

NOMENCLATURE OF CARBOHYDRATES

By R. STUART TIPSON*

National Bureau of Standards, Washington 25, D. C.

In 1939, because of the confusion then existing in the naming of carbohydrates, a committee was appointed by the American Chemical Society Divisions of Sugar Chemistry and Technology, Chemical Education, and Biological Chemistry to develop a rational system of carbohydrate nomenclature. Carbohydrate chemists were widely canvassed for their opinions. The Committee reported the resulting rules in 1944 at the April ACS meeting; they were later modified, were approved by the ACS Council in April 1948, and were published in Chemical and Engineering News.¹

Later, the Division of Carbohydrate Chemistry appointed a new Committee on Carbohydrate Nomenclature, which re-examined the rules, and extended and partly revised them as a result of experience gained and suggestions received in the interim. A final draft of these rules was made in 1952, after excellent collaboration with a counterpart British committee. They were adopted by the ACS on a tentative basis and published,² and were simultaneously adopted and published by The Chemical Society.³

The term "sugar" has been used in two ways. Originally, it was specific for cane or beet sugar (sucrose), which, of course, is still the layman's usage. It then was extended to grape sugar (glucose), fruit sugar (fructose), and eventually to any mono- or oligo-saccharide or derivative thereof.

The "carbohydrates" may be regarded as⁴ comprising (a) aliphatic and alicyclic polyhydroxy aldehydes and polyhydroxy ketones (and derivatives thereof), called simple sugars or monosaccharides; and (b) their condensation products (formed by elimination of water molecules), called oligosaccharides and polysaccharides. The oligosaccharides are subdivided into disaccharides, trisaccharides, tetrasaccharides, etc.

In developing any system of nomenclature, one is confronted immediately with the legacy of the past; and experience has shown that it is usually wise not to tamper with this any more than necessary. Accordingly, we have used as much of it as possible, as a foundation.

First of all, there are common or trivial names. Of course, common names, such as sucrose, lactose, and maltose, will always find use because they are concise and convenient; but, since those names were given to compounds before we knew what they were chemically, the names naturally cannot be rational. To assign a name by a rational system, one must adhere to one great criterion--namely, that, given the

name, the formula can be written, and, given the formula, the name can be written.

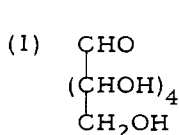
The rules of carbohydrate nomenclature were devised primarily for the English language.

Rule 1 states that "carbohydrate nomenclature should follow the general principles of established organic nomenclature."

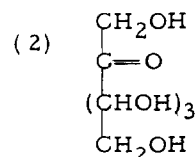
Rule 2 states that "as few changes as possible will be made in terminology universally adopted."

(Rule 3) The reducing sugars fall into two classes. In an aldose, the reducing function is an aldehyde, and its carbon atom is atom number one. In a ketose, the reducing function is a ketone, and the carbonyl carbon atom is assigned the lower possible number.

For indicating the number of carbon atoms in the normal chain, the appropriate one of these names is used: triose, tetrose, pentose, hexose, heptose, octose, nonose, decose, etc.

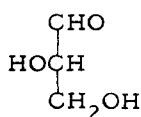


An aldohexose
(1-, NOT 6-)

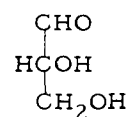


A ketohexose
(2-, NOT 5-)

Now, the simplest possible sugar having an asymmetric carbon atom is a triose, glyceraldehyde. It can exist in two optically active forms



L-Glycerose



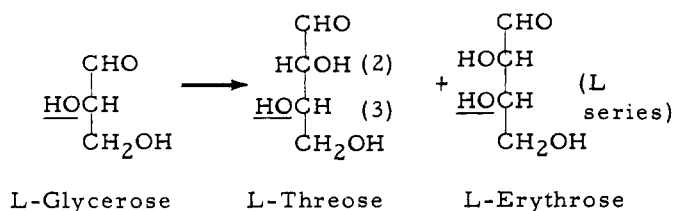
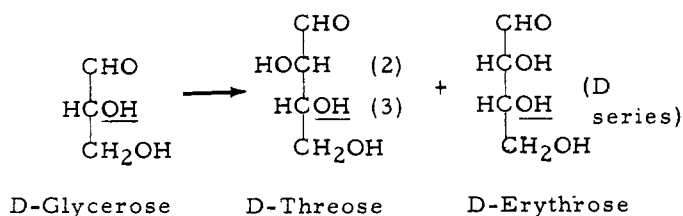
D-Glycerose

The formula on the right was assigned arbitrarily to D-glycerose; this has proved to be a lucky guess, as Bijvoet and co-workers (1949, 1951) have proved that the established convention conforms to reality, and so this formula describes an absolute orientation.

