Representation of Macromolecules and Polymers of Biological Importance*

WALDO E. COHN**
Biology Division, Oak Ridge National Laboratory,*** Oak Ridge, Tenn. 37830

Received May 19, 1969

The structural analysis of natural polymers—proteins, polysaccharides, and nucleic acids—has progressed to the point where long sequences, over 100 in many cases, of nonidentical monomeric units can be accurately positioned. Since even the accepted trivial names of the amino acids, monosaccharides, and nucleotides derived by hydrolysis, and thus considered to be the base units of the polymers, are too long for this situation, 3- or 1-letter contractions are employed in horizontal arrays. Known sequences are indicated by hyphens representing peptide, glycoside, or phosphodiester links, and unknown ones by commas between residues. Substitution on functional groups other than those in the linear main chain are symbolized by vertical bonds at the appropriate symbols. These conventions for the natural polymers have been extended to the polynucleotides recently synthesized by chemical and enzymic means, where homopolymers, repeating copolymers, and interchain associations are encountered. The source-based name, "polymer of," is reduced to either the prefix "poly" or the subscript suffix "n" (e.g., poly A or A_n). Known and random sequences in copolymers utilize the hyphen and comma, respectively [e.g., $(A\text{-}U)_n$ and $(A,U)_n$]. Interchain association (noncovalent) is shown by the center dot [as in $(A)_n \cdot (U)_m$], nonassociation by a plus sign [$(A)_n + (G)_m$], and indefiniteness by a comma between the symbols defining each chain.

Biochemists deal with a variety of chemical entities, from simple ions of the elements up to heterogeneous particles of microscopically visible size. For the former, the classical names and symbols suffice; for the latter, descriptive terms, trivial names, and semisystematic names are coined, become common among specialists, and end up in dictionaries and nomenclature documents. In between there are the organic substances whose complexity and variety require systematic nomenclature, and such has been created by the organic chemist—especially by the IUPAC Commission on Organic Nomenclature and by *Chemical Abstracts*. With relatively few exceptions, the biochemist respects and uses these names; when too cumbersome (in his mind) for repeated use in the text of his papers, he tends to coin *ad hoc* trivial names or contractions or other multi-letter terms rather than resort to the organic chemist's I, II, III, etc. Often he has recourse to the trivial name, sanctioned by biological source or the discoverer's whim or some intriguing property, and his *ad hoc* symbol may be derived from this rather than from the formal organic name.

The most pressing need for an adequate symbolism comes when the biochemist *cum* chemist faces those two special groups of macromolecules, the proteins (or polypeptides) and the nucleic acids (or polynucleotides). Polysaccharides constitute a third, less well-defined group. Detailed chemical knowledge of the primary structure of these substances, which vary in molecular weight from a few hundred to many millions, has come about in the past few decades, beginning with proteins, and it is from conventions first proposed in this area about twenty years ago that the present biochemical nomenclature of polymers has grown.

The fact that biochemical polymer nomenclature began with the naturally-occurring proteins—those essentially "ordered irregular" polymers that, as enzymes, almost define biological chemistry—explains why representation (rather than nomenclature) of biological polymers differs from polymer nomenclature. The polymer chemist wishes to name the product of his own chemical ingenuity, and he is able to choose starting materials, condensing agents, and reaction conditions. The biochemist, at first and continuing, is faced with the need to specify the arrangement of units in substances existing in Nature, over the construction of which he has had no control. The approach to "naming" the latter has given rise to a system of nomenclature that he now applies to *bona fide* synthetic products where the control exercised by the polymer chemist does enter—but not to the extent, at least yet, of inducing the biochemist to change his habits!

The representation of biological polymers of all three kinds (and in the hybrids between them) has three separate parts: the representation of the mers or base units (the

* Presented before the Division of Polymer Chemistry, Symposium on Nomenclature, 156th Meeting, ACS, Atlantic City, N. J., September 1968.

** Senior Biochemist at the Oak Ridge National Laboratory; Director of the NAS-NRC Office of Biochemical Nomenclature (supported by a grant from the National Institutes of Health to NAS-NRC); and Secretary of the IUPAC-IUB Combined Commission on Biochemical Nomenclature.

*** Operated by Union Carbide for the U. S. Atomic Energy Commission.

biochemist has not yet learned to distinguish these; he calls them all monomers); the representation of the linkage between one and the next (ordered, random, branched, or unknown); and the noncovalent associations between polymers or macromolecules.

