IUPAC–IUB Commission on Biochemical Nomenclature. Rules for Naming Synthetic Modifications of Natural Peptides. Tentative Rules*

During the last few years, chemists have made many compounds that are variants of naturally occurring peptides (or proteins) having trivial names. Therefore, the need has arisen for "semitrivial" names to designate these variants without the necessity of designating every residue in the chain.

After discussion with active workers in the field, the following proposals are put forward; they are based on the names used by du Vigneaud and his collaborators (*cf.* Bodanszky and du Vigneaud, *J. Am. Chem. Soc. 81*, 1258 (1959); Popenoe, Lawler, and du Vigneaud, *J. Am. Chem. Soc. 74*, 3713 (1952)) and the symbols introduced by Schwyzer *et al.* (*cf.* Rittel, Iselin, Kappeler, Riniker, and Schwyzer, *Angew. Chem. 69*, 179 (1957); Riniker and Schwyzer, *Helv. Chim. Acta 44*, 685 (1961); see also *J. Biol. Chem. 241*, 2491 (1966)).

This draft has been prepared by a subcommittee consisting of J. S. Fruton, W. Klyne, and R. Schwyzer. The subcommittee is greatly indebted to many colleagues for helpful suggestions, notably to V. du Vigneaud, J. Rudinger, H. B. F. Dixon, and P. E. Verkade, chairman of the IUPAC Commission on Organic Nomenclature.

These proposals are *not* suitable for application to "abnormal" links in a peptide sequence; *e.g.*, to disulfide links or γ-peptide links. They are *only* suitable for modifications involving normal α-peptide links.

Rules

1. Replacement

In a polypeptide of trivial name X, if the qth amino acid residue (starting from the N-terminal end of the chain) is *replaced* by the amino acid residue Abc, the semitrivial name of the modified polypeptide is [q-amino acid]-X and the abbreviated form, chiefly for use in tables, is [Abc^{q}]-X.

Examples

[8-Citrulline]-vasopressin, [Cit^{8}]-vasopressin (Bodanszky and Birkhimer, *J. Am. Chem. Soc. 84*, 4963 (1962)). [5-Isoleucine, 7-alanine]-hypertensin II, [Ile^{5}, Ala^{7}]-hypertensin II (Seu, Smeby, and Bumpus, *J. Am. Chem. Soc. 84*, 3883 (1962)).

Comments

(a) In the full name, the replacement amino acid is designated by its *own full name*, *not* the name of its radical (*cf.* 4 below). This name, and the position of replacement, are given in square brackets [ ], as for isotopic replacement.

(b) In the abbreviated form, the amino acid residues are designated by the standard three-letter symbols (*J. Biol. Chem. 241*, 527, 2491 (1966); *Biochim. Biophys. Acta 121*, 1 (1966)), the first letter *only* being a capital, in square brackets [ ].

(c) In the abbreviated form, the *position* of substitution is indicated in a special fashion, *i.e.*, by a superior numeral ^{q}, to indicate that it is a *residue*, not an individual atom, that is being replaced and also for the reason indicated in comment d.

(d) The nature of the residue replaced is *not* designated in either the full or the abbreviated name. This is contrary to a general principle of organic nomenclature requiring that an atom (or group) that is replaced should (unless it is hydrogen) be clearly designated, as in 2-amino-2-deoxy-D-glucose. It has been decided *not* to insist on the designation of the residue replaced in these semitrivial names in order to keep the names as short as possible, and because the form of nomenclature in Rule 1 clearly differs from ordinary substitution nomenclature.

(e) A partial analogy may be drawn with the form used for isotopic replacement, where the isotope symbol is indicated in square brackets before the name.