(Rule 4) The small capital letters D and L are pronounced "dee" and "ell" and are not to be confused with optical rotation; if this is to be indicated, it is done by adding (dextro) or (levo), or by adding (+) or (-). Racemic modifications are indicated by the prefixes DL or (\pm).

By a process of chain extension, any higher sugar can be built up either from the L- or the D-glycerose, as indicated

*Presented before joint meeting of the Division of Chemical Literature and Division of Organic Chemistry, American Chemical Society Meeting, Chicago, Illinois, September 8, 1958.



Hence, Rule 5 states that "carbohydrates having the same configuration of the highest numbered asymmetric carbon atom as that of D-glycerose belong to the D-configurational series; those having the opposite configuration belong to the L-series."

(Rule 6) Each sugar has one or more asymmetric carbon atoms. Instead of naming each active center individually, as with the steroids, we have retained earlier practice in the sugar field, and we name all the centers collectively. These names were devised by the simple process of lopping off the final "-se" from the names already used for the fifteen aldose sugars from triose through hexose. The configuration of a group of consecutive, but not necessarily contiguous, asymmetric carbon atoms is designated by the appropriate one of these prefixes:

Asymmetric carbon atoms	Prefixes
one	<u>glycero</u>
two	<u>erythro</u> , <u>threo</u>
three	<u>arabino</u> , <u>lyxo</u> , <u>ribo</u> , <u>xylo</u>
four	<u>allo</u> , <u>altro</u> , <u>galacto</u> , <u>gluco</u> , <u>gulo</u> , <u>ido</u> , <u>manno</u> , <u>talo</u>

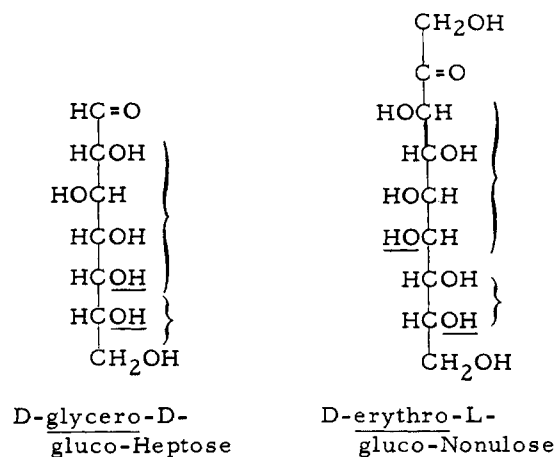
There are thus fifteen configurations named collectively, and each may be D or L, giving a total of thirty. At this point we stop; all higher collections of asymmetric carbon atoms are named (Rule 22) by use of combinations of these prefixes.

In naming a compound, the configurational prefix is added as the last step: it immediately precedes the prefix giving the number of carbon atoms.

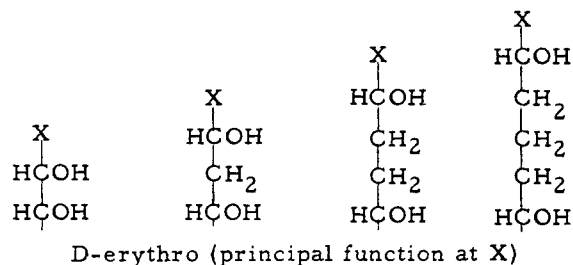
By formalizing the configurational prefixes, we have, in effect, provided, for the fifteen aldoses (from which they were derived), acceptable, systematic names which may be useful to students.

Trivial	Systematic
D-Glycerose	D- <u>glycero</u> -Triose
D-Erythrose	D- <u>erythro</u> -Tetrose
L-Arabinose	L- <u>arabino</u> -Pentose
D-Glucose	D- <u>gluco</u> -Hexose

The following illustrates application to sugars higher than hexoses.