POLYPEPTIDES AND PROTEINS

A simple oligopeptide can be named semi-systematically, e.g., glycylalanylserine. The biochemist will always use the trivial names; the formal, systematic name for even this simple trimer he would find impossible to deal with. For long sequences, three conventions were simultaneously adopted from the earlier literature and now form part of the Tentative Rules of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN):[1] reduce the trivial names to their first three letters (with only rare exceptions); indicate the peptide bonds (-CO-NH-) by hyphens (known sequence) or commas and parentheses (unknown sequence); orient the chain so that the amino-terminal is at left and the carboxyl-terminal at right (Figure 1), or use directional arrows in place of hyphens, if reversed as in Figure 2. Thus, the tripeptide glycylalanylserine becomes Gly-Ala-Ser, or Ser ← Ala ← Gly. If the sequence were unknown, (Gly,Ala,Ser) or (Ala,Gly,Ser) or (Ser,Gly,Ala), etc., would be used— even

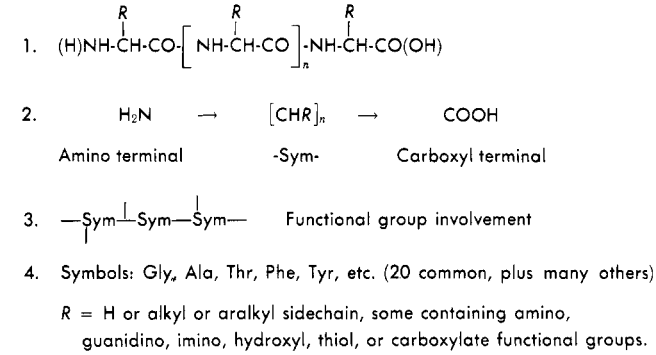


Figure 1. Derivation of symbolic representation of polypeptides

Line 1: the basic structure. Line 2: condensed basic structure. Line 3: symbolic representation (Sym = symbol, taken from line 4; horizontal lines = -CO-NH- groups; vertical lines = covalent links arising from functional groups on sidechains or from H of -CO-NH- group). Line 4: symbols usually derived from first three letters of the trivial names of the amino acids.

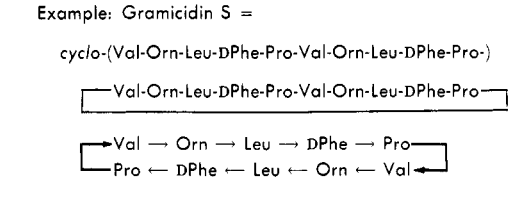


Figure 2. Example of symbolic representation of a cyclic peptide, showing three ways to express direction (polarity) of peptide links

Since all bonds are -CO-NH- and none are through functional groups, all bonds enter and leave each symbol horizontally (cf. Figure 3 for functional group bonds).

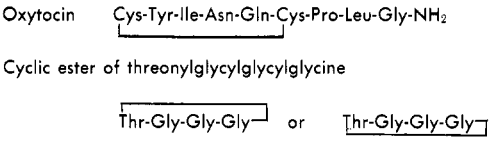
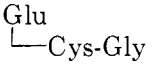


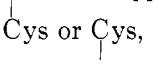
Figure 3. Symbolic representation of peptides containing functional group bonds (vertical lines)

In oxytocin, the -SH groups of two Cys residues are linked. In the tetrapeptide, the functional -OH of Thr is linked to the -COOH of a Gly residue.

in the middle of an otherwise known sequence. Branches in the over-all polypeptide structure resulting from substitution through functional groups (e.g., the ϵ-NH_2 of lysine, the γ-COOH of glutamate, the OH of serine or threonine, the SH of cysteine, or the remaining H of the —CO—NH— peptide bond itself) are indicated by vertical lines originating at the appropriate symbol (Figure 3). Thus, glutathione is depicted as



because it involves the glutamic acid γ-COOH groups and the cysteine α-NH_2 group (not its —SH). A link involving a cysteine —SH would appear as



not as Cys— or —Cys.

Configuration is assumed to the L unless otherwise noted, but the appropriate prefix may be added as required (Figure 2). Terminal functional groups, while implied, may also be spelled out (e.g., line 2 of Figure 1), and nonamino acid substituents, usually abbreviated to three letters,[2] may be bonded to the appropriate positions.

The usefulness of these conventions in aligning sequences for comparison is seen in Figure 4. For computer analyses, as in the search for homologies in long sequences, a one-letter code is desirable. Such has been adapted from the literature by CBN[3] and is demonstrated in Figure 5.

