. Pecoraro, *Inorg. Chem.*, 2011, **50**, 4832.
175 A. V. Pavlishchuk, S. V. Kolotilov, M. Zeller, O. V. Shvets, I. O. Fritsky, S. E. Lofland, A. W. Addison and A. D. Hunter, *Eur. J. Inorg. Chem.*, 2011, 4826.
176 S. Kobayashi, M. Watanabe and T. Chikuma, *Chem. Pharm. Bull.*, 2010, **58**, 620.

177 T. Ueno, S. Abe, T. Koshiyama, T. Ohki, T. Hikage and Y. Watanabe, *Chem. Eur. J.*, 2010, **16**, 2730.
178 J. Dzierzak, E. Bottinelli, G. Berlier, E. Gianotti, E. Stulz, R. M. Kowalczyk and R. Raja, *Chem. Commun.*, 2010, **46**, 2805.
179 D. Yamamoto, T. Nakanishi and T. Osaka, *Electrochim. Acta*, 2011, **56**, 9652.
180 B. B. Prasad, D. Kumar, R. Madhuri and M. P. Tiwari, *Biosens. Bioelectron.*, 2011, **28**, 117.
181 H. Xu, S. L. Gao, Q. W. Liu, D. Pan, L. H. Wang, S. Z. Ren, M. Ding, J. W. Chen and G. Liu, *Sensors.*, 2011, **11**, 10187.
182 G. C. Han, Y. Peng, Y. Q. Hao, Y. N. Liu and F. Zhou, *Anal. Chim. Acta*, 2010, **659**, 238.
183 C. R. Lohani, J. M. Kim and K-H Lee, *Tetrahedron*, 2011, **67**, 4130.
184 G. G. Huang, M. L. Cheng and J. Yang, *J. Chin. Chem. Soc.*, 2011, **58**, 435.
185 Y. Song, Z. He, H. Zhu, H. Hou and L. Wang, *Electrochim. Acta*, 2011, **58**, 757.
186 M. Tabeshnia, M. Rashvandavei, R. Amini and F. Pashaee, *J. Electroanal. Chem.*, 2010, **647**, 181.
187 L. G. Shaidarova, S. A. Ziganshina, A. V. Gedmina, I. A. Chelnokova and G. K. Budnikov, *J. Anal. Chem.*, 2011, **66**, 633.
188 S. Mohr, J. S. Hagele and M. G. Schmid, *Croat. Chem. Acta*, 2011, **84**, 343.
189 F. Kitagawa and K. Otsuka, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2011, **879**, 3078.
190 B. J. V. Verkuijl, A. K. Schoonen, A. J. Minnaard, J. G. de Vries and B. L. Feringa, *Eur. J. Org. Chem.*, 2010, **27**, 5197.
191 D. Jayaraman, M. Atanu, T. Khatija, S. B. Garima and P. R. Chebrolu, *J. Org. Chem.*, 2012, **77**, 371.
192 X. Yang, X. C. Liu, K. Shen, C. J. Zhu and Y. X. Cheng, *Org. Lett.*, 2011, **13**, 3510.
193 H. Braband, Y. Tooyama, T. Fox, R. Simms, J. Forbes, J. F. Valliant and R. Alberto, *Chem. Eur. J.*, 2011, **17**, 12967.
194 N. Ferlin, D. Grassi, C. Ojeda, M. J. L. Castro, E. Grand, A. F. Cirelli and J. Kovensky, *Carbohydr. Res.*, 2010, **345**, 598.
195 L. A. Kochergina and A. V. Emelyanov, *Russ. J. Phys. Chem. A*, 2011, **85**, 1742.
196 A. Dogan, A. D. Ozdel and E. Kilic, *Amino Acids*, 2009, **36**, 373.
197 F. Kiani, A. A. Rostami, F. Gharib, S. Sharifi and A. Bahadory, *J. Chem. Eng. Data*, 2011, **56**, 2830.
198 L. A. Kochergina and O. M. Drobilova, *Russ. J. Inorg. Chem.*, 2010, **55**, 764.
199 L. A. Byrne, M. J. Hynes, C. D. Conolly and R. A. Murphy, *J. Inorg. Biochem.*, 2011, **105**, 1656.
200 L. A. Mertens and E. M. Marzluff, *J. Phys. Chem. A*, 2011, **115**, 9180.
201 R. C. Dunbar, J. D. Steill, N. C. Polfer and J. Oomens, *J. Phys. Chem. A*, 2009, **113**, 10552.
202 F. Menges, C. Riehn and G. Niedner-Schatteburg, *Z. Phys. Chem.*, 2011, **225**, 595.
203 T. Jayasekharan and N. K. Sahoo, *Rapid Comm. Mass. Spectr.*, 2010, **24**, 3562.
204 M. Murariu, E. S. Dragan and G. Drochioiu, *Eur. J. Mass Spectr.*, 2010, **16**, 511.
205 H. M. Watson, J. B. Vincent and C. J. Cassady, *J. Mass Spectr.*, 2011, **46**, 1099.
206 J. Hong, Y. Yiao, J. F. Zhang and W. J. He, *Chinese J. Inorg. Chem.*, 2009, **25**, 2105.

207 H. Mattapalli, W. B. Monteith, C. S. Burns and A. S. Danell, *J. Am. Chem. Soc. Mass Spectr*., 2009, **20**, 2199.
208 E. Constantino, A. Rimola, M. Sodupe and L. Rodriguez-Santiago, *J. Phys. Chem. A*, 2009, **113**, 8883.
209 G. Micera and E. Garibba, *Eur. J. Inorg. Chem*., 2010, 4697.
210 J. Ali-Torres, L. Rodriguez-Santiago and M. Sodupe, *Phys. Chem., Chem. Phys*., 2011, **13**, 7852.
211 T. Shoeib and B. L. Sharp, *Inorg. Chim. Acta*, 2009, **362**, 1925.
212 E. Katsoulakou, G. S. Papaefstathiou, K. F. Konidaris, G. Pairas, C. Raptopoulou, P. Cordopatis and E. Manessi-Zoupa, *Polyhedron*, 2009, **28**, 3387.
213 R. C. Dunbar, J. D. Steill and J. Oomens, *J. Am. Chem. Soc*., 2011, **133**, 9376.
214 V. Jószai and I. Sóvágó, *Polyhedron*, 2011, **30**, 2114.
215 U. Rychlewska, B. Warzajtis, B. D. Glisic, M. D. Zivkovic, S. Rajkovic and M. I. Djuran, *Dalton Trans*., 2010, **39**, 8906.
216 K. Selmeczi, P. Gizzi, D. Champmartin, P. Rubini, E. Aubert, S. Dahoui and D. Henry, *Inorg. Chem*., 2010, **49**, 8222.
217 G. I. Grasso, F. Bellia, G. Arena, G. Vecchio and E. Rizzarelli, *Inorg. Chem*., 2011, **50**, 4917.
218 G. I. Grasso, G. Arena, F. Bellia, G. Maccarrone, M. Parrinello, A. Pietropaolo, G. Vecchio and E. Rizzarelli, *Chem.-Eur. J*., 2011, **17**, 9448.
219 V. Lanza, F. Bellia, R. D Agata, G. Grasso, E. Rizzarelli and G. Vecchio, *J. Inorg. Biochem*., 2011, **105**, 181.
220 V. Castelletto, G. Cheng, B. W. Greenland, I. W. Hamley and P. J. F. Harris, *Langmuir*, 2011, **27**, 2980.
221 J. Brasun, H. Czapor, A. Matera-Witkiewicz, A. Kotynia, A. Sochacka and M. Cebrat, *Dalton Trans*., 2010, **39**, 6518.
222 R. K. Singh, V. C. Srivastava and U. P. Singh, *Protein Peptide Letters*, 2011, **18**, 1280.
223 A. Jancsó, K. Selmeczi, P. Gizzi, N. V. Nagy, T. Gajda and B. Henry, *J. Inorg. Biochem*., 2011, **105**, 92.
224 H. Won, *Bull. Korean Chem. Soc*., 2011, **32**, 4021.
225 J. Brasun, M. Cebrat, L. Jaremko, M. Jaremko, G. Ilc, O. Gladysz and I. Zhukov, *Dalton Trans*., 2009, 4853.
226 J. Brasun, M. Cebrat, M. Jaremko, L. Jaremko, O. Gladysz and I. Zhukov, *J. Inorg. Biochem*., 2009, **103**, 1033.
227 E. Farkas, E. Csapó, P. Buglyó, C. A. Damante and G. Di Natale, *Inorg. Chim. Acta*, 2009, **362**, 753.
228 E. Lodyga-Chruscinska, E. Sochacka, D. Smuga, L. Chruscinski, G. Micera, D. Sanna, M. Turek and M. Gasiorkiewicz, *J. Inorg. Biochem*., 2010, **104**, 570.
229 E. Lodyga-Chruscinska, S. Oldziej, E. Sochacka, K. Korzycka, L. Chruscinski, G. Micera, D. Sanna, M. Turek and J. Pawlak, *J. Inorg. Biochem*., 2011, **105**, 1212.
230 S. Timári, C. Kállay, K. Ősz, I. Sóvágó and K. Várnagy, *Dalton Trans*., 2009, 1962.
231 C. Kállay, K. Várnagy, G. Malandrinos, N. Hadjiliadis, D. Sanna and I. Sóvágó, *Inorg. Chim. Acta*, 2009, **362**, 935–945.
232 A. Matera-Witkiewicz, J. Brasun, J. Swiatek-Kozlowska, A. Pratesi, M. Ginanneschi and L. Messori, *J. Inorg. Biochem*., 2009, **103**, 678.
233 A. Matera-Witkiewicz, J. Brasun and M. Cebrat, *Polyhedron*, 2010, **29**, 3052.
234 J. Brasun, A. Matera-Witkiewicz, E. Kamysz, W. Kamysz and S. Oldziej, *Polyhedron*, 2010, **29**, 1535.
235 M. A. Zoroddu, S. Medici and M. Peana, *J. Inorg. Biochem*., 2009, **103**, 1214.

236 H. Czapor, S. Bielinska, W. Kamysz, L. Szyrwiel and J. Brasun, *J. Inorg. Biochem.*, 2011, **105**, 297.
237 R. I. Stefureac, C. A. Madampage, O. Andrievskaia and J. S. Lee, *Biochem. Cell Biol.*, 2010, **88**, 347.
238 P. Davies, F. Marken, S. Salter and D. R. Brown, *Biochemistry*, 2009, **48**, 2610.
239 F. Guerrieri, V. Minocozzi, S. Morante, G. Rossi, S. Furlan and G. La Penna, *J. Biol. Inorg. Chem.*, 2009, **14**, 361.
240 R. P. Bonomo, G. Di Natale, E. Rizzarelli, G. Tabbi and L. I. Vagliasindi, *Dalton Trans.*, 2009, 2637.
241 L. I. Vagliasindi, G. Arena, R. P. Bonomo, G. Pappalardo and G. Tabbi, *Dalton Trans.*, 2011, **40**, 2441.
242 D. J. Stevens, E. D. Walter, A. Rodriguez, D. Draper, P. Davies, D. R. Brown and G. L. Millhauser, *PLOS Pathogens*, 2009, **5**, 1.
243 F. Stellato, A. Spevacek, O. Proux, V. Minocozzi, G. Millhauser and S. Morante, *Eur. Biophys. J.*, 2011, **40**, 1259.
244 D. Valensin, L. Szyrwiel, F. Camponeschi, M. Rowinsk-Zyrek, E. Molteni, E. Jankowska, A. Szymanska, E. Gaggelli, G. Valensin and H. Kozlowski, *Inorg. Chem.*, 2009, **48**, 7330.
245 C. Migliorini, D. Witkowska, D. Valensin, W. Kamysz and H. Kozlowski, *Dalton Trans.*, 2010, **39**, 8663.
246 G. Di Natale, K. Ősz, Z. Nagy, D. Sanna, G. Micera, G. Pappalardo, I. Sóvágó and E. Rizzarelli, *Inorg. Chem.*, 2009, **48**, 4239.
247 C. Kállay, I. Turi, S. Timári, Z. Nagy, D. Sanna, G. Pappalardo, P. de Bona, E. Rizzarelli and I. Sóvágó, *Monatsh. Chem.*, 2011, **142**, 411.
248 M. Remelli, D. Valensin, D. Bacco, E. Gralka, R. Guerrini, C. Migliorini and H. Kozlowski, *New J. Chem.*, 2009, **33**, 2300.
249 L. Rivillas-Acevedo, R. Grande-Aztatzi, I. Lomeli, J. E. Garcia, E. Barrios, S. Teloxa, A. Vela and L. Quintanar, *Inorg. Chem.*, 2011, **50**, 1956.
250 Y. L. Wang, L. Feng, B. B. Zhang, X. S. Wang, C. Huang, Y. M. Li and W. H. Du, *Inorg. Chem.*, 2011, **50**, 4340.
251 I. Turi, C. Kállay, D. Szikszai, G. Pappalardo, G. Di Natale, P. De Bona, E. Rizzarelli and I. Sóvágó, *J. Inorg. Biochem.*, 2010, **104**, 885.
252 D. Valensin, K. Gajda, E. Gralka, G. Valensin, W. Kamysz and H. Kozlowski, *J. Inorg. Biochem.*, 2010, **104**, 71.
253 D. La Mendola, R. P. Bonomo, S. Caminati, G. Di Natale, S. S. Emmi, Ö. Hannson, G. Maccarrone, G. Pappalardo, A. Pietropaolo and E. Rizzerelli, *J. Inorg. Biochem.*, 2009, **103**, 195.
254 E. Gralka, D. Valensin, K. Gajda, D. Bacco, L. Szyrwiel, M. Remelli, G. Valensin, W. Kamasz, W. Barnaska-Rybak and H. Kozlowski, *Mol. BioSyst.*, 2009, **5**, 497.
255 D. La Mendola, A. Magri, Ö Hansson, R. P. Bonomo and E. Rizzarelli, *J. Inorg. Biochem.*, 2009, **103**, 758.
256 D. La Mendola, A. Magri, T. Campagna, M. A. Campitiello, L. Raiola, C. Isernia, Ö. Hansson, R. P. Bonomo and E. Rizzarelli, *Chem-Eur. J.*, 2010, **16**, 6212.
257 C. A. Damante, K. Ősz, Z. Nagy, G. Pappalardo, G. Grasso, G. Impellizzeri, E. Rizzarelli and I. Sóvágó, *Inorg. Chem.*, 2009, **48**, 10405.
258 S. C. Drew, C. J. Noble, C. L. Masters, G. R. Hanson and K. J. Barnham, *J. Am. Chem. Soc.*, 2009, **131**, 1195.
259 C. A. Damante, K. Ősz, Z. Nagy, G. Grasso, G. Pappalardo, E. Rizzarelli and I. Sóvágó, *Inorg. Chem.*, 2011, **50**, 5342.

260 É. Józsa, K. Ősz, C. Kállay, P. de Bona, C. A. Damante, G. Pappalardo, E. Rizzarelli and I. Sóvágó, *Dalton Trans*., 2010, **39**, 7046.
261 J. Ali-Torres, J. D. Marechal, L. Rodriguez-Santiago and M. Sodupe, *J. Am. Chem. Soc*., 2011, **133**, 15008.
262 J. Ali-Torres, L. Rodriguez-Santiago, M. Sodupe and A. Rauk, *J. Phys. Chem. A*, 2011, **115**, 12523.
263 D. Valensin, C. Migliorini, G. Valensin, E. Gaggelli, G. La Penna, H. Kozlowski, C. Gabbiani and L. Messori, *Inorg. Chem*., 2011, **50**, 6865.
264 G. Droichioiu, *Eur. J. Mass Spectr*., 2009, **15**, 651.
265 G. Droichioiu, M. Manea, M. Dragusanu, M. Murariu, E. S. Dragan, B. A. Petre, G. Mezo and M. Przybylski, *Biophys. Chem*., 2009, **144**, 9.
266 S. A. Kozin, Y. V. Mezentsev, A. A. Kulikova, M. I. Indeykina, A. V. Golovin, A. S. Ivanov, P. O. Tsvetkov and A. A. Makarov, *Mol. Biosyst*., 2011, **7**, 1053.
267 S. Bolognin, L. Messori, D. Drago, C. Gabbiani, L. Cendron and P. Zatta, *Int. J. Biochem. Cell Biol*., 2011, **43**, 877.
268 A. Granzotto, S. Bolognin, J. Scancar, R. Milacic and P. Zatta, *Monat. Chem*., 2011, **142**, 421.
269 M. Murariu, E. S. Dragan, A. Adochitei, G. Zbancioc and G. Droichioiu, *J. Pept. Sci*., 2011, **17**, 512.
270 Y. Miller, B. Y. Ma and R. Nussinov, *Proc. Nat. Acad. Sci. U.S.A*, 2010, **107**, 9490.
271 P. Faller, *ChemBioChem*, 2009, **10**, 2837.
272 A. Olofsson, M. Lindhagen-Persson, M. Vestling, A. E. Sauer-Eriksson and A. Ohman, *FEBS. Journal*, 2009, **276**, 4051.
273 C. J. Sarell, C. D. Syme, S. E. J. Rigby and J. H. Viles, *Biochem*., 2009, **48**, 4388.
274 H. A. Feaga, R. C. Maduka, M. N. Foster and V. A. Szalai, *Inorg. Chem*., 2011, **50**, 1614.
275 J. Shearer and V. A. Szalai, *J. Am. Chem. Soc*., 2008, **130**, 17826.
276 C. Hureau and P. Faller, *Biochimie*, 2009, **91**, 1212.
277 D. L. Jiang, X. J. Li, L. Liu, G. B. Yagnik and F. M. Zhou, *J. Phys. Chem. B*, 2010, **114**, 4896.
278 M. Brzyska, K. Trzesniewska, A. Wieckowska, A. Szczepankiewicz and D. Elbaum, *ChemBioChem*., 2009, **10**, 1045.
279 G. F. Z. da Silva, V. Lykourinou, A. Angerhofer and L. J. Ming, *Biochim. Biophys Acta, Mol. Basis. Dis*., 2009, **1792**, 49.
280 D. Pramanik, C. Ghosh and S. G. Dey, *J. Am. Chem. Soc*., 2011, **133**, 15545.
281 T. Miura, M. Yoda, C. Tsutsumi, K. Murayama and H. Takeuchi, *Yakugaku Zasshi-J. Pharm. Soc. Jap*., 2010, **130**, 495.
282 M. Suwalsky, S. Bolognin and P. Zatta, *J. Alz. Dis*, 2009, **17**, 81.
283 A. M. Mancino, S. S. Hindo, A. Kochi and M. H. Lim, *Inorg. Chem*., 2009, **48**, 9596.
284 A. Rimola, J. Ali-Torres, C. Rodriguez-Rodriguez, J. Poater, E. Matito, M. Sola and M. Sodupe, *J. Phys. Chem. A*, 2011, **115**, 12659.
285 M. Bajda, N. Guzior, M. Ignasik and B. Malawska, *Curr. Med. Chem*., 2011, **18**, 4949.
286 G. Grasso, A. Pietropaolo, G. Spoto, G. Pappalardo, G. R. Tundo, C. Ciaccio, M. Coletta and E. Rizzarelli, *Chem-Eur. J*., 2011, **17**, 2752.
287 B. Y. W. Man, H. M. Chan, C. H. Leung, D. S. H. Chan, L. P. Bai, Z. H. Jiang, H. W. Li and D. L. Ma, *Chem. Sci*., 2011, **2**, 917.
288 D. Valensin, F. Camponeschi, M. Luczkowski, M. C. Baratto, M. Remelli, G. Valensin and H. Kozlowski, *Metallomics*, 2011, **3**, 292.

289 A. Binolfi, E.E. Rodriguez, D. Valensin, N. D'Amelio, E. Ippoliti, G. Obal, R. Duran, A. Magistrato, O. Pritsch, M. Zweckstetter, G. Valensin, P. Carloni, L. Quintanar, C. Griesinger and C.O. Fernandez, *Inorg. Chem.*, 2010, **49**, 10668.
290 Y. Lu, M. Prudent, B. Fauvet, H. A. Lashuel and H. H. Girault, *ACS Chem. Neurosci*, 2011, **2**, 667.
291 L. Perrone, E. Mothes, M. Vignes, A. Mockel, C. Figueroa, M. C. Miquel, M. L. Maddelein and P. Faller, *ChemBioChem*, 2010, **11**, 110.
292 M. Rozga and W. Bal, *Chem. Res. Toxicol.*, 2010, **23**, 298.
293 M. Sokolowska and K. Pawlas, W. Bal, *Bioinorg. Chem. Appl.*, 2010, DOI: 10.1155/2010/725153.]
294 G. Navarra, A. Tinti, M. Leone, V. Militello and A. Torreggiani, *J. Inorg. Biochem.*, 2009, **103**, 1729.
295 A.M.P.C. Nunes, K. Zavitsanos, R. Del Conte, G. Malandrinos and N. Hadjiliadis, *Dalton Trans.*, 2009, 1904.
296 A.M.P.C. Nunes, K. Zavitsanos, G. Malandrinos and N. Hadjiliadis, *Dalton Trans.*, 2010, **39**, 4369.
297 A. M. P. C. Nunes, K. Zavitsanos, R. Del Conte, G. Malandrinos and N. Hadjiliadis, *Inorg. Chem.*, 2010, **49**, 5658.
298 K. Zavitsanos, A.M.P.C. Nunes, G. Malandrinos and N. Hadjiliadis, *J. Inorg. Biochem.*, 2011, **105**, 102.
299 E. Kurowska, A. Bonna, G. Goch and W. Bal, *J. Inorg. Biochem.*, 2011, **105**, 1220.
300 W. M. Tay, A. I. Hanafy, A. Angerhofer and L-J. Ming, *Bioorg. Med. Chem. Letters*, 2009, **19**, 6709.
301 E. Porciatti, M. Milenkovic, E. Gaggelli, G. Valensin, H. Kozlowski, W. Kamysz and D. Valensin, *Inorg. Chem.*, 2010, **49**, 8690.
302 D. Witkowska, S. Bielinska, W. Kamysz and H. Kozlowski, *J. Inorg. Biochem.*, 2011, **105**, 208.
303 A. Jancsó, A. Kolozsi, B. Gyurcsik, N. V. Nagy and T. Gajda, *J. Inorg. Biochem.*, 2009, **103**, 1634.
304 A Kolozsi, A. Jancsó, N. V. Nagy and T. Gajda, *J. Inorg. Biochem.*, 2009, **103**, 940.
305 T. Kowalik-Jankowska, L. Biega, M. Kuczer and D. Konopinska, *J. Inorg. Biochem.*, 2009, **103**, 135.
306 T. Kowalik-Jankowska, J. Jezierska and M. Kuczer, *Dalton Trans.*, 2010, **39**, 4117.
307 M. Pietruszka, E. Jankowska, T. Kowalik-Jankowska, Z. Szewczuk and M. Smuzynska, *Inorg. Chem.*, 2011, **50**, 7489.
308 T. Kowalik-Jankowska, E. Jankowska, Z. Szewczuk and F. Kasprzykowski, *J. Inorg. Biochem.*, 2010, **104**, 831.
309 D. La Mendola, A. Magri, L. I. Vagliasindi, O. Hansson, R. P. Bonomo and E. Rizzarelli, *Dalton Trans.*, 2010, **39**, 10678.
310 A. Travaglia, G. Arena, R. Fattoruso, C. Isernia, D. La Mendola, G. Malgieri, V. G. Nicoletti and E. Rizzarelli, *Chem-Eur. J.*, 2011, **17**, 3726.
311 C. Bure, R. Maget, A. F. Delmas, C. Pichon and P. Midoux, *J. Mass Spectr.*, 2009, **44**, 81.
312 C. Bure, C. Pichon and P. Midoux, *J. Mass Spectr.*, 2009, **44**, 1163.
313 M. Murariu, R. V. Gradinaru, M. Mihai, S. Jurcoane and G. Drochioiu, *Rom. Biotech. Letters*, 2011, **16**, 6242.
314 M. A. Zoroddu, M. Peana, S. Medici and R. Anedda, *Dalton Trans.*, 2009, 5523.
315 J-Y. Kim, M-J. Kim and H-T. Kim, *Bull. Korean Chem. Soc.*, 2010, **31**, 1377.
316 C. Tselepis, S. J. Ford, A. T. McKie, W. Vogel, H. Zoller, R. J. Simpson, J. Diaz Castro, T. H. Iqbal and D. G. Ward, *Biochem. J.*, 2010, **427**, 289.

317 D. Wesenberg, G. J. Krauss and D. Schaumloffel, *Int. J. Mass Spectr.*, 2011, **307**, 46.
318 Z. X. Wu, F. A. Fernandez-Lima and D. H. Russel, *J. Am. Chem. Soc., Mass Spectr.*, 2010, **21**, 522.
319 T. F. Wang, H. J. Andreazza, T. L. Pukala, P. J. Sherman, A. N. Calabrese and J. H. Bowie, *Rapid Comm., Mass Spectr.*, 2011, **25**, 1209.
320 F. Xie, D. E. K. Sutherland, M. J. Stillman and M. Y. Ogawa, *J. Inorg. Biochem.*, 2010, **104**, 261.
321 H. J. Jeong and H. T. Kim, *Eur. J. Mass Spectr.*, 2009, **15**, 67.
322 L. Joly, R. Antoine, A. R. Allouche, M. Broyer, J. Lemoine and P. Dugourd, *J. Phys. Chem. A*, 2009, **113**, 6607.
323 L. Joly, R. Antoine, F. Albrieux, R. Ballivian, M. Broyer, F. Chirot, J. Lemoine, P. Dugourd, C. Greco, R. Mitric and V. Bonacic-Koutecky, *J. Phys. Chem. B*, 2009, **113**, 11293.
324 M. Ngu-Schwemlein, J. K. Merle, P. Healy, S. Schwemlein and S. Rhodes, *Thermochim. Acta*, 2009, **496**, 129.
325 M. Rowynska-Zyrek, D. Witkowska, S. Bielinska, W. Kamysz and H. Kozlowski, *Dalton Trans.*, 2011, **40**, 5604.
326 K. Krzywoszynska, M. Rowynska-Zyrek, D. Witkowska, S. Potocki, M. Luckowski and H. Kozlowski, *Dalton Trans.*, 2011, **40**, 10434.
327 S. Potocki, M. Rowynska-Zyrek, D. Valensin, K. Krzywoszynska, D. Witkowska, M. Luczkowski and H. Kozlowski, *Inorg. Chem.*, 2011, **50**, 6135.
328 M. Rowynska-Zyrek, D. Valensin, L. Szyrwiel, Z. Grzonka and H. Kozlowski, *Dalton Trans.*, 2009, 9131.
329 M. Rowynska-Zyrek, D. Witkowska, D. Valensin, W. Kamysz and H. Kozlowski, *Dalton Trans.*, 2010, **39**, 5814.
330 M. E. Krause, A. M. Glass, T. A. Jackson and J. S. Laurence, *Inorg. Chem.*, 2010, **49**, 362.
331 M. E. Krause, A. M. Glass, T. A. Jackson and J. S. Laurence, *Inorg. Chem.*, 2011, **50**, 2479.
332 O. Seneque and J. M. Latour, *J. Am. Chem. Soc.*, 2010, **50**, 17760.
333 O. Iranzo, T. Jakusch, K. H. Lee, L. Hemmingsen and V. L. Pecoraro, *Chem-Eur. J.*, 2009, **15**, 3761.
334 O. Delalande, H. Desvaux, E. Godat, A. Valleix, C. Junot, J. Labarre and Y. Boulard, *FEBS Journal*, 2010, **277**, 5086.
335 M. Belcastro, T. Marino, N. Russo and M. Toscano, *J. Inorg. Biochem.*, 2009, **103**, 50.
336 K. P. Neupane and V. L. Pecoraro, *J. Inorg. Biochem.*, 2011, **105**, 1030.
337 A. Krezel, J. Wojcik, M. Maciejczyk and W. Bal, *Inorg. Chem.*, 2011, **50**, 72.
338 R. Gusmao, C. Arino, J. M. Diaz-Cruz and M. Esteban, *Analyst*, 2010, **135**, 86.
339 R. Gusmao, S. Cavanillas, C. Arino, J. M. Diaz-Cruz and M. Esteban, *Anal. Chem.*, 2010, **82**, 9006.
340 R. Pal and J. P. N. Rai, *Appl. Biochem. Biotech.*, 2010, **160**, 945.
341 E. Lodyga-Chruscinska, S. Oldziej, D. Sanna, G. Micera, L. Chruscinski, K. Kaczmerak, R. J. Nachman, J. Zabrocki and A. Sykula, *Polyhedron*, 2009, **28**, 485.
342 C. Kállay, Á. Dávid, S. Timári, E. M. Nagy, D. Sanna, E. Garribba, G. Micera, P. de Bona, G. Pappalardo, E. Rizzarelli and I. Sóvágó, *Dalton Trans.*, 2011, **40**, 9711.
343 K. Stokowa, W. Szczepanik, N. Gaggelli, G. Valensin and M. Jezowska-Bojczuk, *J. Inorg. Biochem.*, 2012, **106**, 111.
344 S. Bezer, S. Rapireddy, Y. A. Skorik, D. H. Ly and C. Achim, *Inorg. Chem.*, 2011, **50**, 11920.

345 G. N. Roviello, D. Musumeci, E. M. Bucci and C. Pedone, *Mol. Biosyst.*, 2011, **7**, 1073.
346 P. Comba, N. Dovalil, G. R. Hanson and G. Linti, *Inorg. Chem.*, 2011, **50**, 5165.
347 S. Blasco, M. I. Burguete, M. P. Clares, E. Garcia-Espagna, J. Escorihuela and S. V. Luis, *Inorg. Chem.*, 2010, **49**, 7841.
348 R. S. Joseyphus and M. S. Nair, *Arabian J. Chem.*, 2010, **3**, 195.
349 J. S. Prell, T. G. Flick, J. Oomens, G. Berden and E. R. Williams, *J. Phys. Chem. A*, 2010, **114**, 854.
350 F. Cisnetti, C. Gateau, C. Lebrun and P. Delangle, *Chem-Eur. J.*, 2009, **15**, 7456.
351 S. Rajkovic, M. D. Zivkovic, C. Kállay, I. Sóvágó and M. I. Djuran, *Dalton Trans.*, 2009, 8370.
352 Z. D. Petrovic, V. P. Petrovic, D. Simijonovic and S. Markovic, *Dalton Trans.*, 2011, **40**, 9284.
353 V. Yeguas, P. Campomanes, R. Lopez, N. Diaz and S. Suarez, *J. Phys. Chem. A*, 2010, **114**, 8525.
354 A. Kumar, X. X. Zhu, K. Walsh and R. Prabhakar, *Inorg. Chem.*, 2010, **49**, 38.
355 F. Miskevich, A. Davis, P. Leeprapaiwong, V. Giganti, N. M. Kostic and L. A. Angel, *J. Inorg. Biochem.*, 2011, **105**, 675.
356 A. Krezel, E. Kopera, A. M. Protas, J. Poznanski, A. Wyslouch-Cieszynska and W. Bal, *J. Am. Chem. Soc.*, 2010, **132**, 3355.
357 E. Kopera, A. Krezel, A. M. Protas, A. Belczyk, A. Bonna, A. Wyslouch-Cieszynska, J. Poznanski and W. Bal, *Inorg. Chem.*, 2010, **49**, 6636.
358 R. Mhidia and O. Melnyk, *J. Pept. Sci.*, 2010, **16**, 141.
359 M. M. Ibrahim, A. El-Motaleb and M. Ramadan, *J. Incl. Phen. Macrocyc. Chem.*, 2010, **68**, 287.
360 P. H. Ho, K Stroobants and T. N. Parac-Vogt, *Inorg. Chem.*, 2011, **50**, 12025.
361 E. Dumont, C. Michel and P. Sautet, *ChemPhysChem.*, 2011, **12**, 2596.
362 F. Porcelli, C. Olivieri, L. R. Masterson, Y. Wang and G. Veglia, *J. Biol. Inorg. Chem.*, 2011, **16**, 1197.
363 A. Reis, C. Fonseca, E. Maciel, P. Domingues and M.R.M. Domingues, *Anal. Bioanal. Chem.*, 2011, DPI 10.1007/s00216-011-4688-1
364 A. Barman, W. Taves and R. Prabhakar, *J. Comp. Chem.*, 2009, **30**, 1405.
365 R. K. Singh, S. Prasad and U. P. Singh, *Protein Pept. Letters*, 2010, **17**, 260.
366 T. Kowalik-Jankowska, E. Jankowska and F. Kasprzykowski, *Inorg. Chem.*, 2010, **49**, 2182.
367 C. Hureau, H. Eury, R. Guillot, C. Bijani, S. Sayen, P. L. Solari, E. Guillon, P. Faller and P. Dorlet, *Chem-Eur. J.*, 2011, **17**, 10151.
368 L. C. C. do Lago, A. C. Matias, C. S. Nomura and G. Cerchiaro, *J. Inorg. Biochem.*, 2011, **105**, 189.
369 P. Wang, G. Ohanessian and C. Wesdemiotis, *Eur. J. Mass Spectr.*, 2009, **15**, 325.
370 S. Y. Kwak, H. R. Choi, K. C. Park and Y. S. Lee, *J. Pept. Sci.*, 2011, **17**, 791.
371 A. Hashim, L. Wang, K. Juneja, Y. Ye, Y. F. Zhao and L. J. Ming, *Bioorg. Med. Chem. Letters*, 2011, **21**, 6430.
372 L. J. Ming, *J. Chin. Chem. Soc.*, 2010, **57**, 285.
373 T. Chan, A. Chow, D. W. T. Tang, Q. Li, X. Y. Wang, I. R. Brown and K. Kerman, *J. Electroanal Chem.*, 2010, **648**, 151.
374 C. S. Wang, L. Liu, L. Zhang, Y. Peng and F. M. Zhou, *Biochemistry*, 2010, **49**, 8134.
375 S. Timári, R. Cerea and K. Várnagy, *J. Inorg. Biochem.*, 2011, **105**, 1009.
376 M. F. Tweedle, *Accounts Chem Res.*, 2009, **42**, 958.

377 C. A. Puckett and J. K. Barton, *Bioorg. Med. Chem.*, 2010, **18**, 3564.
378 D. Valensin, P. Anzini, E. Gaggelli, N. Gaggelli, G. Tamasi, R. Cini, C. Gabbiani, E. Michelucci, L. Messori, H. Kozlowski and G. Valensin, *Inorg. Chem.*, 2010, **49**, 4720.
379 A. Kumar, L. Moody, J. F. Olaivar, N. A. Lewis, R. L. Khade, A. A. Holder, Y. Zhang and V. Rangachari, *ACS Chem. Neurosci.*, 2010, **1**, 691.
380 J. Lemke and N. Metzler-Nolte, *J. Organomet. Chem.*, 2011, **696**, 1018.
381 M. B. U. Surfraz, R. King, S. J. Mather, S. Biagini and P. J. Blower, *J, Inorg. Biochem.*, 2009, **103**, 971.
382 K. Zelenka, L. Borsig and R. Alberto, *Org. Biomol. Chem.*, 2011, **9**, 1071.
383 K. Zelenka, L. Borsig and R. Alberto, *Bioconjugate Chem.*, 2011, **22**, 958.
384 G. Gasser, A. M. Sosniak and N. Metzler-Nolte, *Dalton Trans.*, 2011, **40**, 7061.
385 M. F. Yu, J. R. Price, P. Jensen, C. J. Lowitt, T. Shelper, S. Duffy, L. C. Windus, V. M. Avery, P. J. Rutledge and M. H. Todd, *Inorg. Chem.*, 2011, **50**, 12823.
386 F. Schmidt, I. C. Rosnizeck, M. Spoerner, H. R. Kalbitzer and B. Konig, *Inorg. Chim. Acta*, 2011, **365**, 38.
387 T. Saio, K. Ogura, M. Yokochi, Y. Kobashigawa and F. Inagaki, *J. Biomol. NMR*, 2009, **44**, 157.
388 S. Ozcubukcu, K. Mandal, S. Wegner, M. P. Jensen and C. He, *Inorg. Chem.*, 2011, **50**, 7937.
389 C. W. A. Ende, H. Y. Meng, M. Ye, A. K. Pandey and N. J. Zondlo, *ChemBioChem*, 2010, **11**, 1738.
390 M. T. Ma, H. N. Hoang, C. C. G. Scully, T. G. Appleton and D. P. Fairlie, *J. Am. Chem. Soc.*, 2009, **131**, 4505.
391 B. Wang, H. W. Li, Y. Gao, H. Y. Zhang and Y. Q. Wu, *J. Fluorescence*, 2011, **21**, 1921.
392 M. Taki, F. Asahi, T. Hirayama and Y. Yamamoto, *Bull. Chem. Soc. Jpn.*, 2011, **84**, 386.
393 S. S. Sam, J. N. J. N. Chazalviel, A.C.A.C. Gouget-Laemmel, F. F. Ozanam, A. A. Etcheberry and N. E. N. Gabouze, *Nanoscale Res. Letters*, 2011, **6**, DOI: 10.1186/1556-276X-6-412.
394 P. R. Reddy and A. Shilpa, *Ind. J. Chem.*, 2010, **49**, 1003.
395 S. Bradford, Y. Kawarasaki and J. A. Cowan, *J. Inorg. Biochem.*, 2009, **103**, 871.
396 G. Malandrinos, A. M. Nunes, K. Zavitsanos and N. Hadjiliadis, *Pure Appl. Chem.*, 2011, **83**, 1751.
397 A. Chapkanov, Y. Miteva, T. Kolev, M. Spiteller and B. Koleva, *Protein Pept. Letters*, 2010, **17**, 228.

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

Jason Crain
DOI: 10.1039/9781849734677-00119

1 Introduction

The space in which research questions in biophysics directly intersects medicine is currently quite small but it is growing and is beginning to show promise of real downstream clinical benefit. The union is not a particularly easy one: Typically, the focus of molecular biology and medicine is to reveal molecular properties and their links to biological function: Here, chemical detail is all-important and high-resolution structural information in complex systems is essential.[1,2] In the evolving field of soft-condensed matter physics,[3] many working models for complex processes are based on coarse-grained or highly reduced systems in which chemical detail is eliminated in favour of much more generic or vastly simplified models which expose the driving forces and competing interactions that may be relevant in biological phenomena but are otherwise stripped of the chemical specificity that makes them particular to ''real'' systems with biological identity. In this way, for example, the behaviour of colloidal particles having carefully tuned surface properties but no internal structure can be used to model effectively certain aspects of protein crystallization. Similarly, ultra-simple lower alcohols have been shown to be a ''minimal'' model system for biological self-assembly and hyrophobic hydration.[4–6] Here, primitive hydrophobically-driven vesicles form in aqueous solutions and account for anomalous thermodynamic properties of the mixtures. While both these examples represent important advances in understanding, they capture only structural and thermodynamic aspects of the biological systems. As a result, their utility as models is limited because they retain no functional capacity of the real biological systems.

One of the most promising emerging strategies in the study of processes with inherently high levels of structural complexity, multiple degrees of freedom and competing interactions makes use of so-called model systems which represent extreme approximations of ''real'' macromolecular materials. Chosen carefully, and with appropriate combinations of computational and experimental input, coordinated investigations on model systems can be very powerful leading to insight into much more complex systems. The principal aim is to approach a systematic exploration of hydration structure, molecular flexibility and the effect of ions (Hofmeister effects) using a coordinated computational and experimental methodology

University of Edinburgh, Mayfield Road, EH9 3JZ, Scotland, UK & National Physical Laboratory, Hampton Road, Teddington, TW11 OLW. E-mail: j.crain@ed.ac.uk

focused on model systems of extremely reduced complexity. In general, a suitable model system must satisfy the following criteria:

- It must be simpler in terms of size and molecular complexity than the real biomolecules.
- It must retain some of the essential physics and basic interactions present in the real system. Detailed and unambiguous experimental and computational data must be obtainable so that the results of both can be compared directly. In the specific case of the protein folding problem and self-organization phenomena in biomolecules, there are potentially many views as to what level of abstraction relative to a real biomolecule is ''useful'' for a particular purpose.

Ultimately, there must be some compromise found between the degree of ''realism'' of the system and the level of detail at which the system is described. For the purpose of this article where the fundamental questions are general rather than specific to particular biomolecules, we will adopt an extreme simplification of real biological systems. The benefit of this approach is that very sophisticated computational and experimental tools can be deployed together in a coordinated program and the effects of a wide range of thermodynamic and environmental parameters can be investigated in detail. Within the realm of these highly idealized systems the folding of proteins in aqueous solutions can be reduced to a few fundamental issues:

Peptides therefore hold a special or maybe even unique status in this regard: They are a special class of biomolecule which are small and simple enough to serve as model systems in their own right. They can be simulated at atomistic detail, measured by a range of high-resolution methods, and can be synthesized artificially at high purity. These are arguably the required properties of any worthy model system. But in the case of peptides there is something else: Even the very small ones can retain biological functionality in complexes with membranes and other molecules. They are therefore a potentially fruitful ground for developing a serious effort in ''translational'' biophysics wherein physical science methods and thinking are retained but the biological and medical context of the research may be a genuine, primary driver in problem selection.

What we seek to address here is the question of utility of peptides as useful and potentially rich model systems capable of exposing a range of different physical properties relevant to self-organization over different lengthscales including those much larger than the peptide dimensions. Of particular interest in that regard are the use of peptides as potential model systems for tertiary interactions in proteins which may be widely

2 Thermodynamic basics of liquid mixtures and hydrophobicity

Ultimately, folding and self-assembly of complex contacts in peptides is driven in part of by the equilibrium thermodynamics of solvation and mixing. So what happens when we mix two liquids together? Are the liquids miscible (forming a homogeneous mixture) or immiscible? To answer these questions, it is important to introduce the thermodynamics of liquid mixtures starting from the Gibbs free energy $G = H - TS$ where the enthalpy $H = U + PV$ (Fig. 1).[7,8]

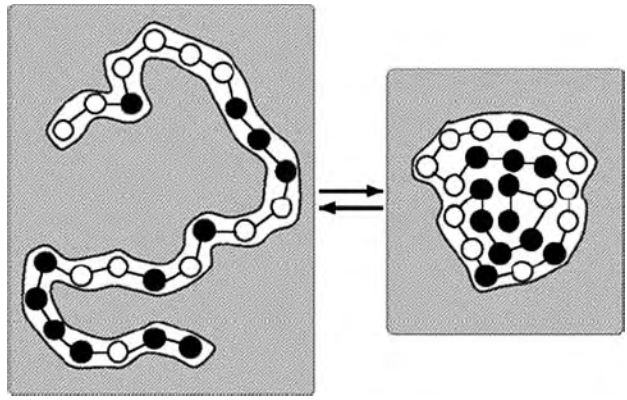


Fig. 1 Folding of a peptide chain to sequester hydrophobic segments from water as an example of strongly non-ideal mixing.

The change in the Gibbs free energy of the system that occurs during a reaction is therefore equal to the change in the enthalpy of the system minus the change in the product of the temperature times the entropy of the system: This change measures the balance between the two driving forces that determine whether a reaction is spontaneous. The H and S terms have different sign conventions: For an ideal solution, when two liquids A and B are mixed the Gibbs free energy of mixing is given by

$$\Delta G_{mix} = nRT(x_A \ln x_A + x_B \ln x_B)$$

where x_A is the mole fraction of liquid A and x_B the mole fraction of liquid B and the mole number n, Ideal systems are defined as those for which intermolecular interactions (enthalpic contributions) are negligible. In that circumstance, there are interactions, but the average A-B interactions in the mixture of liquid A and B are the same as the average A-A and B-B interactions in the pure liquids. Therefore the driving force for mixing is the increasing entropy of the systems as the molecules mingle around and the enthalpy of mixing is zero. So for the entropy of a binary mixture where the two components have mole fractions x and $(1-x)$ we have

$$\Delta S_{mix} = -nR(x\ln(x) + (1-x)\ln(1-x))$$

From which which it is easy to show that the curvature

$$\frac{\partial^2 \Delta S_{mix}}{\partial x^2} = -nR\left(\frac{1}{x} + \frac{1}{1-x}\right)$$

is always negative for all x. Therefore in ideal systems homogeneous mixing is always favoured. Real solutions of course are composed of particles for which A-A, B-B and A-B interactions are all different (Figs. 2 and 3).

Therefore there will be an enthalpy change when the liquids are mixed. There will also be an additional contribution to the entropy arising from the

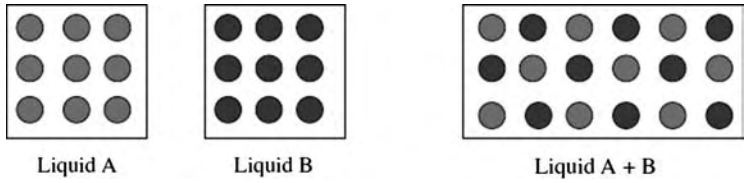


Fig. 2 Schematic illustration of homogeneous mixing of two components.

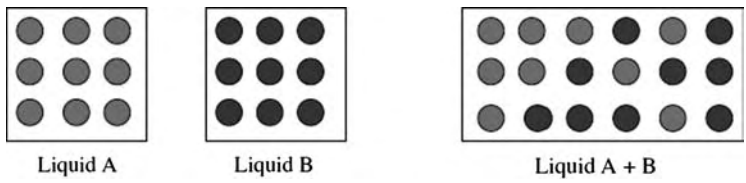


Fig. 3 Schematic illustration of a non-ideal or inhomogeneous two component mixture.

way in which the molecules of one type might cluster together instead of mixing freely with the others.

$$w \propto 2\varepsilon_{AB} - \varepsilon_{AA} - \varepsilon_{BB}$$

This formulation defines the activity cofficients as

$$\ln \gamma_A = x_B^2 \left(\frac{w}{RT}\right)$$

$$\ln \gamma_B = x_A^2 \left(\frac{w}{RT}\right)$$

Together, these lead to the definition of the Gibbs free energy for real, non-ideal mixtures in terms of the activity coefficients as

$$\Delta G_{mix} = RT(n_A \ln \gamma_A x_A + n_B \ln \gamma_B x_B)$$

A useful way of representing the thermodynamic behaviour of liquid mixtures is by the concept of so-called *thermodynamic excess functions*. Thus the excess Gibbs free energy change in a binary mixture is

$$\Delta G^E = \Delta G_{expt} - \Delta G_{ideal}$$

G^E therefore measures the contribution of interactions between liquids A and B. Other thermodynamic excess functions, such as entropy and enthalpy, can be defned in a similar manner (Fig. 4).

Because of the contributions of entropy and enthalpy, two liquids may combine spontanteously to form a single homogeneous single-phase mixture (P = 1) and the liquids are said to be miscible. Or they may combine inhomogeneously (immiscible) resulting in two phases (P = 2). For example, ethanol and water are miscible at room temperature in all proportions whereas benzene and water are immiscible unless one of the two components is present in large excess. In general, if both ΔH_{mix} and ΔS_{mix} are positive the liquids may be immiscible at low temperatures but miscible at

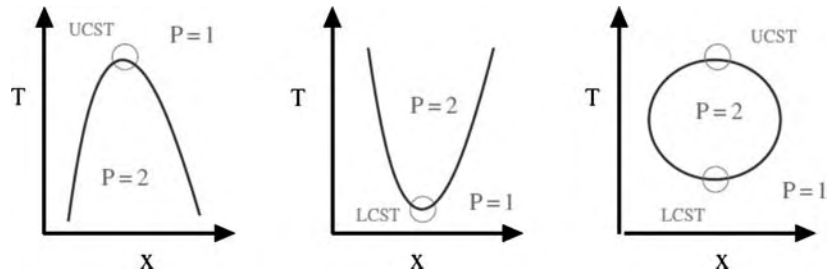


Fig. 4 Illustration of the (temperature *vs* concentration) miscibility regions of a liquid mixture as determined by the Gibbs free energy of mixing: Left panel shows the UCST above which the components of a mixture are miscible in all proportions leading to a single phase region. The centre panel shows the LCST below which there is a single phase region. The presence of both leads to a closed T-x region of immiscibility bounding the two phase (P = 2) region (right panel).

Table 1 Phase behaviour classification of solutes according to the properties of thermodynamic excess functions.

H^E	G^E	TS^E	V^E	$C_p{}^E$	examples	phase behavior
+	+	−	+	−	nitriles, ketones	Completed miscibility or uppercritical demixing
+ or −	+	−	−	+	ether, alcohols, amines	Upper critical demixing, tending to close loop
−	+	−	−	+	tert-amines, cyclic ethers, poltethers	Closed loop
−	−	−	−	≈0	urea, H_2O_2, polyols, sugars	Completely miscible

high temperatures. The minimum temperature at which the liquids are miscible in all proportions is called the upper critical solution temperature (UCST). Below the UCST the liquids are in a two phase region (P = 2), while above the UCST they are in a one phase region (P = 1). Since ΔH_{mix} and ΔS_{mix} are functions of temperature they may change sign as the temperature is varied. As a result, there is nothing to prevent some systems exhibiting both UCST and LCST behaviour resulting in *closed loop immiscibility* regions[9] (Table 1).

The hydrophobic interaction[10–12] refers to the aggregation of hydrophobic species in processes such as protein folding and micelle formation. Thermodynamically, hydrophobic systems are examples of extremely non-ideal solutions. Typically, this is reflected in anomalously high values of thermodynamic molar excess quantities. Still, detailed molecular level descriptions of hydrophobicity are the subject of conflicting models and fundamental issues remain contraversial. From the molecular perspective, an amphipathic molecule consists of two domains with opposing soubility properties in aqueous media: a hydrophilic (polar)domain and a hydrophobic (non-polar) domain. For peptides, the amphipathic character may arise from either the primary structure or the secondary structure: Primary amphipathic peptides have a sequential assembly of a domain of hydrophobic residues with a domain of hydrophilic residues. Secondary amphipathic peptides are generated by the conformational state which allows positioning of hydrophobic and hydrophilic residues on opposite sides of

the molecule. Of course the thermodynamics of mixing is not the whole story for peptides: additional competing interactions arise because amino acids are tethered together by their peptide bonds and backbone linkages as well as constrained by steric clashes. These factors introduce additional competing interactions to that driven by the solvation thermodynamics of their consituent amino acids.

3 Current trends – in what state is the art?

Before embarking on a discussion of example systems and application domains, we note here the sorts of questions for which the study of judiciously chosen model peptides may help to resolve. These are all pertinent to the overarching question of organization on multiple lengthscales including tertiary contacts and self assembly.

A. Accuracy of physical models

A fundamental issue is whether our practical, physical models are capable of accounting for the observed solution state structure of simple peptides: To verify and validate simulation results, a careful comparison of the simulation outcomes directly to experimental data is essential (*e.g.*, obtained by measurements such as nuclear magnetic resonance (NMR), circular dichroism (CD), or infrared spectroscopy) There are ultimately two fundamental concerns regarding theory and related computational simulations representing specific physical models:

(1) computational limitations in the conformational sampling of systems with many degrees of freedom as found in peptides and
(2) inaccurate or poorly transferable force fields which may fail to capture the subtle interactions that determine native folds.

In the case of peptides and their relative structural simplicity, emphasis is almost entirely on the accuracy of the potential descriptions. Within the framework of molecular dynamics force fields, particular importance is directed toward the parameterization of the atomistic interactions, with the functional formulation of the bonded and nonbonded forces often similar among most schemes.[13] There are sustained efforts to improve the accuracy of common empirical force fields such as AMBER,[14] CHARMM,[15] GROMOS,[16] and OPLS[17,18] which tend to revolve around refinement of parameters for the torsional potentials of the protein backbone to balance the conformational equilibrium between extended and helical structures. The basic formulations[19] involve parameterization of the bond stretch, bend, torsion and Lennerd-Jones interactions each represented by the following terms in the energy.

$$E = \sum_{bond} K_r(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)] + \sum_{i<j}^{atoms} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\in R_{ij}} \right]$$

Separate from this is the question of configurational sampling: The study of the assembly or stability of complex supramolecular aggregates in these

solutions by constant temperature molecular dynamics simulations is limited by the efficiency with which methods samples the configurational space.[20,21] There are examples of fast-folding (< millisecond timscales) proteins,[22,23] wherein the sequences reduce energetic frustration leading to a relatively smooth funnel lanscape.[24,25] Wild-type villin folds, for instance, over several microseconds[26] and submicrosecond folding has been observed in certain mutants.[27] But these are exeptions: In general there is the possibility of getting trapped in metastable energy minima meaning that generation of equilibrium configurations over the simulation time is not guaranteed. The general idea of parallel tempering or replica exchange is to simulate *M* replicas of the original system of interest, each replica typically in the canonical (NVT) ensemble, and usually each at a different temperature. The high temperature systems are generally able to sample large volumes of phase space, whereas low temperature systems, whilst having precise sampling in a local region of phase space, may become trapped in local energy minima during the timescale of a typical computer simulation. Parallel tempering achieves good sampling by allowing the systems at different temperatures to exchange complete configurations with a probability of exhange that depends on the energy or temperature difference beteween them.[28,29] Thus, lower temperature systems can access wider regions of configurational phase space. In patricular, the advantage of parallel tempering comes from the fact that it is more than 1/M times more efficient than a standard, single-temperature simulation. This increased efficiency comes firectly from the fact that the lower temperature replicas sample regions of phase space that they would not have been able to access had regular sampling been conducted for a single-temperature simulation that was M times as long (Fig. 5).

The proposed swap move is based on the acceptance criteria defined by the Metropolis algorithm according to

$$S(x \rightarrow x') = 1 \text{ for } \Delta \leq 0 \text{ or } Exp(-\Delta) \text{ for } \Delta \geq 0$$

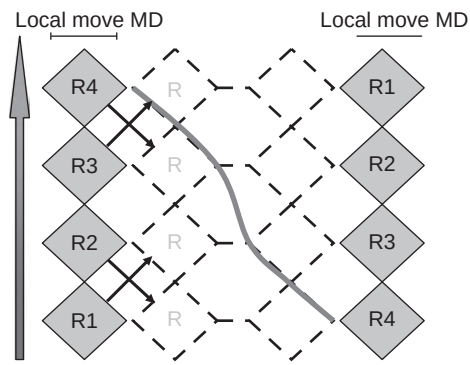


Fig. 5 Illustration of the replica exchange or parallel tempering simulation procedure wherein simulations of identical systems are performed at several temperatures. Standard ''local move'' MD is occasionally interrupted by configuration swaps at different temperatures to improve sampling and accelerate the approach to equilibrium.

where

$$\Delta = \Delta\beta\,\Delta V \text{ and } \beta = \frac{1}{k_B T}$$

If the given swap[30,31] is accepted, we update the permutation functions and rescale or reassign velocities for each replica so they are representative of the new temperatures. The acceptance rule above is appropriate for replica exchange simulations where each replica is simulated in the canonical ensemble. In terms of practical implementation. parallel tempering simulations though computationally demanding make relatively modest communication requirements between the replicas and can be performed on distributed computing clusters with each replica running on single CPU. Although this is sufficient for small to modestly sized molecular systems,[32,33] it becomes a problem when one wants to run large size[34] protein systems for long simulation times. There has been considerable recent progress in this regard starting with initiatives such as the IBM Blue Gene project[35,36]

There are many promising results suggesting that our physical models for peptides is reasonably well developed: Duan and Kollman[37] performed a microsecond molecular dynamics simulation of a 36-residue villin peptide in water starting from an unfolded structure. High-resolution structures of villin have recently been reported by Pande *et al*.[38] Several studies have been reported for the fast-folder 20-residue Trp-cage peptide. Finally Lei and Duan folded the albumin binding domain, a 47-residue, three-helix bundle, to 2.0 Å.[39] The 46-residue protein A folds in silico to within 3.5 Å root-mean-square deviation (rmsd) using Monte Carlo dynamics with an implicit (continuum) solvation model.[40]

The picture is somewhat more equivocal in the case of smaller peptides which offer easier opportunities to examine force field bias and expose systematic variations or deficiencies. Gnanakaran and Garcia[41] studied alanine peptides using a modified AMBER force field and exposed sensitivity to backbone torsional energies.

Much work has also focused on force fields used with implicit solvation, such as the generalized Born surface area (GBSA) method, as implicit models tend to be orders of magnitude faster for large protein simulations. Zhou[42] compared the results of 3 ns explicit and implicit REMD simulations for the protein G hairpin, considering the AMBER ff94, ff96, ff99, and OPLS-AA force fields in combination with explicit (TIP3P for AMBER, SPC for OPLS) or implicit (GBSA for AMBER, surface generalized Born for OPLS) solvents. That study found that implicit solvation tended to destabilize the native structure, most prominently in the AMBER94 and AMBER99 force fields, which showed strong helical biases.

These detailed force field tests show that, despite notable successes, two key issues remain unresolved: balance in backbone conformational preferences and implicit solvation models, particularly in ion-pairing interactions.

Combined experiment-simulation studies focussed on contraversy over the structure of a short fragment of the human HIV-1 membrane glycoprotein gp41. It has has been examined using a combination of parallel

tempering molecular dynamics (PTMD) and far UV circular dichroism spectroscopy. Conflicting reports exist on the solution state conformational bias in this membrane proximal domain spanning the epitope for the 2F5 monoclonal antibody. It was found that $gp_{41}^{659–671}$ exhibits conformational plasticity in which competing folding propensities are present and can be influenced by local microenvironment. Contrary to previous reports, the d_{310} helix does not emerge as a dominant motif from either simulation or experiment, and this peptide is therefore not a model system for this fold type. Other fold groups such as turn motifs are identifiable at elevated temperatures in the PTMD trajectories and are potentially relevant in antibody binding. Here there is also evidence of force field bias with helical populations in pure water are significantly overestimated according to the CHARMM parametrization. However, circular dichroism (CD) data show that helices are promoted in membrane mimetic solvents. As this is a membrane proximal peptide, the helical motif may well have physiological significance.

The balance of all available evidence is that the $gp_{41}^{659–671}$peptide exhibits a number of features that have apparent physiological relvance, and it is therefore a useful model system in that respect. These include at least (1) conformational plasticity in which competing folding propensities are present and can be influenced by local microenvironment and (2) specific fold groups such as turn motifs which are identifiable at elevated temperatures in replica exchange molecular dynamics and potentially relevant in antibody binding and helices (the preferred motifs in the simulated system and promoted in membrane mimicking solvents which is particularly relevant as this is a membrane proximal peptide, and therefore the helical motif may well be physiologially significant).

A recent survey of the thermodynamics of hydration for model compounds related to amino acids suggests that the thermodynamic properties considered are much more dependent on the choice of water force field than on the protein force field. Molecular dynamics water models are generally parameterized to reproduce bulk water properties. However, the accuracy of representation of water properties in such a model is not generally well correlated with its accuracy in combination with solvated species: The issue of water-model choice is complicated by the fact that protein force fields are generally parameterized and tested using a specific model, and thus one cannot convert to a water model even if it has been shown to have superior bulk properties.

B. To what extent do we understand the influence of local solvation environment?

The formation of higher-order organization, tertiary contacts, in solution or in the vicinity of membranes presents a widely varying solvation environment. For example, in the formation of secondary and particularly tertiary structure, a hydrophobic side chain may have strong preference to be buried inside the protein core in the native state. Yet, the same side chain may be partially exposed to water in unfoled or nonnative states. The substantial dielectric constant difference between the two needs to be accounted for in

molecular models and simulations. For meaningful theoretical studies of biomolecular systems, it is necessary to have potential energy functions providing a realistic and accurate representation of both the microscopic interactions and thermodynamic properties (*e.g.* mixing enthalpy). This presents a difficult challenge, particularly when charged or highly polar species are involved: Here, electronic polarization is expected to have a significant role in both structure and energetic properties. One of the limitations of many computational models in current use is that this variation is rarely captured. Most biomolecular potential functions typically neglect real many-body polarization responses. Instead, the effects are "averaged" through parametrization of the atomic partial charges and used for all environements. This results in a compromise between accurate representation of the molecular energies and bulk solvation properties. Such potential functions can yield meaningful results of semiquantitative accuracy, but are not necessarily transferable across various physically relevant solvation environments.[43] Major advances in the representation of biomolecules and particularly the treatment of tertiary contacts and membrane interactions are anticipated if many-body polarization is explicitly incorporated.

At the present time, there is active effort leading to the development of a new generation of force fields for computational studies of biological systems that include induced polarization. Essentially, these models introduce an additional term into the energy expression of the form

$$E_{pol} = -\frac{1}{2}\sum_{atoms}\left(\overrightarrow{\mu_i}\cdot E_i^0\right)$$

where

$$\overrightarrow{\mu_i} = \alpha\overrightarrow{E_i}$$

Here μ_i is the induced dipole. We note that these polarization energies are necessarily many-body – that is, they are not pairwise additive and depend on triplet and higher order correlations.

However, much more work is needed before such potential functions are ready to be used in simulations of heterogenous biological systems. Most initial work is focussed on refining polarizable models for water. Many of the models of water currently used in biomolecular simulations are based on fixed effective partial charges which were adjusted to yield accurate bulk liquid properties. Although such models, which incorporate the average influence of electronic induction in an isotropic liquid, have been and remain exceedingly useful, there are some well-justified concerns about their ability to represent the properties of liquid in inhomogeneous environments Electronic polarization of water is highly sensitive to its environment. In the gas phase, an isolated water molecule has a dipole moment of 1.85 D However, the average molecular dipole is 2.1 D in the water dimer, it

increases in larger water clusters, reaching, in the condensed phase, a value between 2.4–2.6 D.[44]

C. Local *vs*. Nonlocal debate in the formation of complex assemblies:
A second fundamental question regarding the formation of tertiary and higher-order structures refers to the roles of local and non-local interaction in protein folding and stability. A direct way to approach this issue comes from the behaviour of very short which can expose the nature of local interactions and therefore offer a measure of the importance of local interactions in determining protein secondary structure elements. The hypothesis is that if local interactions are dominent, then if folded protein segments having well-defined secondary structure motifs are extracted, the structure isolated fragments should exhibit native-like structures in the absence of tertiary contacts. In general the thermodynamics of peptide folding and larger protein folding differs slightly: Most small proteins show two-state behaviour each molecule is either fully folded or fully unfolded and does not adopt long-lived intermediate structures. Larger multi-domain proteins can differ from pure two-state transitions. An individual molecule can convert from folded to unfolded defining a rate constant K = [folded]/[unfolded] for the two populations in thermodynamic equilibrium. This defines a simple expression for the difference between the Gibbs free energy of the ordered and disordered conformers as $\Delta G^{do} = -RT\ \ln(K)$. Above the transition midpoint, the free energy is positive (*i.e*. unfavorable), yet a population of folded molecules still persists.

A large number of examples exist of helical protein fragments that are able to adopt a significant population of the native a-helix structure in water. The best example is the 13-residue C-peptide of ribonuclease A that adopts native helices in water in high populations. This case is, however, far from being general and there are many cases where native helices have marginal stability at best in short protein fragments.

Of course, some local interactions are very clearly important. Perhaps none more so than the steric repulsions which turn out to be very substantial: In fact, most conformational space is inaccessible. There are only two major allowed regions in the space of the backbone angles: Others result in steric hinderance (*i.e*. bring any two atoms closer than the sum of their respective van der Waals radii) and are considered to be disallowed; all other values are allowed. These two regions correspond to the values assumed by backbone dihedral angles of peptides in α-helix and β-sheet

So how important are non-local interactions in discriminating between the two allowed regions of the Ramachandran plot? Clues to the answer come from experiments on alanine-based peptides. Short alanine-based peptides form stable helices in water. This observation implies that helix formation is energetically favored for main-chain atoms, because poly-L-alanine is essentially pure backbone. (The b carbon can be regarded as a backbone atom because it has no additional degrees of rotational freedom, and all chiral amino acid residues have an equivalent b carbon.) This conclusion is largely indepentent of the ongoing debates over whether short helices are stabilized primarily by hydrogen bonds or van der Waals interactions.

D. Where are the remaining measurement challenges?

Arguably, the most developed experimental data on the assembly process comes from the slower, late phases of folding. Experimental evidence for formation of kinetically observable intermediates during has been accumulating, and techniques such as NMR has provided direct information about the nature of these species (Roder & Wuthrich, 1986; States *et al*., 1987). However, identification of the transient structures that are likely to initiate protein folding and direct the folding pathway(s) is very much more difficult because the polypeptide chain folds rapidly and cooperatively into a compact folded structure.

In fact, the earliest attempts to find structure in peptide fragments of proteins were generally unsuccessful or gave equivocal results and led to the view that short peptide did not contain secondary structure in water solution. For many years the ribonuclease A example was the only exception.

Once it was recognized that short synthetic peptides can induce antibodies that recognize, target sequences in the folded protein (Lerner, 1984; Niman *et al*., 1983) there was renewed effort to explore the structural basis for this recogniction with the assumption being that these peptides must have strong conformational bias either free in solution or on the B-cell receptor, approaching that which it adopts in the native protein. The only other possibility being that the protein becomes disordered through conformational fluctuations (Tainer *et al*., 1984, 1985; Dyson *et al*., 1988) Two-dimensional NMR spectroscopy provides a highly sensitive approach for detection and identification of folded structures in aqueous solutions of peptides. In contrast to many other forms of spectroscopy, such as circular dichroism, infrared, and Raman spectroscopy, NMR spectroscopy provides structural information, once resonances have been assigned, at specific sites throughout the peptide. Several NMR parameters provide information about molecular structure. Information about dihedral angles can be obtained from spin-spin coupling, the existence of hydrogen bonding can be inferred from the temperature coefficients of the amide proton chemical shifts and from amide proton exchange data, and most importantly, direct information on the distances between protons in the peptide can be inferred from the nuclear Overhauser effect (NOE). Interpretation of NOE data for a small linear peptide in water solution is complicated by conformational averaging. Barriers between low-energy minima have been estimated to be only a few kilocalories per mole, and hence peptides are expected to sample a number of backbone conformations on nanosecond time scales. NMR parameters such as the chemical shift and coupling constant are thus a population-weighted average over all conformers, and the NOESY spectrum will be representative of all conformations that have sufficiently high populations.