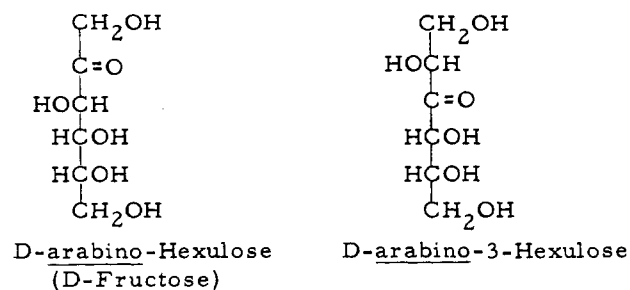


Since the configurational prefix describes a group of consecutive, but not necessarily contiguous, asymmetric carbon atoms, the next examples are all D-erythro.



An inability to apply Rule 6 is responsible for erroneous names applied by some workers to ketoses and deoxy sugars, which I shall next discuss.

(Rule 7) Ketoses having the carbonyl group at carbon atom number two are named by means of the suffix "-ulose." If the carbonyl group is at some other position, the sugar is named by adding the appropriate numeral.



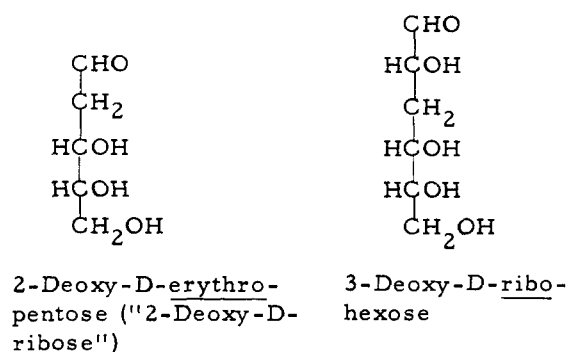
However, certain names are still being used for ketoses which are inacceptable to the organic chemist as definitive names, because they have incorrect structural implications:

<u>Inacceptable</u>	<u>Systematic</u>
Glycerulose	1,3-Dihydroxy-2-propanone
D-Erythrulose	D- <u>glycero</u> -Tetrolulose
D-Threulose	
D-Arabinulose	D- <u>erythro</u> -Pentulose
D-Ribulose	
D-Lyxulose	D- <u>threo</u> -Pentulose
D-Xylulose	

and the corresponding names for the L forms

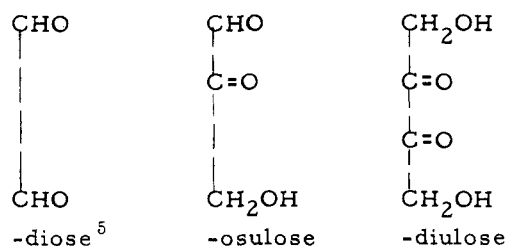
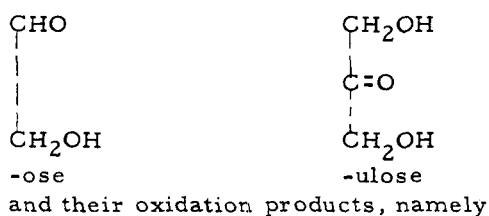
Each name in the left column shows one asymmetric carbon atom more than is actually present!

(Rule 8) The same kind of problem is encountered when we come to naming those sugars in which an alcoholic hydroxyl group is replaced by a hydrogen atom. We have chosen to name these by means of the prefix "deoxy-" preceded by the appropriate numeral.



I have put "2-Deoxy-D-ribose" in quotation marks because it means "that 2-deoxypentose having the D-ribo configuration," which is an impossibility, since a 2-deoxypentose has only two hydroxyl groups on asymmetric carbon atoms.