Adaptation of these conventions to the synthetic polypeptides (polymerized amino acids), which approach the polymer chemist's products, is shown in the examples taken from the Rules[4] in Figure 6. "Polymer of" is represented by a prefix "poly" or by a subscript "*n*." Thus, polyalanine (a source-based name, example 1) is either poly Ala or $(Ala)_n$. All the other conventions are as stated above, with the addition of superscripts to indicate the percentage (molar) composition determined by analysis.

Work on the designation of the conformations of polypeptides and proteins is in progress; for a preliminary report, see *J. Biol. Chem.* **241**, 1004 (1966).

POLYSACCHARIDES

A similar set of conventions[1] governs the representation of polysaccharides, two examples of which are shown in Figure 7. Here the left-right or arrow convention specifies the direction away from the hemiacetal carbon of the link (Figure 8). The monomers are again reduced to the first three letters (in most cases), and configurational (anomeric) letters and locant numbers are affixed in the appropriate places.

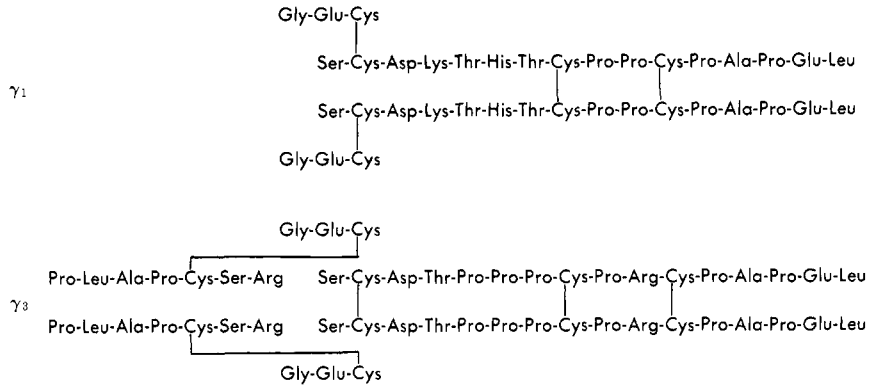
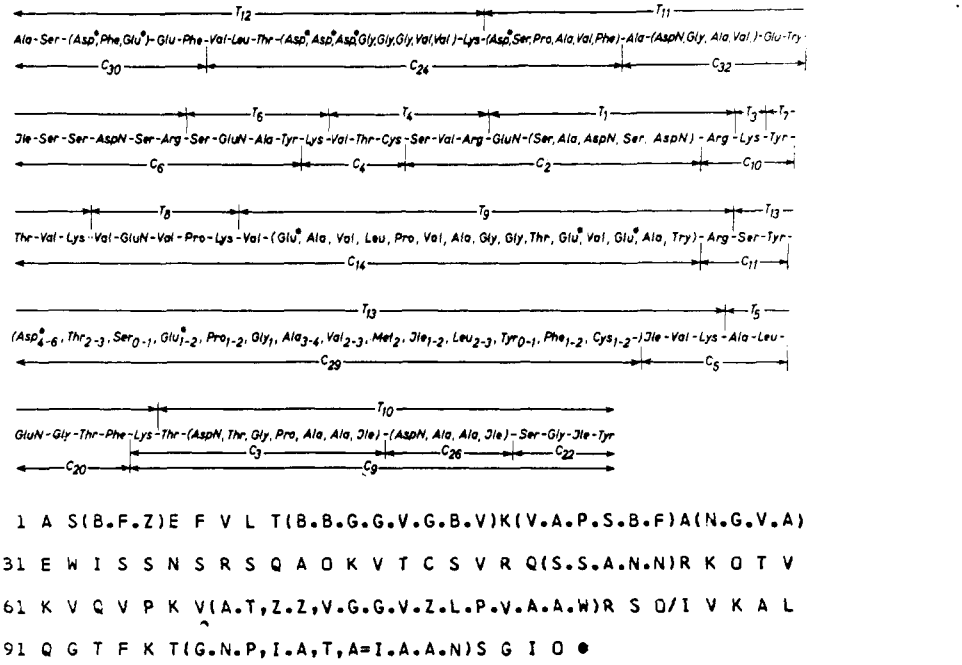


Figure 4. Symbolic representation of complex peptides, showing use in practice of peptide and bond symbols[1]