Synchrotron radiation circular dichroism (SRCD), which has greatly expanded the utility of the method especially as a tool for both structural and functional genomics. Its applications take advantage of the enhanced features of SRCD relative to conventional CD: the ability to measure lower wavelength data containing more electronic transitions and hence more

structural information, the higher signal-to-noise hence requiring smaller samples, the higher intensity enabling measurements in absorbing buffers and in the presence of lipids and detergents, and the ability to do faster measurements enabling high throughput and time-resolved spectroscopy; Linear dichroism (LD) which occurs because molecules fixed in space absorb plane-polarized light maximally in well-defined molecular directions. In the case of electronic transitions, these directions correspond to those of the electric-dipole transition moment, therefore

Solid State Nuclear Magnetic Resonance (ssNMR)[45] overcomes the difficulties in applying solution state NMR to biomolecular complexes and is particularly powerful in the study of membrane-bound peptides. In particular, it overcomes the spectral broadening and reduced tumbling rates which complicate the application of solution state NMR to membrane bound molecules; Small-angle scattering (SAS) covers an extremely broad range of molecule sizes, which is of primary importance for a systemic approach in structural biology. The method provides unique information about the overall structure and conformational changes of native individual proteins, functional complexes, flexible macromolecules and hierarchical systems. However, the absence of standards and the ease with which 3D models can be obtained from 1D scattering data means that an understanding of sample quality, data quality, and modeling assumptions is essential to have confidence in the results.

4 Model systems for self-organization

A. Hydrophobic association in minimal models

One basic problem existing at both small and large lengthscale regimes is the molecular structure of the water-amphiphile interface. Difficulty in finding a satisfactory resolution to this issue in the important biomolecular context is exacerbated by the structural complexity of the relevant molecules.

If a simple alcohol such as methanol or ethanol is mixed with water, the entropy of the system increases far less than expected for an ideal solution of randomly mixed molecules. This well-known effect has been attributed to hydrophobic headgroups creating icy or clathrate structures in the surrounding water,[46,47] although direct experimental support for this hypothesis is scarce.[48] In fact, an increasing amount of experimental and theoretical work suggests that the hydrophobic headgroups of alcohol molecules in aqueous solution cluster together. However, a consistent description of the details of this self-association is lacking. Neutron diffraction with isotope substitution[49] provided the first definitive picture of the molecular scale structure in this highly non-ideal mixture, albiet in a very simple model system. Data on a concentrated methanol-water mixture indicated that most of the water molecules exist as small hydrogen-bonded strings and clusters in an fluid of close-packed methyl groups, with water clusters bridging neighbouring methanol hydroxyl groups through hydrogen bonding. This behaviour suggests that the anomalous thermodynamics of water–alcohol systems arises from incomplete mixing at the molecular level and from retention of remnants of the three-dimensional hydrogen-bonded network structure of bulk water (Fig. 6).

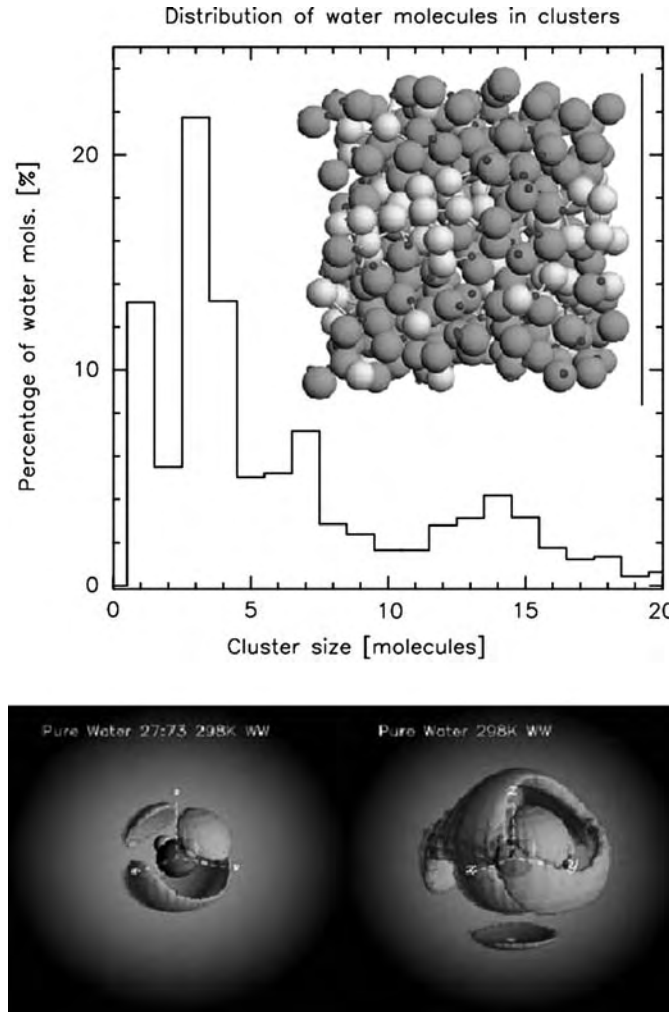


Fig. 6 Cluster distrubution (top) and hydration shell structure (bottom panel) obtained from neutron diffraction measurements and revserse Monte Carlo simulation showing complex assemblies formed in concentrated methanol water mixtures. Top Panel Copyright 2002 Nature Publishing Group.

Interactions between the side chains of hydrophobic amino acids are crucial to the stabilities of proteins and their complexes. To fully understand the kinetics of processes such as protein folding, it is important to know the lifetimes of typical hydrophobic interactions, since they can have two opposing consequences: On one hand, for side chains situated adjacent to each other in the native structure, a long contact lifetime might be advantageous. On the other, for side chains that do not form contacts in the folded state, the formation of long-lived interactions might be disadvantageous. This may be why stretches of contiguous hydrophobic residues are relatively rare in nonmembrane proteins. Molecular dynamics (MD) simulations of different amino acid pairs in aqueous solution have

now been reported to explore the hydrophobic contact lifetimes. These have led to estimates for association constants and demonstrate that current computational resources are now sufficient for full sampling of association events to be achieved for small molecules in aqueous solution. These data are promising for parametrization of new force fields to reproduce aqueous phase binding data.

Thermodynamics of these demixing processes and the origin of their anomalous properties can be understood on the basis of an extension of the fundamental thermodynamics of mixutres.

Let us suppose we have N_j molecules of type j in a box of volume V. The mean number density of these molecules is $\bar{\rho} = N_j/V$, and if the molecules are distributed uniformly around the box, then the local density at any point in the box given by $\rho_j = \bar{\rho}$. This means that the entropy relative to the uniform molecule distribution given by

$$T\Delta S_j = -kT \int dV \rho_j \ln \frac{\rho_j}{\bar{\rho}} \tag{1}$$

is zero by virtue of the fact that $\ln(1) = 0.0$.

If however we now impose density fluctuations on this distribution of molecules the entropy will decrease. Suppose for example that each molecule occupies a volume v_j on average and the molecules form one or more clusters in which they are all in contact with at least one other molecule within the same cluster. Then the density of molecules within the cluster is $\rho_j = 1.0/v_j$ and zero elsewhere, so that the entropy becomes

$$T\Delta S_j = -kTN_j \ln \frac{V}{V_j}$$

where $V_j = N_j v_j$ is the volume occupied by the molecules. Since the molecules are clustered they occupy a smaller volume than the box containing them, so that $V/Vj > 1$ and the entropy falls. For example if the molecules occupy only half the box volume, then the entropy per molecule ($kT = 2.478$ kJ/mole at 298K) $T\Delta S_j = -2.478$ ln 2 kJ/mole $= -1.718$ kJ/mole. Thus there is ample scope within a clustering framework to explain the observed entropy decrease on mixing, at least in principle.

In addition to methanol as a model system for hydrohobic assembly there are other systems which also may exhibit rich associations relevant to biomolecular structure and function: Aqueous proline amino acid for example is known to exhibit unusual physical and colligative properties.[50–52] Of particular interest is the fact that the proline solubility in water is much higher than that of any other amino acid. This is particularly significant for experimental measurements which can be made over a wide concentration range (up to ~ 6.5 M). Above concentrations of about 3.5 M, the solutions acquire viscosities that are unexpectedly large for such low molecular weight species. These mixtures also show rudimentary biochemical functionality: In particular, aqueous proline exhibits hydrotropism – whereby it increases the solubility in water of hydrophobic compounds. Related to this is the suggestion that aqueous proline may act as a simple

chemical chaperone which assists protein folding. Finally, aqueous proline is considered to be a natural bio/cryo-protectant expressed by plants and other organisms (such as bacteria and protozoa) under adverse conditions including low temperature stress. There has been extensive speculation in the literature as to the molecular origin of these properties. One prevailing view proposed by Schobert[53] has been that ordered intermediate range aggregates are formed in solution and that these lead to the peculiar properties of the mixtures. While there has been some indirect evidence for their presence from spectroscopic,[54] light scattering[55] and calorimetric measurements very recent neutron diffraction obtained at intermediate and low Q appear to rule out persistent aggregates having a well-defined characteristic lengthscale. Extensive computer simulations based on classical empirical potentials with sampling efficiency improved by parallel tempering have also ruled out the presence of well-ordered mesoscale structures but do provide evidence for diffuse local heterogeneities arising from concentration fluctuations. Simulations using nonpolarizable water potentials (SPC/E, TIP3P and TIP4P) have been performed and compared to neutron diffraction data through direct evaluation of the static stucture factor $S(Q)$ from the molecular dynamics trajectories. Under ambient conditions there are claims that the mixtures have a complex mesoscale structure in which proline molecules form semi-ordered aggregates in aqueous solution. According to Schobert, these associations, resemble polymer-like ''hydrophilic colloids''. The basic structural elements of these assemblies are assumed to comprise a hydrophobic backbone of stacked pyrrolidine rings with solvent-exposed hydrophilic groups on the surface. Some indirect support for this idea has come also from independent measurements including calorimetric data which revealed evidence for low-temperature eutectic phase separation in aqueous proline at a range of concentrations. A similar inference is drawn from Fourier-transform infrared (FTIR) spectra which show a splitting of the COO- asymmetric stretch band on increasing proline concentration implying that the carboxylate group exists in distinct local environments one of which possibly being the proposed pyrrolidine stack. There is also speculation that significant increases in light scattering in proline solutions above 1.5 M could be explained by proline clustering. Increases in emission intensity of a hydrophobic dye as a function of proline concentration in aqueous solution have also been seen. This was interpreted as additional evidence for the formation of supramolecular assemblies of proline having distinct hydrophobic regions. In the relaxing regime we can attempt to fit to the Rayleigh-Brillouin spectra using a Debeye-type relaxation model. Here the width of the Brillouin spectra at scattering vector q is determined by $D_T q^2$ where D_T is the thermal diffusivity.[56] In the simplest case one assumes a single relaxation time. Then the isotropic scattered intensity (proportional to the dynamic structure factor) takes the form

$$I(\omega) = I_0 \frac{\gamma_0 + M''(\omega)}{\left[\omega^2 - \omega_0^2 + \omega M'(\omega)\right]^2 + \omega^2[\gamma_0 + M''(\omega)]^2}$$

with

$$M' = \frac{-\Delta^2 \omega\tau^2}{1+(\omega\tau)^2}$$

$$M'' = \frac{\Delta^2 \tau}{1+(\omega\tau)^2}$$

Figure 7 shows the temperature dependence of Brillouin light scattering as a function of temperature. Peaks are fit to a generalized hydrodynamic model and illustrate the broadening and subseqent narrowing of the line-width characteristic of viscoelastic dynamics in the material. To the right, is shown the an example bonding motif between water and proline amino acid obtained from MD simulations.

The most extensive recent work shows clear spectrocopic evidence for viscoelatic behaviour and a fast relaxation mode in the glass is found from Rayleigh-Brillouin scattering data. Below 0.07 mole fraction bulk water

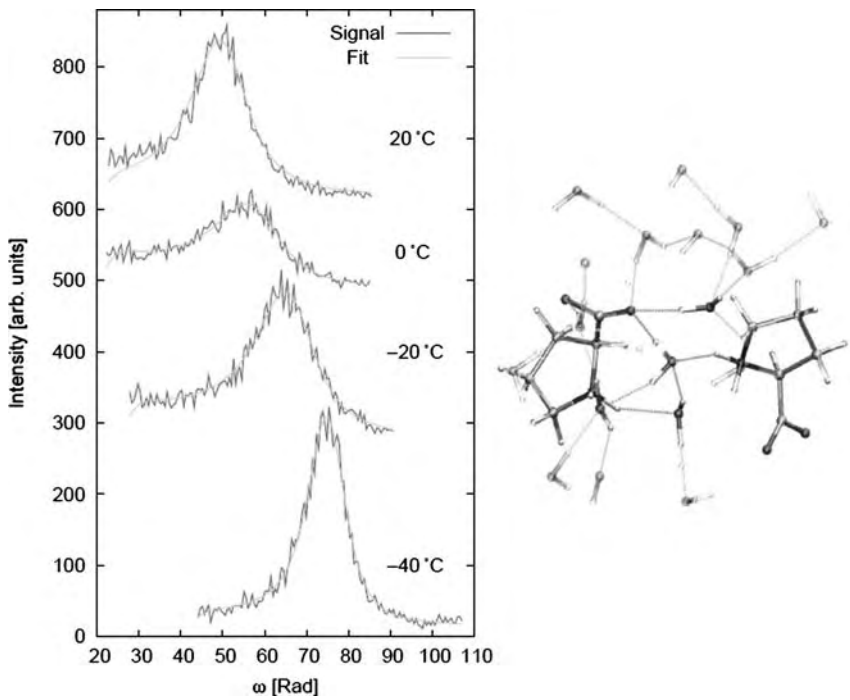


Fig. 7 Glassy dynamics in aqueous proline amino acid at low temperature obtained from Brillouin inelastic light scattering as a function of temperature. The Brillouin data (left) shows a characteristic non-monotonic behaviour with cooling as the relaxation times moves through and then past the Brillouin frequency window. To the right is the local structure and H-bonding motifs obtained from molecular dynamics simulations. From "Molecular mechanisms of cryoprotection in aqueous proline: Light scattering and molecular dynamics simulations" R. Z. Troitzsch, H. Vass, W. J. Hossack, G. J. Martyna and J. Crain, *J. Phys. Chem. B*, **112**, 4290. Copyright 2008 American Chemical Society.

behaviour is observed. A complex dependence of the Landau-Placek ratio on composition and temperature is attributed to separable contributions from concentration fluctuations in the more dilute solutions and to structuural relaxation in more concentrated solutions. The contribution to the Landau-Placzek ratio from concentration fluctuations appears to be largest at the concentration where the temperature dependence of the Brillouin frequencies ceases to be dominated by water and the viscoelastic character emerges. It is assumed this arises because the long-range correlations in water are disrupted at this concentration even though the Raman evidence suggests preservation of local tetrahedral order down to the glass transition where the spectrum appears similar to amorphous ice.

B. Minimal models for hydrogen bonding in peptides: NMA

Hydrogen bonds (H-bonds) are fundamental to the structure and biological function of proteins. They control active configurations by connecting protein structure in dynamic equilibrium. The formation and disruption of hydrogen bonds affects the rates and processes which are responsible for much of the biological activity of proteins.[57–61] A complication is that this behavior is strongly medium dependent, so the action of hydrogen bonds in isolated systems is quite different from that in the aqueous environment. In addition to conventional (N–H . . . O–C) bonding, so-called improper (methyl-donated) hydrogen bonds of the Ca–H. . .O–C type may also participate in the stabilization of important motifs, such as the alpha-helix and beta-sheet as well as in transmembrane inter-helix interactions. However, because the peptide C–O and N–H groups form competing hydrogen bonds to water (See for example Paschek *et al.*[62]) there may be some cancellation so that the net contribution of the peptide H-bond ($CO \cdots HN$) to protein stability might be smaller than expected.Finally, the problem of assessing the entropic contributions to these bonds is a matter of current research.

Recent efforts to study these H-bonding types have so far been driven by the combination of high-resolution X-ray diffraction, NMR and bioinformatics techniques, which yield little information about the intrinsic fluctuations of H-bonds that are important to folding and function. Methods based on atomistic computer simulation are generally limited by the difficulty of sampling the configuration space of large biomolecules and by the inability of empirical force fields to correctly account for the changes in the local environment. An emerging strategy for developing general models of H-bond interactions involves studying selected model systems, such as peptidic fragments. While retaining many of the properties of large biomolecules, such model systems have relatively simple energy landscapes that can be effectively sampled in a computer simulation.

N-methyl-acetamide (NMA) holds special status as the simplest model system of the peptide linkages[63–65] as illustrated schematically in Fig. 8. The molecule (C_3H_7NO) contains a single peptide group terminated by methyl moieties on the carbonyl carbon and on the amino nitrogen. The two methyl groups exhibit near free rotation with their relative orientations corresponding to the conventional main chain dihedral angles of polypeptides and proteins (Fig. 9).

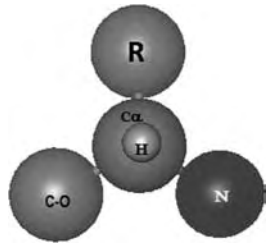


Fig. 8 Schematic illustration of the peptide structured captured by simple model systems.

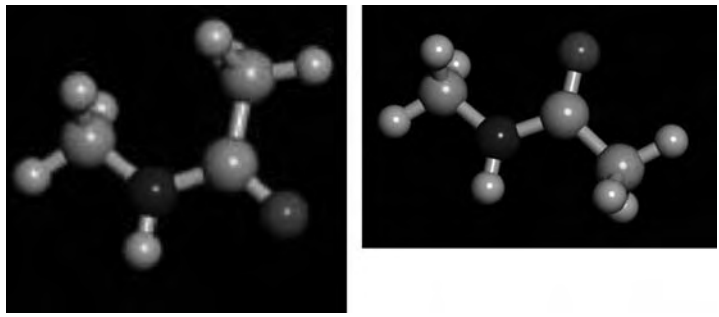


Fig. 9 Illustration cis-trans isomers in the minimal model peptide fragment N-methyl acetamide. This torsional degree of freedom and energy barrier heights are assumed to be representative of main chain dihedral energetics in peptides and proteins.

The results of classical molecular dynamics (MD) simulations and Raman spectroscopy studies of neat liquid N-methylacetamide (NMA),[66,67] predict that near ambient conditions, the molecules form a hydrogen bond network consisting primarily of linear chains. Both the links between molecules within the hydrogen-bonded chains and the associations between chains are stabilized by weak methyl-donated ''improper'' hydrogen bonds. The three-dimensional structural motifs observed in the liquid show some similarity to protein beta-sheets. The temperature and pressure dependence of the hydrogen bond network, as probed by the mode frequency of the experimentally determined amide-I Raman band, blue shifts on heating and red shifts under compression, respectively, suggesting weakened and enhanced hydrogen bonding in response to temperature and pressure increases. Disruption of the hydrogen-bonding network is clearly observed in the simulation data as temperature is increased, whereas the improper hydrogen bonding is enhanced under compression to reduce the energetic cost of increasing the packing fraction. Because of the neglect of polarizability in the molecular model, the computed dielectric constant is underestimated compared to the experimental value, indicating that the simulation may underestimate dipolar coupling in the liquid (Fig. 10).

Pushing the model system analogies further, aqueous NMA is therefore a useful foundation for investigating water/peptide hydrogen bond competition and its thermodynamics. Classical molecular dynamics simulations of aqueous NMA have now been performed across a wide concentration range. The simulations predict considerable NMA self-association even at

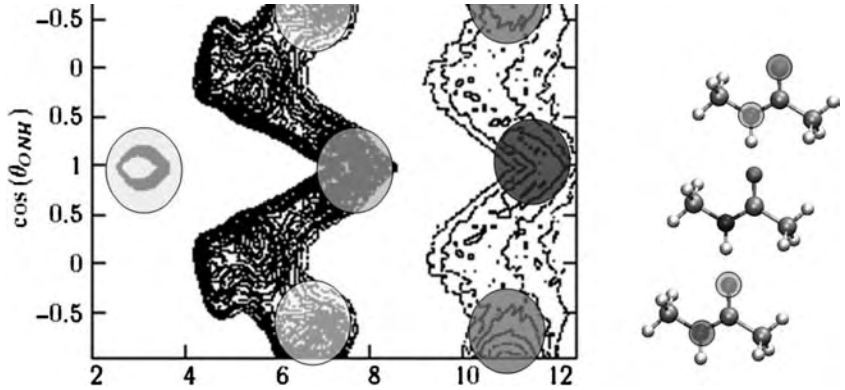


Fig. 10 Illustration of medium range order in neat NMA liquid as obtained from molecular dynamics simulations. The representation is that of a two dimensional distance-angle distribution function showing spatial and orientational correlations linked by hydrogen bonds. The colors and left panel show the dimer and trimer contacts. The further dark blue contact arises from the tetramer with the shortest contact coming from the close N-O contact. [Colour image available on-line]

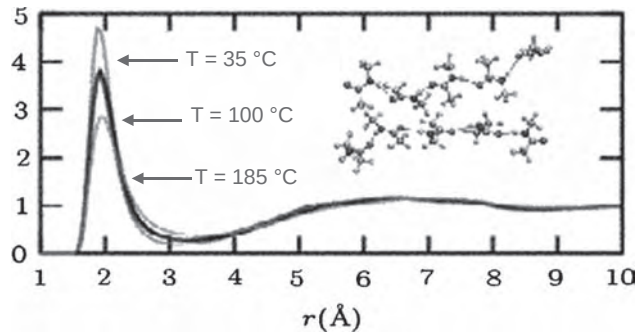


Fig. 11 The integrated radial distribution function for O-H hydrogen bond contacts in liquid NMA as a function of temperature.

low concentrations with a concentration-dependent increase in the ratio of branched to linear clusters co-existing with other more complex hydrogen-boded topologies. Water-mediated NMA contacts are a feature of this regime, manifested by an unexpected increase in the number of short NMA oxygen contacts arising from water bridge motifs. In contrast, bulk water structure is significantly disrupted by the addition of even small quantities of NMA. Liquid configurations characteristic of beta-sheets in proteins are also observed in neat NMA: Like real beta-sheets interchain motifs consists of moieties roughly aligned with one another in a direction transverse to that of the chains and arranged in an approximately planar structure. Unlike genuine beta-sheets of course, the liquid motif obviously lacks both covalent bonding between monomer subunits and well-defined parallel or anti-parallel classes. The liquid motif is stabilized by intermolecular dipole coupling and by improper H bonds (Fig. 11).

A final example where this minimal peptide may offer useful insight into complex tertiary assemblies is to to with so-called interior water in proteins. The picture of protein hydration is of a "dry" hydrophobic interior

stabilised by intra-molecular hydrogen bonds and sidechain packing with a few cavities containing bound water molecules. The exterior of the protein molecule is then made up of hydrophilic groups which hydrogen bond to water molecules thus creating a monolayer of bound surface or hydration water. Recent spectroscopic investigations of water dynamics have complicated this picture and indicate that the protein interior is much more wet than expected and that both the folded and unfolded states contain signicant quantities of bound water molecules buried deeply in the interior of proteins and their presence may be an inevitable consequence of the folding process itself.[68] These are highly conserved features among homologous families and one therefore assumes that they may have structural and functional importance. Typically, these buried water molecules exist in complex local environments forming multiple hydrogen bonds with nearby main chain peptide groups. A very recent statistical analysis suggests some preference for buried water molecules to reside near regions without well developed secondary structures. In some cases, buried waters have been found to form four optimal hydrogen bonds with the surrounding protein but a wide variety of bonding types (ranging from 0 to 4 hydrogen bonds) have been identified. Their average residence times can be accessed experimentally from 1H NMR of nuclear Overhauser effects and multinuclear NMR dispersion measurements. These have revealed buried-to-solvent exhange processes occuring *via* conformational fluctuations on timescales ranging from 10^{-8}s to 10^{-4}s at room temperature for bovine pancreatic trypsin inhibitor (BPTI).[69] Despite their structural and functional significance and intense investigation, the structural properties of buried water molecules and the dynamics of their exchange with the solvent remains poorly understood. This is partly because internal water molecules are challenging to probe experimentally and to simulate. X-ray crystallography techniques have been the most common structural probe though detection of buried water molecules is limited by a weak diffraction signal. Moreover, X-ray studies are sensitive only to bound internal water molecules that are highly ordered at specific binding sites. Structurally disordered interior water is not easliy detectable. A further complication is that hydration structure obtained from protein crystallography may be influenced by the presence of crystallization agents. Magnetic resonance dispersion measurements are also restricted both in spatial resolution and in the sense that only rotationally hindered water molecules contribute to the dispersion signal (Fig. 12).

In general, water molecules can be partitioned into a number of classes based on bonding at the protein surface. This leads naturally to the notion of water molecules bound to surfaces, protein clefts or channels (which may be active sites in the absence of a substrate) and buried water molecules typically occupying protein cavities. A buried water molecule is identified from its bonding topology as one that cannot be connected to bulk water molecules by a continuous hydrogen bonded path. For a recent review see Purkiss *et al*. All currently reported crystal structures of wild-type BPTI identify four buried water molecules deeply sequestered in the groove between the two loop regions. They exhibit extensive hydrogen bond linkages to the protein and are belived to tbe fundamental to the native

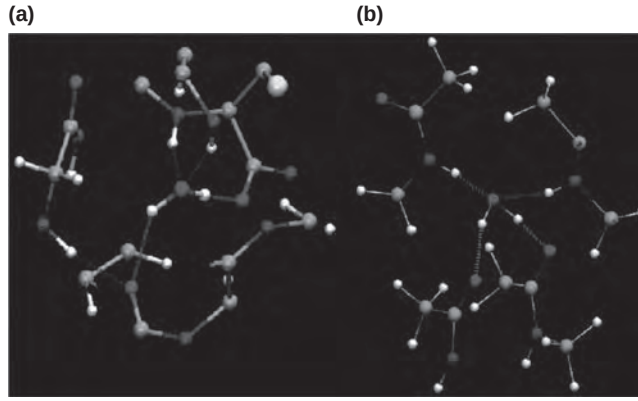


Fig. 12 BPTI showing buried water motif (a) and NMA hydrogen bonding pattern from molecular dynamics trajectory. Structure and hydrogen bonding in neat N-methylacetamide: Classical molecular dynamics and Raman spectroscopy studies of a liquid of peptidic fragments, T. W. Whitfield, G. J. Martyna, S. Allison, S. P. Bates, H. Vass and J. Crain, *J. Phys. Chem. B*, **110**, 3624. Copyright 2006 The American Chemical Society.

structure. Still, they exchange with hydration water on nanosecond to millisecond timescalses with the three mutually H-bonded waters exchanging fastest (mean residence times of 10^{-8} to 10^{-6} s. The isolated buried water is the slowest to exchange doing so with a mean residence time of just under 200 μs. This long residence time for this molecule has been attributed to the proximity of a disulfide bond which suppresses conformational fluctuations. Recently it has been shown that cleavage of this bond significantly accelerates exchange with the residence time dropping by two orders of magnitude

Despite the simplicity of the aqueous NMA system relative to a real protein, analogous partitioning of water molecules into bulk and isolated buried types can be made and their structure and dynamics can be studied in detail as a function of concentration. The figure shows that the hydrogen-bonded interactions of the buried water present in the bovine pancreatic trypsin inhibitor (BPTI) protein bear similarities to the interactions of isolated water molecules in concentrated NMA solutions.More complex classifications of water molecules which form a variety of H-bond combinations with both NMA and water require a more detailed interrogation of bonding topology (Fig. 13).

The peptide bond originates from an electronic delocalization on the three atoms that are constrained to be on the same plane. The delocalization leads to an energetic stabilization. Structurally, proteins are formed from a succession of these planes that may orient differently relative to each other. Another electronic feature of the peptide bond is that it has a dipole moment: The bond is characterized by a delocalisation of negative charge near the carbonyl group and a partial positive charge near the amide group. The dipole moment is not negligible and affects the stabilization of a protein structure, since it is present for each amino acid and its contribution is particularly important for some secondary structures. For steric reasons the trans configuration is preferred by a factor of thousand, except when the residue i + 1 is a proline. In this case, due to the cyclic nature of the residue,

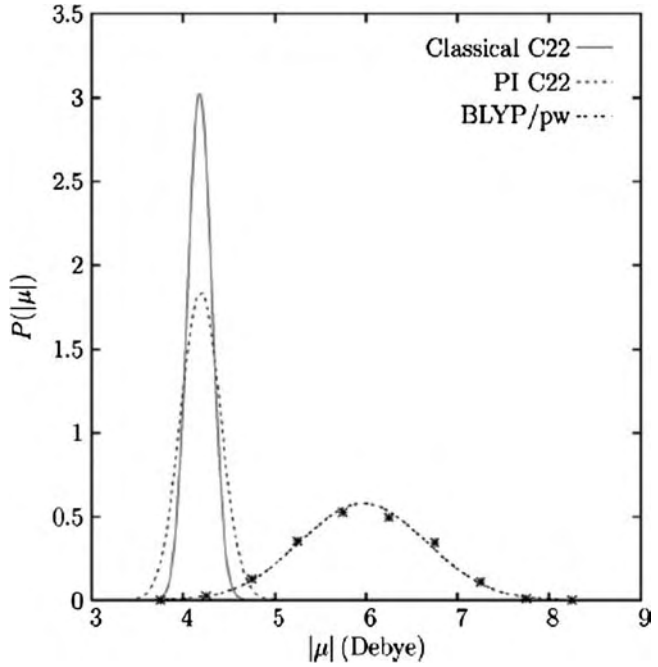


Fig. 13 Dipole enhancement due to hydrogen bond order in neat liquid NMA. From Structural properties of liquid N-methylacetamide *via ab initio*, path integral, and classical molecular dynamics, T. W. Whitfield, J. Crain and G. J. Martyna, *J. Chem. Phys.*, **124**, 094503. Copyright 2006 The American Institute of Physics.

the energy difference is only a factor of 4, so the difference between cis and trans configuration is minimal.

Very naturally, NMA also provides a basis for exploring the electronic properties and polarizability of the peptide bond in a variety of different environments. A key result from molecular dynamics simulations comes from comparison of the molecular dipoles obtained from the ab initio density functional theory simulations to the results of classical models without polarization. The *ab initio* dipole distribution is found to be much broader and the mean value increased substantially by 42%, compared with the nonpolarizable C22 or PIMD C22 models. The average molecular dipole in the polarizable liquid is approximately 6 D, while the experimental gas phase value is 3.73 ± 0.08 D in agreement with the 3.74 D value calculated under BLYP/pw for the structurally optimized monomer. Under the C22 force field, 10 each trans-NMA molecule has an average dipole of approximately 4.2 D.

C. Tertiary contact models in free peptides

The tertiary structure of a protein is pinned by interactions between the the side chains, the R groups. These include (1) ionic interactions/salt bridges which result from the neutralization of an acid and amine on side chains. Some amino acids (such as aspartic acid and glutamic acid) contain an extra -COOH group. Some amino acids (such as lysine) contain an extra $-NH_2$ group. Transfer of a hydrogen ion from the -COOH to the $-NH_2$

group to form zwitterions is possible just as in simple amino acids. Ionic bonds between the negative and the positive groups will form if the chains are folded in such a way that they are close to each other. Ionic bonds between the positive ammonium group and the negative acid group can also form (2) hydrogen bonds between side groups - not between groups actually in the backbone of the chain. Many amino acids contain side chains which have a hydrogen atom attached to either an oxygen or a nitrogen atom. This is a classic situation where hydrogen bonding can occur. For example, the amino acid serine contains an -OH group in the side chain. Hydrogen bonds can exist between two serine residues in different parts of a folded chain. (3) van der Waals/dispersion forces: Several amino acids have large hydrocarbon groups in their side chains. Temporary fluctuating dipoles in one of these groups induce opposite dipoles in another group on a nearby folded chain. The dispersion forces stabilze the tertiary contacts. (4) Disulfide bonds can be formed by oxidation of the sulfhydryl groups on cysteine. (5) Non-Polar hydrophobic interactions: The hydrophobic interactions of nonpolar side chains are believed to contribute significantly to the stabilizing of the tertiary structures (as well as other fold motifs) in proteins (Fig. 14).

Here, a model system is not one which approximates the full three-dimensional fold but rather one that captures the essential physics/chemistry of some particular structural or functional aspect. As we have seen, the question of the extent to which short free peptides can do that is a matter of continuing debate directly linked to the unresolved matter of the relative importance of local *vs*. non-local interactions – see for example Clementi *et al*.[70] While there may be no general answer to this question, there are emerging examples where aspects of structure and maybe even function of proteins are retained in free peptides without the support of the fully folded tertiary contacts. In that sense, even rather complex processes may be amenable to investigation by highly simplified models.

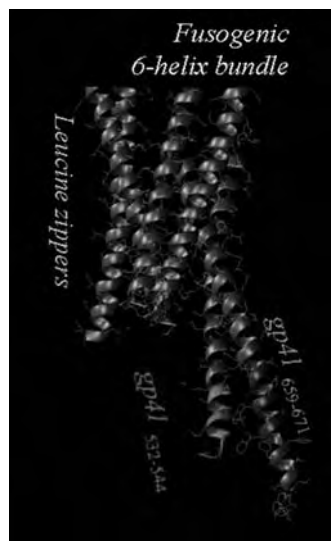


Fig. 14 Showing the tertiary contacts of particualar peptide spans in the HIV viral protein gp41.

As an example, tertiary contacts in the helical peptide bundles relevant in HIV infection include those between the C-terminal membrane proximal external (MPE) and N-terminal fusion peptide proximal (FPP) regions.[71] Their mutual interaction and the extent to which regulation of the local structure occurs *via* tertiary contacts in the native fold motif are important aspects of the formation and properties of the assembly. In principle, a simplified or minimal system consisting only of short peptides spanned by the MPE and FPP regions and a model membrane may exhibit properties that are relevant in the biological context but are amenable to direct structural study and computer simulation[72] as a function of local environment.

In particular free peptides can be chosen to span regions that are widely separated in the primary sequence and can thereby be useful as minimal model systems for tertiary contacts and their local interactions. In this regard the pair of short (13 residues) peptides $gp_{41}^{659–671}$ and $gp_{41}^{532–544}$ are of particular interest: individually, $gp_{41}^{659–671}$ contains the sequential epitope for the monoclonal antibody 2F5 which is thought to block MPE–FPP interactions in the hemifusion phase and has been comprehensively studied under a variety of conditions including solution, micellar and phospholipid environments[73–76] and by MD simulation.[77,78]

The FPP fragment $gp_{41}^{535–539}$ is a highly conserved motif across all HIV-1, HIV-2 and SIV isolates. Both of these spans are terminal extensions of the corresponding heptad repeats. Crucially, in the envelope spike, these stretches are in close proximity, where they are almost aligned sharing the viral membrane environment. This tertiary contact is also evident in the crystal structure determination (Fig. 15).[79]

It is found that free peptides corresponding to the FPP fragment ($gp_{41}^{532–544}$) exert regulatory control over the MPE fragment ($gp_{41}^{659–671}$). Interstingly, the process takes place exclusively in membrane-bound environments: The binding modulation between the two peptides seen in anionic membranes is associated with helix formation. This helical bias is not observed for either of the individual monomers regardless of the membrane environment or their mixtures under different conditions. This observation has led to a proposed membrane-mediated reciprocal interaction between the pair. While the observed modulation observed in isolated peptides requires a membrane environment, it occurs without the extensive conformational support of flanking heptad repeats (HR1 and HR2) present in the tertiary folds of the real protein.

Another example of the use of peptide fragments as models for tertiary or highher order contacts is that of the molecular basis of antibody recognition. As an example, Human chorionic gonadotropin (hCG) is an important biomarker in pregnancy and oncology, and is among the hormones most frequently detected and quantified by specific immunoassays.[81] Intelligent epitope selection is essential to achieving the required assay performance. Recently it has been shown that a loop-specific monoclonal antibody (8G5) can distinguish between $hCG_{66–80}$ and the closely related luteinizing hormone (LH) fragment $LH_{86–100}$, which differ only by a single amino acid residue.[82] Optical spectroscopic measurements and atomistic computer simulations on the free peptides has revealed differences in turn type. It has

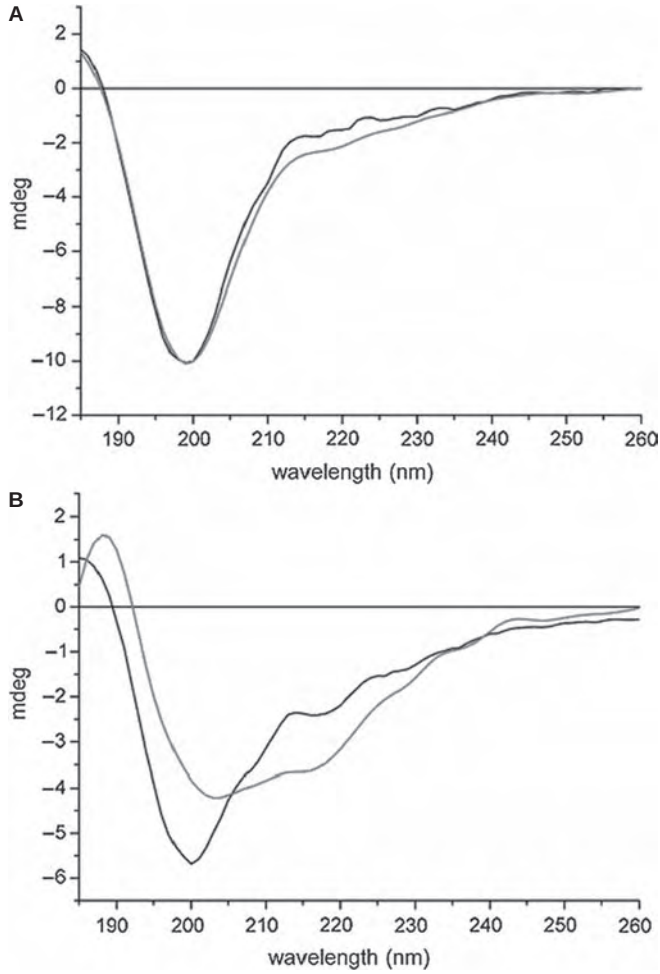


Fig. 15 Circular dichroism spectra showing the perturbation to the secondary structure and rudimentary regulatory behaviour in gp41 fragments in the vicinity of a model membrane. Membrane mediated regulation in free peptides of HIV-1 gp41: minimal modulation of the hemifusion phase. From E. Cerasoli, J. Ravi, C. Gregor, R. Hussain, G. Siligardi, G. Martyna, J. Crain and M. G. Ryadnov, *Phys. Chem. Chem. Phys.*, **14**, 1277. Copyright 2012 The Royal Society of Chemistry.

been suggested that these turn motifs are stabilized by specific hydrogen bonds. The suggestion is that these structural differences are the basis for the observed selectivity in the full protein.

The presence of the hydrogen bond and type I beta-turn (Fig. 16) has been proposed as the molecular origin of the molecular recognition mechanism between the subunits of hCG and LH. These structural differences apprear to persist in the free peptides.

5 Future perspectives

Areas of future research where physics, chemistry and medicine may overlap are expanding and there are numerous potential opportunities to

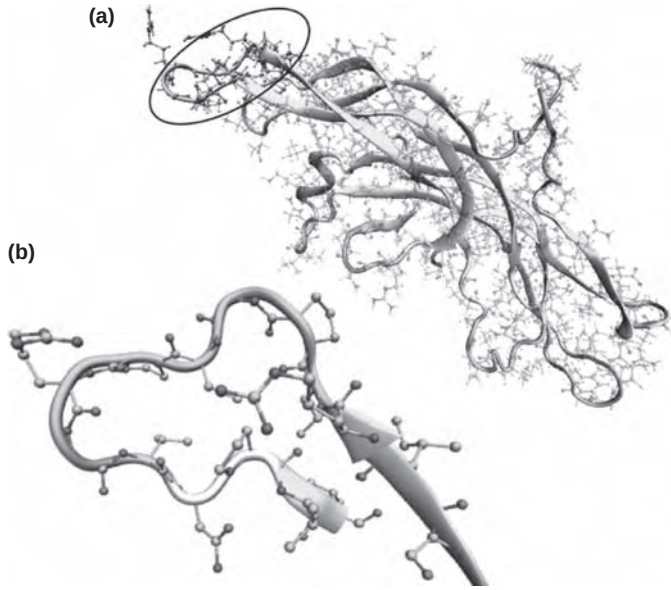


Fig. 16 A small fragment of hCG implicated in the molecular recognition mechanism of monoclonal antibody (8G5). From Antibody Recognition of a Human Chorionic Gonadotropin Epitope (hCG beta(66-80)) Depends on Local Structure Retained in the Free Peptide C. R. Gregor, E. Cerasoli, J. Schouten, J. Ravi, J. Slootstra, A. Horgan, G. J. Martyna, M. G. Ryadnov, P. Davis and J. Crain, *J. Biol. Chem.*, **286**, 25016. Copyright 2011 The American Society for Biochemistry.

develop further the notions discussed in this article. In particular, systems related to peptide-membrane interactions are likely to be of particular interest. From the physical science perspective there are numerous problems of medical relevance defined by an amphipathic, positive charge peptide at physiological pH in contact with a real or model membrane. Despite striking sequence similarity and conformational restrictions to common secondary structure elements (such as helices and beta strands), particularly when in close proximity to the bilayer interface or interior a vast array of processes emerge. These include antimicrobial functionality thought to function by compromising the integrity of the cell membrane, either by forming membrane pores or by inducing non-lamellar structures. As antibiotic resistance increases worldwide, there is an increasing pressure to develop novel classes of antimicrobial compounds to fight infectious disease. Peptide therapeutics represent a novel class of therapeutic agents. Some, such as cationic antimicrobial peptides and have been identified from studies of innate immune effector mechanisms, while others are completely novel compounds generated in biological systems.[83] Currently, only selected cationic antimicrobial peptides have been licensed, and only for topical applications. However, research which exposes the molecular modes of interaction may result in an increased range of peptide therapeutics available in the clinic for broader applications. In viral fusion, peptide-membrane interactions are critical structural components wherein peptides insert into lipid bilayers and act as anchors bringing two lipid bilayers into proximity, promoting infection. Peptide membrane interactions are also

exploited in cell penetration wheren peptide membrane interactions are tuned to facilitate the translocation of small cargo molecules including DNA fragments into cells through the membrane. The process of introducing drugs into cells has always proved a major challenge for the pharmaceutical industry. The cell membrane is selectively permeable and supports no generic mechanism for delivery. These restrictions severely limit the palette of possible drug molecules. CPPs are capable of delivering biologically active cargo to the cell interior and hold significant potential in the field of therapeutics. Attached to a CPP, therapeutic cargo could be delivered to an intracellular target, and circumventing the restrictions normally imposed by the membrane. The mode of action remains controversial and detailed structural data is required to elucidate their entry and transport mechanisms. Many of these themes are covered in detail in other contributions to this volume particularly those concerning peptide design,[84] amino acid and peptide bioconjugates,[85] membrane-active proteins[86] and intrinsically unstructured proteins.[87]

Tackling a problem of this type requires the application of novel metrology, namely atomistic resolution (sub-Å) measurement, to define not only the structural elements of the peptides precisely, but also to characterise the extent of the membrane perturbation induced. Single technique measurement solutions are ill-suited to this broad measurement challenge which spans atomistic to intermediate-range (mesoscopic) lengthscales. Tightly coordinated metrology input is required for advances to addressing the underlying science and technology questions.

References

1 Udo Heinemann, *J. Mol Med*, 2000, **78**, 245–246 (DOI) 10.1007/s001090000120.

2 K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent and T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution., *Nature*, 1997, **389**, 251–260.

3 Claudine E. Williams, Mohamed Daoud (Eds.), *Soft Matter Physics*, Springer.

4 S. Dixit, J. Crain, W. C. K. Poon, J. L. Finney and A. K. Soper, Molecular segregation observed in a concntrated alcohol-water solution., *Nature*, 2002, **416**, 829.

5 S. Dixit, A. K. Soper, J. L. Finney and J. Crain, Water structure and solute association in dilute aqueous methanol, *Europhys. Lett*., 2002, **59**, 377.

6 S. Dixit, W. C. K. Poon and J. Crain, Hydration of methanol in aqueous solutions: a raman spectrscopy study., *J. Phys. Condensed Matter*, 2000, **12**, L323.

7 C. J. Adkins, (1968/1975). *Equilibrium Thermodynamics*, second edition, McGraw-Hill, London, ISBN 0-07-084057-1.

8 D. Kondepudi (2008). *Introduction to Modern Thermodynamics*, Wiley, Chichester, ISBN 978-0-470-01598-8, pages 197–199.

9 F. Wang, S. Saeki and T. Yamaguchi, *Polymer*, 1999, **40**, 2779–2785.

10 P. Wiggins, Hydrophobic hydration, hydrophobic forces and protein folding., *Physica A*, 1997, **238**, 113.

11 S. Garde, G. Hummer, A. E. Garcia, L. Pratt and M. E. Paulaitis, Hydrophobic hydration:inhomogeneous water structure near nonpolar molecular solutes., *Phys. Rev. E*, 1996, **53**, 4310.

12 L. Pratt and A. Pohorille, Hydrophobic effects and modeling of biophysical aqueous solution interfaces, *Chem. Rev*, 2002, **102**, 2671.
13 Peter L. Freddolino, Christopher B. Harrison, Yanxin Liu and Klaus Schulten, *J Nature Physics*, OCTOBER 2010, Vol. **6**.
14 Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang and Peter Kollmann, *Journal of Computational Chemistry*, 2003, **24**, 1999.
15 A. D. MacKerell, Jr., D. Bashford and M. Bellott *et al*., *J. Phys. Chem. B*, 1998, **102**, 3586.
16 W. R. P. Scott, P. H. Huenenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krueger and W. F. van Gunsteren, The GROMOS Biomolecular Simulation Program Package, *J. Phys. Chem. A,* **103**, 3596–3607.
17 W. L. Jorgensen and J. Tirado-Rives, ''The OPLS Force Field for Proteins Energy Minimizations for Crystals of Cyclic Peptides and Crambin''., *J. Am. Chem. Soc*, 1988, **110**(6), 1657–1666.
18 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, ''Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids'', *J. Am. Chem. Soc*, 1996, **118**(45), 11225–11236.
19 S. A. Adcock and J. A. McCammon, Molecular dynamics: Survey of methodsfor simulating the activity of proteins., *Chem. Rev*., 2006, **106**, 1589–1615.
20 C. L. Brooks, J. N. Onuchic and D. J. Wales, Taking a walk on a landscape, *Science*, 2001, **293**, 612–613.
21 K. A. Dill and H. S. Chan, From levinthal to pathways to funnels, *Nature Struc. Bio*., 1997, **4**, 10–19.
22 C. Simmerling, B. Strockbine and A. E. Roitberg, All-atom structure prediction and folding simulations of a stable protein, *J. Am. Chem. Soc*., 2002, **124**, 11258–11259.
23 H. Lei, X. Deng, Z. Wang and Y. Duan, The fast-folding HP35 double mutant has a substantially reduced primary folding free energy barrier., *J. Chem. Phys*., 2008, **129**, 155104.
24 N. D. Socci, H. Nymeyer and J. N. Onuchic, *Exploring the protein folding landscape Physica D*, 1997, **107**, 366–382.
25 M. R. Betancourt and J. N. Onuchic, Kinetics of proteinlike models: the energy landscape factors that determine folding, *J. Chem. Phys*., 1995, **103**, 773–787.
26 M. Wang, *et al*., Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale, *J. Am. Chem. Soc*., 2003, **125**, 6032–6033.
27 J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton and J. Hofrichter, Sub-microsecond protein folding., *J. Mol. Biol*., 2006, **359**, 546–553.
28 Y. Sugita and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett*., 1999, **314**, 141–151.
29 Y. Sugita and Y. Okamoto, Replica-exchange multicanonical algorithm and multicanonical replicaexchange method for simulating systems with rough energy landscape, *Chem. Phys. Lett*., 2000, **329**, 261–270.
30 T. W. Whitfield, L. Bu and J. E. Straub, Generalized parallel sampling, *Physica A-statistical Mechanics and Its Applications - PHYSICA A*, 2002, **305**(1), 157–171.
31 Pu Liu, Byungchan Kim, Richard A. Friesner and B. J. Berne, Replica exchange with solute tempering: A method for sampling biological systems in explicit water, *Proceedings of the National Academy of Sciences*, 2005, **102**, 13749–13754.

32 B. Zagrovic, C. D. Snow, M. R. Shirts and V. S. Pande, Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing, *J. Mol. Biol.*, 2002, **323**, 927–937.
33 D. Paschek and A. E. Garcia, Reversible temperature and pressure denaturation of a protein fragment: A replica exchange molecular dynamics simulation study, *Phys. Rev. Lett*, 2004, Vol. **93**.
34 G. Jayachandran, V. Vishal and V. S. Pande, Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece., *J. Chem. Phys.*, 2006, **124**, 164902.
35 A. Gara, *et al.*, Overview of the Blue Gene/L system architecture, *IBM Journal of Research and Development*, 2005, **49**(2/3), 195–212.
36 Robert S. Germain, Blake Fitch, Aleksandr Rayshubskiy, Maria Eleftheriou, Michael C. Pitman, Frank Suits, Mark Giampapa, and T. J. Christopher Ward. Blue Matter on Blue Gene/L: massively parallel computation for biomolecular simulation. In CODES + ISSS '05: Proceedings of the 3rd IEEE/ACM/IFIP international conference on Hardware/software codesign and system synthesis, pages 207–212, New York, NY, USA, 2005. ACM Press.
37 Y. Duan and P. A. Kollman, Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science*, 1998, **282**, 740–744.
38 Vijay S. Pande, Ian Baker, Jarrod Chapman, Sidney P. Elmer, Siraj Khaliq, Stefan M. Larson, Young Min Rhee, Michael R. Shirts, Christopher D. Snow, Eric J. Sorin and Bojan Zagrovic, *Biopolymers*, 2003, **68**(Issue 1), 91–109.
39 H. Lei and Y. Duan, *J. Chem. Phys.*, 2004, **121**, 12104.
40 W. Kwak and U. H. E. Hansmann, Efficient sampling of protein structures by model hopping, *Phys. Rev. Lett.*, 2005, **95**, 138102.
41 S. Gnanakaran and A. E. Garcia, Helix-coil transition of alanine peptides in water: Force field dependence on the folded and unfolded structures, *Proteins-Structure Function and Bioinformatics*, 2005, **59**(4), 773–782.
42 R. Zhou, *Proteins*, 2003, **53**, 148.
43 E. Harder, A. D. MacKerell and B. Roux, *J. Am. Chem. Soc.*, 2009, **131**, 2760.
44 A. Jones, A. Thompson, J. Crain, M. H. Muser and G. J. Martyna, *Phys. Rev. B*, 2009, **79**, 144119.
45 David D. Laws, Hans-Marcus L. Bitter, and Alexej Jerschow, "Solid-State NMR Spectroscopic Methods in Chemistry", *Angewandte Chemie International Edition (engl.)*, (2002) Vol. 41 , pp. 3096.
46 H. S. Frank and M. W. Evans, Free volume and entropy in condensed systems, *J. Chem. Phys*, 1945, **13**, 507.
47 T. Koop, B. Luo, A. Tsias and T. Peter, Water activity as the determinant for homogeneous ice nucleation in aqueous solutions., *Nature*, 2000, **406**, 611.
48 J. L. Finney and A. K. Soper, Solvent structure and perturbations in solutions of chemical and biological importance., *Chem. Soc. Reviews*, 1994, **70**, 1.
49 A. K. Soper, Empirical potential monte carlo simulation of liquid structure, *Chem. Phys.*, 1996, **202**, 295.
50 R. Paquin and G. Pelletier, *Physiolog. Veg*, 1981, **19**, 103.
51 K. Koster and D. Lynch, *Plant Physiol.*, 1992, **98**, 108.
52 T. M. Nanjo, Y. Kobayashi, Y. Yoshiba, K. Kakubarri, K. Yamaguchi-Shinozaki and K. Shinozaki, *FEBS Lett.*, 1999, **461**, 205.
53 B. Schobert and H. Tschesche, *Biochim Biophys Acta.*, 1978, **541**, 270.
54 A. Rudolph and J. Crowe, *Biophys. Journal*, 1986, **50**, 423.
55 D. Samuel, T. Krishnaswamy, S. Kumar, G. Ganesh, G. Jayaraman, P. Yang, M. Cang, V. Trevidi, S. Wang, K. Hwang, *et al.*, *Protein Science*, 2000, **9**, 344.

56 R. Ma, T. He and C. Wang, *J. Chem. Phys*., 2003, **88**, 1497.
57 M. L. Huggins, *Chem. Rev*., 1943, **32**, 195.
58 G. C. Pimentel, A. L. McClellan, *The Hydrogen Bond*, 1960, W. H. Freeman and Co., San Francisco.
59 R. Taylor and O. Kennard, *J. Am. Chem. Soc*., 1982, **104**, 5063.
60 Z. S. Derewenda, L. Lee and U. Derewenda, *J. Mol. Biol*., 1995, **252**, 248.
61 G. F. Fabiola, S. Krishnaswamy, V. Nagaragan and V. Pattabhi, *D. Acta Cryst*., 1997, **53**, 316.
62 D. Paschek, R. Day and A. E. Garcia, Influence of water-protein hydrogen bonding on the stability of Trp-Cage miniprotein: A comparison between the TIP3P and TIP4P-Ew water models''. *Phys Chem Chem Phys*., 2011.
63 K. B. Wibreg and D. J. Rush, *J. Org. Chem*., 2002, **67**, 826.
64 S. K. Gregurick, G. M. Chaban and R. B. Gerber, *J. Phys. Chem. A*, 2002, **106**, 8696.
65 Y. A. Mantz, H. Gerard, R. Iftimiees and G. J. Martyna, *J. Am. Chem. Soc*., 2004, **126**, 4080.
66 T. W. Whitfield, G. J. Martyna, S. Allison, S. P. Bates, H. Vass and J. Crain, Structure and Hydrogen Bonding in Neat N -Methylacetamide: Classical Molecular Dynamics and Raman Spectroscopy Studies of a Liquid of Peptidic Fragments, *J. Phys. Chem B*, 2006, **110**(no. 8), 3624.
67 T. W. Whitfield, J. Crain and G. J. Martyna, *J. Chem Phys*, 2006, **124**, 094503.
68 M. A. Williams, J. M. Goodfellow and J. M. Thornton, *Protein Sci*., 1994, **3**, 1224.
69 V. Denisov, K. Venu, J. Peters, H. D. Hörlein and B. J. Halle, *Phys. Chem. B*, 1997, **101**, 9380.
70 C. Clementi, A. E. Garcia and J. N. Onuchic, Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: All-atom representation study of protein Llink, *Journal of Molecular Biology*, 2003, **326**(3), 933–954.
71 M. J. Root, M. S. Kay and P. S. Kim, *Science*, 2001, **291**, 884.
72 C. R. Gregor, E. Cerasoli, P. R. Tulip, M. G. Ryadnov, G. J. Martyna and J. Crain, *Phys. Chem. Chem. Phys*., 2011, **13**, 12.
73 J. Coutant, H. Yu, M. J. Clement, A. Alfsen, F. Toma, P. A. Curmi and M. Bomsel, *FASEB J*, 2008, **22**, 4338.
74 Z. Biron, S. Khare, S. R. Quadt, Y. Hayek, F. Naider and J. Anglister, *Biochemistry*, 2005, **44**, 13602.
75 F. M. Brunel, M. B. Zwick, R. M. Cardoso, J. D. Nelson, I. A. Wilson, D. R. Burton and P. E. Dawson, *J. Virol*., 2006, **80**, 1680.
76 G. Barbato, E. Bianchi, P. Ingallinella, W. H. Hurni, M. D. Miller, G. Ciliberto, R. Cortese, R. Bazzo, J. W. Shiver and A. Pessi, *J. Mol. Biol*., 2003, **330**, 1101.
77 C. R. Gregor, E. Cerasoli, P. R. Tulip, M. G. Ryadnov, G. J. Martyna and J. Crain, *Phys. Chem. Chem. Phys*., 2011, **13**, 127.
78 P. R. Tulip, C. R. Gregor, R. Z. Troitzsch, G. J. Martyna, E. Cerasoli, G. Tranter and J. Crain, *J. Phys. Chem. B*., 2010, **114**, 7942.
79 V. Buzon, G. Natrajan, D. Schibli, F. Campelo, M. M. Kozlov and W. Weissenhorn, *PLoS Pathog*., 2010, **6**, 1000880.
80 Eleonora Cerasoli Jascindra Ravi, Craig Gregor Rohanah Hussain Giuliano Siligardi, Glenn Martyna Jason Crain and G. Maxim, *Ryadnov Phys. Chem. Chem. Phys*, 2012, **14**, 1277–1285.
81 C. M. Sturgeon and E. J. McAllister, *Ann. Clin. Biochem*., 1998, **35**, 460–491.
82 Craig R. Gregor, Eleonora Cerasoli, James Schouten, Jascindra Ravi, Jerry Slootstra, Adrian Horgan, Glenn J. Martyna, Maxim G. Ryadnov, Paul Davis

and Jason Crain, *The Journal of Biological Chemistry*, July 15, 2011, **286-**(NO. 28), 25016–25026.
83 Michael Zasloff, *Nature*, 2002, **415**.
84 Ryadnov Prescriptive peptide design, AAPP vol 37.
85 Hudecz, F Amino acid and peptide bioconjugates: synthesis, structure and function AAPP vol 37.
86 Watts, NMR of membrane-active proteins AAPP vol 37.
87 Mac Phee CE Intrinsically unstructured proteins AAPP vol 37.

Protein nanotubes, channels and cages

Jonathan G. Heddle*[a] and Jeremy R. H. Tame[b]
DOI: 10.1039/9781849734677-00151

1 Introduction

One of the most exciting concepts in recent decades has been the idea of building complex structures, materials and even machines at the nanometer scale. This idea, now known as ''nanotechnology'', was first brought to wide attention by Richard Feynman in his now legendary talk at Caltech in December of 1959, entitled ''There's Plenty of Room at the Bottom''.[1] In it, Feynman imagined that robot-like nanomachines could be constructed. Such a technology would be revolutionary, most obviously in manufacturing, heralding the advent of atomically precise, defect-free products. But the potential extends further to areas such as health: Complex, autonomous nano-devices would be the ultimate smart drugs; able to patrol the body, sampling the environment for information on threats such as cancer, or invading bacteria and carry out actions based on a software code. We can define three types of nanotechnology (see Fig. 1). One, which we may call ''hard nanotechnology'' is non-biological and builds structures exactly analogous to the components of large-scale machines (such as gears or axles) by precise placement of individual atoms. The second nanotechnology (''bionanotechnology'') is on a slightly larger scale and uses biological materials such as proteins and DNA which are engineered to form the required structures. The third nanotechnology is cell-based synthetic biology where individual cells themselves are used as ''nanofactories'' which are reprogrammed to synthesize desired materials. This chapter is concerned with bionanotechnology and focuses in particular on the use of proteins. Because this area of research is still at a very early stage it cannot strictly be classified as a technology and we prefer to call it ''bionanoscience''.

Synthetic biology is now a widely researched field which shows much promise.[4] Here, the aim is to manipulate pathways within cells, reprogramming them so that they behave in a desired fashion such as synthesizing a novel compound. A common goal is the production of ''modules'' with particular functions that can be used interchangeably in the design of new ''circuitry'' within the cell. Synthetic biology therefore, while it may involve structural redesign of single proteins is generally concerned with higher levels of abstraction. In general it also relies on existing cellular machinery, meaning that there are limitations as to what can be manufactured due to practical considerations such as the metabolic enzymes that are available in the cell, restrictions imposed by limited cell volume *etc*. Using a bionanoscience approach, some of these limitations can be overcome by building

[a]*Heddle Initiative Research Unit, RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. E-mail: heddle@riken.jp*
[b]*Protein Design Laboratory, Yokohama City University, 1-7—29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045 Japan*

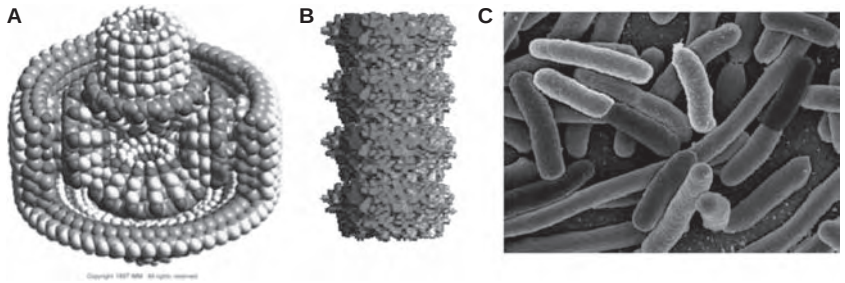


Fig. 1 The three types of nanotechnology. **A:** Hard nanotechnology such as this design for a friction-free bearing made from individual atoms (copyright institute for molecular manufacturing 2012 (www.imm.org)). **B:** Wet nanotechnology which uses self-assembled biological molecules to make new structures such as this model of a protein nanotube[2] (model constructed from PDB ID: 1qaw[3]). **C:** Synthetic biology in which whole cells (such as the *E. coli* cells shown here) are used as factories to synthesize new products.

nanodevice parts *via* the engineering of changes into biological molecules and then self-assembling them extracellularly into structures which may not be viable or possible in intracellular environments. Thus, new structures with new capabilities can be produced.

Hard nanotechnology offers much but is presently a purely theoretical endeavour; it is yet unclear if single atoms can be manoeuvred in three dimensions with the required accuracy to produce complex machine parts. Indeed, the turn of the century saw fierce criticism of the idea by the Nobel Prize winner Richard Smalley who believed that it was not feasible.[5] Even if hard nanomachines can exist, the problem of developing a production technology that can manipulate thousands or millions of single atoms individually with sub-Ångstrom accuracy still remains. In its original form, it was presumed that hard nanotechnology devices would be achieved firstly through the construction of ''assemblers'', nanoscale robots built from non-biological material (usually diamond) which would be able to manipulate individual atoms to construct copies of themselves. Enough nano-assemblers would potentially be able to construct any material with atomic precision and zero defects, ushering in an era of manufacturing in which products are ''grown.'' This form of nanotechnology can be classed as ''hard'' nanotechnology both because of the materials typically used and the almost limitless possibilities that it envisages for the technology. Hard nanotechnology has experienced a number of difficulties in reaching its goal. Most notably the ''Catch-22'' problems associated with how to manipulate individual atoms to construct first generation nano-assemblers without using nano-assemblers, or with attempting to produce highly complex nanoassemblers *via* traditional ''top-down'' methods. While individual atoms can be moved around on surfaces with high accuracy using the tip of a scanning tunnelling microscope (STM),[6] this is an extremely time consuming process and has only ever been used to produce simple patterns consisting of relatively few atoms. Nanoassembler designs call for the precise arrangement of millions of atoms,[7] something which is impractical at present. Furthermore these atoms must be arranged in three dimensions, again this is not possible with current technology. While *in silico* designs

show that, in theory, nanoassemblers, once constructed could be viable,[8] their physical production has not been achieved. Progress is being made towards addressing these goals but it remains to be seen if they will succeed.

An alternative to hard nanotechnology is bionanotechnology (also known as wet nanotechnology or soft nanotechnology). Bionanotechnology piggy-backs on systems existing in nature to overcome many of the hurdles associated with hard nanotechnology. In Nature there already exist numerous precisely defined nanostructures that carry out specific and often highly complex tasks. DNA for example has a well-defined structure and in living cells encodes the sequence of diverse ''machine-like'' proteins including those which maintain, repair and copy the DNA itself. Thus, given a source of energy and raw molecular materials, the DNA and proteins can replicate themselves. Another key point is that the top-down construction problem associated with hard nanotechnology does not exist in bionanotechnology - even complex proteins can spontaneously self assemble into functional structures. Naturally occurring proteins have a wide range of roles including as nanomechanical components (for example actin, which is a major part of eukaryotic cytoskeleton and helps to maintain the structure of the cell), and proteins that act like nanomachines (such as toposiomerases, which are able to carry out intricate, cutting, twisting and ligating of DNA in order to introduce or remove supercoils[9]). Compared to hard nanotechnology systems, bionanotechnology is more limited – it works only in aqueous solutions at a relatively constricted range of temperatures, pressures and ionic strengths and the biological materials themselves may not be suitable for many purposes (such as highly temperature-resistant materials or electrical conductors). The key advantages for bionanosystems are that their practicality has been demonstrated by life processes in living cells and that systems for producing DNA and protein are well characterized, widely available and trivial to implement. Furthermore, the limitations of bionano devices may not be relevant for the likely environments in which they will function. Given the limitations noted above, it is not realistic for example to imagine a consumer electronic device with transistors constructed from DNA or protein. Such transistors would be far slower, less robust and much more expensive than their semiconductor counterparts. However, there are extremely important niche applications where the biological nature of bionano devices would make them ideal candidates, particularly in biomedicine. Smart biosensors for example could benefit from protein or nucleic acid components that directly detect disease-associated signals *in vivo* and perform an appropriate response. They may also provide an important role as a bridging technology where they recognize a biological signal and transduce it to a silicon-based human-machine interface. While unsuitable for direct use as electronic components, biomolecules can help to overcome the self-assembly problem of hard nanotechnology by producing self-assembled scaffolds or moulds which are then used as templates for the assembly of non-biological material. The biological component can be subsequently removed, if necessary, leaving the final nanostructure.

