Journal of Peptide Science
J. Peptide Sci. **9**: 1–8 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.426

JOURNAL OF PEPTIDE SCIENCE

Editorial

A Revised Guide to Abbreviations in Peptide Science and a Plea for Conformity*

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

All other abbreviations should be defined. Previously published usage should be followed unless it is manifestly clumsy or inappropriate. Where it is necessary to devise new abbreviations and symbols, the general principles behind established examples should be followed. Thus, new amino-acid symbols should be of form Abc, with due thought for possible ambiguities (Dap might be obvious for diaminopropionic acid, for example, but what about diaminopimelic acid?). A certain amount of common sense is called for too. It is understandable, for example, that in laboratory conversation the trisyllabic TFM (*tee-eff-em*) drops off the tongue more easily than the six syllables of trifluoromethyl (*try-flew-or-oh-me-thile*), but to use TFM as an abbreviation for CF_3 in print is an absurd obfuscation.

Where alternatives are indicated below, the first is preferred.

AMINO ACIDS

Proteinogenic Amino Acids

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

* The first version of this Guide appeared in *J Peptide Sci* **5**:465–471 (1999). Further detail and guidance is now given as promised there, and by repetition the messages will hopefully be more widely heard.

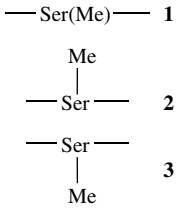
Other Amino Acids

Aad	α-Aminoadipic acid
βAad	β-Aminoadipic acid
Abu	α-Aminobutyric acid
Aib	α-Aminoisobutyric acid; α-methylalanine
β Ala	β-Alanine; 3-aminopropionic acid (avoid Bal)
Asu	α-Aminosuberic acid
Aze	Azetidine-2-carboxylic acid
Cha	β-Cyclohexylalanine
Cit	Citrulline; 2-amino-5-ureidovaleric acid
Dha	Dehydroalanine (also 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
*a*Ile	*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-2-carboxylic acid
Sar	Sarcosine; *N*-methylglycine
Sec	Selenocysteine
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
*a*Thr	*allo*-Threonine; 2*S*, 3*S* in the L-series
Thz	Thiazolidine-4-carboxylic acid, thiaproline
Xaa	Unknown or unspecified (also Aaa)

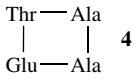
The three-letter symbols should be used in accord with the IUPAC-IUMB recommendations which have been published in many places, and are (2 July 2002) also available with other relevant documents at http://www.chem.qmw.ac.uk/iumb/. See especially *Nomenclature and symbolism for amino acids and peptides (Recommendations 1983)*, which is reproduced at http://www.chem.qmw.ac.uk/iumb/AminoAcid/.

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.

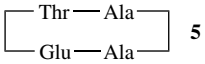
Side chains are understood to be unsubstituted if nothing is shown, but a substituent can be indicated by use of brackets or attachment by a vertical bond up or down. Thus an *O*-methylserine residue could be shown as **1, 2**, or **3**.



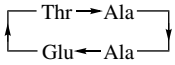
Note that the oxygen atom is not shown: it is contained in the three-letter symbol — showing it, as in Ser(OMe), would imply that a peroxy group was present. Bonds up or down should be used only for indicating side-chain substitution. Confusions may creep in if the three-letter symbols are used thoughtlessly in representations of cyclic peptides. Consider by way of example the hypothetical cyclopeptide threonylalany-l alanylglutamic acid. It might be thought that this compound could be economically represented **4**.



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

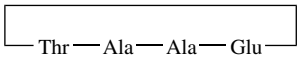


But this is now ambiguous because the convention that the symbols are to be read as having the amino nitrogen to the left cannot be imposed on both lines. The direction of the peptide bond needs to be shown with an arrow pointing from CO to N, as in **6**.



6

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



7

Particular numbered amino acid resides in a peptide chain should be referred to in the style e.g. Leu^5 not 5-Leu or Leu5, or spelt out i.e. leucine-5.

Analogues of peptides should be designated e.g. [Gly^7]-oxytocin, or spelt out 7-glycine-oxytocin for the analogue of oxytocin in which the proline at position 7 has been replaced by glycine. Further abbreviation would be acceptable here to [Gly^7]-OT. Multiple replacements should be indicated e.g. [Ser^4, Gly^7]-OT. Extensions as in glycyloxytocin where an additional glycine residue acylates the *N*-terminal residue of oxytocin should be rendered Gly-OT. Residue deletions, as in oxytocin with the leucine residue at position 8 removed so that the residue 7 in the intact hormone connects directly with the residue at position 9, should be abbreviated des-Tyr^8-OT.

Partial but otherwise natural sequences have been alluded to in various ways; the style e.g. OT-(2-5)-peptide for the four -residue sequence which comprises positions 2 to 5 inclusive of oxytocin, is preferred. Taking this a step further, the sequence PheIleGlnAsn, which corresponds to positions 2-5 of oxytocin with Phe replacing Tyr at position 2, could be abbreviated [Phe^2]-OT-(2-5)-peptide.

