Biochemistry

Volume 7, Number 2 February 12, 1968

IUPAC–IUB Combined Commission on Biochemical Nomenclature. Tentative Rules for Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids)*

The numerous studies on the physical, chemical, and biological properties of synthetic polypeptides have brought with them different ways of describing, in abbreviated form, these products, whose structures are often incompletely known. The use of a variety of nomenclatures complicates the literature; hence, a consistent and clearly defined system for naming such polypeptides is desirable. The proposals set forth here, which represent the consensus of many discussions and suggestions, should aid in systematizing the nomenclature of a wide variety of synthetic polypeptides.

These proposals are based in large part on the abbreviated nomenclature devised by T. J. Gill III[1] and by M. Sela and others.[2] They utilize the abbreviations and conventions set forth in section 2 of *Revised Tentative Rules for Abbreviations and Symbols of Chemical Names of Special Interest in Biological Chemistry*[3] and in *Abbreviated Designation of Amino Acid Derivatives and Peptides*,[4] and they add only those terms or conventions needed for the specification of polymers but not encompassed by these schemes. The abbreviations and conventions of the previous Tentative Rules[3,4] used in this nomenclature system are summarized as follows.

The abbreviations of the amino acid residues and their derivatives or modifications are those indicated in the Tentative Rules[3,4] or formulated according to the principles set out in them. Hyphens or commas between the symbols for residues or groups of residues mean known or unknown sequence, respectively, and involve only the α-NH_2 and α-COOH groups (the peptide link). (Commas may be omitted when other symbols (*e.g.*, subscripts or superscripts) separate symbols in unknown sequences.) Vertical strokes indicate covalent bonds involving functional groups or the remaining hydrogen atom of the peptide bond, depending upon their placement.[4] L-Amino acids and α-peptide links, read from left (NH_2 terminus) to right (COOH terminus), are assumed unless indicated otherwise.[3,4]

Definitions

1. Linear polymer: all amino acids are in an unbranched chain.

2. Graft polymer: one or more polymeric segments are linked to the functional group of a linear chain, thus creating a branch or branches. (Functional groups

---

* Document of the IUPAC–IUB Commission on Biochemical Nomenclature, approved by CBN, IUPAC, and IUB in Oct 1967, and published by permission of the International Union of Pure and Applied Chemistry, the International Union of Biochemistry, and the official publishers of IUPAC, 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. G. Malmström, R. Schwyzer, C. E. 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, Biology Division, Oak Ridge National Laboratory, Box Y, Oak Ridge, Tenn. 37830.

[1] *Biopolymers 2,* 283 (1964). See also *J. Biol. Chem. 240,* 3227 (1965); *Biochim. Biophys. Acta 124,* 374 (1966).

[2] *Advan. Immunol. 5,* 30 (1966).

[3] *J. Biol. Chem. 241,* 527 (1966); *Biochemistry 5,* 1445 (1966); *Biochem. J. 101,* 1 (1966); *Virology 29,* 480 (1966); *Arch. Biochem. Biophys. 115,* 1 (1966); *European J. Biochem. 1,* 259 (1967); *Z. Physiol. Chem. 348,* 245 (1967).

[4] *J. Biol. Chem. 241,* 2491 (1966); *Biochemistry 5,* 2485 (1966); *Biochim. Biophys. Acta, 121,* 1 (1966); *Biochem. J. 102,* 23 (1967); *Arch. Biochem. Biophys. 121,* 1 (1967); *European J. Biochem. 1,* 375 (1967); *Z. Physiol. Chem. 348,* 256 (1967); *Bull. Soc. Chim. Biol. 49,* 121 (1967).

include ϵ-NH$_2$, β- or γ-COOH, etc., and the remaining hydrogen atom of an α-peptide link.)

3. Block polymer: two or more linear or graft polymeric segments are linked to form a larger polymer.

4. Polymeric segment: a polymer that forms a distinct part of a larger polymer (*e.g.*, a block or graft polymer may contain several polymeric segments).

Rules

1. Polymeric segments that contain more than one amino acid symbol are enclosed in parentheses or brackets. A superscript outside of the parentheses indicates the number of repeating sequences per 100 residues of polymer, and it is given to the first decimal place.