2 Current trends

The sophistication of proteins as components of artificially produced nanostructures is increasingly being recognized. The earliest molecular models of protein structures determined by X-ray crystallography were generally interpreted as being rigid. The finding that hemoglobin, the oxygen-carrier of the blood, can adopt two different structures was a great surprise, and there was a belief one ''clicked'' into the other. It is now appreciated that proteins are soft, flexible molecules, and this flexibility is essential to the function of enzymes, enabling them to work as true ''nanomachines''. This flexibility and sophistication means that proteins are perhaps the most promising material for use in bionanotechnology. While the structure and interactions of DNA are extremely well defined and understood, this very simplicity means that DNA is unlikely to be able to produce components as structurally and mechanistically complex as naturally occurring nanomachines such as enzymes. Great strides have been made in the production of programmable two- and three-dimensional DNA structures, but these are largely fixed structures with no programmed function. Rather than interactions between four base pairs, proteins consist of 20 naturally occurring amino acids with very different chemical characters and many possible interactions. This makes predicting the structure of any amino acid sequence of significant length computationally extremely difficult (the ''Levinthal Paradox''[10]), but it also endows proteins with the ability to make structures of astonishing sophistication and complexity. With the advances of solid phase peptide synthesis it is now possible to create easily peptides of significant length containing non-natural amino acids with novel characteristics at chosen points in the sequence, and this has further expanded the repertoire of protein building blocks. Such amino acids may be useful to introduce new functionality at an active site, but with few examples of structures containing these amino acids, it will be difficult to use them to control precisely the shape of a folded protein. There is also a question of cost, since the production of a protein on a large-scale is in general much cheaper with a bacterial expression system, limited to natural amino acids.

Somewhat less ambitious than producing novel proteins of enzyme-level complexity is the use of known protein domains in templating nanometric electronic components. This is attractive because of the requirement of the electronics industry for ever-smaller components coupled with the limitations associated with traditional (lithographic) production techniques. Transistors in electronic components such as microprocessors have continued to decrease in size following Moore's law,[11] allowing faster computing devices. The traditional method of producing the features on the chip is a top-down approach which involves etching with light.[12] However, as feature sizes decrease to only a few tens of nanometres in size or smaller, a limit is reached due to the wavelength of light that can be used.[13] In place of this top-down approach, proteins could potentially self-assemble into designed structures only a small number of nanometres in diameter, from the bottom-up. If such proteins could act as a template for mineralisation by the required semiconductor or other material then they may enable the construction of electronic components of a much smaller size than is currently possible.

Using proteins to build artificial structures for specific tasks is still in its infancy. As protein structures cannot readily be designed *de novo* from amino acid sequences, the most attractive option is to find proteins already existing in nature and utilise them accordingly by exploiting their existing abilities (as is the case for the glucose oxidase used in some blood sugar monitoring devices[14]). A second, more ambitious route is to take existing proteins and engineer them so that they form new structures with new capabilities. To date, research has tended to focus on simple but potentially useful protein fibres. Such fibres are formed by many proteins, and a whole class of diseases arises from protein misfolding into these very stable structures Engineered protein fibres may be useful as templates for nanowires and gels, but for many purposes a hollow structure is required, such as nanotubes, nanochannels or spheres. Cavity-containing artificial nanostructures have two main potential uses – the first is in cases where the cavity is filled with inorganic material having particular magneto-electronic properties, such as conducting, semiconducting or magnetic material. If such proteins are integrated into larger devices, they may then be able to act as conducting nanowires, quantum dots, contrast agents or device components. The second useful cargo is a therapeutic agent. Drug delivery to specific targets is a major area of research, as it may hugely enhance the beneficial effect of a given dose. Placing a drug in the interior cavity of a protein shell may protect it from degradation and may protect the patient receiving it from toxic effects until the target is reached. Specificity (and further reduction of side effects) could be achieved by modification of the outer protein surface with moieties providing localisation to a specific cell/tissue type (*e.g.* an antibody fragment), resulting in a targeted drug delivery vehicle. Tube-like proteins that span membranes (channel proteins) have additional potential in controlling the flow of signals and molecules between isolated volumes. A brief example of a protein nanofibre followed by examples of both natural and artificial protein nanotubes, nanospheres and channels are considered in the following sections.

3 Nanofibres (Amyloid-like fibrils)

Fibrils are highly polymerized protein fibres consisting of stacked β-sheets. They are frequently very stable and resistant to proteases or unfolding. As noted below (see section on domain swapping) domain-swapping may lead to fibre-like structures with indefinite translational symmetry, and domain-swapped proteins can lead to protein deposition in living cells, causing a variety of diseases.[15]

Although in the medical field there is huge interest in blocking fibril formation to prevent diseases such as Alzheimer's and Parkinson's, there is also interest in converting readily available proteins into fibrillar form for biotechnology. These fibrils form viscoeleastic gels at low protein concentration which could find use as tissue scaffolds or in texture-enhancing of foods. One study by Pearce and colleagues used heat-denaturation of purified or crude albumin to find conditions suitable for fibril formation.[16] In fact similar aggregates form in normal food processing through protein unfolding and proteolysis. Fibrils of β-lactoglobulin also form under a

variety of conditions at temperatures from 80–120 °C, and pH 1–3. Kinetic factors appear to dictate which of various possible fibrillar forms dominates after processing, and none of these forms appears to be of a highly regular nature suitable for nano-construction.[17]

Pipich and colleagues have shown that egg-white ovalbumin may be unfolded by biomineralisation by calcium carbonate from a nucleation centre,[18] and the same protein can be induced to form fibrils with KCl and NaCl.[19] In neither case though are the resulting structures ordered over significant distances. Highly structured ribbons of amyloid were produced from globular proteins using hen egg-white lysozyme and β-lactoglobulin. AFM and cryo-EM images show parallel fibrils arranged into flat sheets ordered over approximately 100 nm, but the molecular structure is unknown and no attempt has been made to functionalise the ribbons.[20]

Further work remains to be done on the processes involved in amyloid formation before logical progress can be made towards designer fibrils with specific desired properties using naturally-available proteins as the starting material. From the perspective of nanoscale architecture, a more promising approach involves the production of artificial proteins such as block copolymers carrying domains with different association properties, for example peptide and PEG. Various designs have been realized, generally using short peptides. One of the first attempts to use longer sequences was reported by Smeenk and colleagues who used a repeating alanylglycine-rich motif which is capable of forming β-sheet structures.[22] This peptide was expressed as a protein in *E. coli*, and released by cyanogen bromide from N and C-terminal sequences added to assist purification. PEG was attached *via* cysteine residues at each end of the released peptide. The molecule crystallized into well-defined fibrils, with the possibility of introducing specific amino acid residues at turns in the β-sheets to add extra functionality.

More recently Beun and colleagues[21] have demonstrated the self-assembly behaviour of triblock copolymers (Fig. 2) carrying collagen-based domain (called "C") and a silk-based repeat sequence (called "S"). The silk-like sequence has 24 octapeptide repeats with either a glutamic acid, histidine or lysine residue at the end of the octapeptide. The blocks were arranged in the pattern SCCS, with the silk blocks being of the same type. The three proteins have a filamentous structure with stacked β-rolls. Polymerization occurs when the silk block is uncharged, allowing control through pH of the solution. The coblock combines the qualities of silk, which is flexible and strong, and collagen, which has excellent biocompatibility, and therefore of great interest for tissue engineering. The proteins show slow polymerization into fibres, following first-order kinetics, and giving a narrow length distribution, suggesting that growth occurs from nuclei present in solution. AFM shows straight fibres roughly 10 μm long, which can be dissolved by moving the pH far from the pI of the protein. In this way the coblocks can be polymerized and dispersed repeatedly. The authors suggest from the CD spectrum that the protein units do not immediately adopt the β-roll conformation on neutralization, but only on attachment to a growing fibre, mimicking such processes as the growth of bacterial flagellae. Flagellae are naturally-occurring protein tubular

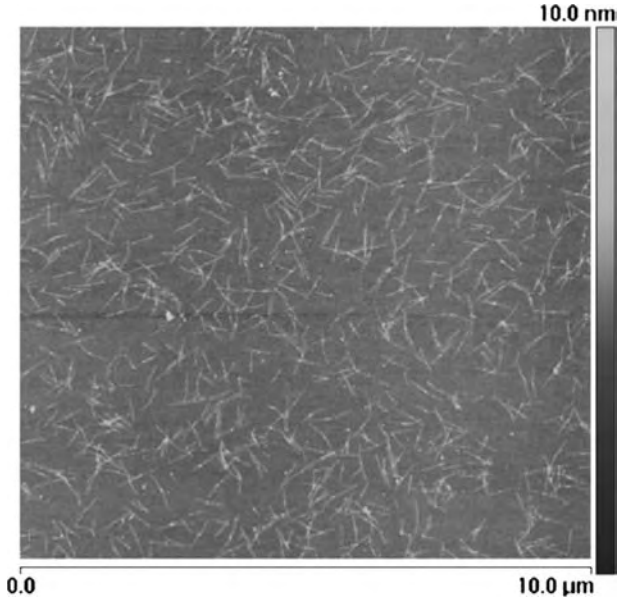


Fig. 2 AFM image showing fibres formed *via* self-assembly.[21] Fibres consist of triblock copolymers carrying a collagen based protein domain and a silk-based amino acid repeat sequence. Reprinted with permission from L. H. Beun, X. J. Beaudoux, J. M. Kleijn, F. A. de Wolf, and M. A. Cohen Stuart, Self-Assembly of Silk-Collagen-like Triblock Copolymers Resembles a Supramolecular Living Polymerization. *ACS Nano*, 2011, **6**, 133–140. Copyright 2012 American Chemical Society.

structures which have also attracted interest as a basic design pattern to be exploited.

4 Nanotubes

Nanotubes are widely recognised as useful structures in nanotechnology. Carbon nanotubes have been very extensively investigated and offer great promise as nanowires[23] and for constructing novel materials[24] with desirable properties such as super conduction and extremely high tensile strength. However, there remain questions about their toxicity.[25] Furthermore, production of carbon nanotubes is energy intensive, typically involving arc discharge or CVD processes. Finally, as every bond in the carbon nanotube structure is fully coordinated, modification of the nanotube (as would be required for biological applications) is difficult to achieve without destroying the structure. Compared to carbon nanotubes, peptide nanotubes may offer greater promise for modification [26,27] and have been used as templates for metallic nanowires[28] but have less structural redundancy when compared to protein nanotubes and thus less potential for extensive modification. Protein nanotubes may be more modifiable than their peptide or carbon counterparts, thus offering greater potential for customization.

As discussed in the previous section, simple protein polymers with a high aspect ratio but lacking a central cavity can be regarded as a precursor to protein nanotubes. Protein polymers are commonly found in nature and if they can be mineralized with conducting or semiconducting materials they

may have the potential to form nano or quantum wires, or other structures such as high density ''electron highways'' in photovoltaic devices. The idea of using proteins to form a nanowire scaffold was first successfully demonstrated by the Lindquist group in 2003, who utilised amyloid fibers which were coated with gold or silver.[29] Since then numerous attempts have been used to produce protein wire scaffolds. However as with the previous work, these usually involve uncontrolled polymerization of a protein without an internal cavity in which conducting elements (typically gold) are attached to the *external surface*. Unfortunately, this provides no control of either the length and shape of the protein wire or the extent of biomineralization of the gold, often resulting in uncontrolled and uneven deposition of metal, making its usefulness for real-life applications questionable.

4.1 Naturally occurring protein nanotubes

Naturally occurring protein nanotubes are not uncommon in nature. Such protein polymers constructed from numerous identical subunits have a tendency to assemble into a hollow helical structure due to the fact that neighbouring subunits are generally slightly rotated relative to one another, resulting in the introduction of a curve as the polymer is extended, and the initial line of subunits meets itself to form a helix. Flagellin, the principal component of bacterial flagellae, is a highly unusual protein which naturally produces curly hollow fibres. Its structure has been the subject of much scrutiny, and it has great medical interest as an antigen, but it has not been very widely used in bionano-technology. Other protein nanotubes that have been used in bionanoscience include tobacco mosaic virus (TMV) and microtubules both of which are considered in more detail below.

4.1.1 Tobacco mosaic virus. Tobacco mosaic virus (TMV; Fig. 3) has a venerable history; not only was it the first virus to be isolated but it in fact

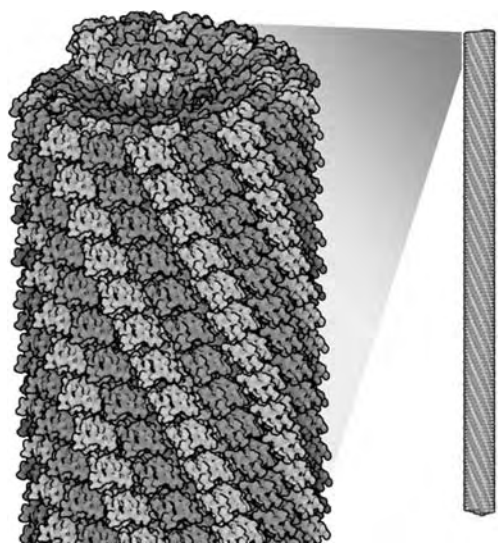


Fig. 3 Image of the helical capsid of tobacco mosaic virus showing repeating protein monomers (purple and cyan) and encapsulated RNA (orange) based on PDB entry 2tmv. Image copyright David S. Goodsell and RCSB PDB. [Colour image available on-line]

played a pivotal role in demonstrating the very existence of viruses themselves. By the end of the 19th century "germ theory" – the idea that infectious diseases were caused by bacteria was well established thanks to the work of scientists such as Louis Pasteur and Robert Koch. However, Russian researcher Dmitrii Iwanowski was able to show that extracts of tobacco plants suffering from the symptoms of TMV infections, passed through a filter small enough to remove bacteria, yet were able to retain their infectivity.[30] This finding is generally regarded to mark the discovery of the viruses. For many years, TMV was at the fore-front of various cutting-edge areas of research including, for example, protein crystallography.[31,32] More recently TMV has become an important focus in bionanoscience research due to the fact that it offers a number of advantages including i) A well characterised function and structure. ii) Monodispersity (see below). iii) It forms a tubular structure which is particularly useful for a number of applications and iv) It can be produced in large quantities (one gram of pure virus can be purified from a kilogram of infected leaf material).

TMV is a RNA virus carrying a single positive-sense strand of RNA. Positive sense refers to the fact that the RNA sequence runs in the same direction as mRNA and can, in principle, be directly translated into an amino acid sequence (although this only occurs for the 5′ proximal genes in TMV[33,34]). The coat of the virus consists of 2130 copies of a 17.5 kDa coat protein that assembles around the RNA in a right-handed helix (see Fig. 3). The RNA itself is 6395 nucleotides in length and as assembly of the protein subunits in a helix only occurs in conjunction with RNA at pHs above 7[35] it acts as a molecular ruler, ensuring that the resulting virions are consistently 300 nm long. The assembly process is well characterised[36] and depends on a specific sequence being present in the ssRNA (the "origin of assembly") which begins at the 876th nucleotide from the 3′ end of the RNA in the common strain of the virus.[36,37] Initially a two-layer disc or short helix consisting of approximately 38 copies of the monomer protein (the so-called "20S aggregate")[35,38] binds to a stem-loop structure formed by the origin and acts as the seed for recruitment of further monomers. Each TMV monomer consists largely of two pairs of alpha helices connected by unstructured loops with residues from both loops contributing to the RNA-binding site.[39] Assembly progresses in the 3′ to 5′ direction[38] and is caused by the RNA binding leading to ordering of unstructured regions into the helical conformation which permits polymerisation.[40] Polymerisation is itself an interesting process covered extensively elsewhere[41] and consists of a loop of RNA at the 5′ end being pulled through the centre of the growing rod which extends by sequential addition of 20S subunits.[42] RNA is bound in a groove between protein layers and so does not obstruct the central cavity. The central cavity of the produced nanotube is 4 nm in diameter versus 18 nm for the diameter of the virion as a whole. The nature of the assembly process has one interesting consequence: The length of the produced TMV can be controlled by altering the length of the RNA[43] (provided the origin of assembly is retained) offering further flexibility to the bionanoscientist.

TMV has been widely used as a bionanoscience tool: Perhaps the earliest example was its use as a mineralisation template in 1999.[44] Mineralisation

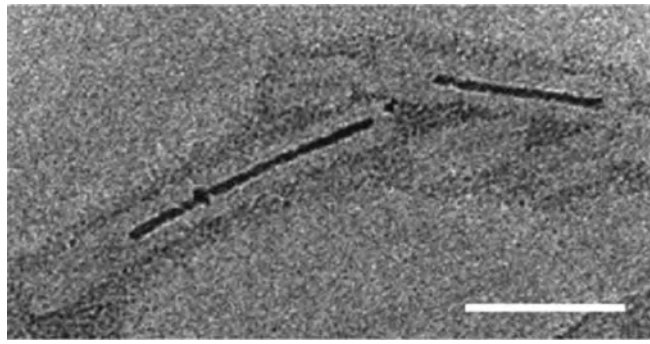


Fig. 4 Electron micrograph showing TMV particles in which the central cavity has been filled with Co-Pt which has been biomineralized *in situ*.[45] Reprinted with permission from R. Tsukamoto, M. Muraoka, M. Seki, H. Tabata, and I. Yamashita, Synthesis of CoPt and FePt3 Nanowires Using the Central Channel of Tobacco Mosaic Virus as a Biotemplate. *Chem. Mater*, 2007, **19**, 2389–2391. Copyright 2012 American Chemical Society.

by proteins is a key field in bionanoscience, allowing self-assembled proteins to dictate and template the production of inorganic nanostrutures typically under benign conditions. In the case of TMV, the external surface of the virus was mineralized with various composites including PbS, CDs and iron oxides.[44] The internal cavity of the virus is also of interest as it potentially provides a more controllable ''nanoreactor'' environment for the production of even narrower wires. To date, the cavity has been used to mineralise nanowires of nickel and cobalt[46] and more recently other materials including CoPt and $FePt_3$[47] (Fig. 4). In bionanoscience, the key concept is of building structures *from the bottom up*. Proteins and protein complexes naturally assemble in this way and an understanding of this process would be useful for the production of artificial protein nanosystems. Here too, TMV has proved useful; the coat protein can form various morphologies depending on buffer conditions. For example, at pHs below approximately 6 the tubular helix dominates, but around a pH of approximately 6.5 these are largely replaced by ''lock washer'' disks. At pHs closer to 7, and if the ionic strength is lowered, ''20S'' discs are seen.[48] TMV research continues to generate interesting results with recent studies producing a ''light harvester''[49] and a gold nano-dot nano-ring with potential metamaterials applications.[50]

TMV may in the future form nano-circuitry parts, connected to larger bio-silicon hybrid devices. Progress has already been made towards this end in controlling the patterning of TMV on semiconductor surfaces. In one report, researchers were able to precisely arrange TMV on a DPN (dip-pen nanolithography) patterned surface; DPN was used to lay down precise patterns of 16-thiohexadecanoic acid to which TMV was able to bind *via* bridging nickel ions.[51]

4.1.2 Flagellin. As mentioned above, bacterial flagellae are protein nanotubes, formed by polymerisation of flagellin protein. Rotation of the flagella provides the driving force for bacterial ''swimming''. The whole of the bacterial flagella is a highly complex and fascinating example of a self-assembled nanomotor.[53] It is in fact highly related to Type III secretion

systems of Gram negative bacteria, which are extremely interesting in their own right, and which form needle complexes used to inject toxic proteins into host cells. So far these Type III needle proteins, which form strong hollow tubes, have not been used for bionanotechnology, but this might be worthy of future study. Mineralization has been carried out on flagella to produce nanotubes of TiO_2.[54] Addition of peptide loops onto the surface of flagella allows them to be coated with various metals to produce templated nanoparticles and nanotubes.[55] Similarly modified flagella can be made to undergo layer-by-layer assembly.[56] There are known mutations in flagellin that result in straight filaments[57,58] which have greater potential in nanoscience applications due to their ability to make regular ordered structures. Very recently flagellae have been used as bio-templates for the fabrication of tunable silica-mineralised nanotubes.[59] This method involved treatment of the flagellae with aminopropyltriethoxysilnae (APTES), followed by tetraethoxysilane (TEOS). By simple modification of the reaction time and conditions, it proved possible to coat the flagellae precisely with a controlled thickness of silica. This is an example of biomorphic mineralization, a field which has been reviewed fairly recently.[60]

4.1.3 Microtubules. Microtubules (MTs) are nanotubular protein structures that constitute part of the cytoskeleton. They are approximately 25 nm in diameter and up to many micrometres in length and are constructed from repeating units of α- and β-tubulin dimer proteins[52] (Fig. 5). They have important specialised roles in neural cells and have even been suggested as the site of quantum processing in the brain.[62] MTs have often attracted attention in bionanoscience because they act as a ''track''

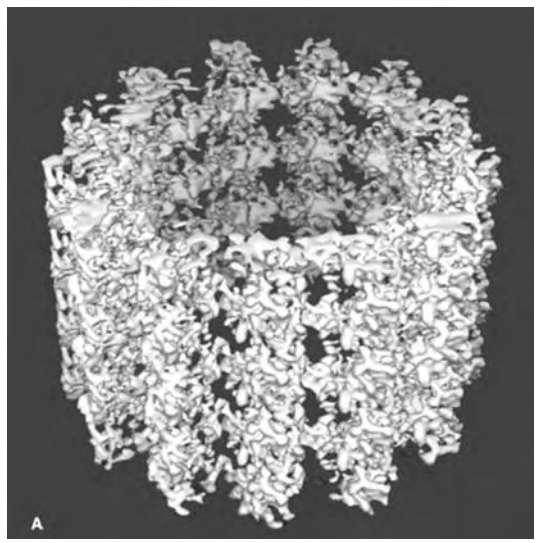


Fig. 5 3D reconstruction of microtubule structure at 8Å obtained from cryo electron microscopy.[52] Reprinted from H. Li, D. J. DeRosier, W. V. Nicholson, E. Nogales and K. H. Downing, Microtubule Structure at 8 Å Resolution, *Structure*, **10**, 1317–1328. Copyright 2012, with permission from Elsevier.

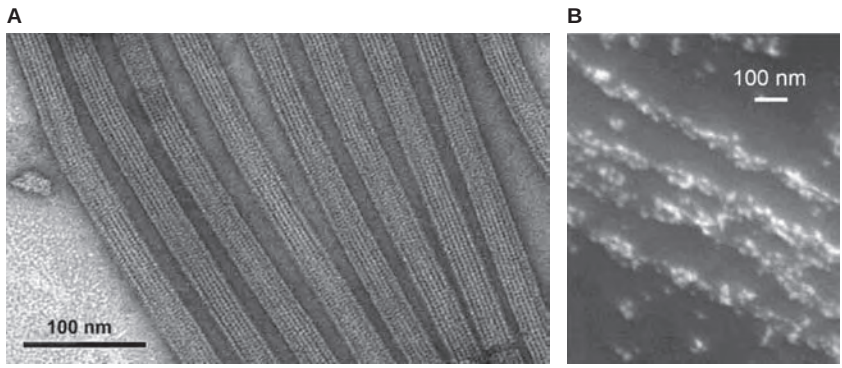


Fig. 6 MTs used to construct bionano arrays.[61] **A:** TEM image showing an array of MTS formed after the addition of tau proteins. **B:** SEM image showing the MT array metallized with gold. Reproduced with permission from J. C. Zhou, Y. Gao, A. A. Martinez-Molares, X. Jing, D. Yan, J. Lau, T. Hamasaki, C. S. Ozkan, M. Ozkan, E. Hu, and B. Dunn, Microtubule-Based Gold Nanowires and Nanowire Arrays. *Small*, 2008, **4**, 1507–1515 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

along which motor proteins carrying cargo can run.[63] Thus they may form part of a nanoscale assembly line in molecular manufacturing. In essence this is their role in the body where kinesin and dynein motors "walk" along the MT tracks carrying a variety or cargoes including proteins and cell organelles. MTs show polarity which allows for directionality of motor protein travel with kinesin travelling towards the "+" end and dynein moving in the opposite direction, towards the "−" end (where the "+" end is defined as terminating in β subunits and the "−" end in α subunits.)

Apart from their role as a transportation track, microtubules in themselves may have additional uses in bionanoscience, and predominantly they have been investigated as nanowire templates. Advantages of microtubules over other systems lies with the fact that polymerization occurs preferentially at the β-tubulin end, so that growth between two desired points (*e.g.* in an electrical circuit) is possible and the potential for this ability to be exploited for the "growing" of interconnects has been demonstrated.[64] In one report MTs were deposited on a Si surface and then coated by the covalent attachment of gold nanoparticles.[61] Current-voltage measurements revealed a 2.5 μm nanowire produced in this way to have a resistivity of 7.3×10^{-5} Ω m. In the same report MTs were used to produce parallel arrays of gold nanowires, in an experiment that demonstrates the power of biological systems to self-assembled into complex systems through addition of new "modules", tau protein was used to regularly space the MTs. Tau protein is a naturally occurring protein from neural cells that binds to MTs.[65] Using this method, a straight, regular array of gold nanowires on a surface was achieved (Fig. 6).

4.2 Artificial Protein Nanotubes

Naturally occurring protein nanotubes have advantages in that they are already proven to produce stable nanotubular structures and may be readily

available in large quantities. However, if nanoscience transitions into a technology that is able to make custom smart drugs and ultimately, programmable ''nanomachines'' then large numbers of components, each designed specifically for the task at hand, will be required. Existing natural structures are unlikely to conform to the exact requirements of dimension, chemical properties and so on in each case. The ability to design protein structures such as nanotubes, or at the very least to modify the properties of known structures, therefore becomes necessary.

While naturally occurring protein tubes often consist of repeating monomers that self-assemble in a helical fashion to form the tube walls, artificial nanotubes are typically constructed from ring-shaped proteins. The concept behind the construction is simple – if rings are stacked on top of one another then tubes will form. This allows a potentially greater control of the position of modifications both in the cavity of the nanotube and the outer surface. Several groups have utilised this technique to construct nanotubes, as described below.

4.2.1 TRAP Nanotube. TRAP (*trp* RNA-binding Attenuation Protein) is a ring-shaped protein found in species of *Bacillus*.[66] It is of interest for bionanoscientists because of its ring shape, its small dimensions (only 8 nm in diameter with a central hole approximately 2 nm across) and its high stability – it is able to tolerate numerous mutations while still maintaining its ring shape, and it is also highly thermostable. The ring is made of 11 identical TRAP monomers, each approximately 8.4 kDa in size and consisting of approximately 75 amino acids depending on species. The ring protein is able to bind up to 55 bases of single-stranded RNA around its outer perimeter and is also able to bind a tryptophan ligand in a buried binding pocket present on each of its constituent monomers and the crystal structure of both tryptophan-bound and RNA-bound forms have been solved.[67,68] Both of these abilities are related to its *in vivo* role in the control of tryptophan synthesis.

TRAP has previously been the subject of ''symmetry engineering'' whereby, by production of tandem repeats of the protein, the normal 11-membered ring was forced to form a 12-membered ring.[69] The cavity of the protein has also been used to capture gold nanoparticles and deliver and constrain them on a silicon oxide surface where they were used to form part of a floating nanodot gate memory prototype.[70] In order to promote stacking of TRAP rings into the tubular form, cysteine residues were inserted at positions 69 and 55 on opposite faces of the ring.[2] These residues are approximately the same radial distance from the centre of the ring and so should align when one ring is directly above the other (Fig. 7). As TRAP has no naturally occurring cysteines, it was hoped that the ring stacking would occur and that rings would be held together by the formation of disulfide bonds between opposing cysteine side chains. Initial results showed that the mutant proteins did indeed form long, straight nanotubes, often up to a micrometre or more in length (a high aspect ratio given the height of a single ring is a mere 3 nm). Single particle analysis of the constructed tubes was able to elucidate the order of the packing between individual rings (Fig. 8) and suggested a model whereby one face of each ring was connected to its

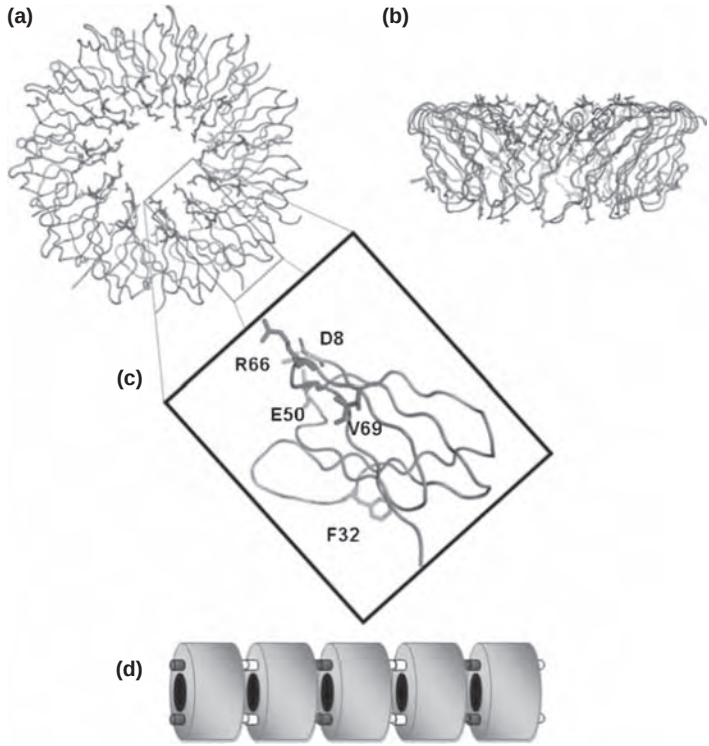


Fig. 7 Constructing a protein nanotube from TRAP. **A:** Crystal structure of the TRAP ring (based on PDB ID: 1qaw[3]) looking down on the 11-fold rotational symmetry axis and, in **B:** orthogonal to the axis. **C:** shows an enlargement of one of the eleven TRAP monomer proteins from the ring. A number of amino acid residues are highlighted. Residues 50 and 69, which are on opposite faces of the ring but approximately the same distance from the centre, were mutated to cysteine to make the nanotube. **D:** Shows a cartoon of nantotube formation. Mutations on each face of the ring are shown in either white or red. If these mutations are self-binding then the protein rings will assemble into a tube as shown. In the case of TRAP it was envisaged that pairs of cysteines on both faces would form disulphide bonds although the mechanism of tube formation appears to be somewhat more complex. Figure from Miranda *et al.*,[2] Copyright © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, used with permission. [Colour image available on-line]

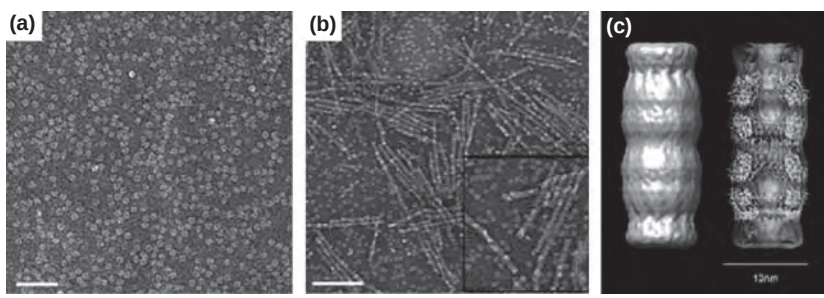


Fig. 8 Evidence of tube formation by mutant TRAP. **A:** shows A TEM image of standard TRAP rings. **B:** shows mutant TRAP which has formed protein nanotubes up to several hundred nanometres in length with an inset showing some tubes enlarged 1.5 times. **C:** Iso-surface image of single particle reconstruction of the assembled tube showing the surface (left) and divided down the longitudinal axis to reveal the interior. The scale bar in A is 50 nm, in B 100 nm and in C 10 nm. A and B are panels from Fig. 2 of Miranda *et al.*,[2] C is a panel from Fig. 7 of the same. Both are copyright © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, used with permission.

neighbour *via* disulfide bridges while the other may be due to disulfides and/or hydrophobic packing interactions. It is hoped that the TRAP nanotube may provide the basis for furthur development into more sophisticated structures. Like TMV, it has obvious potential as a nanowire template for example. Furthermore its additional abilities such as the wrapping of RNA may allow it to easily interface with RNA-based structures. Finally, a protein known as "Anti-TRAP"[71,72] can bind around the outer circumference of the protein[73] allowing potential additional "hooks" for attachment of further structures.

4.2.2 Hcp Nanotube. Hcp protein offers an example of an artificial protein nanotube which in many ways resembles the TRAP system. Hcp is a ring-shaped hexamer from *Pseudomonas aeruginosa* where it constitutes part of the Type VI secretion system. The ring has a total diameter of 9 nm with a central hole approximately 4 nm in diameter.[74] As with TRAP, the ring protein was modified to promote stacking by the addition of cysteine residues on the two flat external faces (Fig. 9). The resulting tubes were up to 100 nm in length and could be capped by spiking the polymerisation mixture with rings containing cysteine residues on one face only,[75] providing statistical length control and resulting in the production of hollow nanocapsules.

4.2.3 SP1 Nanotube. SP1 (Stable Protein 1) has similarities to both TRAP and hcp. It is a stable, stress response protein originally isolated from aspen plants.[76–78] It has a ring shape and is constructed from multiple subunits (twelve identical monomers in this case[79]). The X-ray crystal structure of the protein has been solved[80] and shows that the SP1 ring in fact resembles a double ring with 6 protein monomers in each "layer". Overall, the ring is approximately 11 nm in diameter and 4 nm in depth with the central hole being approximately 3 nm in diameter. The N-terminus of each monomer lines the internal cavity. This fact was utilised in order to make a chain-like protein nanotube with variable spacing between each SP1 building block. In the experiment, reported by Medalsy *et al.*,[81] the gene encoding the SP1 monomer was modified so that a six histidine tag was added to the N-terminus. The modified protein still formed a ring and was added to 1.8 nm diameter gold nanoparticles decorated with Ni-NA ligands which bind His-tags. In this way each SP1 ring was able to bind two gold nanoparticles, each entering the central hole from opposite faces on the ring. Each gold particle protruded from the hole and thus was able to bind to the histidines of a second SP1 ring. In this way a chain-like nanotube could be formed, bridged by gold nanoparticles (Fig. 10). The distance between each ring in the chain could be decreased by truncating the N-terminus to which the his-tag was attached.

Interestingly the SP1 ring was recently used to produce a simple nanoscale "Set-Reset" circuit[82] (a fundamental building block of logic circuits). Because the N-terminus of the protein lines the central cavity, genetic fusions can be made to add additional amino acids to endow the cavity with new binding properties. In this case silicon-binding peptide[83–85] was used. The resulting ring protein[86] was thus able to bind a single 5 nm diameter silicon nanoparticle in its central cavity. To function well, the

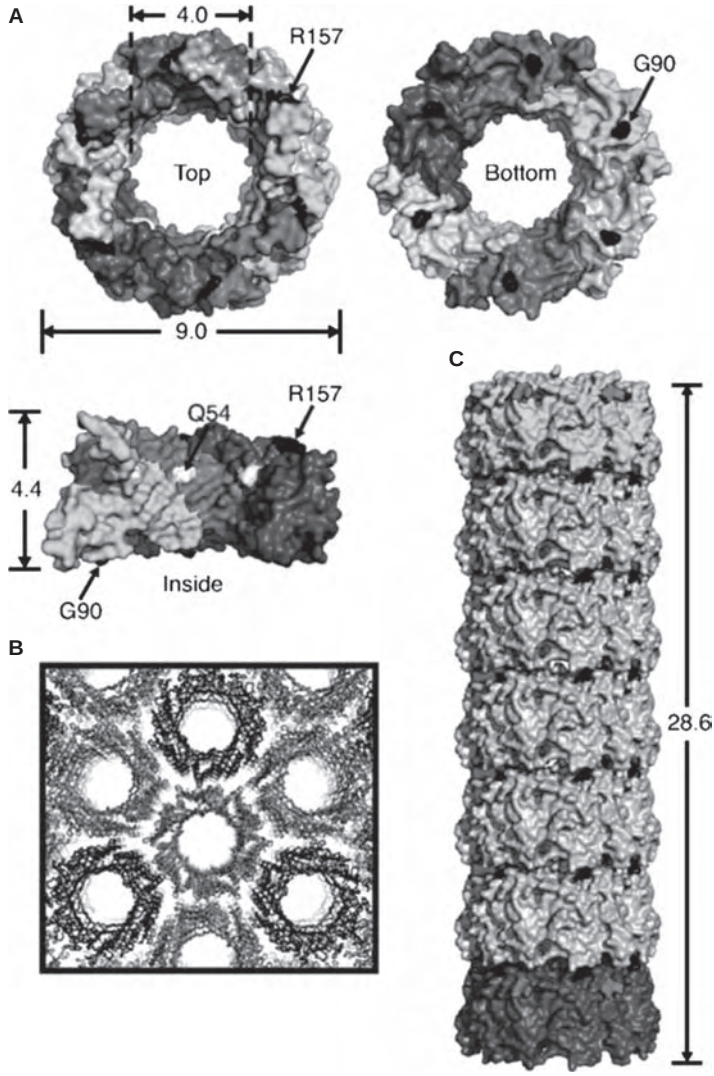


Fig. 9 Hcp can be engineered to form a protein nanotube.[75] **A:** The crystal structure of hcp (PDB ID: 1y12[74]) shown looking in both directions down the rotational symmetry axis and orthogonal to it. **B:** In the crystal structure, the rings pack into a tubular structure, thus providing guidance for positioning of mutations to link rings into tube. **C:** The resulting tube formed from stacked rings that results from disulfide bond formation between residues C157 and C90. Reproduced with permission from E. R. Ballister, A. H. Lai, R. N. Zuckermann, Y. Cheng, and J. D. Mougous, *In vitro* self-assembly of tailorable nanotubes from a simple protein building block, 2008, **105**, 3733–3738. Copyright 2008 National Academy of Sciences, U.S.A.

silicon particle needs to be able to store charge in a stable fashion, yet still be addressable. This was achieved by the protein ring, which isolated the nanoparticle from the conducting surface but still allowed access to the particle by the tip of a conductive atomic-force microscope. An AFM tip can therefore be used to alter and read the charge on the silicon nanoparticle.

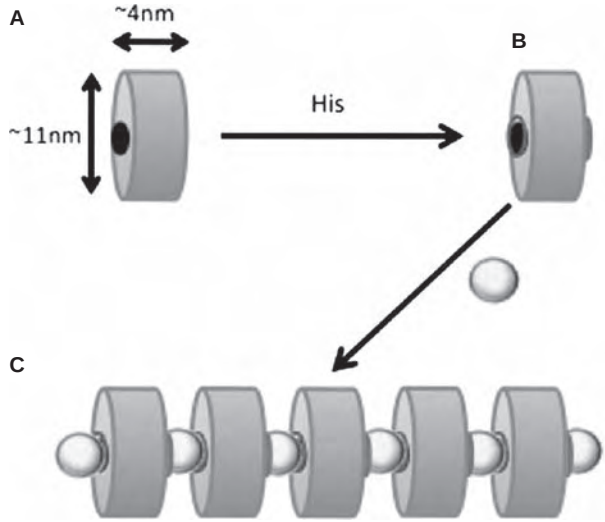


Fig. 10 A chain-like spaced nanotube can be built from modified SP1 protein. **A:** A wild type SP1 ring is modified so that the interior of the cavity contains multiple histidines (shown by the green modification in **B**, depicted outside the cavity for clarity). Gold nanoparticles modified with Ni-NTA are able to bind to the histidines in the cavity from both faces of the ring, allowing them act as a bridge holding neighbouring rings together into a chain-like tube (**C**). Based on information in Medalsy *et al.*, 2008.[81] [Colour image available on-line]

4.3 Protein nanotube summary

Protein-based nanotubes offer a range of possibility both as templates for inorganic wires, drug delivery systems or as components of larger devices. In each case the ability to control precisely the length of the produced nanotubes is of vital importance. Inorganic wires have potential uses as interconnects between electrodes in circuits or biosensors. The distances to be bridged must be well-defined to build reliable circuitry reproducibly, and for this nanotubes of the correct length will be necessary. For drug delivery systems, enclosed protein cages are required, so a nanotube would need to be sealed at each end after loading with the therapeutic agent of choice. One can imagine a system which opens on sensing a target cell has been reached, but it may be simpler to trigger uptake of the nanotube by the target cell through endocytosis. The length of the protein tube will probably affect a range of parameters such as diffusion of drug from the cavity of the tube and biodistribution. If nanotubes are used as part of a larger device such as a complex ''nanomachine'' then, as with any machine, the dimensions of each part must be engineered within certain defined tolerances if the machine is to work correctly.

Precisely defining and controlling the length of protein nanotubes is a difficult problem as they typically form by a process of polymerization of numerous identical subunits to give a straight structure with simple translational symmetry. There is, in general, no difference in energy to adding one more subunit to a long structure or to a short one because of this symmetry, and so long structures will not naturally grow to some fixed maximum length. In practice the tube length is distributed over a range of sizes. This range can be controlled and somewhat narrowed by controlling the reaction conditions or by ''spiking'' the reaction mixture with monomers

that do not support further polymerization (as in the case of Hcp). However, this cannot produce a monodisperse sample. One theoretical solution to the problem is producing protein nanotubes from the creation of a tandem gene, which fuses together a number of copies of the monomer. If these fusion proteins naturally associate to lie alongside each other in register in hollow bundles, then the length of the resulting tube will be determined by the number of copies of the protein subunit encoded in the fusion gene. This method has obvious difficulties even for short tubes, as there is no simple way of encouraging the fusion proteins to lie in register, unless each subunit has different preferred associations Even so, it seems unlikely that this method could produce very long tubes, partly because the DNA construct itself would be unstable, and partly because long flexible strings of beads have many conformations available to them, and self-association into the desired tube would require assistance in the form of molecular scaffolds or other chaperones. TMV, which naturally produces nanotubes of a consistent (300 nm) length, shows perhaps the best way to achieve length-programmability. In this case RNA, act as a molecular ruler upon which the protein subunits assemble. A similar system of molecular rulers may need to be developed for other protein nanotubes.

Once a protein nanotube of suitable shape and reproducibility has been found, decoration with required materials and connection to other components will be necessary to incorporate it into a functional device. Decoration of protein nanotubes may be useful in a variety of situations, most obviously the production of nanowires. Again TMV shows possible approaches, and it has been decorated with conducting materials both in the inner and outer cavity as mentioned above. Filling the inner cavities of protein nanotubes with materials such as metals poses particular difficulties as material filling from the ends may mineralize prematurely, thus blocking the further entry of metal ions, resulting in an unfilled central portion. This can be addressed by alteration of reaction conditions so that metal deposition occurs only once the tube has filled with metal ions. A further possibility in the case of protein nanotubes constructed from ring-shaped components could be the filling of the individual rings prior to assembly into a tube. If protein nanotubes of the correct length and decorated/filled with the required material can be constructed, the final challenge is to endow them with the means to connect to other components such as other proteins, nanowires and so on in order to build up more complex structures with sophisticated functions.

A bionanomachine, like any complex machine will be made of modules each with its own specific role in the whole and it is unlikely that a single protein will be able to carry out such functions. Therefore multiple components will be necessary. Currently bionanoscience lacks the ability to routinely connect together individual functional modules in a reproducible way and the development of ''universal joints'' would be a significant advance.

5 Cage proteins

As with nanotubes, nano-size protein cages offer a cavity that can be filled with molecular ''cargo''. However, in such proteins the cavity may be

completely enclosed from the outside, offering potential advantages when the cargo or contents must be protected from the surrounding environment or released at an appropriate time (for example in nanoscale reaction vessels or in drug delivery devices). Cages also offer only a single "external" and single "internal" face, whereas nanotubes may have polarity. The symmetry of cage proteins however ensures they are self-limiting in size, and since they are also generally very stable their reproducibility and monodispersity are guaranteed.

5.1 Naturally occurring cage proteins

As with tubular proteins, cage proteins are common in nature and their utility is self-evident for storage or protection; they can fully encapsulate a space and in doing so store the contents in isolation from the surrounding environment. This is of particular use from the bionanoscience perspective when the cargo is potentially fragile or toxic.

5.1.1 Virus capsids. The use of virus capsids in bionanoscience is a large area of research. Capsids have a wide variety of sizes (from 18–500 nm) and morphologies ranging from tubular to icosahedral[87] (which account for the majority and can, to a rough approximation, be regarded as spherical). Viruses devoid of genetic material (Virus-like particles; VLPs) are being investigated for applications in which either immunogenicity is required (as vaccines or vaccine components) or where no immunogenic response is preferred (such as drug delivery and imaging).

5.1.1.1 Plant Viruses. Plant viruses are becoming popular subjects for research due to the fact that they are generally regarded as non-infectious for humans, removing the possibility of reversion to forms capable of replication. This assumption is supported by the fact that biotoxicity in mice appears to be low.[90] They have been investigated as drug delivery vehicles, nano-reactors and mineralisation templates amongst others. Amongst the most widely used virus capsids are cowpea chlorotic mottle virus (CCMV) and cowpea mosaic virus (CPMV; Fig. 11). Both of these icosahedral virus capsids have N or C termini exposed on the outer or inner surfaces meaning that additional peptides or even whole protein domains can be added to these surfaces using straightforward genetic engineering techniques. Additional exposed amino acid side-chains may be genetically modified to provide further functionality. For example, exposed carboxylate groups on the surface of CPMV have been reacted with methyl(aminopropyl)viologen to produce redox active nanoparticles[89] and they have been modified with peptides for mineralisation of a variety of minerals and metals[91–94] including FePt (Fig. 12). Recently, surface lysines have been modified by the addition of succinamate groups. This charge modification allowed facile deposition of cobalt or iron oxide to give evenly coated 32 nm diameter particles.[93]

For use as vaccines, plant viruses, like other viruses, offer multiple repeating units meaning that a single particle may be able to expose numerous copies of an epitope, ensuring a strong immune response. Furthermore, while plant viruses themselves are non-infectious to mammals,

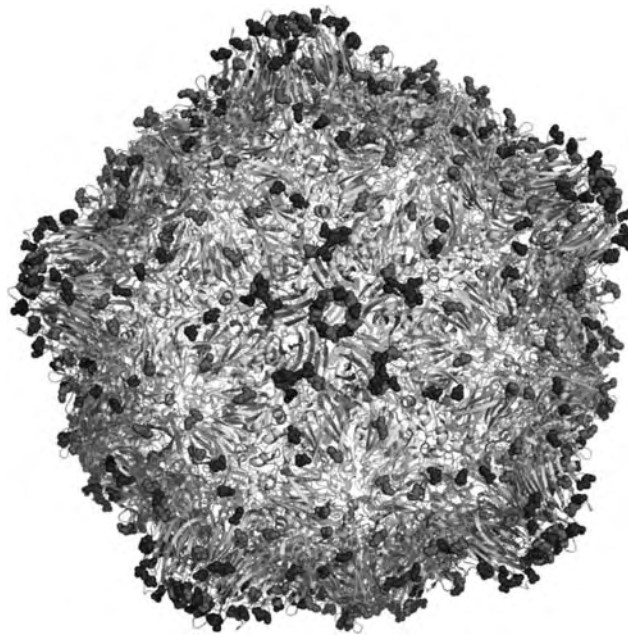


Fig. 11 Crystal structure of CPMV (PDB code: 1NY7[88]). Structure is shown in cartoon format, looking down the 5-fold axis with the small subunit coloured cyan and the large subunit coloured green. Surface accessible carboxylate groups, some or all of which can be modified[89] (excepting the C-terminal residues) are shown as blue spheres. Surface accessible, reactive lysines are shown as red spheres. [Colour image available on-line]

they may trigger an innate immune response further enhancing the generation of immunity to whatever foreign epitopes the virus particles may be carrying. One of the most widely studied viruses in this area is cowpea mosaic virus (CPMV). CPMV is an icosahedral comovirus approximately 30 nm in diameter, constructed of 120 copies of two protein subunits (''small'' and ''large'', Fig. 11). Its use as a potential vaccine has focused on the insertion of peptide epitopes into one of the unstructured loops which decorate the surface. Peptides of less than 40 amino acids in length with a pI below 9.0 are generally well tolerated.[95] The ability of CPMV modified to carry alien epitopes to confer immunity *in vivo* has been demonstrated: CPMV carrying a peptide from the capsid of mink enteritis virus (MEV) was shown to confer protection to it in both mink and dogs.[97,98]

While the relative inertness of plant viruses with regard to mammalian cells is an advantage for some applications, for others such as targeted drug delivery, it presents a potential problem. However the stability of the virus capsid and the relative ease of manipulation in combination with its tolerance of large peptide inserts means that it can be engineered to target cells of interest. For example Hibiscus chlorotic ringspot virus (HCRSV) is a plant virus with an icosahedral capsid similar in size and structure to CPMV. In one instance folic acid was attached to the surface of the virus *via* carbodiimide coupling while the central cavity was filled with doxorubicin, an anti-cancer agent. The modified capsids were found to be preferentially taken up by ovarian cancer cells (which over expresses folic acid receptors

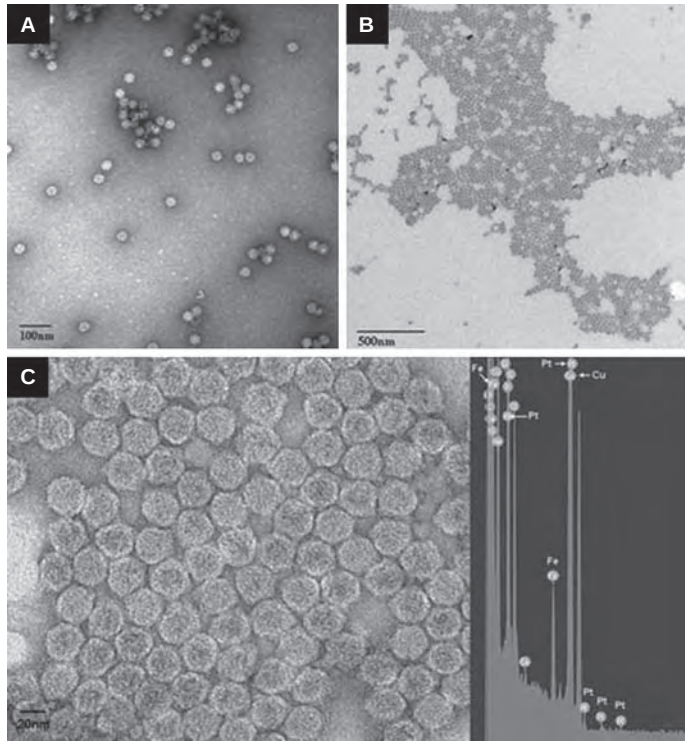


Fig. 12 CPMV particles can be mineralized with a thin metallic coating.[92] **A:** TEM image showing uranyl acetate stained CPMVFePt chimaera prior to mineralization. B and **C:** unstained modified particles coated with FePt. The right hand side of C, shows an EDXS spectrum of the mineralized particles confirming FEPt mineralization. Reproduced from S. N. Shah, N. F. Steinmetz, A. A. A. Aljabali, G. P. Lomonossoff, and D. J. Evans, Environmentally benign synthesis of virus-templated, monodisperse, iron-platinum nanoparticles. *Dalton Transactions*, 2009, 8479–8480 with permission of The Royal Society of Chemistry.

on their surfaces) where they had a cell-killing effect.[99] Similarly CPMV has been used to target tumours *via* decoration with F56f, a ligand that targets vascular endothelial growth factor receptor-1 (VEGFR-1), which is expressed on a number of cancer cells. CPMB modified in this way (along with a fluorophore) was observed to bind to tumour cells expressing VEGFR-1.[100] In the cases of drug delivery, targeting to cells maybe required but immunogenicity unwanted. In such cases the ease with which plant virus capsids can be modified means that their immunogenicity can be passivated by for example, trivial attachment of molecules of polyethylene glycol (PEG). TMV is different from many of the viruses used in this field as it is a rod-shaped tube rather than an icoashedron (see above). However, under certain conditions TMV can also form spherical structures. Recent results have shown that heating of the virus to 94 °C followed by cooling (''thermal remodelling'') results in the formation of spherical protein nanoparticles the diameter of which depends on the concentration of protein.[101] The structural properties of these spheres remain to be determined. Such switches in morphology depending on buffer conditions are not uncommon in viruses (see SV40 below and Fig. 13).

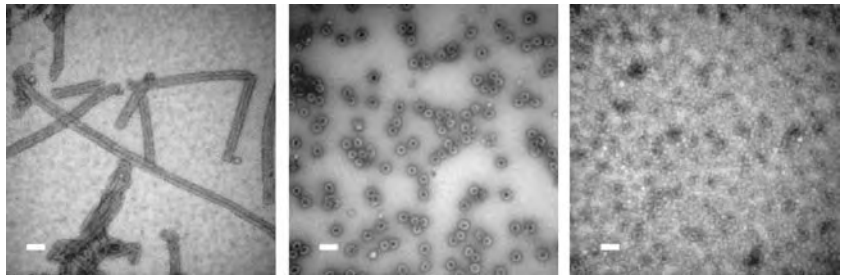


Fig. 13 Some viruses can adopt discrete morphologies depending on conditions. SV40, shown here, can exist as sphere or tubes on buffer conditions. Left: SV40 VP1 pentamers form a tubular structure *in vitro*. Middle: SV40 VP1 pentamers forming a 40 nm spherical particles *in vitro* in the presence of DNA. Right: VP1 forming VLPs (Virus Like Particles) within insect cells. Scale bar is 100 nm throughout. Images courtesy of Hiroko Tsukamoto. Reproduced with permission from Heddle *et al.*, 2008.[96]

5.1.1.2 Animal viruses: SV40. Simian Virus 40 (SV40)[102,103] is another example of a virus that is able to adopt different shapes depending on buffer conditions. It predominantly consists of a spherical capsid 40–45 nm in diameter. This *in vivo* form is constructed from 72 pentamers of the major capsid protein VP1 and 72 copies of minor capsid proteins (VP2 or VP3). At pH 5.0 long tubes form, whereas at high salt concentrations spheres as small as 20 nm are most common. Switching on the assembly of spherical particles is achieved by VP2 with VP1 at neutral pH[104] and EGTA can reversibly disassemble the capsid.[105] Further modulation of the structure is achieved by DNA, which is required for assembly under physiological conditions and which causes a change from tubular to spherical forms at pH 5.[45] A number of different morphologies are shown in Fig. 13. SV40 represents something of a double-edged sword regarding it potential medical use. On the one hand its ability to infect very efficiently a wide range of human cells, a potential weakness of plant-viruses, may be an advantage for use as a drug or gene delivery agent. However it is also carcinogenic. In recent work SV40 was conjugated with hGEF and was taken up by human epithelial carcinoma cells.[106] It may prove to be useful as a vector in gene therapy.[107]

5.1.2 Ferritin. Iron is absolutely required for life because of its unique redox chemistry.[108] Higher animals also require very considerable amounts of iron in the form of haem, the oxygen binding component in haemoglobin, the oxygen carrier of the blood However, its reactivity makes iron dangerous in the cell since it can generate reactive oxygen species: Ferrous iron is known to react with hydrogen peroxide in the Fenton reaction to produce highly reactive hydroxyl radicals which cause damage to cellular components.[109] There is therefore a need to store iron as a contingency for periods when iron may be missing from the diet, but this iron must be stored in a way that isolates it from the rest of the cell and holds it in a relatively inert form. Ferritin is the protein that achieves this. Ferritin is a highly conserved protein found in all domains of life and is a hollow sphere consisting of 24 protein subunits (Fig. 14).[110,111] In higher animals, two very similar protein subunit types constitute the ferritin sphere: H type, containing the catalytic ferroxidase site[112] responsible for the conversion of iron from free radical

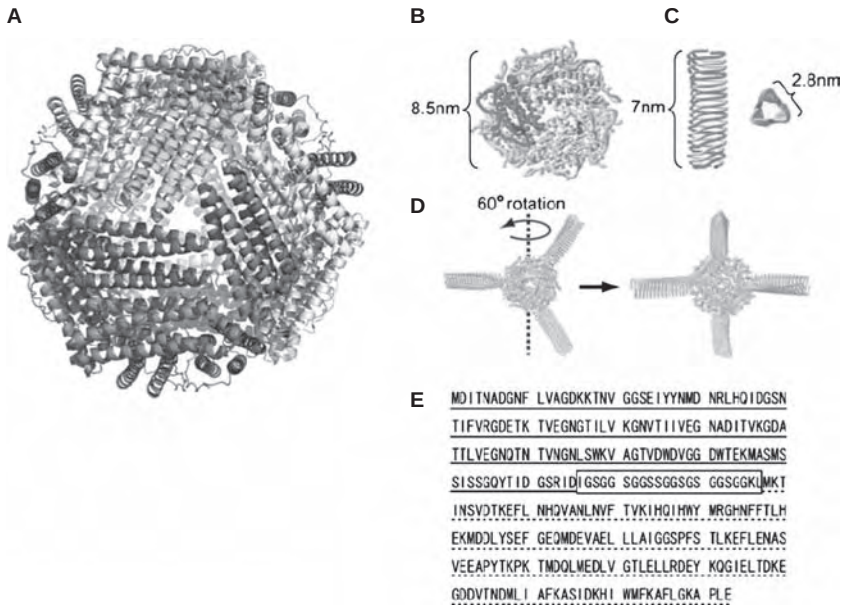


Fig. 14 **A:** Crystal structure of ferritin from *Thermotoga Maritima* (PDB ID: 1VLG), looking down the 3-fold axis. Each monomer is shown in a different colour. **B:** Structure of LisDPS looking along the three-fold axis with the N-terminal residue shown in red. C: "side" view and "top" view of gp5C trimer with each monomer in a different colour. The C-terminal residues are coloured red. D: Model of the assembled ball-and-spike protein (BSPS). E: Amino acid sequence of a monomer of BSPS. The gp5C sequence is underlined with a solid line and the LisDps with a dotted line. The linker between the two is boxed**.** Panels B-E is a renumbered figure reproduced with permission from K. Sugimoto, S. Kanamaru, K. Iwasaki, F. Arisaka, and I. Yamashita, Construction of a Ball-and-Spike Protein Supramolecule13. *Angew. Chem. Int. Ed.*, 2006, **45**, 2725–2728 and is copyright © 2006 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. [Colour image available on-line]

producing ferrous (2+) to ferric (3+) forms, and L type. The ferritin sphere stores a maximum of about 4500 ions per ferritin molecule. Cavities lined with acidic residues in the wall of the sphere connect the external environment with the central cavity. Fe(II) in solution is able to enter the cavity *via* these channels and is reduced to Fe (II) at the ferroxidase sites on the H-chain.[112] Iron ions then nucleate to form the iron oxide core. Nature has therefore created a highly efficient protein for creating metallic nanodots of fixed size.

A large amount of work has been carried out to modify ferritin for use in various bionanoscience applications. This was pioneered by the work of Mann and colleagues who were able to demonstrate the mineralization of iron oxides and manganese amongst others.[113–117] Yamashita and colleagues have expanded this work, mineralizing a wide variety of materials in the inner cavity of ferritin while adding various other modifications to its external surface. Materials mineralized to date include nickel hydroxide,[118] zinc selenide[119] and gold sulfide.[120] Ferritin has also been used to transfer these nanodots to silicon surfaces, and working devices have been demonstrated.[121–123] Modified ferritins could have a number of potential uses as electronic components and their potential as a reporter in MRI has been documented.[124]

5.1.3 DPS. Dps (DNA-binding protein from starved cells) is a protein that was found to be commonly expressed in response to oxidative damage in *E. coli* and which was able to bind to DNA.[125–128] Dps is a member of the ferritin superfamily of proteins, and like ferritins contains a ferroxidase centre. The crystal structure of Dps protein from *Listeria innocua*, the form commonly used in bionanoscience, has been solved[129] and forms a hollow spherical structure reminiscent of apo-ferritin except that dps consists of 12 monomers, rather than the 24 seen in ferritin, with each monomer being approximately 18 kDa. Consequently dps is smaller, with an overall diameter of approximately 8.5 nm and a central cavity 5 nm across. Like ferritin, dps is able to uptake and store iron ions in its central cavity and has also been used as a nano-reaction chamber in which materials such as Co_3O_4[130] and CdS[131] are mineralised.

Recently dps was coated with two different peptides – one capable of binding carbonaceous material[132] and a second, titanium binding peptide.[133] The resulting protein was able to form spheres that could bind to and coat the exterior of carbon nanotubes, and then subsequently add a layer of mineralised titanium.[134] This approach is promising as carbon nanotubes are touted as potential energy storing and conducting materials, and have been used together with virus proteins in prototype batteries.[135] Titanium oxide has great promise in photovoltaic cells and in the catalytic splitting of water, and its reactivity may be enhanced by carbon nanotubes.[136]

Perhaps the most intriguing utilization of dps from a bionano-technological point of view was its inclusion in the construction of a ball-and-spike supramolecule.[137] This work demonstrated a potential first step in the modular assembly approach of constructing complex biological nanodevices. The ball-and-spike structure consisted of the dps ball with four long ''spikes'' protruding from its surface at positions equivalent to the vertices of a tetrahedron. At these positions, dps has a 3-fold symmetry axis where the N-termini from three monomers meet. The spike protein was the C-terminal portion of gp5 protein (gp5c) from bacteriophage T4.[138] Three copies of this protein assembled into a triangular prism. Thus, each gpc5 protein could be fused to the N-terminus of each dps monomer and when the dps assembled, three copies of gp5c would be brought together, and, providing a long enough linker was present connecting it to the dps, would assemble into the expected triangular prism (Fig. 14).

