

1 Nomenclature, definitions and geometry of basic structure elements

In the past, several independent definitions were used to describe the structural parameters of the nucleic acids and of their constituents. There have been two major proposals made by the "IUPAC-IUB Commission on Biochemical Nomenclature"¹⁾ which will be followed throughout this volume. One, concerned with the chemical nomenclature dates back to 1970 [70I1], the other is much more recent, 1983, and used to describe the three-dimensional structure of nucleosides, nucleotides and nucleic acids [83I1].

1.1 Abbreviations and symbols

There are two kinds of nucleic acids,

ribonucleic acid (RNA) and
deoxyribonucleic acid (DNA).

The nucleic acids are linear polymers composed of four different building blocks, the nucleotides, which are linked by phosphodiester bonds. The individual nucleotide consists of a furanoside-type ribose (in RNA) or 2'-deoxyribose (in DNA) connected by a C1'-N glycosyl bond with one of four different bases, and by ester bond(s) to a phosphate group (Tables 1 and 2, Fig. 1).

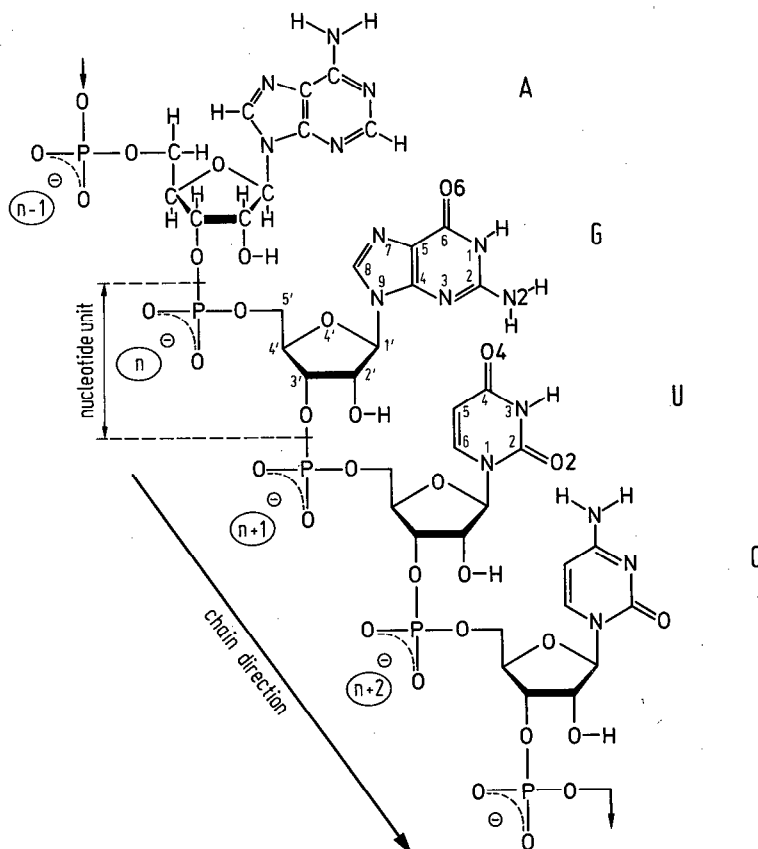


Fig. 1. Fragment of ribonucleic acid (RNA) with sequence adenosine (A), guanosine (G), uridine (U), cytidine (C) linked by 3',5'-phosphodiester bonds. Chain direction is from 5'- to 3'-end as shown by arrow. Atom numbering scheme is indicated in one framed nucleotide unit, 5'-GMP. All hydrogen atoms are drawn in A and only functional hydrogens in other nucleotides. In short notation, this fragment would be pApGpUpCp or pAGUCp. In deoxyribonucleic acid (DNA), the hydroxyl attached to C2' is replaced by hydrogen and uracil, by thymine [84S1].

¹⁾ IUPAC=International Union of Pure and Applied Chemistry. IUB=International Union of Biochemistry.