Homologous sequences around the interchain bridges in γG1, γG3, and γG4 myeloma proteins.




```
 1 A S(B.F.Z)E F V L T(B.B.G.G.V.G.B.V)K(V.A.P.S.B.F)A(N.G.V.A)
31 E W I S S N S R S Q A O K V T C S V R Q(S.S.A.N.N)R K O T V
61 K V Q V P K V(A.T,Z.Z,V.G.G.V.Z.L.P.V.A.A.W)R S O/I V K A L
91 Q G T F K T(G.N.P,I.A,T,A=I.A.A.N)S G I O ●
```

Figure 5. A protein, represented by 3-letter and 1-letter symbols (the segments delineated are those derived from partial hydrolyses), demonstrating the bond symbols used to designate known and unknown sequences (hyphen and comma, respectively, in the 3-letter system[1] and by space and period, respectively, in the 1-letter system.[3])

POLYNUCLEOTIDES AND NUCLEIC ACIDS

It is in the relatively new and rapidly-expanding field of nucleic acids that the inventiveness of the biochemical nomenclaturist is put to its most severe test. While one complication—branching—has so far not been encountered, others compound the problem, such as: ordered synthetic *vs.* irregular natural polymers; noncovalent interchain associations (hydrogen bonding); and natural and synthetic modifications of the (relatively few) fundamental monomers (purine and pyrimidine nucleotides).

A typical polynucleotide sequence is shown in Figure 9. In general, it is a simple poly(ribose phosphate) chain, the only complexity being the variation in the aglycones (purines and pyrimidines) attached to the C-1′ of the sugar moieties. The four most common bases are shown (in the deoxyribose compounds, uracil is replaced by 5-methyluracil), but both *N*- and *O*-substituted derivatives of these nucleotides are known to occur naturally—in addition to thio replacements and isomeric forms.

Although three-letter symbols, similar to those already discussed, have been constructed and used,[1] it is conventional in the field to use one-letter symbols.

Although the nucleotide—composed of a heterocyclic base, a pentose residue, and a phosphoric acid residue bound to each monosaccharide unit—may properly be

IUPAC-IUB Tentative Rules: Nomenclature of Synthetic Polypeptides

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 $(DLAla\text{-}Lys)_n$

4. Linear sequence of unknown order [Composition: 56% Glu, 38% Lys, and 6% Tyr (Σ = 100%)]:
poly $Glu^{56}Lys^{38}Tyr^{6}$ or $(Glu^{56}Lys^{38}Tyr^{6})_n$ (all L)

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 $(Glu^{56})_x$-$(Lys^{44})_y$

6. 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%)]:
$(poly\ (Glu\text{-}Lys)^{25}\text{-}(poly\ Ala\text{-}Tyr_2\text{-}Glu)^{12.5}$

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):
(poly Ala-poly Tyr) — (poly DLAla.Lys) or [$(Ala)_n$-$(Tyr)_n$] — $(DLAla.Lys)_n$

or (poly DLAla.Lys) — (poly Ala-poly Tyr)

Figure 6. Excerpt from IUPAC-IUB CBN Tentative Rules governing representation of synthetic polypeptides (chemically polymerized amino acids), indicating representation of block and graft polymers, branched chains, and composition by analysis[4]

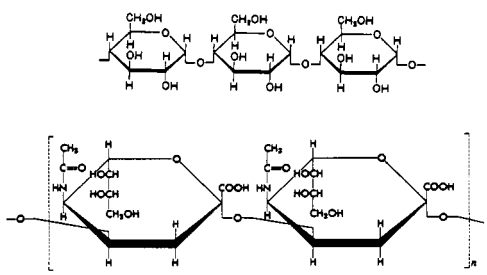


Figure 7. Segments of two natural homopolysaccharides, cellulose, a poly(α-D-glucopyranose), and colominic acid—a poly(*N*-acetylneuramic acid)

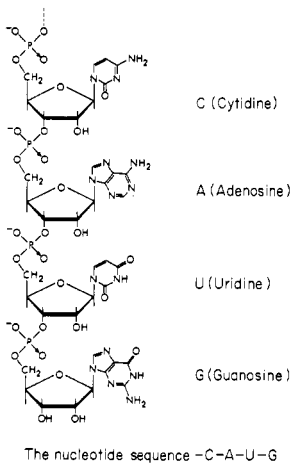


Figure 9. A tetraribonucleotide sequence, structurally and in one-letter symbolism[1]

