

Nucleoside and Nucleotide Nomenclature

As with any area of chemistry, the utilization of proper nomenclature when referring to nucleosides, nucleotides, and nucleic acids provides a clarity that facilitates understanding by other scientists. This chemical direction began in the nineteenth century with the isolation and identification of various constituents of nucleic acids, and continues unabated today with, for example, the chemical synthesis of oligonucleotide analogs of various lengths and with all manner of modifications. The term “nucleotide” was first used by P.A. Levene (Levene and Mandel, 1908); “nucleoside” was introduced by the same author the following year (Levene and Jacobs, 1909). The specific components of nucleic acids, as they were isolated from natural sources, were given names related to those sources that have remained with them over the years. For example, guanine was given its name because it was first isolated from bird excrement, i.e., guano (Unger, 1846a,b). Adenine was named by Kossel (1885, 1886) after isolation from bovine pancreas, the root being derived from the Greek word for gland. Interestingly, the first nucleotide that was actually isolated was inosinic acid, by Liebig (1847), and the first nucleoside isolated was guanosine, which was initially given the name vernine (Schulze and Bosshard, 1885, 1886).

The main purpose of this appendix is to provide pertinent references that will direct the reader to the relevant guidelines or evolving nomenclature as described in the literature. When additional suggestions or guidance are appropriate, those comments will be included as well. One of the beauties and challenges of science as it evolves is the process of keeping pace with accurate, concise, and flexible systems of nomenclature. Chemists have always been able to discover or synthesize compounds that render existing systems inadequate.

Current nomenclature in the area of nucleosides, nucleotides, and nucleic acids comprises a mixture of (1) common names that have gained official recognition, (2) guidelines that have been derived and officially recommended by the International Union of Pure and Applied Chemistry (IUPAC)/International Union of Biochemistry and Molecular Biology (IUBMB), and (3) evolving usage that is derived by individual scientists and laboratories and subjected to peer review through publication. A working group has been commissioned

(1998) by IUBMB to review guidelines for nucleotide (including oligonucleotide) nomenclature. As those guidelines are developed and made available, they will be referenced in future updates of this appendix.

As noted earlier, the first nomenclature arose out of natural product isolation and identification, and thus common names were utilized for adenosine, guanosine, cytidine, uridine, and thymidine. The purine numbering system arose from that era, and it has been kept intact (see below) with the usual site of carbohydrate attachment at N-9. The pyrimidine numbering system was changed, at least as applied to the standard nucleosides, and in the current system the site of attachment of the carbohydrate in normal nucleosides is N-1 (earlier literature has the numbering such that attachment is at N-3). The accepted system of numbering for both purines and pyrimidines is illustrated in Figure A.1D.1. Nomenclature systems for the carbohydrate and nitrogen base moieties in nucleosides and nucleotides remain unchanged, except that primed numbers are used to refer to the carbohydrate atoms when the entire molecule is being considered.

The fact that DNA and RNA differ mainly in the presence or absence of a 2'-hydroxyl in the carbohydrate pentofuranosyl moiety has simplified nomenclature. The only complicating factor is that uridine, which occurs in RNA, is replaced by thymidine in DNA, and unfortunately, thymidine, isolated first from thymus DNA, was not given a 2'-deoxy designation when named by P.A. Levene (Levene and London, 1929; Levene and Tipson, 1935). Nomenclature guidelines have maintained the original usage, with thymidine referring to a 2'-deoxynucleoside, while in all other cases the 2'-deoxy is included within the name for the DNA component. Abbreviations that are used, however, can take into account the 2'-deoxy nature of thymidine quite readily. For example, the abbreviations for 2'-deoxyadenosine, 2'-de-

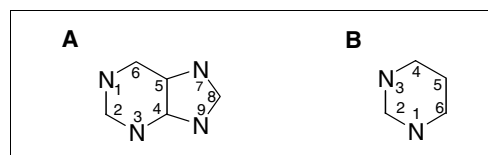


Figure A.1D.1 Numbering system for (A) purines and (B) pyrimidines.

Standard
Nomenclature,
Data, and
Abbreviations

A.1D.1

Contributed by John A. Secrist III

Current Protocols in Nucleic Acid Chemistry (2000) A.1D.1.1-A.1D.3

Copyright © 2000 by John Wiley & Sons, Inc.

Table A.1D.1 Available Nomenclature Recommendations on Nucleic Acids

Title	Available in
Nomenclature of Carbohydrates	McNaught (1996, 1997a,b,c)
Abbreviations and Symbols for Nucleic Acids, Polynucleotides, and their Constituents (published 1970)	IUPAC-IUB (1970a,b, 1971a,b,c,d, 1972, 1974)
Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains (published 1982)	IUPAC-IUB (1983a,b)
Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences (published 1984)	NC-IUB (1985a,b, 1986a,b,c)
Nomenclature of Junctions and Branchpoints in Nucleic Acids (published 1994)	NC-IUBMB (1995a,b, 1996)

oxyguanosine, 2'-deoxycytidine, and thymidine are dA, dG, dC, and dT.

Nomenclature Documents

Two excellent Web sites are available for learning about nomenclature documents coming from IUPAC, IUBMB, and the Joint Commission on Biochemical Nomenclature. The documents mentioned below are also included on these websites: <http://www.chem.qmw.ac.uk/iubmb> and <http://www.chem.qmw.ac.uk/iupac/>.

Chemists wishing to learn about nucleoside nomenclature have several sources to consult. To obtain information on carbohydrate nomenclature, a recent revision/updating of the IUPAC recommendations "Nomenclature of Carbohydrates," is available. For these recommendations as well as for others listed below, the literature citations will include all the journals that have reproduced them, so that readers can consult the one most readily available to them. For example, the carbohydrate nomenclature document is available in McNaught (1996, 1997a,b,c).

