

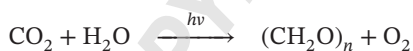
1

Structure of Saccharides

1.1 Structure of Monosaccharides

Carbohydrates comprise one of the largest and most fascinating groups of natural products. They are the most abundant macromolecules on Earth and represent the most structurally diverse class of biomolecules. We encounter them everywhere in nature: as monosaccharides, oligo- and polysaccharides and glycoconjugates, in which carbohydrates are glycosidically linked to other natural products such as proteins (glycoproteins), lipids (glycolipids) and many other biologically active molecules.

Every year, billions of tons of carbohydrates are produced by photosynthesis in plants and cyanobacteria. Photosynthesis consists of light-dependent reactions, which generate ATP (adenosine triphosphate) and reducing equivalents in the form of NADPH (nicotinamide adenine dinucleotide phosphate), and light-independent reactions, in which CO₂ is fixed into carbohydrates via the Calvin cycle. In this process, the chemical energy of ATP and the reducing power of NADPH from the light-dependent reactions are used to reduce CO₂ to sugars, while water is oxidised to oxygen.



When the general formula of the photosynthetic products, (CH₂O)_n, is rearranged as C_n(H₂O)_n, it illustrates why these compounds were historically called 'hydrates of carbon'. This led to the modern term 'carbohydrate' and in German to the word 'Kohlenhydrat'. For example, the well-known aldopentoses, such as ribose, have the molecular formula C₅(H₂O)₅, and the well-known aldohexoses, such as glucose and mannose, have the molecular formula C₆(H₂O)₆. Today, the term 'carbohydrate' is used synonymously with the expression 'saccharide', 'glycan', or even just 'sugar'.

Carbohydrates perform a wide variety of functions in nature. They serve as energy source and storage molecules (e.g. starch and glycogen), structural material (e.g. cellulose and chitin) and information carriers (complex oligosaccharides and glycoconjugates). Thus, the biological functions of carbohydrates span a broad spectrum, ranging from energy metabolism and structural stability to the regulation of cellular processes. Consequently, progress in glycoscience requires an

understanding of carbohydrate structures, their (bio)synthesis and their biological functions (glycobiology).

Chemically, carbohydrates are polyhydroxycarbonyl compounds, where polyhydroxyaldehydes are called ‘aldoses’ and polyhydroxyketones ‘ketoses’. Aldoses are derived from glyceraldehyde by a formal insertion of a stereogenic (HCOH)-group between the carbonyl carbon atom and the α -C-atom of glyceraldehyde. This leads to the aldotetroses, and further successive (HCOH)-group insertions between the carbonyl group and the adjacent stereogenic centre lead to aldopentoses and aldohexoses. Thus, aldoses can be of different chain length. Ketoses, by analogy, derive from 1,3-dihydroxyacetone, leading to tetruloses, pentuloses and hexuloses (Figure 1.1).

Each aldose and each ketose exist as two enantiomers. To distinguish between the two, the prefixes ‘D’ and ‘L’, respectively, are used. D and L are relative designations for the configuration of the highest-numbered stereogenic centre of a monosaccharide, in other words the one most distant from the carbonyl group, which is numbered 1 in aldoses and 2 in ketoses. This asymmetrically substituted carbon atom is called the ‘configurational reference atom’. The relative nomenclature is based on the configuration of the two enantiomeric glyceraldehyde molecules in the case of the aldose family and on the stereochemistry of erythrulose for ketoses. When the hydroxyl group at the stereogenic centre of glyceraldehyde (and erythrulose, respectively) is positioned left in the Fischer projection, the molecule is named the L- (latin: *laevus* = left) enantiomer. When this hydroxyl is positioned right, it

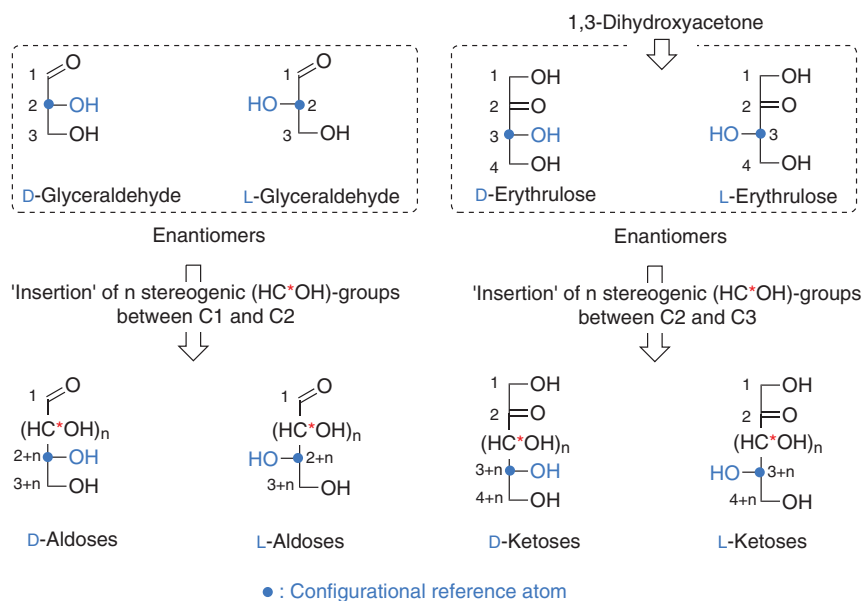


Figure 1.1 Aldoses and ketoses can be of different chain length. The group of aldoses is derived from glyceraldehyde by the formal insertion of stereogenic (HCOH)-groups. Ketose can be built up from 1,3-dihydroxyacetone in the same way. The structures are shown as Fischer projections. The highest-numbered asymmetrically substituted carbon atom is called the ‘configurational reference atom’.

is called the D- (latin: *dexter* = right) form. Certainly, D- and L-sugars could also be described using the (R)- and (S)-designations according to the Cahn–Ingold–Prelog (CIP) rules. The historical D- and L-conventions, however, are advantageous in carbohydrate nomenclature and have, therefore, been maintained. In addition to the D- and L-designations, carbohydrates are often specified by the sign of their optical rotation, which certainly bears no logical relationship to the D- and L-conventions. If the sign of the optical rotation is indicated, a (+)- or (–)-sign is added before the configurational prefix. Racemic forms are indicated by (±).

