Pure & Appl. Chem., Vol.54, No.8, pp.1517-1522, 1982.
Printed in Great Britain.

0033-4545/82/081517-06$03.00/0
Pergamon Press Ltd.

PROVISIONAL

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

and

INTERNATIONAL UNION OF BIOCHEMISTRY

JOINT COMMISSION ON BIOCHEMICAL NOMENCLATURE*

ABBREVIATED TERMINOLOGY OF OLIGOSACCHARIDE CHAINS

Comments on these recommendations are welcome and should be sent within 8 months from August 1982 to the Secretary of the Commission

Dr. H. B. F. DIXON
Department of Biochemistry
University of Cambridge
Tennis Court Road
Cambridge CB2 1QW
UK

Comments from the viewpoint of languages other than English are especially encouraged. These may have special significance regarding the publication in various countries of translations of the nomenclature eventually approved by IUPAC.

*Membership of the Commission for 1979-81 was as follows:

Chairman: P. KARLSON (FRG); *Secretary:* H. B. F. DIXON (UK); *Members:* B. L. HORECKER (USA); Y. JEANNIN (France); C. LIÉBECQ (Belgium — as Chairman of IUB Committee of Editors of Biochemical Journals); B. LINDBERG (Sweden); K. L. LOENING (USA); G. P. MOSS (UK); J. REEDIJK (Netherlands); S. F. VELICK (USA); J. F. G. VLIEGENTHART (Netherlands).

Additional contributors to the formulation of these recommendations:

Nomenclature Committee of IUB (those additional to JCBN): H. BIELKA (GDR); W. B. JAKOBY (USA); B. KEIL (France); E. C. WEBB (Australia).

Expert Subcommittee: B. LINDBERG (Convener, Sweden); D. HORTON (USA); late W. KLYNE (UK); K. L. LOENING (USA); D. J. MANNERS (UK); W. G. OVEREND (UK); H. PAULSEN (FRG); D. A. REES (UK); R. S. TIPSON (USA).

PROVISIONAL

Abbreviated Terminology of Oligosaccharide Chains

Recommendations 1980*

IUB-IUPAC Joint Commission on Biochemical Nomenclature (JCBN)

CONTENTS

Systematic names and structural formulas of oligosaccharides and of oligosaccharide chains in polysaccharides and in glycoconjugates become unwieldy with increasing molecular size, and for large molecules there is a need for an abbreviated terminology. The recommendations below, based upon a report by a subcommittee,* follow in the main what has become established practice in the carbohydrate literature.

In the recommendations given herein, which are, as far as possible, formulated within the framework of the IUPAC-IUB Tentative Rules for Carbohydrate Nomenclature (1), all available structural information is given explicitly. When a more condensed set of abbreviations is desired, the recommendations given in Section Lip-3 of the Nomenclature of Lipids (2) should be used. These recommendations, with some examples of their application, are given in the "Appendix" to the present document.

DEFINITIONS

An oligosaccharide is a molecule containing a small number (2 to about 10) of monosaccharide residues, connected by glycosidic linkages. A carbohydrate containing two such residues is a disaccharide, a carbohydrate containing three such residues is a trisaccharide, and so on.

* Document of the IUB-IUPAC Joint Commission on Biochemical Nomenclature (JCBN) whose members are P. Karlson (Chairman), H. B. F. Dixon, Y. Jeannin, C. Liébecq (as Chairman of the IUB Committee of Editors of Biochemical Journals), B. Lindberg, K. L. Loening, G. P. Moss, and S. F. Velick in consultation with J. F. G. Vliegenthart and the Nomenclature Committee of IUB, whose additional members are H. Bielka, W. B. Jakoby, B. Keil, and E. C. Webb. Comments or suggestions for modifications may be sent to the secretary of JCBN, H. B. F. Dixon, Department of Biochemistry, Tennis Court Road, Cambridge, United Kingdom CB2 1QW, or to any member. JCBN thanks the expert subcommittee of B. Lindberg (convener), D. Horton, the late W. Klyne, K. L. Loening, D. J. Manners, W. G. Overend, H. Paulsen, D. A. Rees, and R. S. Tipson for drafting these proposals.