5.2 Artificial cage proteins

Two important developments in this field are that of nanohedra[139] (Fig. 15) and polyhedra.[140] Nanohedra, like the ball-and-spike protein covered earlier, utilise symmetry matching of two connected proteins to construct a protein cage building block. The approach of Padilla *et al.*[139] fused together a trimeric protein (bromoperoxidase[141]) and a dimeric protein (the M1 matrix protein from influenza virus[142]), producing a 49 kDa fusion protein. The two domains were held together in a fixed relative orientation by a bridging alpha helix. The symmetry of the resulting cage depended on the angle between the subunits with a tetrahedral symmetry being chosen in this report. Equilibrium sedimentation and electron microscopy analysis seemed to suggest that the desired structure was sucessfully assembled. In principle

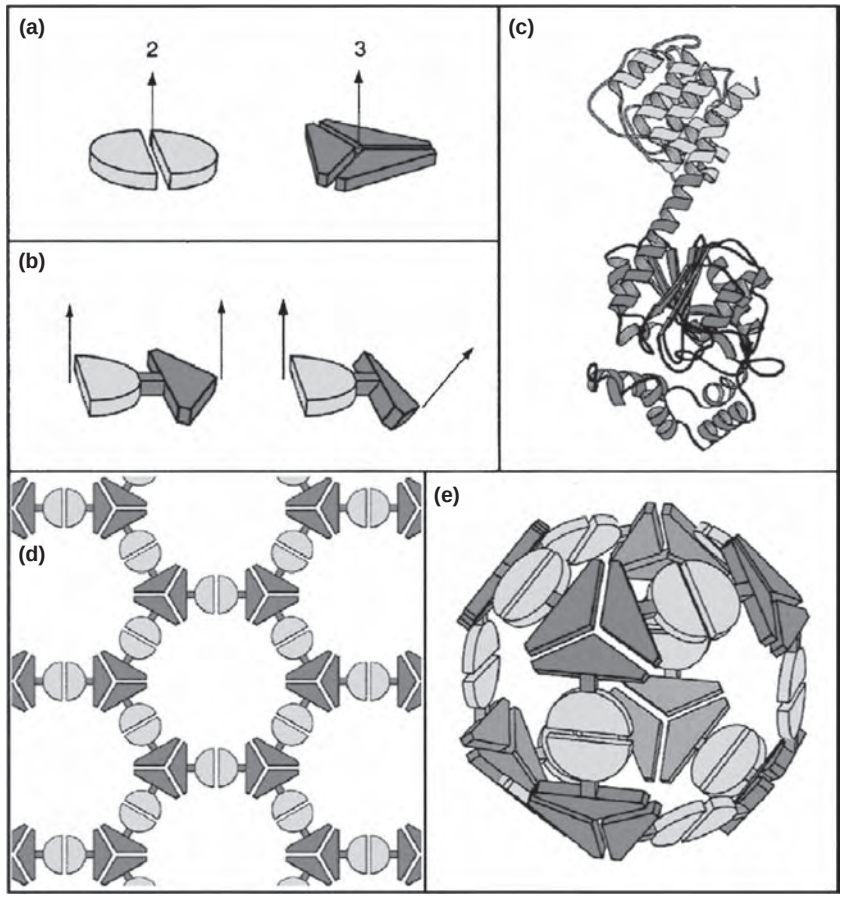


Fig. 15 A symmetry-matching approach to construction of artificial protein nanohedra.[139] Green semi-circles and red triangles represent dimer and trimer proteins respectively. Panel **c** shows a ribbon structure representation of two such proteins fused together *via* a short amino acid linker (shown in blue). With the correct geometry and symmetry, the proteins will assemble in the pattern shown in (d) to form a cage (e). Figure reproduced with permission from J. E. Padilla, C. Colovos, and T. O. Yeates, Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proceedings of the National Academy of Sciences*, 2001, **98**, 2217–2221. Copyright 2012 National Academy of Sciences, U.S.A. [Colour image available on-line]

this method can be repeated with different proteins to produce cages with a range of sizes and properties. A similar approach has been used that utilises a long synthetic peptide as the building block providing the required symmetry matching for construction of a polyhedral shell.[140] A more in depth treatment of nanohedra and polyhedra is included in the review of Lamarre and Ryadnov.[143]

5.2.1 Spherical TRAP. A recent discovery has demonstrated that TRAP protein (see above) can be remodelled so that it forms a hollow cage structure reminiscent of a virus capsid.[144] In order to achieve this effect, a point mutation was first made in the TRAP protein. Residue lysine 35 was mutated to cysteine to give a protein known as TRAPK35C. TRAP has no natural cysteines and the mutation results in 11 cysteine residues,

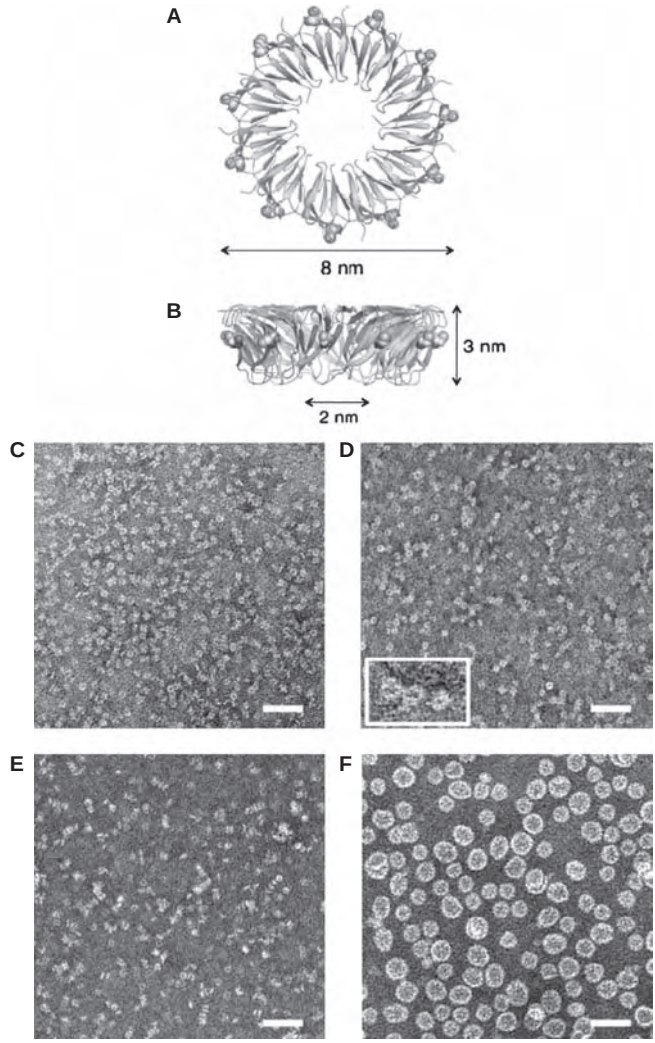


Fig. 16 Conversion of a ring protein into a capsid-like protein cage. **A** and **B** show the crystal structure of TRAP (PDB ID: 1qaw[145]) in cartoon format looking along and orthogonal to the rotation symmetry axis respectively. The residue mutated to a cysteine (K35) is shown as coloured balls. **C:** Purified wild type TRAP protein appears as rings. **D:** TRAP containing the K35C mutation also appears as standard rings. **E:** addition of gold nanoparticles to the wild type protein causes no significant change to the structure. **F:** Addition of gold nanoparticles results in the ring proteins being converted into large capsid-like spherical structures. Scale bars in **E**–**F** are 40 nm. Reprinted with permission from A. D. Malay, J. G. Heddle, S. Tomita, K. Iwasaki, N. Miyazaki, K. Sumitomo, H. Yanagi, I. Yamashita and Y. Uraoka, *Nano Lett*., 2012, **12**, 2056–2059. Copyright 2012 American Chemical Society. [Colour image available on-line]

which protrude out from the outer circumference of the ring (Fig. 16). The resulting protein is stable and has a ring-shaped structure indistinguishable to that of the wild-type protein. When the TRAPK35C is mixed with 1.4 nm diameter gold nanoparticles it is remodelled into a new structure.

The new structure is completely different from the original ring – it is an approximately 20 nm diameter hollow sphere (Fig. 16). Such a drastic

remodelling of one protein structure into another is unheard of – and is all the more surprising given the very high stability of the starting protein. The exact mechanism of the remodelling is unknown, but it is clear that the cysteine mutation is necessary (the change does not occur when the wild-type TRAP is used) and the gold is also required. Interestingly, gold appears not to be a necessary component of the final structure, with some of the protein clusters showing no association with gold nanoparticles, and apparently acts as a catalyst in this spontaneous conversion. 1.4 nm gold nanoparticles consist of 55 gold atoms and are known to have special electronic properties and catalytic abilities very different from those of the bulk metal.[146,147] It is likely that catalysis by the gold proceeds initially by binding of the thiol group on the cysteine side chain to the gold nanoparticle surface followed by a rearrangement possibly involving disulfide formation between cysteines of neighbouring TRAP rings which results in the sphere being formed.

The spherical structure produced is called a capsid-like structure (CLS) because of its superficial resemblance to icosahedral virus capsids. The CLS can actually be produced as two discrete sizes depending on the ratio of gold to protein. When the concentration of gold is high a 16 nm CLS is produced. When it is low a 22 nm CLS results. A similar result is seen in some viruses where a variable T-number (a measure of quasi-symmetry in the capsid) results in the production of viruses of discrete sizes.[148,149] Despite these similarities, the TRAP protein has no known sequence or evolutionary relationship to virus proteins. Much work remains to be carried out to understand the details of the assembly procedure of the CLS. However, one thing that is known is that assembly takes several minutes after the addition of gold. This offers the possibility that the CLS could be filled with material of choice, simply by adding it to the assembly solution before the addition of gold nanoparticles in the hope that it will be automatically encapsulated.

5.3 Cage protein summary

Protein cages such as virus capsids and capsid-like particles have potential uses in a staggering range of applications, from imaging, to electronics to drug delivery. Their monodisperse size and the fact that their central cavity limits precisely the size of any material that is used to fill it are clear advantages over other protein structures. The large number of repeating units on the surface of these proteins can present difficulties in making modifications that are not repeated over all symmetry-equivalent positions. However in cases where high density coverage is desired (such as in vaccines) this can be an advantage. The relative ease of modification of the surface of many spherical proteins combined with the availability of various symmetry arrangements allows the addition of further protein and other structures to the surface raising the prospect that these particles may truly become all-purpose building blocks for more complex materials and devices. For example, the potential of producing nanoscale particles of metal or semi-conductor that self-assembled into a 3D matrix with the distance between each particle being small and precisely defined is intriguing could have potential uses in photovoltaic and 3D memory applications.

6 Channel proteins

In nature channel proteins span cellular lipid membranes and provide a channel through which water-soluble molecules can pass. By controlling the opening and closing of pores, cells can regulate concentration gradients and the flow of molecular information between the cell and environment. A close relative of protein nanotubes, channel proteins offer different but equally intriguing possibilities. For example, controllable nanopores that could be embedded in artificial membranes could have applications as diverse as biosensing and energy storage.

6.1 α-hemolysin

Bionano-technology has largely but not entirely proceeded through the development of known protein structures, either by fusing them with peptides with desirable properties or simply mutating them to alter an intrinsic function. One of the most successful examples of both techniques is α-hemolysin, a bacterial toxin that forms a heptameric transmembrane pore, and whose structure is known in detail from X-ray crystallographic studies (Fig. 17). A variety of techniques including genetic engineering and the introduction of non-natural amino acids have been used by the group of Bayley to endow the pore with different functional properties. The work is too broad to discuss in detail here, and the interested reader is directed to the review by Bayley.[151]

Through the development of patch-clamp technology and other advances, it is now possible to measure the ion carrying properties of single molecules of transmembrane channels with great accuracy. The Bayley group has developed hemolysin into a sensitive stochastic biosensor for a wide range of analytes, from simple ions to complex compounds. For example, by introducing a metal ion binding site into hemolysin, the protein can be made in to a sensitive detector of those ions.[152] More recently, by creating a mutant hemolysin carrying one copy of a short protein kinase inhibitor peptide, a biosensor could be made which detects

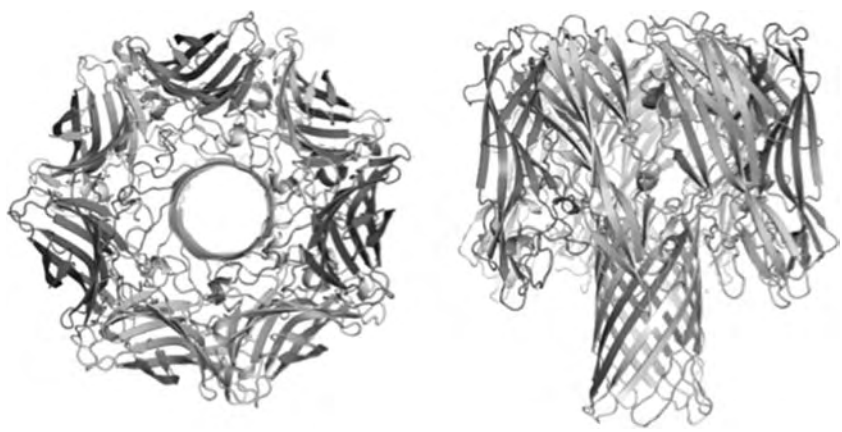


Fig. 17 Crystal structure of hemolysin[150] shown from the "top" (left) and "side" (right) as a ribbon-style Ca trace (PDB 7AHL) Each chain is coloured a different colour. The transmembrane beta-barrel allows the passage of water-soluble substrates roughly 2 nm across. [Colour image available on-line]

cAMP-dependent protein kinase. The most profound advance by the Bayley group was perhaps the discovery that hemolysin could be made to swallow cyclodextrin.[153] Cyclodextrins are, as the name implies, simple cyclic sugars which have proved to be of enormous industrial importance. They also led to the development of host-guest chemistry, as molecules which are able to solubilise a very wide range of lipophilic molecules in water. β-cyclodextrin has seven glucopyranoside residues linked by 1,4 bonds, and therefore shares the same seven-fold rotation symmetry as hemolysin. Simple mutagenesis of hemolysin allows the protein to bind β-cyclodextrin so that it constricts the pore running through the centre of the protein. The cyclodextrin itself can be modified, allowing the nature of channel to be controlled.[154] This work with hemolysin has led to the founding of Oxford Nanopore Technologies (http://www.nanoporetech.com), whose products are not aimed solely at analyte detection but also rapid DNA sequencing.[155]

All of these advances, using a known stable protein framework and the established technology of voltammetry, have been conceptually simple, but remain very impressive for their realisation and breadth of scope. Hemolysin perhaps is the most successfully exploited protein building block in bionano-technology, and is continuing to be developed, most recently through the development of a variant with 8-fold symmetry.[156] Hemolysin is a rare example of bionano-technology success in the commercial arena.

7 Examples of useful principles and methods

How best to assemble artificial protein structures including those discussed in this chapter is not a trivial problem. The use of symmetry has already been discussed in section 4. In this section we consider domain swapping and click chemistry.

7.1 Domain swapping

Domain swapping, the exchange of a polypeptide strand between two or more copies of a protein, is widely found in Nature. In fact, despite the name, there is no particular restriction on the structure of the protein part which is exchanged. This is not necessarily folded into a separate domain, and may be a simple β-strand or extended loop. The phenomenon is of relevance to the assembly of amyloid-like fibres and similar polymers associated with disease, and was reviewed in 2002 when about 40 domain-swapped structures were known[157] and also more recently.[158] In the last decade much progress has been made in understanding the thermodynamic aspects of this phenomenon.[159] The "domain" swapped between adjacent copies of the protein is almost always found at the N or C terminus, and has some inherent flexibility. RNase A is highly unusual since it can form domain-swapped structures either through its N or C terminus in mildly acid solutions. There is now a curated database of domain-swapped structures[160] (see http://caps.ncbs.res.in/3dswap/index.html). About 60 such models are currently present in PDB.

Cadherins are cell-cell adhesion proteins which exchange their N-terminal β-strands. Conserved tryptophan residues fit into complementary pockets

on the partner protein to form specific cadherin-cadherin dimers.[161] Cadherins within the same family have very similar sequences, yet the differences between them are significant, and strongly influence cellular interactions. The difference in binding affinity of N- and E-type cadherins is approximately 1 kcal/mol. This seemingly small energy difference is in fact highly conserved and required for cell-cell binding specificity. Domain swapping allows cadherins to bind to each other with low affinity but high specificity. It is partly driven by strains within the monomer form and anchoring interactions at each end of the β-strand which is exchanged.[162] In fact domain swapping is so sensitive to small energy changes that fairly conservative single site mutations can affect it, changing a monomer to a dimer. Domain swapping has an associated entropic cost of tying two or more protein molecules to each other, but this may be paid for by the relief of strain (or other unfavourable interactions) in the monomer.[163]

The possibilities of using domain swapping to engineer supramolecular assemblies are obvious[164] but so far there have been few examples published of successes. Although Nature has exploited this mechanism with great success, it depends upon a delicate thermodynamic balance which requires considerable experimentation to understand. One very interesting example is that of ubiquitin and barnase. Ha and colleagues[165] inserted ubiquitin (Ub) into one of six surface loops of barnase (Bn). In folded Ub, the distance between the N and C termini is 38 Å, which prevents the barnase from folding as a stable monomer. These ''BU'' proteins forms dimers and higher oligomers. The X-ray structure of one variety of the mutant shows the barnase folds to form a relaxed, domain-swapped oligomer, the oligomer sharing the same symmetry as the crystal. One can imagine that a single polymer molecule extends across the crystal itself. In this case, the isolated dimer form re-equilibrated with longer oligomers. Some control over oligomerization was possible by changing the length of the linkers between Ub and Bn; long linkers allowed the protein to form a stable monomer, and very short linkers were highly destabilizing to the monomer.[166] Control over higher order oligomerization was not demonstrated.

Domain-swapping is known to lead to fibrillogenesis in many cases, and Nagarkar and colleagues have discussed domain-swapped protein fibres and described the design of domain-swapping peptides to create hydrogels with desired properties.[167] The basic design is a peptide with exchangeable and non-exchangeable parts, which forms a β-turn. The dimer has hydrophilic and hydrophobic faces and can further associate to form a bilayer by hydrophobic interactions. The process is controllable by temperature since the hydrophobic effect is temperature-sensitive. The length of the exchangeable portion of the peptide influenced the local structure of the fibrils which form, and their mechanical properties. It was suggested that this type of peptide may be responsive to environmental changes such as redox potential or metal ion concentration. It is clear however that domain-swapping has great potential as a tool for constructing novel protein and peptide-based materials, either for use as sensors, biogels or nano-construction blocks of very high local order which readily

self-assemble into large arrays and present suitable functional groups for templating reactions.

7.2 Click chemistry

"Click chemistry" is a general term for methods of linking reliably different chemical entities. This modular approach to synthesis generally gives very high yields and uses simple reaction conditions. Although the idea was inspired by natural polymers such as peptides, the term is also more specifically used to indicate cycloadditions such as the Huisgen 1,3-dipolar reaction in which copper (I) ions are used as a catalyst to react an azide with an acetylene group. This has proved a very selective and reliable means of covalently attaching different substrates, and has found widespread uses outside organic chemistry, notably for labeling DNA, proteins, carbohydrates and other polymers.[168,169]

Click chemistry has been shown to provide a convenient means of bioconjugation through azide groups on proteins in a single step. Cysteines have been exploited for many years to add functional groups to proteins through covalent bonds, but electrophiles that react with cysteines are also prone to attack by amines and so are not always optimal. In any case, a completely orthogonal means of attaching any independent entity to a protein at a specific location leaves the experimenter free to use surface cysteine residues for other uses. To date there have not been many reports of protein decoration or modification by click chemistry. Several examples have been published[170,171] and the trend seems set to increase, however most examples use non-natural amino acids such as para-azidophenylalanine or homo-azidoalanine as the point of modification. The modified residue can be introduced by a variety of methods, but remains an inconvenience at best. An interesting advance is the facile transfer of azides to lysine residues on proteins, which allows either native or simply prepared mutant proteins to be modified directly. This reaction has traditionally been carried out in organic solvents unsuitable for proteins, but the introduction of imidazole-1-sulfonyl azide hydrochloride[172,173] opens up the possibility of conjugating folded proteins in aqueous solution. The only drawback is that high pH maybe required to activate the surface amine groups, and the azide transfer is not selective and tends to work with any free lysine. More recently, it has been shown that azides can be introduced into proteins in metal-free conditions using pH control.[174] This has particular advantages since copper ions may be difficult to remove from protein and are also toxic. Moreover it was found that at pH 8.5 the reaction is more specific, and only amines of suitable pKa are modified (such as the N-terminal amine). In the case of lysozyme the N-terminus was found not to be modified, presumably because of its low surface accessibility. There is a cost in efficiency however, compared to the metal-catalysed high pH reaction, which could only be partly compensated by excess reagent and increased reaction times. Importantly, the metal-free, low pH reaction conditions are compatible with the folding of most proteins, unlike the very high pH (around 11) used in earlier work. While optimization may require some experimentation in each case, it appears click chemistry could emerge as a general protein modification technique. It also offers a convenient means of attaching proteins (such as enzymes) to a particular scaffold.

8 *Ab initio* design

Proteins offer great potential for the production of nanoscale materials and functional devices/drugs, but this potential could perhaps best be unlocked if it were possible to design proteins quickly and reliably. To date, however, *ab initio* protein design has remained a very challenging area, the considerable effort devoted to this end notwithstanding. Even the design of simple protein folds remains difficult, despite the huge advances in computer speed and the rapidly growing number of protein structures. It is however possible to predict secondary structures with considerable accuracy, and it is known that these structures show preferred means of self-associating. The coiled-coil is one example of a well-understood protein-protein interaction which can be engineered successfully.[175] Recently a novel 32-residue peptide called CC-Hex has been designed using coiled-coils as a template, leading to a stable hexameric nano-tube structure with a central channel of 6 Å diameter.[176] This protein in fact has a new fold. The central channel is lined by leucine and isoleucine residues, and these can be replaced by polar residues such as aspartic acid or histidine. In the 1.75 Ångstrom resolution X-ray structure of the form carrying aspartic acid residues along the cavity it appears to be filled with water molecules. CC-Hex appears to be a potentially useful new scaffold, either as a simple tube or a membrane-spanning channel. It should be noted however that the designers expected the peptides to form a tetramer, not a hexamer. This underlines the fact that even protein designs built with super-secondary structures believed to obey known rules may produce surprises. Protein science is a long way from creating *ab initio* protein structures with confidence, and the use of naturally-occurring domains seems likely to dominate bionano-technology for some time to come. Other super-secondary structures which may be useful as nano building blocks are leucine zippers, which are a type of coiled-coil found in the DNA binding domain of a family of transcription factors.[177] Banta and colleagues have described the use of leucine zippers, EF-hand domains and eleastin-like polypeptides to create novel, functional hydrogels.[178]

9 Future perspective and applications

There are a number of promising avenues for the future of protein bionanoscience. These include:

1. The arrangement and patterning of inorganic materials into functional electronic devices. While protein and biological-molecule systems are unlikely ever to directly replace the components of mass-produced electronics which are already in the few tens of nanometer size regime, have incredibly high speed of function and far superior stability, protein may nevertheless be able to provide scaffolding and interfacing properties for specific situations, particularly biomedical applications such as biosensors.
2. Biodegradable smart drugs with the ability to sense their surroundings, and carry out an appropriate therapeutic response.
3. Continued development with other bionanotechnologies such as DNA origami and biomorphic mineralisation.

4. Fusion with semiconductor electronic devices – providing a human-machine interface functional at an extremely intimate level (might protein nanostructures for example be able to guide electrodes from silicon devices to specific points on single cells?) For medical use, such as in the production of smart drugs and medical nanodevices. Such technology could have applications in biosensing, locating diseased tissue, and treatment. Electronic devices may need to interface with cells with high fidelity to replace the function of damaged biological systems (for example in restoring hearing loss).

In the immediate future, progress seems likely to continue steadily using established, naturally-occurring protein scaffolds and motifs. A move towards completely artificial protein structures requires a very significant advance in our fundamental understanding of proteins, and even so may require computational and skill resources available to only a few groups. Although outwardly aimed at direct, practical objectives, bionanoscience may in fact provide novel insights into protein structure and thermodynamics; the refolding of the protein TRAP is one example. As hemolysin has shown however, there is already room in the market-place for bionanotechnology, and more examples of using proteins as sensors, interfaces and storage devices are likely to emerge over the coming years.

References

1 R. P. Feynman, *A transcript of this talk is available online (*www.zyvex.com/nanotech/feynman.html*)*, 1959.
2 F. F. Miranda, K. Iwasaki, S. Akashi, K. Sumitomo, M. Kobayashi, I. Yamashita, J. R. H. Tame and J. G. Heddle, *Small*, 2009, **5**, 2077–2084.
3 X. Chen, A. A. Antson, M. Yang, P. Li, C. Baumann, E. J. Dodson, G. G. Dodson and P. Gollnick, *J. Mol. Biol.*, 1999, **289**, 1003–1016.
4 E. Andrianantoandro, S. Basu, D. K. Karig and R. Weiss, *Mol Syst Biol*, 2006, **2**.
5 R. E. Smalley, *Sci Am*, 2001, **285**, 76–77.
6 D. M. Eigler and E. K. Schweizer, *Nature*, 1990, **344**, 524–526.
7 E. Drexler, *Nanosystems: Molecular Machinery, Manufacturing and Computation*, John Wiley & Sons Inc, New York, 1992.
8 R. A. Freitas and R. C. Merkle, *Kinematic Self-Replicating Machines*, Landes Bioscience, Georgetown, 2004.
9 J. J. Champoux, *Annual Review of Biochemistry*, 2001, **70**, 369–413.
10 C. Levinthal, *J. Med. Phys.*, 1968, **65**, 44–45.
11 R. R. Schaller, *IEEE Spectrum*, 1997, **34**, 52–59.
12 T. Ito and S. Okazaki, *Nature*, 2000, **406**, 1027–1031.
13 M. Lundstrom, *Science*, 2003, **299**, 210–211.
14 J. D. Newman and A. P. F. Turner, *Biosensors and Bioelectronics*, 2005, **20**, 2435–2453.
15 M. J. Bennett, M. R. Sawaya and D. Eisenberg, *Structure*, 2006, **14**, 811–824.
16 F. G. Pearce, S. H. Mackintosh and J. A. Gerrard, *J Agric Food Chem*, 2007, **55**, 318–322.
17 S. M. Loveday, X. L. Wang, M. A. Rao, S. G. Anema and H. Singh, *J Agric Food Chem*, 2011, **59**, 8467–8474.
18 V. Pipich, M. Balz, S. E. Wolf, W. Tremel and D. Schwahn, *Journal of the American Chemical Society*, 2008, **130**, 6879–6892.

19 P. Pal, M. Mahato, T. Kamilya, B. Tah, R. Sarkar and G. B. Talapatra, *J Phys Chem B*, 2011, **115**, 4259–4265.
20 C. Lara, J. Adamcik, S. Jordens and R. Mezzenga, *Biomacromolecules*, 2011, **12**, 1868–1875.
21 L. H. Beun, X. J. Beaudoux, J. M. Kleijn, F. A. de Wolf and M. A. Cohen Stuart, *ACS Nano*, 2011, **6**, 133–140.
22 J. M. Smeenk, M. B. Otten, J. Thies, D. A. Tirrell, H. G. Stunnenberg and J. C. van Hest, *Angew Chem Int Ed Engl*, 2005, **44**, 1968–1971.
23 Q. Cao and J. A. Rogers, *Advanced Materials*, 2009, **21**, 29–53.
24 E. T. Thostenson, Z. Ren and T.-W. Chou, *Composites Science and Technology*, 2001, **61**, 1899–1912.
25 K. Kostarelos, *Nat Biotech*, 2008, **26**, 774–776.
26 X. Gao and H. Matsui, *Advanced Materials*, 2005, **17**, 2037–2050.
27 J. D. Hartgerink, E. Beniash and S. I. Stupp, *Science*, 2001, **294**, 1684–1688.
28 M. Reches and E. Gazit, *Science*, 2003, **300**, 625–627.
29 T. Scheibel, R. Parthasarathy, G. Sawicki, X. M. Lin, H. Jaeger and S. L. Lindquist, *Proc Natl Acad Sci USA*, 2003, **100**, 4527–4532.
30 Z. Milton, in *Discoveries in Plant Biology*, eds. S.-D. Kung and S.-F. Yang, World Scientific Publishing, Singapore, 1998, pp. 105–110.
31 J. D. Bernal and I. Fankuchen, *J. Gen. Phys*, 1941, **25**, 147–165.
32 P. J. G. Butler, *Journal of General Virology*, 1984, **65**, 253–279.
33 R. N. Beachy and M. Zaitlin, *Virology*, 1977, **81**, 160–169.
34 T. R. Hunter, T. Hunt, J. Knowland and D. Zimmern, *Nature*, 1976, **260**, 759–764.
35 A. C. H. Durham, J. T. Finch and A. Klug, *Nature New Biol*., 1971, **229**, 510.
36 Y. Okada, *Adv. Biophys*., 1986, **22**, 95–149.
37 D. Zimmern, *Cell*, 1977, **11**, 463–482.
38 T. M. Schuster, R. B. Scheele, M. L. Adams, S. J. Shire, J. J. Steckert and M. Potschka, *Biophys J*, 1980, **32**, 313–329.
39 B. Bhyravbhatla, S. J. Watowich and D. L. Caspar, *Biophys J*, 1998, **74**, 604–615.
40 D. L. Caspar, *Biophys J*, 1980, **32**, 103–138.
41 G. Lebeurier, A. Nicolaieff and K. E. Richards, *Proceedings of the National Academy of Sciences of the United States of America*, 1977, **74**, 149–153.
42 D. R. Turner, C. J. McGuigan and P. J. Butler, *J Mol Biol*, 1989, **209**, 407–422.
43 A. Mueller, A. Kadri, H. Jeske and C. Wege, *Journal of Virological Methods*, 2010, **166**, 77–85.
44 W. Shenton, T. Douglas, M. Young, G. Stubbs and S. Mann, *Adv. Mater*., 1999, **11**, 253–256.
45 H. Tsukamoto, M.-a. Kawano, T. Inoue, T. Enomoto, R.-u. Takahashi, N. Yokoyama, N. Yamamoto, T. Imai, K. Kataoka, Y. Yamaguchi and H. Handa, *Genes to Cells*, 2007, **12**, 1267–1279.
46 M. Knez, A. M. Bittner, F. Boes, C. Wege, H. Jeske, E. Maib and K. Kern, *Nano Lett*., 2003, **3**, 1079–1082.
47 R. Tsukamoto, M. Muraoka, M. Seki, H. Tabata and I. Yamashita, *Chem. Mater*., 2007, **19**, 2389–2391.
48 A. Klug, *Philos Trans R Soc Lond B Biol Sci*, 1999, **354**, 531–535.
49 R. A. Miller, A. D. Presley and M. B. Francis, *J. Am. Chem. Soc*., 2007, **129**, 3104–3109.
50 O. K. Zahr and A. S. Blum, *Nano Letters*, 2011.
51 R. A. Vega, D. Maspoch, K. Salaita and C. A. Mirkin, *Angewandte Chemie*, 2005, **117**, 6167–6169.

52 H. Li, D. J. DeRosier, W. V. Nicholson, E. Nogales and K. H. Downing, *Structure*, 2002, **10**, 1317–1328.
53 R. M. Macnab, *Journal of Bacteriology*, 1999, **181**, 7149–7153.
54 W. R. Hesse, L. Luo, G. Zhang, R. Mulero, J. Cho and M. J. Kim, *Materials Science and Engineering: C*, 2009, **29**, 2282–2286.
55 M. T. Kumara, B. C. Tripp and S. Muralidharan, *Chemistry of Materials*, 2007, **19**, 2056–2064.
56 M. T. Kumara, B. C. Tripp and S. Muralidharan, *Biomacromolecules*, 2007, **8**, 3718–3722.
57 H. C. Hyman and S. Trachtenberg, *Journal of Molecular Biology*, 1991, **220**, 79–88.
58 S. Kanto, H. Okino, S.-I. Aizawa and S. Yamaguchi, *Journal of Molecular Biology*, 1991, **219**, 471–480.
59 W. Jo, K. J. Freedman, D. K. Yi and M. J. Kim, *Nanotechnology*, 2012, **23**, 055601.
60 T.-X. Fan, S.-K. Chow and D. Zhang, *Progress in Materials Science*, 2009, **54**, 542–659.
61 J. C. Zhou, Y. Gao, A. A. Martinez-Molares, X. Jing, D. Yan, J. Lau, T. Hamasaki, C. S. Ozkan, M. Ozkan, E. Hu and B. Dunn, *Small*, 2008, **4**, 1507–1515.
62 R. Penrose, *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*, Oxford University Press, 1989.
63 A. Gennerich and R. Vale, 2009, **21**, 59.
64 Y. Yang, P. A. Deymier, L. Wang, R. Guzman, J. B. Hoying, H. J. McLaughlin, S. D. Smith and I. N. Jongewaard, *Biotechnology Progress*, 2006, **22**, 303–312.
65 N. Hirokawa, Y. Shiomura and S. Okabe, *The Journal of Cell Biology*, 1988, **107**, 1449–1459.
66 P. Babitzke and P. Gollnick, *J. Bacteriol.*, 2001, **183**, 5795–5802.
67 A. A. Antson, E. J. Dodson, G. Dodson, R. B. Greaves, X. Chen and P. Gollnick, *Nature*, 1999, **401**, 235–242.
68 A. A. Antson, J. Otridge, A. M. Brzozowski, E. J. Dodson, G. G. Dodson, K. S. Wilson, T. M. Smith, M. Yang, T. Kurecki and P. Gollnick, *Nature*, 1995, **374**, 693–700.
69 J. G. Heddle, T. Yokoyama, I. Yamashita, S.-Y. Park and J. R. H. Tame, *Structure*, 2006, **14**, 925–933.
70 J. G. Heddle, I. Fujiwara, H. Yamadaki, S. Yoshii, K. Nishio, C. Addy, I. Yamashita and J. R. H. Tame, *Small*, 2007, **3**, 1950–1956.
71 M. B. Shevtsov, Y. Chen, P. Gollnick and A. A. Antson, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 17600–17605.
72 D. Snyder, J. Lary, Y. Chen, P. Gollnick and J. L. Cole, *J Mol Biol*, 2004, **338**, 669–682.
73 M. Watanabe, J. G. Heddle, K. Kikuchi, S. Unzai, S. Akashi, S. Y. Park and J. R. Tame, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 2176–2181.
74 J. D. Mougous, M. E. Cuff, S. Raunser, A. Shen, M. Zhou, C. A. Gifford, A. L. Goodman, G. Joachimiak, C. L. Ordonez, S. Lory, T. Walz, A. Joachimiak and J. J. Mekalanos, *Science*, 2006, **312**, 1526–1530.
75 E. R. Ballister, A. H. Lai, R. N. Zuckermann, Y. Cheng and J. D. Mougous, 2008, **105**, 3733–3738.
76 D. Pelah, O. Shoseyov and A. Altman, *Tree Physiol*, 1995, **15**, 673–678.
77 D. Pelah, O. Shoseyov, A. Altman and D. Bartels, *Journal of Plant Physiology*, 1997, **151**, 96–100.

78 D. Pelah, W. X. Wang, A. Altman, O. Shoseyov and D. Bartels, *Phsiol. Plantarum*, 1997, **99**, 153–159.
79 W.-X. Wang, D. Pelah, T. Alergand, O. Shoseyov and A. Altman, *Plant Physiol.*, 2002, **130**, 865–875.
80 O. Dgany, A. Gonzalez, O. Sofer, W. Wang, G. Zolotnitsky, A. Wolf, Y. Shoham, A. Altman, S. G. Wolf, O. Shoseyov and O. Almog, *J. Biol. Chem.*, 2004, **279**, 51516–51523.
81 I. Medalsy, O. Dgany, M. Sowwan, H. Cohen, A. Yukashevska, S. G. Wolf, A. Wolf, A. Koster, O. Almog, I. Marton, Y. Pouny, A. Altman, O. Shoseyov and D. Porath, *Nano Lett.*, 2008, **8**, 473–477.
82 I. Medalsy, M. Klein, A. Heyman, O. Shoseyov, R. D. F. Remacle, R. D. Levine and D. Porath, *Nat Nano*, 2010, **5**, 451–457.
83 T. Hayashi, K.-I. Sano, K. Shiba, Y. Kumashiro, K. Iwahori, I. Yamashita and M. Hara, *Nano Lett.*, 2006.
84 K.-I. Sano, H. Sasaki and K. Shiba, *Langmuir*, 2005, **21**, 3090–3095.
85 K.-I. Sano, H. Sasaki and K. Shiba, *Journal of the American Chemical Society*, 2006, **128**, 1717–1722.
86 A. Heyman, I. Medalsy, O. BetOr, O. Dgany, M. Gottlieb, D. Porath and O. Shoseyov, *Angewandte Chemie International Edition*, 2009, **48**, 9290–9294.
87 T. Douglas and M. Young, *Science*, 2006, **312**, 873–875.
88 T. Lin, Z. Chen, R. Usha, C. V. Stauffacher, J. B. Dai, T. Schmidt and J. E. Johnson, *Virology*, 1999, **265**, 20–34.
89 N. F. Steinmetz, G. P. Lomonossoff and D. J. Evans, *Langmuir*, 2006, **22**, 3488–3490.
90 P. Singh, D. Prasuhn, R. M. Yeh, G. Destito, C. S. Rae, K. Osborn, M. G. Finn and M. Manchester, *Journal of Controlled Release*, 2007, **120**, 41–50.
91 N. F. Steinmetz, S. N. Shah, J. E. Barclay, G. Rallapalli, G. P. Lomonossoff and D. J. Evans, *Small*, 2009, **5**, 813–816.
92 S. N. Shah, N. F. Steinmetz, A. A. A. Aljabali, G. P. Lomonossoff and D. J. Evans, *Dalton Transactions*, 2009, 8479–8480.
93 A. A. A. Aljabali, J. E. Barclay, O. Cespedes, A. Rashid, S. S. Staniland, G. P. Lomonossoff and D. J. Evans, *Adv. Funct. Mater.*, 2011, **21**, 4137–4142.
94 A. A. Aljabali, J. E. Barclay, G. P. Lomonossoff and D. J. Evans, *Nanoscale*, 2010, **2**, 2596–2600.
95 C. Porta, V. E. Spall, K. C. Findlay, R. C. Gergerich, C. E. Farrance and G. P. Lomonossoff, *Virology*, 2003, **310**, 50–63.
96 J. G. Heddle, *Nanotechnology, Science and Applications*, 2008, **1**, 67–78.
97 J. P. Langeveld, F. R. Brennan, J. L. Martinez-Torrecuadrada, T. D. Jones, R. S. Boshuizen, C. Vela, J. I. Casal, S. Kamstrup, K. Dalsgaard, R. H. Meloen, M. M. Bendig and W. D. Hamilton, *Vaccine*, 2001, **19**, 3661–3670.
98 K. Dalsgaard, A. Uttenthal, T. D. Jones, F. Xu, A. Merryweather, W. D. Hamilton, J. P. Langeveld, R. S. Boshuizen, S. Kamstrup, G. P. Lomonossoff, C. Porta, C. Vela, J. I. Casal, R. H. Meloen and P. B. Rodgers, *Nat Biotechnol*, 1997, **15**, 248–252.
99 Y. Ren, S. M. Wong and L. Y. Lim, *Bioconjug Chem*, 2007, **18**, 836–843.
100 F. M. Brunel, J. D. Lewis, G. Destito, N. F. Steinmetz, M. Manchester, H. Stuhlmann and P. E. Dawson, *Nano Letters*, 2010, **10**, 1093–1097.
101 J. Atabekov, N. Nikitin, M. Arkhipenko, S. Chirkov and O. Karpova, *Journal of General Virology*, 2011, **92**, 453–456.
102 F. A. Anderer, H. D. Schlumberger, M. A. Koch, H. Frank and H. J. Eggers, *Virology*, 1967, **32**, 511–523.
103 R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin and S. C. Harrison, *Nature*, 1991, **354**, 278–284.

104 M.-a. Kawano, T. Inoue, H. Tsukamoto, T. Takaya, T. Enomoto, R.-u. Takahashi, N. Yokoyama, N. Yamamoto, A. Nakanishi, T. Imai, T. Wada, K. Kataoka and H. Handa, *J. Biol. Chem.*, 2006, **281**, 10164–10173.
105 S. N. Kanesashi, K. Ishizu, M. A. Kawano, S. I. Han, S. Tomita, H. Watanabe, K. Kataoka and H. Handa, *J. Gen. Virol.*, 2003, **84**, 1899–1905.
106 Y. Kitai, H. Fukuda, T. Enomoto, Y. Asakawa, T. Suzuki, S. Inouye and H. Handa, *Journal of Biotechnology*, 2011, **155**, 251–256.
107 M. Vera and P. Fortes, *DNA Cell Biol*, 2004, **23**, 271–282.
108 F. W. Outten and E. C. Theil, *Antioxid Redox Signal*, 2009, **11**, 1029–1046.
109 E. S. Henle and S. Linn, *Journal of Biological Chemistry*, 1997, **272**, 19095–19098.
110 S. H. Banyard, D. K. Stammers and P. M. Harrison, *Nature*, 1978, **271**, 282–284.
111 K. Yoshizawa, Y. Mishima, S. Y. Park, J. G. Heddle, J. R. Tame, K. Iwahori, M. Kobayashi and I. Yamashita, *J. Biochem. (Tokyo)*, 2007, **142**, 707–713.
112 D. M. Lawson, P. J. Artymiuk, S. J. Yewdall, J. M. A. Smith, J. C. Livingstone, A. Treffry, A. Luzzago, S. Levi, P. Arosio, G. Cesareni, C. D. Thomas, W. V. Shaw and P. M. Harrison, *Nature*, 1991, **349**, 541–544.
113 T. Douglas, D. P. Dickson, S. Betteridge, J. Charnock, C. D. Garner and S. Mann, *Science*, 1995, **269**, 54–57.
114 S. Mann, J. M. Williams, A. Treffry and P. M. Harrison, *J Mol Biol*, 1987, **198**, 405–416.
115 F. C. Meldrum, T. Douglas, S. Lei, P. Arosio and S. Mann, *J. Inorg. Biochem.*, 1995, **58**, 59–68.
116 F. C. Meldrum, B. R. Heywood and S. Mann, *Science*, 1992, **257**, 522–523.
117 F. C. Meldrum, V. J. Wade, D. L. Nimmo, B. R. Heywood and S. Mann, *Nature*, 1991, **349**, 684–687.
118 M. Okuda, K. Iwahori, I. Yamashita and H. Yoshimura, 2003, **84**, 187–194.
119 K. Iwahori, K. Yoshizawa, M. Muraoka and I. Yamashita, *Inorg. Chem.*, 2005, **44**, 6393–6400.
120 K. Yoshizawa, K. Iwahori, K. Sugimoto and I. Yamashita, *Chem. Lett.*, 2006, **35**, 1192.
121 A. Miura, T. Hikono, T. Matsumura, H. Yano, T. Hatayama, Y. Uraoka, T. Fuyuki, S. Yoshii and I. Yamashita, *Jpn. J. Appl. Phys.*, 2006, **45**, L1–L3.
122 K. Ohara, Y. Uraoka, T. Fuyuki, I. Yamashita, T. Yaegashi, M. Moniwa and M. Yoshimaru, *JPN. J. APPL. PHYS*, 2009, **48**, 04C153.
123 M. Uenuma, K. Kawano, B. Zheng, N. Okamoto, M. Horita, S. Yoshii, I. Yamashita and Y. Uraoka, *Nanotechnology*, 2011, **22**, 215201.
124 B. Cohen, H. Dafni, G. Meir, A. Harmelin and M. Neeman, *Neoplasia*, 2005, **7**, 109–117.
125 M. Almiron, A. J. Link, D. Furlong and R. Kolter, *Genes Dev*, 1992, **6**, 2646–2654.
126 O. L. Lomovskaya, J. P. Kidwell and A. Matin, *J Bacteriol*, 1994, **176**, 3928–3935.
127 A. Martinez and R. Kolter, *J Bacteriol*, 1997, **179**, 5188–5194.
128 S. Altuvia, M. Almiron, G. Huisman, R. Kolter and G. Storz, *Mol Microbiol*, 1994, **13**, 265–272.
129 A. Ilari, S. Stefanini, E. Chiancone and D. Tsernoglou, *Nat Struct Biol*, 2000, **7**, 38–43.
130 D. A. Resnick, K. Gilmore, Y. U. Idzerda, M. T. Klem, M. Allen, T. Douglas, E. Arenholz and M. Young, *Journal of Applied Physics*, 2006, **99**, 08Q501.
131 K. Iwahori, T. Enomoto, H. Furusho, A. Miura, K. Nishio, Y. Mishima and I. Yamashita, *Chem. Mater.*, 2007, **19**, 3105–3111.

132 D. Kase, J. L. Kulp, M. Yudasaka, J. S. Evans, S. Iijima and K. Shiba, *Langmuir*, 2004, **20**, 8939–8941.
133 K. Sano and K. Shiba, *J. Am. Chem. Soc*., 2003, **125**, 14234–14235.
134 I. Inoue, B. Zheng, K. Watanabe, Y. Ishikawa, K. Shiba, H. Yasueda, Y. Uraoka and I. Yamashita, *Chemical Communications*, 2011, **47**, 12649–12651.
135 Y. J. Lee, H. Yi, W. J. Kim, K. Kang, D. S. Yun, M. S. Strano, G. Ceder and A. M. Belcher, *Science*, 2009.
136 D. Eder, *Chemical Reviews*, 2010, **110**, 1348–1385.
137 K. Sugimoto, S. Kanamaru, K. Iwasaki, F. Arisaka and I. Yamashita, *Angew. Chem. Int. Ed.*, 2006, **45**, 2725–2728.
138 S. Kanamaru, P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka and M. G. Rossmann, *Nature*, 2002, **415**, 553–557.
139 J. E. Padilla, C. Colovos and T. O. Yeates, *Proceedings of the National Academy of Sciences*, 2001, **98**, 2217–2221.
140 S. Raman, G. Machaidze, A. Lustig, U. Aebi and P. Burkhard, *Nanomedicine: Nanotechnology, Biology and Medicine*, 2006, **2**, 95–102.
141 H. J. Hecht, H. Sobek, T. Haag, O. Pfeifer and K. H. van Pee, *Nat Struct Mol Biol*, 1994, **1**, 532–537.
142 B. Sha and M. Luo, *Nat Struct Mol Biol*, 1997, **4**, 239–244.
143 B. Lamarre and M. G. Ryadnov, *Macromolecular Bioscience*, 2011, **11**, 503–513.
144 A. D. Malay, J. G. Heddle, S. Tomita, K. Iwasaki, N. Miyazaki, K. Sumitomo, H. Yanagi, I. Yamashita and Y. Uraoka, *Nano Lett*., 2012, **12**, 2056-2059.
145 X. Chen, A. A. Antson, M. Yang, P. Li, C. Baumann, E. J. Dodson, G. G. Dodson and P. Gollnick, *J Mol Biol*, 1999, **289**, 1003–1016.
146 M. B. Cortie and E. van der Lingen, *Mater. Forum*, 2002, **26**, 1–14.
147 G. Schmid, *Chem. Soc. Rev*., 2008, **37**, 1909–1930.
148 M. A. Krol, N. H. Olson, J. Tate, J. E. Johnson, T. S. Baker and P. Ahlquist, *Proc. Natl. Acad. Sci. U.S.A*., 1999, **96**, 13650–13655.
149 J. Tang, J. M. Johnson, K. A. Dryden, M. J. Young, A. Zlotnick and J. E. Johnson, *J. Struct. Biol*., 2006, **154**, 59–67.
150 L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley and J. E. Gouaux, *Science*, 1996, **274**, 1859–1866.
151 H. Bayley, *Mol Biosyst*, 2007, **3**, 645–647.
152 O. Braha, B. Walker, S. Cheley, J. J. Kasianowicz, L. Song, J. E. Gouaux and H. Bayley, *Chem Biol*, 1997, **4**, 497–505.
153 L. Q. Gu, O. Braha, S. Conlan, S. Cheley and H. Bayley, *Nature*, 1999, **398**, 686–690.
154 W. W. Li, T. D. Claridge, Q. Li, M. R. Wormald, B. G. Davis and H. Bayley, *J Am Chem Soc*, 2011.
155 D. Stoddart, A. J. Heron, J. Klingelhoefer, E. Mikhailova, G. Maglia and H. Bayley, *Nano Lett*, 2010, **10**, 3633–3637.
156 A. F. Hammerstein, L. Jayasinghe and H. Bayley, *J Biol Chem*, 2011, **286**, 14324–14334.
157 Y. Liu and D. Eisenberg, *Protein Sci*, 2002, **11**, 1285–1299.
158 A. M. Gronenborn, *Curr Opin Struct Biol*, 2009, **19**, 39–49.
159 A. Malevanets, F. L. Sirota and S. J. Wodak, *J Mol Biol*, 2008, **382**, 223–235.
160 K. Shameer, G. Pugalenthi, K. K. Kandaswamy, P. N. Suganthan, G. Archunan and R. Sowdhamini, *Bioinform Biol Insights*, 2010, **4**, 33–42.
161 O. J. Harrison, F. Bahna, P. S. Katsamba, X. Jin, J. Brasch, J. Vendome, G. Ahlsen, K. J. Carroll, S. R. Price, B. Honig and L. Shapiro, *Nat Struct Mol Biol*, 2010, **17**, 348–357.

162 J. Vendome, S. Posy, X. Jin, F. Bahna, G. Ahlsen, L. Shapiro and B. Honig, *Nat Struct Mol Biol*, 2011, **18**, 693–700.
163 J. W. O'Neill, D. E. Kim, K. Johnsen, D. Baker and K. Y. Zhang, *Structure*, 2001, **9**, 1017–1027.
164 T. O. Yeates and J. E. Padilla, *Curr Opin Struct Biol*, 2002, **12**, 464–470.
165 J. H. Ha, J. M. Karchin, N. Walker-Kopp, L. S. Huang, E. A. Berry and S. N. Loh, *J Mol Biol*, 2012, **416**, 495–502.
166 T. A. Cutler, B. M. Mills, D. J. Lubin, L. T. Chong and S. N. Loh, *J Mol Biol*, 2009, **386**, 854–868.
167 R. P. Nagarkar, R. A. Hule, D. J. Pochan and J. P. Schneider, *Biopolymers*, 2010, **94**, 141–155.
168 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew Chem Int Ed Engl*, 2001, **40**, 2004–2021.
169 J. E. Moses and A. D. Moorhouse, *Chem Soc Rev*, 2007, **36**, 1249–1262.
170 S. Schoffelen, M. H. Lambermon, M. B. van Eldijk and J. C. van Hest, *Bioconjug Chem*, 2008, **19**, 1127–1131.
171 S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. G. Davis, *Nature*, 2007, **446**, 1105–1109.
172 E. D. Goddard-Borger and R. V. Stick, *Organic Letters*, 2007, **9**, 3797–3800.
173 S. F. M. van Dongen, R. L. M. Teeuwen, M. Nallani, S. S. van Berkel, J. J. L. M. Cornelissen, R. J. M. Nolte and J. C. M. van Hest, *Bioconjugate Chemistry*, 2008, **20**, 20–23.
174 S. Schoffelen, M. B. van Eldijk, B. Rooijakkers, R. Raijmakers, A. J. R. Heck and J. C. M. van Hest, *Chemical Science*, 2011, **2**.
175 M. G. Ryadnov, A. Bella, S. Timson and D. N. Woolfson, *J Am Chem Soc*, 2009, **131**, 13240–13241.
176 N. R. Zaccai, B. Chi, A. R. Thomson, A. L. Boyle, G. J. Bartlett, M. Bruning, N. Linden, R. B. Sessions, P. J. Booth, R. L. Brady and D. N. Woolfson, *Nat Chem Biol*, 2011, **7**, 935–941.
177 M. G. Ryadnov, D. Papapostolou and D. N. Woolfson, eds. E. Gazit and R. Nussinov, Humana Press, 2008, pp. 35–51.
178 S. Banta, I. R. Wheeldon and M. Blenner, *Annu Rev Biomed Eng*, 2010, **12**, 167–186.

Prescriptive peptide design

Maxim G Ryadnov
DOI: 10.1039/9781849734677-00190

1 Introduction

This chapter highlights developments in peptide design over the last years to the time of its submission, with background information covering an unlimited timeframe. An emphasis is made on prescriptive or de novo design of polypeptide sequences in relation to specific folds, topologies and functions, as seen in native systems. Therefore, the choice of reviewed designs in this chapter is biased towards biologically relevant structures that span molecular and nano- to micrometer length scales, and towards native strategies for their construction, that is self-assembly. The chapter reviews literature sourced from different databases including Web of Science and PubMed. Individual sections are arranged according to structural preferences and definitions and consequently to the types of relevant de novo designs. A section describing basic design principles provides introduction to the problem of peptide design and structure-function relationships.

2 Generic considerations for prescriptive peptide design

2.1 Peptide design versus protein design

The last 20 years have witnessed considerable progress in designing novel peptides and peptide-based materials.[1–3] The progress provides a growing repertoir of validated rules enabling the construction of increasingly complex and functionally sophisticated structures. This has already earned peptide design a status of an independent discipline which deserves a dedicated attention.

A question remains with regard to what makes *peptide design* different from *protein design*.

Both approaches aim at finding a rationale that will allow for a better control over polypeptide structure. Both study structure-function relationships that are free from the constraints of natural selection. Both deal with protein folding.

One factor that discriminates between the two is the extent to which these strategies and relationships can be applied. It may be fairly straightforward to predict the sequence of a simple folding motif, *i.e.* an α-helix or β-sheet, but predicting sequence-folding pathways for a functional protein comprising several different motifs, each of which can be unique, is a formidable task.

In a way, peptide design is a simplification of protein design. This simplification concerns with the limited sequence space of peptides when compared to proteins. Indeed, the folding and oligomerisation of shorter

National 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

sequences are relatively easy to assign. However, at large existing design rules for peptides are based on sequence stretches found in native proteins. Not only these fragments are shorter than full protein sequences, but often they cannot fold autonomously, that is, without the support of the rest of the protein structure, which serves as a self-constraining scaffold. Particularly challenging this is for very short sequences (< 10 amino-acid residues) which may lose their original biological activity when used in a free form.

For example, many unrelated proteins have strong affinity for integrin receptors by virtue of containing one universal epitope, the RGD sequence, originally discovered in fibronectin (Fig. 1).[4,5] All these proteins provide loop structures that expose this peptide for binding by integrins and integrin associated receptors. In addition, the sequence has a rigid turn-like conformation which is fixed by the framework of the whole folded protein.[6] When cut from the protein and used as is the sequence is less active suggesting the loss of the biologically active conformation. In contrast, constraining it in a cyclopeptide and forcing a predominant turn conformation can give rise to a superpotent integrin antagonist.[4–6]

Peptide design has a number of advanages over protein design. Firstly, the synthesis of peptides up to 50 amino acids long is now routine and affordable. The turnaround time from the design stage to a pure material can be under one week. Secondly, peptides are more amenable to iterative designs with a smaller number of alternative sequences. Thirdly, empirically it is more straightforward to impose the principles of negative (away from undesired structures)[7] and positive (towards desired structures)[7,8] designs to the selection of individual folding motifs than to fully folded proteins.

More importantly, it is becoming increasingly evident that rationally designed peptides can deliver functions with efficiencies that are at least similar to those of proteins. This stems from that those proteins that can execute various functions allocate specific fragments to these functions in their sequences, *i.e.* peptides.

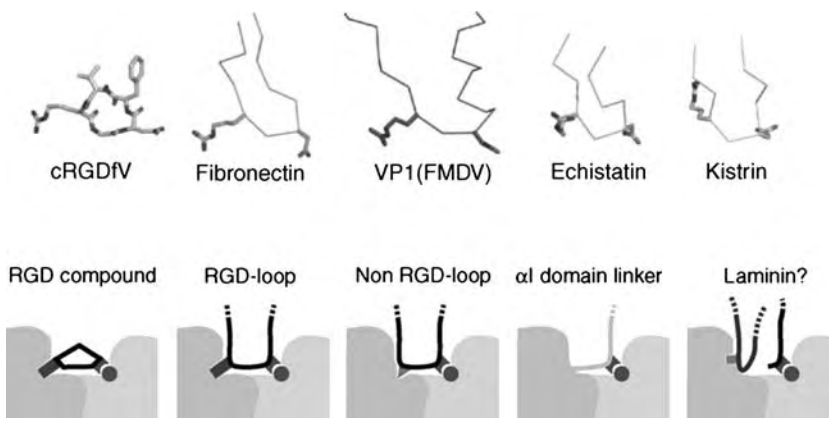


Fig. 1 Structures of RGD-containing loops from known integrin ligands and a cyclic RGD peptide cilengitide (Reprinted from J. Takagi, Structural basis for ligand recognition by integrins, *Curr. Opin. Cell Biol.*, **19**, 557, Copyright (2007), with permission from Elsevier).[5]

2.2 *De novo* design versus biomimetic design

Using the above example, further improvements and structural optimisations to the RGD epitope can be made either in its primary structure or at the secondary structure level.

The RGD peptide is a small molecule with the molecular weight of less than 500 Da. This perfectly adjusts to the Lipinski's rule[9] making the peptide an attractive drug candidate. However, given that peptides are notoriously unstable against enzymatic degradation main optimisation targets for the sequence concern with creating a more stable chemical structure. In addition, side chains and their positional arrangements are crucial determinants for the biological activity of peptides. Therefore, in generating more active compounds particular attention is given to the enhanced enzymatic resistance of the polypeptide backbone while retaining the position of side chain pharmacophores. This underpins peptidomimetic designs. Figure 2 gives examples of peptidomimetics derived from the parent RGD sequence.[10]

An alternative route is to focus on the bioactive conformation of the epitope. Here the target is to fix or lock this conformation by imposing a secondary structure template, which often implies the incorporation of additional residues to induce the right side-chain topology or in other words the right exposure of the pharmacophore (Fig. 3).[6] Cyclization of epitope sequences extended with conformation-inducing residues provides efficient templates for stable interactions between the side chains of the epitope and the target protein.

However, the spatial fitting of side chains may not be sufficient as activity often requires interactions from the peptide backbone.[11] For instance, during the development of a powerful angiogenisis inhibitor cilengitide it was noted that peptides with the same side-chain topology differed in their

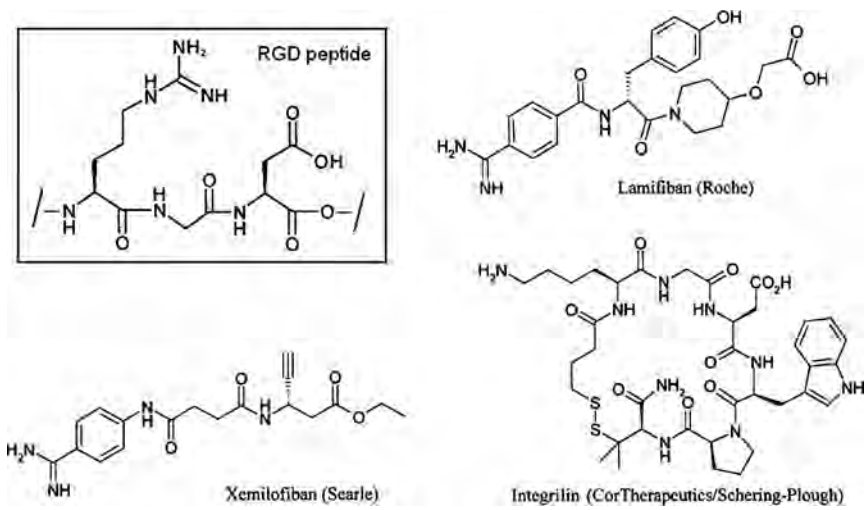


Fig. 2 Chemical structures of RGD peptidomimetics developed by the pharmaceutical industry. (Springer and G. P. Curley, H. Blum and M. J. Humphries, *Cell. Mol. Life Sci.*, **56**, 1999, 427, Integrin antagonists, Fig. 4, original copyright notice is given to the publication in which the material was originally published, by adding: "With kind permission from Springer Science and Business Media").[10]

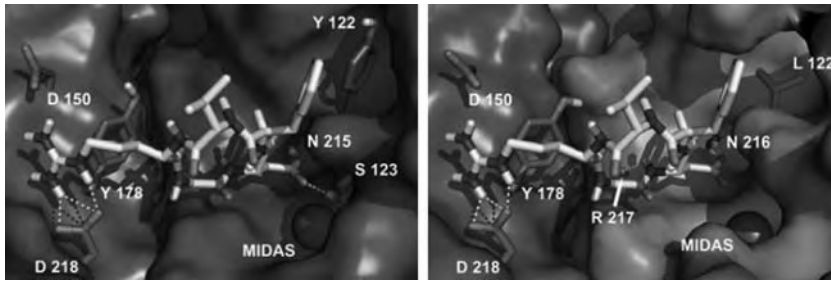


Fig. 3 Cilengitide (stick model) bound to αvβ3 (left) and to αvβ5 (right) integrins (reproduced with permission from C. Mas-Moruno, F. Rechenmacher and H. Kessler, Cilengitide: the first anti-angiogenic small molecule drug candidate. Design, synthesis and clinical evaluation, *Anticancer Agents Med Chem*. 2010, **10**, 753).[6]

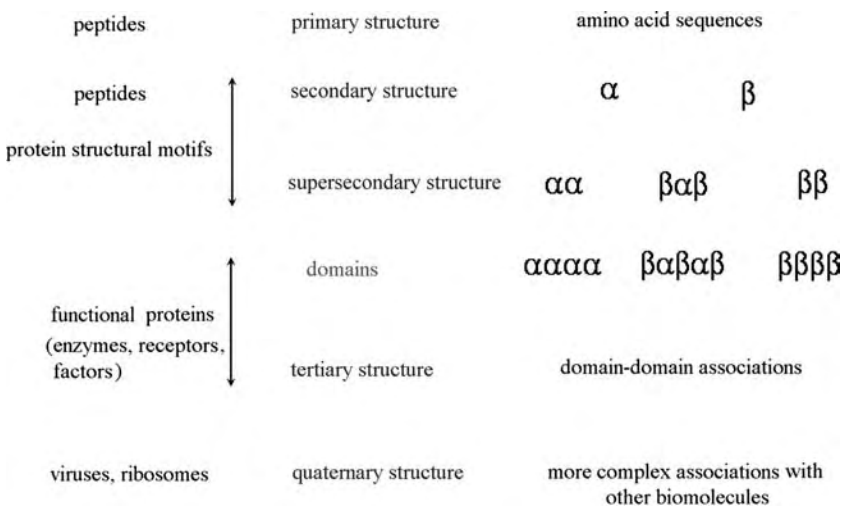


Fig. 4 Schematic representation of protein folding hierarchy.

ability to bind the targeted $\alpha_v\beta_3$ integrin.[12] The only difference between these compounds was the directionality of the peptide bonds: parent peptide and its retro-inverse analogue. This indicates that it is necessary for the peptide backbone to interact with the receptor, and that these interactions are synergestic with those furnished by the corresponding side chains.[11,12] This emphasises the importance of sequential relationships at the primary and secondary structure levels for biological function. Such relationships may be more critical and certainly are much better understood than those that occur at higher hierarchical levels of protein folding, *i.e.* between secondary and supersecondary, tertiary and quaternary structures (Fig. 4).

The current understanding of how the primary structure of a peptide defines its secondary structure and mechanistic function certainly is not complete, but proves to be satisfactory to explore novel regular structures and perhaps more importantly the ways these relate to native structural

motifs. In this respect, designing peptide sequences "from scratch" represents a fairly efficient tool for revealing new rules that will enable exploitable structure-function predictions. This is the realm of de novo peptide design.

3 Structural space of de novo design

De novo peptide design is primarily focused on the construction of secondary and supersecondary structures. Designed peptide sequences are long enough to fold into regularly repeating structures and are short enough to be expected to fold predictably. Thus, the same reasoning which distinguishes between peptide and protein designs applies to de novo peptide and protein designs.

Two relatively independent design routes are being followed. One targets autonomously folded structures (secondary) and another one concerns with designing higher order but also autonomous (super-secondary) structures. These often overlap due to the inherent limitations of de novo design which remains to be predominantly empirical and hence semi-iterative. Further structures are designed to construct self-assembling systems as a basis for nanostructured materials, which will be discussed later on in the chapter. De novo design ultimately seeks new approaches to extend our understanding of protein folding and function and to broaden the repertoir of functional structures for applications. Therefore, such a classification rather reflects upon the state of the art than represents a systemic approach to rational design. The types of possible secondary structures are restricted to those that already exist in native systems, which directly conforms to the very nature of the peptide bond and to the steric constraints of amino-acid side chains that influence conformations of torsion angles of a polypeptide chain. The sequence composition of α-amino acids is the main variable which guides primary structure into a specific fold. It is not surprising therefore that alternative routes using non-natural amino-acid backbones are being tried. Drastically different and purely artificial conformations can be obtained from sequences composed of β- and γ-amino acids. This has already led to the emergence of a new class of folding elements, dubbed foldamers, which are expanding the repertoir of peptide folds available for design (Fig. 5).[13–15]

There exist no native counterparts for foldamers making them de novo structures by definition.[16] This also draws a clear separation line for relevance to native biology and whereupon only de novo α-peptide designs will be discussed.

3.1 Secondary and supersecondary structure motifs

In practical terms, de novo design provides a strategy to emulate specific native folding motifs. Proteins and their domains are geometrically arranged oligomers of secondary structure elements. Few types of elements and arrangements that are common for many proteins exist. The arrangements are sequential, that is one polypeptide sequence encodes for all elements within one arrangement. This is in marked contrast to non-covalent oligomers which can propagate further with the formation of self-assembled nano-to-microscopic architectures. Similar elements and arrangements can

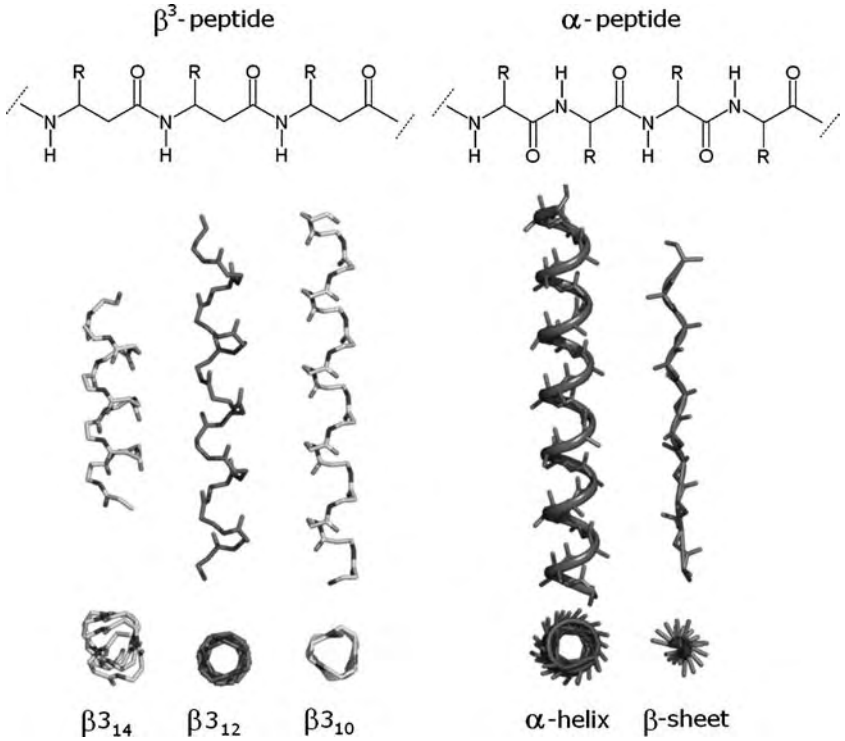


Fig. 5 Peptide foldamers. (Top) chemical structures of polypeptide backbones composed of β and α amino acids, β- and α- peptides, respectively. The β-peptide has the R groups attached to the β-carbon. Examples of β-peptide helices (left) in comparison with conventional α-helix and β-sheet (right). Side chains and hydrogens are not shown for clarity (reproduced with permission from S. J. Shandler, M. V. Shapovalov, R. L. Dunbrack and W. F. DeGrado. Development of a rotamer library for use in beta-peptide foldamer computational design, *J. Am. Chem. Soc.*, **132**, 7312. Copyright (2010) American Chemical Society).[14]

be found in different proteins and can elicit different functions. However, all are recognised by conserved patterns that relate sequential and spatial organisations of secondary structure elements. This forms the basis for the topological descriptions of proteins, which help to classify proteins and provide a design framework for de novo sequences. Therefore, in simplified terms it can be said that secondary elements constitute one type of folding motifs that directly derive from primary structure, whereas supersecondary arrangements composed of secondary motifs give another type.