As explained earlier, a structural tree of aldoses and ketoses is created from glyceraldehyde and dihydroxyacetone, respectively, of which every member is characterised by a distinct sequence of stereogenic centres, each of which was attributed an individual name. These names form the fundament of carbohydrate nomenclature (Figures 1.2 and 1.3).

The structural sugar tree of aldoses and ketoses comprises an enormous stereochemical variety of structures, providing a fruitful area for the study of stereochemistry. D-Glucose, for example, the most abundant monosaccharide found in nature, contains six carbon atoms and four stereogenic centres in the open-chain form. Consequently, it is just one stereoisomer from a group of 16 stereoisomeric aldohexoses, all of which have the molecular formula $C_6H_{12}O_6$. Most of the widely distributed monosaccharides belong to this group of hexoses. In addition, there

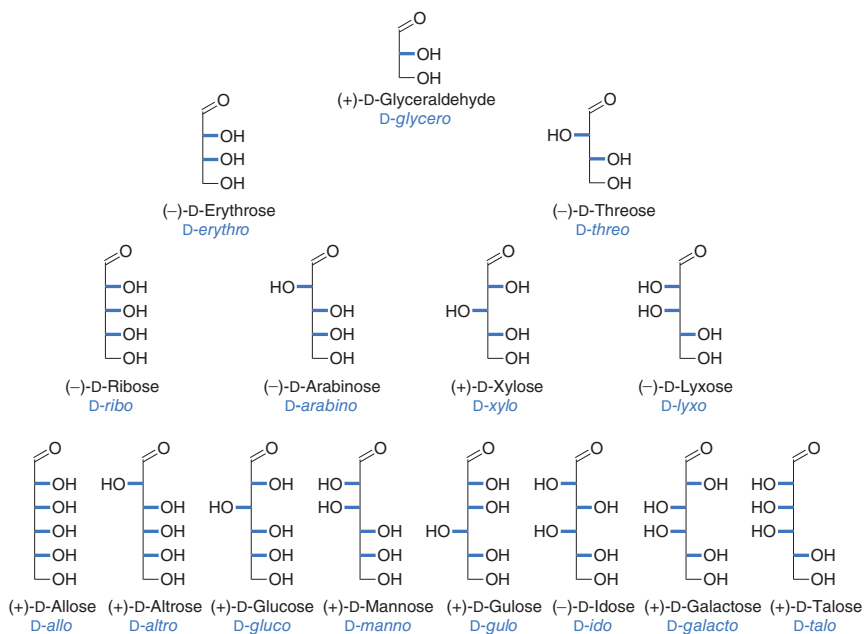


Figure 1.2 The structural tree of the acyclic forms of D-aldoses with their names and signs of optical activity, drawn as their Fischer projections. In the D-series, two aldotetroses, four aldopentoses, and eight aldohexoses exist. The structural elements highlighted in blue represent the sequence of stereocentres, corresponding to the respective configurational prefix used for the systematic naming of carbohydrates.

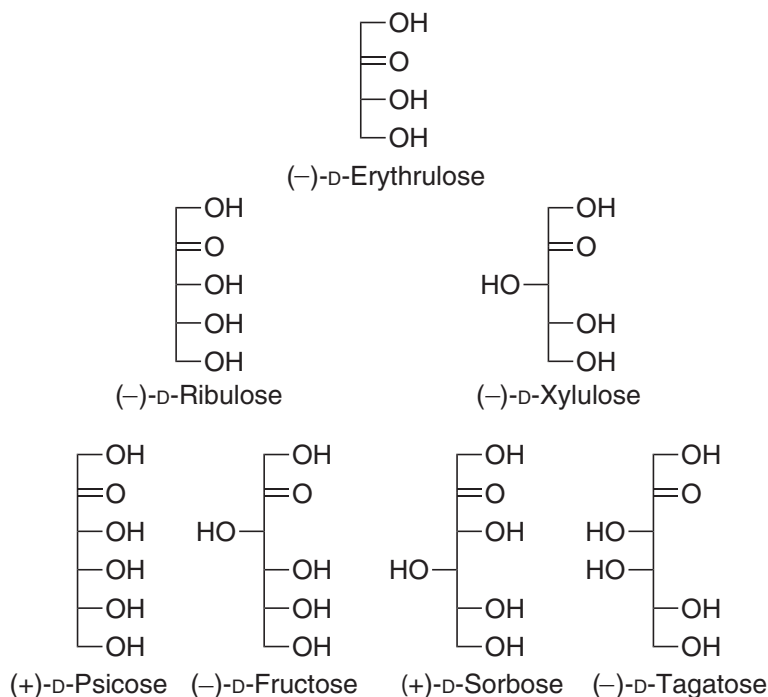


Figure 1.3 The structural tree of the acyclic forms of D-ketoses with their names and signs of optical activity, drawn as their Fischer projections.

exist eight stereoisomeric hexuloses (ketoses with six carbon atoms) which are constitutional isomers to the aldohexoses. The 16 stereoisomeric aldohexoses consist of eight enantiomeric pairs, one enantiomer being the L-form, its mirror image comprising the D-enantiomer in each case. Aldohexoses, which are not enantiomeric to each other, are called diastereomeric. In carbohydrate chemistry, the term 'epimeric' is also frequently used. Epimers are diastereomeric compounds, which differ only in their stereochemistry at one stereogenic carbon atom; hence, such isomeric sugars can be named 2-, 3-, 4-, or 5-epimers, respectively.

Whereas some aldoses and ketoses of the sugar trees occur abundant in nature, others are only found in limited quantities and are called rare sugars.¹ Examples of rare sugars are psicose, sorbose and tagatose, as well as altrose, allose, idose, gulose and talose.

