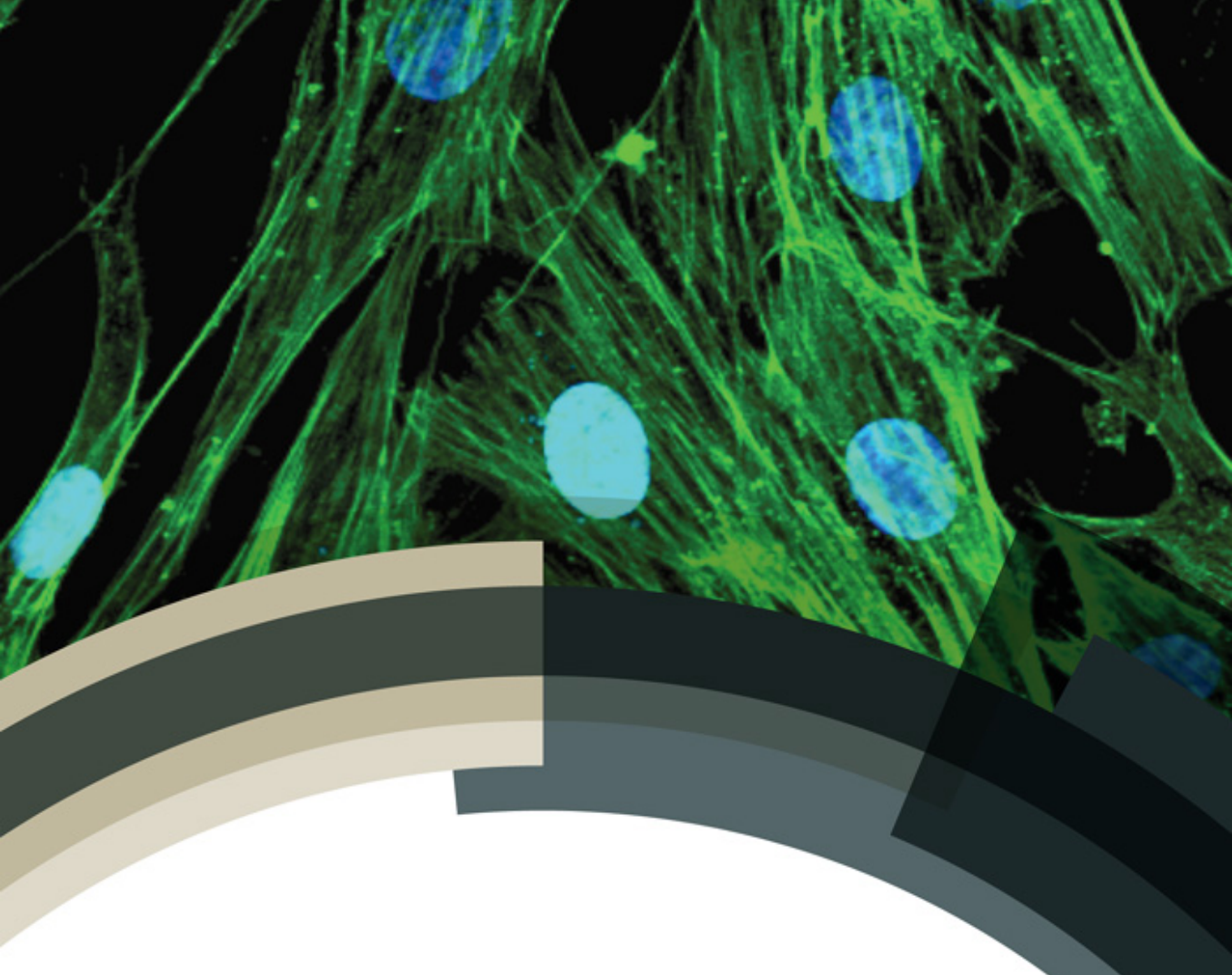


Specialist Periodical Reports

Amino Acids, Peptides and Proteins

Volume 43

Edited by Maxim Ryadnov and Ferenc Hudecz

ROYAL SOCIETY OF CHEMISTRY

Amino Acids, Peptides and Proteins

Volume 43

A Specialist Periodical Report

Amino Acids, Peptides and Proteins

Volume 43

Editors
Maxim Ryadnov, *National Physical Laboratory, UK*
Ferenc Hudecz, *Eötvös Lorand University, Hungary*