Secondary motifs. Secondary motifs are typically confined to four main categories – helix, sheet, turn and loop with inclusion of some more specific structures such as polyproline helices (Fig. 6).

Regardless of a particular category all these are subject to the hydrophobic effect which drives the formation of hydrophobic interfaces by burying hydrophobic residues in the interiors. The highly polar peptide bond is disfavoured in a hydrophobic core which requires localised structuring within the peptide backbone. Because the peptide bond is regularly repeated this leads to ordering as well, which is realised through hydrogen bonds neutralising the backbone polar groups. Electrostatic interactions

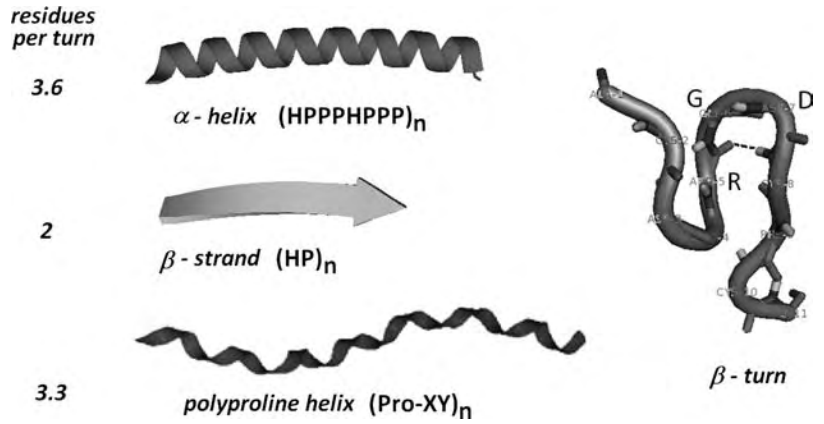


Fig. 6 Secondary motifs and their basic sequence patterns. α-helix, β-strand, polyproline (collagen) helix and an RGD-based turn, (2ZTA, 1ICO, 1CGD and 1FUL and 2LKW PDB entries rendered with PyMol).

and van der Waals forces are also important in this regard, and particularly so in furnishing contacts with side chain polar groups. Secondary structures thus can be referred to as local structures as opposed to global structures of fully folded proteins. However, it is the amino-acid sequence that specifies both the secondary structure and the type of hydrogen bonding. For this reason, to design secondary structures the main focus is made on (i) sequence patterns that can ensure geometries characteristic of a particular element (α-helix, β-strand) or (ii) on a specific hydrogen bond (turn).

For example, canonical α-helix and β-strand have 3.6 and 2 residues per turn, respectively, which are elemental repeats in contiguous segments stabilised by backbone hydrogen bonding. In helices this bonding is *intra*-backbone and is maintained in each *i, i + 4* amino-acid pair. This allows a helix to extend indefinetely in length from under ten amino acids to hundreds. In contrast, a β-strand having fewer residues per turn is of a more extended conformation which cannot be stabilised by hydrogen bonds within its backbone and requires *inter*-backbone hydrogen bonding. β-strands therefore tend to extend laterally by marrying other β-strands into β-sheets, and need not to be longer than ten residues. A real challenge is to design an individual β-strand and a homogeneous α-helix oligomer.

In designing turn-like structures, one can make assumptions of conformational preferences based on the sequence length between the two end residues of a given turn. All turns are very short motifs consisting of two (γ), three (β) or four (α) peptide bonds. However, for reasons outlined in 2.2 free peptides forming stable turn conformations need to be constrained. Therefore, it is somewhat more reliable to stabilise a particular *intra*-turn hydrogen bond in a particular turn arrangement.[17] Here the challenge is not only to design a stable structure, but also to differentiate between single or simple turns, which can also be inverse, and multiple turns within the same backbone.[6,11] The importance of turn structures cannot be stressed upon too highly. Being reversals of polypeptide chains turns are the motifs that determine the preference of proteins for globularity over linearity. Perhaps for this reason and to organise similar secondary motifs in different globular

shapes turns can be subdivided into several types. For example, there are at least nine distinct β-turn types.[17]

Turns prove to be important in the design of other secondary motifs different from but related to canonical α-helices and β-strands. Typical examples are hairpins which can be viewed as turns with extended ends. Hairpins can exist in both forms – autonomous and oligomerising, and as it is challenging to design one particular form for other motifs it is as challenging for hairpins.

Encouraginaly enough, by following very basic sequence principles of secondary motifs it is possible to prescribe a particular structure. Sequences with alternating hydrophobic *(H)* and polar *(P)* residues form β-strands, $(HP)_n$. The same patterns can afford turns with the inclusion of glycine or proline residues at turn points as these residues are known efficient breakers of extended secondary structures with the exception of polyproline helices which as the name suggests are built of repeating proline patterns as, for example, glycine-proline-proline triads in collagen triple helices. Incorporating hydrophobic residues at $i+4$ positions creates helix-promoting patterns $(HPPPHPPP)_n$ which are further stabilised with the inclusion of polar or small (alanine) residues between the hydrophobic residues (Fig. 6).

Thus, once a pattern is chosen it can be filled with specific residues which can also be specified according to their conformational preferences.[18] For example, glutamines, lysines and leucines favour helices, whereas tryptophans, threonines and valines strongly prefer β-structures.

These basic rules underpin first design principles underlying protein structure-function relationships. As it stands, the set of these principles is far from completion to enable prediction approaches and algorithms which would make de novo design a routine operation.[19] Establishing peptide sequence-structure links is crucial at secondary structure levels, the niche that can be fulfilled only by one's ability to create de novo secondary structures.

Super-secondary motifs. In the description and design rules of super-secondary motifs the main emphasis is made on the ways secondary motifs are arranged with respect to each other. A much lesser stress is made upon how the motifs are linked together. Because the linkages are represented by turn and loop structures their specific roles and impact are completely omitted in the topological descriptions of proteins. A main assumption in design terms is therefore on the existence of given super-secondary structures exemplified by native proteins, which can also serve as justified reasoning for generalised comparisons. In this respect, it is important to highlight the main types of super-secondary motifs, their topologies and principles of their description.

Protein topology is used as one of the main properties for the classification and comparison of protein structures. A number of databases for structural classifications are available, with each using topology principles. Amongst the most comprehensive and regularly updated are SCOP,[20] CATH,[21] FSSP[22] and specifically dedicated to protein topology PTGL[23] and TOPS++FATCAT.[24] Conceived as theoretical procedures for

single-type structures (α-helices or β-sheets) protein topologies found use in reconstructing folding pathways from predicted sequences of unfolding events. Starting from fully extended conformations of defined sequences *ab initio* algorithms are being attempted to predict fragment topologies including secondary and super-secondary structures.[25,26]

One early example is LINUS.[27] This is a simulation method that is solely based on amino-acid sequence inputs and is a computational implementation of hierarchic models supporting most designs to date. LINUS operates at short and medium range interactions omitting disufide bonds and prosthetic groups. A simulation run starts with an amino acid sequence, S, where $S = s_1 - s_2 - s_3 - \ldots s_N$, with s being amino acid residues with N residues in the sequence. The algorithm follows current and trial conformations progressing one residue at a time sequentially generating trial conformations which are rejected if any two atoms in the entire sequence overlap. This allows the algorithm to reveal a favorable structure in fixed intervals of allowed interactions followed by repeating the process in stages as the interval size increases. Recognised favourable conformations are constrained to persist at later stages thereby affording hierarchic build-ups. Helix, sheet and turns are represented with only three types of interaction used; repulsive to measure steric overlaps and attractive to measure hydrogen bonding and the hydrophobic clustering of apolar residues.

When applied to crystal protein structures the algorithm gives promising results for matching secondary and super-secondary motifs in protein fragments (Fig. 7).

Although the prediction outputs of super-secondary motifs remain to be moderate, topology descriptions have become a routine tool providing fairly accurate representations of secondary and super-secondary motifs from a biological point of view. In principle, any protein fold can be visulalised using topology diagrams, which are essentially two-dimensional folding graphs. In these graphs cylinders and flat arrows denote helices and β-strands, respectively. Loop-like lines are not structure specific and are used to illustrate sequentially adjacent secondary motifs. This way, all motifs can be shown and similar motifs in different proteins can be distinguished.

A number of well recognised motifs reflecting upon specific folding patterns exist and can be categorised into three basic groups – those that comprise only α motifs (*all α motifs*) or only β motifs (*all β motifs*) and those that contain both types (*α/β and α + β motifs*). The number and nature of motifs within and between these groups vary and depend on the nature of comprising units – helices and β-strands. It is logical to see that all β motifs is a relatively more diverse group than the others. Partly, this is because a stable single β-strand cannot exist and can readily associate with other strands as opposed to helices that are individual structures stabilised by interactions along their backbones; partly, because principles for β-sheet folding outside the protein context are less understood and as a result autonomous β-sheets are more difficult to design. This in turn requires more careful recognition between all-β motifs leading to the seemingly more diverse group of unique super-secondary structures. All the other groups can be described using the same topology and classification terms with

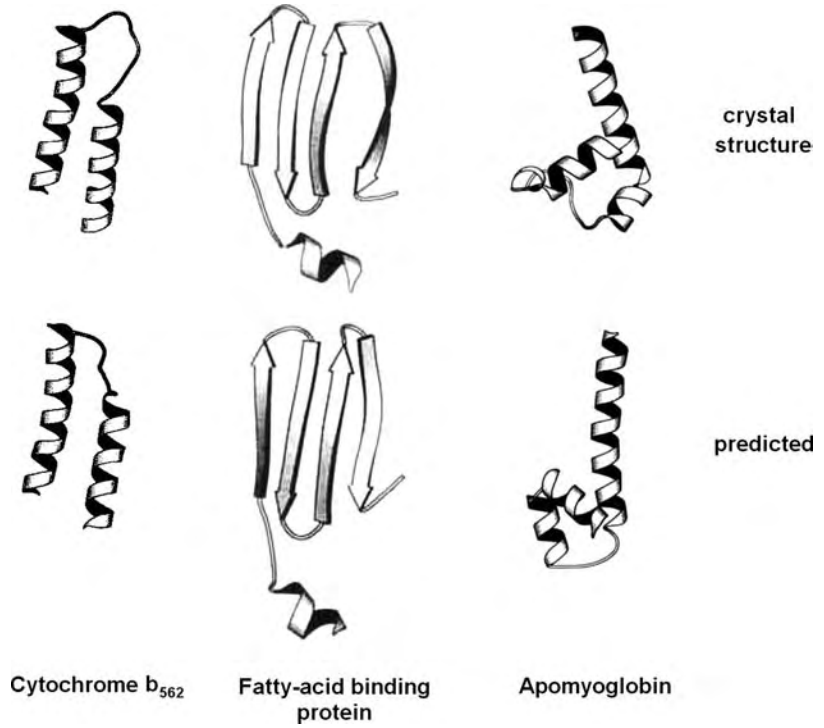


Fig. 7 Molscript drawings of crystal structures versus structures predicted by LINUS for sequence (26-75) fragments of cytochrome b_{562}, intestinal fatty-acid binding protein and apomyoglobin, (reproduced with permission from R. Srinivasan and G. D. Rose, LINUS: A hierarchic procedure to predict the fold of a protein, *Proteins*, 1995, **22**, 81. Copyright Wiley-VCH Verlag GmbH and Co. KGaA. Reproduced with permission).[27]

additional descriptions resulting from unique features of particular secondary motifs. Without the understanding of such differences de novo design is impracticle. Therefore, representative motifs for each group starting with all-β motifs are discussed in the following sections of structural rationales and de novo design which highlight current state in relating secondary motifs to polypeptide topologies and to main trends in de novo design.

4 Current trends: structural rationales

All β motifs

Individual β-strands can be viewed as extended helix-like structures stabilised in super-secondary associations. Because these are sequence motifs, one sequence can in principle be designed to fold autonomously into several strands arranged in a unique way. For example, in one motif three consecutive anti-parallel β-strands connected *via* two short loops are arranged into two hairpins. The hairpins are linked with a fourth strand *via* a longer loop such that the N-terminal and C-terminal strands are placed next to each other. This so-called *Greek key* motif is said to have a 4123 or 1432 topology depending on the position of the N-terminal strand (Fig. 8).

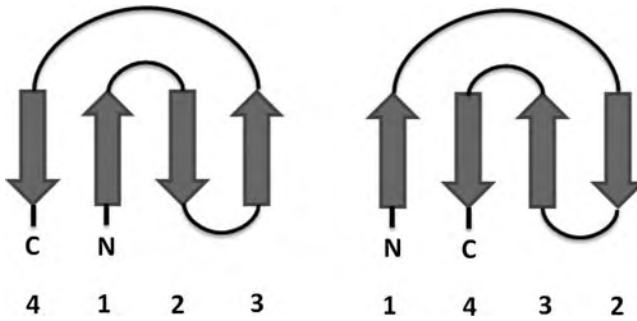


Fig. 8 Topologies of the Greek key motif.

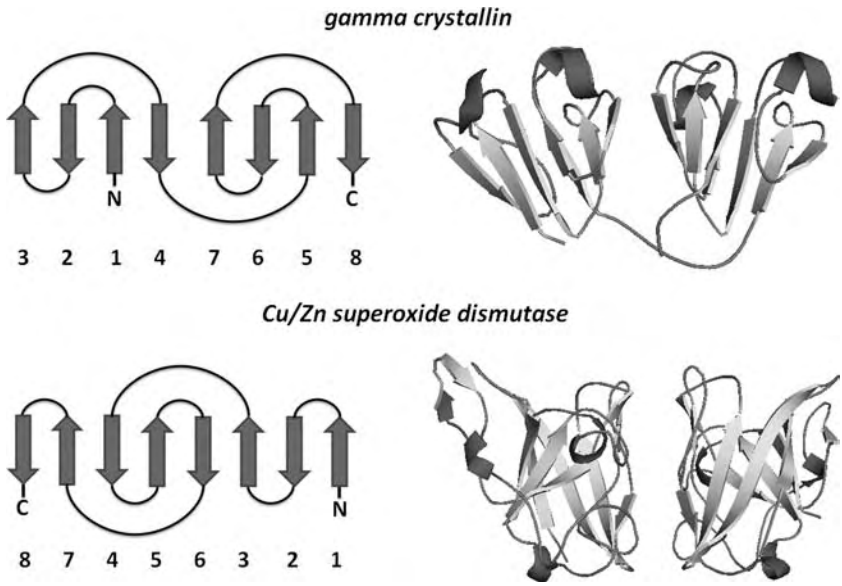


Fig. 9 Examples of Greek key containing proteins. Topologies and structures of gamma crystallin and Cu/Zn superoxide dismutase (1AMM and 2APS PDB entries rendered with PyMol).

This motif is very common in proteins and can be a part of larger and more extented arrangments.[28] Figure 8 shows two different topologies incorporatinng the Greek key motif from two different proteins. Using such an up-down topology one can desribe a backbone ensemble of β-strands as to the number of strands, their order and their orientation with respect to one another, parallel or anti-parallel. Strand alignments denote also intra-strand patterns of hydrogen bonding. For example, strands 4 and 3 for 4123 (or 1 and 2 for 1432) in the Greek key motif are edge strands that are laterally connected *via* hydrogen bonds with strands 1 and 2 (or 4 and 3). Thus, a network of hydrogen bonds spans through all strands between the edges forming a sheet.

The Greek motif might be most common in protein architectures (Fig. 9). However, other motifs of similar complexity are also abundant. Examples include β-meander, in which, unlike in the Greek key, β-strands run

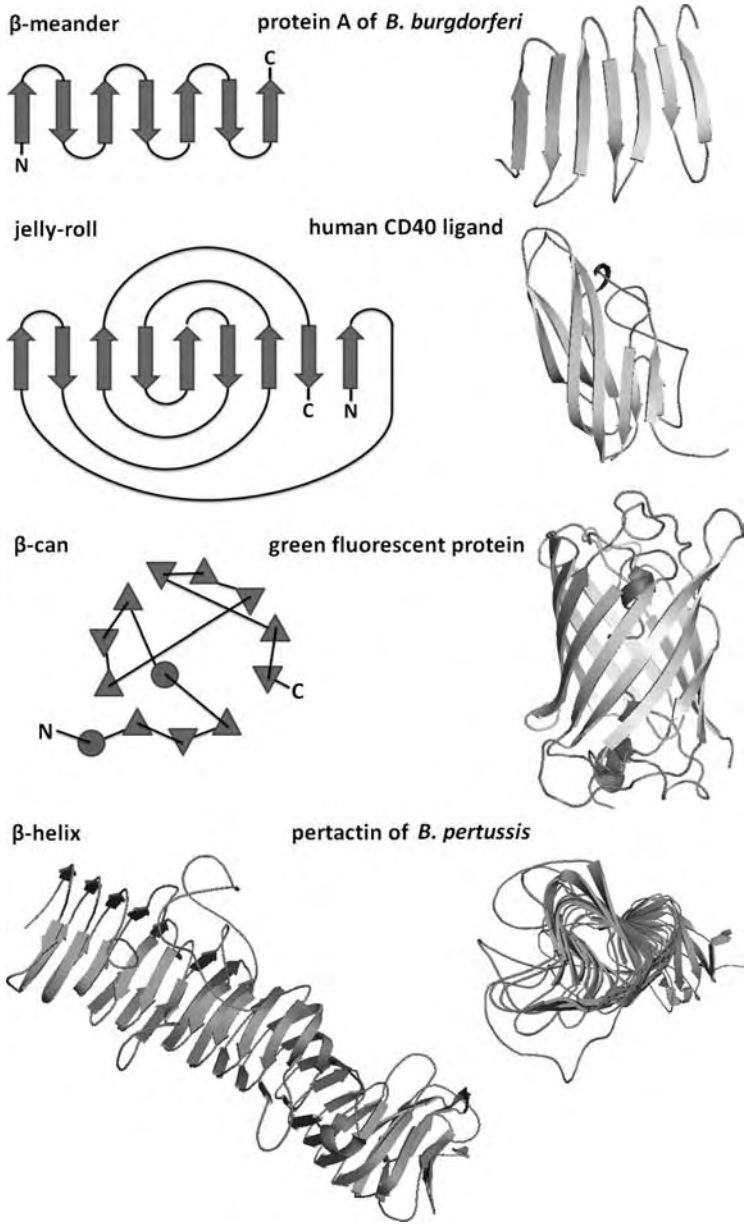


Fig. 10 Structures and simplified topologies of all-β motifs. The topology diagram of β-can shows α-helices and β-strands as circles and triangles oriented perpendicular to the plane of the paper, respectively. Structures are from 1OSP, 1ALY, 1EMA and 1DAB PDB entries rendered with PyMol.

sequentially in the backbone and in the fold; β-corners, which are β-strands bent in the middle over a β-bulge to give hairpin-like anti-parallel arrangements; or ''jelly-roll'' folds, which are usually described as Greek key motifs with an additional twist (Fig. 10).[29–35] All these structures tend to be anti-parallel arrangements of β-strands into sheets. In some cases, a formed β-sheet can assemble by interacting with the edges of other sheets

leading thus to extended sheets. The edges of one sheet can also close up with forming β-barrels as, for example, in green fluorescent protein whose fold was dubbed β-can to reflect upon the barrel-like arrangement of β-strands around an α-helix (Fig. 10).[36,37] β-sheets can also pack face to face organising into circular structures reminiscent of the blades of a propeller.[38]

Somewhat different from all these structures are relatively recently described β-helices.[39] These are spiral or helix-like arrangements of parallel β-strands, which are also *inter*-strand associations whose main characteristic is a repeating unit of three strands forming a ring structure (Fig. 10). β-helices have been found in several proteins predominantly of microbial origin. From the design point of view, the structures are fascinating due to their unique regularity and repetitive sequence motifs, spatial extension and precise biological functions. Because β-helices can span several nanometers in length as one monodisperse construct and can associate laterally or in an end-to-end fashion they are of certain interest as building blocks for higher order oligomers or as templates for hybrid protein assemblies.[40]

All α motifs

Helical motifs can be topologically similar to all-β motifs. Helix hairpins or helix-hairpin-helix (HhH), helix-turn-helix (HtH) and helix-loop-helix (HlH) structures are typical examples of sequence all-α motifs (Fig. 10).[41–45] These motifs play key roles in different biological processes serving as important functional elements as, for example, in DNA-bindng proteins including those with enzymatic activities (*e.g.* DNA polymerases, transcription factors) and structural functions (*e.g.* ribosomal proteins).[41–45] As their names imply each of these motifs is a pair of anti-parallel helices connected *via* a linker whose exact nature determines a paticular motif type. The length of a linker in each case varies and directly effects conformational preferences. Indeed, in HhH motifs hairpin linkers can have two, three and four residues. In two-residue hairpins only one conformation is allowed, whereas four-residue hairpins offer a choice between two different conformations. A consensus *GHG* repeat (where *H* is a hydrophobic amino acid) that is characteristic of HhH motifs packs helices at acute angles, which is in contrast to HtH motifs that have α-helices arranged at a right angle (Fig. 11).[41,46] Further, linkers in HlH motifs are longer than in HtH and can span several residues, hence loops. This allows HlH motifs to accommodate additional functions in the loops as, for example, in EF hands – a structural domain found in calcium-binding proteins.[47] The loops in EF hands arrange into a specific bipyramidal configuration to coordinate Ca^{2+} ions. Typically it takes 12 residues to chelate one calcium ion and to form a stable hydrophobic core, for which both anionic and hydrophobic amino acids are used.[48] By tuning the coordination pattern of the loop, namely XxYxZx-Yx-Xxx-Z, where the upper case denotes ligands involved in metal coordination and the lower case are other residues, it is possible to engineer environments capable of competitive binding of other metal ions including lanthanides, magnesium and zinc.[49]

Unlike of β-hairpins and more generally β-sheets, the classification of helical hairpin-like structures is less straightforward. This can be attributed

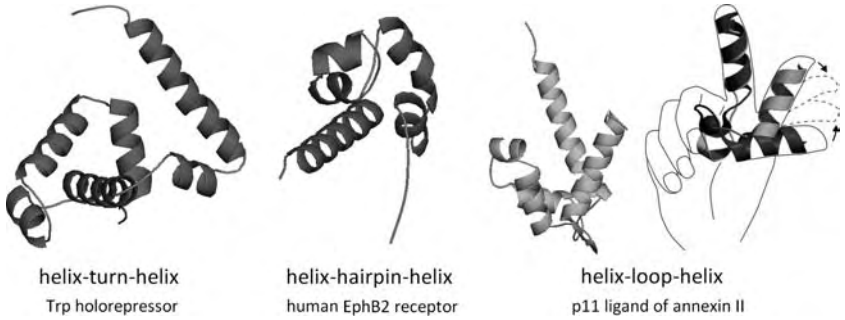


Fig. 11 All α sequence motifs. (From left to right) structures of *E.coli* Trp holorepressor (HtH), SAM domain of human EphB2 receptor (HhH) and p11 (S100A10) ligand of annexin II (HlH) (2OZ9,1B4F and 1A4P PDB entries rendered with PyMol). (Right) a symbolic schematic of the EF-hand motif, in which one helix (E) occupies the index finger, whereas another helix (F) winds up the thumb. Upon binding Ca^{2+} helix F undergoes a transition from the close (apoprotein, light grey) to the open (holoprotein, dark grey) conformation (reprinted from A. Lewit-Bentley and S. Rety, EF-hand calcium-binding proteins, *Curr. Opin. Struct. Biol.*, **10**, 637, Copyright (2000), with permission from Elsevier)[48]).

to that proteins containing these motifs do not tend to share a common fold or a common minimal motif and that the geometry of linkers between helices is less defined. For example, HhH motifs are separate, and not necessarily periodic, sub-units of larger domains, whilst helical hairpins are intrinsically less constrained than β-hairpins as their formation and stabilisation are not reliant on inter-strand interactions which provides them with a greater choice for different strand registries. Nevertheless, there appear to exist different turn structures in helical hairpin motifs which prefer well defined geometric and conformational parameters. In design terms these take a form of helix-terminating or capping motifs which derive from the basic parameters of helix formation.[50–54]

Specifically, helix-capping motifs result from the so-called edge effects and are found at or near the ends of helices. The first four amide protons and last four carbonyl groups lack intrahelical hydrogen bonds which cannot be formed with other amide and carbonyl groups within the backbone. Instead, these terminal groups can be capped by alternative hydrogen bonds furnished by polar groups flanking the ends of the helix. Moreover, these termini positions exhibit certain amino acid preferences, *i.e.* capping motifs (Fig. 12).[54,55]

Thus, helix-ending residues are referred to as Ncap and Ccap and full capping-motifs are denoted as N''-N'-Ncap-N_1-N_2-N_3-$\cdots\cdots$-C_3-C_2-C_1-Ccap-C'-C''-..., where Ncap-N_3 and Ccap-C_3 are the terminal residues lacking hydrogen bond partners.[53–55] The capping motifs occur and are promoted by interhelical turns that are formed by the primed residues bracketing the helix at either end. These are the short linker regions in HtH motifs that join helices perpendicular to one another creating αα-corners.[50,51] The turns have their first residue adopting the left-handed α-helix conformation (α_L), which terminates the first right-handed helix in a specific hydrogen bonding pattern, and is followed by the next residue in a β-strand conformation. Taken together these give α_L-β interhelical turns (Fig. 12).[52–56]

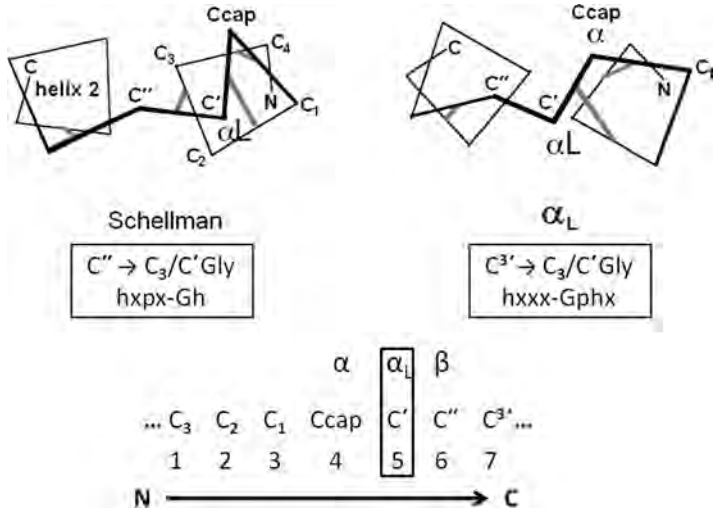


Fig. 12 Helix-capping in HtH motifs. Examples of two C-terminal capping motifs, Schellman and α_L motifs. Backbone chains are black contiguous lines. Hydrogen bonds are crossover lines joining corresponding residues. α_L, α, and β denote left-handed helix, right-handed helix and β-structure conformations. Numeric numbers indicate the sequence positions of the motif residues. Both popular names and systematic nomenclatures (in squares) for the motifs are given. h, x, and p stand for hydrophobic, any and polar. Hyphens mark the helix boundary, *i.e.* Ccap-C′ (the diagrams are reproduced from S. J. Lahr, D. E. Engel, S. E. Stayrook, O. Maglio, B. North, S. Geremia, A. Lombardi and W. F. DeGrado, Analysis and Design of Turns in α-Helical Hairpins, *J. Mol. Biol.*, **346**, 1441, Copyright (2005), with permission from Elsevier).[56]

The first type originally described by Schellman, Schellman motif,[53] relies on the formation of hydrogen bonds between residues at positions 1 and 6, and between residues at positions 2 and 5, with α_L occupying position 5. An α_L turn is a variation of this motif involving only helical hydrogen bonds up to the α_L residue, whereas in Shellman motifs additional hydrogen bonds are possible.[52,54] Both motifs are of ...α-α_L-β... conformations that are described as right-handed αα-corners. To form a left-handed corner it is necessary to turn helix 2 by 180° (Fig. 12), which is highly unfavourbale as this imposes strong steric hindrance.[50,56] As a consequence, left-handed αα-corners in solved protein structures are not common.

Typically, the C-terminal position of helix 1 (C′) should lack a side chain, which is readily accommodated by glycine.[54] The following position, the first residue of helix 2, (C″) is taken by a polar amino acid which has a strong β-propensity and furnishes hydrogen bonding within the helix. C_3 position is predominantly hydrophobic and is buried with other hydrophobic side chains at the interface of the contacting termini of the two helices. Thus, helix-capping motifs can be classified based on contributions from hydrogen bonding and hydrophobic interactions. Each capping motif can be named for the closest pair of interacting hydrophobic residues. For example, in the termination sequence C″→ C_3 the arrow points from the hydrophobic residue, which is external to the helix (C″) to the hydrophobic residue within the helix (C_3). In some cases, when a particular residue prefers a particular position one- or three-letter code of this residue is used preceded by the position and a slash, as in C″→ C_3/C′Gly (Fig. 12). Using

this nomenclature all capping motifs, both N- and C-terminal, can be named systematically.

αα-corners appear to be stable and can be found in both short peptides where they are extended with short irregular structures and in large globular proteins, in which they pack hierarchically by adopting aligned, orthogonal or slanted arrangements of helices.[57,58] Given the strict geometry of αα-corners the number of possible seeding structures for hierarchical growth is limited to a few. Minimal structures can be viewed as one-helix extensions of a corner (Fig. 13).

The folding and geometry of HtH, HhH and HlH motifs depend on and are determined by their loop structures. Therefore, the design of these motifs often comes down to choosing the right turn.[56] Understandably, for native proteins this is imposed by the requirement of keeping helices apart thus avoiding lengthwise contacts, which is necessary for function. Consequently, sequence preferences in these motifs favour independently folded and evenly polar helices. This is in contrast to another class of all α motifs which strongly depend on *inter*-helical contacts. These are helix bundles – defined sequence and assembly oligomers resulting from amphiphilic helical folding.[59–61] In these motifs individual helices arrange with the formation of one hydrophobic interface by folding into amphipathic structures having two distinctive seams of hydrophobic amino acids, those that get buried in the interface facing each other, and polar amino acids that are exposed to their aqueous environment.

Helix bundles are rationalised by the number of helices and the number of sub-oligomers in a given bundle giving rise to different motifs with topologies similar to those observed for all β structures. The lowest oligomers, two-helix bundles, tend to be unstable and dimerise into four-helix bundles, whereas three-, four and five helix bundles are common and very stable (Fig. 14).[59–61]

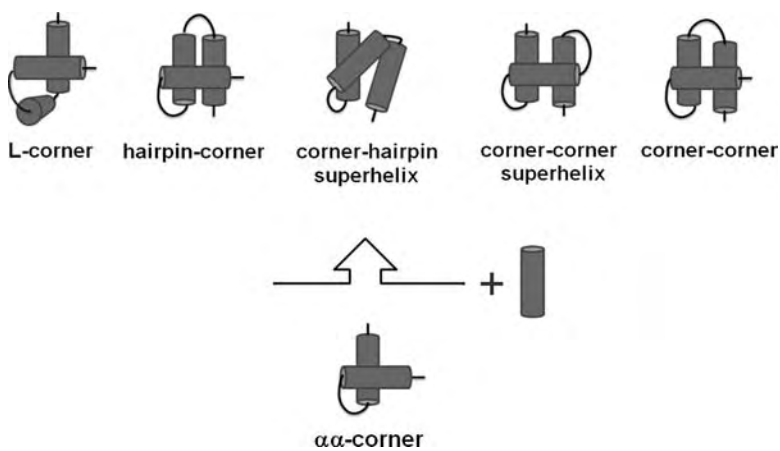


Fig. 13 αα-corner-derived motifs. Schematic representations of minimal motifs as one-helix extensions of a αα-corner. Both directions of the polypeptide chain are possible for each structure (only one direction is shown) (adapted from A. V. Efimov, A structural tree for α-helical proteins containing α-α-corners and its application to protein classification, *FEBS Lett.*, **391**, 167, Copyright (1996), with permission from Elsevier).[57]

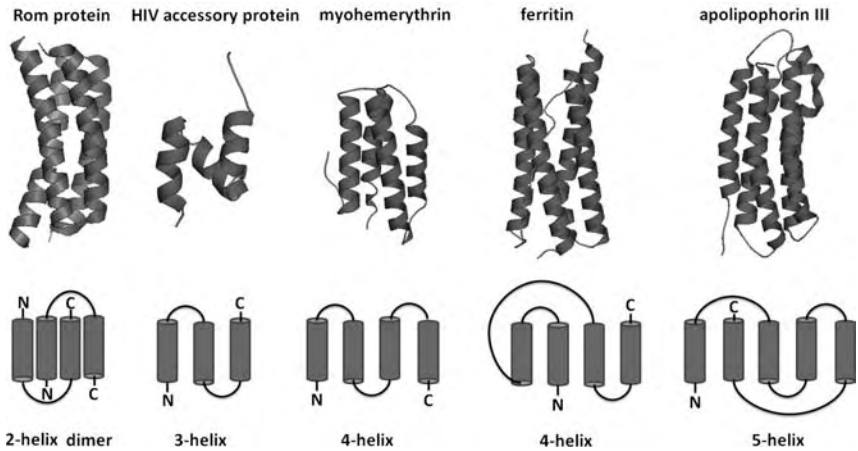


Fig. 14 All-α sequence motifs. Up-and-down topologies and structures of the RNA one modulator (Rom) protein, a headpiece of the HIV accessory protein, myohemerythrin, ferritin and apolipophorin III (2IJK, 1F4I, 2MHR, 1KRQ and 1EQ1 PDB entries rendered with PyMol).

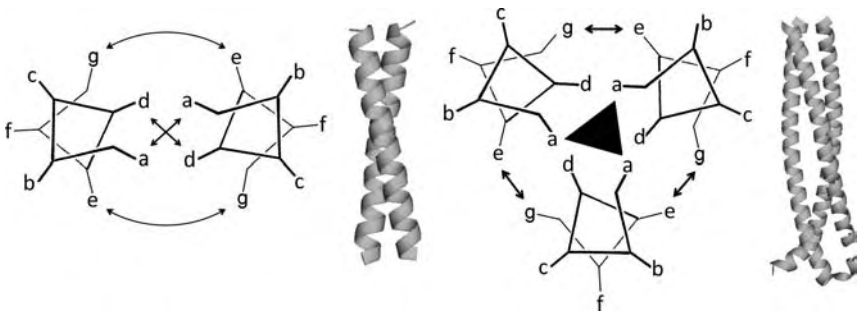


Fig. 15 Coiled coils. Helical-wheel representations and structures of leucine zipper domains: *dimeric* transcriptional activator GCN4 and *trimeric* cyclic nucleotide-gated channel (2ZTA and 3SWF PDB entries rendered with PyMol). *abcdefg* repeats are configured into helical wheels with 3.5 residues per turn. Electrostatic *e-g* interactions and hydrophobic *a-a* and *d-d* pairs are indicated by double-headed arrows. The black triangle shows the segregation of hydrophobic side chains in *a* positions – *a* layer (segregation in *d* positions is omitted for clarity).

Helix bundles share many similarities with other folding motifs – coiled coils.[62–64] In the oligomerisation context coiled coils can be defined as bundles of unconnected helices as opposed to helix bundles that are typically single-stranded structures. The independence of coiled coils on loops and turns renders them very diverse and readily amenable to design (Fig. 15).[65–69]

However, such independence requires another means of constraining helices into specific bundle formations. A unifying determinant to ensure this for all coiled coils is the interdigitation of side chains of adjacent helices in the interface of a bundle. Thus, it is the hydrophobic residues along each polypeptide chain that predominantly determine a particular coiled-coil architecture. Usually referred to as ''knobs-into-holes'' packing such an arrangement can be supported by different patterns, canonical and non-canonical (Fig. 16).[68,70]

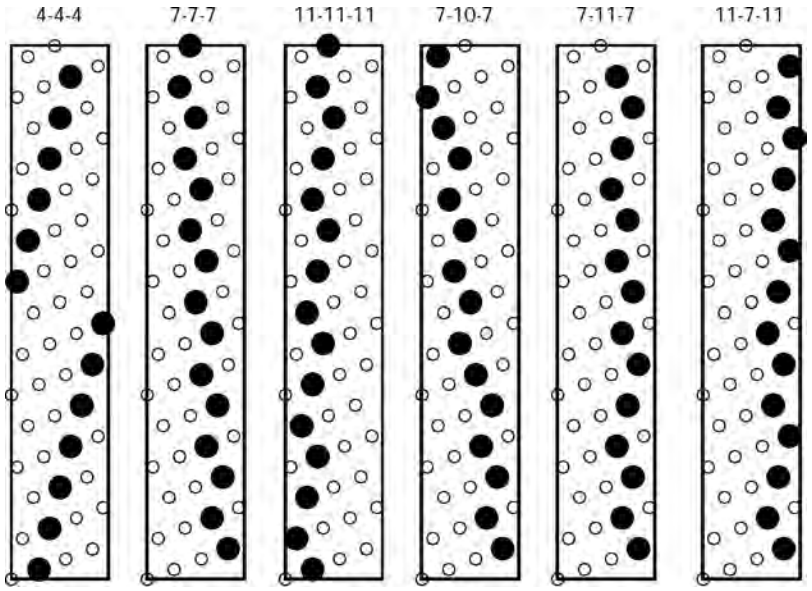


Fig. 16 Coiled-coil patterns. Helical net diagrams for $i+4$ (4) patterns for helices in globular proteins (4-4-4) and for 3,4 patterns for coiled coils (all the others). The diagrams are opened out cylinders wrapped around an α-helix showing C_α atoms (open circles) and hydrophobic residues (black circles) projected onto the outside of the cylinder. 4, 7, 10 and 11 denote tetrads, heptads, decads and undecads respectively (Reprinted from M. R. Hicks, D. V. Holberton, C. Kowalczyk and D. N. Woolfson, Coiled-coil assembly by peptides with non-heptad sequence motifs, *Fold Des*., **2**, 149, Copyright (1997), with permission from Elsevier).[70]

Canonical patterns dominate and are built from heptad repeats in which hydrophobic side chains alternate three and four residues apart, $(HPPHPPP)_n$.[62,64,68] Normally denoted *abcdefg*, with *a* and *d* shaping the packing and being hydrophobic, this 3,4 hydrophobic pattern is in fact a prerequisite for the coiled-coil packing which distinguishes it from another type observed in globular proteins known as the ''ridges-into-grooves'' packing.[71] Indeed, non-canonical patterns equally incorporate and maintain the 3,4 pattern as, for instance, in decads and undecads that show 3,4,3 (*abcdefghij*) and 3,4,4 (*abcdefghijk*) combinations of hydrophobic repeats, respectively.[68,70] Similarly, the alignment of 3,4 patterns in partnering helices is crucial for correct (in-register) interactions. The rationale for the 3,4 patterns is, on the one hand, to create a contiguous hydrophobic seam on each helix, which would guarantee stable and discrete super-helix associations, and, on the other hand, to maximise burial of the seams *via* their precise matching in bundles – knobs-into-holes packing. Because the average spacing of hydrophobic residues along the sequence (3.5 residues) fall short of one complete turn of an α-helix (3.6 residues) (Fig. 6), the seam adopts a left-handed twist with respect to the right-handed helix, which allows the association of helices in the bundle with left-handed helix-crossing angles (Fig. 16).[62,68] The exact number of helices and their type (hetero- or homotypic) in the bundle are defined by amino acid preferences of hydrophobic and electrostatic pairs (Fig. 15).[64,69] For example, a classical leucine zipper motif has leucines in *d* sites, which direct the assembly of parallel dimers.

Higher oligomers for helix bundels and coiled coils are also possible in native systems. Predominantly, these are complex structures of lower oligomers. Six-helix bundles present a notable example.[72] This type is preferred by enveloped viruses such as retroviruses and paramyxoviruses to catalyze their fusion with host cells.[73] For instance, gp41, which mediates fusion of immunodeficiency viruses, upon binding a cellular membrane undergoes a cascade of conformational transitions leading to its three monomeric units pinned by a coiled-coil formation, which in turn promotes the assembly of a six-helix bundle.[74] Motifs along this pathway are of particular interest for therapeutic intervention by means of blocking tertiary contacts in the bundle, namely in and between membrane proximal regions. Promising fusion inhibitors have been reported for this purpose including a gp41 fragment T20, also known as fuzeon and enfuvirtide, which is now used as an anti-HIV drug.[75]

Stand-alone bundles are limited to pentamers and some globular proteins are known to contain hexamers.[76,77] Although there is still scarce information on these and larger oligomers the known bundles provide inspiration for new designs, with the emergence of engineered pentamers,[78] hexamers[79] and heptamers[80] (Fig. 17).

The outlined all-α helix motifs may not represent all possibilities that exist in different protein folds, however, they should provide a sufficient basis as fundamental structures for a parameterised description of protein folding and are equally instrumental for de novo helical designs. Further along these lines and perhaps of more importance are exploitable relationships between the basic secondary structures in native proteins. These are expressed in α/β motifs, those motifs that incorporate both α-helices and β-strands together with associated loop and turn structures.

α-β motifs

Motifs containing both major types of secondary structure can be divided into different categories reflecting upon two basic arrangements – α + β and α/β. The main difference between the two types is that in α + β proteins α-helices and β-strands are separated along the backbone, whereas in α/β they alternate. In α + β folds helices cluster and are often wrapped by a β-sheet, whereas in α/β folds helices organise into exteriors of the fold. β-strands can adopt parallel, anti-parallel or mixed orientations in α + β proteins, but are parallel in α/β folds.

Thus, α + β motifs are characteristic of complex non-wound α-β topologies which are difficult to interpret and assign rationally. Many proteins impose further complicated variations of α + β arrangements and often are singled out as unique mixed folds (Fig. 18). This represents a qualitatively extreme level for de novo design even considering relatively small and simple folds like ubiquitin, which puts de novo α + β structures in the far end of the spectrum of emerging designs.

Typical α/β motifs are repeating β-α-β units, in which β-strands are kept in parallel by hydrogen bonding and form an inner or central layer, a β-sheet. α-helices pack against this layer. Because these α/β motifs are sequence motifs (i) helices pack anti-parallel to the β-sheets, which in turn requires (ii) relatively long loops to allow the cross-over of the helices on the

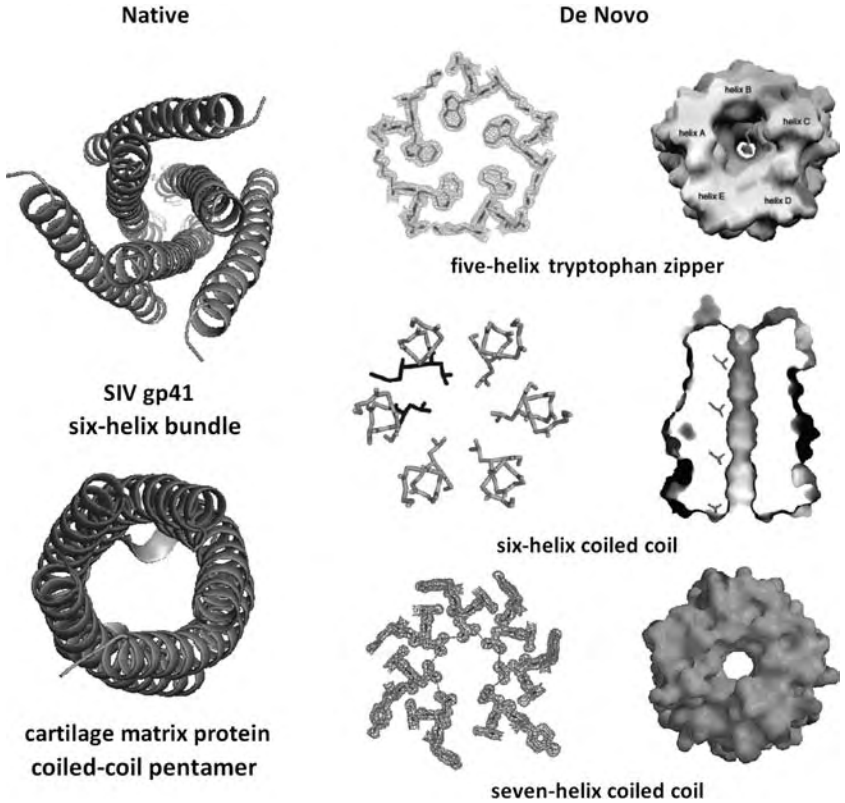


Fig. 17 Structures of complex and stand-alone helical oligomers. (Left column) SIV gp41 ectodomain and the assembly domain of cartilage oligomeric matrix protein representing a complex six-helix bundle and a stand-alone coiled-coil pentamer, respectively (2EZS and 1VDF PDB entries rendered with PyMol). (Right column) Designed coiled-coil oligomers. Electron density maps showing cross-sections in *a* layers and (far right) molecular surface representations of the assembled coiled coils showing central channels. (reproduced with permission from J. Liu, W. Yong, Y. Deng, N. R. Kallenbach and M. Lu. Atomic structure of a tryptophan-zipper pentamer, *Proc. Natl. Acad. Sci. USA*, **101**, 16156. Copyright (2004) National Academy of Sciences, U.S.A.,[78] reprinted by permission from Macmillan Publishers Ltd: (N. R. Zaccai, B. Chi, A. R. Thomson, A. L. Boyle, G. J. Bartlett, M. Bruning, N. Linden, R. B. Sessions, P. J. Booth, R. L. Brady and D. N. Woolfson, A de novo peptide hexamer with a mutable channel, *Nat. Chem. Biol.*, **7**, 935), copyright (2011),[79] and reproduced with permission from J. Liu, Q. Zheng, Y. Deng, C.-S. Cheng, N. R. Kallenbach and M. Lu, A seven-helix coiled coil, *Proc. Natl. Acad. Sci. USA*, **103**, 15457. Copyright (2006) National Academy of Sciences, U.S.A.[80]).

outer layers (Fig. 18). A rationale for such aligned packing is that helical $i+4$ residues create surfaces that interface complementarily with those formed by β-strands. The interface is the result of the right-handed twists adopted by each side, α and β. This is the complementarity of the twists which is the hallmark of all α/β motifs.

A classical subunit of α/β motif is often called the Rossmann fold.[81] This is a linear β-α-β-α-β arrangement. In those cases when two terminal β-strands close up *via* side-side interactions parallel barrel-like structures can be formed resulting in α/β barrel folds.[82,83] α/β barrels or, as they are also called, TIM (triosephosphate isomerase) barrels are uniformly wound structures, *i.e.* the direction of their twisted core β-sheets is constant. An

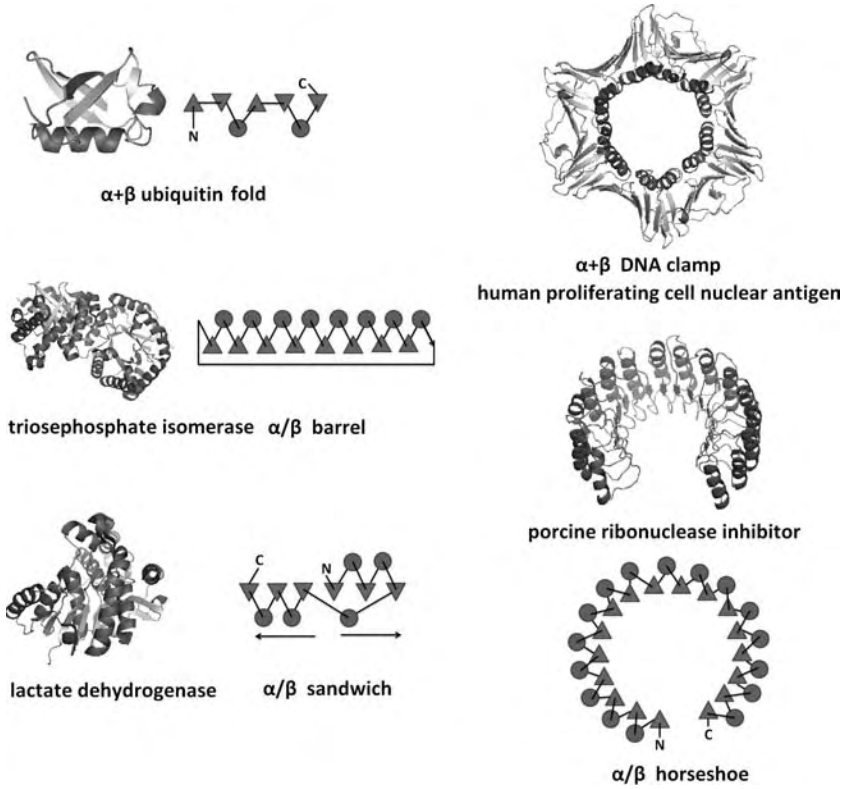


Fig. 18 α-β motifs. Structures of ubiquitin, triosephosphate isomerase, lactate dehydrogenase of *P. falciparum,* human proliferating nuclear antigen and porcine ribonuclease inhibitor are from 1UBQ, 1TIM, 1T25, 1AXC and 2BNH PDB entries rendered with PyMol. Simplified topology diagrams show α-helices and β-strands as circles and triangles oriented perpendicular to the plane of the paper, respectively. Arrows denote the direction of the polypeptide chain.

alternative arrangement occurs when a barrel opens up which results in a doubly wound α/β sandwich where helices form two layers following two different directions (Fig. 18).

α/β arrangements are stable and most frequent as domain types in enzymes and nucleotide-binding proteins that require tight complexes to encapsulate the active site of an enzyme or to furnish a specific nucleotide-recognition stretch. Proteins also employ Rossmann-like folds to build structures as extensive as, for example, an α/β horseshoe found in ribonuclease inhibitors (Fig. 18). This is an open β-sheet fold formed from seventeen parallel strands with sixteen helices forming an outer layer. Such an unusural shape affords the formation of a versatile structural framework with a strong hydrophobic core between the α and β layers, which is predominaltly occupied by leucine residues giving rise to leucine-rich repeats.[84,85] Other leucine-rich-repeat proteins adopt similar folds and structurally are variations in relation to the number of strands and helices. Differences in the length and number of repeats define the curvature of the leucine-rich-repeat domains to enable interactions with a rich variety of ligands.

Different folding modes for α-helices and β-strands and their complementarity in α-β motifs make the design of an α-β motif a formidable

endeavour. A design requires a multi-parameter approach and may become as complex as a multi-step assembly process favoured by specific conditions – the situation that is mirrored by the extent to which one lacks empirical and prediction models. Therefore, small and simpler protein folds of two or three secondary blocks present more attractive models. These include zinc fingers and conformational switches that remain amongst the most popular examples.[86,87]

5 Current trends: de novo motifs

De novo design of autonomously folded (secondary and super-secondary) motifs is of constant interest. However, its pace has yet to match that with which new solved protein structures appear. Indeed, given the state of the art in synthetic, computational and analytical capabilities which now allow the generation and probing of new sequences within weeks or months, in record time when compared to that a few decades ago, one should expect a stronger impact of de novo design. Admittedly, the impact is uneven across different motif types: helical motifs are being relatively well exploited, whereas the design of autonomously folded β-structures continue to stumble over the lack of control over their stability and aggregation properties. β-turn designs may stand out by having benefited from having been a major focus of research over many years thanks to their intimate relevance to protein-protein interactions and hence drug design per se. This field shows truly substantial advancements and as outlined above aims nowadays at structural and conformational mimetics as new receptor antagonists and inhibitors. However, even such a clear focus on function does not guarantee a robust and standard design rationale which continues largely to rely on intuition and empirical approaches.

Autonomously folded structures may belong to early events of protein folding, but should not necessarily represent folding intermediates or lead to more complex topologies or a compact folded form. Although it may not be correct and is difficult to partition protein folding into separate stages autonomously folded designs are defined here as relatively stable secondary and super-secondary structures or, in other words, those structures whose folding transitions are limited to the formation of a target structure.

Apart from the fundamental problems of protein folding there remain certain limitations to such designs which require more detailed investigations and understanding of various factors that can have significant effects on overall and local polypeptide structure. Correlation between stabilising forces and increasing complexity and size of a given design or correlation between hydrophobicity and folding propensities and aggregation and solubility are all those important factors that are not trivial to predict.

All β motifs

A continuous legacy of β-hairpin designs spanning over several years is perhaps most appropriate to demonstrate this. Early designs that include short hairpins (8–16 residues) revealed several critical parameters ranging from the stability of strands and their length limits to the importance of turn regions and their chirality (Fig. 19).[88–93]

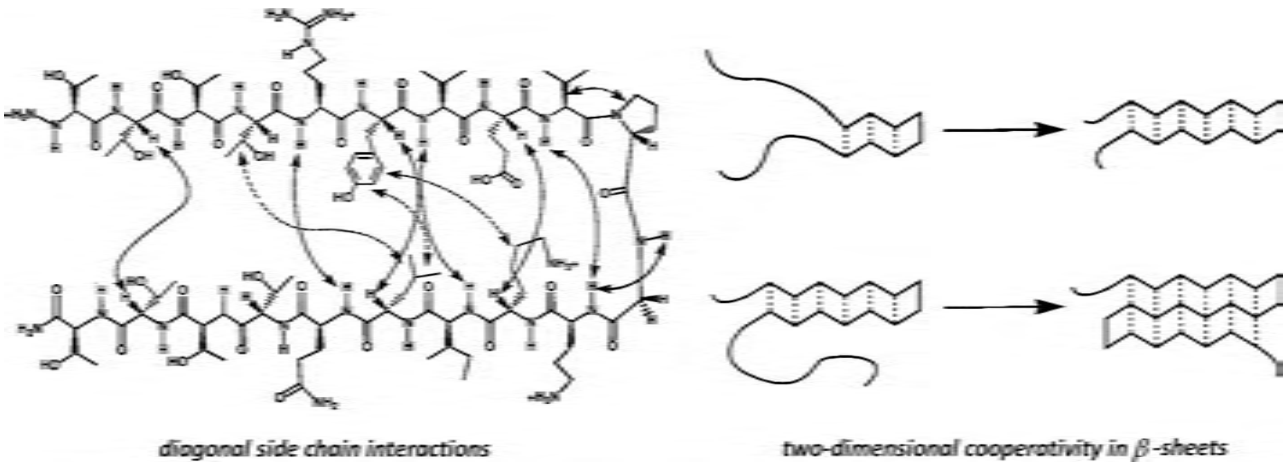


Fig. 19 Design-probed stability factors of anti-parallel β-sheet formation. (Left) NOEs between backbone protons (solid arrows) and side chain networks (dashed arrows) in a designed β-hairpin. (Right) length-dependent cooperativity along and perpendicular to the strand direction (H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Giriat, T. Muir and S. H. Gellman, Length-dependent stability and strand length limits in antiparallel β-sheet secondary structure, *Proc. Natl. Acad. Sci. USA*, **98**, 12015. Copyright (2001) National Academy of Sciences, U.S.A.[93]).

From the start β-hairpins have been used as minimalist structures to modulate β-sheet formation. Immediate questions that resulted from the first[88] and subsequent designs concerned with stabilising factors that support autonomous hairpin folding. Initial approaches therefore were based on using native hairpins as, for example, from tendamistat,[88] B1 domain of protein G,[89] ubiquitin[90] and ferredoxin.[91] Two important outcomes of these early designs were to experimentally show that monomeric hairpins in aqueous solution are possible and that hairpins apparent in the tertiary context of a fully folded protein are not necessarily stable in the isolated form. This raised questions of the rational structural control of designed β-hairpins. Particular strategies were sought to afford β-hairpins with two-residue loops at defined positions. Two-residue loops would be tight enough to provide the minimal number of non-strand residues thereby maximizing the proportion of β-structure in the design. One significant rationale came from an observation that such loops adopt specific type I′ and II′ β-turn conformations which are diastereomeric to types I and II respectively in native proteins.[92,94] This tendency was consistent with that both β-strands in native sheets and type I′ and II′ β-turns show right-handed twists. This is in contrast to type I and II turns which prefer a left-handed twist. Such an unusual correlation between the right-handedness of β-strands and turns in designed hairpins led to an efficient strategy for designing β-hairpins. It was shown that loops with one incorporated residue in D configuration adopt I′ and II′ β-turns, and that D-prolines (^{D}Pro) in ^{D}Pro-X two-residue loops strongly promote autonomous β-hairpin formation in solution as opposed to ^{L}Pro-X containing peptides that appeared to be disordered.[95–98]

These findings permitted other designs that proved to be instrumental to probe the effect of *inter*-strand side chain contacts on the conformational preferences of β-sheet formation. Using a series of model β-hairpins with ^{D}Pro-X loops, it was possible to demonstrate that the stability of β-sheets can be regulated by inter-strand juxtapositions of side chains.[92,93] More specifically, it was shown that ''lateral'' side-chain interactions, those that are adjacent with respect to the hydrogen bonding register, and ''diagonal'', those that are not, are stabilising for β-sheets.[93]

The importance of the results is three-fold. Firstly, they show that side chain interactions are defining for β-sheet stability. Secondly, they emphasise that two-residue ^{D}Pro-X loops despite their strong preference for hairpin folding cannot alone support high populations of β-hairpins. Thirdly, they highlight the critical importance of the ''two-dimensional'' stability of β-sheets when compared to other secondary motifs such as α-helices whose stability is determined only in one dimension – along the helix axis. Collectively, this leads to a conjecture of a length-dependent cooperativity in β-sheet folding which occur along and perpendicular to the strand direction (Fig. 19).

Other investigations along similar lines revealed strong dependencies of cooperative stability on lateral electrostatic side-chain interactions and on cross-strand hydrophobic and aromatic interactions.[99–101] In the latter case, cross-strand tryptophan-tryptophan pairs were shown to stabilise β-hairpins to the extent of reversible thermal transitions (midpoints at $\sim$ 50 °C) and remarkable thermodynamic properties showing folding free energies

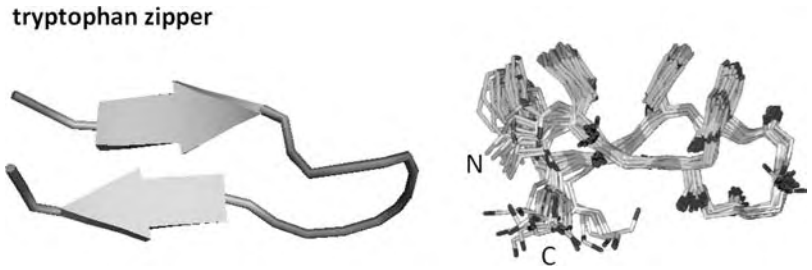


Fig. 20 Tryptophan zipper (*trpzip*) motif. An NMR structure (left, 1LEO PDB entry rendered with PyMol) and an ensemble of optimally aligned NMR structures of the motif (right) (A. G. Cochran, N. J. Skelton and M. A. Starovasnik, Tryptophan zippers: stable, monomeric β–hairpins, *Proc. Natl. Acad. Sci. USA*, **98**, 5578. Copyright (2001) National Academy of Sciences, U.S.A.[100]).

similar to those of fully folded proteins.[100] Alhtough such designs, dubbed tryptophan zippers (*trpzip*), were not found to have analogous sequences in native proteins, they continue to have strong appeal because of their design simplicity and of being stable monomeric β-hairpins that are cooperatively folded in water (Fig. 20). A typical design is a type II′ turn sequence (EGNK) extended with a WTW sequence at each end and with an additional residue to promote the formation of cross-strand hydrogen bonds between the N and C termini. When folded the hairpin adopts a common turn geometry and edge-to-face packing motif, which can be farther optimised.[101]

Following upon similar principles de novo β-meander motifs have been recently introduced. One approach looked to promote β-sheet face inversions at the expense of the promiscuous nature of β-sheet edges.[102] By specifically maximizing hydrogen bonding between β-strands and by punctuating these strands with strong turn motifs the specific β-topology which can be controlled to a reasonable extent was generated without taking into account side-chain interactions. Such purely backbone designs can be tailored or augmented with funcitonal side-chain optimisation (Fig. 21).

Often in new designs it is necessary to avoid alternative structures. Typically it is a choice between stabilising a desired structure and destabilising or discriminating against undesired structures. The concept is known as positive (towards desired) and negative (away from alternative) design.[7,8] A classical example is the design of a β-sandwich model reported as a β-doublet.[7] In this design a Greek key topology represented alternative structures that were disfavoured by introducing short β-turns (XDGX) and by choosing the central hairpin with hydrophobic side chains placed across wide rather than narrow pairs of hydrogen bonds. Sequences were generated and analysed for localised regions of strain which were relieved by amino acid substitutions. Following repeated evaluations of van der Waals contacts and bond geometries a final sequence was chosen to fold as a β-doublet (Fig. 21).