Today, the stereochemical structures of all monosaccharides are known, but from a historical perspective one can appreciate how difficult the path to get there was. First, the concept of stereoisomers had to be realised. In 1874, both Jacobus van't Hoff and Joseph Le Bel suggested that the presence of asymmetric carbon atoms in organic compounds had to be the cause for their optical activity. Furthermore, they stated that in case of the aldohexoses, where four of the carbon atoms are asymmetric, there should exist $2^4 = 16$ stereoisomers. Which is indeed the case, as we have seen before. But which structure belongs to which carbohydrate? It

was Emil Fischer² who discovered the interrelationships of the large family of monosaccharides. He elucidated the structures of glucose and its isomers at a time when NMR spectroscopy and modern separation techniques were not yet available. His work, carried out in the late nineteenth and early twentieth century, was then, and is today, recognised as one of the outstanding achievements of early structural work.³

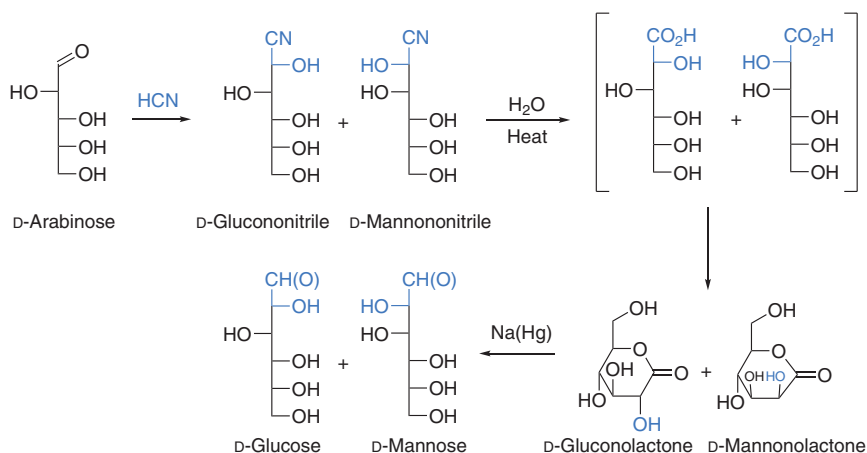
Fischer's Elucidation of the Basic Stereochemical Interrelationships of Monosaccharides

As there was no way of determining absolute configurations at that time, Fischer derived the configurations of all aldohexoses relative to one enantiomer of glyceraldehyde. He arbitrarily chose the D-configuration for (+)-glyceraldehyde, one of two possibilities, and luckily selected correctly. This was established by Johannes Bijvoet and co-workers in 1951 by X-ray diffraction analysis. Consequently, the molecular structures of monosaccharides used, as represented since Fischer's time, denote not only the correct relative configurations but also the correct absolute configurations.

The formal (HCOH)-group insertions, which have led us to the family of aldoses starting with glyceraldehyde, were practically carried out by Fischer applying a method, which became known as the Kiliani–Fischer cyanohydrin synthesis. Hydrocyanic acid reacts with aldehydes and ketones under acid or base catalysis to give an isomeric pair of α -hydroxynitriles, called cyanohydrins, hydrolysis of which produces α -hydroxy acids. These quickly cyclise in an intramolecular reaction into two diastereomeric lactones, which can be separated. Then, their reduction with sodium amalgam under mildly acidic conditions delivers the corresponding aldoses, each having one additional hydroxymethylene group between the aldehydic carbon and its adjacent carbon atom compared to the starting material. Thus, four tetroses, eight pentoses and 16 aldohexoses are obtained from the simplest aldose glyceraldehyde (Scheme 1.1).

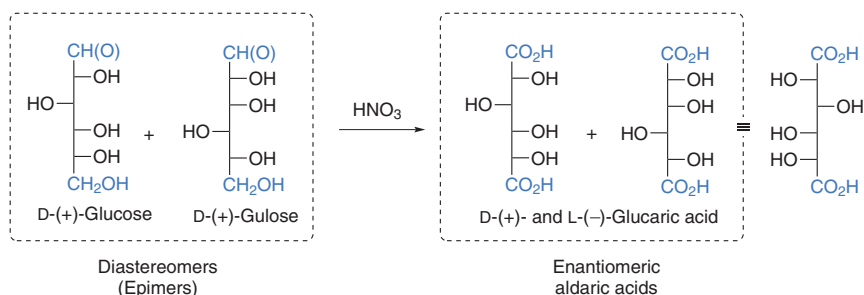
Since each step of the Kiliani–Fischer synthesis gives rise to two products, a chemical method had to be developed to distinguish them. This was done by converting each aldose into a derivative with identical terminal groups and determining whether the products were optically inactive meso forms or optically active, thus asymmetric. Therefore, aldoses were treated with strong nitric acid to give the corresponding α,ω -dicarboxylic acids, called aldaric acids. In this reaction, erythrose, ribose, xylose, allose and galactose give the internally compensated meso dicarboxylic acids (e.g. tartaric acid from erythrose), whereas from the other aldoses optically active, asymmetric acids are derived. This technique, taken together with the structures of the sugars from which they were derived, allows the five named sugars and their epimers to be structurally characterised.

Only glucose, mannose, gulose and idose cannot be assigned structures on this basis. However, Fischer noted that D-glucose and D-gulose give enantiomeric derivatives after appropriate oxidation. There was only one pair of acids having this relationship, which could be derived from these four hexoses. It was also known that D-glucose is derived from D-arabinose, and consequently glutaric acid must be the



Scheme 1.1 The Kiliani–Fischer synthesis is the practical method to insert a (HCOH)-group between C1 and C2 of an aldose. By this way, the aldopentose D-arabinose, for example, is converted into two diastereomeric aldohexoses, here D-glucose and D-mannose.

oxidation product of D-glucose. Based on this conclusion, the structure of D-glucose could be assigned. D-Mannose must be its epimer. The structures of D-gulose and D-idose were similarly deduced. Following this reasoning, Fischer solved the inter-relationships between the monosaccharides like a stereochemical crossword puzzle based on optical activity and rotation values (Scheme 1.2).



Scheme 1.2 The oxidation of the epimeric aldoses D-glucose and D-gulose (both of which show dextrorotatory optical activity) leads to enantiomeric aldaric acids, which exhibit optical rotations of equal magnitude but opposite sign.

Excursus: Measuring Optical Rotation with the Polarimeter

Polarimetry was of fundamental importance for the analysis of sugars in the early days of carbohydrate chemistry, and rotation values are still reported today in the literature for new carbohydrate derivatives. Polarimetry is useful to establish the non-racemic nature of a sample of a chiral compound, and it still finds a use today in the estimation of the anomeric purity of glycosides for example, because α - and β -glycosides normally show very different rotation values.

Thus, we are inserting a little methodological detour at this point to explain how rotation values are measured and reported.