Authors
A. Chianese, *University of Campania Luigi Vanvitelli, Italy*
G. Franci, *University of Campania Luigi Vanvitelli and CIRPEB, Italy*
M. Galdiero, *University of Campania Luigi Vanvitelli and CIRPEB, Italy*
S. Galdiero, *University of Naples Federico II and CIRPEB, Italy*
Ferenc Hudecz, *Eötvös Loránd University, Hungary*
Naresh Kumar, *National Physical Laboratory, UK*
G. Morelli, *University of Naples Federico II and CIRPEB, Italy*
Eugeni M. Ryadnov, *University College London, UK*
Maxim G. Ryadnov, *National Physical Laboratory, UK*
D. Stelitano, *University of Campania Luigi Vanvitelli, Italy*
Beáta G. Vértessy, *BME Budapest University of Technology and Economics, Hungary*

ISBN: 978-1-78801-367-3
PDF ISBN: 978-1-78801-385-7
EPUB ISBN: 978-1-78801-866-1
ISSN: 1361-5904
DOI: 10.1039/9781788013857

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

The Royal Society of Chemistry is a charity, registered in England and Wales, Number 207890, and a company incorporated in England by Royal Charter (Registered No. RC000524), registered office: Burlington House, Piccadilly, London W1J 0BA, UK, Telephone: +44 (0) 20 7437 8656.

Visit our website at www.rsc.org/books

Printed in the United Kingdom by CPI Group (UK) Ltd, Croydon, CR0 4YY, UK

Preface

DOI: 10.1039/9781788013857-FP005

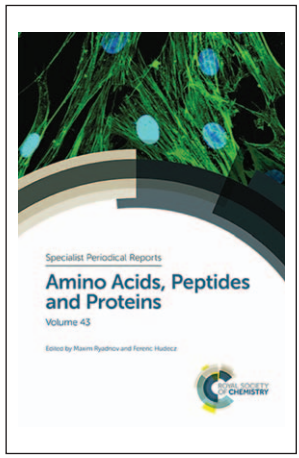
Amino acids, peptides and proteins is a book series that was launched in 1969 to provide a systemic overview of contemporary research efforts in protein and peptide science. Since its inception the series has covered recent developments in the chemistry and biology of proteins, peptides and their precursors without a bias of topical polularity. This has helped to keep abreast with achievements in different areas. Most recently, the series has started to explore topics that are complementary to protein chemistry to better reflect inter-disciplinarity of this specialist area and the importance of developing innovative measurement techniques and methods. Therefore, individual reports in this series are presented as accounts of both existing and emerging research areas relevant to protein science in a broad sense. This volume continues this tradition and brings new and established science together. The book reviewes literature predominantly published over the last few years, while each chapter also outlines underpinning concepts and terminology, which may have been introduced earlier but remain valid to date.

This 43rd volume opens with a detailed discussion of versatile proteins that recognise a non-canonical nucleobase uracil – a common error in human DNA – and induce highly precise and effective uracil-excision repair mechanisms (Vértessy). The proteins thus act in a ''search and destroy'' fashion helping to restore normal cellular functions. Elevated levels of uracil in cells can lead to the emergene of stable uracilated DNA that is implicated in a number of health-related conditions. For instance, HIV can readily convert deoxyuridine phosphate to build its retro-transcribed DNA in a uracilated form that helps regulate integration of the viral genetic information. The second chapter follows upon the discussion of viral infectivity while making a particular emphasis on viral membrane glycoproteins and their role in mediating the uptake of the herpes simplex virus (HSV) into human cells (Stelitano, Franci, Chianese, Galdiero, Morelli and Galdiero). Despite significant progress made in the understanding of precise mechanisms by which the virus overcomes and hijacks immune responses to infect cells, a vaccine for HSV-1 and HSV-2 has yet to emerge. It is becoming increasingly apparent however that it is the HSV glycoproteins themselves that can offer promising candidates for the engineering of effective vaccines against HSV. In this light, the chapter sets out to present the state of the art in vaccine development and highlight new insights into the structure–function relationships of HSV glycoproteins. The third chapter diversifies the discussion around biologically functional proteins by introducing branched polypeptide topologies. The role and importance of predictable physico-chemical, structural and compositional characteristics of such topologies are detailed starting with chemical synthesis (Hudecz). Here the focus is shifted towards capitalising on the versatility of peptide chemistry to assemble branched constructs exhibiting defined exploitable properties