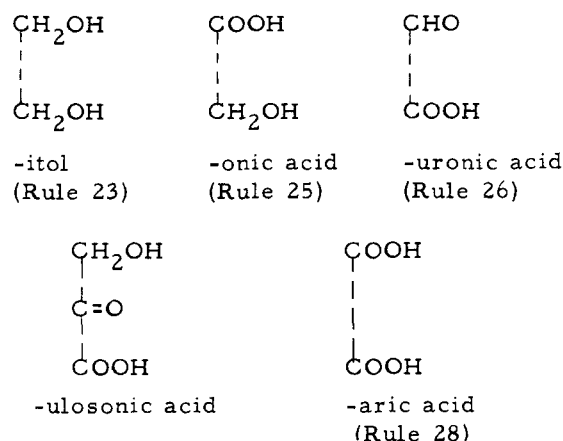
To return to Rule 7 for the sugars, the Committee is at present considering the proposal that a sugar which is both an aldose and a ketose will be named by replacing the final "e" in the name of the (formally) parent aldose by the suffix "-ulose," preceded by the appropriate configurational prefixes. (This and the next two proposals are purely tentative, and are not to be regarded as rules, as they have not been officially accepted.)



Another reasonable proposal under consideration is that ketoses having two ketonic carbonyl groups will be named by means of the suffix "-diulose."

A third proposal under consideration is that dialdehydes, formed from aldoses by oxidation of the terminal hydroxymethyl group to -CHO, be named by use of the suffix "-diose."⁵

We may now consider the names assigned to the products obtained on (a) reduction of an aldose or (b) further oxidation of the above compounds.



None of these endings is new to sugar chemistry. We merely recommend using them systematically, and many very old names would remain unchanged: for example, mannitol, mannose, mannonic acid, and mannuronic acid. However, for use in discussions of sugar chemistry, certain well-established names would require change. And, to be completely fair, I will give you the example which would need almost complete changes--the series, derived from D-glyceraldehyde, of compounds which have been thoroughly studied outside sugar chemistry. The names would be:

<u>Suggested Name</u>	<u>Usual Name</u>
Glyceritol	Glycerol
D-Glycerose	D-Glyceraldehyde
Glycerodiose ⁵	Tartronaldehyde
D-Glyceronic acid	D-Glyceric acid
D-Glyceruronic acid	Tartronaldehydic acid
Glyceraric acid	Tartronic acid

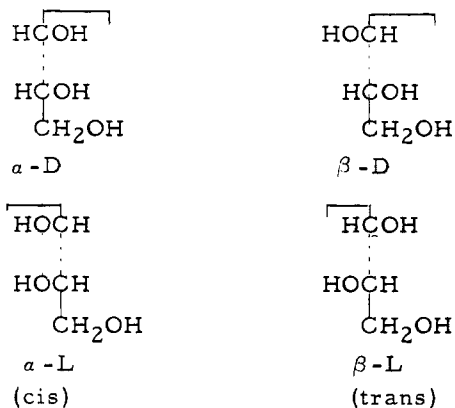
In certain areas of overlap between specialized sugar chemistry and general organic chemistry, two acceptable names for a single

compound may be desirable--one for the sugar chemist and the other for the general organic chemist.

(Rule 17) The acyclic nature of a sugar or derivative containing an uncyclized CHO or CO group as the primary function is indicated by inserting the italicized prefix aldehydo or keto, respectively, immediately before the configurational prefix and stem name, as in: aldehydo-D-glucose pentaacetate; keto-D-fructose pentabenzoate.

(Rule 15) Now, so far, I have depicted the sugars as open chains; but, as you know, they and their derivatives can exist as ring forms, any one of which has two anomeric forms, called α and β . This anomeric prefix, which can only be used in conjunction with a configurational prefix (D or L), immediately precedes the latter.

In the D-series, the more dextrorotatory anomer is designated α and the hemiacetal hydroxyl group of the anomeric carbon atom is written to the right in the Fischer projection formula. The less dextrorotatory anomer is designated β and the hydroxyl group is written to the left. The reverse holds for the α and β anomers in the L series.



The steric arrangement of the groups on the α -D anomeric carbon atom is the same as that on the β -L anomeric carbon atom; that on the α -L anomeric carbon atom is the same as that on the β -D anomeric carbon atom. The α forms are cis, and the β forms are trans, as regards the relation of the anomeric to the configurational carbon atom.

