Biochemistry

 Volume 5, Number 8 August 11, 1966

IUPAC–IUB Commission on Biochemical Nomenclature. Abbreviated Designation of Amino Acid Derivatives and Peptides. Tentative Rules*

These Tentative Rules are an attempt to achieve a broad systematization of various types of abbreviated notation already in use (*e.g.*, Brand and Edsall, *Ann. Rev. Biochem. 16*, 224 (1947); Report of the Committee on Abbreviations of the American Society of Biological Chemists, December 18, 1959; Report of the Committee on Nomenclature of the European Peptide Symposium, Pergamon Press, 1963, pp 261–269; "Tentative Rules for Abbreviations and Symbols of Chemical Names of Special Interest in Biological Chemistry," IUPAC Information Bulletin No. 20, July 1963; pp 16–18, 1965 revision of the latter[1]). They seek to reconcile the needs of the protein chemist, *i.e.*, indication of amino acid sequences, with those of persons concerned more with the chemical reactions of proteins and the synthesis of polypeptides, *i.e.*, the need of conveying more detailed chemical information in abbreviated form.[2]

1. General Considerations

1.1 The symbols chosen are derived from the trivial names or chemical names of the amino acids and of chemicals reacting with amino acids and polypeptides. For the sake of clarity, brevity, and listing in tables, the symbols have been, wherever possible, restricted to three letters, usually the first letters of the trivial names.

1.2 The symbols represent not only the names of the compounds but also their structural formulas.

1.3 The amino acid symbols by themselves represent the amino acids. The use of the symbols to represent the free amino acids is not recommended in textual material, but such use may occasionally be desirable in tables, diagrams, or figures. Residues of amino acids are represented by addition of hyphens in specific positions as indicated in Section 3.

1.4 Heteroatoms of amino acid residues (*e.g.*, O^{β} and S^{β} of serine and cysteine, respectively, N^{ϵ} of lysine, N^{α} of glycine, etc.) do not explicitly appear in the symbol; such features are understood to be encompassed by the abbreviation.

1.5 Amino acid symbols denote the L configuration unless otherwise indicated by D or DL appearing before the symbol and separated from it by a hyphen. When it is desired to make the number of amino acid residues appear in a clearer manner, the hyphen between the configurational prefix and the symbol may be omitted (see 5.3.1.1, *et seq.*). (Note: The designation of an amino acid residue as DL is inappropriate for compounds having another amino acid residue with an asymmetrical center.)

1.6 Structural formulas of complicated features may be used along with the abbreviated notation wherever necessary for clarity.

* Document CBN 1964-19/4 of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN), approved by CBN in July 1965 and published by permission of the International Union of Pure and Applied Chemistry, the International Union of Biochemistry, and the official publishers to the International Union of Pure and Applied Chemistry, Messrs. Butterworths Scientific Publications.

[1] This set of rules should be regarded as an extension, not a replacement, of Section 2 (Polypeptides and Proteins) of "Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry," *Biochemistry 5*, 1445 (1966).

[2] Comments on these Tentative Rules may be sent to any member of CBN: O. Hoffmann-Ostenhof (Chairman), W. E. Cohn (Secretary), A. E. Braunstein, J. S. Fruton, B. Keil, W. Klyne, C. Liébecq, B. G. Malmström, R. Schwyzer, E. C. Slater, or corresponding member, N. Tamiya.

2. Abbreviations for Amino Acids

2.1 *Common Amino Acids.*

Alanine	Ala	Leucine	Leu
Arginine	Arg	Lysine	Lys
Asparagine[3]	Asn[3]	Methionine	Met
Aspartic acid	Asp	Ornithine	Orn
Cysteine	Cys	Phenylalanine	Phe
Glutamic acid	Glu	Proline	Pro
Glutamine[3]	Gln[3]	Serine	Ser
Glycine	Gly	Threonine	Thr
Histidine	His	Tryptophan	Trp
Isoleucine	Ile	Tyrosine	Tyr
		Valine	Val

2.2 *Less Common Amino Acids.* Abbreviations for less common amino acids should be defined in each publication in which they appear. The following principles and notations are recommended.

2.2.1 HYDROXYAMINO ACIDS.

Hydroxylsine	Hyl
3-Hydroxyproline	3Hyp
4-Hydroxyproline	4Hyp

2.2.2 *allo*-AMINO ACIDS.

allo-Isoleucine	*a*Ile
allo-Hydroxylysine	*a*Hyl

2.2.3 "NOR" AMINO ACIDS. "Nor" (*e.g.*, in norvaline) is not used in its accepted sense (denoting a lower homolog) but to change the trivial name of a branched-chain compound into that of a straight-chain compound (compare with "iso," paragraph 2.1). "Nor" should therefore be treated as part of the trivial name without special emphasis.

