

# System for Classification of Structurally Related Carbohydrates<sup>1</sup>

Horace S. Isbell

A system is presented for the classification of structurally and configurationally related carbohydrates. Each substance is assigned a code number that defines the structure and configuration. By inspection of the code numbers, or by a punched-card technique, groups of structurally related carbohydrate derivatives can be selected readily from a heterogeneous collection. The structures, configurations, and conformations of pyranose and furanose derivatives are discussed, and classifications are made on the basis of a few fundamental structures.

Certain ambiguous and objectionable features in the classification of pyranose ring conformations in the C1 and 1C categories of Reeves are pointed out. In this paper, C1 denotes the chair conformation in both the D and the L aldohexose series in which the reference group attached to carbon 5 of the ring is in the equatorial position. Similarly, C2 denotes the chair conformation in both series in which the reference group attached to carbon 5 of the ring is in the axial position. Xylose and sorbose are classified like glucose; lyxose and tagatose like mannose; arabinose and fructose like galactose; and ribose and psicose like talose. Because the designations are independent of the D or L series, they avoid the erroneous classification of enantiomorphs in different conformations.

## 1. Introduction

During the past several years the infrared absorption spectra of a large group of carbohydrate derivatives have been measured at the Bureau with the object of providing reference spectra and data for structurally related materials. A classification system was devised to show the structure and configuration of the compounds by means of numbers suitable for coding and separating by punched-card techniques. Although the system was developed primarily for comparing infrared absorption spectra, it can be used for classifying structurally related carbohydrates for any purpose.

The code numbers are assigned by means of a key outlined in tables 1 and 2. A decimal point is inserted after the second digit to separate figures that provide broad generic classification from those that show definite structure. For example,  $\alpha$ -D-glucopyranose is given the code number 10.2110. Reading from left to right, 1 shows that the substance is a monosaccharide; 0, that the hydroxyl groups are predominantly unsubstituted; 2, that it has a primary 6-carbon structure; 1, that it has the glucose configuration; 1, that it has a C1 pyranose ring with an axial glycosidic group, and 0 that the glycosidic hydroxyl is not substituted. The code numbers for all  $\alpha$ -D-glucopyranose structures will have the .211 sequence. Each structural grouping has a characteristic sequence of numbers. Hence by inspection of the numbers, or by punched-card technique, groups of structurally and configurationally related carbohydrate derivatives can be readily selected. The classifications are not intended for general nomenclature and no changes are proposed in the rules for naming carbohydrates. In subse-

quent paragraphs the structural elements involved in classification will be considered in order, and directions will be given for assigning code numbers.

## 2. Components of the Code Number

### 2.1. Type of Carbohydrate

The first digit in the code number divides the carbohydrates into groups of increasing complexity: 1 signifies a monosaccharide; 2, a disaccharide; 3, an oligosaccharide; 4, a polysaccharide composed of one type of unit; and 5, a polysaccharide composed of a variety of units. Substances in types 1 to 4 are given definitive code numbers but no attempt is made to classify substances in 5, beyond coding the structural units present.

### 2.2. Major Substituent

The second digit is assigned to show substitution on the polyol structure. When half or more of the hydroxyls are free, the classification number is 0; when more than half of the hydroxyls are substituted the number corresponds to the substituent present in predominating amount. If two classifications are possible, the one with the lower number is used.

### 2.3. Carbon Skeleton

The first digit to the right of the decimal point defines the skeleton of the fundamental carbon chain. Alcohols, aldoses, aldonic acids, uronic acids, etc., are classified as primary; ketoses and related compounds as secondary. Cyclitols and higher sugars are included in a miscellaneous group, 9. The aldohexoses are placed in two groups designated as 7a and 7b. Substances in the 7a group have like configurations for the two configurational units given in the conventional ACS name [1];<sup>2</sup> those in the 7b

<sup>1</sup> The work reported in this publication was sponsored by the Office of Naval Research as part of a program on the investigation of the structure, configuration, and ring conformation of the sugars and their derivatives by infrared absorption measurements (NRO 55208). The classification system presented here has been used to classify the infrared spectra of about 200 carbohydrate derivatives. The spectra and reference numbers will appear in forthcoming publications.

<sup>2</sup> Figures in brackets indicate the literature references at the end of this paper.