RECOMMENDATIONS

1. Trivial Names

(*a*) Certain trivial names firmly established in the literature are specific for particular structures, and their continued use is allowed in instances where the full name may be unwieldy.

Examples: *Disaccharides:* Cellobiose, chitobiose, gentiobiose, kojibiose, lactose, melibiose, sophorose, sucrose, α,α-trehalose, turanose.
Tri- and Oligosaccharides: Melezitose, panose, raffinose, stachyose.

Further examples are given in Section Lip-3 of the Nomenclature of Lipids (2).

(*b*) Such disaccharide names as xylobiose and mannobiose, which are ambiguous, should only be used when there is no risk of confusion. The systematic name should be given together with the trivial name, at the first mention.

(*c*) The accepted trivial names for disaccharides should only be extended to higher oligosaccharides when the latter contain a single type of sugar residue and linkage.

Examples: Cellotetraose, maltotriose.

When, however, the trivial name is derived from the name of a carbohydrate that contains two or more different sugars, or types of linkage, or both, extension of the names to higher oligosaccharides is *not* recommended.

Examples: *Not* recommended are such names as agarotetraose and nigerotriose.

2. Systematic Names

These should be assigned as indicated in Carb-39 and Carb-40 (1).

3. Abbreviated Names for Use for Oligosaccharides

(*a*) The symbols chosen are derived from the trivial names of the constituent sugars. For the sake of clarity, brevity, and listing in tables, the symbols have, wherever possible, been restricted to three letters, usually the first three letters of the trivial name.

Allose	= All	Arabinose	= Ara	Rhamnose	= Rha
Altrose	= Alt	Lyxose	= Lyx	Fucose	= Fuc
Galactose	= Gal	Ribose	= Rib		
Glucose	= Glc	Xylose	= Xyl	Fructose	= Fru
Gulose	= Gul			Neuraminic acid	= Neu
Idose	= Ido			Muramic acid	= Mur
Mannose	= Man				
Talose	= Tal				

Symbols derived from less common trivial names may be used, but the systematic name should be given together with the trivial name and the abbreviation at the first mention.

Examples: 3,6-Dideoxy-D-*xylo*-hexose (abequose = Abe)
6-Deoxy-D-glucose (quinovose = Qui)
3-*C*-(Hydroxymethyl)-D-*glycero*-aldotetrose (D-apiose = Api)

(*b*) The symbols represent the structural formulas of the compounds and also their names.

Reproduced from *J. Biol. Chem.*, Vol.257, No.7, pp.3347-3351 (1982)
by courtesy of American Society of Biological Chemists, Inc.

(*c*) The symbols represent the individual sugars or their residues. The use of the symbol to represent the free sugar is not recommended in textual material, but such use may occasionally be desirable in tables, diagrams and figures.

(*d*) The ring form is indicated, where necessary, by using the first letter of furanose, pyranose, or septanose, italicized and uncapitalized, added after the abbreviated name of the monosaccharide.

Examples: Arabinofuranose = Ara*f*
Glucopyranose = Glc*p*

(*e*) *Uronic Acid*—The suffix A is added to the symbol for the parent monosaccharide.

Examples: Glucuronic acid = GlcA
Galactopyranuronic acid = Gal*p*A

(*f*) *Deoxy Sugars*—Rational names, but no abbreviations, are recommended. One exception is 2-deoxy-D-*erythro*-pentose (deoxyribose) which is abbreviated dRib.

(*g*) *Aminodeoxy Sugars*—For 2-amino-2-deoxy sugars, the suffix N is added to the symbol of the parent monosaccharide. If the latter is *N*-acetylated, the suffix becomes NAc.

Examples: 2-Amino-2-deoxy-D-glucopyranose = D-Glc*p*N
2-Amino-2,6-dideoxy-L-galactose = L-FucN
2-Acetamido-2-deoxy-D-mannopyranose = D-Man*p*NAc

For other, less common, aminodeoxy sugars, the appropriate locants are added before N.

Example: 3,6-Bis(acetamido)-3,6-dideoxymannose = Man3,6(NAc)$_2$

(*h*) For anhydro sugars, the appropriate locants and the prefix An are added before the symbol of the parent monosaccharide.