ranging from responsive binding to phospholipids to multivalent conjugation with chemotherapeutic agents, molecular reportes and epitopes. Synthetic protocols are assessed in terms of the impact branching heterogeneity might have on conformational preferences of resulting polymers and subsequently on cell viability, immunoreactivity and biodistribution. Indeed, conformational changes play a critical role in normal and pathological processes of the cell. Notably, implications of incorrect folding or misfolding can be far reaching for proteopathies such as Alzheimer's and Parkinson's diseases, the rate of which is set to increase with the ageing population in the developed countries. Often the onset of these diseases is attributed to the structural alterations and modifications in three main protein classes – tau proteins, amyloid-β and presenilins. These proteins are discussed in the penultimate report of this volume which reviews the challenge of elucidating their derivatives that are considered and sought as the primary causative agents of neuropathological conditions (Ryadnov and Ryadnov). Innovative methods and measurement approaches that are being applied to capture and characterise these derivatives as well as their precursors constitute a complementary part of this discussion. Ultimately, the biological properties of proteins and peptides are linked to the chemistry of their primary structures and our ability to accurately determine and measure it. This is where the development of new techniques based on various physical phenomena can impact most. The closing chapter extends the recently introduced tendency of this series to discuss measurement capabilities, both established and emerging, and their impact in advancing protein science (Kumar). This review critically assesses pros and cons of tip-enhanced Raman spectroscopy for the chemical characterisation of amino acids, peptides and proteins at the nanoscale. Current methods and innovative solutions are presented and necessary comparisons are made with other tehniques that are more common for protein measurements. The report concludes the volume with exemplar attempts of applying the technique to establish empirical correlations between the chemistry, aggregation and morphology of amyloid oligomers – a feat that remains inaccessible to other methodologies.

Maxim Ryadnov and Ferenc Hudecz

CONTENTS



Cover

Front cover features a fluorescence micrograph of human dermal fibroblasts proliferating on a designed extracellular matrix. Cytoskeleton and nuclei are highlighted in green and blue, respectively. Image courtesy of Nilofar Faruqui and Mike Shaw.

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

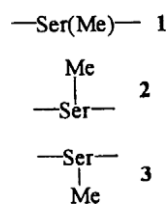
Aad	α-Aminoadipic acid
βAad	β-Aminoadipic acid
Abu	α-Aminobutyric acid
Aib	α-Aminoisobutyric acid; α-methylalanine
βAla	β-Alanine; 3-aminopropionic acid (avoid Bal)
Asu	α-Aminosuberic acid
Aze	Azetidine-2-carboxylic acid
Cha	β-cyclohexylalanine
Cit	Citrulline; 2-amino-5-ureidovaleric acid
Dha	Dehydroalanine (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)

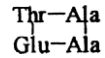
The three-letter symbols should be used in accord with the IUPAC-IUB conventions, which have been published in many places (*e.g. European J. Biochem.* 1984; **138**: 9–37), and which are (May 1999) also available with other relevant documents at: http://www.chem.qnw.ac.uk/iubmb/iubmb.html#03

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**.

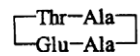


Note that the oxygen atom is not shown: it is contained in the three-letter symbol – showing it, as in Ser(OMe), would imply that a peroxy group was present. Bonds up or down should be used only for indicating side-chain substitution. Confusions may creep in if the three-letter symbols are used thoughtlessly 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.



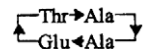
4

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**.



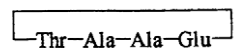
5

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-chlorobenzyloxycarbonyl and so on

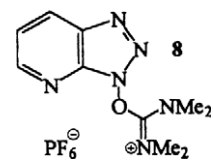
Amino Acid Derivatives

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

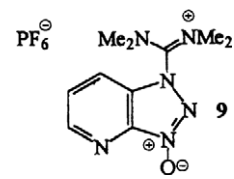
Reagents and Solvents

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