TABLE 1. Classification key for code numbers

Code	1. Type of carbohydrate	2. Major substituent	3. Carbon skeleton	4. Configuration	5. Characteristic unit	6. Substituent on main functional group <sup>a</sup>
0	Monosaccharide	Not substituted	<5 primary	Racemic	Anomeric mixture	Not substituted.
1	Disaccharide	Methylated	5 primary	<i>gluco</i>	a-Pyranose C1	[See table 2 for list of substituents.]
2	Oligosaccharide	Acetylated	6 primary	<i>manno</i>	e-Pyranose C1	
3	Polysaccharide	Nitrated	7a primary	<i>galacto</i>	a-Pyranose C2	
4	(simple)		7b primary	<i>talo</i>	e-Pyranose C2	
5	Polysaccharide (complex)		<5 secondary	<i>ido</i>	Other pyranose	
6			5 secondary	<i>gulo</i>	Furanose	Acid derivative
7			6 secondary	<i>altro</i>	Glycitol	
8			7 secondary	<i>allo</i>	Aldehyde or keto derivative.	
9	Miscellaneous	Miscellaneous	Miscellaneous	Miscellaneous	Acid derivative	

<sup>a</sup> Asterisk is used to indicate linkage in higher saccharides.

TABLE 2. Numbers for substituent groups and auxiliary structures

Code	Substituent group	Code	Substituent group	Code	Substituent group	Code	Auxiliary structure
0	None.	30	Miscellaneous esters.	50	Nitrogen derivatives.	70	Carboxylic acid.
10	OH group.	31	Phosphate.	51	Amino.	71	Carboxylic ester.
11	O-Methyl.	32	Nitrate.	52	Acetamido.	72	Carboxylic $\gamma$ -lactone.
12	O-Ethyl.	33	Sulfate.	53	N-Glycosyl.	73	Carboxylic $\delta$ -lactone.
13	O-Cyclohexyl.	34	O-Tosyl.	54	Phenylhydrazine derivative.	74	Carboxylic amide.
14	O-Benzyl.	35	O-Mesyl.	55	Oxime.	75	Carboxylic salt.
15	O-Phenyl.	36		56		76	Oxo derivative.
16	O-Trityl.	37		57		77	
17	Hemiacetal.	38		58		78	
18	Aldehyde.	39		59		79	
19		40	Linkages other than O or N.	60	Cyclic derivatives.	80	Deoxy.
20	Carboxylic ester.	41	Fluoride.	61	Isopropylidene.	81	Deoxyamino.
21	O-Acetyl.	42	Chloride.	62	Benzylidene.	82	Deoxyacetamido.
22	O-Benzoyl.	43	Bromide.	63	1-Methoxyethylidene.	*	
23		44	Iodide.	64	1-Methoxybenzylidene.	*	
24		45	Mercaptal.	65	Ethylene oxide.		
25		46		66	Butylene oxide.		
26		47		67	Amylene oxide.	97	Organic solvent of crystallization.
27							
28		48		68	Hexylene oxide.	98	Salt of crystallization.
29		49		69		99	Water of crystallization.

group have unlike configurations. Thus a *D-glycero-D-gluco*-aldoheptose or an *L-glycero-L-gluco*-aldoheptose is classified 7a, whereas a *D-glycero-L-gluco*-aldoheptose or an *L-glycero-D-gluco*-aldoheptose is classified 7b. The first pair of compounds as well as the second pair are enantiomorphous, and for evaluation of infrared absorption, do not require separate code numbers.

#### 2.4. Configuration

The second digit to the right of the decimal signifies the configuration of the main structural unit. Configurations listed in table 1 are assigned numbers 1 to 8 according to the scheme outlined in table 3. No distinction is made between *D* and *L* modifications because enantiomorphous substances give like absorption. Racemic substances are grouped under 0, and compounds not covered elsewhere are classified under 9. Because the pentoses and ketohexoses have one less asymmetric carbon than the aldohexoses, each can be considered to be related to two aldohexoses that differ in the configuration of the terminal asymmetric carbon. On the basis of similarity in properties, the pentoses and ketohexoses are classified under configurations 1 to 4 rather than 5 to 8 [3].

TABLE 3. Classification of configuration

Aldotetrose		Aldopentose		Ketohexose		Related aldohexoses	
Code	Configuration	Code	Configuration	Code	Configuration	Code	Configuration
2	Threose.	1	Xylose.	1	Sorbose.	1	Glucose.
2	Erythrose.	2	Lyxose.	2	Tagatose.	2	Mannose.
		3	Arabinose.	3	Fructose.	3	Galactose.
		4	Ribose.	4	Psicose.	4	Talose.
						5	Idose.
						6	Gulose.
						7	Altrose.
						8	Allose.

#### 2.5. Structure

The digit third to the right of the decimal represents the characteristic structure of the substance. Products containing anomeric forms in equilibrium, as, for example, amorphous sugars and materials not subject to specific classification, are represented by 0; pyranose and furanose structures by 1 to 6; glycitols by 7; open-chain aldehyde and ketone derivatives by 8; and aldonic acids by 9 regardless of whether they are present as the acid, salt, ester, amide, or lactone. Assignment of code numbers for the pyranose and furanose derivatives requires review of some problems of nomenclature and ideas of configurationally related carbohydrates (see sec-

tion 4). Briefly, however, pyranoses are classified according to ring conformation and the position of the glycosidic group with respect to the plane of the ring. Pyranoses a and e have, respectively, axial and equatorial glycosidic groups.<sup>3</sup> All furanose derivatives are represented by 6.