Sugars, like all asymmetric compounds, cause rotation in the plane of polarisation of linearly polarised light because the two circularly polarised components travel at different speeds through the sugar solution. The optical activity of an asymmetric sample in solution is observed in a polarimeter and measured as a rotation value, α . It is reported as specific rotation $[\alpha]$ obtained at a certain temperature and wavelength λ . The observed α and the specific rotation $[\alpha]$ are related in the manner shown in the following equation:

$$[\alpha]_{\lambda}^{\circ C} = \frac{\alpha_{\lambda}^{\circ C}}{l \cdot c}$$

where l is the length of the cell in decimetres (cells are often 1 dm in length) and c is the concentration of the compound given in grams per mL.

Consequently, the unit for specific rotation values is $\text{deg} \cdot \text{mL} \cdot \text{dm}^{-1} \cdot \text{g}^{-1}$, not degrees.

The observed rotation value α depends on the concentration of the sample, the temperature and wavelength used, with the latter usually being the wavelength of the sodium-D line (a spectral doublet at 589.0 and 589.6 nm). Specific rotation values are, therefore, always reported together with these specifications. In practice, specific optical rotations are reported as dimensionless numbers, with the units of the specific rotation being implicit and omitted from the reported value. The sign of '+' or '-' has to be added before the value of optical rotation, for example:

$$[\alpha]_{\text{D}}^{20} = +160.5 \quad (c = 1.1 \text{ in MeOH})$$

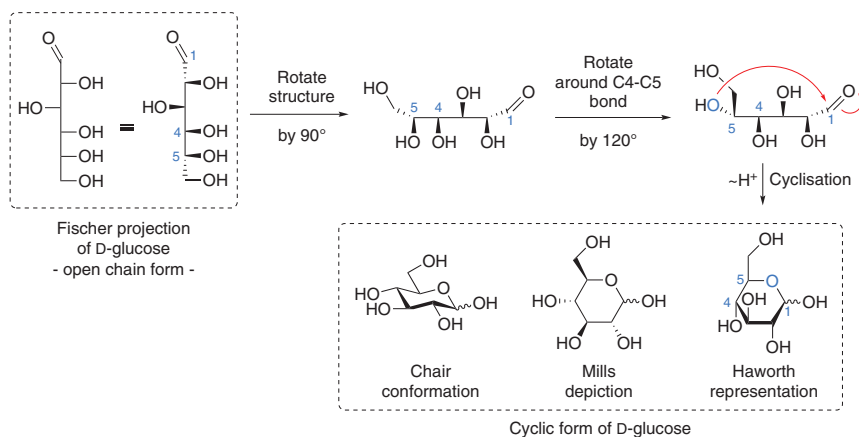
The specific rotation value has the same sign as the observed α of a particular compound. Optically active compounds, which turn the plane of linear polarised light to the right, are assigned a positive sign for the observed rotation value. The enantiomer of this compound gives the same rotation value in the polarimeter but with the opposite sign.

Rotation values of compounds of different molecular mass can be compared on the basis of their molar rotation values $[M]$, which are calculated using the relation $[M] = M \cdot [\alpha]$. Here, M is the molecular mass of the compound and $[\alpha]$ the specific rotation as defined earlier.

The Cyclic Structure of Monosaccharides

Carbohydrates in fact do not exist in the open-chain form, but mainly as cyclic hemiacetals and hemiketals, respectively. These are formed by the intramolecular nucleophilic addition of a hydroxyl group to the carbonyl function of a monosaccharide. Whereas acyclic hemiacetals and hemiketals are labile and decompose to deliver the starting materials, the cyclic forms are stable because the intramolecular hemiacetal formation is entropically favoured compared to the respective intermolecular reaction.

It is a bit tricky to convert the stereochemical information which is nicely depicted in the structural tree of Fischer projections into cyclic structures. One way of arriving at the right cyclic structure is to first consider that according to convention, in a Fischer projection, horizontally projected covalent bonds are pointing towards the viewer, whereas vertically projected bonds are pointing backwards. Next, the structure can be rotated by 90° , followed by rotation around the bond between C4 and C5 by 120° , as shown for D-glucose in Scheme 1.3. After the structure has been reoriented such, the intramolecular cyclisation can be performed, leading to the so-called Haworth representation of the cyclic hemiacetal where the stereochemical information is exactly the same as in the Fischer projection of the open-chain form. In other words, the configuration at C2–C5 remains the same, as it should be. Whether the stereoinformation from the acyclic Fischer to the cyclic Haworth form has been transmitted correctly can be checked by determining the configuration at each stereocentre according to the CIP rules.



Scheme 1.3 The stereoinformation contained in the Fischer projection of open-chain sugars, illustrated here for D-glucose, must be correctly translated into the structures of the cyclic hemiacetals, which can be depicted as a Haworth representation, a Mills depiction, or, best of all, the chair conformation. The wavy line at the anomeric centre indicates a mixture of anomers.

Another practical hint for the transformation of a structure given in the Fischer projection into a Haworth or chair representation is following a rule according to which a hydroxyl group located left in the Fischer projection translates into an upward position in the Haworth (chair)–representation and a Fischer–right–hydroxyl translates into a Haworth (chair)–down–OH (Scheme 1.3).

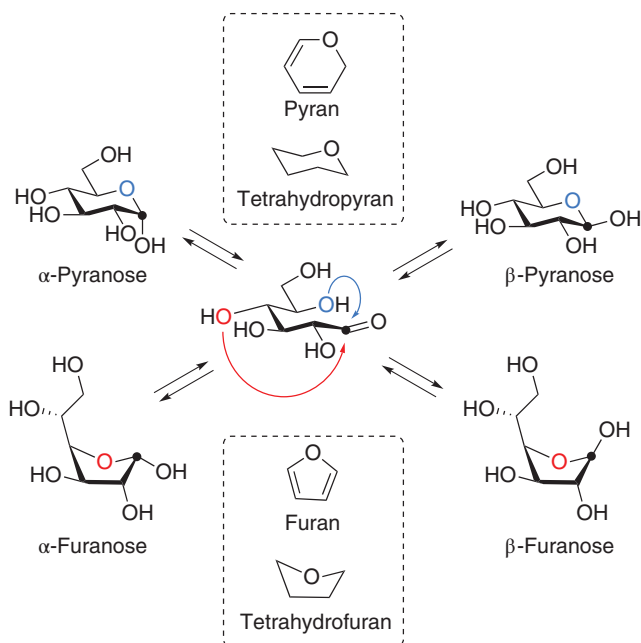
Cyclic sugars can be represented by several different semiotic descriptions. In addition to the Haworth representation, the so-called Mills depiction is used. It is the favoured way of representing carbohydrate structures in other areas of natural product chemistry, when monosaccharides are part of much larger structures which contain additional non-carbohydrate rings. It is also a favourable representation when synthetic pathways from carbohydrates to chiral non-carbohydrate building blocks have to be described.