Anti-parallel associations of β-strands are predominant in proteins. This is not surprising as these are backbone structures, in which N-terminus of one strand always faces the C-terminus of its neighbouring strand. For this

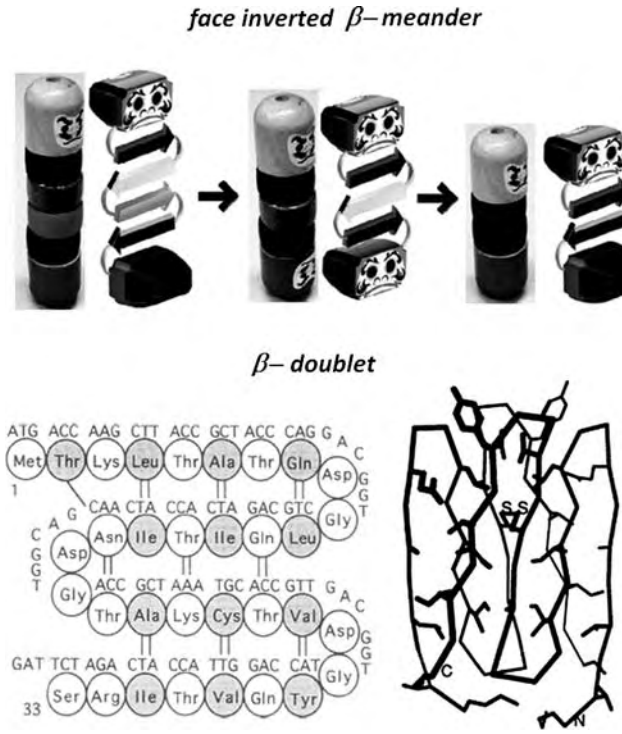


Fig. 21 Designed all-β motifs. (Top) schematic representation of β-sheet face inversion. Deletion of the second strand from the bottom makes the sheet turn over (I), whereas the further deletion of second strand from the top restores the original orientation. (Bottom) gene and peptide sequences and a stereo representation of the β-doublet formed by two identical dimerised sheets. Shaded and unshaded residues form the internal core and are solvent exposed, respectively. Solid lines show hydrogen bonds supporting the four-stranded β-sheet. (Reproduced with permission from reprinted with permission from K. Makabe and S. Koide. The promiscuity of beta-strand pairing allows for rational design of beta-sheet face inversion, *J. Am. Chem. Soc*., **130**, 14370. Copyright (2008) American Chemical Society[102] and T. P. Quinn, N. B. Tweedy, R. W. Williams, J. S. Richardson and D. C. Richardson, Betadoublet: de novo design, synthesis, and characterization of a beta-sandwich protein, *Proc. Natl. Acad. Sci. USA*, **91**, 8747. Copyright (1994) National Academy of Sciences, U.S.A.[7]).

reason, autonomous parallel β-sheets from one sequence are not possible. The problem of designing these can be solved with the help of diacid or diamine linkers which by covalently joining two strands at their same termini can provide a parallel arrangement. Also called C-to-C and N-to-N linkers these structures can promote the formation of parallel β-hairpins analogously to turns in anti-parallel hairpins (Fig. 22).[103,104]

Proline residues in only one configuration were shown to be able to support anti-parallel orientation and, when used in C-to-C linkers such as two-residue loops comprising ^{D}Pro and 1,1-dimethyl-1,2-diaminoethane (DADME), can nucleate parallel β-hairpin formation. This is in marked contrast to cyclohexanedicarboxylic acid (CHDA), both enantiomeric forms of which were shown to be favouring parallel β-sheet formation.[103,104] CHDA-Gly linkers connecting the N-termini of two strands can serve as a main determinant for designing parallel folding, which intrinsically is subject to attractions between the strands and does not require the

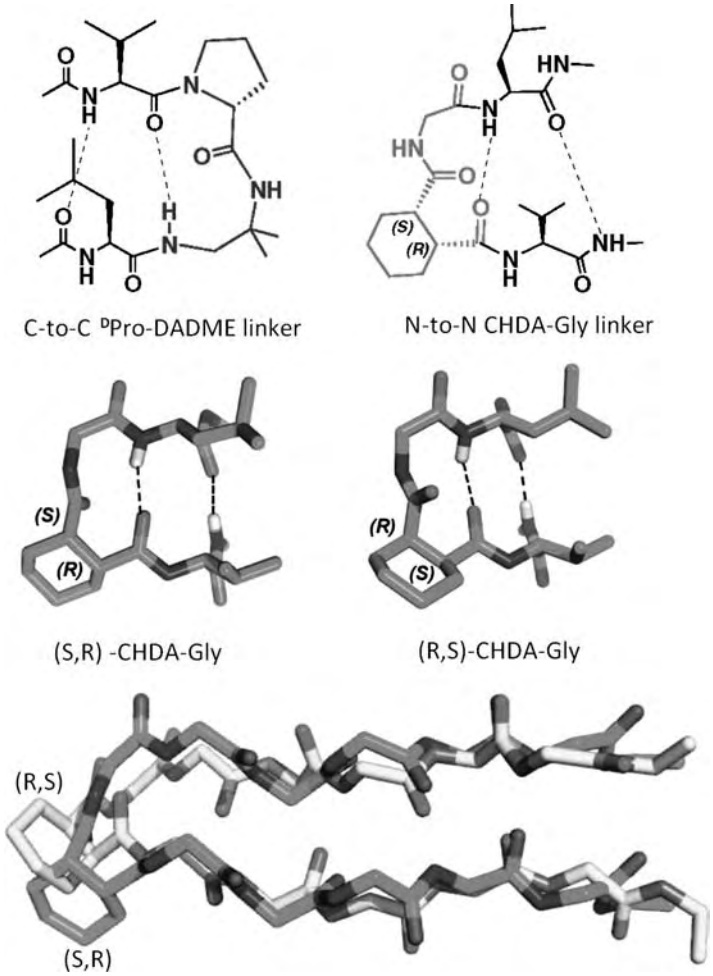


Fig. 22 Parallel β-hairpin formation. (Top) C-to-C and N-to-N linkers in tetrapeptide frameworks. (Middle) solid-state conformations of CHDA-based linkers. Hydrogen atoms only for NH of Val and Leu are shown. (Bottom) overlay of NMR structures of two hairpins supported by two CHDA configurations (reprinted with permission from F. Freire, J. D. Fisk, A. J. Peoples, M. Ivancic, I. A. Guzei and S. H. Gellman. Diacid linkers that promote parallel beta-sheet secondary structure in water, *J. Am. Chem. Soc.*, **130**, 7839. Copyright (2008) American Chemical Society).[104]

support of tertiary contacts (Fig. 22). Owing to their predictable nature the designs are adapted for generating artificial folds such as those based on D-amino acids.[105]

β-structure designs tend to aggregate which is often associated with undesired amyloidogenic properties. This constitutes their main disadvantage re-emphasising the importance and challenge of designing monomeric β-structures. It alludes also to ample opportunities for rational strategies to control aggregation. Early evidence can come from observations of amyloid-like properties of designed β-hairpin structures which has led to a conclusion that folded β-sheets can polymerise through edge-by-edge interactions and exhibit helical modes of assembly (Fig. 23).[106,107]

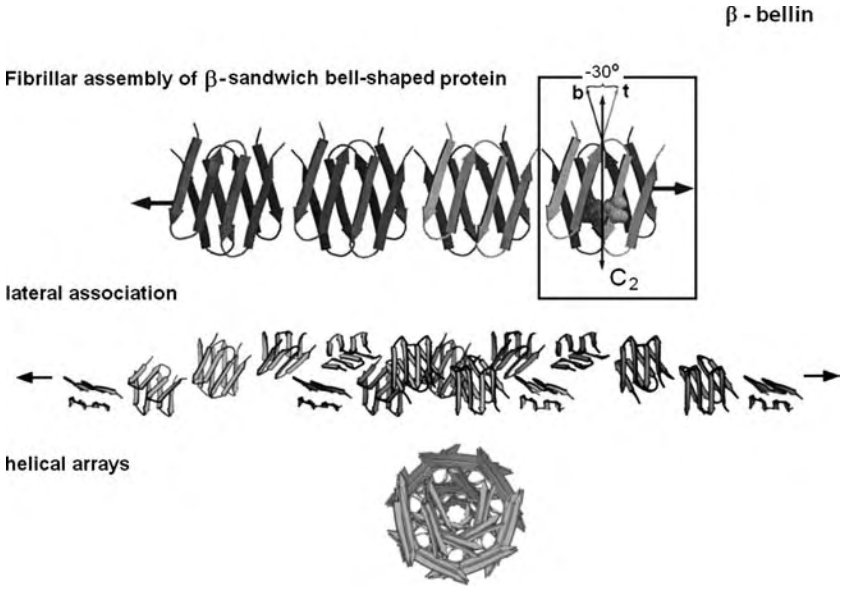


Fig. 23 Fibrillar polymerisation of β-sandwich bell-shaped proteins (β-bellins). (Top) β-sandwich monomers associate laterally. A vertical double-headed arrow denotes the C_2 axis of rotation through the monomer (in square). The dihedral angle between the bottom strand axis *b* and the top strand axis *t* is $-30°$. (Middle) β-sandwich monomers associate *via* the formation of a double helix. Constituent single helices are shown in the left and right parts. The double helix is shown in the centre. (Bottom) Projection along the fibril direction from the double helical part. (reproduced with permission from A. Lim, M. J. Saderholm, A. M. Makhov, M. Kroll, Y. Yan, L. Perera, J. D. Griffith and B. W. Erickson, Engineering of betabellin-15D: a 64 residue beta sheet protein that forms long narrow multimeric fibrils, *Protein Sci.*, 1998, **7**, 1545. Copyright Wiley-VCH Verlag GmbH and Co. KGaA. Reproduced with permission,[106] and reprinted from Biophys J, 83, H. Inouye, J. E. Bond, S. P. Deverin, A. Lim, C. E. Costello and D. A. Kirschner, Molecular Organization of Amyloid Protofilament-Like Assembly of Beta-bellin 15D: helical array of *β*-sandwiches, 1716, Copyright (2002), with permission from Elsevier[107]).

Similar tendencies were found for single β-strands, linear peptide amphiphiles and two-strand β-hairpins, all of which were shown to form fibres with regular nanoscale features characteristic of paracrystalline phases.[108–110] In the case of hairpins the importance of particular turn arrangements, namely the inclusion of ^{D}Pro-X, was demonstrated to influence the assembly.[110] By choosing between ^{D}Pro and ^{L}Pro it proved to be possible to maintain the parallel or anti-parallel orientation of the strands and to tailor specific fibrillar morphologies. The strategy represents an appealing rationale which allows controlling the properties of peptide-based materials from the bottom up (Fig. 24). The effect of turns was apparent to underpin externally triggered functions including hydrogelation for cell encapsulation and drug release and membrane-mediated anti-microbial responses,[111–113] and can also be extended to thermally reversible hydrogels by making β-strands in the hairpin structure exchangeable or swapping.[114]

With their morphological and dimensional properties being reminiscent of those for native collagen and fibrin extracellular matrices, designer fibres find use as cell-supporting scaffolds that are viewed as critical for tissue

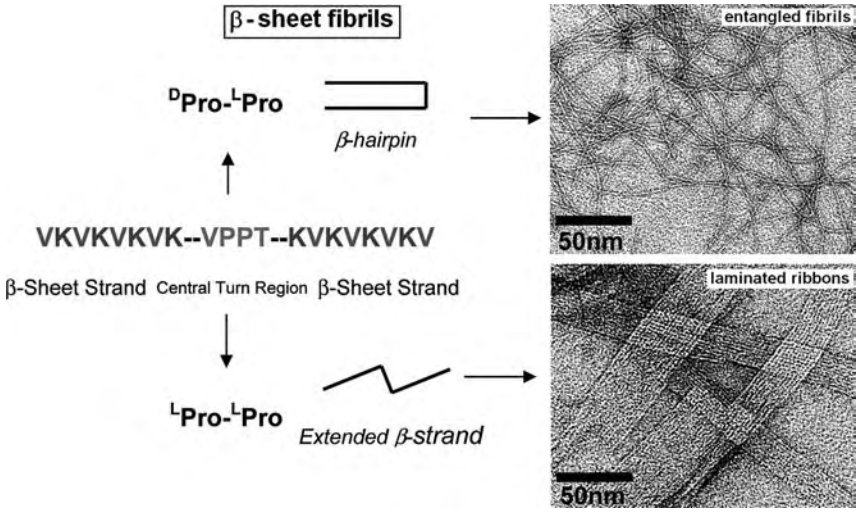


Fig. 24 Effect of turns on tuning fibrillar materials. Schematics (left) and electron micrographs (right) for ^{D}Pro-^{L}Pro turns promoting the formation of β-hairpins that reversibly self-assemble into β-sheet-rich hydrogels, and for ^{L}Pro-^{L}Pro turns imposing extended conformations that assemble into irreversible laminated ribbons. (reprinted with permission from M. S. Lamm, K. Rajagopal, J. P. Schneider and D. J. Pochan. Laminated morphology of nontwisting beta-sheet fibrils constructed *via* peptide self-assembly, *J. Am. Chem. Soc.*, **127**, 16692. Copyright (2005) American Chemical Society[110]).

formation and repair. Different designed systems have been tried for tissue engineering. Most of these systems possess strong gelation properties and the ability to support 3D cell culture, with some systems being injectable with demonstrated potential for *in vivo* applications.[115–118]

Encouraged by the clear success in constructing β-sheet fibres other self-assembling morphologies have been targeted by all β designs. Notablly, tailored hairpin architectures as well as orthogonal covalent arrangements of β-strands were shown to be able to generate capsule- or cage-like assemblies (Fig. 25).[119,120] These nanoscale structures result from the C_3-symmetric reversible assembly of the β-sheet monomers resembling thus the virus architecture and hence holding promise for controlled macromolecular transport inside living cells. However, the lack of examples of using such structures in gene and drug delivery or biomedicine in general is also apparent, which probably reflects upon the fact that spherical nanostructures are strongly prone to aggregation and to environmentally-triggered transitions to fibrillar phases, which in turn results from the less robust and stable assembly pathways of short peptide blocks when compared to native cage-like systems such as viruses and protein cages. Nevertheless, the reported examples of self-assembled encapsulants show great potential for the design of novel materials and should be particularly conducive to those applications where small (<100 nm) and uniformly sized nanoparticles are needed.

Factors discriminating for one morphology type against another are not entirely understood. Polypeptide sequence and structure have an ultimate answer to this. Irrespective of their size and complexity protein structures are programmed at the primary structure level. This does not necessarily

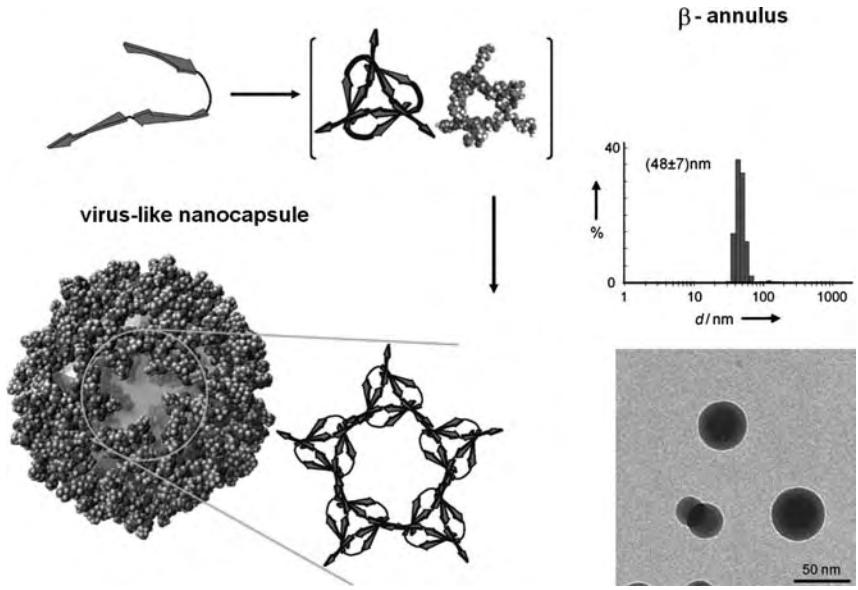


Fig. 25 Hypothetical assembly of virus-like capsules from β-hairpins *via* the formation of a β-annulus (in brackets). (Right) size distribution and morphology of assembled particles determined by dynamic light scattering and electron microscopy (K. Matsuura, K. Watanabe, T. Matsuzaki, K. Sakurai and N. Kimizuka, Self-assembled synthetic viral capsids from a 24-mer viral peptide fragment, *Angew Chem. Int. Ed. Engl.*, 2010, **49**, 9662. Copyright Wiley-VCH Verlag GmbH and Co. KGaA. Reproduced with permission.[120]).

accommodate assembly imperfections and mismatches caused by external stimuli which can lead to the partial or complete loss of programmed assembly properties. Complementary routes exist to compensate these and to even provide empirical control over the hierarchical building-up of nano-to-microscale structures. The involvement of chemical and ezymatic reactions is plausible for structural motifs to exhibit specific roles of creating spatial confinements favouring particular types of assembly nucleation and growth.

This route would require system designs rather than structure designs and should encompass both sequences capable of self-assembly and selection strategy allowing for self-correction and amplification of the most stable conformers. One solution to this can be provided by the use of reversible enzyme-catalysed reactions which enables to direct β-sheet self-assembly under thermodynamic control (Fig. 26).[121] Such system designs can access structurally diverse materials by rectifying kinetic defects that occur at higher rates for higher order structures. Directly correlated with enzyme concentrations various self-assembled populations can be empirically selected by manipulating enzyme clusters and their sizes which can serve as one variable for designs based on a single motif.[122]

All α motifs

In contrast to β-structures, helical designs are characterised by a greater structural plasticity or by the ability to accommodate internal structural conflicts which allows them to choose between different conformational states, *i.e.* from partially helical to a β- or mixed form (Fig. 27),[123–126] from

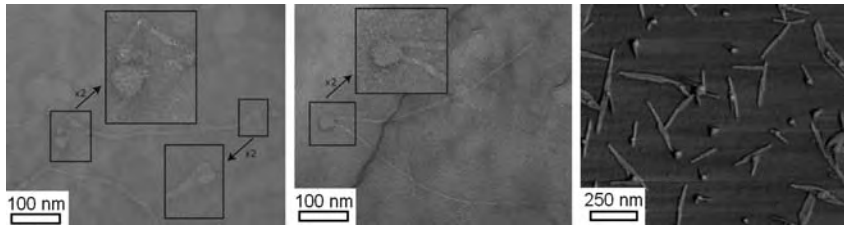


Fig. 26 Spatial confinement of enzyme-catalysed nucleation and growth. Electron and atomic force micrographs demonstrating spherical structures (30 nm in diameter) each containing several enzyme molecules (2 nm in diameter) from which fibres propogate over time (Reprinted by permission from Macmillan Publishers Ltd: Nat. Nanotechnol. (R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das and R. V. Ulijn, Enzyme-assisted self-assembly under thermodynamic control, 4, 19), copyright (2009)).[121]

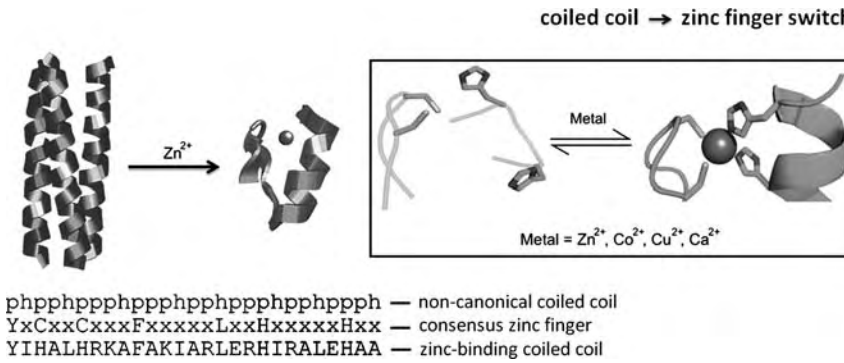


Fig. 27 Desinged zinc-bindng coiled coil. Two folding patterns are incorporated into one sequence – a non-canonical coiled coil pattern and a consensus zinc finger (p – polar, h – hydrophobic, x – any). A three-helix bundle assembles at benign conditions but switches into a zinc finger upon the addition of Zn^{2+} ions (reprinted with permission from E. Cerasoli, B. K. Sharpe, D. N. Woolfson, ZiCo: a peptide designed to switch folded state upon binding zinc, *J. Am. Chem. Soc.*, **127**, 15008. Copyright (2005) American Chemical Society[124]). (In brackets) schematic rearrangement of a zinc finger with two sulfhydryl groups (cysteines) and two imidazole nitrogen atoms (histidines) coordinating divalent metal ions (Reproduced from Y. Berezovskaya, C. T. Armstrong, A. L. Boyle, M. Porrini, D. N. Woolfson and P. E. Barran, Metal binding to a zinc-finger peptide: a comparison between solution and the gas phase. *Chem. Commun.*, 2011, **47**, 412, with permission of The Royal Society of Chemistry[126]).

one helical motif to another[127] and from a random coil to a helix.[128] This constitutes one direction in designing helix-based motifs, that is those structures that can switch between α- and β-conformations and between unstructured and helix states.

To render a given sequence switching one has to consider a trigger which will mediate the transition of one folded state into a thermodynamically more favourable conformation.[129] Typically it is done in two steps. Firstly, two different states are superimposed into one sequence by incorporating the patterns of two corresponding motifs. Secondly, one motif is designed to prevail at benign conditions but to become ''frustrated'' when an external trigger is introduced thereby giving way to the folding pattern of the second motif. Reported examples include coiled-coil designs that convert into helical hairpins under a reversible oxidative stress,[127] coiled coils[124,126] or partial helices[128] that upon binding metal ions fold into α,β-mixed hairpins

and more pronounced helices, respectively. Designed coiled coils can also transform into nanostructured fibres, which retain their α-helix structure, by pH tuning[130] or metal complexation.[131] Single sequences can encode different nanoscale morphologies resulting from controlled changes in pH and peptide concentrations and can be either α-helical or derive from a random coil or a β-sheet.[132] Elevated temperatures are also known to trigger $\alpha \rightarrow \beta$ transitions. However, these, albeit very common, lead to amyloid-like structures and are difficult to revert.[123,133,134]

Another direction in the de novo design of all helical motifs, which is often complementary to the first, is focused on helix bundles, be these stand-alone bundles, templated bundles or self-assembling helices. As described above the challenge in designing stand-alone bundles relates to the ability of a peptide to switch its oligomerisation state and to remain in a precise oligomeric form. Oligomerisation normally occurs spontaneously,[135–140] but can be also triggered by a chemical stimulus.[141] Designed bundles follow native hairpin-like topologies (Fig. 14). Sequence hairpins,[136–138] dimerised two-helix hairpins[140] and dimerised domain-swapped four-helix bundles[139] have been reported (Fig. 28).

Oligomerised monomeric helices extend the repertoire of coiled-coil designs including homo- and hetero-oligomers with helices arranged in parallel and anti-parallel fashions.[63,141–147] To date, oligomeric designs containing up to seven helices have been described.[78–80] Larger structures can be found amongst native systems belonging to membrane-integrated proteins that provide well-defined interfacial channels.[148,149] Helical bundles and coiled coils are tubular or barrel-like architectures with central channels running through their centres. However, only oligomers comprising five or more helices can be regarded as having evident pores that can be tailored by design.[79] The diameters of such channels roughly correlate

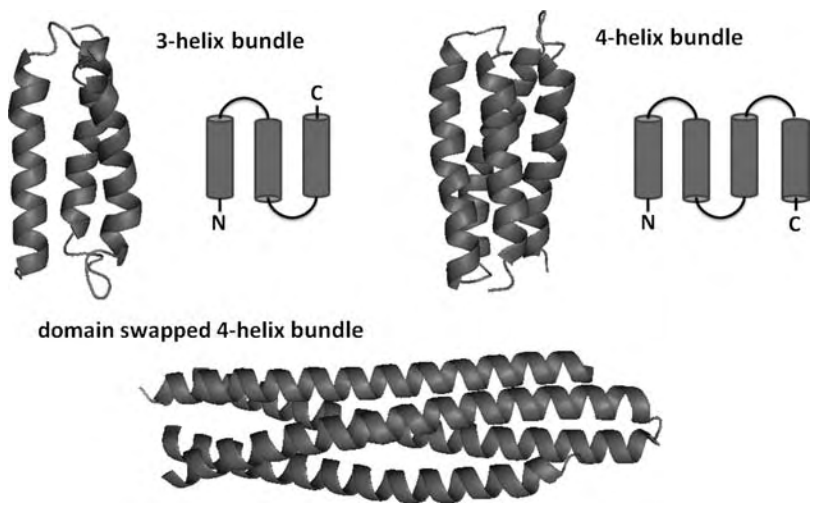


Fig. 28 De novo helix bundles. The domain-swapped bundle comprises two four-helix bundles. Each monomeric bundle has two long α-helices that intertwist with the helices of the other monomeric bundle (shown). Structures for three-helix, four-helix and domain swapped bundles are from 2A3D, 2JUA and 3VJF PDB entries rendered with PyMol.

with the oligomer size increasing by 1 Å per helix, from 5 Å for pentamers to 7 Å for hexamers.[78–80] The stability of designed channels is strongly linked to the cooperative folding of the bundles themselves. Compromised stability can lead to aggregation. One approach to avoid this is use template-assembled synthetic proteins (TASPs).[150,151] A TASP has a non-liner topology and is an array of branched polypeptides orthogonally attached to a molecular template. Folding in TASPs is pre-determined by the template and the covalent attachment. Although it is more straightforward to define the number of branches and hence the type of an oligomer, *e.g.* three or four chains to shape three- and four-helix bundles, TASPs can still self-associate and are not necessarily rigid arrangements to provide uninterrupted interfaces.[151,152] Besides, TASP designs have their own inherent drawbacks. They are more difficult to synthesise than linear polypeptides and are purely artificial.

De novo design may have reached a limit as to the number of helices possible for monodisperse bundles. However, oligomers above heptamers will not lose their appeal particularly given the ever-pressing demand for new design rules. In parallel with this, others see compromised stand-alone bundles as an opportunity to rationalise the design of self-assembling structures at equilibrium. Self-assembled helical oligomers are dominated by fibre structures.[153] In part, this is because of the fairly straightforward design rules of fibre assemblies, in part, because of the important roles helix-based filamentous structures play in cellular processes. Different concepts have been proposed for helical fibre assemblies ranging from sticky-end or axial staggering of α-helices[154,155] and coiled-coil dimers[156] to dendrimer displays,[157] domain-swapped dimers[158] and bifaceted leucine zippers.[159] Owing to them offering a structurally permissive background[153] helical fibres are being developed as extracellular scaffolds for tissue engineering applications.

Two main approaches are being taken. One, which is analogous to β-structured scaffolds, focuses on the gelation properties of the fibres,[160] whereas another one puts major stress on the control of the fibre geometry.[161] A series of studies revealed a considerable potential of coiled coils to accommodate varios fibre morphologies introduced at whim.[161–164] Because coiled-coil interfaces offer well understood intermolecular relationships their predictable manipulation is possible. For example, the incorporation of competitive but complementary binders can generate alternative interfaces through which the binders can direct the hierarchical assembly of fibre-forming monomeric blocks.[161] This gives rise to different shapes of individual fibres and inter-fibre arrangements such as branching and crisscrossings that underpin polygonal networks and porous matrices.[161–165] Thus, alternative interfaces, which are essentially the same coiled-coil interfaces but introduced by alternative topologies of monomeric blocks, elicit additional geometric constraints into the assembly (Fig. 28). At the level of an individual fibre this is very appealing as it lends itself nicely to nanoscale engineerng.[166–169] However, more robust approaches are needed to create microscopic matrices that would span an adequate amount of space to serve as a substrate for cell adhesion and migration.

An arbitrary assembly of monomeric blocks that would incorporate the properties of both fibre-forming coiled coils and alternative binders can

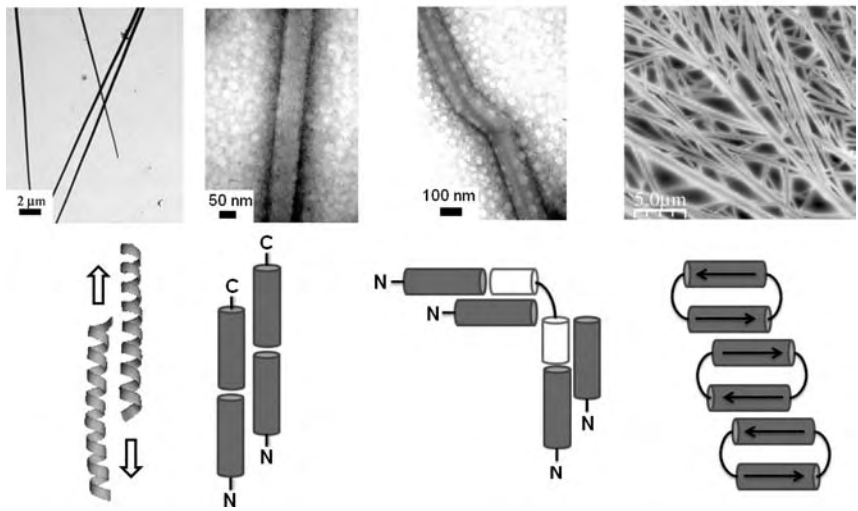


Fig. 29 De novo fibre designs. (From left to right, top) electron micrographs of straight and kinked fibres and an AFM image of microscopic matrices (reprinted by permission from Macmillan Publishers Ltd: (M. G. Ryadnov and D. N. Woolfson, Engineering the morphology of a self-assembling protein fibre, *Nat. Mater.*, **11**, 37), copyright (2003),[161] reprinted with permission from M. G. Ryadnov and D. N. Woolfson, MaP peptides: programming the self-assembly of peptide-based mesoscopic matrices. *J. Am. Chem. Soc*, **127**, 12407. Copyright (2005) American Chemical Society,[163] and A. Bella, S. Ray, M. Shaw and M. G. Ryadnov, Arbitrary self-assembly of peptide extracellular microscopic matrices, *Angew Chem. Int. Ed. Engl.*, 2012, **51**, 428. Copyright Wiley-VCH Verlag GmbH and Co. KGaA. Reproduced with permission[170]). (From left to right, above) longitudinal ''sticky-end'' topology giving straight fibres, competitive binder topology with a complementary hairpin (open cylinders) reversing the sticky-end assembly into fibre kinking and a bifaceted ''locked'' topology providing arbitrary microscopic matrices. Black arrows indicate the N-C direction of the polypeptide chain, open arrows show assembly directions. Sticky-ended dimer was rendered from 1UNW PDB entry rendered with PyMol.

provide just an efficient solution.[170] To make this happen monomers must be locked into a rigid backbone topology that would provide interfacial interactions with a significant selection freedom which cannot be achieved through the strict alignment of the helices in the bundle usually employed in fibre designs. Thus, the helix registry must be variable. This can be done by the backbone cyclisation of two complementary coiled-coil domains into one bifaceted block (Fig. 29). Although the domains are complementary to one another they cannot interact within the same block. Their arrangement in the block is anti-parallel, whereas coiled-coil interactions they can form require a parallel orientation. Therefore, helical heterodimers formed by the domains are between different blocks. This enables the lateral propagation of the blocks. Longitudinal and hence mixed propagation is induced by domain sequences comprising oppositely charged heptad repeats used as alternating modules[165] to provide minimal one-heptad shifts between the domains of different blocks (Fig. 28).[170]

The ultrastructure of the described matrices is helical. This type of assembly is readily reversible and accommodative of functional molecular modifications or decorations such as fibre recruitment of folded proteins or cell adhesion motifs to promote matrix-receptor or matrix-biomarker

interactions.[170–172] Indeed, designed helical scaffolds make a far-reaching claim in supporting cell growth and proliferation. Although tissue repair requires much more than mere cell adhesion and growth the performance of these scaffolds proves to be comparable with that of native collagene matrices *in vitro*.[160,170] This is particularly notable because native matrices are complex systems that are composed of several fully folded proteins, including structural and functional proteins, whereas de novo matrices can assemble from just a single polypeptide.

The all-helix fibre topologies have inspired designs based on other protein motifs, notably the collagen triple helix. Collagen is the main constituent of the native extracellular matrix which makes it an obvious candidate for engineering biomimetic matrices.[173] However, despite its simplicity or perhaps because of its simplicity the collagene assembly still presents a considerable challenge for design. This can be attributed to that the fibrous collagen assembly, firstly, requires long polypeptide chains that, secondly, tend to pre-organise into very stable triple helices at any hierarchical level.[174] Consequently, collagen fibres exhibit very specific and conserved nanoscale order, whereas collagen designs are inherently characterised by the high promiscuity of potential interactions. This paradox leads to a conjecture that an artificial collagen can be tackled by providing an additional capacity, which is not typical for native collagen, but which would drive fibrillar assembly from shorter sequences.[175] The latter is also critical in the light of synthetic issues associated with collagen sequences, with short oligomerising peptides being of clear preference. Thus, a staggered strand set-up in the collagen triple helix appears to be inevitable for efficient designs. Different variations of the same rationale have been proposed including covalent tethering of collagen fragments enabling strand overhangs,[175] an axial staggered orientation of strands containing a hydrophobic core with oppositely charged flanking domains[176] and sticky-ended heterotrimers.[177]

Fibre designs exemplified by different folding motifs are in fact interchangeable and complementary. The influence of all-α topologies on all-β fibre structures is as significant as that of all-β designs on all-α fibres. The same is true for collagen-based structures. It would be difficult to identify one generic concept that would underpin all known and possible future designs. Therefore, those design trends that are actively pursued for one type of secondary structure will always be the same for another type. This is as natural as encouraging for reaching the general goal of establishing robust peptide design rules. Similarly, as it is with β-structured motifs, all-α oligomers are being explored as building blocks for different nanoscale morphologies. Virus mimetics,[178–180] vaccine designs[181–183] and structured nanoscale environments for chemical reactions[184] are potential applications targeted by the most recent designs, which are primarily focused on spherical self-assemblies relying on coiled-coil formation.

Designing one-type motifs can relate to the ability of a polypeptide chain to fold and assemble on biological surfaces, membranes. It is in fact another type of conformational switching which is induced by selective binding which is in turn believed to provide a mechanistic rationale for host defence regulation of microbial invasions. Therefore, unlike other switching

Fig. 30 Clustering anitmicrobial backbone folds. Each peptide structure is represented by a 16-dimensional vector coding for the prevalence of backbone torsion angles. (Reprinted by permission from Macmillan Publishers Ltd: (C. D. Fjell, J. A. Hiss, R. E. Hancock and G. Schneider, Designing antimicrobial peptides: form follows function, *Nat. Rev. Drug Discov.*, **11**, 37), copyright (2011)).[185]

mechanisms, this type can be regarded as a direct biological response mediated by folding. A large class of peptide sequences and consequently motifs that possess the ability to adopt regular structures upon membrane binding are correspondingly called host defence peptides or antimicrobial peptides.[185] Antimicrobial peptides represent one of the richest collections of sequences that are covered in numerous reviews and deposited in databases whose number is growing.[186–188] Meanwhile, all antimicrobial sequences can be clustered into distinctive categories according to their secondary and supersecondary structure (Fig. 30). Synthetic designs aim to improve therapeutic indexes, the ratios between increased antimicrobial efficiencies and reduced toxicity. Many designs are alterations of native sequences that contribute to the elucidation of quantitative structure-activity relationships. De novo peptides are monomeric backbone designs which are used to give new insights into host-defence mechanisms and to provide basis for most enzymatically stable mimetics.[189–197]

Antimicrobial peptides are typically amphipathic cationic structures.[185] This structural preference allows them to preferentially bind anionic microbial membranes and assemble within them with the formation of supramolecular lipid-bound assemblies. Different views exist as to the exact mechanisms of the action of antimicrobial peptides ranging from poration to carpet mechanisms.[185–193] There also exist sequences that do not exhibit membrane-mediated activities, do not fold in the presence of membranes or otherwise, but instead translocate across the membranes to bind intracellular targets.[198] However, first design approximations remain constant and concern with creating sequences that can differentiate between bacterial and mammalian membranes and that fold when interacting with microbial surfaces. Furthemore, certain antimicrobial mechanisms are taken as

guiding to provide new design rationales. For example, it is thought that in order to form pores or channels in membranes membrane-bound peptides re-structure and self-assemble into transient tubular structures that allow the passage of water molecules and ions thus disbalancing and disrupting membranes.[199,200] Stabilising such channels can provide an efficient rationale for new antibiotics. The main caveat here is that the channels are not stand-alone or crystallisable assemblies. Whilst helical peptides cannot form hydrogen-bonded barrel structures and tend to be stabilised by hydrophobic interactions that in membranes are provided by lipids, β-strands can form barrel structures through hydrogen bonding without lipids. Another aspect is whether formed barrels are contiguous enough to span the membranes or strong enough to cause membrane rupture through aggregation. One strategy that can be controlled by design is to build an extensive network of hydrogen bonds which will run symmetrically through a hollow cylinder (Fig. 31).[197]

The cylinder can be formed by stacking peptide monomers whose conformation is locked into a flat ring having two faces promoting the formaion of hydrogen bonds laterally. The faces have to be equal to generate a tube with a central core running through all monomers. The arrangement of alternating D- and L-amino acids in the peptide backbone enables this and ensures the positioning of side chains on the outside surface of the assembly.

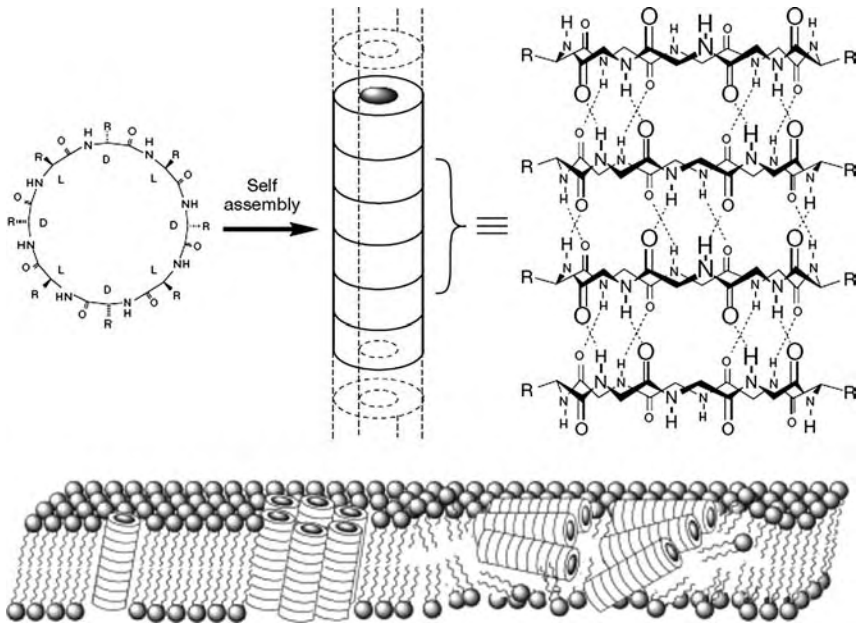


Fig. 31 Self-assembled antimicrobial nanotubes. (Top) β-sheet-like hollow architectures arranged by inter-backbone hydrogen bonds of cyclic D,L-peptides, with side chains (R) forming the outside of the architecture. (Above) potential membrane permeation mechanisms: (from left to right), intramolecular pore, barrel stave and carpet-like. Cyclic peptides are shown as rings. (Reprinted by permission from Macmillan Publishers Ltd: (S. Fernandez-Lopez, H. S. Kim, E. C. Choi, M. Delgado, J. R. Granja, A. Khasanov, K. Kraehenbuehl, G. Long, D. A. Weinberger, K. M. Wilcoxen and M. R. Ghadiri, Antibacterial agents based on the cyclic D,L-alpha-peptide architecture, *Nature*, **412**, 452), copyright (2001)).[197]

Indeed, such an architecture exerts strong antimicrobial activities and represents somewhat a hybrid assembly; it is rich in β-sheet and forms nanotubes without membranes but readily incorporates into the membranes with side chains capable of interacting with lipids (Fig. 30).[197]

The ability of peptides to span membranes and to form autonomous assemblies can provide design tools for probing regulatory mechanisms of antimicrobial action. For example, anti-antimicrobial sequences can be designed to neutralise antimicrobial sequences thus suggesting possible resistance mechanisms developed by bacteria against antimicrobial peptides which would not invole enzymatic degradation or changes in lipid compositions.[201,202] Another aspect of this synergistic ability to assemble and possess membrane-associated activities can be exploited in the design of cell-penetrating sequences[203] for specific gene and drug delivery.[204–206] This can further be explored for related switching mechanisms using polymer- and nanoparticle-based chimeras employing reversible coiled-coil assembly.[206–209]

Conformational plasticity of helical folds finds other uses. Two-helix hairpins that take somewhat an intermediate niche between bundle designs and conformational switches are designed to mimic native motifs such as EF-hands or as search frameworks for stable interhelical turns and new interfacial interaction patterns. Helix-turn-helix geometries prove to be perfectly fit for purpose. These include designed lanthanide-binding turns that retain αα-corner structures characteristic of HtH motifs. The binding is associated with increases in helicity which can be tuned and monitored thereby paving the way to artificial metallonucleases.[210,211] Due Ferri (di-iron or two iron) designs capitalise on helix-loop-helix arrangements that dimerise and bind two iron ions.[212] The structure of the proteins even without the bound ions exhibits ligand environments preorganised with sub-Å accuracy within the core of the protein. Upon addition of Fe (II) ions the diferrous centres react with oxygene forming oxo-bridged di-Fe (III) molecules. A step increase in design complexity is to mimic natural enzymes such as phenol oxidases that cycle between di-Fe (II) and di-Fe (III) states and use oxidizing equivalents for phenol-based conversions. A de novo modification of binding sites is required to enable the binding and accurate positioning of a phenol molecule as well as product release with entry to a new reaction cycle. One of the main design challenges is to provide an efficient access to the binding site without compromising its stability. Careful tuning in the sequence of the interhelical turn, which adopts an α-α_L-β conformation (Fig. 12), involving the stabilisation of hydrogen-bonded C-capping interactions of helix 1 and N-capping interactions with helix 2 (Fig. 12), led to a de novo phenol oxidase (Fig. 32).[213]

α-β motifs

As the main focus on application-oriented designs becomes stronger the boundaries between designs based on different motif types become less defined. Antimicrobial designs and designs aimed at nanotechnology applications, where more importance is given to the end product rather than to how it is made, present typical examples. Nonetheless, there remains strong interest in meeting the challenge of designing autonomous structures.

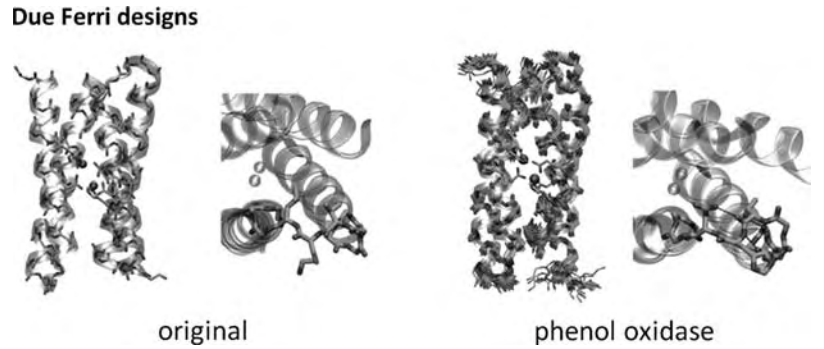


Fig. 32 De novo iron-binding (Due Ferri) helix-turn-helix designs. Two HtH motifs dimerised in anti-parallel fashion hosting iron ions (balls) in the dimer interface. Structures of original (left) and phenol oxidase (right) designs with corresponding loop structures. N- and C-capping interactions are shown for phenol oxidase (Reprinted by permission from Macmillan Publishers Ltd: (M. Faiella, C. Andreozzi, R. T. de Rosales, V. Pavone, O. Maglio, F. Nastri, W. F. DeGrado and A. Lombardi. An artificial di-iron oxo-protein with phenol oxidase activity, *Nat. Chem. Biol.*, **5**, 882), copyright (2009)).[213]

Particularly so this is deemed for de novo α-β motifs that constitute a necessary step towards functional proteins. Mixed motifs require covalent assemblies of different secondary structures. Understandably, such designs are more complex and are not as abundant as α or β designs. α-β structures require a stricter stereochemical control over the nucleation of different secondary motifs as this pre-defines the cooperativity of different hydrogen-bond patterns shared in one motif.[214] The use of conformationally constrained residues can facilitate in stabilising the patterns locally and in joining them into one mixed structure. For instance, α-aminoisobutyric acid (Aib) and ^{D}Pro that nucleate helical and β-hairpin structures, respectively, can be used as transition points between α-helix and β-sheet patterns in one sequence, whereas single achiral Gly residues can introduce helix-capping motifs, such as a Schellman motif,[53] thereby controlling the transition.[214] Cooperative folding in mixed motifs can require additional stabilising measures, which also help avoid aggregation. Stabilising tertiary contacts can be introduced to drive the formation of interfacial cores between secondary-structure segments as, for example, can be done by locking hydrophobic residues into a hydrophobic core between α-helix and β-sheet segments (Fig. 33).[215]

Alternatively, by dividing the conformational space of one fold into two distinct conformational spaces of similar complexity, namely backbone and side-chain conformations, and by treating them separately one can design a backbone template with the most favourable combination of side-chain rotamers to generate a target structure. Following this approach a topology design whose geometric parameters can be adjusted with specific turn structures is tackled first. A side-chain sequence can then be selected with each residue classified for its preference in core, surface or boundary positions with respect to the backbone surface. This strategy permits a combination of both rational and combinatorial approaches which can also incorporate computational elements.[217] In many cases new proteins are templated on existing native sequences to provide a designable backbone

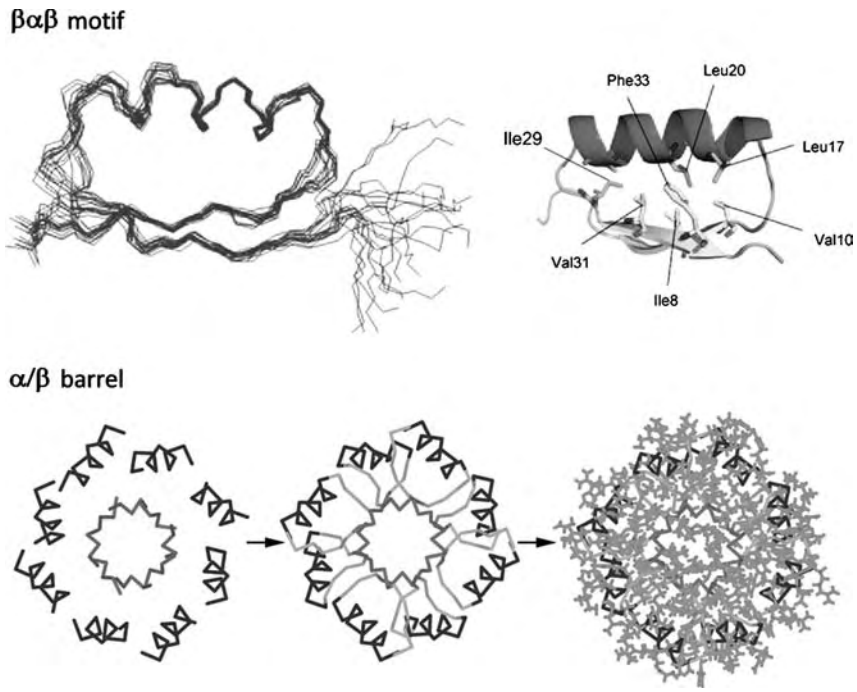


Fig. 33 De novo α-β designs. (Top) NMR structures of an $\beta\alpha\beta$ motif. A hydrophobic core is highlighted between the helix and β-sheet portions of the fold. (Bottom) schematics of the idealized backbone and side-chain ensemble (far right) of a β-sheet barrel surrounded with an α-helix barrel. Short αβ1, αβ3 and βABα turns connecting the barrels are shown in gray (reproduced with permission from H. Liang, H. Chen, K. Fan, P. Wei, X. Guo, C. Jin, C. Zeng, C. Tang and L. Lai. De novo design of a beta alpha beta motif, *Angew Chem. Int. Ed. Engl*., 2009, **48**, 330. Copyright Wiley-VCH Verlag GmbH and Co. KGaA. Reproduced with permission,[215] and reprinted from F. Offredi, F. Dubail, P. Kischel, K. Sarinski, A. S. Stern, C. van de Weerdt, J. C. Hoch, C. Prosperi, J. M. François, S. L. Mayo and J. A. Martial, De novo backbone and sequence design of an idealized alpha/beta-barrel protein: evidence of stable tertiary structure. *J. Mol. Biol*., **163**, Copyright (2003), with permission from Elsevier[217]).

with a residual sequence memory and some bias towards the native structures.[218,219] De novo proteins have no sequences to be based on and no bias to define a target backbone. It may be the case then that a search of nearby conformational space in addition to sequence space is necessary.[146] Computational procedures appear to give promising leads and up to the point of de novo α/β topologies that can show remarkably precise and superposable correlations with crystal structures.[220] This is possibly the shortest route to ensure a traceable impact of de novo design on detailed prediction and elucidation of protein structure-function relationships. De novo α/β topologies are elite designs and when endowed with biological functions will provide evidence for working synergies between experimental and computational approaches.

6 Future perspectives

In this chapter current trends in prescriptive peptide design have been outlined. The chapter does not aim at a full coverage of the area which is

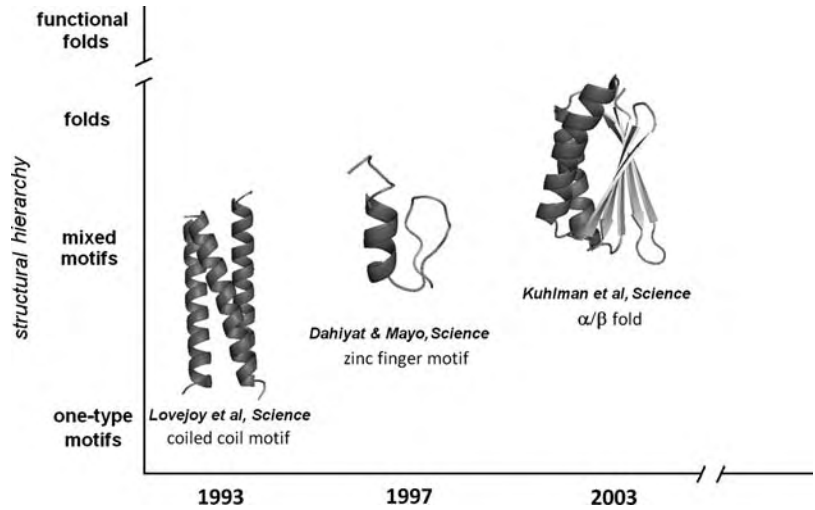


Fig. 34 A progress timeline of de novo designs versus structural hierarchy. Designed structures are from 1COS, 1PSV and 1QYS PDB entries (left to right) rendered with PyMol.

doubtfully achievable in a single volume. Nonetheless, the given highlights should be deemed sufficient for the reader to follow and possibly to contribute to future developments. It would be fair to say that the existing trends in de novo design will remain focused on exploiting traditional or similar routes based on known motifs and topologies. It is less straightforward to contemplate what would be the exact direction that will bear fruit, and whether such a direction exists or has yet to emerge (Fig. 34).

New ideas can come as a logical consequence of developing more advanced measurement methods and tools or as a disruptive technology. Peptide self-assembly can serve as a good example. Not defined at all or considered as an experimental side effect just two decades ago, peptide self-assembly is dominating and expanding the field now. It is the ease with which simple self-assembling designs access functions and applications that could not be provided before. Self-assembly indeed opens new possibilities for exploitation, which will certainly lead to new discoveries and are anticipated to stimulate commercialisation already finding unique niches in industry. However, two main challenges will remain constant for the foreseeable future. These are rational drug design and functional protein design. Selective inhibitors and promoters of very specific biomolecular interactions, enzymes and protein machinery created from scratch are the end products of de novo design that all directly relate to autonomous structures functioning in dynamic environments. This is the key reason why new designs have tremendous potential in biology, medicine and technology and must be tried.

References

1 A. L. Boyle and D. N. Woolfson, *Chem Soc Rev*., 2011, **40**, 4295.
2 J. B. Matson, R. H. Zha and S. I. Stupp, *Curr Opin Solid State Mater Sci*., 2011, **15**, 225.

3 B. A. Smith and M. H. Hecht, *Curr Opin Chem Biol*., 2011, **15**, 421.
4 E. Ruoslahti, *Matrix Biol*., 2003, **22**, 459.
5 J. Takagi, *Curr Opin Cell Biol*, 2007, **19**, 557.
6 C. Mas-Moruno, F. Rechenmacher and H. Kessler, *Anticancer Agents Med Chem*, 2010, **10**, 753.
7 T. P. Quinn, N. B. Tweedy, R. W. Williams, J. S. Richardson and D. C. Richardson, *Proc. Natl.Acad. Sci. USA*, 1994, **91**, 8747.
8 J. R. Beasley and M. H. Hecht, *J Biol Chem*, 1997, **272**, 2031.
9 C. A. Lipinski, B. F. Lombardo, B.W. Dominy and P.J. Feeney, *Adv Drug Del Rev*., 2001, **46**, 3.
10 G. P. Curley, H. Blum and M. J. Humphries, *Cell. Mol. Life Sci*., 1999, **56**, 427.
11 D. Heckmann and H. Kessler, *Meth. Enzymol*., 2007, **426**, 463.
12 J. Wermuth, S. L. Goodman, A. Jonczyk and H. Kessler, *J. Am. Chem. Soc*., 1997, **119**, 1328.
13 C. M. Goodman, S. Choi, S. Shandler and W. F. DeGrado, *Nat Chem Biol*., 2007, **3**, 252.
14 S. J. Shandler, M. V. Shapovalov, R. L. Dunbrack and W. F. DeGrado, *J Am Chem Soc*., 2010, **132**, 7312.
15 G. Guichard and I. Huc, *Chem Commun*., 2011, **47**, 5933.
16 P. G. Vasudev, S. Chatterjee, N. Shamala and P. Balaram, *Chem Rev*., 2011, **111**, 657.
17 E. G. Hutchinson and J. M. Thornton, *Protein Sci*., 1994, **3**, 2207.
18 P. Koehl and M. Levitt, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 12524.
19 J. Zhang, Y. Liang and Y. Zhang, *Structure*, 2011, **19**, 1784.
20 http://scop.mrc-lmb.cam.ac.uk/scop/
21 http://www.cathdb.info
22 http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page + LibInfo + -id + 73IcK1g5KuB + -lib + FSSP
23 http://ptgl.uni-frankfurt.de
24 http://fatcat.burnham.org/TOPS/main.html
25 K. Drew, P. Winters, G. L. Butterfoss, V. Berstis, K. Uplinger, J. Armstrong, M. Riffle, E. Schweighofer, B. Bovermann, D. R. Goodlett, T. N. Davis, D. Shasha, L. Malmström and R. Bonneau, *Genome Res*., 2011, **21**, 1981.
26 R. Srinivasan, P. J. Fleming and G. D. Rose, *Methods Enzymol*., 2004, **383**, 48.
27 R. Srinivasan and G. D. Rose, *Proteins*, 1995, **22**, 81.
28 D. N. Woolfson, P. A. Evans, E. G. Hutchinson and J. M. Thornton, *Protein Eng*., 1993, **6**, 461.
29 K. Makabe, S. Yan, V. Tereshko, G. Gawlak and S. Koide, *J. Am. Chem. Soc*., 2007, **47**, 14661.
30 R. Khayat and J. E. Johnson, *Structure*, 2011, **19**, 904.
31 C. Abad-Zapatero, S. S. Abdel-Meguid, J. E. Johnson, A. G. Leslie, I. Rayment, M. G. Rossmann, D. Suck and T. Tsukihara, *Nature*, 1980, **286**, 33.
32 V. Efimov, *FEBS Lett*., 1991, **284**, 288.
33 R. M. Hughes and M. L. Waters, *Curr Opin Struct Biol*., 2006, **16**, 514.
34 K. S. Rotondi and L. M. Gierasch, *Biopolymers*, 2006, **84**, 13.
35 M. S. Searle, *Biopolymers*, 2004, **76**, 185.
36 J. W. Fairman, N. Noinaj and S. K. Buchanan, *Curr. Opin. Struct. Biol*., 2011, **21**, 523.
37 F. Yang, L. G. Moss and G. N. Phillips, *Nat. Biotechnol*., 1996, **14**, 1246.
38 C. K. Chen, N. L Chan and A. H. Wang, *Trends Biochem Sci*., 2011, **36**, 553.
39 E. K. Leinala, P. L. Davies and Z. Jia, *Structure*, 2002, **10**, 619.
40 H. H. Tsai, K. Gunasekaran and R. Nussinov, *Structure*, 2006, **14**, 1059.

41 X. Shao and N. V. Grishin, *Nucl. Acids Res.*, 2000, **28**, 2643.
42 L. Aravind, V. Anantharaman, S. Balaji, M. M. Babu and L. M. Iyer, *FEMS Microbiol. Rev.*, 2005, **29**, 231.
43 C. L. Santos, F. Tavares, J. Thioulouse and P. Normand, *FEMS Microbiol. Rev.*, 2009, **33**, 411.
44 C. Murre, G. Bain, M. A. van Dijk, I. Engel, B. A. Furnari, M. E. Massari, J. R. Matthews, M. W. Quong, R. R. Rivera and M. H. Stuiver, *Biochim. Biophys. Acta*, 1994, **1218**, 129.
45 C. A. Grove, F. De Masi, M. I. Barrasa, D. E. Newburger, M. J. Alkema, M. L. Bulyk and A. J. Walhout, *Cell*, 2009, **138**, 314.
46 A. J. Doherty, L. C. Serpell and C. P. Ponting, *Nucleic Acids Res.*, 1996, **24**, 2488.
47 Y. Zhou, W. Yang, M. Kirberger, H. W. Lee, G. Ayalasomayajula and J. J. Yang, *Proteins*, 2006, **65**, 643.
48 A. Lewit-Bentley and S. Rety, *Curr. Opin. Struct. Biol.*, 2000, **10**, 637.
49 J. P. Davis, J. A. Rall, P. J. Reiser, L. B. Smillie and S. B. Tikunova, *J Biol. Chem.*, 2002, **277**, 49716.
50 A. V. Efimov, *Protein Eng.*, 1991, **4**, 245.
51 A. V. Efimov, *FEBS Lett.*, 1984, **166**, 33.
52 R. Aurora and G. D. Rose, *Protein Sci.*, 1998, **7**, 21.
53 C. Schellman, In *Protein Folding* (P. Jaenicke, ed.), 1980, 53.
54 R. Aurora, R. Srinivasan and G. D. Rose, *Science*, 1994, **264**, 1126.
55 M. Sagermann, L. G. Mårtensson, W. A. Baase and B. W. Matthews, *Protein Sci.*, 2002, **11**, 516.
56 S. J. Lahr, D. E. Engel, S. E. Stayrook, O. Maglio, B. North, S. Geremia, A. Lombardi and W. F. DeGrado, *J Mol Biol.*, 2005, **346**, 1441.
57 A. V. Efimov, *FEBS Lett.*, 1996, **391**, 167.
58 F. H.C. Crick, *Acta Crystallogr.*, 1953, **6**, 689.
59 S. Kamtekar and M. H. Hecht, *FASEB J.*, 1995, **9**, 1013.
60 W. F. DeGrado, H. Gratkowski and J. D. Lear, *Protein Sci.*, 2003, **12**, 647.
61 R. B. Hill, D. P. Raleigh, A. Lombardi and W. F. DeGrado, *Acc Chem Res.*, 2000, **33**, 745.
62 F. H. C. Crick, *Acta Cryst*, 1953, **6**, 689.
63 P. B. Harbury, T. Zhang, P. S. Kim and T. Alber, *Science*, 1993, **262**, 1401.
64 D. N. Woolfson, *Adv. Protein Chem.*, 2005, **70**, 79.
65 B. Apostolovic, M. Danial and H.-A. Klok, *Chem. Soc. Rev.*, 2010, **39**, 3541.
66 M. G. Ryadnov, *Biochem. Soc. Trans.*, 2007, **35**, 487.
67 H. Robson Marsden and A. Kros, *Angew. Chem. Int. Ed.*, 2010, **49**, 2988.
68 A. N. Lupas and M. Gruber, *Adv. Protein Chem.*, 2005, **70**, 78.
69 P. Burkhard, J. Stetefeld and S. V. Strelkov, *Trends Cell Biol.*, 2001, **11**, 82.
70 M. R. Hicks, D. V. Holberton, C. Kowalczyk and D. N. Woolfson, *Fold Des.*, 1997, **2**, 149.
71 C. Chothia, M. Levitt and D. Richardson, *J. Mol. Biol.*, 1981, **145**, 215.
72 F. Naider and J. Anglister, *Curr. Opin. Struct. Biol.*, 2009, **19**, 473.
73 R. M. Markosyan, M. Y. Leung and F. S. Cohen, *J. Virol.*, 2009, **83**, 10048.
74 V. Buzon, G. Natrajan, D. Schibli, F. Campelo, M. M. Kozlov and W. Weissenhorn, *PLoS Pathog.*, 2010, **6**, e1000880.
75 M. J. Root, M. S. Kay and P. S. Kim, *Science*, 2001, **291**, 884.
76 V. N. Malashkevich, R. A. Kammerer, V. P. Efimov, T. Schulthess and J. Engel, *Science*, 1996, **274**, 761.
77 J. H. Jeoung, D. A. Pippig, B. M. Martins, N. Wagener and H. Dobbek, *J. Mol. Biol.*, 2007, **368**, 1122.
78 J. Liu, W. Yong, Y. Deng, N. R. Kallenbach and M. Lu, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 16156.

79 N. R. Zaccai, B. Chi, A. R. Thomson, A. L. Boyle, G. J. Bartlett, M. Bruning, N. Linden, R. B. Sessions, P. J. Booth, R. L. Brady and D. N. Woolfson, *Nat Chem. Biol*., 2011, **7**, 935.
80 J. Liu, Q. Zheng, Y. Deng, C.-S. Cheng, N. R. Kallenbach and M. Lu, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 15457.
81 S. Rao and M. Rossmann, *J Mol Biol*., 1973, **76**, 241.
82 A. Ochoa-Leyva, F. Barona-Gómez, G. Saab-Rincón, K. Verdel-Aranda, F. Sánchez and X Soberón, *J Mol. Biol*., 2011, **411**, 143.
83 G. Saab-Rincón, L. Olvera, M. Olvera, E. Rudiño-Piñera, E. Benites, X. Soberón and E. Morett, *J Mol. Biol*., 2012, **416**, 255.
84 J. Jorda and A. V. Kajava, *Adv Protein Chem Struct Biol*., 2010, **79**, 59.
85 B. Kobe and A. V. Kajava, *Curr. Opin. Struct. Biol*., 2001, **11**, 725.
86 J. H. Laity, B. M. Lee and P. E. Wright, *Curr. Opin. Struct. Biol*., 2001, **11**, 39.
87 S. S. Krishna, I. Majumdar and N. V. Grishin, *Nucleic Acids Res*., 2003, **31**, 532.
88 F. J. Blanco, M. A. Jimenez, J. Herranz, M. Rico, J. Santoro and J. L. Nieto, *J Am. Chem. Soc*., 1993, **115**, 5887.
89 F. J. Blanco, G. Rivas and L. Serrano, *Nat. Struct. Biol*., 1994, **1**, 584.
90 M. S. Searle, D. H. Williams and L. C. Packman, *Nat Struct. Biol*., 1995, **2**, 999.
91 M. S. Searle, R. Zerella, D. H. Williams and L. C. Packman, *Protein Eng*., 1996, **9**, 559.
92 F. A. Syud, H. E. Stanger and S. H. Gellman, *J. Am. Chem. Soc*., 2001, **123**, 8667.
93 H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Giriat, T. Muir and S. H. Gellman, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 12015.
94 B. L. Sibanda, T. L. Blundell and J. M. Thornton, *J Mol. Biol*., 1989, **206**, 759.
95 T. S. Haque, J. C. Little and S. H. Gellman, *J Am. Chem. Soc*., 1996, **118**, 6975.
96 S. R. Raghothama, S. K. Awasthi and P. Balaram, *J Chem. Soc. Perkin Trans*, 1998, **2**, 137.
97 T. S. Haque and S. H. Gellman, *J Am. Chem. Soc*., 1997, **119**, 2303.
98 I. L. Karle, S. K. Awasthi and P. Balaram, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 8189.
99 B. Ciani, M. Jourdan and M. S. Searle, *J Am Chem. Soc*., 2003, **125**, 9038.
100 A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 5578.
101 A. G. Cochran, R. T. Tong, M. A. Starovasnik, E. J. Park, R. S. McDowell, J. E. Theaker and N. J. Skelton, *J Am Chem Soc*, 2001, **123**, 625.
102 K. Makabe and S. Koide, *J Am Chem Soc*, 2008, **130**, 14370.
103 J. D. Fisk, M. A. Schmitt and S. H. Gellman, *J Am. Chem. Soc*., 2006, **128**, 7148.
104 F. Freire, J. D. Fisk, A. J. Peoples, M. Ivancic, I. A. Guzei and S. H. Gellman, *J Am Chem. Soc*., 2008, **130**, 7839.
105 S. Annavarapu and V. Nanda, *BMC Struct. Biol*., 2009, **9**, 61.
106 A. Lim, M. J. Saderholm, A. M. Makhov, M. Kroll, Y. Yan, L. Perera, J. D. Griffith and B. W. Erickson, *Protein Sci*., 1998, **7**, 1545.
107 H. Inouye, J. E. Bond, S. P. Deverin, A. Lim, C. E. Costello and D. A. Kirschner, *Biophys J*., 2002, **83**, 1716.
108 A. Aggeli, I. A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. McLeish, A. N. Semenov and N. Boden, *Proc Natl Acad Sci USA*, 2001, **98**, 11857.
109 J. D. Hartgerink, E. Beniash and S. I. Stupp, *Proc Natl Acad Sci USA*, 2002, **99**, 5133.
110 M. S. Lamm, K. Rajagopal, J. P. Schneider and D. J. Pochan, *J Am Chem Soc*., 2005, **127**, 16692.

111 C. Sinthuvanich, A. S. Veiga, K. Gupta, D. Gaspar, R. Blumenthal and J. P. Schneider, *J Am Chem. Soc*., 2012, **134**, 6210.
112 A. Altunbas, S. J. Lee, S. A. Rajasekaran, J. P. Schneider and D. J. Pochan, *Biomaterials*, 2011, **32**, 5906.
113 C. M. Micklitsch, P. J. Knerr, M. C. Branco, R. Nagarkar, D. J. Pochan and J. P. Schneider, *Angew Chem Int Ed Engl*, 2011, **50**, 1577.
114 R. P. Nagarkar, R. A. Hule, D. J. Pochan and J. P. Schneider, *J Am Chem Soc*., 2008, **130**, 4466.
115 V. Jayawarna, M. Ali, T.A. Jowitt, A.F. Miller, A. Saiani, J.E. Gough and R.V. Ulijn, *Adv. Mater*., 2006, **18**, 611.
116 M. J. Webber, J. Tongers, C. J. Newcomb, K. T. Marquardt, J. Bauersachs, D. W. Losordo and S. I. Stupp, *Proc Natl Acad Sci USA*, 2011, **108**, 13438.
117 R. G. Ellis-Behnke, Y. X. Liang, S. W. You, D. K. Tay, S. Zhang, K. F. So and G. E. Schneider, *Proc. Natl. Acad. Sci. USA*, **103**, 5054.
118 G. A. Silva, C. Czeisler, K. L. Niece, E. Beniash, D. A. Harrington, J. A. Kessler and S. I. Stupp, *Science*, 2004, **303**, 1352.
119 K. Matsuura, H. Hayashi, K. Murasato and N. Kimizuka, *Chem Commun*, 2011, **47**, 265.
120 K. Matsuura, K. Watanabe, T. Matsuzaki, K. Sakurai and N. Kimizuka, *Angew Chem Int Ed Engl*., 2010, **49**, 9662.
121 R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das and R. V. Ulijn, *Nat Nanotechnol*., 2009, **4**, 19.
122 A. R. Hirst, S. Roy, M. Arora, A. K. Das, N. Hodson, P. Murray, S. Marshall, N. Javid, J. Sefcik, J. Boekhoven, J. H. van Esch, S. Santabarbara, N. T. Hunt and R. V. Ulijn, *Nat Chem*., 2010, **2**, 1089.
123 B. Ciani, E. G. Hutchinson, R. B. Sessions and D. N. Woolfson, *J Biol. Chem*., 2002, **277**, 10150.
124 E. Cerasoli, B. K. Sharpe and D. N. Woolfson, *J Am Chem Soc*., 2005, **127**, 15008.
125 X. I. Ambroggio and B. Kuhlman, *J Am Chem Soc*., 2006, **128**, 1154.
126 Y. Berezovskaya, C. T. Armstrong, A. L. Boyle, M. Porrini, D. N. Woolfson and P. E. Barran, *Chem. Commun*., 2011, **47**, 412.
127 M. J. Pandya, E. Cerasoli, A. Joseph, R. G. Stoneman, E. Waite and D. N. Woolfson, *J Am Chem Soc*., 2004, **126**, 17016.
128 D. J. Cline, C. Thorpe and J. P. Schneider, *J Am Chem Soc*., 2003, **125**, 2923.
129 B. A. Krantz and T. R. Sosnick, *Nat Struct Mol Biol*, 2001, **8**, 1042.
130 S. N. Dublin and V. P. Conticello, *J Am Chem Soc*., 2008, **130**, 49.
131 Y. Zimenkov, S. N. Dublin, R. Ni, R. S. Tu, V. Breedveld, R. P. Apkarian and V. P. Conticello, *J Am Chem Soc*, 2006, **128**, 6770.
132 K. Pagel, S. C. Wagner, K. Samedov, H. von Berlepsch, C. Böttcher and B. Koksch, *J Am Chem Soc*., 2006, **128**, 2196.
133 M. O. Steinmetz, Z. Gattin, R. Verel, B. Ciani, T. Stromer, J. M. Green, P. Tittmann, C. Schulze-Briese, H. Gross, W. F. van Gunsteren, B. H. Meier, L. C. Serpell, S. A. Müller and R. A. Kammerer, *J Mol Biol*, 2008, **376**, 898.
134 K. Pagel and B. Koksch, *Curr Opin Chem Biol*., 2008, **12**, 730.
135 R. B. Hill, D. P. Raleigh, A. Lombardi and W. F. DeGrado, *Acc Chem Res*., 2000, **33**, 745.
136 S. T. Walsh, H. Cheng, J. W. Bryson, H. Roder and W. F. DeGrado, *Proc Natl Acad Sci USA*, 1999, **96**, 5486.
137 M. H. Hecht, J. S. Richardson, D. C. Richardson and R. C. Ogden, *Science*, 1990, **249**, 884.
138 A. Go, S Kim, J. Baum and M. H. Hecht, *Protein Sci*., 2008, **17**, 821.