## 2.6. Main Functional Group

The figures occupying the fourth and fifth positions to the right of the decimal in the code number show the character of the main functional group, usually the glycosidic group. Because there are more than 10 common substituents, two spaces are allowed for this structural feature. The numbers given in table 2 are used to describe substitution not only of the main functional group, but also of other parts of the molecule, as described in section 3. The numbering system can be readjusted to fit any collection of carbohydrate derivatives.

The numbers representing the substituent groups are used with structures of diverse type. Thus a methyl group when used in conjunction with structures 1 to 6 (column 5 of table 1) represents a glycoside; with 7, an ether; with 8, an acetal; and with 9, an ester. Application of the system can be seen from the examples that follow:

Examples	Code
Methyl $\alpha$ -D-xylopyranoside.....	10. 11111
Methyl $\alpha$ -D-glucopyranoside.....	10. 21111
$\alpha$ -D-Glucopyranose 1-acetate.....	10. 21121
Methyl D-gluconate.....	10. 21911
$\alpha$ -D-Mannopyranosylamine.....	10. 22151
D-Mannose phenylhydrazine.....	10. 22854
D-Mannonic phenylhydrazide.....	10. 22954
D-Mannonic $\gamma$ -lactone.....	10. 22966
1,6-Anhydro- $\beta$ -D-altropyranose.....	10. 27268
$\beta$ -D-Mannopyranose 1,2-(methyl ortho-acetate).....	10. 22263

The numbers selected by the described methods constitute the *primary code number* and are sufficient for classifying simple substances. Various schemes are used for more complicated substances.

## 3. Representation of Complex Structures

Substitution on the main functional group is shown directly in the primary code number. Substitution on the carbon chain is indicated by the class number and shown more specifically by numbers following the primary code number. The class number suffices to show complete substitution by a single substituent. Thus 2,3,4,6-tetra-*O*-acetyl- $\alpha$ -D-glucose is 12.2110. A partially substituted structure is shown by a group of numbers to the right of the primary code number. The first number or numbers, given in parentheses, show the position of

substitution, the next number defines the substituent. Several substituents can be listed consecutively.

Examples	Code
2,3,4,6-Tetra- <i>O</i> -acetyl- $\alpha$ -D-glucose.....	12.2110
3- <i>O</i> -Methyl- $\alpha$ -D-glucopyranose.....	10.2110(3)11
2,3,4-Tri- <i>O</i> -acetyl- $\alpha$ -D-glucose.....	12.2110(6)10
2,3-Isopropylidene- $\beta$ -D-lyxopyranose.....	10.1220(2,3)61
Methyl $\beta$ -D-galactopyranoside 6-nitrate.....	10.23211(6)32
2- <i>O</i> -Methyl-6- <i>O</i> -trityl- $\alpha$ -D-glucopyranose.....	10.2110(2)11(6)16

Auxiliary or additional structures are treated like substituents. The main functional group is considered first, and the second (auxiliary structure) is shown by suitable numbers following its position in parentheses.

Examples	Code
2-Oxo-D-gluconic acid.....	10.2190(2)76
$\alpha$ -D-Glucofuranurono- $\gamma$ -lactone.....	10.2160(6)72
Salt of $\alpha$ -D-glucopyranuronic acid.....	10.2110(6)75
Ester of methyl $\alpha$ -D-glucopyranuronide.....	10.21111(6)71
2-Deoxy-2-amino- $\alpha$ -D-glucopyranose.....	10.2110(2)81
6-Deoxy-L-mannopyranose.....	10.2210(6)80

Structures in the main functional group and similar structures in auxiliary groups do not always have the same number. Thus, the gamma lactone of an aldonic acid is represented by . - - 66, whereas a gamma lactone present as an auxiliary group is indicated by 72 following the location given in parentheses. The first number, 66, shows the presence of the butylene oxide ring in a structure already defined, whereas 72 following the parentheses represents a complete lactone structure. Ordinarily, anhydrides involving the glycosidic carbon do not require numbers defining the point of substitution; thus  $\beta$ -D-glucopyranose 1,6-anhydride is 10.21268, in which 68 shows the presence of a hexylene oxide ring involving the main functional group. D-Fructofuranose-1,2-anhydride 10.736(1,2)65, however, requires the numbers in parentheses, to locate the oxide ring on the carbon skeleton. In any case of ambiguity, positions of substitution should be given.

Oligosaccharides are represented by a series of numbers corresponding to the units present. The glycosidic linkage is shown by an asterisk at the right of the code number for each unit. This shows that the structure defined by the preceding number is joined in glycosidic linkage to another carbohydrate unit. The parentheses are followed by the code number for the next unit. The names for the oligosaccharides used here are obtained by ACS nomenclature rule 35 [1]. For convenience in typing, hyphens are used in the code number instead of the arrows used in the names.