In the Haworth projection, the carbohydrate ring is assumed to be planar, and the hydroxyl groups are arranged above and below the plane of the ring. Consequently,

the Haworth projection conveys no information about the actual three-dimensional conformation of carbohydrates. The six-membered ring of monosaccharides is certainly not planar, but adopts a chair conformation and thus, the cyclic forms of carbohydrates are most accurately drawn as chairs. This representation also facilitates the distinction between axial and equatorial ring positions, a difference which is of some importance in carbohydrate chemistry. Therefore, in carbohydrate chemistry, the chair conformation is the most accurate and most meaningful representation of the sugar ring. Accordingly, conformational representations of sugars are also used throughout this text.

Two aspects of the cyclisation of the open-chain form of monosaccharides have not yet been addressed. First, not only six-membered but also five-membered rings can be formed depending on which hydroxyl group of the monosaccharide chain reacts with the carbonyl group. The cyclic forms of monosaccharides are then called furanoses and pyranoses, respectively, in analogy to their unfunctionalised heterocycle analogues, tetrahydrofuran and tetrahydropyran. Second, from the nucleophilic attack of a hydroxyl group at the prostereogenic carbonyl carbon, two different configurations at C1 may result in the newly formed sugar ring. The two possible stereoisomeric cyclic hemiacetals are referred to as anomers and designated as α or β . The C1 carbon in cyclic monosaccharides is called the anomeric centre.

Hence, depending on which configuration is created upon cyclisation, the hemiacetals are named α - or β -furanoses and α - and β -pyranoses, respectively. They all are in equilibrium with each other through the open-chain form (Scheme 1.4).



Scheme 1.4 Mutarotation of *D*-glucose in solution produces a mixture of α - and β -pyranoses and α - and β -furanoses. The formation of five- and six-membered monosaccharide rings (furanoses and pyranoses) is exemplified by the aldose *D*-glucose, with different anomeric configurations.

The fact that all forms of a specific monosaccharide, α - and β -furanoses and α - and β -pyranoses, are in equilibrium with each other through the open-chain form is the cause of a phenomenon known as mutarotation. In solid form, each sugar is a discrete stereoisomer. When the crystals are dissolved in water, the hemiacetal ring opens and the equilibrium of pyranoses and furanoses forms. This equilibration occurs with all reducing saccharides and is accompanied by a change in optical rotation called mutarotation. This equilibration can be base- or acid-catalysed.

Thus, the observed rotation value of a freshly prepared sugar solution changes over time until mutarotation is complete and an equilibrium value is reached. If either pure α -D-glucopyranose with a specific rotation value $[\alpha]$ of +112 or pure β -D-glucopyranose with $[\alpha] = +19$ is dissolved in water, the rotation value keeps changing until an equilibrium value of +52.7 is reached after about 3 hours, corresponding to an anomeric mixture of glucopyranoses, 37% α - and 63% β -pyranose. In the various equilibria of free sugars, the proportion of pyranose and furanose forms varies, depending on their structure (Table 1.1).

To avoid any confusion, all structures of cyclic α -D-aldoses (pyranoses and furanoses) are summarised in Figure 1.4.

The Anomeric Configuration

The new stereogenic centre generated by cyclisation of open-chain monosaccharides is called the anomeric centre. In aldoses, the anomeric centre is C1, and in ketoses,

Table 1.1 The percentage compositions of free sugars in aqueous solution at equilibrium.⁴

Sugar	α -Pyranose	β -Pyranose	α -Furanose	β -Furanose	Temperature (°C)
Allose	14.0	77.5	3.5	5.0	31
Altrose	27.0	43.0	17.0	13.0	22
Arabinose	60.0	35.5	2.5	2.0	31
Fructose	2.0	70.0	5.0	23.0	30
Fucose	28.0	67.0		5.0	31
Galactose	30.0	64.0	2.5	0.2	31
Glucose	37.3	62.6		3.0	22
Gulose	16.0	81.0		3.0	22
Idose	38.5	36.0	11.5	14.0	31
Lyxose	70.0	28.0	1.5	0.5	31
Mannose	63.7	35.5	0.6	0.2	36
Psicose	22.0	24.0	39.0	15.0	27
Ribose	21.5	58.5	6.5	13.5	31
Sorbose	93.0	2.0	4.0	1.0	31
Tagatose	79.0	16.0	1.0	4.0	27
Talose	42.0	29.0	16.0	13.0	22
Xylose	36.5	63.0		< 1.0	31

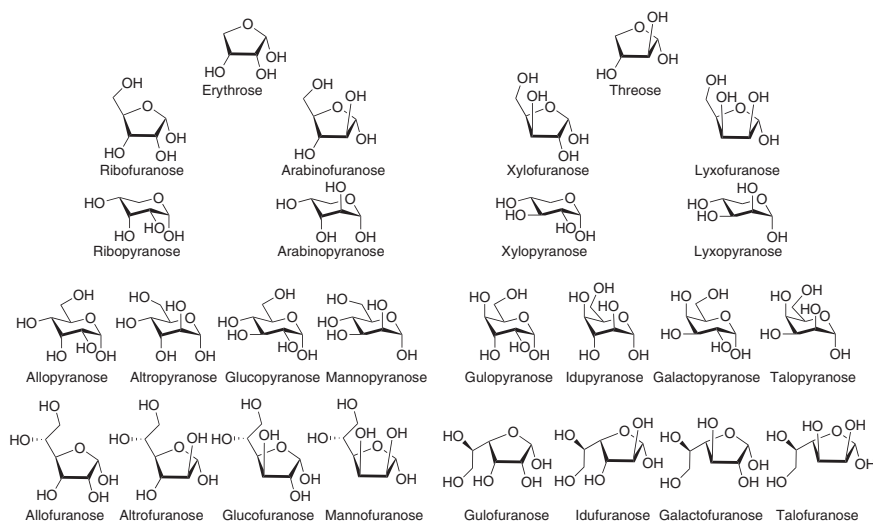


Figure 1.4 Cyclic forms of the α -D-aldoses.

it is C2. The two possible stereoisomeric cyclic hemiacetals in cyclic hemiacetals are referred to as anomers. Anomers differ only in the configuration at the anomeric centre; thus, they are diastereomers.