(f) The replacement of an amino acid residue by its enantiomer may be shown logically by the application of this rule as follows: the replacement in X of L-alanine at position 7 by D-alanine results in [7-D-alanine]-X with the abbreviation [D-Ala^{7}]-X. An example may be found in Boissonnas, Guttman, and Pless (*Experientia 22* (1966)), dealing with the D-Ser^{1} ..... derivative of β-corticotropin; the natural compound has L-serine in position 1. Another example is the [α-D-Asp^{1}]-

* Document of the IUPAC–IUB Commission on Biochemical Nomenclature (CBN), approved by CBN in July 1966 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. 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. C. Malmström, R. Schwyzer, E. C. Slater, or corresponding member, N. Tamiya. Reprints of these Tentative Rules may be obtained from Waldo E. Cohn, Director, NAS–NRC Office of Biochemical Nomenclature, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tenn. 37830.

hypertensin II of Riniker and Schwyzer (*Helv. Chim. Acta 47*, 2357 (1964)).

2. Extension

The compounds obtained by the extension of polypeptide X at either (a) the N-terminal end or (b) the C-terminal end are designated by the kinds of names and abbreviations shown below; these are in accordance with the general principles of polypeptide nomenclature (*J. Biol. Chem. 241*, 2491 (1966); *Biochim. Biophys. Acta 121*, 1 (1966)).

Examples

(a) Extension at N-terminal end:

	Aminoacyl-X	Abc-X
e.g.,	Valyl-X	Val-X
or	Valylglycyl-X	Val-Gly-X (for extension by two residues)

(b) Extension at C-terminal end:

	X-yl-amino acid	X-yl-Abc
e.g.,	X-yl-leucine	X-yl-Leu

(where X-yl is the trivial name of polypeptide X with the ending -yl).

Comment

This rule is not applicable to the extension at the C-terminal of natural peptides having a terminal α-carboxamido group, as in the case of oxytocin or α-melanophore-stimulating hormone (α-MSH). It has been suggested that new names be given to the peptides having a free terminal α-carboxyl group (*e.g.*, oxytocinoic acid) and that extension at the C-terminal end be denoted as in the example given above (*e.g.*, oxytocinoyl-Abc).

3. Insertion

The compound obtained by the *insertion* of an additional amino acid residue Abc in the position between the qth and (q + 1)th residues of a polypeptide X is named qa-endo-amino acid-X (abbreviated form, endo-Abc^{qa}-X).

Example

4a-endo-tyrosine-hypertensin II; endo-Tyr^{4a}-hypertensin II.

Comments

(a) This form has analogies in other fields where endo implies the insertion of something into a structure (*e.g.*, endo-methylene). The prefix or index qa is based on analogies with the steroids where the atoms inserted in a ring after atom no. q are designated qa, qb, etc.

(b) The prefix homo is *not* suitable for designating the insertion of a whole residue, since it is commonly used to modify the names of *individual* amino acids, *e.g.*, homoserine.

(c) Multiple insertions, and insertion of two or more residues together in the same place in the chain, are shown by a logical extension of this rule. For example, the insertion into the polypeptide X of threonine between residues 4 and 5, and of valine and glycine (*in that order*) between residues 6 and 7, is shown by the name "endo-4a-threonine,6a-valine,6b-glycine-X" and the abbreviation "endo-Thr^{4a}, (Val^{6a}-Gly^{6b})-X."

4. Removal

The compound obtained by the formal *removal* of an amino acid residue from a polypeptide X in position q is designated by the name des-q-amino acid-X, abbreviated des-Abc^{q}-X.

Example

des-7-proline-oxytocin; des-Pro^{7}-oxytocin (Jacquenoud and Boissonas, *Helv. Chim. Acta 45*, 1462 (1962)).

Comment

(a) Removal of a whole residue is indicated as is the removal of a ring in steroids, *e.g.*, des-A-androstane.

(b) "de" is *not* suitable as a prefix because it is easily confused, in speaking, with D (for configuration).