<sup>3</sup> In accord with the suggestion of Barton, Hassel, Pitzer, and Prelog [2], the terms axial and equatorial, abbreviated by a and e respectively, are used to denote positions of groups with respect to the plane of the ring. Bonds connecting axial groups are perpendicular to the mean plane of the ring; those connecting equatorial groups are directed outward at a small angle above or below the plane of the ring.

Some examples are

$\alpha$ -Cellortiose:

$\beta$ -D-Glucopyranosyl-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 4)- $\alpha$ -D-glucopyranose is 30.212\*(1-4).212\*(1-4).2110.

$\alpha$ -Maltotriose:

$\alpha$ -D-Glucopyranosyl-(1 $\rightarrow$ 4)- $\alpha$ -D-glucopyranosyl-(1 $\rightarrow$ 4)- $\alpha$ -D-glucopyranose is 30.211\*(1-4).211\*(1-4).2110.

Raffinose:

$\alpha$ -D-Galactopyranosyl-(1 $\rightarrow$ 6)- $\alpha$ -D-glucopyranosyl-(1 $\rightarrow$ 2)- $\beta$ -D-fructofuranoside is 30.231\*(1-6).211\*(1-2).736\*.

Sucrose:

$\alpha$ -D-Glucopyranosyl-(1 $\rightarrow$ 2)- $\beta$ -D-fructofuranoside is 20.211\*(1-2).736\*.

Lactose:

$\beta$ -D-Galactopyranosyl-(1 $\rightarrow$ 4)- $\alpha$ -D-glucopyranose is 20.232\*(1-4).2110.

A branched structure is represented by use of braces. For example,  $\alpha$ -D-glucopyranosyl-(1 $\rightarrow$ 4)-[ $\alpha$ -D-galactopyranosyl-(1 $\rightarrow$ 6)]- $\alpha$ -D-mannopyranose is

$$\left. \begin{array}{l} 30.211*(1-4) \\ .231*(1-6) \end{array} \right\}.2210.$$

Polysaccharides composed of one type of structural unit are represented by 4 in the first digit of the code number and a sequence of numbers showing the skeleton, configuration, and structure of the monosaccharide unit, followed by numbers showing the position, or positions of the glycosidic linkages. For example:

amylose is 40.211\*(1-4),  
amylopectin is 40.211\*(1-4)\*(1-6),  
dextran is 40.211\*(1-6)\*(1-3)\*(1-4).

The numbers in parentheses for dextran show the presence of glycosidic linkages at carbons 6, 3, and 4. The polymeric character of the substance is indicated by 4 at the left of the code number.

Hetero-polysaccharides are classified according to the various structures present. Thus a polysaccharide composed of  $\alpha$ -D-galactopyranose units,  $\beta$ -D-mannopyranose units, and  $\beta$ -L-arabinofuranose units is 50.231\*.222\*.136\*. The examples cover only a few of the many substances that can be classified; others will be given in future publications.

## 4. Discussion of Configuration and Conformation

### 4.1. Classification of Anomers

According to present nomenclature, when an anomeric pair of sugars or sugar derivatives belongs in the D configurational series, the more dextrorotatory member is called alpha, and the less dextrorotatory, beta. If the substances belong in the L series, the more levrorotatory member is called alpha. Assignment of the substance to the D or L series

depends on the configuration of the highest numbered carbon in the configurational unit. This carbon is number 4 in the pentose series and number 5 in the hexose series. Basing the alpha and beta names on different reference carbons results in a lack of correlation between the alpha or beta name and the structure of the anomeric carbon, especially for the pentoses, ketohexoses, and furanoses [3, 4, 5]. A change in the names ordinarily used for these substances is not desirable but for correlation of structurally related compounds a more fundamental relationship must be used. The stereomeric problems for furanose and pyranose derivatives are different; hence classification of the two structures will be treated separately.

### 4.2. Classification of Furanose Structures

The dimensions and normal bond angles of the atoms comprising the furanose ring are such that a planar ring can be formed with little strain. Although very little is known concerning the conformation of furanose rings, intramolecular, and crystal-line forces can cause distortion. Thus X-ray studies have shown that the furanose ring in crystalline sucrose is nonplanar [6]. There are few furanose structures available for study and no means for predicting probable conformation. Hence for the present it seems desirable to consider the ring as substantially planar, with considerable flexibility. With a planar furanose ring the different configurations for carbons 1 to 4 give 16 isomers, 8 in the D series, and their 8 enantiomorphs in the L series. The 8 representatives of either series consist of two modifications of each of 4 fundamental ring structures, that have, respectively, the *xylo*, *arabo*, *lyxo*, and *ribo* configurations. The two modifications of each of these, differing in the configuration of carbon-1, can be classified as c and t on the basis of a *cis* or *trans* configuration for the glycosidic group and the hydroxyl on the adjacent ring carbon. This procedure is applicable to all furanose derivatives and avoids the complications that arise in classification by conventional nomenclature.