Table 1. The constituents of nucleosides and nucleotides.

| | | |
|--|------------|--|
| Base | Purine | { adenine guanine |
| | Pyrimidine | { cytosine uracil (in RNA), thymine (in DNA) |
| Nucleoside = Base + Sugar | | { D-ribose (in RNA) 2'-deoxy-D-ribose (in DNA) |
| Nucleotide = Base + Sugar + Phosphate | | { linked by phosphodiester bond at 2', 3' or 5' hydroxyl in RNA 3' or 5' hydroxyl in DNA |

Table 2. Abbreviations and symbols for bases, nucleosides, and nucleotides¹⁾.

| Base | | Nucleoside | | Nucleotide | |
|--|--------|------------------------------|------------|----------------------|----------------|
| Name | Symbol | Name | Symbol | Name | Symbol |
| Ribonucleosides and -nucleotides | | | | | |
| Uracil | Ura | uridine | Urd or U | uridylic acid | 5'-UMP or pU |
| Cytosine | Cyt | cytidine | Cyd or C | cytidylic acid | 5'-CMP or pC |
| Adenine | Ade | adenosine | Ado or A | adenylic acid | 5'-AMP or pA |
| Guanine | Gua | guanosine | Guo or G | guanylic acid | 5'-GMP or pG |
| 2'-Deoxyribonucleosides and -nucleotides²⁾ | | | | | |
| Thymine | Thy | deoxythymidine ³⁾ | dThd or dT | deoxythymidylic acid | 5'-dTMP or pdT |
| Cytosine | Cyt | deoxycytidine | dCyd or dC | deoxycytidylic acid | 5'-dCMP or pdC |
| Adenine | Ade | deoxyadenosine | dAdo or dA | deoxyadenylic acid | 5'-dAMP or pdA |
| Guanine | Gua | deoxyguanosine | dGuo or dG | deoxyguanylic acid | 5'-dGMP or pdG |

Other examples**Nucleotides²⁾**

uridine 2'-monophosphate (2'-UMP)
 uridine 3'-monophosphate (3'-UMP, Up)
 cytidine diphosphate (CDP, ppC)
 cytidyl-(3',5')-uridine (CpU)
 adenosine triphosphate (ATP, pppA)
 guanosine 2',3'-cyclic phosphate
 (2',3'-GMP, G > p, cGMP)

Polynucleotides

polyadenylic acid (poly A); alternate copolymer of dA and dT, poly(deoxyadenylate-deoxythymidylate), poly [d(A-T)], or poly(dA-dT) or (dA-dT)_n or d(A-T)_n; the same but randomly distributed dA, dT: replace hyphen by comma, poly d(A, T) etc. A complex between poly(A) and poly(U) is designated poly(A)·poly(U) alanine-specific transfer RNA from *E. coli* (tRNA^{Ala} (*E. coli*))

¹⁾ Adapted from [70A1], taken from [84S1].

²⁾ The symbols for 2'-deoxyribonucleosides and -tides are as for ribonucleosides and -tides with the prefix d.

³⁾ Since thymidine occurs as a ribonucleoside in tRNA, use of the prefixes d for deoxyribose and r for ribose is recommended [70I1].

a) Bases and nucleosides

In the atom designation, base atoms are described by letter and numerals with or without parentheses, e.g. N(9), N9 or N₉, and sugar atoms are distinguished by primed numerals, e.g. C(2'), C2' or C_{2'}.

The free purine bases adenine, guanine bear a hydrogen atom at position 9, which in the nucleosides is substituted by C1' of ribose or deoxyribose in a β -type glycosyl link (Fig. 2). The same holds for the free hydrogen atom in position 1 of the pyrimidine bases cytosine, uracil (in RNA) or the equivalent thymine (in DNA).

b) Nucleotides

Nucleosides can be phosphorylated in three (ribose) or two (deoxyribose) sugar hydroxyl positions to form a number of different nucleotides:

| | | |
|----------------------|---|----------------------------------|
| ribo-nucleoside | { | nucleoside-5'-phosphate |
| | | nucleoside-3'-phosphate |
| | | nucleoside-2'-phosphate |
| | | etc. |
| deoxyribo-nucleoside | { | deoxyribonucleoside-5'-phosphate |
| | | deoxyribonucleoside-3'-phosphate |
| | | etc. |

Nucleosides can also be di- or triphosphorylated (see Fig. 2):

diphosphorylated at two positions, e.g.:

adenosine-3',5'-diphosphate, A-3',5'-P₂, or A-3':5'-P₂, or 3',5'ADP, or pAP

di- or triphosphorylated at one position, e.g.:

adenosine-5'-diphosphate, ADP

adenosine-5'-triphosphate, ATP.