Norvaline	Nva
Norleucine	Nle

2.2.4 HIGHER UNBRANCHED AMINO ACIDS. We suggest the following general rules for guidance in forming abbreviations: the functional prefix "amino" should be included in the symbol as the letter "A," diamino as "D."

The trivial name of the parent acid should be abbreviated to leave no more than two or three letters, as convenient and necessary for clarity. The word "acid" ("*-saure*," etc.) should be omitted from the symbol as carrying no significant information. Unless otherwise indicated (see paragraph below), single groups are in the α position, two amino groups in the α,ω (monocarboxylic acids) or α,α' positions (dicarboxylic acids). The location of amino acids in positions other than α and ω is shown by the appropriate Greek letter prefix.

Examples:

α-Aminobutyric acid	Abu
α-Aminoadipic acid	Aad
α-Aminopimelic acid	Apm
α,γ-Diaminobutyric acid	Dbu
α,β-Diaminopropionic acid	Dpr
α,α'-Diaminopimelic acid	Dpm
β-Alanine	βAla
ϵ-Aminocaproic acid	ϵAcp
β-Aminoadipic acid	αAad

2.2.5 N^{α}-ALKYLATED AMINO ACIDS. N^{α}-Methylamino acids are becoming more and more common (*e.g.*, in the large group of depsipeptides). This justifies special symbols.

Examples:

N-Methylglycine (sarcosine)	MeGly or Sar
N-Methylisoleucine	MeIle
N-Methylvaline, etc.	MeVal, etc.
N-Ethylglycine, etc.	EtGly, etc.

3. Amino Acid Residues

3.1 *Lack of Hydrogen on the α-Amino Group.* The α-amino group is understood to be at the left-hand side of the symbol when using hyphens, and—in special cases—at the point of the arrow when using arrows to indicate the direction of the peptide bond (–CO→NH–, –NH←CO–).

Examples:

–Gly: $-HNCH_2COOH$

–Ala: $-HNCH(CH_3)COOH$

>Gly or $^{\perp}$Gly: $>NCH_2COOH$

>Ala or $^{\perp}$Ala: $>NCH(CH_3)COOH$

3.2 *Lack of Hydroxyl on the α-Carboxyl Group.* The α-carboxyl group is always understood to be on the right-hand side of the symbol when using hyphens, and—in such special cases as 5.3.1.3—at the tail of the arrow when using arrows to indicate the direction of the peptide bond (–CO→NH, –NH←CO–).

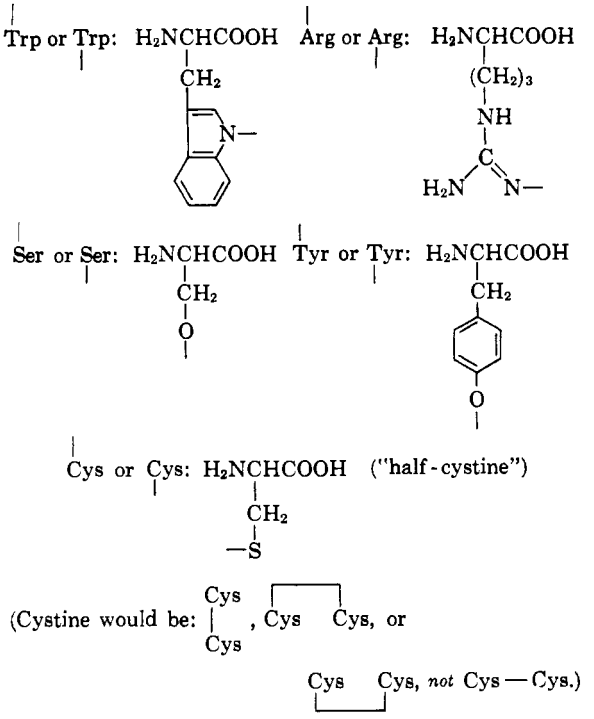
Example: Gly–: $H_2NCH_2CO–$.