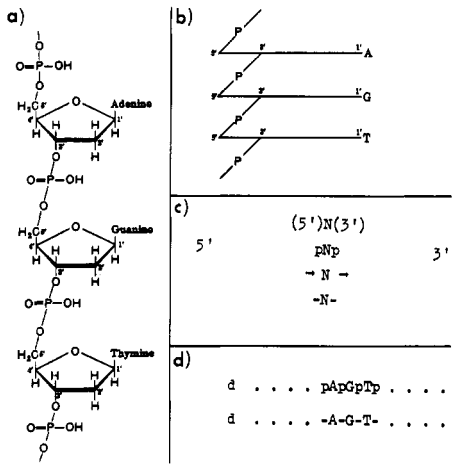


Figure 10. Derivation of symbolic representation of polynucleotides

A segment of a deoxyribonucleic acid (DNA) chain showing (a) structure of chain, (b) pictorial representation of the same segment, (c) symbolic representation of a homopolymer (N = nucleoside, p = phosphoric acid residue), (d) the sequence of (a) and (b) in one-letter symbols. The arrows in (c) are replaced by hyphens when the phosphodiester linkage is from 3′-OH at its left to 5′-OH at its right. The same convention holds in (d).

Maltose: Glcpα1 → 4Glc
Lactose: Galpβ1 → 4Glc or Galpβ1-4Glc
Stachyose: Galpα1-6Galpα1-6Galpα1-2βFruf
Branched chain: Glcβ1-3Galβ1-4Glc
2
↑
1αFuc
Inulin: DFruf2-(1DFruf)$_{35\text{-}2}$2-3DGlcp1-(1DFruf)$_x$2-1DGlcp
Xylan: βDXylp1-(4βDXylp1)$_x$-4βDXylp1-(4βDXylp1)$_y$-βDXylp
3
↑
1
βDXyl1-(4βDXylp)$_z$

Figure 8. Derivation of symbolic representation of polysaccharides[1]

Top line shows three-letter symbols separated by the glycoside link (arrow), pointing away from the hemiacetal carbon. The examples in the last four lines indicate how the left-to-right arrow may be replaced by a hyphen, how configurational prefixes may be inserted, and how branches are indicated.

regarded as the monomer, it is convenient to symbolize the nucleoside (base-sugar) in this regard and to represent the internal —PO_2H— or terminal —PO_3H_2 residues separately, as "p" or hyphen (bond). The small p is chosen to afford contrast with the (single) capital letters used for the nucleosides. Biochemists traditionally use P to represent the —PO_3H_2 radical in abbreviated names, as in ATP, glucose-P, etc. Thus pCpApUpG, or -C-A-U-G, represents the segment shown in Figure 9, which could be named -cytidylyladenylyluridylylguanosine. The directional convention is 3′ → 5′; without the arrow, it is so understood. Also understood are β base-sugar linkages,

Polymerized nucleotides

$$n\ (pppN) \xrightarrow{E} (pN)_n + n(pp)$$

$$n(ppN) \xrightarrow{E} (pN)_n + n(p)$$

	Name	Symbol
	polynucleotide	poly N or N_n
	polyadenylic acid; polyadenylate	poly A or A_n
[alternating]	poly(adenylate-cytidylate)	poly (A-C) or $(A\text{-}C)_n$
[random, 1:2]	poly(guanylate,2 uridylate)	poly (G,U_2) or $(G,U_2)_n$

Figure 11. Representation of synthetic polynucleotides[1]

As in Figure 10, N = nucleoside residue, p = phosphoric residue. As in Figure 5, the hyphen indicates a known sequence, the comma an unknown one. Without arrows, the direction is from 5′-terminal to 3′-terminal, left to right.

D configurations, and 1 or 9 locant numbers for the pyrimidines and purines, respectively, those being the preponderant order of affairs in nature. Without the prefix "d," the single capital letters for the nucleosides represent the ribosyl (as in Figure 9), not the 2-deoxyribosyl nucleosides, these comprising natural RNA and DNA, respectively. A segment of a DNA chain is shown in Figure 10.

Known and unknown sequences are described, as in the peptides, by hyphens (or arrows) and commas, respectively—the unknown sequences again being placed within parentheses. Thus G-A-U(C,C,U)Gp is a heptanucleotide containing a trinucleotide of unknown sequence between the first U and the terminal Gp residue.