The relative stereodescriptors α or β are used in carbohydrate nomenclature to describe the configuration at the anomeric centre by relating it to the anomeric reference atom, which is the most distant stereogenic centre. The descriptor α is used when the OH group at the anomeric centre projects in the same direction as the OH group at the anomeric reference atom in the Fischer projection, and β is used when these two hydroxyl groups are pointing in opposite directions.

Thus, for D-glucose and all aldohexoses of the D-series, α -anomers have the hydroxyl group at the anomeric centre projecting downwards in Haworth or chair formulae; α -L-compounds have this group projecting upwards. The β -anomers have the opposite configurations at the anomeric centres, i.e., the hydroxyl group projects upwards for β -D-sugars and downwards for β -L-compounds. A wavy line is used for the anomeric bond when the anomeric configuration is not specified (Figure 1.5).

In simple cases, the anomeric reference atom is the same as the configurational reference atom, as we have seen earlier for α - and β -glucose. However, sometimes, assigning the anomeric configuration can be tricky. This might well be exemplified by the pentopyranose arabinose and the similarly configured 6-deoxy sugar fucose. In case of L-fucose, the anomer with the anomeric hydroxyl pointing upwards is named the α -anomer, whereas D-arabinopyranose carrying an axial anomeric hydroxyl group is the β -anomer. The same applies to the furanose form of arabinose (Figure 1.6).

The Anomeric Effect

The equatorially positioned substituents of a carbohydrate ring are, for steric reasons, the most energetically favoured, compared to their axial counterparts, as is

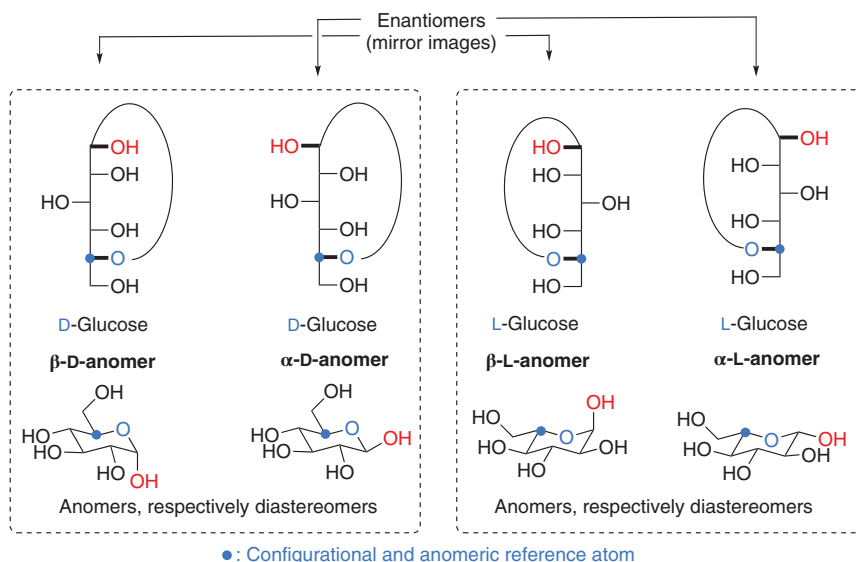


Figure 1.5 Structures of α - and β -anomers of *D*-glucopyranose (left) and *L*-glucopyranose (right), shown as Fischer projections (top) and chair conformations (bottom). In these examples, the configurational and reference atoms are identical. The anomeric configuration is defined by whether the OH group at the anomeric centre (red) projects in the same direction as the OH group at the anomeric reference atom in the Fischer projection (blue) (α -anomers) or in opposite direction (β -anomers).

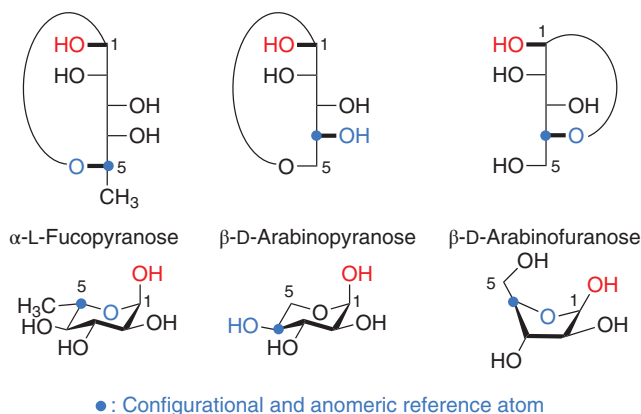
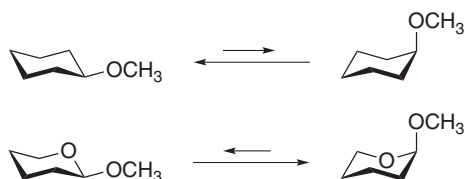


Figure 1.6 *L*-Fucopyranose and *D*-arabinopyranose appear very similar; however, fucose is an aldohexose whereas arabinose is an aldopentose and hence, their anomeric reference atoms are different (C5 in fucose and C4 in arabinose). As an example of a furanose structure, β -*D*-arabinofuranose is also shown.

the case in every molecule with a chair conformation. However, the anomerically bound groups in carbohydrates do not follow this rule completely. For *D*-pyranoses, *D*-pyranosides and especially carbohydrate derivatives with electronegative groups at the anomeric centre, the anomeric α -configured derivatives with the anomeric group located in an axial position are often more stable than would be predicted from the steric interactions they have with adjacent substituents. An aqueous solution of *D*-glucose, for example, contains the α - and the β -form in a ratio of 38:62, and the

effect is even more pronounced for D-mannose, where the $\alpha:\beta$ ratio is 65:34. The unusual preference of the sterically unfavoured axial position over the equatorial position at the anomeric centre has been termed the ‘anomeric effect’ by Raymond Lemieux, who first recognised it in carbohydrates.

The anomeric effect has since been shown to be of general importance in molecules where two heteroatoms are attached to a tetrahedral centre. It can also be generalised to any cyclohexyl or even acyclic system containing a structural motif C–Y–C–X, where Y is a heteroatom bearing one or more lone pairs (N, O, S) and X is an electronegative atom or group, such as a halogen or an aglycone group (–OR) (Scheme 1.5).