Example: 3,6-Anhydrogalactose = 3,6AnGal

(*i*) *Configuration*—The configurational symbol (D or L) is included before the symbol for the monosaccharide, and is separated therefrom by a hyphen.

Examples: D-Glucopyranose = D-Glc*p*
L-Arabinofuranose = L-Ara*f*
3,6-Anhydro-D-galactose = 3,6An-D-Gal

(*j*) *Anomeric Configuration*—The anomeric symbol (α or β) is included before the configurational symbol and separated therefrom by a hyphen.

Examples: α-D-Glucopyranose = α-D-Glc*p*
β-L-Arabinofuranose = β-L-Ara*f*

(*k*) Structural formulas may be used for complicated features together with the abbreviated notation whenever necessary for clarity.

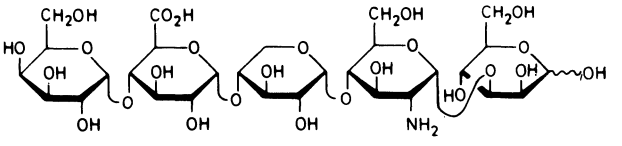
4. *Unbranched Oligosaccharides*

Between the symbol (abbreviated name) of one monosaccharide group or residue and the next are placed two locants that indicate the respective positions involved in this glycosidic union. These locants are separated by an *arrow* (directed from the locant corresponding to the glycosyl carbon atom to the locant corresponding to the carbon atom carrying the hydroxyl group involved) and are enclosed in parentheses (see Rule Carb-40 (1)). For nonreducing oligosaccharides, *double-headed arrows* are used between the locants of the appropriate glycosyl carbon atoms when the symbols are used, but not when the names are spelled out.

The hyphens, except that separating the configurational symbol and the symbol for the monosaccharide, may be omitted, *e.g.*

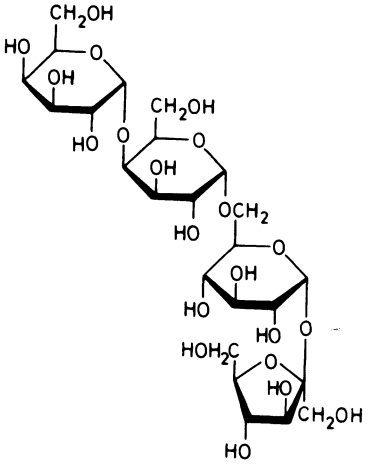
αD-Gal*p*(1 → 4)αD-Glc*p*A(1 → 4)αD-Xyl*p*(1 → 4)αD-Glc*p*N(1 → 3)D-Man

Examples:



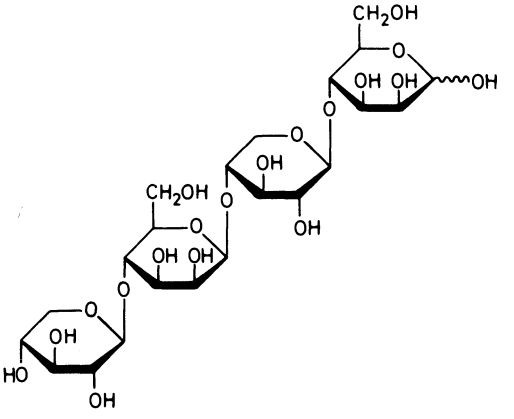
α-D-Gal*p*-(1 → 4)-α-D-Glc*p*A-(1 → 4)-α-D-Xyl*p*-(1 → 4)-α-D-Glc*p*N-(1 → 3)-D-Man*p*

I



α-D-Gal*p*-(1 → 4)-α-D-Gal*p*-(1 → 6)-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*

II



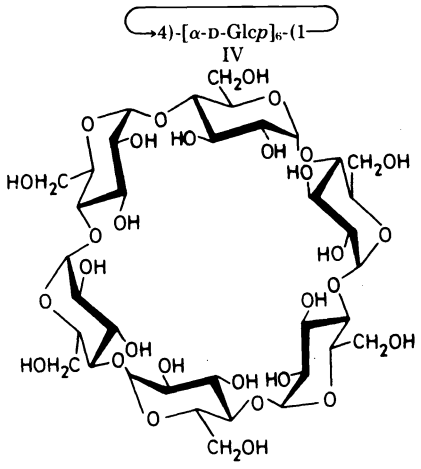
β-D-Xyl*p*(1 → 4)-β-D-Man*p*-(1 → 4)-β-D-Xyl*p*-(1 → 4)-D-Man*p*

III

5. *Cyclic Oligosaccharides*

A cyclic oligosaccharide is symbolized as illustrated below.