The synthetic polymers, whether produced by chemical or enzymic means (Figure 11, top two lines), are further abbreviated by one of two means (Figure 11, bottom 4 lines). Two points are obvious: the internucleoside p is replaced by the hyphen (although an end p may be indicated); the subscript "*n*," chosen in place of the p recommended by the IUPAC Polymer Nomenclature Commission in 1952, was recommended[5] by the American Chemical Society Polymer Nomenclature Committee and, in 1967, viewed favorably by the IUPAC Macromolecular Nomenclature Commission.[6] Copolymers utilize the hyphen and comma between the symbols of the monomeric units to indicate repeating (alternating) or random sequence, respectively. Thus, poly (dA-dT) or $(dA\text{-}dT)_n$ indicates the repeating dA-dT-dA-dT----, and poly (dA,dT) or $(dA,dT)_n$ indicates the random sequence. The prefix d may also be placed before the parentheses, to give poly d(A-T) or $d(A\text{-}T)_n$, etc. This symbolism for the regular polymer differs by the presence of the hyphen from poly dAT, used in the paper[7] that initially described this regular copolymer, and fairly widely used since. Unfortunately, juxtaposed symbols have also been widely used for the irregular copolymers. The hyphen thus relieves potential ambiguity. Composition of the random copolymer, and size of either kind can be indicated by appropriate numbers or symbols.

Association between two or more independent—i.e., not covalently linked—polynucleotides is represented by the center dot (a hyphen would imply a covalent bond, and

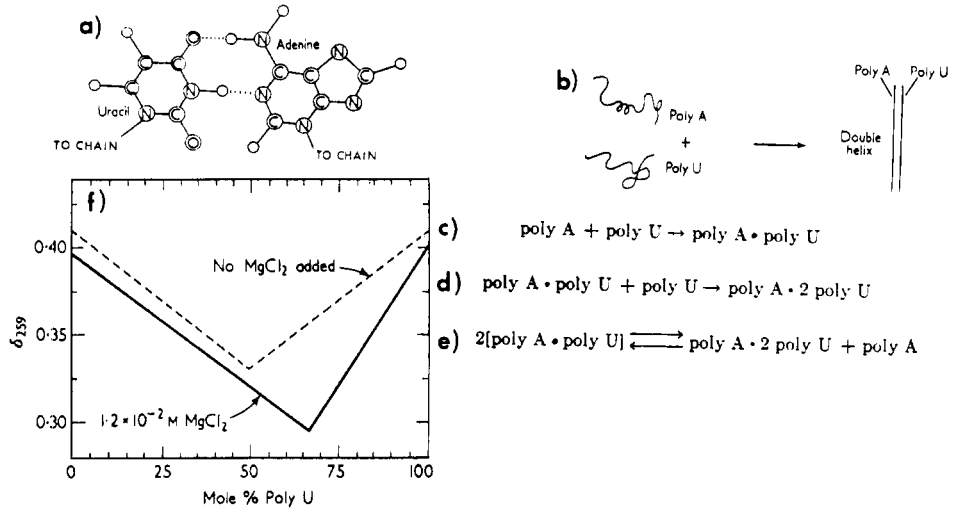


Figure 12. Symbolic representation of associated (hydrogen bonded) polynucleotide chains

(a) structure of the H-bonded monomeric units; (b) pictorial representation of the two chains (poly A and poly U) nonassociated and associated; (c) the reaction of (b) in symbolic terms; (d) reaction between the double-stranded product of (c) and another molecule of poly U; (e) a reaction of the product of (c) brought about by proper physical conditions, wherein two double-stranded complexes yield one triple-stranded complex—the product of (d)—and a single-stranded poly A; (f) the evidence for reactions (c) (broken line) and (d) (solid line); a plot of extinction change *vs.* mole % poly U added to poly A (From Michelson, A. M., J. Massoulié, and W. Guschlbauer, *Prog. Nucleic Acid Res. Mol. Biol.* 6, 81 (1967).

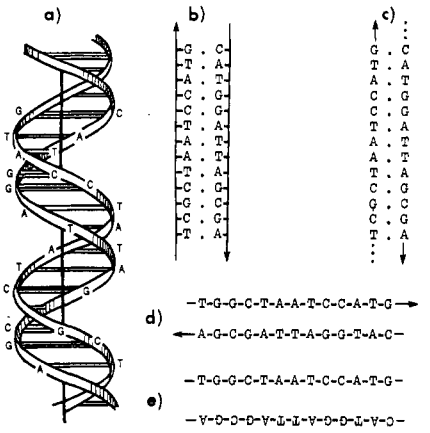


Figure 13. Segment of a DNA double helix, showing five ways of representing the structure with one-letter symbols for each nucleoside residue, hyphens for the phosphodiester link, and dots for the hydrogen bonds

The polarity convention (Figure 10) is assumed to hold when the letters are viewed right-side-up [as in (e)]. In (d), reverse polarity in the lower chain is shown by the arrow.