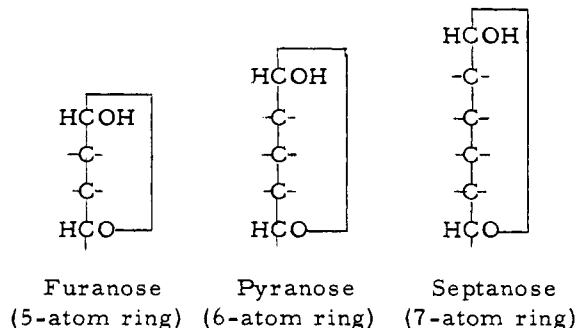
(Rule 19) The root "glyc" is used in a generic sense, to denote any sugar or derivative thereof; as in "glycose" and "glycoside."

(Rule 18) The size of the ring in the heterocyclic forms of monosaccharides may be indicated by replacing, in the sugar name, the letters "se" by

"furanose" for the 5-atom ring
 "pyranose" for the 6-atom ring, and
 "septanose" for the 7-atom ring

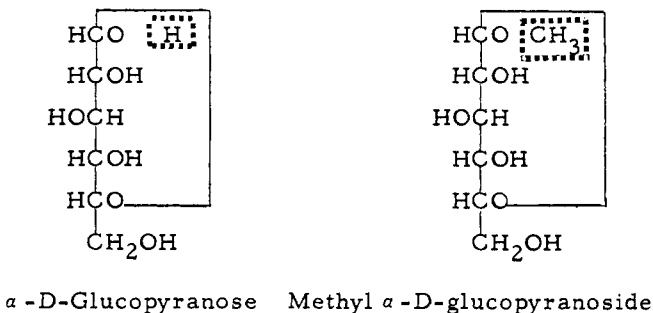
Likewise, for the glycosides, the size of the ring may be shown by replacing the syllable

"-side" by "furanoside," "pyranoside," or "septanoside."

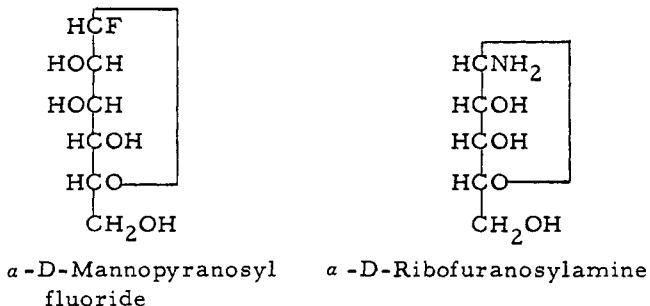


If the ring size is known, the appropriate one of these should be used, unless the numbering of substituents makes it obvious. For example, 2,3,5-tri-O-benzoyl-D-ribofuranose can be contracted to 2,3,5-tri-O-benzoyl-D-ribose.

(Rule 20) A glycoside is a mixed acetal resulting from the exchange of an organic radical for the hydrogen atom of the hemiacetal hydroxyl group of a cyclic form of a sugar. It is named by substituting "ide" as a suffix in place of the terminal "e" of the corresponding sugar name; before this word is placed the name of the organic substituent, separated by a space (Rule 12).



(Rule 21) If the hemiacetal hydroxyl group is detached from a cyclic modification of a sugar, the residue is a glycosyl radical; it may be glycofuranosyl, glycopyranosyl, or glycoseptanosyl. Note that a glycosyloxy radical is identical with a glycoside radical.



(Rule 9) Now, as regards naming of substituted sugars, it so happens that substitution is almost always on oxygen, and only occasionally on carbon or nitrogen. When the hydrogen atom of an alcoholic hydroxyl group of a carbohydrate

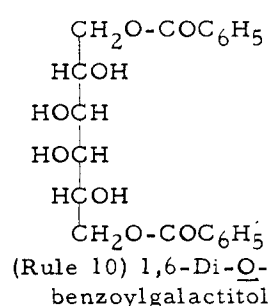
is substituted, an italic capital letter *O* (for oxygen) is attached by a hyphen directly before the substituting prefix. The *O* prefix need not be repeated for multiple substitution by one group. Similar principles apply to substitution on nitrogen (prefix, *N*), and the prefix *C* may be used to indicate substitution on carbon, to avoid possible ambiguity.