139 R. Arai, N. Kobayashi, A. Kimura, T. Sato, K. Matsuo, A. F. Wang, J. M. Platt, L. H. Bradley and M. H. Hecht, *J Phys Chem B*., 2012 DOI: 10.1021/jp212438h.
140 S. F. Betz, P. A. Liebman and W. F. DeGrado, *Biochemistry*, 1997, **36**, 2450.
141 R. S. Signarvic and W. F. DeGrado, *J Mol Biol*., 2003, **334**, 1.
142 D. A. D. Parry, R. D. B. Fraser and J. M. Squire, *J. Struct. Biol*., 2008, **163**, 258.
143 G. Grigoryan and W. F. DeGrado, *J. Mol. Biol*., 2011, **405**, 1079.
144 G. Grigoryan and A. E. Keating, *Curr. Opin. Struct. Biol*., 2008, **18**, 477.
145 E. Moutevelis and D. N. Woolfson, *J. Mol. Biol*., 2009, **385**, 726.
146 P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber and P. S. Kim, *Science*, 1998, **282**, 1462.
147 S. Nautiyal, D. N. Woolfson, D. S. King and T. Alber, *Biochemistry*, 1995, **34**, 11645.
148 V. Koronakis, A. Sharff, E. Koronakis, B. Luisi and C. Hughes, *Nature*, 2000, **405**, 914.
149 M. Mueller, U. Grauschopf, T. Maier, R. Glockshuber and N. Ban, *Nature*, 2009, **459**, 726.
150 M. Mutter, P. Dumy, P. Garrouste, C. Lehmann, M. Mathieu, C. Peggion, S. Peluso, A. Razaname and G. Tuchscherer, *Angew. Chem. Int. Ed. Engl*., 1996, **35**, 1482.
151 A. S. Causton and J. C. Sherman, *J Pept Sci*., 2002, **8**, 275.
152 J. O. Freeman and J. C. Sherman, *Chem. Eur J*., 2011, **17**, 14120.
153 D. N. Woolfson and M. G. Ryadnov, *Curr Opin Chem Biol*., 2006, **10**, 559.
154 M. J. Pandya, S. M. Spooner, M. Sunde, J. R. Thorpe, A. Rodger and D. N. Woolfson, *Biochemistry*, 2000, **39**, 8728.
155 S. A. Potekhin, T. N. Melnik, V. Popov, N. F. Lanina, A. A. Vazina, P. Rigler, A. S. Verdini, G. Corradin and A. V. Kajava, *Chem Biol*., 2001, **8**, 1025.
156 H. Dong, S. E. Paramonov and J. D. Hartgerink, *J Am Chem Soc*., 2008, **130**, 13691.
157 M. Zhou, D. Bentley and I. Ghosh, *J Am Chem Soc*, 2004, **126**, 734.
158 N. L. Ogihara, G. Ghirlanda, J. W. Bryson, M. Gingery, W. F. DeGrado and D. Eisenberg, *Proc Natl Acad Sci USA*, 2001, **98**, 1404.
159 D. E. Wagner, C. L. Phillips, W. M. Ali, G. E. Nybakken, E. D. Crawford, A. D. Schwab, W. F. Smith and R. Fairman, *Proc Natl Acad Sci USA*, 2005, **102**, 12656.
160 E. F. Banwell, E. S. Abelardo, D. J. Adams, M. A. Birchall, A. Corrigan, A. M. Donald, M. Kirkland, L. C. Serpell, M. F. Butler and D. N. Woolfson, *Nat Mater*., 2009, **8**, 596.
161 M. G. Ryadnov and D. N. Woolfson, *Nat Mater*., 2003, **2**, 329.
162 M. G. Ryadnov and D. N. Woolfson, *Angew Chem Int Ed Engl*., 2003, **42**, 3021.
163 M. G. Ryadnov and D. N. Woolfson, *J. Am Chem. Soc*, 2005, **127**, 12407.
164 M. G. Ryadnov and D. N. Woolfson, *J Am Chem Soc*., 2007, **129**, 14074.
165 M. G. Ryadnov, A. Bella, S. Timson and D. N. Woolfson, *J Am Chem Soc*., 2009, **131**, 13240.
166 D. Papapostolou, A. M. Smith, E. D. Atkins, S. J. Oliver, M. G. Ryadnov, L. C. Serpell and D. N. Woolfson, *Proc Natl Acad Sci USA*, 2007, **104**, 10853.
167 M. G. Ryadnov, D. Papapostolou and D. N. Woolfson, *Methods Mol Biol*., 2008, **474**, 35.
168 S. C. Holmström, P. J. King, M. G. Ryadnov, M. F. Butler, S. Mann and D. N. Woolfson, *Langmuir*, 2008, **24**, 11778.
169 M. G. Ryadnov and D. I. Cherny, *Macromol Biosci*., 2012, **12**, 195.

170 A. Bella, S. Ray, M. Shaw and M. G. Ryadnov, *Angew Chem Int Ed Engl*., 2012, **51**, 428.
171 M. G. Ryadnov and D. N. Woolfson, *J Am Chem Soc*., 2004, **126**, 7454.
172 A. Bella, H. Lewis, J. Phu, A. R. Bottrill, S. C. Mistry, C. E. Pullar and M. G. Ryadnov, *J Med Chem*., 2009, **52**, 7966.
173 M. D. Shoulders and R. T. Raines, *Annu Rev Biochem*., 2009, **78**, 929.
174 M. D. Shoulders, K. A. Satyshur, K. T. Forest and R. T. Raines, *Proc Natl Acad Sci USA*, 2010, **107**, 559.
175 F. W. Kotch and R. T. Raines, *Proc Natl Acad Sci USA*, 2006, **28**(103), 3028.
176 S. Rele, Y. Song, R. P. Apkarian, Z. Qu, V. P. Conticello and E. L. Chaikof, *J Am Chem Soc*., 2007, **129**, 14780.
177 L. E. O'Leary, J. A. Fallas, E. L. Bakota, M. K. Kang and J. D. Hartgerink, *Nat Chem*., 2011, **3**, 82.
178 G. Grigoryan, Y. H. Kim, R. Acharya, K. Axelrod, R. M. Jain, L. Willis, M. Drndic, J. M. Kikkawa and W. F. DeGrado, *Science*, 2011, **332**, 1071.
179 T. Riedel, A. Ghasparian, K. Moehle, P. Rusert, A. Trkola and J. A. Robinson, *Chembiochem*, 2011, **12**, 2829.
180 B. Lamarre and M. G. Ryadnov, *Macromol Biosci*., 2011, **11**, 503.
181 U. Schroeder, A. Graff, S. Buchmeier, P. Rigler, U. Silvan, D. Tropel, B. M. Jockusch, U. Aebi, P. Burkhard and C. A. Schoenenberger, *J Mol Biol*., 2009, **386**, 1368.
182 T. A. Pimentel, Z. Yan, S. A. Jeffers, K. V. Holmes, R. S. Hodges and P. Burkhard, *Chem Biol Drug Des*., 2009, **73**, 53.
183 F. Boato, R. M. Thomas, A. Ghasparian, A. Freund-Renard, K. Moehle and J. A. Robinson, *Angew Chem Int Ed Engl*., 2007, **46**, 9015.
184 M. G. Ryadnov, *Angew Chem Int Ed Engl*, 2007, **46**, 969.
185 C. D. Fjell, J. A. Hiss, R. E. Hancock and G. Schneider, *Nat Rev Drug Discov*., 2011, **11**, 37.
186 http://www.bbcm.units.it/~tossi/pag1.htm
187 http://aps.unmc.edu/AP/main.php
188 http://www.bicnirrh.res.in/antimicrobial/
189 D. L. Lee and R. S. Hodges, *Biopolymers*, 2003, **71**, 28.
190 M. G. Ryadnov, O. V. Degtyareva, I. A. Kashparov and Y. V. Mitin, *Peptides*, 2002, **23**, 1869.
191 Y. Shai and Z. Oren, *Peptides*, 2001, **22**, 1629.
192 Y. Shai, *Curr Pharm Des*, 2002, **8**, 715.
193 G. N. Tew, R. W. Scott, M. L. Klein and W. F. Degrado, *Acc. Chem. Res*., 2010, **43**, 30.
194 A. Ivankin, L. Livne, A. Mor, G. A. Caputo, W. F. Degrado, M. Meron, B. Lin and D. Gidalevitz, *Angew Chem Int. Ed Engl*., 2010, **49**, 8462.
195 E. A. Porter, X. Wang, H. S. Lee, B. Weisblum and S. H. Gellman, *Nature*, 2000, **404**, 565.
196 N. P. Chongsiriwatana, J. A. Patch, A. M. Czyzewski, M. T. Dohm, A. Ivankin, D. Gidalevitz, R. N. Zuckermann and A. E. Barron, *Proc Natl Acad Sci USA*, 2008, **105**, 2794.
197 S. Fernandez-Lopez, H. S. Kim, E. C. Choi, M. Delgado, J. R. Granja, A. Khasanov, K. Kraehenbuehl, G. Long, D. A. Weinberger, K. M. Wilcoxen and M. R. Ghadiri, *Nature*, 2001, **412**, 452.
198 M. Scocchi, A. Tossi and R. Gennaro, *Cell Mol Life Sci*., 2011, **68**, 2317.
199 L. Yang, T. M. Weiss and H. W. Huang, *Biophys J*, 2000, **79**, 2002.
200 S. Qian, W. Wang, L. Yang and H. W. Huang, *Proc Natl. Acad. Sci. USA*, 2008, **105**, 17379.

201 M. G. Ryadnov, G. V. Mukamolova, A. S. Hawrani, J. Spencer and R. Platt, *Angew Chem Int Ed Engl*., 2009, **48**, 9676.
202 A. Peschel and H.-G. Sahl, *Nat Rev Microbiol*., 2006, **4**, 529.
203 F. Heitz, M. C. Morris and G. Divita, *Br. J. Pharmacol*., 2009, **157**, 195.
204 B. Apostolovic, M. Danial and H. A. Klok, *Chem Soc Rev*., 2010, **39**, 3541.
205 B. Lamarre, J. Ravi and M. G. Ryadnov, *Chem Commun*., 2011, **47**, 9045.
206 S. P. Deacon, B. Apostolovic, R. J. Carbajo, A. K. Schott, K. Beck, M. J. Vicent, A. Pineda-Lucena, H. A. Klok and R. Duncan, *Biomacromolecules*, 2011, **12**, 19.
207 M. G. Ryadnov, B. Ceyhan, C. M. Niemeyer and D. N. Woolfson, *J Am Chem Soc*., 2003, **125**, 9388.
208 D. Aili, P. Gryko, B. Sepulveda, J. A. Dick, N. Kirby, R. Heenan, L. Baltzer, B. Liedberg, M. P. Ryan and M. M. Stevens, *Nano Lett*., 2011, **11**, 5564.
209 H. R. Marsden, J. W. Handgraaf, F. Nudelman, N. A. Sommerdijk and A. Kros, *J Am Chem Soc*., 2010, **132**, 2370.
210 J. T. Welch, W. R. Kearney and S. J. Franklin, *Proc Natl Acad Sci USA*, 2003, **100**, 3725.
211 P. Caravan, J. M. Greenwood, J. T. Welch and S. J. Franklin, *Chem Commun*, 2003, **20**, 2574.
212 O. Maglio, F. Nastri, R. F. Martin de Rosales, M. Faiella, V. Pavone, W. F. DeGrado and A. Lombardi, *C. R. Chimie*, 2007, **10**, 703.
213 M. Faiella, C. Andreozzi, R. T. de Rosales, V. Pavone, O. Maglio, F. Nastri, W. F. DeGrado and A. Lombardi, *Nat Chem Biol*, 2009, **5**, 882.
214 I. L. Karle, C. Das and P. Balaram, *Proc Natl Acad Sci USA*, 2000, **97**, 3034.
215 H. Liang, H. Chen, K. Fan, P. Wei, X. Guo, C. Jin, C. Zeng, C. Tang and L. Lai, *Angew Chem Int Ed Engl*., 2009, **48**, 330.
216 M. Levitt, M. Gerstein, E. Huang, S. Subbiah and J. Tsai, *Annu. Rev. Biochem*., 1999, **66**, 549.
217 F. Offredi, F. Dubail, P. Kischel, K. Sarinski, A. S. Stern, C. van de Weerdt, J. C. Hoch, C. Prosperi, J. M. François, S. L. Mayo and J. A. Martial, *J Mol Biol*, 2003, **325**, 163.
218 K. Raha, A. M. Wollacott, M. J. Italia and J. R. Desjarlais, *Protein Sci*., 2000, **9**, 1106.
219 B. I. Dahiyat and S. L. Mayo, *Science*, 1997, **278**, 82.
220 B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard and D. Baker, *Science*, 2003, **302**, 1364.

Targeting alpha-helix based protein interactions; nuclear receptors as a case study[†]

Lech-Gustav Milroy, Lidia Nieto and Luc Brunsveld*
DOI: 10.1039/9781849734677-00238

1 Introduction

In nature, α-helices are ubiquitous protein secondary structures and essential components of many protein-protein interactions (PPIs). Their suitability as protein recognition elements can be explained by the unique structural properties and the specific nature of the protein targets; the result of thousands of years of evolutionary pressure. In contrast to, for example, enzymatic active sites or transcription factor ligand binding domains, which typically consist of a deep solvent excluded lipophilic binding cleft, to which small drug molecules or hormones can bind, the target interface at which proteins interact are by nature shallower, more extended and hydrophilic in order to facilitate the multivalent interactions between predominantly large polar protein domains.[1–5] In this arena, the nuclear receptor (NR)-coactivator interaction is especially unique among α-helix based PPIs given the shortness of the α-helical coactivator LXXLL motif, which is defined by a so-called charge clamp located at the NR surface. The fundamental importance of nuclear receptors for numerous (patho-)physiological processes combined with the unique features of the NR-coactivator interaction make it a promising target for drug discovery and thus an intriguing case study to lead a discussion on α-helix mediated protein interactions. Attempts to inhibit the NR-coactivator interaction, or indeed any α-helix-mediated PPI, typically starts with a thorough examination of the native α-helical binding motif. While clearly an excellent drug lead, the natural unadultered NR co-activator peptide sequence is a poor drug itself, especially given that the NR target sits intracellular, but also because of the general metabolic instability of peptides. The challenge is therefore either to develop α-helical mimetics using synthetic non-peptide chemical scaffolds with improved drug-like properties or, alternatively, to improve the so-called 'druggability' of α-helical peptides through minimal chemical modifications that do not compromise the excellent target selectivity and affinity inherent to the native sequence. Either way, the development of novel inhibitors of protein-protein interactions based on α-helical structures first requires an understanding of α-helicity and how the principles of α-helicity have been generally applied to the design of novel PPI inhibitors: peptide-based or otherwise.

As mentioned previously, α-helical secondary structures play an important role in a host of PPIs. On this subject there are a number of excellent

Laboratory of Chemical Biology, Department of Biomedical Engineering, Technische Universiteit Eindhoven, Den Dolech 2, 5612 AZ, Eindhoven, The Netherlands. E-mail: l.brunsveld@tue.nl
[†] Funded by ERC grant 204554 – SupraChemBio.

recent reviews,[6–11] to which the authors wilfully direct the reader's attention. This first section of this chapter will describe the key characteristics of the α-helix and briefly highlight how the principles of α-helicity have been applied to the rational design of novel PPI inhibitors, including those based on oligomeric and foldamer scaffold structures as well as small molecule and peptide based inhibitors. In this regard, the NR-coactivator interaction can be treated separately from other α-helix-derived PPIs based on the shortness of the co-activator α-helix and the well-defined charge clamp at the NR surface, for which longer oligomer or foldamer based approaches appear to be less well suited. Approaches based on the rational design of small drug inhibitors including small oligomer-like scaffolds (biphenyls), the HTS of small drug molecules, and more recently modified peptides have until now achieved the most success.

α-Helix first principles

In the classical case, one turn of an α-helix is composed of 3.6 amino acid sub-units, with back-bone dihedral angles approximately $\Psi = -41^{\circ}$ and $\Phi = -62^{\circ}$.[8] This gives rise to distances of 1.5 Å per residue or 5.4 Å per α-helical turn. A signature aspect of the α-helical structure is the network of stabilizing hydrogen-bonding between the C=O oxygen of amino acid *i* and the amide NH proton of amino acid $i + 4$ (Fig. 1), which extends along the peptide back-bone. An important consequence of this well-ordered

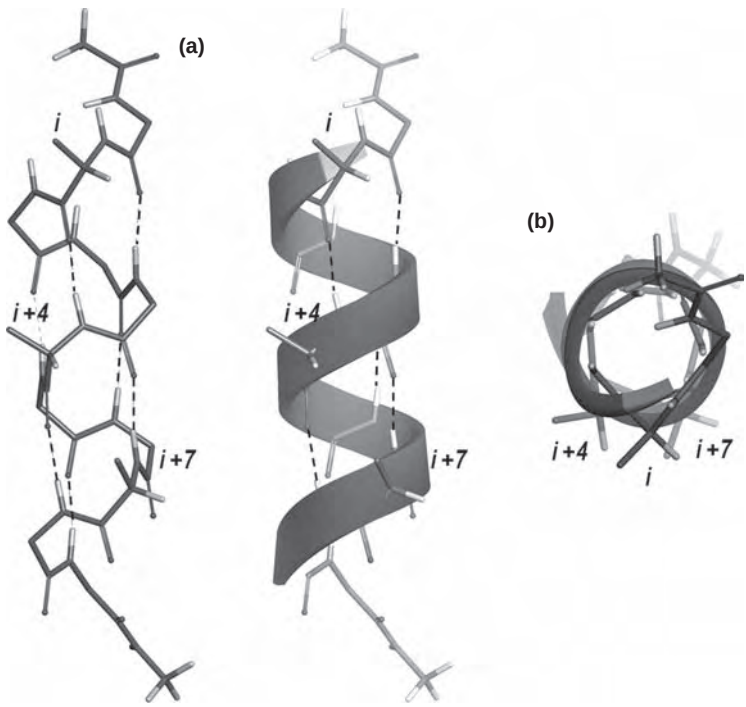


Fig. 1 Front stick and ribbon (a) and top ribbon (b) perspective of an idealized α-helix and the arrangement of relevant residues *i*, $i + 4$ and $i + 7$. A canonical α-helix has 3.6 amino acids per turn of the helix, which places the C=O group of amino acid *i* exactly in line with the H-N group of amino acid $i + 4$ (and C=O $i + 3$ with H-N $i + 7$).

structural arrangement is the precise three-dimensional orientation of the amino-acid side-chain residues.

In the broader sense, α-helices are critically involved in a number of important protein-protein interactions.[7] For example, the tumour suppressor p53 interacts with hDM2 at a hydrophobic groove *via* three hydrophobic residues – Phe19, Trp23 and Leu26, presented as a short α-helix interaction – which are important for the regulation of p53's anti-apoptotic activity.[12–14] The B-cell lymphoma 2 (Bcl-2)-protein family also regulates cell apoptosis *via* an α-helix-mediated PPI. The Bcl-X_L protein, in particular, recognises related pro-apoptotic proteins such as BAK, BIM and NOXA B, which are of different helix length, each possessing an array of hydrophobic residues (*e.g.* Val74-Leu78-Ile81-Ile85 in the case of BAK) as an integral part of the peptide sequence.[15] In another example, the C-terminal transactivation domain of hypoxia-inducible factor 1α or HIF-1α – a gene transcription factor – bears two short α-helical domains, which bind to coactivator protein p300 or the homologous CREB binding protein, CBP. These two interactions regulate the HIF-1α dependent expression of vascular endothelial growth factor (VEGF) and its receptor VEGFR2, which are engaged in vasculature formation (angiogenesis) in solid tumors.[16–18] This has rendered the HIF-1α/p300 or HIF-1α/CBP an important target for the inhibition of cancer progression and metastasis using α-helix mimetics.[19] Finally, the regulation of transcription factor β-catenin by B-cell CLL/lymphoma 9 – important for cell cycle regulation, apoptosis and the tumorigenesis of several types of human cancer[20] – is governed through the binding of an α-helical segment of BCL9 to a large binding groove present on β-catenin.[21] Down-regulation of β-catenin's transcriptional activity is possible through inhibition of the β-catenin/BCL9 interaction by metabolically stable peptidomimetics,[22,23] and thus represents a highly promising approach for the treatment of cancer in cases where β-catenin has become constitutively upregulated.

For the identification of novel inhibitors of PPIs, the high throughput screening of large compound libraries represents a viable approach;[24] α-helix-mediated PPIs included.[25] However, the predictable well-defined three-dimensional topology of the α-helix lends α-helix-mediated PPIs to a more rational approach to designing drug inhibitors, which makes use of molecular modelling and the natural peptide sequence as a lead structure. Inhibition of PPIs by small lipophilic drug molecules represents a challenging though highly viable strategy, with numerous benefits in contrast to peptide-based inhibitors, including good cell permeability and metabolic stability. In the case of α-helix mediated PPIs, a number of notable recent examples have been reported in the literature, which perfectly highlight the validity of the approach. For example, 1,2-benzodiazepine-2,5-diones (Fig. 2A, **1**) have been developed as selective inhibitors of the PPI between HMD2 and the tumour suppressor protein p53.[26] Here, the benzodiazepine scaffold precisely mimics the orientation of the α-helical Phe19 (*i*), Trp23 ($i+4$) and Leu26 ($i+7$) residues of the p53 protein. Inhibition of this important PPI will prove to be beneficial in determining its role in tumour growth, and may form the basis of future anti-cancer therapies. More recently, 4-benzoylaminobenzoic acids (Fig. 2A, **2**) were developed by Craik *et al.* to specifically target a dimeric protease essential for the human

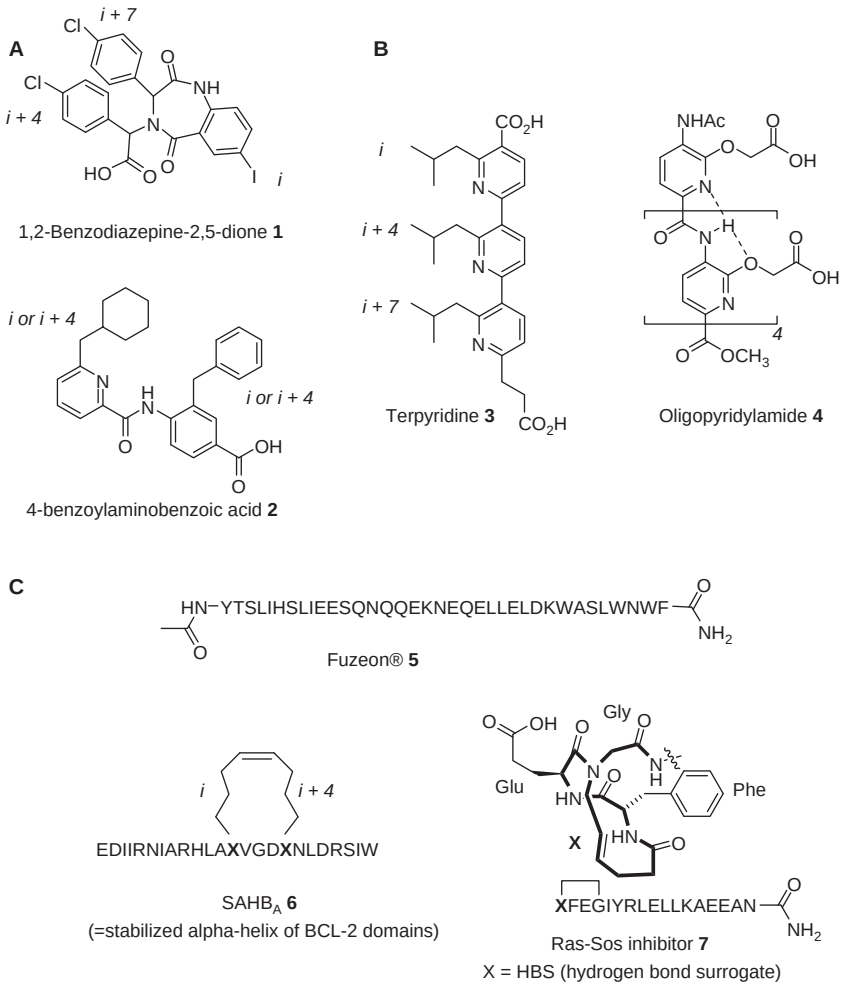


Fig. 2 Selected examples of inhibitors of α-helix mediated PPIs: **A.** Small lipophilic drug molecules: 1,2-bezodiazepine-2,5-dione **1** (HDM2-p53 inhibitor)[26] and 4-benzoylaminobenzoic acid **2** (dimeric protease essential for the human Kaposi's sarcoma-associated herpesvirus);[27] **B.** non-hydrogen bond mediated oligomer **3** (Bak/Bcl-x_L interaction)[28–33] and hydrogen bond-mediated oligomer **4** (islet amyloid polypeptide (IAPP) aggregation)[28,34] **C.** modified peptides: Fuzeon®, **5** (HIV-1 fusion),[53–55,57,58] stapled peptide **6** (BID/Bcl-x_L interaction) SAHB$_A$, (mimic of the BH3 domain of the BID pro-apoptotic protein)[35] and HBS-derived peptidomimetic **7** (Ras-Sos interaction).[36,37]

Kaposi's sarcoma-associated herpesvirus (KSHV).[27] Here, **2** inhibited protease dimerization at helix 5 by mimicking the interaction by the *i* and *i + 4* residues Met197 and Ile201 of one protein monomer with the Trp109 residue on the second.

Oligomeric scaffold structures in principle overcome many of the limitations inherent to lipophilic small molecules (Fig. 2B). Their extended iterative structures are specifically engineered with the multivalency of long α-helical structures in mind, while their rigid backbones – biaryl, hydrogen-bond mediated or otherwise – ensure a precise three-dimensional

orientation of the side-chain substituents in a manner reminiscent of the natural folded peptide sequence. In this sense, therefore, oligomeric scaffold structures could be considered more biologically relevant than small lipophilic drug molecules in their ability to more closely mimic especially extended α-helices for targeting the extended shallow surfaces common to many important protein-protein interactions.

In this regard, Hamilton and co-workers have succeeded in developing a number of biologically significant PPI inhibitors based on oligomeric scaffold structures. Examples include promising low-micromolar inhibitors of the cancer-inducing Bak BH3/Bcl-x_L interaction[38,39] and dual-serving modulators of islet amyloid polypeptide (IAPP) aggregation – a process important for type II diabetes –[8] based on terpyridyl and oligopyridylamide structures (Fig. 2B, **3** and **4**, respectively). Other related subclasses of oligomeric scaffold structures not discussed here include terphenyl,[40] biphenyl 4-4′-dicarboxamides,[41] enaminones[42] or examples where hydrogen-bonding has been integrated within the scaffold scaffold such as benzoylureas.[43] Much like the oligomeric scaffold structures cited above, other foldamers aim for greater biological relevance by assuming the molecular characteristics typically associated with peptide-based α-helical structures.[44–51]

Although natural peptides do not typically make good drugs, minor chemical modifications have been shown to dramatically improve the biological properties of a natural α-helical peptide lead compound. For example, the highly efficacious anti-HIV drug Fuzeon® (enfuvirtude),[52–55] is an improvement on the natural peptide (HIV-1 gp41 (643-678) through simple primary amidation and acetylation of the C- and N-termini, respectively (Fig. 2C, **5**). The result is a more stable α-helical structure through moderate attenuation of fixed charge/helix macrodipole interactions. In addition, more potent truncated variants of Fuzeon® have also been developed, bearing the same capped termini, but with added side-chain diamide linkages between glutamic acid residues at the i and $i+7$ positions on the same face of the α-helix.[56] Aside from the bis amide macrocycle, disulfide-bridging has also been used to improve the stability of the gp41 C-peptide sequence using D- and L-cysteines located at the i and $i+3$ positions.[57] One of the most significant recent advances in the field of peptide based inhibitors of α-helix mediated PPIs has been the ability to reengineer peptides with improved cell permeability, increased metabolic stability, and high affinity binding. Verdine and co-workers were one of the first to demonstrate this in their work on the development of $SAHB_A$ or 'stabilized alpha-helix of Bcl-2 domain, analogue A′ (Fig. 2C, **6**) targeting the BID BH3/Bcl-x_L interaction.[35] In this case, the introduction of a stable artificial hydrocarbon staple *via* ruthenium catalyzed ring-closing metathesis increased α-helicity, leading to enhanced protease resistance and serum stability, as well as improved cell permeability and enhanced Bcl-2 binding affinity. In a similar way, Arora *et al.* also used chemical modification to improve the properties of a natural peptide inhibitor using a 'hydrogen-bond surrogate-based' (HBS) approach (Fig. 2C, **7**) to mimic the guanine nucleotide exchange factor Sos and its interaction with the cell signal transducer, Ras.[37] Interference of this interaction causes a down-regulation of Ras signalling in response to receptor tyrosine kinase activation, leading to

inhibition of cancer cell growth, differentiation and survival. In this case, rather than introducing the staple at a certain point along the peptide chain, as in **6**, the N-terminal *i* residue was instead covalently linked to the $i + 4$ amino acid *via* a hydrocarbon chain – using a ruthenium catalysed metathesis strategy – in order to mimic the N-terminal *i* and $i + 4$ hydrogen bond (C=O---N-H). This modification serves to nucleate and stabilize the α-helix, thereby reducing the entropic penalty for, and thus favouring, binding to the Ras protein surface, as well as improving on cell permability and the metabolic stability of the peptide inhibitor. There are numerous other examples in the literature of similar side-chain cyclization strategies to improve the properties of peptide inhibitors, which could not be covered in this chapter due to space restrictions.[58–65]

2 Nuclear receptors

Nuclear Receptors (NRs) are a superfamily of 48 transcription factors that play widespread and important roles in the regulation of numerous biological processes, including development, reproduction, metabolism, homeostasis and disease.[66–68] Indeed, nuclear receptors can be considered as quintessential α-helical proteins, given their α-helix rich structures – especially in the ligand binding domain – which are fundamental to the mechanisms of the gene transcription machinery.

Historically, the nuclear receptor superfamily has been classified into different subfamilies based on their dimerization and DNA binding properties (Fig. 3).[69] One group consists of the steroid hormone receptors (SR), which function as ligand induced homodimers that bind to DNA half-sites organized as inverted repeats. These receptors include the estrogen receptor (ER), the androgen receptor (AR), the progesterone receptor (PR), the glucocorticoid receptor (GR) and the mineralocorticoid receptor (MR). A second group consists of receptors that form heterodimers with the retinoid X receptor (RXR), which characteristically bind to direct repeats (although some bind to symmetrical repeats as well). With the exception of steroid receptors, this group includes all other known ligand dependent receptors, such as the vitamin D receptor (VDR), the retinoic acid receptors (RARs), the thyroid hormone receptors (TRs) and the peroxisome proliferator activated receptors (PPARs). Some NRs, such as the orphan receptors, can also bind DNA efficiently as monomers while others bind primarily to direct repeats as homodimers.

The classical model of ligand-activated NR signalling involves activation of the receptors through the binding of natural ligands, which enables NRs to regulate the transcriptional activity of target genes through interaction with specific DNA sequences in the promoter and recruitment of specific coregulators.[70] For their effective function, therefore, NR proteins must be able to interact with a number of coregulatory proteins, which function either by enhancing (co-activators) or repressing (co-repressors) gene activation. In addition, to achieve transcriptional activation of target genes, there is a plethora of other factors involved in NR-associated signal transduction, such as dimerization partners, promoter regions, chaperones, ubiquitin ligases, kinases, phosphatases, and others.[71]

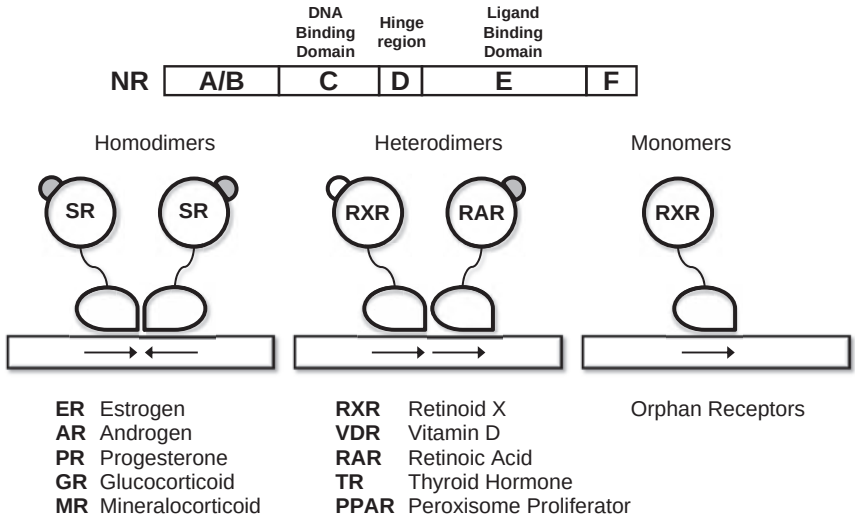


Fig. 3 *Top*. Schematic representation of the characteristic structural organization of NRs, consisting of a highly variable N-terminal region (A/B), a central DNA binding domain (C), a short hinge region (D), the ligand binding domain (E), and the C-terminal F region, which is not present in all NRs. *Bottom*. Schematic illustration of examples of NR subfamilies based on their dimerization and DNA binding properties. *From left to right*, homodimeric receptors, RXR heterodimers and monomeric orphan receptors. Arrows indicate the consensus NR recognition sequence on the double stranded DNA (*e.g.* direct repeats of AGGTCA for non-steroid receptors and the majority of orphan receptors, while steroid receptors such as ER bind as homodimers to inverted repeats of AGGTCA.) SR = steroid receptor.

Despite the functional diversity exhibited by this class of transcription factors, they share a remarkable structural and functional similarity.[72,73] Most NR exhibit the characteristic modular organization of eukaryotic NR transcription factors consisting of a highly variable N-terminal region, a central DNA binding domain, a short-hinge region, and the ligand binding domain, which is a C-terminal domain containing the hormone binding pocket (Fig. 3). Although there is undoubtedly interplay between the domains, each domain has a separate crucial signal transduction function. The N-terminal region (A/B domain) contains the ligand-independent activation function (AF-1) and is involved in potentially multiple protein-protein interactions.[74] The most conserved domains of this protein family are the DNA binding domain (DBD), localized in the middle of the primary structure, which mediates specific interaction with the target DNA sequences, and the ligand binding domain (LBD), which is responsible for ligand recognition, dimerization, interaction with co-activators and ligand-dependent transcriptional activation. These different domains are linked by a less conserved region (D domain) that behaves as a flexible hinge between both domains, and contains a nuclear localization signal (NLS). Furthermore, the C-terminal ligand binding domain contains a ligand-dependent transcriptional activation function (AF-2). The C-terminal F region, which is contiguous with the E domain, is not present in all receptors, and its function is more poorly understood.

Over the past years, structural and biochemical studies of a number of NR have provided a wealth of information on ligand recognition, receptor

activation, and recruitment of coregulators. The resolution of the crystal structures of several ligand-free (apo) and ligand-occupied (holo) NR-LBD alone or in a complex with coactivators have allowed for a much better understanding of the molecular details involved in these ligand-activated transcription factors.[75–80] It is remarkable to note, in this context, the first crystal structure of the fully-intact PPARγ-RXRα dimer complexed with DNA[76] and the first cryo-EM structure of the full RXR-VDR heterodimer complex with its DNA response element recently solved.[80] In this chapter emphasis will be placed on the biological function and structural features of the estrogen receptor and androgen receptor, as prototypical models for helix-mediated PPIs in transcription and a discussion about new approaches to target these interactions.

ER and AR

At the cellular level, and in the absence of an agonistic ligand, nuclear receptors repress transcription through a dominant association with corepressor complexes. In the unliganded state some NRs, such as AR, are held in the cytosol of target cells, where the LBD is often associated with heat shock chaperone proteins forming an inactive complex.[81] Upon ligand binding, such as the natural hormone estradiol for ER or testosterone for AR, a significant conformational change within the LBD occurs that releases it from the inactive complex. This results in NR dimerization, phosphorylation, translocation to the nucleus and ultimately binding to NR response elements within the regulatory regions of target genes. These NR response elements have a general half-site consensus of RGGTCA (for which R = purine). Once bound to DNA, NRs act by recruiting general transcription factors or intermediary factors (coactivators). Binding of a coactivator protein is believed to be one of the key events in initiating transcriptome assembly and subsequent gene transcription.

To date, the two known subtypes of ER – ERα and ERβ – have afforded a great variety of crystal structures with bound natural and synthetic ligands, making a complete dissection of the mechanisms of action of these ligands possible.[82–86] In the late 1990s, several X-ray crystallographic studies were the first to shed light on how ligands alter NR-LBD conformation, furthermore revealing the structural determinants of agonist and antagonist recognition. The first structure of a NR LBD was described for the unliganded retinoid X receptor α subtype (RXRα).[75] Brzozowski *et al.* determined the structure of human ERα-LBD in complex with its endogenous hormone, 17β-estradiol (E2) and with the selective antagonist raloxifene (RAL).[77] Shortly after, the 3D structure of ERα-LBD bound to the agonist diethylstilbestrol (DES) and to the selective antagonist 4-hydroxytamoxifen (OHT) were solved.[85] In addition, the structure of the ligand-binding domain of ERβ in the presence of a partial agonist genistein (GEN) and a full antagonist raloxifen (RAL) was later reported.[87] These and more recent structural studies have successfully delineated the structural basis of NR-ligand and -coactivator interactions.[82]

Helix 12: the α-helical molecular switch

Based on the determined structures of several liganded NR-LBD, the overall fold for the LBD has proven to be prototypic for many NR. They

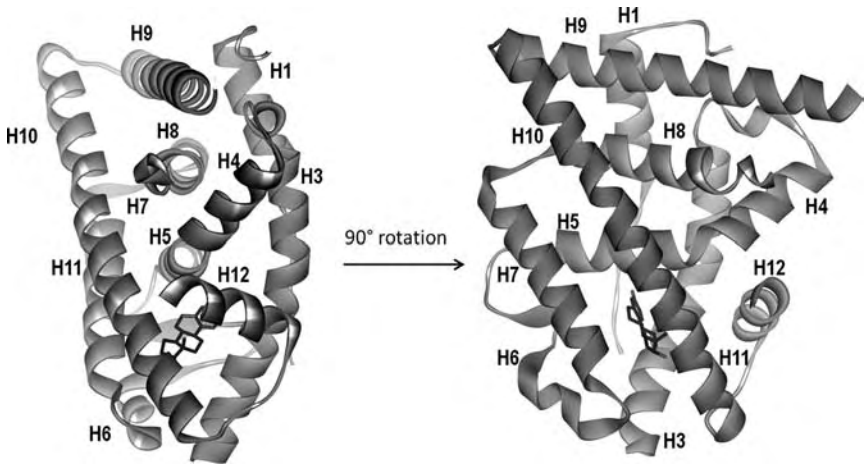


Fig. 4 Crystallographic structure of the ER-LBD monomer complexed to 17β-estradiol (PDB accession code: 1ere).[77] The ligand, shown with sticks, is completely embedded in the binding pocket. Left. Front view. Right. Side view. The LBDs exhibit a similar overall topology evident for other NR LBDs.

comprise 12 α-helices (numbered H1-H12) for ER and RXR and 11 α-helices for AR (H2 is not present in AR LBD) and two short β-turn, arranged in three anti-parallel layers that form an "α-helical sandwich" (Fig. 4). Helix 12 (H12), often referred to as the AF-2 core, is a conserved structural element essential for ligand-dependent transcriptional activation. These NR LBDs contain three structurally distinct surfaces: firstly, a dimerization surface, mediated primarily by the N-terminus of H11 and helices H9 and H10, all of which contribute to the specificity of DNA binding by correctly spacing and orienting the DBD subunits. Secondly, the ligand-binding pocket (LBP), which occupies a relatively large portion of the hydrophobic core in the lower half of the domain, and interacts with diverse lipophilic small molecules, and finally the activation function helix AF-2 (including the H12, see above), which binds to regulatory protein complexes and modulates ligand-dependent transcriptional activity to a positive or negative extent.

Similar to other steroid receptors, ER isoforms and AR share the same highly conserved functional domains and show some similarity in their amino acid sequence. The DBD domains of ERα and ERβ are highly convergent (95% similarity) whereas the LBD is approximately 55% homologous. However, while the central AR-DBD shares greater than 50% sequence identity with ER-DBD, the LBD is much less conserved (25%) also compared to other steroid receptors (22–55%). Furthermore, the volume of the LBP varies slightly among the different receptors: approximately 420 $Å^3$ for AR and 450 $Å^3$ for ERα, compared with a LBP volume of $\sim$1300 $Å^3$ measured for the peroxisome proliferator-activated receptors (PPARs). Therefore, the specificity of NR ligand-binding depends on the size and the shape of this LBP and the amino acid residues that are in proximity to the ligand forming direct interactions.

A comparison between the liganded and unliganded structures of RXR revealed striking differences in the repositioning of the C-terminal helices 11

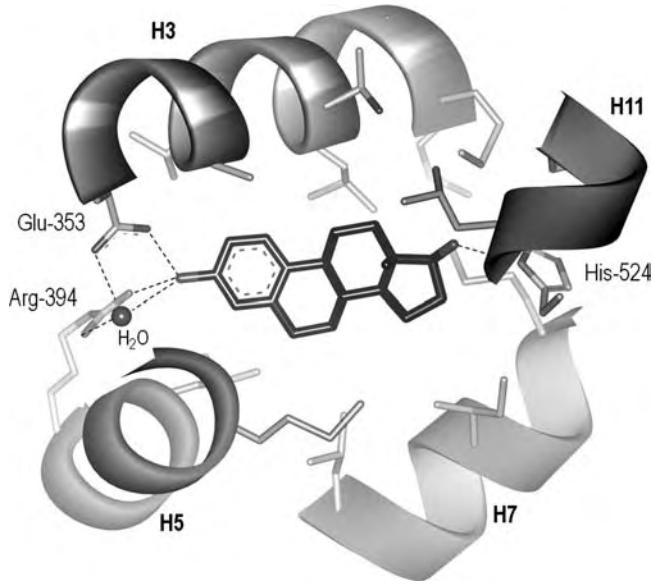


Fig. 5 Schematic representation of 17β-estradiol and its interactions with the hERα-LBD ligand binding pocket. (PDB accession code: 1ere).[77] Residues within the binding cavity that make significant contacts with the ligand are shown. Intramolecular hydrogen bonds are displayed as dashed lines and amino acid residues are numbered according to the hER crystal structure template.

and 12.[75,88–90] In the apo-form, H12 extends away from the LBD, while in the holo-LBD H12 folds back over the LBP, with the ligand entirely buried within the LBP. Agonists and antagonists are typically hydrophobic molecules, which bind at the same site at the LBP within the core of the LBD (Fig. 5). Again the LBD H12 is the crucial structural element in this case, which is capable of adopting multiple positions depending on the nature of the bound ligand. Agonistic ligand binding triggers a conformational change that results in the repositioning of H12 across the ligand binding pocket on the core of the LBD, commonly referred to as the 'mouse trap' mechanism.[72] Together with H3 and H4, H12 forms a favourable shallow, hydrophobic groove surrounded by charged amino acids, which completely buries the ligand within the core of the LBD (Fig. 5). This reorganization of the AF-2 helix creates a new platform for the recognition of coactivator proteins (Fig. 6). These coactivator proteins bind to the LBD *via* a leucine enriched α-helical LXXLL motif (where L is leucine and X any other amino acid), termed 'NR boxes'.[85,91–93] In effect, deletion and point mutation studies have shown that the AF-2 activation helix is absolutely essential for the formation of the coactivator binding groove and a necessity for ligand induced transcriptional activation.[94]

In contrast to agonist ligands, steroid antagonists induce an alternate conformation in AF-2 that occludes the coactivator binding site and recruits corepressors that can actively silence steroid responsive genes (Fig. 6).[72] Contrary what might be expected, the activation of a NR by these small molecule ligands is much more complex than a two-state (on and off)

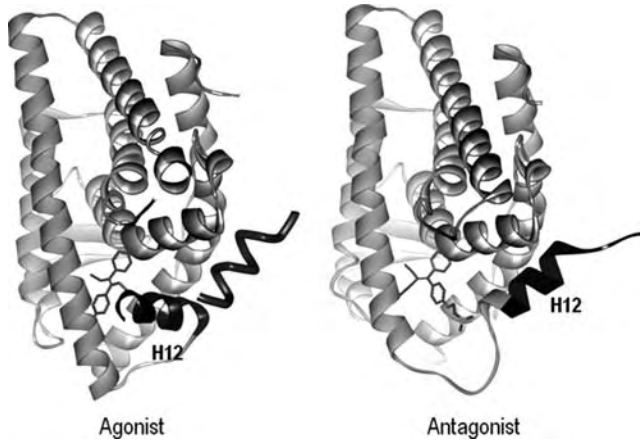


Fig. 6 Agonist and antagonist conformations of ERα ligand binding domain. Agonistic ligand binding triggers a conformational change that results in the repositioning of H12 across the ligand binding pocket on the core of the LBD and the binding of coactivator proteins (coactivator peptide highlighted in tube format). Antagonists mimic the agonist-binding mode, but induce an alternate conformation, which prevents H12 from capping the pocket, thereby blocking the binding of coactivator proteins. Highlighted in black is the displaced H12. The models were rendered from Brookhaven, PDB accession codes: 1ere (ER-LBD/17β-estradiol/Nuclear Receptor Box II),[77] and 3ert (ER-LBD/tamoxifen).[85]

process.[95] Control over the NR conformation can thus be exerted *via* the ligand, but interdomain interactions, for example *via* the DNA binding domain, also play an important role in this respect.

H12 Dynamics

Crystal structures of NR LBDs complexed with various ligands, both agonist and antagonist, have shown that the LBD-H12 can adopt an array of alternate positions (Fig. 6). However, little is known about the dynamics of H12 in the LBD bound ligand state, less about the organization and dynamics of H12 within the ER and AR LBDs in the absence of ligand, where H12 seems to be highly mobile and exposed to the solvent.

Despite the structural information available, little is not much known about whether and to what extent H12 dynamics influences and maybe even regulates cofactor binding. Recent Hydrogen/Deuterium Exchange (HDX) mass spectrometry experiments have revealed that ligand binding does not change H12 dynamics for ERα and ERβ in the case of agonist, antagonist, or selective ER modulator (SERM) ligands.[96,97] In contrast, the opposite behaviour has been reported for other HDX studies based on the thyroid receptor (TR) conformation, showing that H12 remains relatively protected from solvent in the apo-form, as well as in the presence of agonist or antagonist ligands.[98] What seems apparent though is that the ligand-induced conformation of a NR ligand-binding domain – in particular the H12 of the LBD – plays a critical role in the activation mechanism of transcription and in determining the pharmacological response. Importantly, post-translational modifications (PTMs), such as phosphorylation, can also finely regulate the molecular interactions of transcription factors with coactivators and ultimately regulate the overall transcriptional activity.[99–106]

Coregulator binding

Nuclear Receptors use multiple mechanisms to activate gene transcription. This transcriptional activity can be achieved by at least two activation domains: the ligand-independent activation function AF-1, located in the N-terminal region and the ligand-dependent AF-2, located at the C-terminus of the LBD. Both AF-1 and AF-2 regions of NRs can activate one another in a promoter- and cell-context specific manner by interacting with a large plethora of transcription coactivators. The AF-1 domain is the major transactivation domain in the case of AR.

In effect, recruitment of transcriptional coactivators following ligand activation is the second crucial step in NR-mediated target gene expression. Coactivator recruitment is mediated *via* a PPI between the liganded NR surface at the LBD and at least one α-helical motif on the coactivator protein, which presents mostly an LXXLL motif within the polypeptide chain.[107] This short signature LXXLL motif (NR box) is necessary and sufficient to permit the interaction between receptors and coactivators.[91] Since the first coactivator – the steroid receptor coactivator-1 (SRC-1) – was identified in the 1990s, more than 200 known proteins, heterogeneously expressed in the different cell and tissue types, have been described to interact with NRs. The most common coactivators are the three widely expressed p160 family members: SRC-1, GRIP1/TIF/SRC-2 and pCIP/ AIB1/ SRC-3.[108]

In addition, intra-receptor N-terminal/C-terminal interactions between the N-terminal domain and the LBD have been shown to be necessary for effective ER and AR transactivation.[109–112] In these interactions, the AR LBD preferentially binds FXXLF motifs present in the AR N-terminal domain (AF-1 domain).

AF-1 domain

AF-1 domain is the least conserved region among NRs, and in contrast to AF-2, AF-1 shows a weak propensity for intrinsic structure formation.[113] Although this lack of structure limits the available information on structure-function information, it has been shown that the structural plasticity of the AF-1 domain is important for receptor function.[114] The AF-1 domain is the major transactivation domain in the case of AR, being the AR N-terminal domain (NTD) and the main recruiting surface for coactivators.[115,116] Dependent on androgen binding for activation, FXXLF motifs present in the AR NTD also bind the AR-LBD through the so-called NH2-/ COOH-terminal domain interaction (N/C interaction), which makes AR unique among NRs. The NH2-terminal region involved in this interaction with the LBD corresponds to the first 36 amino acids within the NTD, which contain the ^{23}FQNLF^{27} motif. Also a second motif in the NTD with the sequence ^{433}WHTLF^{437} has been mapped to bind to a different region of the LBD.[136] AR LBD also binds AR-specific cofactors and some LXXLL motifs of nuclear receptor co-activators.[117]

AF-2 domain

Within the AF-2, the integrity of a conserved amphiphatic α-helix is necessary for ligand-dependent transactivation and coactivator recruitment.

Structural information obtained from X-ray crystallography has provided evidence that coactivator LXXLL motifs adopt an α-helical conformation that binds to the AF-2 surface of the liganded LBD. Coactivators bind to a groove on the NR LBD surface formed by charged and hydrophobic amino acids in helices 3, 4, 5 and 12. Only on proper ligand-dependent repositioning, is H12 able to complete the formation of the recognition surface for the binding of the NR box to the LBD (Fig. 6). Then, coactivator peptides adopt an α-helical conformation where the three conserved leucine side chains in positions *i*, *i* + *3*, and *i* + *4* are oriented along one face of the α-helix, projected into a hydrophobic area on the surface of the NR. In contrast, the XX residues at positions + 2 and + 3 (relative to the first L) are solvent exposed, oriented away from the groove on the opposite half of the amphiphilic helix (Fig. 7).

The features of the LBD interface that interact with coactivators are similar among members within the NR family. The NR coregulator interface is an L-shaped hydrophobic cleft comprised of three distinct subsites

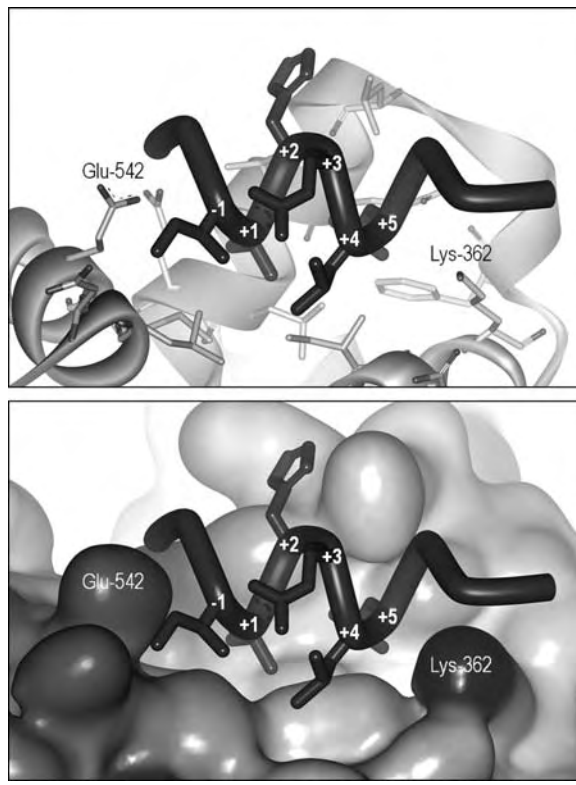


Fig. 7 Binding of the Nuclear Receptor Box II LXXLL motif to hERα; PDB accession code: 1ere (ER-LBD/17β-estradiol/Nuclear Receptor Box II).[77] Close-up of the Nuclear Receptor Box II LXXLL motif (tube) on the surface of the ER-LBD coactivator binding site, where key residues participating in these PPIs are shown. The Nuclear Receptor box II peptide backbone is represented as a tube with the leucine side-chains of the LXXLL motif shown (at positions + 1, + 4 and + 5 relative to the first L), oriented on one face of the α-helix, projected into the hydrophobic area on the surface of the ER. While the two XX residues at positions + 2 and + 3 are solvent exposed, oriented away from the groove on the opposite half of the amphiphilic helix. The two ERα residues K362 and E542, which define the 'charge clamp', are also depicted.

that bind to hydrophobic groups at the +1, +4, and +5 positions in cognate peptides. These protein-protein interactions are mediated by hydrophobic interactions with the leucine side chains, but also by hydrogen bonding between conserved charged amino acid residues. Particularly important in this case is a conserved lysine residue in H3 and a highly conserved glutamate residue in H12 of the LBD of NR, which form a 'charge clamp' that stabilizes the peptide LXXLL helix *via* hydrogen bonding and packing of side chain residues into the hydrophobic cavity (Fig. 7). Crucially, the 'charge clamp' governs the NR-coactivator interaction, making this system quite unique amongst almost all other α-helix mediated PPIs. In effect, point mutation studies have verified the importance of this charge clamp in the functional interaction between NRs and SRC NR boxes.[118] It is also believed that the 'charge clamp' determines the precise length of the helical motif for optimal docking at the surface cleft. Therefore, the distance between these two opposite charges within the LBD and the specific length of the helix formed by the LXXLL motif are very important for proper coactivator recognition and binding.

Several works have demonstrated that an efficient interaction not only depends on the LXXLL motif but also on sequences in the flanking regions, which may make differential contacts with helices 3, 4 and 5 in the receptor's ligand binding domain. The role of flanking sequences in influencing these PPIs has been evaluated by different biochemical techniques, suggesting that non-conserved residues adjacent to the LXXLL motif of coactivators make a significant contribution to the selectivity and affinity of a specific NR.[119–121] Additional charged residues near AF-2 and flanking the LXXLL motif regions also appear to be significant in this case.[122] Some of these studies will be dissected in the following section of this chapter, focusing mainly on the LXXLL and FXXLf motifs.

LXXLL motif

A phage display technique has been used to investigate the interaction of coregulators with different liganded NRs.[119,123–126] This technique has emerged as a powerful method for isolating short peptides with characteristic sequence motifs that mimic the surfaces of these protein-protein interactions (PPI). Based on random and focused peptide library screening (containing the core LXXLL motif), different classes of peptides have been found to display specific receptor preferences. Most of the AF-2 interacting coactivators found using random phage display contain an LXXLL motif. While the two internal XX residues flanked by leucines within the NR core do not appear to be important for receptor binding,[91,92] residues in the −1 and −2 positions with respect to the LXXLL motif have been shown to govern receptor-binding affinity and specificity.[119,123,127–130] Generally, positively charged residues are favoured at position −1, −2, or −3 relative to the LXXLL motif, while negatively charged residues are found at position +8 or +9. Neutral or positively charged residues occur at positions +2 and +3 within the core motifs, with neutral or negatively charged residues at +6 and +7.[122]

In the case of ERα-interacting peptides, some of the peptides found by phage display could be grouped into three classes based upon sequence

conservation of the two amino acids immediately prior to on the N-terminal side of the LXXLL motif.[119] Class I peptides contain a conserved serine residue (S) at the −2 position and a positively charged arginine residue (R) at the −1 position. Class II peptides have a proline residue (P) occupying the −2 position and a hydrophobic leucine residue (L) directly preceding the LXXLL motif. Whereas, class III peptides share a conserved serine (S) or threonine residue (T) at the −2 position followed by a hydrophobic leucine (L) or isoleucine residue (I) at the −1 position (Table 1). All three classes are found in known ERα-interacting cofactors. Additionally from peptide phage display screening, the peroxisome proliferator-activated receptor, subtype γ1 (PPARγ1) was discovered to bind preferentially to peptides with a very strong consensus for HPLLXXLL,[126] resembling somewhat the class II peptides mentioned above. Furthermore, a proline residue at the −2 position appeared to be favoured for other receptors such as TR, VDR, RXR and ERβ.[119] Moreover, phage display and other techniques revealed differences in LXXLL binding between ERα and ERβ. For instance, the addition of an isoleucine residue (Ile) at the −1 position appeared to favour greater selectivity for ERβ over ERα.[131] Some phage peptides have also been reported to selectively disrupt ERβ but not ERα-mediated reporter gene expression.[119] Selectivity for ERα over ERβ has also been reported in the case where random mutagenesis analysis identified S884 of coactivator TRBP at −3 position as a key residue for receptor selectivity for TRβ, RXRα, and ERβ versus ERα.[132] Here, a S884Y mutant showed a significant increase in binding to TRβ, RXRα, and ERβ, but decreased binding to ERα, which could be explained by an interaction between the S884 residue of TRBP and the E380 residue on ERα, which was

Table 1 Sequences of LXXLL and FXXLF motif containing peptides discovered by phage display screening. LXXLL-containing sequences derived from the human nuclear receptor coactivator proteins SRC-1 are also shown.

Source		*Sequence* +1 +4 +5
SRC-1	NR1	YSQTSHKLVKLLTTTAEQQ
	NR2	LTARHKILHRLLQEGSPSD
	NR3	ESKDHQLLRYLLDKDEKDL
ER interacting[119]		
ERα Class I	D2	GSEPKSRLLELLSAPVTDV
	D11	VESGSSRLMQLLMANDLLT
	D30	HPTHSSRLWELLMEATPTM
ERα Class II	D14	QEAHGPLLWNLLSRSDTDW
	D47	HVYQHPLLLSLLSSEHESG
	C33	HVEMHPLLMGLLMESQWGA
ERβ selective	Peptide-293	SSIKDFPNLISLLSR
AR interacting[140]		
	FLET1	SSRFESLFAGEKESR
	FLET2	SSKFAALWDPPKLSR
	FLET3	SRWQALFDDGTDTSR
	FLET4	SRWAEVWDDNSKVSR
	FLET5	SSNTPRFKEYFMQSR
	FLET6	SRFADFFRNEGLSGSR
	FLET7	SSRGLLWDLLTKDSR

not witnessed in the case of the other NRs. The AR AF-2 domain can also bind a set of LXXLL motif peptides.[133,134] Unlike other NRs, such as ER, however, this interaction is much weaker for AR.[135] The AR LBD preferentially binds FXXLF motifs present in the AR N-terminal domain (*vide supra*) and other AR-specific cofactors.[136–138]

FXXLF motif and other F-rich motifs

Phage display screening has shown that phenylalanine-rich motifs preferentially bind to the AR LBD.[130,139,140] Furthermore, X-ray crystallography has provided detailed structural information on the AR LBD coactivator-binding groove and coactivator interactions, where both the AR-NTD and AR cofactors, containing either an FXXLF or LXXLL motif, form part of an amphiphatic α-helix that interacts with AF-2 in the LBD (Fig. 8).[136,141,143] Like other NRs, the charge clamp, consisting of a glutamic acid residue at position 897 in the H12 region and a lysine residue at 720 in the H3, contribute largely to the stabilization of these PPIs.

Fletterick and co-workers have shown how the AR AF-2 surface recognizes coactivator peptides with varying hydrophobic sequences containing either LXXLL or F/WXXLF/W/Y motifs. X-ray crystallography studies

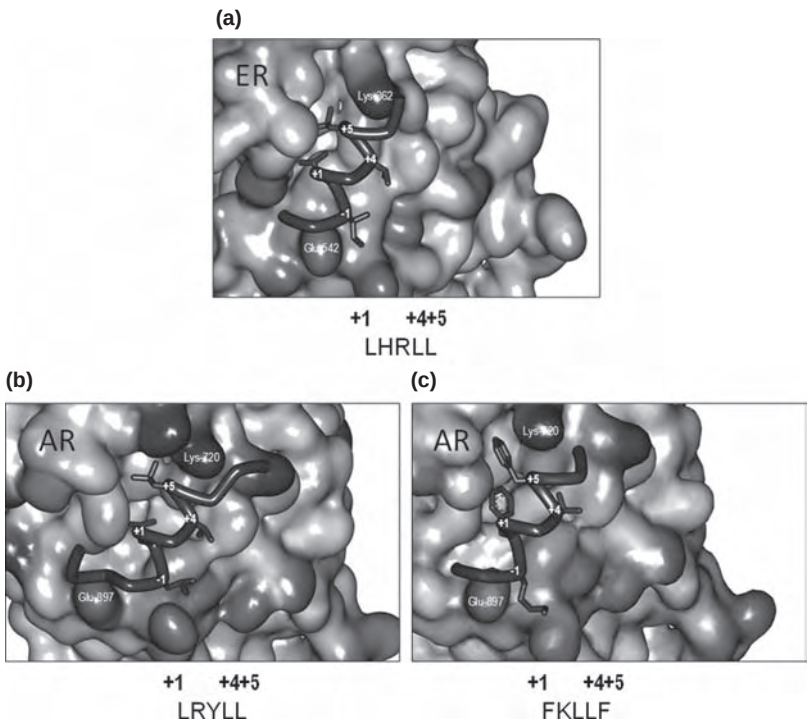


Fig. 8 Crystal structures of the ERα and AR LBDs bound to LXXLL and FXXLF-containing coactivator peptide sequences. a) ERα co-crystallized with an agonist and an SRC2 NR box II peptide (PDB accession code: 3erd).[85] b) Agonist-bound AR co-crystallized with an SRC2 NR box III peptide (PDB accession code: 1t63).[143] c) Agonist-bound AR co-crystallized with DHT and a peptide derived from its physiological coactivator ARA70 (PDB accession code: 1t5z).[143]

have shown that the AR AF-2 cleft is unique in its ability to reorganize and adapt to the binding of different coactivator sequence motifs.[140] As previously reported,[136,137] it could be shown that LXXLL motif peptides make fewer hydrophobic contacts with the surface of the LBD, and AR-LBD preferentially binds to the AR-NTD and coactivators with bulky hydrophobic residues, allowing for a better fit at the surface pocket of the AF-2 domain (Fig. 8).[140,141] AR favours F residues at +1 and +5, and prefers E at +2 and positively charged K or R at +3, what results in a stronger interaction and improved AR specificity.[142]

Other motifs

An antagonist peptide not containing the LXXLL consensus motif, which bound specifically to ERα in a ligand independent manner, was also isolated from a random peptide phage display screen.[119,124] Subsequent X-ray crystallography studies showed that this peptide, through binding to ERα-LBD in a unusual extended conformation, recognizes a different site from the classic coregulator recruitment (AF-2) region.[144] Importantly, this unique interaction surface, located in a different region of the LBD, is not affected by the different conformational effects induced by receptor agonist and antagonists.

Understanding the recognition principles underlying the α-helix mediated NR-coactivator interaction will surely help in the rational design of novel peptidomimetics blocking receptor-mediated gene transcription.

3 α-Helix-inspired drug discovery

One of the main driving forces behind the development of small molecule mimics of α-helical structures has been the need for novel chemotherapies, in particular to overcome drug resistance, for the treatment of cancer and metabolic diseases. Given that peptide α-helices represent the most abundant protein secondary structures, the temporal modulation of α-helix-mediated PPIs by small drug molecules would represent a viable therapeutic strategy. In the previous section, nuclear receptors (NRs) were discussed as quintessential α-helix based proteins, and some of the key structural parameters which determine α-helix protein-protein interactions were identified, in particular between NRs and coactivator proteins. This section will highlight efforts to selectively modulate the NR-coactivator interaction, starting with the development of small drug-like lipophilic inhibitors, focusing in particular on de novo design and high throughput screening approaches. Next, the discussion will shift to more α-helix-like mimetics, such as oligomeric scaffold structures and then finish by highlighting how the research field has turned full circle, back to the natural co-activator peptide sequence and its consideration as a lead structure. Here, the focus will be on methods for improving on contemporary weaknesses of peptide drugs, such as low proteolytic stability, and poor cell permeability.