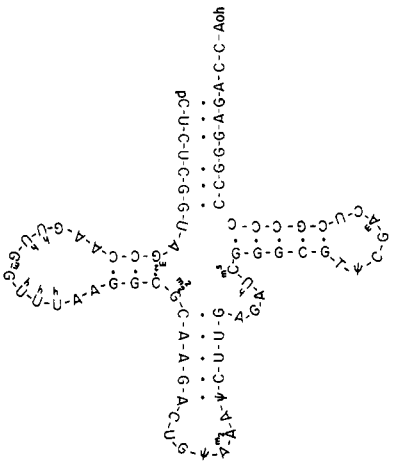


Figure 15. Symbolic representation of a ribonucleic acid

The same transfer RNA (Figure 14) folded to show hydrogen-bonding possibilities (dots between residues). Symbols and conventions as in Figures 14 *et ante*.

thus a block or graft polymer), e.g., poly d(A-C)·poly d(T-G), because of the hydrogen bonding known to occur between A and T and between G and C (Figure 12a). A colon was originally proposed[7] and used for this, but the use of the colon in other potentially conflicting contexts led to the choice of the center dot by CBN. The absence of association between two such chains is indicated by the plus sign, as in the reactions shown in Figure 12b, c, d, e; it indicates merely a single solution with two independent polymers. Indefiniteness as to association utilizes a comma between the symbols of the chains. In all cases, the optional $(\text{symbol})_n$ may replace the poly (symbol) form. Triple-stranded complexes are represented as, e.g., $(A)_n \cdot 2(U)_m$ or $(A)_n \cdot (U_m)_2$ or $(A)_n \cdot (U)_m(U)_m$, as in Figure 12d and e. Figure 13 illustrates how a segment of the H-bonded DNA chain may be symbolized.

The problem of representing adducts on individual residues assumes some importance in systems as abbreviated as these, especially when sequences are to be compared by aligning one below the other. Obviously, a 1-methyladenosine residue cannot be represented as 1-methylA, or even as 1-MeA. Not only are four additional spaces required, upsetting the alignment (as well as the "rhythm") of the monomeric units, but the principle of one capital letter per monomeric unit is violated. Hence the trend to reduce the more common adducts to single, lower case letters and to devise a new system for locant numbers, giving m^1A for 1-methyladenosine, m_2^2G for N^2-dimethylguanosine, Am for 2′-*O*-methyladenosine; the m follows the symbol when it is the ribosyl moiety rather than on the base. When spacing is important, these can become

$$A^{m^1}, \quad G^{m^2_2}, \quad \text{and} \quad A^{m}$$

respectively; see Figures 14 and 15 for linear and two-dimensional arrays, respectively. Chemical substituents generally utilize their usual symbols, but in lower case; e.g., an for anisoyl, bz for benzoyl, ac for acetyl—again to eliminate capital letters except for the nucleosides themselves.

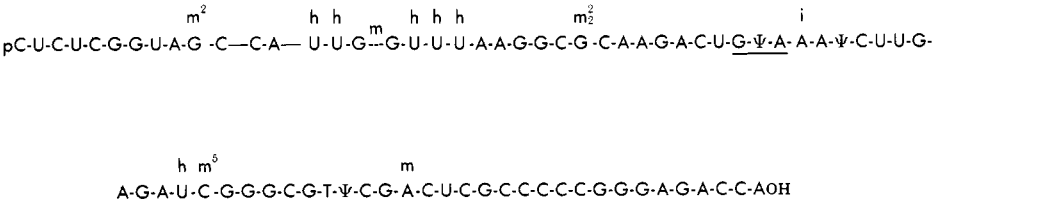


Figure 14. Symbolic representation of a ribonucleic acid

A transfer RNA in linear array, with end groups indicated (5′-phosphate at left, 3′-hydroxyl at right) and nucleoside substituents shown by lower-case symbols (m^2 = 2-methyl, m_2^2 = 2-dimethyl, h = 5,6-dihydro, right-hand superscript m = 2′-O-methyl, i = isopentenyl, m^5 = 5-methyl). Compare Figure 15. (From "Handbook of Biochemistry," H. A. Sober, Ed., pp. H62-63, Chemical Rubber Co., 1968.)