Table 4 shows that structurally related tetroses, pentoses, hexoses, and higher sugars differ merely with respect to R, the substituent group attached to carbon 4 of the ring. In the hexofuranoses, which have 5 asymmetric carbons, the different configurations of carbon 5 give rise to two isomers for each of the c and t modifications of the four fundamental ring configurations. To emphasize configurational relationships the hexofuranoses in table 4 have been given systematic names based on the configuration of the groups comprising the ring structure, in addition to the conventional names. The systematic names are presented to aid in the correlation of related structures and are not proposed as alternate names for the substances. The ketofuranohexoses differ from the aldofuranopentoses in that the former have a CH<sub>2</sub>OH group in place of the hydrogen of carbon-1 of the latter. Like the aldofuranopentoses, the ketofuranohexoses can be classified as c or t accord-

TABLE 4. Furanose structures <sup>a</sup>

Code.....	..16	..26	..36	..46
Aldoses				
R=H.....	L-Threofuranose.....	L-Erythofuranose.....	( <sup>b</sup> )	( <sup>c</sup> )
R=CH <sub>2</sub> OH.....	D-Xylofuranose.....	D-Lyxofuranose.....	D-Arabinofuranose.....	D-Ribofuranose.
$\begin{array}{c} \text{H} \\   \\ \text{R}=\text{CH}_2\text{OH}-\text{C}- \\   \\ \text{OH} \end{array}$	D-Glucofuranose (D-glycero-D-xylo-aldohexofuranose).	D-Mannofuranose (D-glycero-D-lyxo-aldohexofuranose).	D-Altrofuranose (D-glycero-D-arabo-aldohexofuranose).	D-Allofuranose (D-glycero-D-ribo-aldohexofuranose).
$\begin{array}{c} \text{OH} \\   \\ \text{R}=\text{CH}_2\text{OH}-\text{C}- \\   \\ \text{H} \end{array}$	L-Idofuranose (L-glycero-D-xylo-aldohexofuranose).	L-Gulofuranose (L-glycero-D-lyxo-aldohexofuranose).	L-Galactofuranose (L-glycero-D-arabo-aldohexofuranose).	L-Talofuranose (L-glycero-D-ribo-aldohexofuranose).
Ketoses (CH <sub>2</sub> OH on carbon on right of formula)				
R=H.....	L-Xylulofuranose.....	L-Ribulofuranose.....	D-Xylulofuranose.....	D-Ribulofuranose.
R=CH <sub>2</sub> OH.....	D-Sorbofuranose.....	D-Tagatofuranose.....	D-Fructofuranose.....	D-Psicose.

<sup>a</sup> All of the compounds exist in enantiomorphic forms, one member of which is represented here.

<sup>b</sup> The product, D-threofuranose, is classified with L-threofuranose under code ..16.

<sup>c</sup> The product, D-erythofuranose, is classified with L-erythofuranose under code ..26.

ing to a *cis* or a *trans* relationship of the glycosidic group with respect to the hydroxyl of the adjacent ring carbon. There are few furanose derivatives, and hence for the present separate code numbers are not given to the *c* and *t* modifications. Classification of all furanose derivatives in only four groups makes comparisons possible with relatively few compounds.

#### 4.3. Conformation and Classification of Pyranose Structures

X-ray investigations and other evidence indicate that the atoms forming the pyranose ring do not lie in a single plane and that the rings can assume several shapes or conformations [7]. It was found in 1937 that if the pyranose sugars are divided into two groups, alpha when carbon 1 and carbon 5 have like configurations and beta when they have unlike configurations, then on bromine oxidation there is marked similarity in the behavior of the alpha sugars on the one hand and of the beta sugars on the other. This shows that there is an important structural difference in the alpha and beta modifications. The difference was ascribed to the existence of a strainless ring structure in which the alpha and beta positions are not symmetrically located with respect to the carbon-oxygen ring. Several strainless ring forms (or conformations) seemed possible [3]. But at the time this work was done there was no satisfactory method for deciding which structure was actually present. Subsequent work by Hassel

and Ottar [9], Reeves [10], and others [11] has shown that the pyranose ring invariably assumes one of two chair conformations, given in figure 1, unless the ring is altered by bridging to a boat form.