3.3 *Lack of Hydrogen on Amino, Imino, Guanidino, Hydroxyl, and Thiol Functions in the Side Chain.*

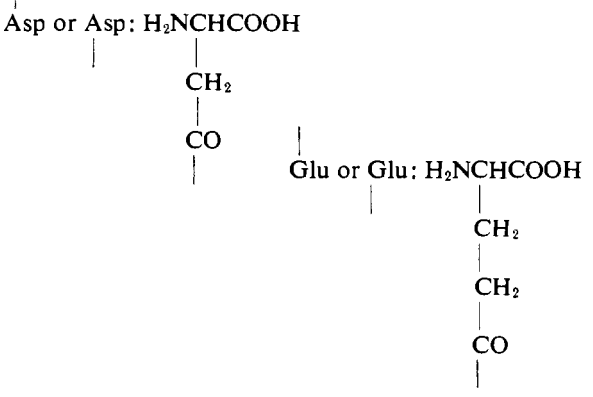
Lys or Lys: $H_2NCH(COOH)(CH_2)_4NH–$

His or His: $H_2NCH(COOH)CH_2$–(imidazole) or $H_2NCH(COOH)CH_2$–(imidazole)

[3] Asparagine and glutamine may also be denoted as Asp(NH_2) or Asp, and Glu(NH_2) or Glu, respectively.

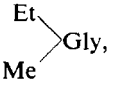


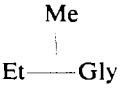
3.4 *Lack of Hydroxyl on Carboxyl Groups in the Side Chain.*



4. Substituted Amino Acids

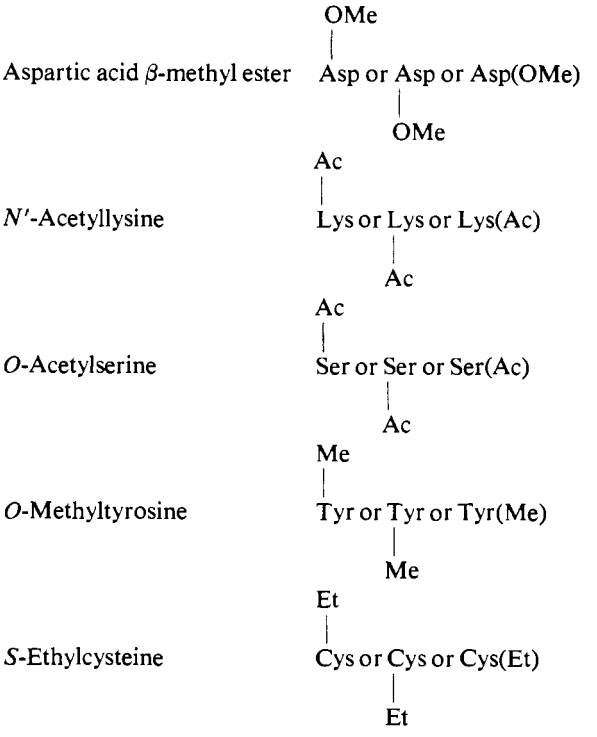
4.1 *Substitution in the α-Amino and α-Carboxyl Groups.* This follows logically from 3.1 and 3.2. The following examples will make the usage clear.

N-Acetylglycine	Ac-Gly
Glycine ethyl ester	Gly-OEt
N^{α}-Acetyllysine	Ac-Lys
Serine methyl ester	Ser-OMe
α-Ethyl *N*-acetylglutamate	Ac-Glu-OEt
Isoglutamine	Glu-NH_2
N-Ethyl-*N*-methylglycine	EtMeGly, 



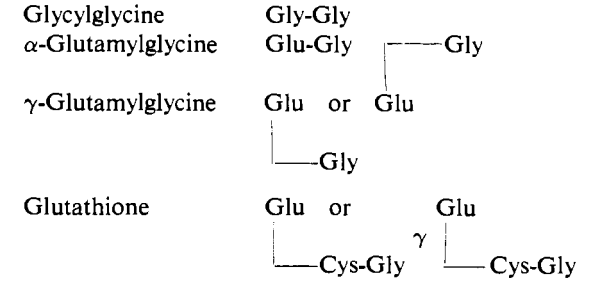
4.2 *Substitution in the Side Chain.* Side-chain substituents may be portrayed above or below the amino acid symbol, or by placing the symbol for the substituent in parentheses immediately after the amino acid symbol.

The use of parentheses should be reserved for a *single* symbol denoting a side-chain substituent. Where a more complex substituent is involved, it is recommended that the vertical stroke and a two-line abbreviation be used.