Example: Cyclomaltohexaose (the older names cyclohexaamylose, Schardinger α-dextrin or α-cyclodextrin, are not recommended).



→4)-[α-D-Glc*p*]$_6$-(1→

IV

6. *Branched and Substituted Oligosaccharides*

(*a*) Substituents may be symbolized as follows:

acetyl	= Ac	methyl	= Me
benzoyl	= Bz	phenyl	= Ph
benzyl	= Bzl, $PhCH_2$ (Bn)	*p*-toluenesulfonyl tosyl	= Tos (Ts)
ethyl	= Et		
glycolyl	= Gl	trimethylsilyl	= Me_3Si
methanesulfonyl mesyl	= $MeSO_2$ (Ms)	trityl	= Trt, Ph_3C (Tr)

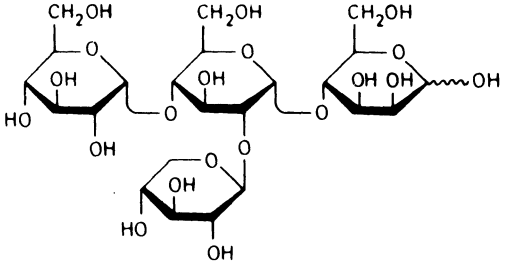
Most of these symbols are in the lists of those recommended for biochemical use (3) or for symbolizing derivatives of amino acids (4). Some symbols that are commonly used in the carbohydrate literature are given in parentheses. Substituents will not carry prefixes if attached to oxygen or nitrogen, but will be preceded by an italicized *C* if attached to carbon. The position of the substituent will be shown by the appropriate numeral. Substituents directly follow the abbreviation for the monosaccharide residue.

Examples:

Ethyl D-glucopyranuronate	= D-Glc*p*A6Et
β-D-Galactopyranose 4-sulfate	= β-D-Gal*p*4SO_3^-
2-*C*-Methyl-D-xylose	= D-Xyl2*C*Me
3,4-Di-*O*-methyl-L-rhamnose	= L-Rha3,4Me_2
N-Acetylneuraminic acid	= Neu5Ac
N-Acetyl-2-deoxyneur-2-enaminic acid	= Neu2en5Ac

(*b*) For branched oligosaccharides, the main chain and oligosaccharide side chains will be depicted as outlined for unbranched oligosaccharides. The position of a branch is indicated above, or below, the main chain, with the numerals and an *arrow* indicating the glycosidic linkage.

Examples:



α-D-Glc*p*-(1 → 4)-α-D-Glc*p*-(1 → 4)-D-Man*p*

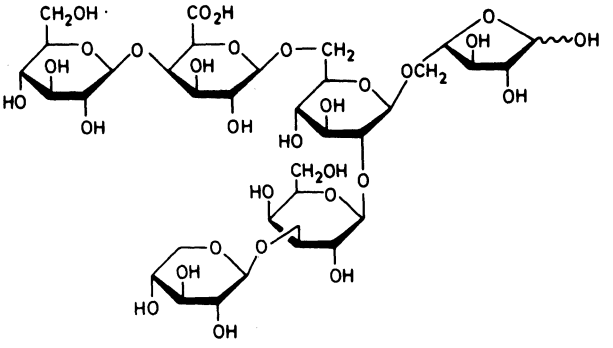
2

↑

1

β-D-Xyl*p*

V



β-D-Glc*p*-(1 → 4)-β-D-Gal*p*A-(1 → 6)-β-D-Glc*p*-(1 → 5)-L-Ara*f*

2

↑

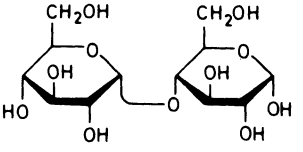
1

β-D-Xyl*p*-(1 → 3)-β-D-Gal*p*

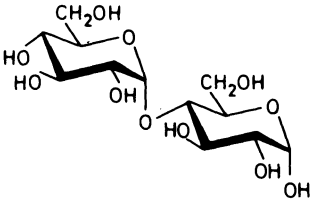
VI

(*c*) In drawing structural formulas, Haworth perspective formulas (formula 1), conformational formulas (formula 2) or Mills (5) formulas (formula 3) may be used (1, 6). It should be borne in mind that the detailed conformation indicated by a conformational formula has not always been established. Mills formulas are particularly useful in describing synthetic work. Use of a Fischer projection (formula 4) is sometimes advisable.

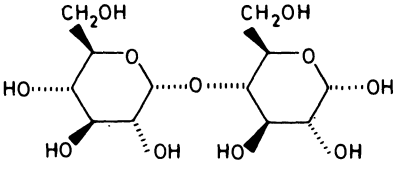
Examples:



Formula 1

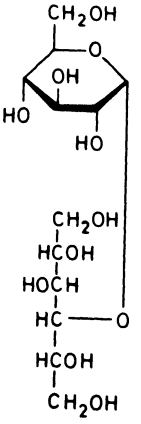


Formula 2



Formula 3

α-Maltose



Maltitol, 4-*O*-α-D-glucopyranosyl-D-glucitol

Formula 4

Appendix

The Condensed System of Symbolism of Sugar Residues in Oligosaccharides and Oligosaccharide Chains

In the condensed system the common configuration and ring size are implied in the symbol. Thus, Glc means D-glucopyranose; Fru, D-fructofuranose; and Fuc, L-fucopyranose. Whenever the configuration or ring size is found to differ from the common one, or is to be emphasized, this may be indicated by using the appropriate symbols from the extended system.

The anomeric descriptor indicates the configuration of the glycoside linkage, and is therefore placed before the locant if the direction of the bond is to the right, or after the locant if the direction of the bond is to the left. The two locants are separated by a hyphen. No hyphens are used between the symbol for the sugar and the parentheses indicating the glycosidic bond.

Example: Raffinose = Gal(α1-6)Glc(α1-2β)Fru

The parentheses may be omitted in representing branched oligosaccharides, when parentheses are used to indicate the branches. In this way it is possible to write branched sequences on one line, as shown in the examples.

Comment—The main difference between the extended form and the condensed form is the place of the anomeric descriptor, α or β. In the extended form the anomeric descriptor is considered to be part of the name (symbol) of the sugar unit; this system is preferred by carbohydrate chemists. In the condensed system the anomeric descriptor specifies the type of glycosidic linkage. This usage, first codified in *Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry* (7), is preferred by many biochemists. JCBN has been unable to reach a consensus as to which system should be recommended for general use, so gives both systems here as optional.

As in the extended system, placing a hyphen or parenthesis to the right of the symbol for a monosaccharide residue signifies removal of OH from the reducing carbon. Thus amygdalin may be represented:

Extended: β-D-Glc*p*-(1 → 6)-β-D-Glc*p*-O-CH(CN)-Ph
Condensed: Glc(β1-6)Glc(β)-O-CH(CN)Ph

This also applies in representing glycolipids (Ref. 2, Section Lip-3).

The following examples illustrate the use of the two different systems. They refer to the numbered structural formulas given above.

Comment—For long or multiple branches it may be advisable to use the two-line notation even in the condensed system.

Structure I

Extended: α-D-Gal*p*-(1 → 4)-α-D-Glc*p*A-(1 → 4)-α-D-Xyl*p*-(1 → 4)-α-D-Glc*p*N-(1 → 3)-D-Man*p*
Condensed: Gal(α1-4)GlcA(α1-4)Xyl(α1-4)GlcN(α1-3)Man

Structure II

Extended: α-D-Gal*p*-(1 → 4)-α-D-Gal*p*-(1 → 6)-α-D-Glc*p*-(1 ↔ 2)-β-D-Fru*f*
Condensed: Gal(α1-4)Gal(α1-6)Glc(α1-2β)Fru

Structure III

Extended: β-D-Xyl*p*-(1 → 4)-β-D-Man*p*-(1 → 4)-β-D-Xyl*p*-(1 → 4)-D-Man*p*
Condensed: Xyl*p*(β1-4)Man(β1-4)Xyl*p*(β1-4)Man

Note—In this case Xyl*p* is used in the condensed system to stress the pyranose form.