Scheme 1.5 The conformational equilibrium between the two complementary chair forms changes markedly when the cyclohexane ring is replaced by a pyranose ring containing an endocyclic oxygen atom.

The anomeric effect can be explained in several ways. It partly involves a dipole–dipole effect based on intramolecular electrostatic interactions of two dipoles next to the anomeric centre. One of the two dipoles arises from the two lone electron pairs of the endocyclic carbohydrate ring oxygen. The other dipole points along the polarised bond between the anomeric carbon atom and its bound atom X. Anomeric configurations, where the two dipoles partially neutralise each other, are favoured over the diastereomers, where the anomeric configuration leads to partial intramolecular addition of the two dipoles. The anomeric effect virtually ensures the axial configuration of an electronegative substituent at the anomeric centre such as in the case of acetobromoglucose, where the β -anomer cannot be isolated (Figure 1.7).

Unfavourable lone pair–lone pair interactions have also been used to explain the anomeric effect. Most importantly, however, the anomeric effect is a stereoelectronic effect, which can be explained by molecular orbital theory. A lone pair of electrons located in an n-molecular orbital of the atom Y overlaps with the antibonding

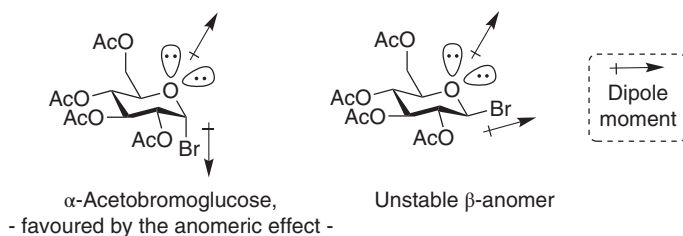


Figure 1.7 The anomeric effect explains why the β -form of acetobromoglucose is unstable and only the α -form can be isolated.

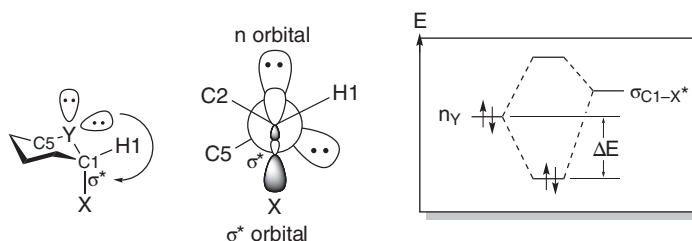


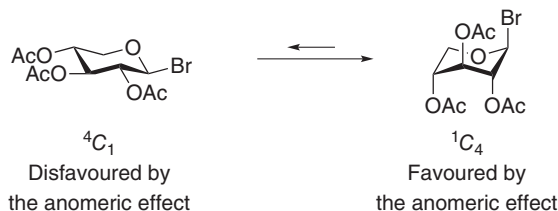
Figure 1.8 Favourable anomer due to $n_Y \rightarrow \sigma^*$ delocalisation of nonbonding electrons, which is possible with an anti-periplanar arrangement of the involved orbitals. A significant effect is obtained when the σ_{CX}^* orbital is of low energy. The lone pair electrons on the endocyclic oxygen are shown as in sp^3 orbitals.

σ^* -orbital of the C—X bond. This favourable $n_Y \rightarrow \sigma^*$ delocalisation of nonbonding electrons (negative hyperconjugation) is only possible when the C—X bond is axial with an anti-periplanar arrangement of the involved orbitals, as in α -acetobromoglucose and α -glycosides. In β -glycosides, the nonbonding electrons of the endocyclic oxygen atom are delocalised into the antibonding C—H orbital, which is of higher energy than the C—O σ^* orbital. The overlap of the lone pair is therefore more efficient in the α -anomer than in the β -anomer.

The $n_Y \rightarrow \sigma^*$ delocalisation of electrons is also reflected by bond length changes, slightly shortening the Y—C1 bond while lengthening the C1—X bond (Figure 1.8).

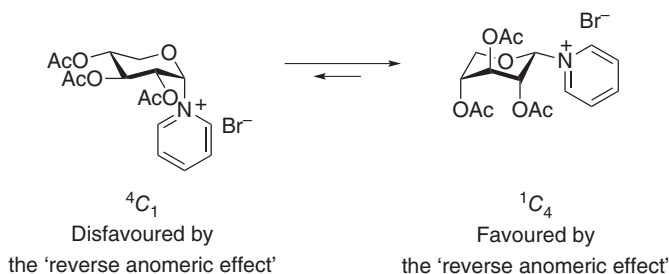
The anomeric effect is of a different size for every specific case. It is strongly influenced by the substituent at C2. When this is equatorial, as in glucose and galactose, the anomeric effect is weakened, whereas in the case of an axial C2 substituent as in mannose, it is enhanced. Moreover, the nature of the anomeric group is of crucial influence for the anomeric effect, as it is proportional to the electronegativity of the anomERICALLY bound atom. Solvents also influence the anomeric effect, such that increased polarity of the solvent used decreases the influence of the anomeric effect on the equilibration of the two alternative anomers in solution.

The anomeric effect may even induce a ring flip. In 2,3,4-tetra-*O*-acetyl- β -xylopyranosyl bromide, for example, the *gauche* effect—the preference for certain *gauche* arrangements of substituents over anti-periplanar (1,3-diaxial) arrangements—favors the 4C_1 chair. However, the anomeric effect opposes this preference, stabilising the 1C_4 conformation of β -xylopyranosyl bromide (Scheme 1.6).



Scheme 1.6 Peracetylated β -xylopyranosyl bromide adopts a 1C_4 conformation due to stabilisation by the anomeric effect.

If the substituent at the anomeric centre is clearly electropositive compared to the anomeric carbon, such as a positively charged nitrogen atom, the same electrostatic considerations as stated for the anomeric effect lead to the stabilisation of the anomer with the equatorially positioned anomeric group. This effect, which is causally no different from the anomeric effect, has been termed the ‘reverse anomeric effect’.⁵ It is assisted by the fact that an equatorial ring position is energetically favoured due to steric reasons, especially in the case of a large substituent like a pyridinium group (Scheme 1.7).



Scheme 1.7 An anomeric pyridinium group, for example, leads to a reverse dipole at the anomeric carbon, and consequently a ‘reverse anomeric effect’ is observed.