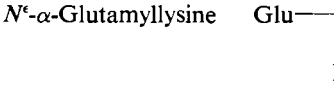


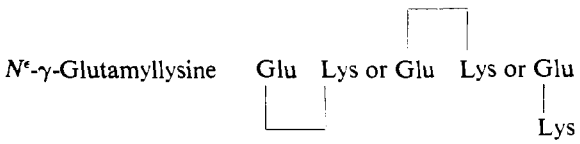
5. Polypeptides

5.1 *Polypeptide Chains.* Polypeptides may be dealt with in the same manner as substituted amino acids, *e.g.*

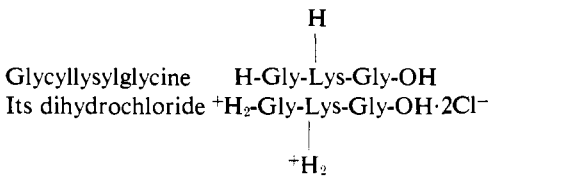


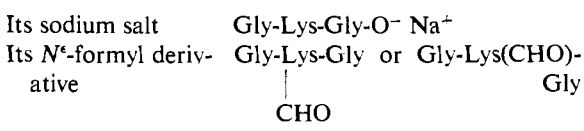
(Note that Glu would represent the corresponding Cys-Gly thiolester with a bond between the γ-carboxyl of glutamic acid and the thiol group of cysteine.)





The presence of free, substituted, or ionized functional groups can be represented (or stressed) as follows:

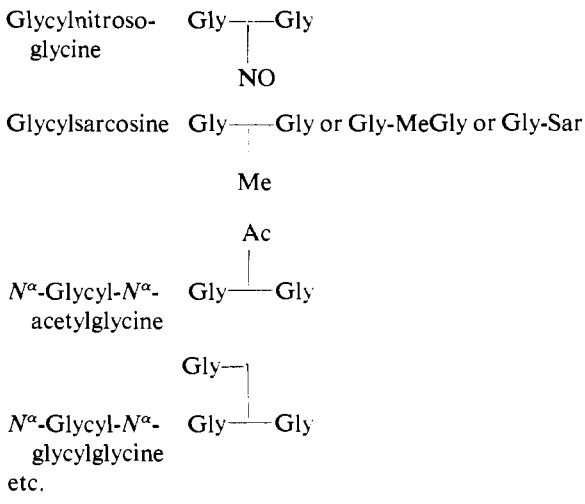




etc.

5.2 *Peptides Substituted at* N^{α}.

Examples:



etc.

5.3 *Cyclic Polypeptides.*

5.3.1 HOMODETIC CYCLIC POLYPEPTIDES (the ring consists of amino acid residues in peptide linkage only).

Three representations are possible:

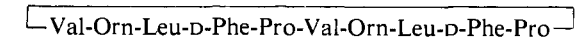
5.3.1.1 The sequence is formulated in the usual manner but placed in parentheses and preceded by (an italic) *cyclo.*

Example: Gramicidin S =

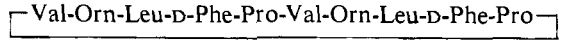
cyclo-(-Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-) or (see 1.5, sentence 2)

cyclo-(Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro-)

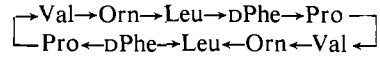
5.3.1.2 The terminal residues may be written on one line, as in 5.3.1.1, but joined by a lengthened bond. Using the same example in the two forms (see 1.5):



or



5.3.1.3 The residues are written on more than one line, in which case the CO→NH direction must be indicated by arrows, thus (in the optional manner of 1.5):

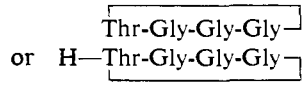


5.3.2 HETERODETIC CYCLIC POLYPEPTIDES (the ring consists of other residues in addition to amino acid residues in peptide linkage). These follow logically from the formulation of substituted amino acids.

Example:

etc.

Cyclic ester of threonylglycylglycylglycine

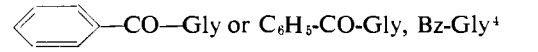


etc.

6. Abbreviations for Substituents

Groups substituted for hydrogen or for hydroxyl may be indicated either by their structural formulas or by (accepted) abbreviations, *e.g.*

Benzoylglycine or (hippuric acid):



Glycine methyl ester: $Gly\text{-}OCH_3$ or Gly-OMe

Suggestions for the abbreviations of protecting groups common in polypeptide chemistry follow. All such symbols (except those allowed by individual journals, *e.g.*, Bz, Ac, Ph, Me, Et, etc.) *should be defined in each paper.* Although symbolization by the use of capital letters throughout would be useful for distinguishing these symbols from those of the amino acids, we propose the use of one capital letter followed by lower case letters in order not to increase the flood of capital letter abbreviations in biological chemistry.