表 2. MSH の 構 造 と 活 性

種々の動物の MSH と関連ペプチド

α-MSH[9] (ブタ, ウシ, ウマ, サル)	Ac·Ser·Tyr·Ser·Met·Glu·His·Phe·Arg·Try·Gly·Lys·Pro·Val·NH_2
β-MSH (ブタ)[33]	H·Asp·Glu·Gly·Pro·Tyr·Lys·Met·Glu·His·Phe·Arg·Try·Gly·Ser·Pro·Pro·Lys·Asp·OH
(ウシ, ヒツジ)[34]	H·Asp·Ser·Gly·Pro·Tyr·Lys·Met·Glu·His·Phe·Arg·Try·Gly·Ser·Pro·Pro·Lys·Asp·OH
(ウマ)[35]	H·Asp·Glu·Gly·Pro·Tyr·Lys·Met·Glu·His·Phe·Arg·Try·Gly·Ser·Pro·Arg·Lys·Asp·OH
(サル)[36]	H·Asp·Glu·Gly·Pro·Tyr·Arg·Met·Glu·His·Phe·Arg·Try·Gly·Ser·Pro·Pro·Lys·Asp·OH
(ヒト)[37]	H·Ala·Glu·Lys·Lys·Asp·Glu·Gly·Pro·Tyr·Arg·Met·Glu·His·Phe·Arg·Try·Gly·Ser·Pro·Pro·Lys·Asp·OH

1. 活性との相関関係

脳下垂体中葉のメラニン細胞刺激ホルモン (MSH) には動物の種類によって変化しない α-MSH と，種属間に差の見いだされている β-MSH との存在が知られており，α-MSH は表1に示したようにN末端基がアセチル化されているが，β-MSH のN末アミノ基は遊離している。しかし β-MSH では α-MSH に比べてN末端側に3～7個のアミノ酸からなるペプチドが余分に結合している。これらの関係が表2に示されている。

表2中の α-MSH と β-MSH の構造を比較すると，N末端側では α-MSH の Ac·Ser·Tyr·Ser— が β-MSH の H·Asp·{Ser/Glu}·Gly·Pro·Tyr·{Ser/Lys}— あるいは H·Ala·Glu·Lys·Lys·Asp·Glu·Gly·Pro·Tyr·Arg— に相当していることがわかる。そうすると末端アセチル は Phe の anticodon とはなんの関係もないものかもしれない。Phe の RNA を濃縮してこのような配列の有無をみればよいが，この仕事はわれわれの手で進行中である。

マップで検出できる範囲では $A_pA_pA_pX_p$ の X_p は U_p の系統のものに限られるが，多量の t-RNA またはある程度分画した t-RNA の RNase 消化物をカラムクロマトグラフすると $A_pA_pA_pG_p$ が微量であるが見いだされる[13]。$A_pA_pA_pG_p$ の微量の存在は他の報告にもある[8,13,14]。

Figure 16. Indication of the international use of the IUPAC-IUB CBN symbols in both polypeptide and polynucleotide representations

Montage of two papers appearing in "Proteins, Nucleic Acids, and Enzymes"

The extent to which the symbolism described in this paper has become an international language is indicated by the segments of papers reproduced in Figure 16.

The rapid advances in biochemical aspects of polymer and macromolecule chemistry and the difficulties in assigning proper biradical names to their base units make it appear that biochemical polymers will continue to be named on a source basis and will utilize symbols for the names of the sources.

LITERATURE CITED

(1) "Abbreviations and Symbols for Chemical Substances of Special Interest in Biological Chemistry" (1965 Revision), *J. Biol Chem.* **241**, 527 (1966); *Biochemistry* **5**, 1445 (1966); and in 8 other journals in 4 languages.

(2) "Abbreviated Designation of Amino Acid Derivatives and Peptides," *J. Biol. Chem.* **241**, 2491 (1966); *Biochemistry* **5**, 2485 (1966); and 7 other journals in 4 languages.

(3) "One-Letter Notation for Amino Acid Sequences," *J. Biol. Chem.* **243**, 3557 (1968); *Biochemistry* **7**, 2703 (1968); and other journals in 7 languages.

(4) "Abbreviated Nomenclature of Synthetic Polypeptides (Polymerized Amino Acids)," *J. Biol Chem.* **243**, 2451 (1968); *Biochemistry* **7**, 483 (1968); and 7 other journals in 4 languages.

(5) *Macromolecules* **1**, 193 (1968).

(6) IUPAC Macromolecular Nomenclature Commission, Unpublished summary of 1967 meeting.

(7) Inman, R. B., and R. L. Baldwin, *J. Mol. Biol.* **5**, 172 (1962).