2. The molar percentage of a single type of amino acid residue within a copolymer, derived from the amino acid analysis and assuming copolymerization, is indicated by a superscript attached to the symbol of the residue. The molar percentages are given in whole numbers and should total 99–101 %.

3. Designation of Polymeric Segments or Linear Polymers. The prefix "poly" or the subscript "*n*" indicates "polymer of." It is attached to each main chain and is repeated in each polymeric segment within a larger polymer unless there is sufficient indication of size and of structure to make this repetition unnecessary. For example, poly Glu or (Glu)$_n$ for polyglutamic acid, and (Glu)$_{10}$ for a decapeptide of glutamic acid.

COMMENT. "*n*" replaces the "p" as originally, but no longer (1967 revision), used in the polymer nomenclature scheme devised by the IUPAC Subcommission on the Nomenclature of Macromolecules.[5] It is used in designating polynucleotides (see section 5 of *Abbreviations and Symbols of Chemical Names of Special Interest in Biological Chemistry, loc cit.*), and it is chosen in place of "p" in order to avoid confusion with the "p" used for a phosphoric acid residue in the latter scheme. The "*n*" may be replaced by a definite number (*e.g.*, 10 above), an average (*e.g.*, $\overline{10}$), or a range (*e.g.*, 8–12), as appropriate.

4. Designation of Branches and Branch Points. Branches (side chains) connected to the main chain can be designated in one of three ways: by a vertical line joining the main chain and the branch (side chain); by an extended bond joining the appropriate residues with the main chain written first; or by a horizontal double dash (not preferred). *The branch points* are indicated by the origin and terminus of the vertical line. If the origin is *unknown*, the line originates at the "p" in "poly," if "poly" is used, or at the first parenthesis (bracket), if the subscript "*n*" is used (see Rule 1). If the origin is *known*, the line originates: (a) vertically at the initial letter of the appropriate symbol, if functional groups other than α-NH$_2$ or α-COOH residues are involved; (b) vertically at the position of the appropriate link, if substitution for the remaining hydrogen atom of a peptide link is involved; or (c) horizontally to the left or right of the symbol, respectively, if α-NH$_2$ or α-COOH groups are involved. The same rules apply to the termination of the line. Thus, the linkage between a side-chain functional group and an α-NH$_2$ or α-COOH group in the main chain is indicated by two perpendicular lines with the vertical line originating in the functional group and the horizontal line in the α-NH$_2$ or α-COOH group. A number in parentheses lying beside the line indicates the number of such links per 100 residues of polymer, if known.

COMMENT. A limitation of the double dash as a connecting link lies in its inability to originate or to terminate definitively in a specific residue. Either the arrangement of the symbols must be such that connected ones are adjacent, or the information must be given independently.

Examples

1. Simple homopolymer:

poly Ala or (Ala)$_n$

2. Linear copolymer, random sequence, composition unknown:

poly DLAla,Lys or (DLAla,Lys)$_n$

3. Linear copolymer, alternating sequence, composition unknown:

poly DLAla-Lys or (DLAla-Lys)$_n$

4. Linear sequence of unknown order (composition: 56 % Glu, 38 % Lys, 6 % Tyr (Σ = 100 %)):

a. poly Glu^{56}Lys^{38}Tyr^{6} or (Glu^{56}Lys^{38}Tyr^{6})$_n$ (all L)
b. poly DGlu^{56}DLys^{38}Tyr^{6} (only Tyr is L)
c. poly DLGlu^{56}Lys^{38}DTyr^{6} (Glu is DL, Tyr is D)

5. Block polymer of poly Glu combined through the α-COOH terminus to the α-NH$_2$ terminus of poly Lys (composition: 56 % Glu, 44 % Lys (Σ = 100 %)):

poly Glu^{56}-poly Lys^{44} or (Glu^{56})$_n$-(Lys^{44})$_n$

6. a. Known, repeating sequence (a polymer of Glu-Lys-Lys-Tyr):

poly Glu-Lys$_2$-Tyr or (Glu-Lys$_2$-Tyr)$_n$

b. Known, repeating sequences within each of two constituent blocks of a linear polymer (composition: 37.5 % Glu, 25 % Lys, 25 % Tyr, 12.5 % Ala (Σ = 100 %)):