Reeves [10] has given these two forms the designations C1 and 1C respectively. Thus "the Sachse strainless ring chair form in which the sixth carbon atom and the ring oxygen project on the same side of the plane formed by carbon atoms 1,2,4, and 5 is

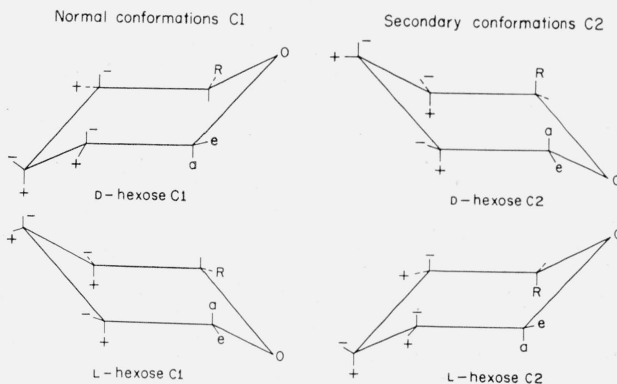


FIGURE 1. Conformations of the pyranose ring.

Positions marked + or - show, respectively, the location of the reference group for a D or an L configuration of the individual carbon. A reference group is defined as any group other than hydrogen. For the historical development of the designation of configuration, see the excellent review by C. S. Hudson [8].

designated C1, and its mirror image 1C." This definition is ambiguous, inasmuch as the mirror image of a *whole* molecule in the C1 conformation likewise meets Reeves' criterion for a C1 conformation. Unfortunately, "mirror image" was used in the definition in the sense of the ring skeleton only, but was then applied to the whole molecule. This led to a different classification for the D and L forms of derivatives of galactose and of arabinose. To avoid the term "mirror image" in its special sense, C1 is used in the present classification to represent the conformation of the pyranose ring in which carbon 6 is in an equatorial position, and C2 is used to represent the conformation in which carbon 6 is in an axial position. With this classification, the mirror image of D-galactopyranose C1 is L-galactopyranose C1 and the "inverted" conformation of the ring for either enantiomorph is designated by C2. The two conformations differ in that the groups in axial positions of one are in equatorial positions of the other, and the reverse.

Figure 2 shows the C1 structures for the aldohexopyranoses of the D-configurational series.<sup>4</sup> Table 5 lists the reference groups attached to the ring according to whether they are in axial or equatorial positions. On the left side of the table, the C1 conformations are listed with their code numbers. Opposite each of the C1 conformations is a C2 conformation with the same arrangement of reference groups, except for the one attached to carbon 5. The arrangement of the reference groups common to both substances is given in the center of the table. For example, either  $\alpha$ -D-glucopyranose C1 or  $\alpha$ -L-glucopyranose C1 has the a e e e conformation which is like the a e e e a conformation of either  $\beta$ -L-idopyranose C2 or  $\beta$ -D-idopyranose C2 except for the arrangement of the groups at carbon 5. In the classification system code numbers --1 and --2 are given to substances with the C1 conformation having axial and equatorial glycosidic groups, respectively. Code numbers --3 and --4 are given to substances with the C2 conformation having axial and equatorial glycosidic groups, respectively. Pyranose derivatives having a conformation other than the C1 or C2 are given the code number --5, regardless of configuration at the anomeric carbon.

Undoubtedly the conformation of the ring is not the same for all pyranose derivatives and in some cases several conformations may exist. This may account, as previously pointed out [3], for anomalous values for the differences in molecular rotations obtained by application of Hudson's rules of isorotation [12]. Some sugar derivatives exist that for stereomeric reasons are limited to certain conformations. For example, 1,4-anhydroglucopyranose requires a boat-shaped ring in which carbons 1 and 4 are directed toward each other. However, the consistent relationship found between the rates of ox-

<sup>4</sup> The L forms are not given because they are mirror images and can be constructed from the diagrams of figure 1.

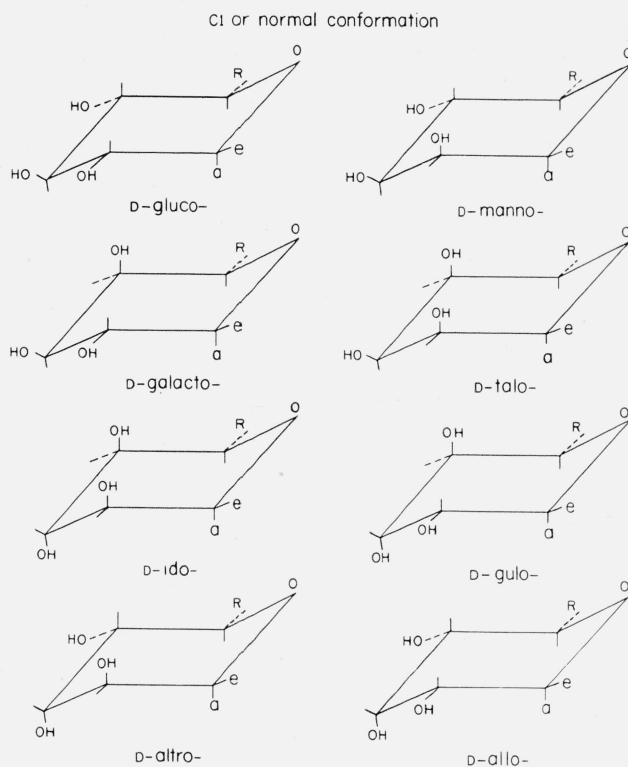


FIGURE 2. Fundamental pyranose types.