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 (care! confusion with benzyl is common)
Bzl	Benzyl (also Bn); Bzl(No_2) = 4-nitro benzyl and so on, the substituent location only being indicated if it is not *para*, e.g. Bzl($2NO_2$), 2-nitrobenzyl. If a substituted protecting group is used on a side chain, leading to brackets within brackets, a hierarchy should be used e.g. Tyr[Bzl($2NO_2$)].
Cha	Cyclohexylammonium salt
Clt	2-Chlorotrityl

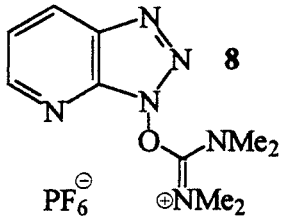
Dcha	Dicyclohexylammonium salt
Dde	1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl
Ddz	2-(3,5-Dimethoxyphenyl)-isopropoxycarbonyl
Dnp	2,4-Dinitrophenyl
Dns	5-Dimethylaminonaphthalene-1-sulfonyl (dansyl)
Dpm	Diphenylmethyl (also Bzh. benzhydryl)
Dpp	Diphenylphosphinyl
Et	Ethyl
Fmoc	9-Fluorenylmethoxycarbonyl
For	Formyl
Hmb	2-Hydroxyl-4-methoxy-benzyl
Mbh	4, 4′-Dimethoxydiphenylmethyl, 4, 4′-Dimethoxybenzhydryl
Mbs	4-Methoxybenzenesulfonyl
Me	Methyl
Mob	4-Methoxybenzyl
Mtr	2,3,6-Trimethyl,4-methoxybenzenesulfonyl
Nps	2-Nitrophenylsulfanyl
OAll	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	Methoxycarbonylsulfanyl
TBDMS	*t*-Butyldimethylsilyl
Tf	Trifluoromethanesulfony-l (triflyl)
TMS	Trimethylsilyl
Pmc	2,2,5,7,8-Pentamethylchroman-6-sulfonyl
Pr^i	*i*-Propyl
Pr^n	*n*-Propyl
Tfa	Trifluoroacetyl
Tos	4-Toluenesulfonyl (also Ts)
Troc	2,2,2-Trichloroethoxycarbonyl
Trt	Trityl, triphenylmethyl
Xan	9-Xanthydryl
Z	Benzyloxycarbonyl (also Cbz). Z(2Cl) = 2-chlorobenzyloxycarbonyl and so on: see also remarks on usage with substituted protecting groups under Bzl.

AMINO ACID DERIVATIVES

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

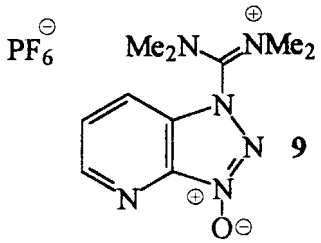
REAGENTS AND SOLVENTS

BOP	1-Benzotriazolyloxy-tris-dimethylamino-phosphonium hexafluorophosphate
CDI	Carbonyldiimidazole
DAST	Diethylaminosulfur trifluoride
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
DME	Dimethoxyethane
DMF	Dimethylformamide
DMS	Dimethylsulfide
DMSO	Dimethylsulfoxide
DPPA	Diphenylphosphoryl azide
EDTA	Ethylenediamine tetraacetic acid
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.



8

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.



9

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.

HFIP	Hexafluoroisopropanol
HMP	Hexamethylphosphoric triamide (also HMPA, HMPTA)
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	1-Hydroxybenzotriazole
HOCt	1-Hydroxy-4-ethoxycarbonyl-1,2,3-triazole
NBS	*N*-Bromosuccinimide
NDMBA	*N*, *N*′-Dimethylbarbituric acid
NMM	*N*-Methylmorpholine

PAM	Phenylacetamidomethyl resin
PEG	Polyethylene glycol
PEGA	Polyethylene glycol dimethylacrylamide co-polymer
PPA	Polyphosphoric acid
PyBOP	1-Benzotriazolyloxy-tris-pyrrolidinophosphonium hexafluorophosphate
SDS	Sodium dodecylsulfate
TBAF	Tetrabutylammonium fluoride
TBTU	See remarks under HATU above
TEA	Triethylamine
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TFMSA	Trifluoromethanesulfonic acid
THF	Tetrahydrofuran
WSC1	Water soluble carbodiimide: 1-ethyl-3-(3′-dimethylaminopropyl)-carbodiimide hydrochloride (also EDC)

TECHNIQUES

These abbreviations may be combined in obvious ways, preferably with a hyphen, e.g. ESI-MS.

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
OD	Optical density
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
AVP	Arginine vasopressin
BK	Bradykinin; Bn has also been used but should be avoided as it has been used for bombesin too.
BSA	Bovine serum albumin
CCK	Cholecystokinin
CNS	Central nervous system
CSF	Cerebiospinal fluid
DNA	Deoxyribonucleic acid
ee	Enantiomeric excess
FSH	Follicle stimulating hormone
GH	Growth hormone
HIV	Human immunodeficiency virus
i.m.	intramuscular
i.p.	intraperitoneal
i.v.	intravenous
LHRH	Luteinizing hormone releasing hormone
MAP	Multiple antigen peptide (care! has also been used for "multiple-anchored protein")
NPY	Neuropeptide Y
MCD	Mast cell degranulating peptide
MIC	Minimum inhibitory concentration
OT	Oxytocin
PNA	Peptide nucleic acid
PTH	Parathyroid hormone
QSAR	Quantitative structure-activity relationship
RNA	Ribonucleic acid
SAR	Structure activity relationship
TASP	Template-assembled synthetic protein
TRH	Thyrotropin releasing hormone
VIP	Vasoactive intestinal peptide
VP	Vasopressin

Complete conformity to an agreed set of rules and conventions is an editorial pipe-dream. Peptide scientists come from too many scientific subcultures and lands for that. Some flexibility and allowance for taste can reluctantly be tolerated on pragmatic grounds. But the inventive and idiosyncratic approach has to be discouraged, not least because of the increasing use of electronic searching — for that reason particular care and conservatism is essential in the composition of abstracts. And the nearer we can get to uniformity, the fewer the confusions and the smoother the editorial and publishing process will be.

JOHN H JONES
Editor-in-Chief