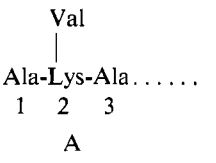
5. Substitution Forming a Side Chain

The compound formed by the substitution of an additional amino acid residue as a side chain into a polypeptide X is named by applying the ordinary rules of nomenclature to the trivial name.

(a) If the substitution is on a side-chain *amino* group of polypeptide X, the name of the additional amino *residue* is written (with the termination "yl") and prefixed by symbols indicating the position of substitution (residue number and atom).

Example

An imaginary compound (A)

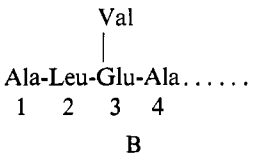


in which a valyl group is substituted at the ϵ-amino group of lysine at position 2 of the chain of a peptide X is named $N^{\epsilon 2}$-valyl-X (abbreviated $N^{\epsilon 2}$-Val-X).

(b) If the substitution is on a side-chain *carboxyl* group of polypeptide X, the additional amino acid having a free α-carboxyl group, the substituted derivative is named by specifying the position of substitution (residue number, and atom) and is given the designation "X-yl-amino acid."

Example

An imaginary compound (B)



in which a valine residue is substituted into the δ-carboxyl group of glutamic acid in position 3 of the chain of a peptide X would be named $C^{\delta 3}$-X-yl-valine (abbreviated $C^{\delta 3}$-X-yl-Val).

Comment

Note the importance of clear distinction from *replacement* as indicated in Rule 1.

6. Partial Sequences (Fragments)

Polypeptide sequences that form fragments of a longer sequence that already has a trivial name may be designated as follows. The *trivial name* is followed by numbers giving the positions of the first and last amino acids, and then the usual *Greek* designation giving the number of amino acid units in the fragment; thus

Trivial name (-X–Y-).....peptide.

Example: from α-MSH

Ac-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-
1 2 3 4 5 6 7 8 9 10 11 12

-Val-NH_2 α-MSH
13

we may have
Met-Glu-His-Phe-Arg-Trp-Gly
4 10
α-MSH-(4–10)-heptapeptide
and
His-Phe-Arg-Lys-Pro-Val-NH_2
6 8 11 13
α-MSH-(6–8)-(11–13)-hexapeptide amide

The last example illustrates the nomenclature for a composition sequence of two fragments, and also for an amide-terminal group.

SUMMARY WITH EXAMPLES: The Systematic Application of These Principles to the Name of an Imaginary Pentapeptide "Iupaciubin"[a] May Illustrate the Symbolism.

Rule	Operation	Short Name	Structure
	(Fundamental name)	Iupaciubin	1 2 3 4 5 Ala-Lys-Glu-Tyr-Leu
1.	Replacement	[Phe^4]-iupaciubin[b]	4 Ala-Lys-Glu-Phe-Leu
2a.	Extension (N terminal)	Arginyl-iupaciubin, Arg-iupaciubin	1 5 Arg-Ala-Lys-Glu-Tyr-Leu
2b.	Extension (C terminal)	Iupaciubyl-methionine, iupaciubyl-Met	1 5 Ala-Lys-Glu-Tyr-Leu-Met
3.	Insertion	Endo-Thr^{2a}-iupaciubin	2 2a 3 Ala-Lys-Thr-Glu-Tyr-Leu
4.	Removal	Des-Glu^3-iupaciubin	2 4 Ala-Lys-Tyr-Leu
5a.	Side-chain substitution on amino group	$N^{\epsilon 2}$-Val-iupaciubin	Val |ε Ala-Lys-Glu-Tyr-Leu 2
5b.	Side-chain substitution on carboxyl group	C-$^{\delta 3}$-Iupaciubyl-valine	Val |δ Ala-Lys-Glu-Tyr-Leu 3
6.	Partial sequence	Iupaciubin-(2–4)-tripeptide	2 3 4 Lys-Glu-Tyr

[a] To symbolize the harmonious cooperation of IUPAC and IUB. [b] Note that only for *replacement* are square brackets required.