TABLE 5. Axial and equatorial positions of reference groups in aldohexopyranose structures

C1, or normal conformation (reference group at carbon 5 equatorial)		Character of reference group				C2, or secondary conformation (reference group at carbon 5 axial)	
Code	Configuration	Carbon-				Configuration	Code
		1	2	3	4		
--11	$\alpha$ -gluco-	a	e	e	e	$\beta$ -ido-	--53
--12	$\beta$ -gluco-	e	e	e	e	$\alpha$ -ido-	--54
--21	$\alpha$ -manno-	a	a	e	e	$\beta$ -gulo-	--63
--22	$\beta$ -manno-	e	a	e	e	$\alpha$ -gulo-	--64
--31	$\alpha$ -galacto-	a	e	e	a	$\beta$ -altro-	--73
--32	$\beta$ -galacto-	e	e	e	a	$\alpha$ -altro-	--74
--41	$\alpha$ -talo-	a	a	e	a	$\beta$ -allo-	--83
--42	$\beta$ -talo-	e	a	e	a	$\alpha$ -allo-	--84
--51	$\alpha$ -ido-	a	a	a	a	$\beta$ -gluco-	--13
--52	$\beta$ -ido-	e	a	a	a	$\alpha$ -gluco-	--14
--61	$\alpha$ -gulo-	a	e	a	a	$\beta$ -manno-	--23
--62	$\beta$ -gulo-	e	e	a	a	$\alpha$ -manno-	--24
--71	$\alpha$ -altro-	a	a	a	e	$\beta$ -galacto-	--33
--72	$\beta$ -altro-	e	a	a	e	$\alpha$ -galacto-	--34
--81	$\alpha$ -allo-	a	e	a	e	$\beta$ -talo-	--43
--82	$\beta$ -allo-	e	e	a	e	$\alpha$ -talo-	--44

dation with bromine and the configuration of the anomeric carbon in the aldohexopyranose series is evidence of the prevalence of a common ring conformation in at least 10 (out of a possible 16) aldohexopyranoses studied by bromine oxidation [3, 4]. The relative stability of the two conformations depends on a number of factors, particularly the tendency of large groups to take an equatorial rather than an axial position [9, 10]. The  $\text{CH}_2\text{OH}$  group of the aldohexopyranoses has a strong tendency to assume the equatorial position, and this accounts for the fact that most aldohexopyranoses have the C1 conformation.

The pyranose forms of the pentoses and ketohexoses differ from those of the aldohexoses in that at carbon 5 they do not have a  $\text{CH}_2\text{OH}$  group and hence the ring-forming carbon is nonasymmetric. Thus each pentose or ketohexose is structurally related to two aldohexoses. The relationship for  $\alpha$ -D-xylose, a typical pentose, is shown in figure 3. The structure for  $\alpha$ -D-xylopyranose could be derived by replacing the  $\text{CH}_2\text{OH}$  group of either  $\alpha$ -D-glucopyranose or  $\beta$ -L-idopyranose by hydrogen. The conformation of  $\alpha$ -D-xylopyranose produced by this hypothetical procedure would depend on the conformation of the parent hexose.  $\alpha$ -D-Glucopyranose C1 and  $\beta$ -L-idopyranose C2 would yield a form of  $\alpha$ -D-xylopyranose having equatorial hydroxyls at carbons 2, 3, and 4, and an axial hydroxyl at carbon 1. But  $\alpha$ -D-glucopyranose C2 and  $\beta$ -L-idopyranose C1 would yield a form of  $\alpha$ -D-xylopyranose having axial hydroxyls at carbons 2, 3, and 4, and an equatorial hydroxyl at carbon 1.