Furthermore, (Rule 10) the name of the parent compound may be retained as the root for the substituted compound. We thus have such names as

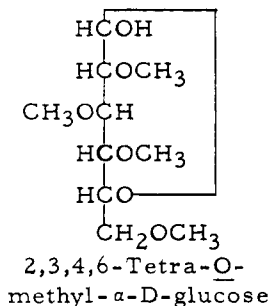
1,6-di-*O*-benzoylgalactitol

2,3,4,6-tetra-*O*-methyl- α -D-glucose

(Rule 11) Esters also may be named in an alternative way; and so may ethers (Rule 13).



(Rule 11) Galactitol
1,6-dibenzoate



(Rule 13) α -D-Glucopyranose tetramethyl ether

Rule 14 states that substituents shall be written into the name in alphabetical order: acetyl-anhydro-deoxy-, etc. (regardless of their number: tri-*O*-acetyl-dianhydro-(mono)deoxy-) and Rule 16 states that hyphens shall be used in conformity with established practice.

I shall omit consideration of Rules 27 and 30-34, which are concerned with rather specialized problems in sugar nomenclature, and pass to the final rule (Rule 35) which is concerned with naming disaccharides, trisaccharides, etc.

Non-reducing disaccharides are named as glycosyl glycosides.

REFERENCES

¹Chem. Eng. News, 26, 1623 (1948).

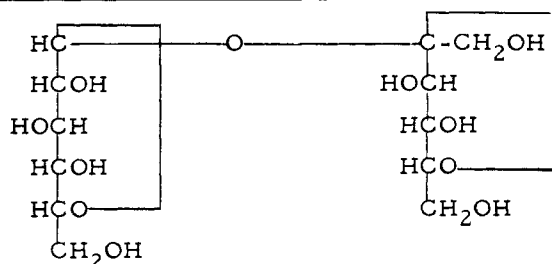
²Chem. Eng. News, 31, 1776 (1953).

³J. Chem. Soc., 5108 (1952).

⁴On matters not explicitly covered by the Rules, the opinions expressed are those of the author.

⁵Recently changed to "dialdose."

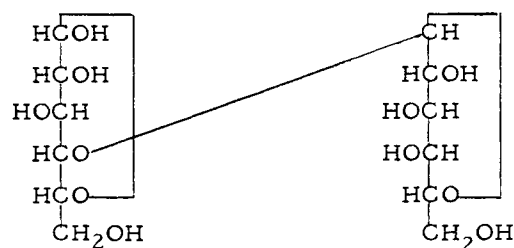
Non-reducing Disaccharide



Sucrose: β -D-Fructofuranosyl α -D-glucopyranoside
or α -D-Glucopyranosyl β -D-fructofuranoside

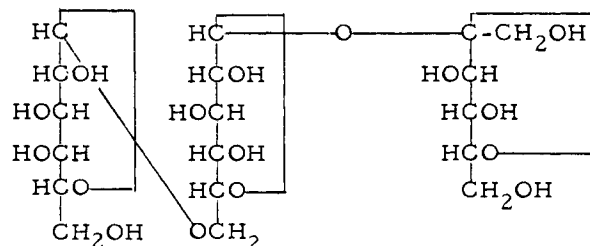
A reducing disaccharide is named as an *O*-glycosylglycose.

Reducing Disaccharide



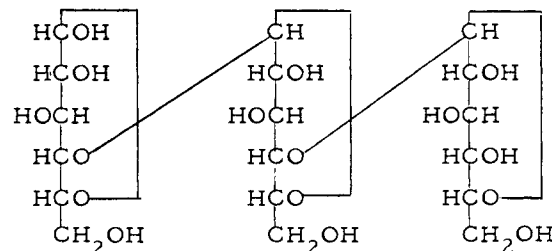
α -Lactose: 4-*O*- β -D-Galactopyranosyl- α -D-glucopyranose
or *O*- β -D-Galactopyranosyl-(1 \rightarrow 4)- α -D-glucopyranose

Non-reducing Trisaccharide



Raffinose: *O*- α -D-Galactopyranosyl-(1 \rightarrow 6)- α -D-glucopyranosyl β -D-fructofuranoside

Reducing Trisaccharide



α -Cellotriose: *O*- β -D-Glucopyranosyl-(1 \rightarrow 4)-*O*- β -D-glucopyranosyl-(1 \rightarrow 4)- α -D-glucopyranose