(poly Glu-Lys)^{25}-(poly Ala-Tyr$_2$-Glu)$^{12.5}$

or

(Glu-Lys)$^{25}_n$-(Ala-Tyr$_2$-Glu)$^{12,5}_n$

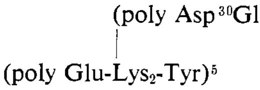
(The connection between the polymeric segments shown here is from the α-COOH of Lys to the α-NH$_2$ of Ala. Origin or termination in any other residue or

---

[5] *J. Polymer Sci. 8*, 257 (1952); 1967 revision in press.

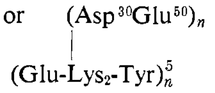
functional group can be shown by rearranging the order of residues and by the orientation of the connecting line at its origin and terminus (see examples 7–9).)

c. Known, repeating sequence in the main chain connected by the ε-NH$_2$ of a lysine (which of the two is not known) to an unknown point in an unknown sequence in the side chain (composition: 30% Asp, 55% Glu, 10% Lys, 5% Tyr (Σ = 100%)):



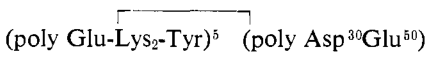
$$\begin{array}{l}(\text{poly Asp}^{30}\text{Glu}^{50})\\ \quad|\\ (\text{poly Glu-Lys}_2\text{-Tyr})^5\end{array}$$

or



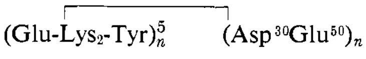
$$\begin{array}{l}(\text{Asp}^{30}\text{Glu}^{50})_n\\ \quad|\\ (\text{Glu-Lys}_2\text{-Tyr})_n^5\end{array}$$

or



$$(\text{poly Glu-Lys}_2\text{-Tyr})^5\ (\text{poly Asp}^{30}\text{Glu}^{50})$$

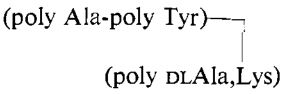
or



$$(\text{Glu-Lys}_2\text{-Tyr})_n^5\quad(\text{Asp}^{30}\text{Glu}^{50})_n$$

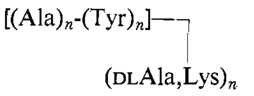
(Note: The double-hyphen system is not applicable here.) If the lysine residue connected to the side chain were known, the main chain would be written in the form, *e.g.*, (poly Glu-Lys-Lys-Tyr)^5 or (poly Glu-Lys-Lys-Tyr)^5, as appropriate.

7. Graft polymer with the main chain of DL-alanine and L-lysine connected through the ε-NH$_2$ group of lysine to the α-COOH group of L-tyrosine in the side chain, which consists of a block polymer of L-tyrosine and L-alanine (no analytical data for the main chain):



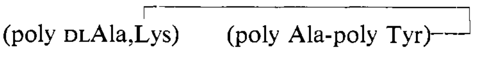
$$\begin{array}{l}(\text{poly Ala-poly Tyr})\text{—}\\ (\text{poly DLAla,Lys})\end{array}$$

or



$$\begin{array}{l}[(\text{Ala})_n\text{-}(\text{Tyr})_n]\text{—}\\ (\text{DLAla,Lys})_n\end{array}$$

or



$$(\text{poly DLAla,Lys})\quad(\text{poly Ala-poly Tyr})\text{—}$$

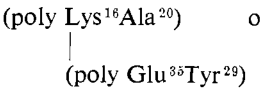
or

(poly DLAla,Lys)--(poly Ala-poly Tyr)

(Note: The points of attachment of Lys and Tyr cannot be specified in the last example.)