Apparently the pentoses tend to assume the conformation with the hydroxyl of carbon 3 in an equatorial rather than an axial position. In this connection Isbell and Frush wrote as follows in 1940 [5]: "As may be noted from a space model, carbon 3 lies opposite the oxygen of the ring and its

attached groups appear to be in a particularly favorable position to influence the conformation of the ring. Perhaps this accounts for the unexpectedly large influence of the configuration of carbon 3 on the relative reactivities of the alpha and beta modifications. Another observation which may depend on the effect of the configuration of carbon 3, is that in the pentose series where the ring forming carbon is not asymmetric, each pentose resembles the hexose in which carbons 3 and 5 have opposite configurations." Thus xylose, lyxose, arabinose, and ribose derivatives resemble the corresponding derivatives of glucose, mannose, galactose, and talose, respectively, more closely than they resemble the derivatives of idose, gulose, altrose, and allose [3, 4, 5]. Hence for classification purposes the aldopentoses are given code numbers corresponding to those of the first-named group of substances. For example, the code number of  $\alpha$ -D-xylopyranose (a e e e) is .1110, whereas, the code number for the hypothetical "inverted" or C2 conformation (e a a a) is .1140. The pyranose derivatives of the related ketohexoses, sorbose, tagatose, fructose, and psicose are likewise given code numbers corresponding to the glucose, mannose, galactose, and talose configurations, respectively.

The ketoses differ from the aldoses in that the glycosidic carbon has a  $\text{CH}_2\text{OH}$  group in place of the aldehydic hydrogen. The presence of this group greatly alters the equilibrium for the anomeric modifications in solution. Apparently the  $\text{CH}_2\text{OH}$  group tends to assume an equatorial position analogous to that taken by the  $\text{CH}_2\text{OH}$  group in the C1 aldohexopyranoses. This accounts for the fact that equilibrium solutions of fructose, sorbose, tagatose, mannoheptulose, glucoheptulose, and other ketoses contain almost exclusively the pyranose modification having an equatorial  $\text{CH}_2\text{OH}$  group at carbon 2.

At present any attempt to assign definite conformations to all pyranose derivatives is premature. There is need, however, for consideration of the special stereomeric factors associated with the ring conformation most likely to be present in any substance. This is particularly true for the evaluation of infrared absorption, because this property is extremely sensitive to the molecular structure. Hence some method of classification is desirable. A definitive classification requires that the structures of the compounds shall have been established unequivocally. As this is not true with pyranose derivatives in general, it is necessary to make an arbitrary classification. By comparison of the infrared spectra of substances classified in the arbitrary manner, exceptional properties may be discovered and ultimately explained in terms of conformation, or of molecular structures at present unknown. As a working hypothesis, in the absence of unequivocal evidence to the contrary, the C1 conformation is assumed as a first choice and the C2 only as a second choice. With provisional assignments, infrared absorption and other data can be examined for correlations based on either the C1 or the C2 conformation.

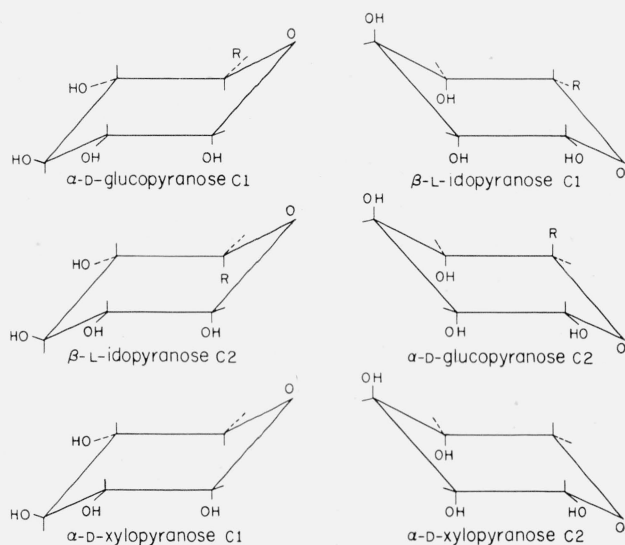


FIGURE 3. Structurally related hexoses and pentoses.

## 5. References

- [1] ACS Rules on Carbohydrate Nomenclature, Chem. Eng. News **31**, 1776 (1953).
- [2] D. H. R. Barton, O. Hassel, K. S. Pitzer, and V. Prelog, Nature **172**, 1096 (1953).
- [3] H. S. Isbell, J. Research NBS **18**, 505 (1937) RP990.
- [4] H. S. Isbell and W. W. Pigman, J. Research NBS **18**, 141 (1937) RP969.
- [5] H. S. Isbell and H. L. Frush, J. Research NBS **24**, 125 (1940) RP1274.
- [6] C. A. Beevers and W. Cochran, Nature **157**, 872 (1946).
- [7] O. L. Sponsler and W. H. Dore, Ann. Rev. Biochem. **5**, 63 (1936).
- [8] C. S. Hudson, Advances in Carbohydrate Chem. **3**, 1 (1948).
- [9] O. Hassel and B. Ottar, Acta Chem. Scand. **1**, 929 (1947).
- [10] R. E. Reeves, Advances in Carbohydrate Chem. **6**, 107 (1951).
- [11] J. A. Mills, Advances in Carbohydrate Chem. **10**, 1 (1955).
- [12] C. S. Hudson, J. Am. Chem. Soc. **52**, 1680 (1930).

WASHINGTON, May 29, 1956.