The phosphate can be attached to two hydroxyls of the same nucleoside to form a cyclic phosphate (Fig. 2), e.g.:

adenosine-3',5'-cyclic phosphate, Ado-3',5'-P, Ado-3':5'-P, 3',5'-AMP, A > p, cAMP

adenosine-2',3'-cyclic phosphate, Ado-2',3'-P, Ado-2':3'-P, 2',3'-AMP, A > p, cAMP.

A special case is the coenzyme nicotinamide-adenine-dinucleotide, NAD⁺, which contains nicotinamide riboside (Nir) and Ado separated by a pyrophosphate group: Ado-5'PP5'-Nir.

c) Oligo- and polynucleotides

These are designated with the one-letter code (Table 2), with prefix d if in the DNA series. The "polarity" is in the direction 5' → 3' (Fig. 1), if not otherwise indicated. The phosphodiester linkage is the common 3'p5', and is specified if different.

The oligonucleotide guanylyl-3',5'-cytidyl-3',5'-uridine can be abbreviated GpCpU or, shorter G-C-U or GCU, with G the 5'-end and U the 3'-end of the chain.

If the oligomer contains terminal phosphate groups, these are specified:

ApGpUp (or A-G-Up or AGUp) has a 3'-terminal phosphate

ApGpU > p (or A-G-U > p or AGU > p) has a terminal 2'3' (or 2':3') cyclic phosphate

pApGpU (or pA-G-U or pAGU) has a 5'-terminal phosphate

dApdGpCpU (or dApdGpCpU or dAdGCU or d(AG)r(CU))

contains deoxyribo- and ribonucleotides in the same oligomer.

In the base-paired complementary oligonucleotides, the nomenclature is:

A C U A G C
U G A U C G

or

A → C → U → A → G → C
U ← G ← A ← U ← C ← G

or

Ap Cp Up Ap Gp C
Up Gp Ap Up Cp G

In the deoxyribo series, the prefix d is used but it can be omitted if it is not necessary:

d(pGATCGAT) or pGATCGAT.

In polymer nucleic acids which are mostly obtained synthetically, the prefix poly is used, meaning "polymer of":

polyadenylic acid or polyadenylate or poly(A)

alternating copolymer poly(adenylate-cytidylate): poly(A-C)

random copolymer of the same type poly(A,C).

In the deoxy series, the prefix d is used:

poly(dA); and for the alternating copolymer: poly[d(A-T)] or poly(dA-dT).

Complementary duplex formation is indicated by a dot symbol poly(A)·poly(U) and a triple helix is: poly(A)·2poly(U). The alternating copolymer duplex in the deoxyribonucleoside series is described as:

poly[d(G-C)]·poly[d(G-C)], or poly(dG-dC)·poly(dG-dC).

d) Modified bases and sugars

Bases and sugars can be modified by different substituents in different positions.

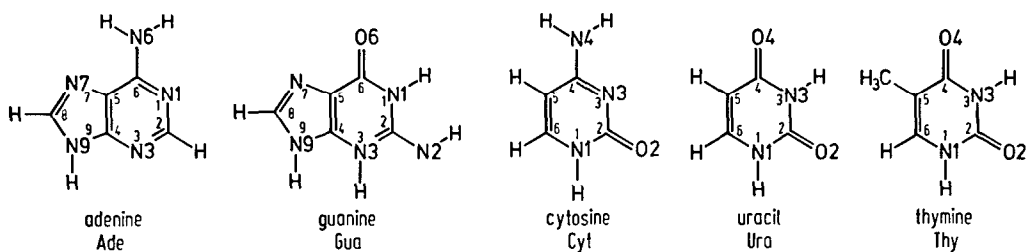
The substituents are abbreviated as:

| | |
|----------------|--|
| m, e, ac | methyl, ethyl, acetyl |
| z, c | aza (N replaces C), deaza (C replaces N) |
| h | dihydro (hU = dihydrouridine) |
| hm, ho (or oh) | hydroxymethyl, hydroxy |
| aa | aminoacyl |
| f | formyl |
| i | isopentenyl |
| s | thio or mercapto (sU = thiouridine) |
| fl, cl, br, io | fluoro, chloro, bromo, iodo. |

The positions are indicated by superscripts, multipliers by subscripts. Some examples:

| | |
|-------------------------------|--|
| m ₂ A | dimethyladenosine or N ⁶ -dimethyladenosine |
| m ₂ ⁵ A | ribosyl-6-(dimethylamino)purine |
| ac ⁴ C | N ⁴ -acetylcytidine |
| s ² U | 2-thiouridine. |

Frequently, riboses are methylated at the O2' position. The prefix 2'-O-Me is used, or replaced by the suffix m, e.g.: 2'-O-MeC is written as Cm.