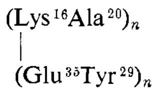
8. Graft polymer with an unknown sequence in the main chain and in the side chain (composition: 16% Lys, 20% Ala, 35% Glu, 29% Tyr (Σ = 100%)):

a. Number and position of the points of attachment in the main chain unknown, but terminating in the lysine residues of the side chain:



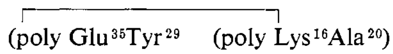
$$\begin{array}{l}(\text{poly Lys}^{16}\text{Ala}^{20})\\ \quad|\\ (\text{poly Glu}^{35}\text{Tyr}^{29})\end{array}$$

or



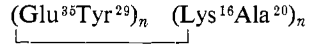
$$\begin{array}{l}(\text{Lys}^{16}\text{Ala}^{20})_n\\ \quad|\\ (\text{Glu}^{35}\text{Tyr}^{29})_n\end{array}$$

or



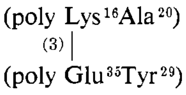
$$(\text{poly Glu}^{35}\text{Tyr}^{29}\quad(\text{poly Lys}^{16}\text{Ala}^{20})$$

or



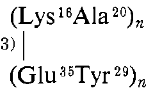
$$(\text{Glu}^{35}\text{Tyr}^{29})_n\quad(\text{Lys}^{16}\text{Ala}^{20})_n$$

b. Same, but attachments are three in number and connect the ε-NH$_2$ groups of the lysine residues in the side chain and the γ-COOH groups of the glutamic acid residues in the main chain:



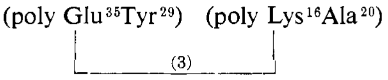
$$\begin{array}{l}(\text{poly Lys}^{16}\text{Ala}^{20})\\ \quad{}^{(3)}|\\ (\text{poly Glu}^{35}\text{Tyr}^{29})\end{array}$$

or



$$\begin{array}{l}(\text{Lys}^{16}\text{Ala}^{20})_n\\ \quad{}^{(3)}|\\ (\text{Glu}^{35}\text{Tyr}^{29})_n\end{array}$$

or

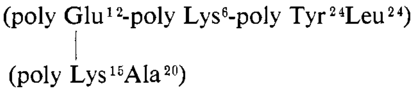


$$(\text{poly Glu}^{35}\text{Tyr}^{29})\quad(\text{poly Lys}^{16}\text{Ala}^{20})\quad(3)$$

or

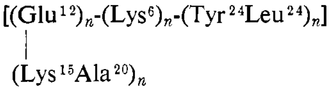
$(\text{poly Tyr}^{29}\text{Glu}^{35})\overset{\gamma\ \epsilon}{\underset{(3)}{-}}(\text{poly Lys}^{16}\text{Ala}^{20})$

9. Graft polymer with a block polymer and an unknown sequence in the side chain (upper) attached to an unknown sequence in the main chain (lower); the points of attachment are between the γ-COOH groups of glutamic acid in the side chain and the ε-NH$_2$ groups of lysine in the main chain (composition: 12% Glu, 21% Lys, 24% Tyr, 24% Leu, 20% Ala (Σ = 101%)):



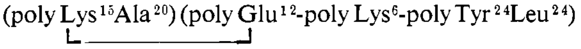
$$\begin{array}{l}(\text{poly Glu}^{12}\text{-poly Lys}^{6}\text{-poly Tyr}^{24}\text{Leu}^{24})\\ \quad|\\ (\text{poly Lys}^{15}\text{Ala}^{20})\end{array}$$

or



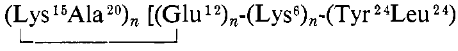
$$\begin{array}{l}[(\text{Glu}^{12})_n\text{-}(\text{Lys}^{6})_n\text{-}(\text{Tyr}^{24}\text{Leu}^{24})_n]\\ \quad|\\ (\text{Lys}^{15}\text{Ala}^{20})_n\end{array}$$

or



$$(\text{poly Lys}^{15}\text{Ala}^{20})\,(\text{poly Glu}^{12}\text{-poly Lys}^{6}\text{-poly Tyr}^{24}\text{Leu}^{24})$$

or



$$(\text{Lys}^{15}\text{Ala}^{20})_n\ [(\text{Glu}^{12})_n\text{-}(\text{Lys}^{6})_n\text{-}(\text{Tyr}^{24}\text{Leu}^{24})_n]$$

or

(poly Ala^{20}Tyr^{15})-εγ-(poly Glu^{12}-poly Lys^{6}-poly Tyr^{24}Leu^{24})