The high number of co-crystal structures of different NR ligand binding domains (LBDs) complexed with different ligands have provided a clear picture of some of the mechanisms involved in NR ligand binding and transactivation. For instance, the conformational change in the NR brought about by

ligand binding and the consequent array of protein–protein interactions play a crucial role in triggering and driving the multitude of signal transduction events. Therefore, with the increasing structural knowledge on ligand-driven NR conformation, a large spectrum of synthetic ligands has been developed to selectively modulate steroid NR action. Ligands with full, partial agonist or antagonist activities have been developed, as well as compounds that selectively modulate different receptor subtypes and act in a cell-selective manner. However, the development of resistance to current endocrine therapy and the serious side effects of many of these drugs limit their utility and safety.[145] Therefore, there is a significant medical need for new and improved nuclear receptor drugs in order to overcome these limitations.

Non peptide-based small molecule inhibitors

The motivation for targeting non peptide-based α-helical mimics is that they in principle marry the high target selectivity of natural peptide sequences with the high bioavailability typically associated with small lipophilic drug molecules. However, mimicking the rigidity and optimal arrangement of side chain functionality associated with a lead peptide sequence can be an arduous process.[146–148] To overcome this, high-throughput screening (HTS) technologies have been used to screen large compound libraries (typically $>10{,}000$) in search of novel lead compounds.[149,150] Alternatively, attempts have been made to target some of the key structural features of the α-helical LXXLL peptide motif in search of novel scaffold structures, which can then be rationally optimized during subsequent iterations of synthesis and biological testing.[151] One of the principle advantages of non peptide-based structures is that by not being limited to the side-chain diversity of natural amino acids, as is the case of natural peptide sequences, there is greater scope for structural variation based around a flexible core scaffold. The scaffold is then typically modified in a modular fashion (in as few steps as possible) towards greater side-chain variation, thus facilitating the fine-tuning of target affinity and selectivity.

Small lipophilic drug-like molecules

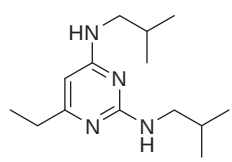
In contrast to other types of α-helical mimetics, small lipophilic drug-like molecules tend to be smaller ($MW < 500$ Da) and structurally more 'compact', which means that they typically cover less of the protein surface, especially compared to the natural peptide sequence. Small drug-like molecules are designed to be more lipophilic than peptides (LogP values ranging from 2 to 5), in an effort to improve bioavailability and increase their potential druggability.[152]

However, given the small, compact nature of small drug molecules relative to the large shallow surface of the target protein-protein interaction, it has proved to be very challenging to design novel scaffold structures. Nevertheless, one successful approach has been to rationally target some of the key structural features of the coactivator α-helical LXXLL motif – quite literally to outfit a rigid core-structure with lipohilic side chains reminiscent of those found in the LXXLL motif – followed by a rational optimization of target properties through subsequent iterations of synthesis and biological testing. Indeed, one of the advantages of using small lipophilic drug

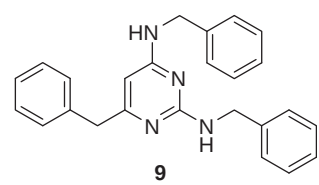
molecules in contrast to peptide-based structures is the unending scope for structural variation based around a flexible core scaffold, which can be modified in a modular fashion using a medicinal chemistry approach.

De novo design: pyrimidine scaffolds targeting ER and AR. One of the first examples of this approach was reported by Katzenellenbogen and co-workers, initially targeting the ER-coactivator interaction using a pyrimidine-based core (Fig. 9). More specifically, 2,4,6-trisubstituted pyrimidines were synthesized bearing small linear or branched alkyl substiuents, which served to mimic the *i*, *i* + *3*, and *i* + *4* arrangement of the leucine side-chains of the LXXLL motif.[151,153,154] In the first instance, low micromolar levels of inhibition was achieved, with selectivity for ER over AR, using isobutyl substituent groups (Fig. 9, **8**).[155] The same scaffold structure was successfully reengineered to selectively inhibit the AR-coactivator interaction over ER by replacing the isobutyl side-chains with benzylic groups (Fig. 9, **9**).[155] This modification was motivated by an awareness that AR preferentially binds to coregulator proteins and peptides presenting tryptophan and phenylalanine amino acids in place of leucine in the consensus α-helical motif (*e.g.* ^{23}FQNLF^{27} and ^{433}WHTLF^{437} sequences located in the AR N-terminal domain of the AR LBD). Furthermore, a point mutation (T877A) at the NR surface was made, which rendered AR-positive responsive LNCaP cells sensitive towards the isobutyl analogue, **8**. This effect was rationalized in terms of a reduction in lipophilicity and steric blockade at the NR surface, which enabled both analogues to be accommodated equally well.

Pyridylpyridones targeting the LXXLL motif. Recently, Hamilton and co-workers reported on the use of a pyridylpyridone biaryl system as a rigid molecular structure to inhibit the ER-coactivator interaction.[156] An interesting difference from the seminal triazine, pyrimidine and trithiane rings systems developed by Katzenellenbogen, where in this case the flat polar heteroaromatic core is appended with flexible lipophilic side-chain groups (*vide supra*), is that the pyridylpyridone ring system enables an out-of-plane projection of alkyl substituents in the NR box through rotation about the biaryl bond. Indeed, X-ray structure analysis of **23** (Fig. 10) clearly showed the two aromatic rings perpedicular to one another, thereby limiting



8



9

IC_{50} (μM)		**8**	**9**
	ERα	7.9	NB
	AR	NB	3.5
	AR (T877A)	7.4	18.9

Fig. 9 2,4,6-trisubstituted pyrimidine-based inhibitors of the NR-coactivator interaction: ER selective analogue **8**, modelled on the LXXLL consensus motif, and AR selective analogue **9**, modelled on aromatic-rich consensus motifs (NB = no binding).[155]

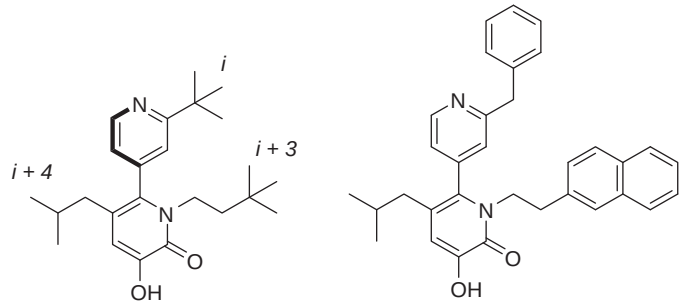


Fig. 10 Biaryl-based inhibitors of the ER-coactivator interaction, mimicking the *i*, $i+3$, and $i+4$ leucine residues of the natural LXXLL peptide motif.[156]

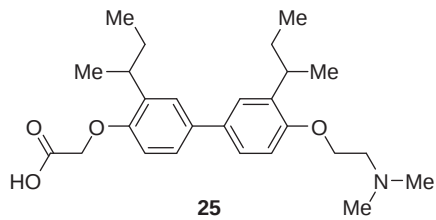


Fig. 11 Katzenellenbogen's biaryl inhibitor of the ER surface 'charge clamp'.[157]

steric repulsion between the pyridyl ring and the two lipophilic *ortho*-substituents at $i+3$ and $i+4$. An *in vitro* fluorescence polarization assay was used to study the interaction of a number of synthetic derivatives leading to a series of novel inhibitors with K_i values in the mid-to-low micromolar range, with the best result achieved in the case of **24** (Fig. 10), which measured a K_i value of 4.2 μM. The structure of biphenyls **23** and **24** are based on the oligomeric structures developed by Hamilton and co-workers, introduced previously in this book chapter (*vide supra*).

Biphenyl inhibitors targeting the charge clamp. Katzenellenbogen *et al.* developed biaryl-derived inhibitors of the ER-coactivator interaction with specific designs for the surface 'charge-clamp', which brackets the NR box and is in many cases essential for coactivator binding.[157] Branched alkyl chains were additionally introduced in a manner similar to Hamilton *et al.* in their work on pyridylpyridone biphenyl system, to mimic the leucine side chains of the natural LXXLL peptide motif. From this work, compound **25** (Fig. 11) was identified as being the most promising candidate for further development, with a K_i of 33 μM. The same compound was reported to be active in a reporter gene assay with an $IC_{50} \approx 2\,\mu M$. However, the efficacy was found to be lower relative to the the positive control, presumably due to issues of cell permeability.

Scaffold structures detected by high-throughput screening (HTS)

An alternative approach makes use of large libraries of structurally diverse drug molecules, and emphasises the need for robust screening platforms,

including a reliable assay for detecting inhibition of the NR-coactivator interaction in a high-throughput manner.

Anti-inflammatory drugs, 3,3′,5-triiodothyroacetic, pyrazolopyrimidines and indoles targeting AR. Fletterick *et al*. combined a functional X-ray screening approach to identify molecules that inhibit coactivator binding in two distinct non-classical ways: namely, direct coactivator inhibition through binding at the AF2 surface, and allosteric inhibition through binding to BF3, which, *via* an alternative allosteric mechanism, produces a more disordered state in the coactivator LXXLL/FXXLF peptide motif in the crystal state (Fig. 12).[149]

For this work, a diverse 55,000-strong compound library – including a number of off-patent drug molecules – was screened using an automated fluorescence polarization assay, which was run simultaneous with X-ray soaking studies. From this screen, four 'hit' compounds were identified (Fig. 12), with IC_{50} values (concentration of small molecule, that leads to a 50% inhibition of the maximum polarization value) typically in the 50 μM regime: namely, three nonsteroidal anti-inflammatory drugs, flufenamic acid (**10**), and 3,3′,5- triiodothyroacetic acid (**11**), and a low-abundance thyroid hormone (**12**). One of the advantages of this combined screening/soaking approach is the potential for identifying novel modes of inhibition. This was highlighted by the discovery of a series of allosteric inhibitors of coactivator binding, including RB1 (**13**) and 2-methylindole (**14**), which were found to occupy a shallow groove distinct from the AF-2 domain, termed the BF-3 domain. The extent of their binding to AR is highlighted in Fig. 13. Compared with the natural LXXLL peptide motif, molecules **10-12** cover significantly less of the AR surface, but fill the deep lipophilic groove of the AF-2 domain, which is typically occupied by the phenyl and leucine side-chain residues of the LXXLL and FXXLF coactivator binding motifs. Nevertheless, compounds **10**–**14** represent molecularly efficient fragment molecules that will likely assist in the future development of more potent AR-coactivator inhibitors.

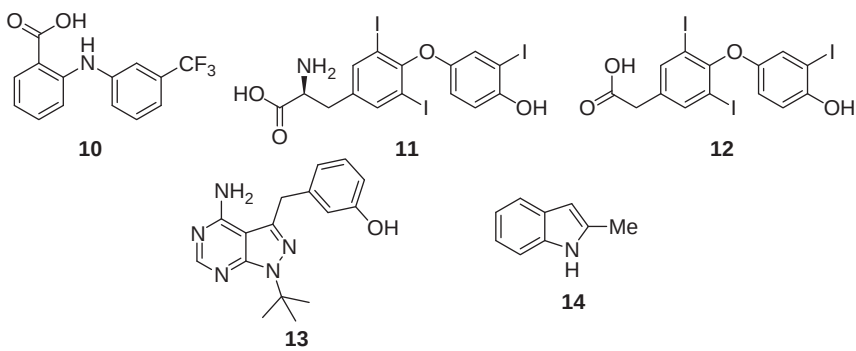


Fig. 12 Small molecules that inhibit the AR-coactivator interaction: Compounds **10**–**12** were first identified in a high-throughput screen of the AR LBD/SRC2–3 interaction, and later confirmed in an X-ray crystallography screen. By contrast, compounds **13** and **14** were identified alone by the the same X-ray crystallography screen.[149]

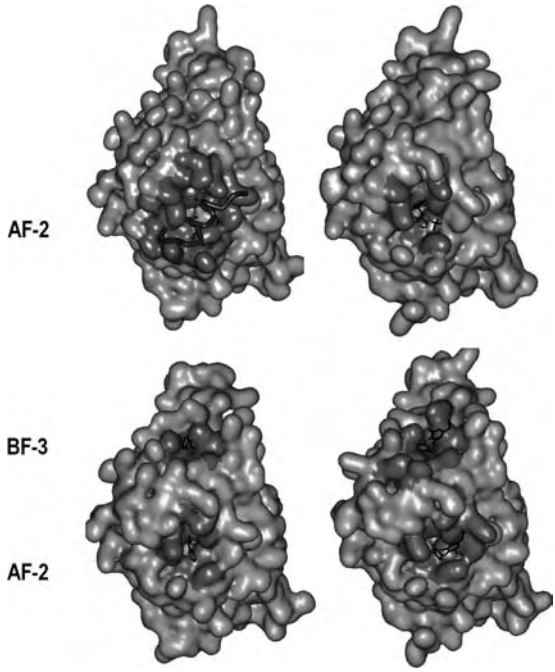


Fig. 13 Comparison of the AR LBD surface interactions (<4.5 Å) with the naturally-derived Src2 NR Box III peptide (*top left*) (PDB accession code: 1t63)[143] and synthetic small molecules: RB1 (**13**, *top right*, PDB accession code: 2piq),[149] 2-methylindole (**14**, *bottom left*, PDB accession code: 2pio)[149] and a thyroid hormone (**12**, *bottom right*, PDB accession code: 2piu).[149] The AR protein is rendered as a filled surface (grey). The peptide and small molecules are represented in tube and stick format, respectively.[149]

Guanylhydrazones targeting ER. Katzenellenbogen and co-workers recently characterized guanylhydrazone derivatives as ER coactivator binding inhibitors.[158] Following up on initial reports of **15** as a non-conventional ER antagonist in mammalian two-hybrid and ER transactivation assays,[159] a focussed set of sixteen analogues was synthesized (Fig. 14). Results from this work revealed that the focused compound set targeted the ER coactivator binding site in a concentration dependent fashion, achieving comparable or slightly lower micromolar IC_{50} values than **15** in reporter gene and mammalian two-hybrid assays (for example, **16**, Fig. 14). Whereas 4-hydroxytamoxifen (a potent classical ER antagonist) inhibited estradiol's binding to ER in a concentration dependent fashion, the effects of compound **16** in the same reporter gene assay were effectively the same at 1 nM and 100 nM estradiol concentrations. This result provided clear evidence that **16** specifically targeted the NR-coactivator interaction as opposed to the ligand binding pocket.

Quinazolinones and benzothiazoles targeting ER. Recently, high-throughput screening of a large library of diverse small molecules by Katzenellenbogen and co-workers[150] identified a number of novel scaffold structures as inhibitors of the ER-coactivator interaction. For this work, a highly efficient assay was developed, based on estradiol binding to europium-labeled ERα LBD, leading to recruitment of a Cy5-labeled fragment

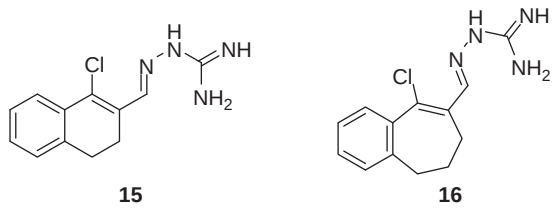


Fig. 14 Guanylhydrazone 'hit' compound (**15**) and an active analogue (**16**) derived from follow-up optimization studies of novel inhibitors of the ER-coactivator interaction.[158]

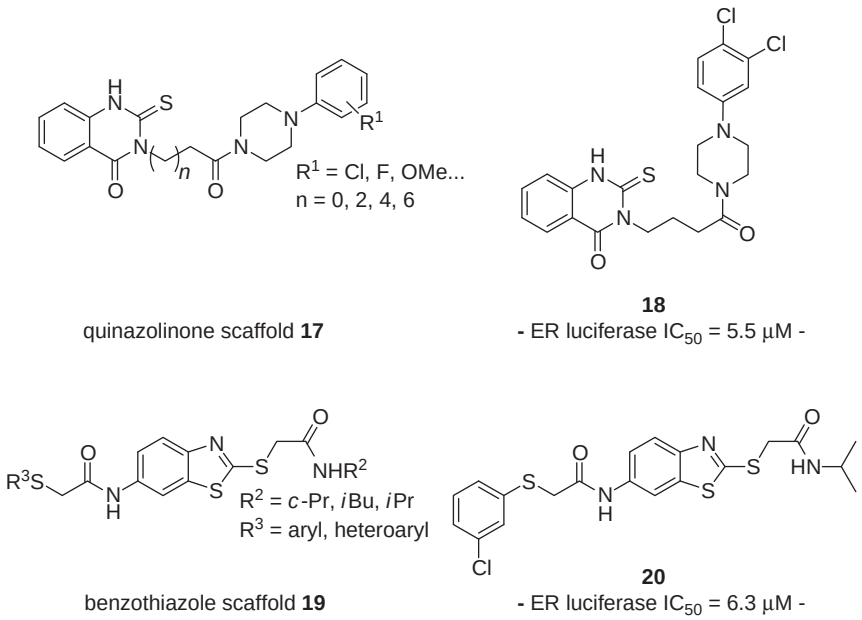


Fig. 15 Compound libraries based on quinazolinone and benzothiazole scaffold structures (**17** and **19**, respectively) with one of the most potent analogues from each series (**18** and **20**, respectively).[150]

of SRC-3.[160] This binding event could be monitored by TR-FRET, and was used to screen an 86,000-strong compound library for specific disruption of this interaction, manifest as a decrease in the FRET signal. The results from this screen were verified in follow-up assays, which led to the identification of four novel 'hits'. Two of these active compounds – one quinazolinone, one benzothiazole (Fig. 15, based on scaffolds **17** and **19**) – were further elaborated in sub-sequent rounds of synthesis and biological screening to enable structure-activity-relationship (SAR) studies. This resulted in the identification of a number of derivatives with luciferase IC_{50} values typically in the low micromolar range including compounds **18** and **20** (Fig. 15) – two of the most potent molecules from both series. A cell-based ERα-mediated luciferase reporter gene assay was also performed to highlight the cellular activity of the compounds. Furthermore, as with the guanylhydrazone series (*vide supra*), the discovery that different low and high concentrations of the potent endogenous ER agonist, estradiol, did not affect coactivator

inhibition provided convincing evidence that these molecules do indeed inhibit at the NR surface and not at the ligand binding pocket.

According to modelling studies (Figs. 16 and 17), these two inhibitor molecules (**18** and **20**) are thought to interact with the NR surface to a much smaller degree than the model peptide: the ILHRLL sequence of the SRC NR box peptide. Therefore, the inability to improve on the micromolar inhibition achieved until now might be explained by an entropic factor, whereby only a fraction of the water molecules are displaced by the ligand on binding to the protein surface (Fig. 17). This finding might provide an important structural explanation for why potent non-peptide based drug inhibitors of the NR-coactivator interaction are difficult to find. The high throughput screening of extensive, structurally diverse compound libraries has undoubtedly proven itself as a powerful approach to finding novel low micromolar active compounds. However, subsequent iterations of medicinal chemistry optimization, while clearly successful in making incremental

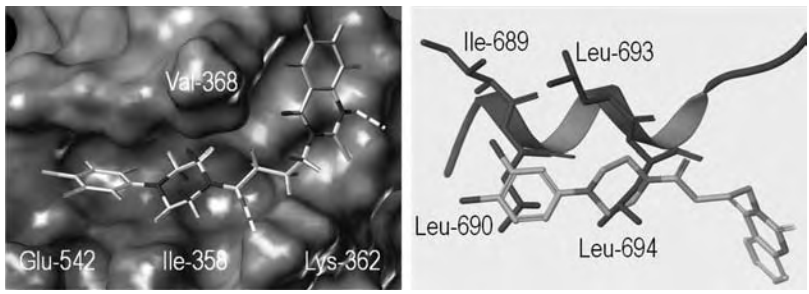


Fig. 16 *Left*. ERα binding mode hypothesis for quinazolinone analogue **18** – the most potent from the series – at the binding site (PDB accession code: 3erd). Proposed hydrogen bond interactions inidicated as dashed lines. *Right*. Quinazolinone analogue **18** aligned with the α-helical coactivator peptide (ribbon; PDB accession code: 3erd).[85,150] A. Sun, T. W. Moore, J. R. Gunther, M.-S. Kim, E. Rhoden, Y. Du, H. Fu, J. P. Snyder and J. A. Katzenellenbogen, Discovering small-molecule estrogen receptor α/coactivator binding inhibitors: high-throughput screening, ligand development, and models for enhanced potency, *ChemMedChem*, 2011, **6**, 654. Copyright (2011) John Wiley & Sons, Inc.

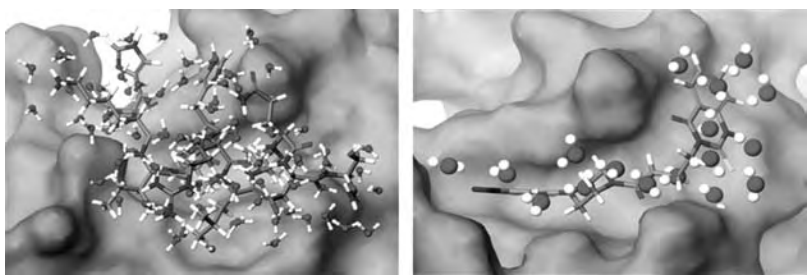


Fig. 17 Computational model explaining differences in experimental binding affinities between the ILHRLL sequence natural SRC coactivator peptide (*left*) and quinazolinone analogue **18** (*right*). In this case, the more lipophilic molecule displace a fraction of the water molecules compared with the natural peptide sequence.[85,150] A. Sun, T. W. Moore, J. R. Gunther, M.-S. Kim, E. Rhoden, Y. Du, H. Fu, J. P. Snyder and J. A. Katzenellenbogen, Discovering small-molecule estrogen receptor α/coactivator binding inhibitors: high-throughput screening, ligand development, and models for enhanced potency, *ChemMedChem*, 2011, **6**, 654. Copyright (2011) John Wiley & Sons, Inc.

improvements on the initial activity, have so far proved incapable of making the critical leap from micromolar to nanomolar potency levels. It could be suggested, therefore, that the screening of compound libraries with a bias towards higher molecular weight compounds or structures that by design more closely mimic the coactivator α-helix might lead to further success.

SJ-AK: a covalent irreversible inhibitor of TR. Work carried out in the laboratories of Guy and co-workers led to the identification of a novel series of β-aminoketones as inhibitors of the Thyroid Receptor (TR)-coactivator PPI.[161,162] Subsequent optimization of this compound class resulted in a second generation of more potent and less toxic analogues, including SJ000311413 or SJ-AK (**21**) (Fig. 18).[163,164] Rather than mimicking the α-helical coactivator in a reversible manner, it instead behaves irreversibly by forming an electrophilic enone, *in situ*, which is subjected to a 1,4,-Michael addition by cysteine C298 located at the coactivator binding site (AF-2) of the TR ligand binding domain.[165] The result is a covalent 'Michael' adduct, which prevents coactivator binding. In this study, the cellular mechanism of **21** and classical ligand-competitive inhibitor of TR, NH-3 (**22**), were directly compared and **21** was shown to act in a more restricted manner at the genomic level than classical antagonists (Fig. 18), which target the ligand binding pocket. While itself not an α-helix mimetic, as defined by other examples in this book chapter, **21** clearly shows that specifically targeting coactivator inhibition can potentially lead to more selective NR drug therapies.

Peptide-based small molecule inhibitors

NRs are regulated by various factors within the cellular milieu including interactions with coregulator proteins leading to receptor activation or repression and subsequent expression or silencing of target genes.[109] The importance of these interactions between the NR and small peptide motifs within coactivators has consequently rendered them attractive targets for

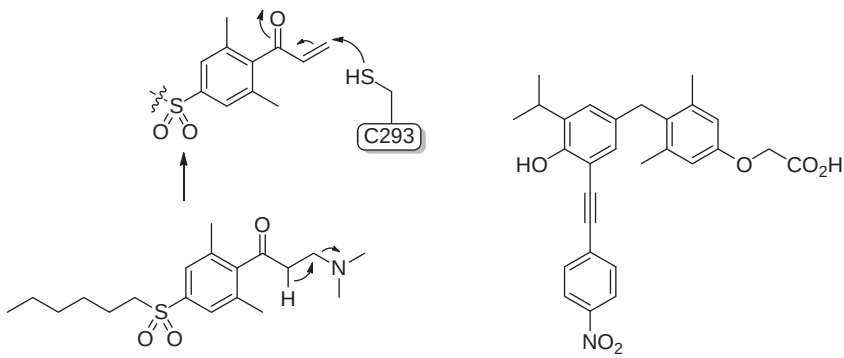


Fig. 18 SJ-AK (**21**) is a β-aminoketone, which, rather than targeting the TR ligand binding pocket, as in the case of NH3 (**22**), instead inhibits the recruitment of coactivators at target gene promoters, which has the effect of reversing TR induced gene expression and regulating the effects of the hormone downstream.[166]

the development of alternative drug therapies, which target the NR-coactivator interaction.[110] These coactivator binding inhibitors (CBI) are capable of blocking gene activation by direct disruption of the NR-coactivator interaction. Short synthetic peptides that appropriately mimic the structural features of the consensus LXXLL sequence have thus emerged as potential drug candidates.[111]

In view of the fact that natural α-helical structures are purely peptide-derived, the simplest way to inhibit their interactions has been through the development of mimics based on natural peptide sequences. This approach has in the past proved to be useful for determining the structural requirements of α-helix mediated PPI inhibition, leading to the development of useful small peptide-derived probes for use in cellular studies. From a drug discovery perspective, however, progress in the development of peptide-based inhibitors based on natural amino acids has been hindered by their reduced metabolic stability, due to the ease of recognition by endogenous proteases.[167] Furthermore, their low cell permeability has contrived their bioavailability and lower the therapeutic index. For this reason, small lipophilic drug molecules have tended to be preferred. However, the NR-coactivator interaction and other α-helix mediated PPI interactions are difficult targets, which are not easily addressed by small molecules other than natural peptides. A current trend in the PPI field therefore is to reconsider natural peptide sequences and address the lack of metabolic stability and cell permeability inherent to peptide-based inhibitors using chemical means.

Thus, a number of strategies have been developed to circumvent these issues, including, for example, α-helix containing miniproteins,[168,25] LXXLL-cell-penetrating peptide conjugates,[169] *N*-methylation, the use of unnatural D- or D/L-peptide sequences, the development of α-helices incorporating β-amino acids at repeating intervals[170] and even cyclization strategies such as macrocyclization, or side-chain disulfide or unnatural C-C bond ''stapling''.[171]

Spatola and co-workers have identified potent inhibitors of the ER-coactivator interaction based on short peptides with side-chain to side-chain covalent linkages.[131] On the one hand, intramolecular disulfide-bridged peptides involving different combinations of cysteine configurations and ring sizes were studied. A ($i \rightarrow i+3$) D-Cys, L-Cys disulfide cyclization resulted in the most favourable combination yielding the highest potencies (**26**, Fig. 19A), with added selectivity for ERα over ERβ. However, the same study showed that higher helicity did not necessarily translate into better surface binders. The potential of ($i \rightarrow i+4$) lactam-bridged peptides as stabilized α-helical peptide binders was also demonstrated, with selectivity for ERβ (for example, **27**, Fig. 19B).

Geistlinger and Guy reported on the selective inhibition of individual nuclear hormone receptors using lactam bridging peptide sequences modelled on the steroid receptor coactivator 2 (SRC2-2).[172,173] In the first instance, selective inhibition between ER and TR was achieved using a library of modified peptides by exploiting structural differences between respective NR boxes through modifications exclusively at the leucine amino acid residues of the LXXLL motif.[172] Later on, the same peptide library was

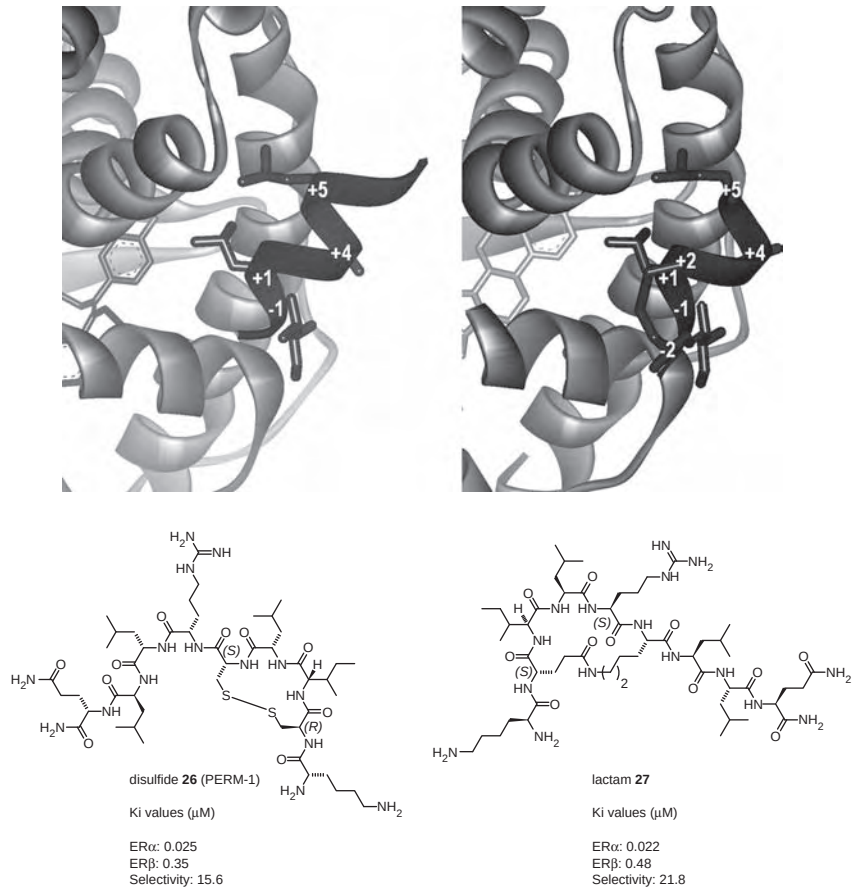


Fig. 19 Bridged inhibitors of the ER-coactivator interaction developed by Spatola and co-workers.[131] *Top*. Close-up of the coactivator binding site of the crystal structure ER/DES/GRIP1 (PDB accession code: 3erd)[85] (*left*) and ER/estradiol/PERM-1 – an ERα-selective (*i*→*i*+*3*) disulfide bridged peptide (**26**) – (PDB accession code: 1pcg)[131] (*right*). Both GRIP1 NR box II and **26** are shown in tube format. *Bottom*. Structures of modified peptides PERM-1, **26**, and (*i*→*i*+*4*) lactam bridged peptide **27**.[131]

used to achieved ER subtype selective co-activator inhibition for different ER agonist ligands.[173] This work showed that discrimination between NRs can be realised through manipulation within the LXXLL motif, a result that may have important implications for the future design of small lipophilic drug molecules.

In another approach reported by Phillips *et al*., the ER-coactivator interaction was inhibited using peptides stabilized by a hydrocarbon staple.[174] The staple was introduced at the solvent-exposed side of the α-helix by means of intramolecular ruthenium catalyzed metathesis of terminal olefin side-chains. In this way, peptides stapled at positions (*i*, *i*+*3*), (*i*, *i*+*4*) or (*i*, *i*+*7*) were prepared (*e.g*. *i*→*i*+*4* stapled peptide **29**, Fig. 20), which achieved improved binding affinities compared with the equivalent unstapled peptides (*e.g*. the natural GRIP1 peptide, **28**, Fig. 20), due to their increased α-helical character. In addition to the classical LXXLL binding interaction, detailed X-ray co-crystallography studies revealed non-classical

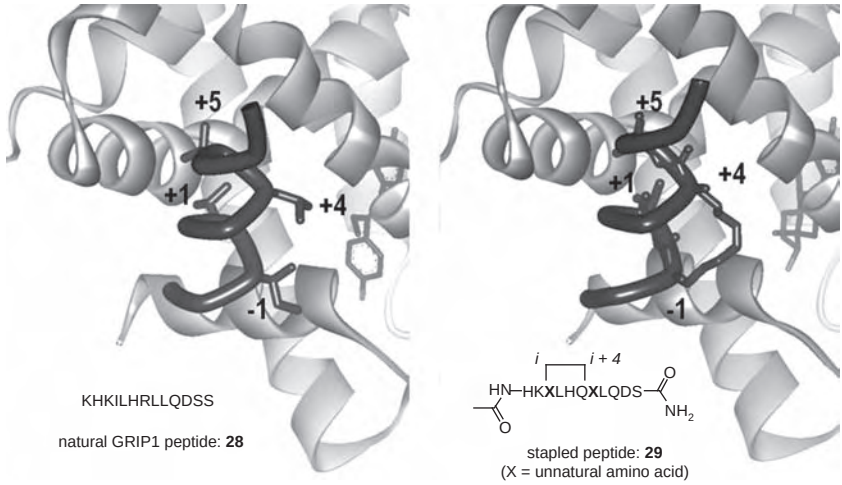


Fig. 20 Close-up of the coactivator binding site of the crystal structure ER/DES/GRIP1 (PDB accession code: 3erd)[85] (*left*) and ERα/estradiol/peptide **29** – a hydrocarbon ($i \rightarrow i+4$) stapled peptide (PDB accession code: 2yja)[174] (*right*). Both GRIP1 NR box II and **29** are shown in tube format.

van der Waals interactions between the lipophilic staple and the hydrophobic groups at the coactivator binding groove. In some cases as well, the introduction of the staple caused a turn in the α helix and a shift in the natural LXXLL binding register. Thus, the hydrocarbon staple significantly changed the physicochemical properties of the peptide and can in some cases lead to alternative modes of binding (Fig. 20).

Conclusions

The examples provided in this chapter have clearly shown that it is possible to inhibit α-helix-based protein-protein interactions (PPIs) using small molecules. Specifically for NRs, while the high-throughput screening (HTS) approach has clearly increased the frequency with which active compounds have been discovered (by sheer weight of numbers), it has not yet, however, led to a significant increase in binding affinities, especially compared with active compounds discovered by de novo design, and is a reflection on the fact that protein-protein interactions are challenging drug targets. There is a suggestion, therefore, that the type of molecules being screened might be of the wrong kind, and that future efforts in this direction might consider focusing on compounds capable of achieving multiple points of contact, for example, larger macrocyclic or long rigid multidomain heterocyclic-based or oligomeric type structures of the kind discussed at the beginning in this chapter, or even a return to peptide-based inhibitors. By using peptides as starting points it is possible to control target selectivity through simple modification of the amino acid side-chains. Furthermore, peptides have the right molecular properties to enable high affinity binding to shallow protein surfaces, such as greater hydrophilicity, multivalency and excellent side-chain packing. Recent efforts have focussed on translating excellent peptide lead structures into more druggable entities by increasing the metabolic

stability by, for example, stapling methods such as disulfide bond formation or all-carbon linkers *via* ring-closing metathesis. Strong ongoing research on modified peptide-based inhibitors, both for NRs and other PPIs, is expected to alternatively lead to novel molecular approaches to target these α-helix-based PPIs.

References

1 E. Giralt, M. W. Peczuh, X. Salvatella, *Protein Surface Recognition, Approaches for Drug Discovery*, Wiley, Weinheim, 2011.
2 M. R. Arkin and J. A. Wells, *Nat. Rev. Drug Discov.*, 2004, **3**, 301.
3 C. Katz, L. Levy-Beladev, S. Rotem-Bamberger, T. Rito, S. G. D. Rüdiger and A. Friedler, *Chem. Soc. Rev.*, 2011, **40**, 2131.
4 S. Zhong, A. T. Macias and A. D. MacKerell Jr., *Curr. Top. Med. Chem.*, 2007, **7**, 63.
5 A. Reayi and P. Arya, *Curr. Opin. Chem. Biol.*, 2005, **9**, 240.
6 J. Becerril, J. M. Rodriguez, P. N. Wyrembak and A. D. Hamilton, Chapter 5: Inhibition of Protein-Protein Interactions by Peptide Mimics, 2011, pp. 105 from Protein Surface Recognition, Approaches for Drug Discovery, Editors: E. Giralt, M. W. Peczuh, X. Salvatella, Wiley, Weinheim, 2011.
7 T. A. Edwards and A. J. Wilson, *Amino Acids*, 2011, **41**, 743.
8 M. J. Adler, A. G. Jamieson and A. D. Hamilton, *Curr. Top. Microbiol. Immunol.*, 2011, **348**, 1.
9 J. M. Davis, L. K. Tsoub and A. D. Hamilton, *Chem. Soc. Rev.*, 2007, **36**, 326.
10 A. L. J. Jochim and P. S. Arora, *ACS Chem. Biol.*, 2009, **5**, 919.
11 A. L. J. Jochim and P. S. Arora, *Mol. BioSyst.*, 2009, **5**, 924.
12 F. Bernal, A. F. Tyler, S. J. Korsmeyer, L. D. Walensky and G. L. Verdine, *J. Am. Chem. Soc.*, 2007, **129**, 2456.
13 J. R. Porter, M. R. Helmers, P. Wang, J. L. Furman, S. T. Joy, P. S. Arora and I. Ghosh, *Chem. Commun.*, 2010, **46**, 8020.
14 F. Campbell, J. P. Plante, T. A. Edwards, S. L. Warriner and A. J. Wilson, *Org. Biomol. Chem.*, 2010, **8**, 2344.
15 C. R. Braun, J. Mintseris, E. Gavathiotis, G. H. Bird, S. P. Gygi and L. D. Walensky, *Chem. Biol.*, 2010, **17**, 1325.
16 G. L. Semenza, *Nat. Rev. Cancer*, 2003, **3**, 721.
17 S. J. Freedman, Z. Y. Sun, F. Poy, A. L. Kung, D. M. Livingston, G. Wagner and M. J. Eck, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 5367.
18 S. A. Dames, M. Martinez-Yamout, R. N. De Guzman, H. J. Dyson and P. E. Wright, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 5271.
19 L. K. Henchey, S. Kushal, R. Dubey, R. N. Chapman, B. Z. Olenyuk and P. S. Arora, *J. Am. Chem. Soc.*, 2010, **132**, 941.
20 P. Polakis, *Genes Dev.*, 2000, **14**, 1837.
21 J. Sampietro, C. L. Dahlberg, U. S. Cho, T. R. Hinds, D. Kimelman and W. Q. Xu, *Mol. Cell*, 2006, **24**, 293.
22 K. I. Takemaru, M. Ohmitsu and F. Q. Li, *Handb. Exp. Pharmacol.*, 2008, **186**, 261.
23 S. A. Kawamoto, A. Coleska, X. Ran, H. Yi, C.-Y. Yang and Shaomeng Wang, *J. Med. Chem.*, 2012, **55**, 1137.
24 C. García, Chapter 6: Discovery of Inhibitors of Protein-Protein Interactions by Screening Chemical libraries, pp 133, from E. Giralt, M. W. Peczuh, X. Salvatella, *Protein Surface Recognition, Approaches for Drug Discovery*, Wiley, Weinheim, 2011.

25 T. Phan, H. D. Nguyen, H. Göksel, S. Möcklinghoff and L. Brunsveld, *Chem. Commun*., 2010, **46**, 8207.
26 M. D. Cummings, C. Schubert, D. J. Parks, R. R. Calvo, L. V. LaFrance, J. Lattanze, K. L. Milkiewicz and T. B. Lu, *Chem. Biol. Drug Design*, 2006, **67**, 201.
27 T. Shahian, G. M. Lee, A. Lazic, L. A. Arnold, P. Velusamy, C. M. Roels, R. K. Guy and C. S. Craik, *Nat. Chem. Biol*., 2009, **5**, 640.
28 J. M. Davis, L. K. Tsoub and A. D. Hamilton, *Chem. Soc. Rev*., 2007, **36**, 326.
29 O. Kutzki, H. S. Park, J. T. Ernst, B. P. Orner, H. Yin and A. D. Hamilton, *J. Am. Chem. Soc*., 2002, **124**, 11838.
30 H. Yin and A. D. Hamilton, *Angew. Chem., Int. Ed*., 2005, **44**, 4130.
31 J. M. Davis, A. Truong and A. D. Hamilton, *Org. Lett*., 2005, **7**, 5405.
32 I. C. Kim and A. D. Hamilton, *Org. Lett*., 2006, **8**, 1751.
33 H. Yin, G. I. Lee, K. A. Sedey, J. M. Rodriguez, H. G. Wang, S. M. Sebti and A. D. Hamilton, *J. Am. Chem. Soc*., 2005, **127**, 5463.
34 J. A. Hebda, I. Saraogi, M. Magzoub, A. D. Hamilton and A. D. Miranker, *Chem. Biol*., 2009, **16**, 943.
35 L. D. Walensky, A. L. Kung, I. Escher, T. J. Malia, S. Barbuto, R. D. Wright, G. Wagner, G. L. Verdine and S. J. Korsmeyer, *Science*, 2004, **305**, 1466.
36 D. Wang, K. Chen and P. S. Arora, *J. Am. Chem. Soc*., 2006, **128**, 9248.
37 A. Patgiri, K. K Yadav, P. S. Arora and D. Bar-Sagi, *Nat. Chem. Biol*., 2011, **7**, 585.
38 A. Kazi, J. Sun, K. Doi, S.-S. Sung, Y. Takahashi, H. Yin, J M. Rodriguez, J. Becerril, N. Berndt, A. D. Hamilton, H.-G. Wang and S. M. Sebti, *J. Biol. Chem*., 2011, **286**, 9382.
39 J. T. Ernst, J. Becerril, H. S. Park, Hang Yin and A. D. Hamilton, *Angew. Chem. Int. Ed*., 2003, **42**, 535.
40 B. P. Orner, J. T. Ernst and A. D. Hamilton, *J. Am. Chem. Soc*., 2001, **123**, 5382.
41 J. M. Rodriguez, L. Nevola, N. T. Ross, G. Lee and A. D. Hamilton, *Chembiochem*, 2009, **10**, 829–833.
42 J. M. Rodriguez and A. D. Hamilton, *Tetrahedron Lett*., 2006, **47**, 7443.
43 J. M. Rodriguez and A. D. Hamilton, *Angew. Chem. Int. Ed*., 2007, **46**, 8614.
44 T. Hara, S. R. Durell, M. C. Myers and D. H. Appella, *J. Am. Chem. Soc*., 2006, **1**, 1995.
45 R. Hayashi, D. Wang, T. Hara, J. A. Iera, S. R. Durell and D. H. Appella, *Bioorg.Med. Chem*., 2009, **17**, 7884.
46 J. A. Kritzer, J. D. Lear, M. E. Hodsdon and A. Schepartz, *J. Am. Chem. Soc*., 2004, **126**, 9468.
47 M. Werder, H. Hauser, S. Abele and D. Seebach, *Helv. Chim. Acta*, 1999, **82**, 1774.
48 Z. Z. Brown, J. Alleva and C. E. Schafmeister, *Biopolymers (Pept. Sci.)*, 2011, **96**, 578.
49 C. M. Goodman, S. Choi, S. Shandler and W. F. DeGrado, *Nat. Chem. Biol*., 2007, **3**, 252.
50 S. Hecht and I. Huc, *Foldamers: Structure, Properties, and Applications*, Wiley, Weinheim, 2007.
51 W. S. Horne and S. H. Gellman, *Acc. Chem. Res*., 2008, **41**, 1399.
52 C. Wild, T. Oas, C. McDanal, D. Bolognesi and T. Matthews, *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 10537.
53 C. Wild, D. C. Shugars, T. K. Greenwell, C. B. Mcdanal and C. J. Matthews, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 9770.
54 C.-H. Chen, T. J. Matthews, C. B. McDanal, D. P. Bolognesi and M. L. Greenberg, *J. Virol*., 1995, **69**, 3771.

55 C. Wild, T. Greenwell and T. Matthews, *AIDS Res. and Human Retrovir*., 1993, **9**, 1051.
56 J. K. Judice, J. Y. K. Tom, W. Huang, T. Wrin, J. Vennari, C. J. Petropoulos and R. S. McDowell, *Proc. Natl. Acad. Sci. U. S. A*., 1997, **94**, 13426.
57 K. L. Myung, K. K. Hee, Y. L. Tae, K.-S- Hahm and L. K. Kil, *Exp. Mol. Med*., 2006, **38**, 18.
58 M. S. Wolfe, W. P. Esler and C. Das., *J. Mol. Neurosc*., 2002, **19**, 83.
59 C. Blaines-Mira, M. T. Pastor, E. Valera, G. Fernández-Ballester, J. M. Merino, L. M. Gutierrez, E. Perez-Payá and A. Ferrer-Montiel, *Biochem. J*., 2003, **375**, 159.
60 J. C. Phelan, N. J. Skelton, A. C. Braisted and R. S. McDowell, *J. Am. Chem. Soc*., 1997, **119**, 455.
61 G. Udugamasooriya, D. Saro and M. R. Spaller, *Org. Lett*., 2005, **7**, 1203.
62 J. C. Calvo, K. C. Choconta, D. Diaz, O. Orozco, M. M. Bravo, F. Espejo, L. M. Salazar, F. Guzman and M. E. Patarroyo, *J. Med. Chem*., 2003, **46**, 5389.
63 D. Y. Jackson, D. S. King, J. Chmielewski, S. Singh and P. Schultz, *J. Am. Chem. Soc*., 1991, **113**, 9391.
64 H. E. Blackwell and R. H. Grubbs, *Angew. Chem. Int. Ed*., 1998, **37**, 3281.
65 H. Zhang, F. Curreli, X. Zhang, S. Bhattacharya, A. A Waheed, A. Cooper, D. Cowburn, E. O Freed and A. K. Debnath, *Retrovirol*., 2011, **8**, 1.
66 D. J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon and R. M. Evans, *Cell*, 1995, **83**, 835.
67 N. J. McKenna and B. W. O'Malley, *Cell*, 2002, **108**, 465.
68 H. Gronemeyer, J. A. Gustafsson and V. Laudet, *Nat. Rev. Drug Discov*., 2004, **3**, 950.
69 P. Germain, B. Staels, C. Dacquet, M. Spedding and V. Laudet, *Pharmacol. Rev*., 2006, **58**, 685.
70 G. K. Whitfield, P. W. Jurutka, C. A. Haussler and M. R. Haussler, *J. Cell Biochem*., 1999, 110.
71 T. W. Moore, C. G. Mayne and J. A. Katzenellenbogen, *Mol. Endocrinol*., 2010, **24**, 683.
72 D. Moras and H. Gronemeyer, *Curr. Opin. Cell Biol*., 1998, **10**, 384.
73 M. Robinson-Rechavi, H. Escriva Garcia and V. Laudet, *J. Cell Sci*., 2003, **116**, 585.
74 D. N. Lavery and I. J. McEwan, *Biochem. J*., 2005, **391**, 449.
75 W. Bourguet, M. Ruff, P. Chambon, H. Gronemeyer and D. Moras, *Nature*, 1995, **375**, 377.
76 P. F. Egea, A. Mitschler, N. Rochel, M. Ruff, P. Chambon and D. Moras, *EMBO J*., 2000, **19**, 2592.
77 A. M. Brzozowski, A. C. Pike, Z. Dauter, R. E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G. L. Greene, J. A. Gustafsson and M. Carlquist, *Nature*, 1997, **389**, 753.
78 A. K. Shiau, D. Barstad, J. T. Radek, M. J. Meyers, K. W. Nettles, B. S. Katzenellenbogen, J. A. Katzenellenbogen, D. A. Agard and G. L. Greene, *Nat. Struct. Biol*., 2002, **9**, 359.
79 V. Chandra, P. Huang, Y. Hamuro, S. Raghuram, Y. Wang, T. P. Burris and F. Rastinejad, *Nature*, 2008, **456**, 350.
80 I. Orlov, N. Rochel, D. Moras and B. P. Klaholz, *EMBO J*., 2011, **31**, 291.
81 W. B. Pratt and D. O. Toft, *Endocr. Rev*., 1997, **18**, 306.
82 W. Bourguet, P. Germain and H. Gronemeyer, *Trends Pharmacol. Sci*., 2000, **21**, 381.

83 A. C. Pike, A. M. Brzozowski and R. E. Hubbard, *J. Steroid Biochem. Mol. Biol.*, 2000, **74**, 261.
84 P. Huang, V. Chandra and F. Rastinejad, *Annu. Rev. Physiol.*, 2010, **72**, 247.
85 A. K. Shiau, D. Barstad, P. M. Loria, L. Cheng, P. J. Kushner, D. A. Agard and G. L. Greene, *Cell*, 1998, **95**, 927.
86 A. C. W. Pike, A. M. Brzozowski, J. Walton, R. E. Hubbard, A. G. Thorsell, Y. L. Li, J. A. Gustafsson and M. Carlquist, *Structure*, 2001, **9**, 145.
87 A. C. Pike, A. M. Brzozowski, R. E. Hubbard, T. Bonn, A. G. Thorsell, O. Engstrom, J. Ljunggren, J. A. Gustafsson and M. Carlquist, *EMBO J.*, 1999, **18**, 4608.
88 M. I. Dawson and Z. Xia, *Biochim. Biophys. Acta*, 2012, **1821**, 21.
89 A. R. de Lera, W. Bourguet, L. Altucci and H. Gronemeyer, *Nature Rev. Drug Discov.*, 2007, **6**, 811.
90 J. P. Renaud, N. Rochel, M. Ruff, V. Vivat, P. Chambon, H. Gronemeyer and D. Moras, *Nature*, 1995, **378**, 681.
91 D. M. Heery, E. Kalkhoven, S. Hoare and M. G. Parker, *Nature*, 1997, **387**, 733.
92 R. T. Nolte, G. B. Wisely, S. Westin, J. E. Cobb, M. H. Lambert, R. Kurokawa, M. G. Rosenfeld, T. M. Willson, C. K. Glass and M. V. Milburn, *Nature*, 1998, **395**, 137.
93 A. Warnmark, E. Treuter, J. A. Gustafsson, R. E. Hubbard, A. M. Brzozowski and A. C. Pike, *J. Biol. Chem.*, 2002, **277**, 21862.
94 P. S. Danielian, R. White, J. A. Lees and M. G. Parker, *EMBO J.*, 1992, **11**, 1025.
95 J. T. Moore, J. L. Collins and K. H. Pearce, *ChemMedChem*, 2006, **1**, 504.
96 S. Y. Dai, M. J. Chalmers, J. Bruning, K. S. Bramlett, H. E. Osborne, C. Montrose-Rafizadeh, R. J. Barr, Y. Wang, M. Wang, T. P. Burris, J. A. Dodge and P. R. Griffin, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 7171.
97 S. Y. Dai, T. P. Burris, J. A. Dodge, C. Montrose-Rafizadeh, Y. Wang, B. D. Pascal, M. J. Chalmers and P. R. Griffin, *Biochemistry*, 2009, **48**, 9668.
98 A. C. Figueira, D. M. Saidemberg, P. C. Souza, L. Martinez, T. S. Scanlan, J. D. Baxter, M. S. Skaf, M. S. Palma, P. Webb and I. Polikarpov, *Mol. Endocrinol.*, 2011, **25**, 15.
99 T. Hunter and M. Karin, *Cell*, 1992, **70**, 375.
100 L. C. Murphy, S. V. Seekallu and P. H. Watson, *Endocr. Rel. Cancer*, 2011, **18**, R1.
101 A. J. Whitmarsh and R. J. Davis, *Cell. Mol. Life Sci.*, 2000, **57**, 1172.
102 W. Zwart, A. Griekspoor, V. Berno, K. Lakeman, K. Jalink, M. Mancini, J. Neefjes and R. Michalides, *EMBO J.*, 2007, **26**, 3534.
103 S. F. Arnold, M. Melamed, D. P. Vorojeikina, A. C. Notides and S. Sasson, *Mol. Endocrinol.*, 1997, **11**, 48.
104 Z. Guo, B. Dai, T. Jiang, K. Xu, Y. Xie, O. Kim, I. Nesheiwat, X. Kong, J. Melamed, V. D. Handratta, V. C. Njar, A. M. Brodie, L. R. Yu, T. D. Veenstra, H. Chen and Y. Qiu, *Cancer Cell*, 2006, **10**, 309.
105 S. Kraus, D. Gioeli, T. Vomastek, V. Gordon and M. J. Weber, *Cancer Res.*, 2006, **66**, 11047.
106 S. J. Han, D. M. Lonard and B. W. O'Malley, *Trends Endocrinol. Metab.*, 2009, **20**, 8.
107 R. S. Savkur and T. P. Burris, *J. Pept. Res.*, 2004, **63**, 207.
108 N. J. McKenna, R. B. Lanzn and B. W. O'Malley, *Endocr. Rev.*, 1999, **20**, 321.
109 W. L. Kraus, E. M. McInerney and B. S. Katzenellenbogen, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 12314.
110 Y. Merot, R. Metivier, G. Penot, D. Manu, C. Saligaut, F. Gannon, F. Pakdel, O. Kah and G. Flouriot, *J. Biol. Chem.*, 2004, **279**, 26184.

111 R. Metivier, G. Penot, G. Flouriot and F. Pakdel, *Mol. Endocrinol.*, 2001, **15**, 1953.
112 E. B. Askew, J. T. Minges, A. T. Hnat and E. M. Wilson, *Mol. Cell Endocrinol.*, 2012, **348**, 403.
113 A. Warnmark, E. Treuter, A. P. Wright and J. A. Gustafsson, *Mol. Endocrinol.*, 2003, **17**, 1901.
114 I. J. McEwan, D. Lavery, K. Fischer and K. Watt, *Nucl. Recept. Signal*, 2007, **5**, e001.
115 C. L. Bevan, S. Hoare, F. Claessens, D. M. Heery and M. G. Parker, *Mol. Cell Biol.*, 1999, **19**, 8383.
116 J. A. Tan, S. H. Hall, P. Petrusz and F. S. French, *Endocrinology*, 2000, **141**, 3440.
117 J. A. Simental, M. Sar, M. V. Lane, F. S. French and E. M. Wilson, *J. Biol. Chem.*, 1991, **266**, 510.
118 W. Feng, R. C. J. Ribeiro, R. L. Wagner, H. Nguyen, J. W. Apriletti, R. J. Fletterick, J. D. Baxter, P. J. Kushner and B. L. West, *Science*, 1998, **280**, 1747.
119 C. Chang, J. D. Norris, H. Gron, L. A. Paige, P. T. Hamilton, D. J. Kenan, D. Fowlkes and D. P. McDonnell, *Mol. Cell Biol.*, 1999, **19**, 8226.
120 B. D. Darimont, R. L. Wagner, J. W. Apriletti, M. R. Stallcup, P. J. Kushner, J. D. Baxter, R. J. Fletterick and K. R. Yamamoto, *Genes Dev.*, 1998, **12**, 3343.
121 E. M. McInerney, D. W. Rose, S. E. Flynn, S. Westin, T. M. Mullen, A. Krones, J. Inostroza, J. Torchia, R. T. Nolte, N. Assa-Munt, M. V. Milburn, C. K. Glass and M. G. Rosenfeld, *Genes Dev.*, 1998, **12**, 3357.
122 B. He and E. M. Wilson, *Mol. Cell Biol.*, 2003, **23**, 2135.
123 J. D. Norris, L. A. Paige, D. J. Christensen, C. Y. Chang, M. R. Huacani, D. Fan, P. T. Hamilton, D. M. Fowlkes and D. P. McDonnell, *Science*, 1999, **285**, 744.
124 L. A. Paige, D. J. Christensen, H. Gron, J. D. Norris, E. B. Gottlin, K. M. Padilla, C. Y. Chang, L. M. Ballas, P. T. Hamilton, D. P. McDonnell and D. M. Fowlkes, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 3999.
125 H. J. Huang, J. D. Norris and D. P. McDonnell, *Mol. Endocrinol.*, 2002, **16**, 1778.
126 N. B. Mettu, T. B. Stanley, M. A. Dwyer, M. S. Jansen, J. E. Allen, J. M. Hall and D. P. McDonnell, *Mol. Endocrinol.*, 2007, **21**, 2361.
127 J. M. Hall, C. Y. Chang and D. P. McDonnell, *Mol. Endocrinol.*, 2000, **14**, 2010.
128 D. P. McDonnell, C. Y. Chang and J. D. Norris, *J. Steroid. Biochem Mol. Biol.*, 2000, **74**, 327.
129 C. Y. Chang, J. D. Norris, M. Jansen, H. J. Huang and D. P. McDonnell, *Methods Enzymol.*, 2003, **364**, 118.
130 C. Y. Chang, J. Abdo, T. Hartney and D. P. McDonnell, *Mol. Endocrinol.*, 2005, **19**, 2478.
131 A. M. Leduc, J. O. Trent, J. L. Wittliff, K. S. Bramlett, S. L. Briggs, N. Y. Chirgadze, Y. Wang, T. P. Burris and A. F. Spatola, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 11273.
132 L. Ko, G. R. Cardona, T. Iwasaki, K. S. Bramlett, T. P. Burris and W. W. Chin, *Mol. Endocrinol.*, 2002, **16**, 128.
133 X. F. Ding, C. M. Anderson, H. Ma, H. Hong, R. M. Uht, P. J. Kushner and M. R. Stallcup, *Mol. Endocrinol.*, 1998, **12**, 302.
134 B. He, J. A. Kemppainen, J. J. Voegel, H. Gronemeyer and E. M. Wilson, *J. Biol. Chem.*, 1999, **274**, 37219.
135 V. Christiaens, C. L. Bevan, L. Callewaert, A. Haelens, G. Verrijdt, W. Rombauts and F. Claessens, *J. Biol. Chem.*, 2002, **277**, 49230.

136 B. He, J. A. Kemppainen and E. M. Wilson, *J. Biol. Chem.*, 2000, **275**, 22986.
137 K. Steketee, C. A. Berrevoets, H. J. Dubbink, P. Doesburg, R. Hersmus, A. O. Brinkmann and J. Trapman, *Eur. J. Biochem.*, 2002, **269**, 5780.
138 D. J. van de Wijngaart, H. J. Dubbink, M. E. van Royen, J. Trapman and G. Jenster, *Mol. Endocrinol.*, 2012, **352**, 57.
139 C. L. Hsu, Y. L. Chen, S. Yeh, H. J. Ting, Y. C. Hu, H. Lin, X. Wang and C. Chang, *J. Biol. Chem.*, 2003, **278**, 23691.
140 E. Hur, S. J. Pfaff, E. S. Payne, H. Gron, B. M. Buehrer and R. J. Fletterick, *PLoS Biol.*, 2004, **2**, E274.
141 B. He, R. T. Gampe, Jr., A. J. Kole, A. T. Hnat, T. B. Stanley, G. An, E. L. Stewart, R. I. Kalman, J. T. Minges and E. M. Wilson, *Mol. Cell*, 2004, **16**, 425.
142 H. J. Dubbink, R. Hersmus, A. C. Pike, M. Molier, A. O. Brinkmann, G. Jenster and J. Trapman, *Mol. Endocrinol.*, 2006, **20**, 1742.
143 E. Estebanez-Perpina, J. M. R. Moore, E. Mar, E. Delgado-Rodrigues, P. Nguyen, J. D. Baxter, B. M. Buehrer, P. Webb, R. J. Fletterick and R. K. Guy, *J. Biol. Chem.*, 2005, **280**, 8060.
144 E. H. Kong, N. Heldring, J.-Å. Gustafsson, E. Treuter, R. E. Hubbard and A. C. W. Pike, *Proc. Nat. Acad. Sci. U.S.A.*, 2005, **102**, 3593.
145 J. Millour, D. Constantinidou, A. V. Stavropoulou, M. S. C. Wilson, S. S. Myatt, J. M.-M. Kwok, K. Sivanandan, R. C. Coombes, R. H. Medema, J. Hartman, A. E. Lykkesfeldt and E. W.-F. Lam, *Oncogene*, 2010, **29**, 2983.
146 J. Vagner, H. Qu and V. J. Hruby, *Curr. Opin. Chem. Biol.*, 2008, **12**, 292.
147 M. W. Peczuh and A. D. Hamilton, *Chem. Rev.*, 2000, **100**, 2479.
148 P. L. Toogood, *J. Med. Chem.*, 2002, **45**, 1543.
149 E. Estébanez-Perpiñá, L. A. Arnold, P. Nguyen, E. Delgado Rodrigues, Ellena Mar, R. Bateman, P. Pallai, K. M. Shokat, J. D. Baxter, R. K. Guy, P. Webb and R. J. Fletterick, *Proc. Nat. Acad. Sci. U.S.A.*, 2007, **104**, 16074.
150 A. Sun, T. W. Moore, J. R. Gunther, M.-S. Kim, E. Rhoden, Y. Du, H. Fu, J. P. Snyder and J. A. Katzenellenbogen, *ChemMedChem*, 2011, **6**, 654.
151 A. A. Parent, J. R. Gunther and J. A. Katzenellenbogen, *J. Med. Chem.*, 2008, **4**, 435.
152 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 2001, **46**, 3.
153 H. Zhou, M. L. Collins, J. R. Gunther, J. S. Comninos and J. A. Katzenellenbogen, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 4118.
154 A. L. Rodriguez, A. Tamrazi, M. L. Collins and J. A. Katzenellenbogen, *J. Med. Chem.*, 2004, **47**, 600.
155 J. R. Gunther, A. A. Parent and J. A. Katzenellenbogen, *ACS Chem. Biol.*, 2009, **4**, 435.
156 J. Becerril and A. D. Hamilton, *Angew. Chem. Int. Ed.*, 2007, **46**, 4471.
157 A. B. Williams, P. T. Weiser, R. N. Hanson, J. R. Gunther and J. A. Katzenellenbogen, *Org. Lett.*, 2009, **11**, 5370.
158 A. L. LaFrate, J. R. Gunther, K. E. Carlson and J. A. Katzenellenbogen, *Bioorg. Med. Chem.*, 2008, **16**, 10075.
159 D. Shao, T. J. Berrodin, E. Manas, D. Hauze, R. Powers, A. Bapat, D. Gonder, R. C. Winneker and D. E. Frail, *J. Steroid Biochem. Mol. Biol.*, 2004, **88**, 351.
160 J. R. Gunther, Y. Du, E. Rhoden, I. Lewis, B. Revennaugh, T. W. Moore, S. H. Kim, R. Dingledine, H. Fu and J. A. Katzenellenbogen, *J. Biomol. Screen.*, 2009, **14**, 181.

161 L. A. Arnold, E. Estébanez-Perpiñá, M. Togisha, N. Jouravel, A. Shelat, A. C. McReynolds, P. Nguyen, J. D. Baxter, R. J. Fletterich, P. Webb and R. K. Guy, *J. Biol. Chem.*, 2005, **280**, 43048.
162 L. A. Arnold, E. Estébanez-Perpiñá, M. Togisha, A. Shelat, C. A. Ocasio, A. C. McReynolds, P. Nguyen, J. D. Baxter, R. J. Fletterich, P. Webb and R. K. Guy, *Sci. STKE 2006*, 2006, pl3.
163 L. A. Arnold, A. Kosinski, E. Estébanez-Perpiñá, R. J. Fletterich and R. K. Guy, *J. Med. Chem.*, 2007, **50**, 5269.
164 J. Y. Hwang, L. A. Arnold, F. Zhu, A. Kosinski, T. J. Mangano, V. Setola, B. L. Roth and R. K. Guy, *J. Med. Chem.*, 2009, **52**, 3892.
165 E. Estébanez-Perpiñá, L. A. Arnold, N. Jouravel, M. Togashi, J. Blethrow, E. Mar, P. Nguyen, K. J. Phillips, J. D. Baxter, P. Webb, R. K. Guy and R. J. Fletterick, *Mol. Endocrinol.*, 2007, **21**, 2919.
166 P. Sadana, J. Y. Hwang, R. R. Attia, L. A. Arnold, G. Neale and R. K. Guy, *ACS Chem. Biol.*, 2011, **6**, 1096.
167 J. J. Nestor, Jr., *Curr. Med. Chem.*, 2009, **16**, 4399.
168 B. Vaz, S. Möcklinghoff and S. Folkertsma, S. Lusher, J. de Vlieg and Luc Brunsveld, *Chem. Commun.*, 2009, 5377.
169 M. Carraz, W. Zwart, T. Phan, R. Michalides and L. Brunsveld, *Chem. Biol.*, 2009, **16**, 702.
170 E. F. Lee, B. J. Smith, W. S. Horne, K. N. Mayer, M. Evangelista, P. M. Colman, S. H. Gellman and W. D. Fairlie, *ChemBioChem*, 2011, **12**, 2025.
171 C. E. Schafmeister, J. Po and G. L. Verdine, *J. Am. Chem. Soc.*, 2000, **122**, 5891.
172 T. R. Geistlinger and R. K. Guy, *J. Am. Chem. Soc.*, 2003, **125**, 6852.
173 T. R. Geistlinger, A. C. McReynolds and R. K. Guy, *Chem. Biol.*, 2004, **11**, 273.
174 C. Phillips, L. R. Roberts, M. Schade, R. Bazin, A. Bent, N. L. Davies, R. Moore, A. D. Pannifer, A. R. Pickford, S. H. Prior, C. M. Read, A. Scott, D. G. Brown, B. Xu and S. L. Irving, *J. Am. Chem. Soc.*, 2011, **133**, 9696.