6.1 *N-Protecting Groups of the Urethan Type.*

Benzyloxycarbonyl	Z–
p-Nitrobenzyloxycarbonyl	$Z(NO_2)$–
p-Bromobenzyloxycarbonyl	Z(Br)–

[4] Bz– is the symbol generally used for *benzoyl* in organic chemistry. Its use for *benzyl* (which has become rather common in polypeptide chemistry) should be discouraged. We propose *Bzl*– for benzyl.

p-Methoxybenzyloxycarbonyl	Z(OMe)–
p-Methoxyphenylazobenzyloxycarbonyl	Mz–
p-Phenylazobenzyloxycarbonyl	Pz–
t-Butyloxycarbonyl	Boc–
Cyclopentyloxycarbonyl	Poc–

6.2 *Other N-Protecting Groups.*

Acetyl	Ac–
Benzoyl	Bz–
Tosyl	Tos–
Trifluoroacetyl	Tfa–
Phthalyl	Pht–
Benzyl	Bzl–
Trityl	Trt–
Tetrahydropyranyl	Thp–
Dinitrophenyl	Dnp–
Benzylthiomethyl	Btm–
o-Nitrophenylsulfenyl	Nps–

6.3 *Carboxyl-Protecting Groups.*

Methoxy (methyl ester)	–OMe
Ethoxy (ethyl ester)	–OEt
Tertiary butoxy (*trty*-butyl ester)	–OBu^t
Benzyloxy (benzyl ester)	–OBzl
Diphenylmethoxy (benzhydryl ester)	–OBzh
p-Nitrophenoxy (*p*-nitrophenyl ester)	–ONp
Phenylthio (phenyl thiolester)	–SPh
p-Nitrophenylthio	–SNp
Cyanomethoxy	–OCH$_2$CN

Note: Contrary to the symbols for amino acid residues, the position of the dashes in the symbols for substituents carries no significant information.

A New N-Terminal Blocking Group Involving a Schiff Base in Hemoglobin A_{Ic}*

W. R. Holmquist and W. A. Schroeder

ABSTRACT: Hemoglobin A_{Ic} is a minor hemoglobin component in the erythrocytes of normal adult humans. Chemically, it is the condensation product, a Schiff base, between one molecule of hemoglobin A and one molecule of a ketone or aldehyde R=O. The pK_a of this Schiff base is 6.64. The point of linkage of R=O to hemoglobin A to form hemoglobin A_{Ic} is at the N terminus of one of the two β chains. Other than this, no difference has been found between the primary amino acid sequence of the α and β chains of hemoglobins A_{Ic} and A.

The ketone or aldehyde R=O has a molecular weight of approximately 280, is not an aromatic aldehyde, is not a steroid, and does not contain phosphorus, carbohydrates, or amino acids. Esterified nonketo acyl groups of less than five carbon atoms are absent. It is probable that R=O contains neither nitrogen nor free carboxyl groups. There is as yet no basis for excluding the possibility that R=O is a long-chain aliphatic aldehyde or ketone. Methods are described for isolating A_{Ic} as well as smaller peptides which contain the blocking group R, such as R-Val-His. The lability of the Schiff base linkage at alkaline pH has limited the experimental procedures. This lability can be circumvented by reduction with $NaBH_4$. A practical method for calculating ionization constants from paper electrophoretic data is described and used to obtain the pK_a of the Schiff base nitrogen in A_{Ic}.

Hemoglobin A_{II}, which has the subunit structure $\alpha_2\beta_2$, is the major hemoglobin protein of most adult humans. The major component has been variously designated as A, A_{II}, A_0, and A_1. Partial correlations between different systems of nomenclature are given by Schnek and Schroeder (1961), Huehns and Shooter (1965), and Schroeder and Jones (1965). The only sure guide to the identity of a hemoglobin is an understanding of the methods by which it was isolated and characterized. This paper will designate the main component as hemoglobin A_{II} to be consistent with earlier publications from this laboratory. The amino acid sequences of both the α and β polypeptide chains and much of their secondary, tertiary, and quaternary structures are known.[1] Whereas there is only a single major component, A_{II}, several minor components exist,

* From the Division of Chemistry and Chemical Engineering,† California Institute of Technology, Pasadena, California 91109. *Received March 7, 1966.* This work was supported in part by grants (HE-02558 and GM1262) from the National Institutes of Health, U. S. Public Health Service, the National Science Foundation, and the Du Pont Chemical Co.

† Contribution No. 3355.

[1] References to the original literature can be found in the reviews by Huehns and Shooter (1965) and Schroeder and Jones (1965).