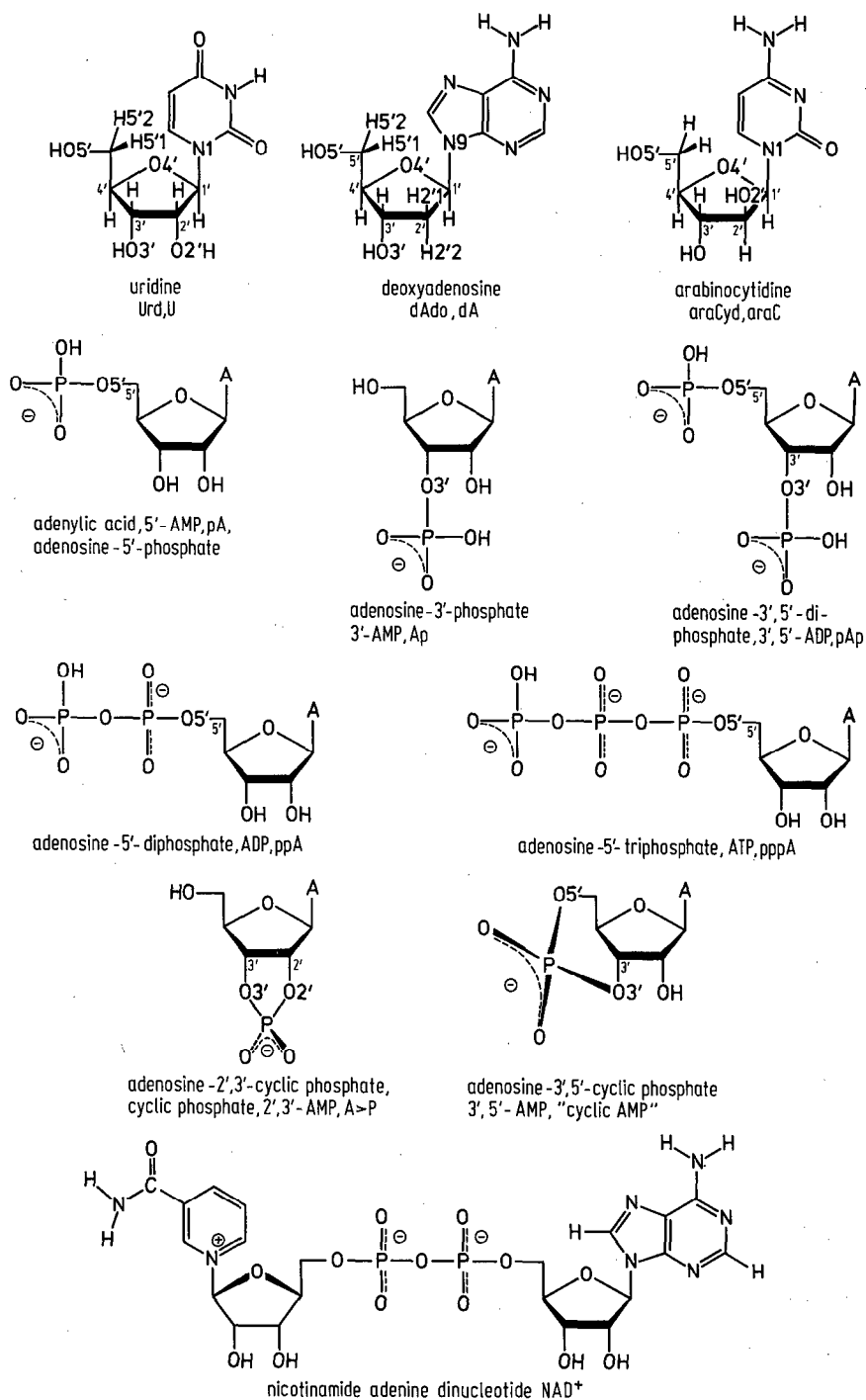


Fig. 2. Chemical structure of some bases, nucleosides, nucleotides and the coenzyme NAD^+ [84S1]. The word arabino describes the sugar moiety which is derived from the arabinose. In this context, all the nucleosides could be described more fully as e.g. ribouridine or deoxyribouridine. Because the common nucleosides are of the ribo form, the word ribo is usually omitted in the nomenclature.