Structure IV

Extended: →4)-[α-D-Glc*p*]$_n$-(1 or [4Glcα1]$_n$
Condensed: [→4Glcα1→]$_n$

Structure V

Extended: α-D-Glc*p*-(1 → 4)-α-D-Glc*p*-(1 → 4)-D-Man*p*
2
↑
1
β-D-Xyl*p*
Condensed: Glcα1-4(Xylβ1-2)Glcα1-4Man

Structure VI

Extended: β-D-Glc*p*-(1 → 4)-β-D-Gal*p*A-(1 → 6)-β-D-Glc*p*-(1 → 5)-L-Ara*f*
2
↑
1
β-D-Xyl*p*-(1 → 3)-β-D-Gal*p*
Condensed: Glcβ1-4GalAβ1-6(Xylβ1-3Galβ1-2)Glcβ1-5Ara
or
Glcβ1-4GalAβ1-6Glcβ1-5Ara
2
↑
Xylβ1-3Galβ1

REFERENCES

1. *Tentative Rules for Carbohydrate Nomenclature, Part I, 1969. Biochemical Nomenclature and Related Documents*, 3rd Ed, 1978, IUB, pp. 174-195; *Biochem. J.* **125**, 673-695 (1971); *Biochemistry* **10**, 3985-4004 and 4995 (1971); *Biochim. Biophys. Acta* **244**, 223-302 (1971); *Eur. J. Biochem.* **21**, 455-477 (1971); *J. Biol. Chem.* **247**, 613-635 (1972)
2. *The Nomenclature of Lipids, 1976. Biochemical Nomenclature and Related Documents*, 3rd Ed, 1978, IUB, pp. 122-132; *Biochem. J.* **171**, 21-35 (1978); *Eur. J. Biochem.* **79**, 11-21 (1977); *Hoppe-Seyler's Z. Physiol. Chem.* **358**, 617-631 (1977)

3. *Abbreviations and Symbols. Biochemical Nomenclature and Related Documents,* 3rd Ed, 1978, IUB, pp. 57-61; *Eur. J. Biochem.* **74,** 1-5 (1977)
4. *Symbols for Amino-Acid Derivatives and Peptides, 1971. Biochemical Nomenclature and Related Documents,* 3rd Ed, 1978, IUB, pp. 78-84; *Arch. Biochem. Biophys.* **150,** 1-8 (1972); *Biochem. J.* **126,** 773-780 (1972); *Biochemistry* **11,** 1726-1732 (1972); *Biochim. Biophys. Acta* **263,** 205-212 (1972); *Eur. J. Biochem.* **27,** 201-207 (1972); *J. Biol. Chem.* **247,** 977-983 (1972); *Pure Appl. Chem.* **40,** 315-331 (1974)
5. Mills, J. A. (1955) *Adv. Carbohydr. Chem.* **10,** 1-53
6. *Section E, Stereochemistry* in *Nomenclature of Organic Chemistry* (1979) pp. 473-490, Pergamon, Oxford; also in *Biochemical Nomenclature and Related Documents,* 3rd Ed, 1978, IUB, pp. 1-18
7. *Abbreviations and Symbols for Chemical Names of Special Interest in Biological Chemistry, 1965. Biochemical Nomenclature and Related Documents,* 3rd Ed, 1978, IUB, pp. 52-56; *Arch. Biochem. Biophys.* **136,** 1-12 (1970); *Biochem. J.* **101,** 1-7 (1966); *Biochemistry* **5,** 1445-1453 (1966); *Bull. Soc. Chim. Biol.* **50,** 3-20 (1968) in French; *Eur. J. Biochem.* **1,** 259-266 (1967); *Hoppe-Seyler's Z. Physiol. Chem.* **348,** 245-255 (1967) in German; *J. Biol. Chem.* **241,** 527-533 (1966); *Mol. Biol.* **1,** 872-885 (1967) in Russian.