The basic set of recommendations for nucleosides, nucleotides, and nucleic acids, entitled "Abbreviations and Symbols for Nucleic Acids, Polynucleotides, and Their Constituents," was published in 1970 and is available in IUPAC-IUB (1970a,b, 1971a,b,c,d, 1972, 1974). This broadly-based document covers normal nucleosides and their abbreviations, as well as nucleotides, oligonucleotides, and polynucleotides. Information on names and abbreviations for modified nucleosides that occur in various nucleic acids, including tRNA, is also included.

Available nomenclature recommendations on nucleic acids are listed in Table A.1D.1.

A summary of the modified nucleosides in RNA is available, along with symbols and notes and references relevant to naming these nucleosides (Limbach et al., 1994). Updates of this reference are available through the following Web site: <http://medlib.med.utah.edu/RNAmods/>.

LITERATURE CITED

- IUPAC-IUB. 1970a. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Eur. J. Biochem.* 15:203-208.
- IUPAC-IUB. 1970b. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *J. Biol. Chem.* 245:5171-5176.
- IUPAC-IUB. 1971a. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Arch. Biochem. Biophys.* 145:425-436.
- IUPAC-IUB. 1971b. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Biochem. J.* 120:449-454.
- IUPAC-IUB. 1971c. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Biochemistry* 9:4022-4027.
- IUPAC-IUB. 1971d. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Biochim. Biophys. Acta* 247:1-12.
- IUPAC-IUB. 1972. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Eur. J. Biochem.* 25:1.
- IUPAC-IUB. 1974. Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents. *Pure Appl. Chem.* 40:277-290.
- IUPAC-IUB. 1983a. Abbreviations and symbols for the description of conformations of polynucleotide chains. *Eur. J. Biochem.* 131:9-15.
- IUPAC-IUB. 1983b. Abbreviations and symbols for the description of conformations of polynucleotide chains. *Pure Appl. Chem.* 55:1273-1280.
- Kossel, A. 1885. Über eine neue Base aus dem Thierkörper. *Ber.* 18:79-81.

- Kossel, A. 1886. Weitere Beiträge zur Chemie des Zellkerns. *Z. Physiol. Chem.* 10:248-264.
- Levene, P.A. and Mandel, J.A. 1908. Über die Konstitution der Thymo-nucleinsäure. *Chem. Ber.* 41:1905-1909.
- Levene, P.A. and Jacobs, W.A. 1909. Über die Hefe-Nucleinsäure. *Chem. Ber.* 42:2474-2478.
- Levene, P.A. and London, E.S. 1929. The structure of thymonucleic acid. *J. Biol. Chem.* 83:793.
- Levene, P.A. and Tipson, R.S. 1935. The ring structure of thymidine. *J. Biol. Chem.* 109:623.
- Liebig, J. 1847. Über die Bestandtheile der Hüsigkeiten des Heisches. *Ann.* 62:257-369.
- Limbach, P.A., Crain, P.F., and McCloskey, J.A. 1994. Summary: The modified nucleosides of RNA. *Nucl. Acids Res.* 22:2183-2196.
- McNaught, A.D. 1996. Nomenclature of carbohydrates. *Pure Appl. Chem.* 68:1919-2008.
- McNaught, A.D. 1997a. Nomenclature of carbohydrates (recommendations 1996). *Adv. Carbohydr. Chem. Biochem.* 52:43-177.
- McNaught, A.D. 1997b. Nomenclature of carbohydrates. *J. Carbohydr. Chem.* 16:1191-1280.
- McNaught, A.D. 1997c. Nomenclature of carbohydrates. *Carbohydr. Res.* 297:1-92.
- NC-IUB. 1985a. Nomenclature for incompletely specified bases in nucleic acid sequences. *Biochem. J.* 229:281-286.
- NC-IUB. 1985b. Nomenclature for incompletely specified bases in nucleic acid sequences. *Eur. J. Biochem.* 150:1-5.
- NC-IUB. 1985c. Nomenclature for incompletely specified bases in nucleic acid sequences. *Nucl. Acids Res.* 13:3021-3030.
- NC-IUB. 1986a. Nomenclature for incompletely specified bases in nucleic acid sequences. *J. Biol. Chem.* 261:13-17.
- NC-IUB. 1986b. Nomenclature for incompletely specified bases in nucleic acid sequences. *Mol. Biol. Evol.* 3:99-108.
- NC-IUB-1986c. Nomenclature for incompletely specified bases in nucleic acid sequences. *Proc. Nat. Acad. Sci. U.S.A.* 83:4-8.
- NC-IUBMB. 1995a. Nomenclature of junctions and branchpoints in nucleic acids. *Eur. J. Biochem.* 230:1-2.
- NC-IUBMB. 1995b. Nomenclature of junctions and branchpoints in nucleic acids. *Nucleic Acid Res.* 23:3363-3364.
- NC-IUBMB. 1996. Nomenclature of junctions and branchpoints in nucleic acids. *J. Mol. Biol.* 554-555.
- Schulze, E. and Bosshard, E. 1885. Zur Kenntniss des Vorkommens von Allantoin, Asparagin, Hypoxanthin und Guanin in den Pflanzen. *Z. Physiol. Chem.* 9:420-444.
- Schulze, E. and Bosshard, E. 1886. Über einen neuen stickstoffhaltigen Pflanzenbestandtheil. *Z. Physiol. Chem.* 10:80-89.
- Unger, B. 1846a. Bemerkungen zu obiger Notiz. *Ann.* 58:18-20.
- Unger, B. 1846b. Das Guanin und seine Verbindungen. *Ann.* 59:58-68.

Contributed by John A. Secrist III
Southern Research Institute
Birmingham, Alabama