The reverse anomeric effect, however, has long been debated, since the substituents reported to produce it are often bulky aromatic rings, and the phenomenon could therefore be explained as a steric effect. Recent investigations, nevertheless, have provided evidence that the effect does exist.⁶

In alkyl glycopyranosides, the anomeric effect operates not only along the endocyclic C1—oxygen bond but also along the exocyclic C1—oxygen bond. The anomeric effect leading to preferred conformations of the exocyclic alkoxy group is called ‘*exo*-anomeric effect’. Again, it is an anti-periplanar arrangement of a lone pair and σ^* orbital which determines the favoured conformation, in this case the lone pair on the aglycone oxygen and the σ^* orbital of the C1—O5 bond. In axially configured acetals, the *exo*-anomeric effect is less important because it operates in the opposite direction than the ‘*endo*-anomeric effect’. However, in an equatorial acetal, the *exo*-anomeric effect is dominant and dictates the preferred conformation of the aglycone alkoxy group.

Briefly summarised, the *exo*-anomeric effect stabilises the axial anomer of a glycoside, whereas the *endo*-anomeric effect stabilises conformations in which an oxygen lone pair of the exocyclic alkoxy group is anti-periplanar to the $\sigma^*(C1O5)$ orbital (Figure 1.9).

The subject of the anomeric effect is important in carbohydrate chemistry but still under some debate.

Conformations of Monosaccharides

Like cyclohexane, the six-membered ring of monosaccharides can adopt two isomeric chair conformations, designated 1C_4 and 4C_1 , respectively. The conformation

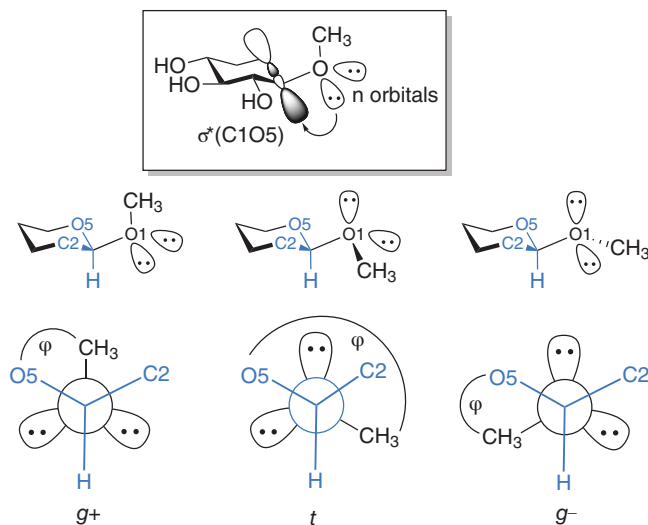
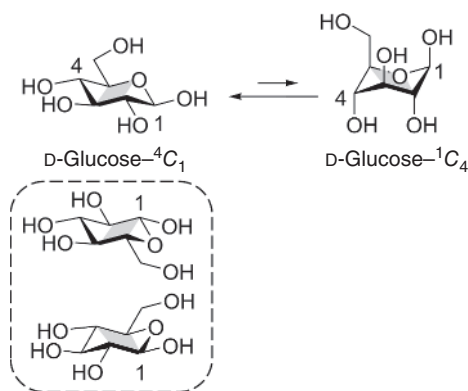


Figure 1.9 Three conformations of the aglycone in methyl β -D-xylopyranoside in 4C_1 conformation where an anti-periplanar arrangement of the lone pair on the aglycone oxygen and the σ^* orbital of the C1–O5 bond allows hyperconjugation. The lone pair electrons on the exocyclic oxygen O1 are shown as in sp^3 orbitals. The corresponding Newman projections along the glycosidic bond C1–O1 are shown. The symbols g^+ ($+60^\circ$), t (180°) and g^- (-60°) are used to denote the values of the torsion angle $\phi = C(\text{methyl})\text{--O1--C1--O5}$.

is indicated by an italic capital letter that specifies the type of ring shape, accompanied by numerals that distinguish the variants. Thus, for example, the complementary chair conformations of D-glucose are denoted as 4C_1 and 1C_4 . Here, the letter *C* stands for ‘chair’ and the numbers refer to the carbon atoms located above or below the reference plane of the chair, which is defined by atoms C2, C3, C5 and the ring oxygen.

In the style, monosaccharides are drawn throughout this book, the reference plane can be unambiguously assigned according to International Union of Pure and Applied Chemistry (IUPAC) rules. These require that the reference plane is defined by two parallel sides of the ring, chosen such that the lowest-numbered ring carbon atom, C1 in pyranoses, is exoplanar.⁷ With this in mind, it is evident that alternative drawings of the D-glucose- 4C_1 conformer, such as those shown in Scheme 1.8, are not suitable for denoting the chair conformation, since in both cases C1 is not located *exo* to the reference plane that can be drawn in the ring (Scheme 1.8).

Other conformations of monosaccharide rings are also denoted by specific capital letters: *B* for boat, *S* for skew, *H* for half-chair and *E* for envelope in six-membered rings. In five-membered rings, envelope (*E*) conformations are observed, and twist conformations are denoted by *T*. To precisely specify a particular conformation, in some of these cases, the positions of heteroatoms (mostly oxygen) must also be indicated, using either subscript or superscript atomic symbols in addition to the numerical labels for ring carbon atoms.



Scheme 1.8 The two chair conformations of *D*-glucose. The equilibrium lies on the side of *D*-glucose- 4C_1 . The boxed structures also describe *D*-glucose- 4C_1 correctly, but are not suitable for defining the conformation according to IUPAC rules (C_1 is not located *exo* to the reference plane).

To name other than chair conformers, the reference plane is defined as follows. In boat conformers, it is spanned by two parallel sides of the ring. In skew conformers, it is formed by three adjacent atoms and one non-adjacent atom, chosen such that the lowest-numbered atom (or the next highest) is *exoplanar*. For the envelope conformation, the reference plane includes five adjacent coplanar atoms, whereas in the half-chair it includes four adjacent coplanar atoms.

The mathematical description of sugar puckering was established by Cremer and Pople, who introduced a spatial polar coordinate system to quantify ring bending.⁸ In this representation, the complementary chair conformations are located at the north and south poles of a sphere, while the other conformers are distributed between them. The path between 4C_1 and 1C_4 chairs during ring inversion is determined by the local energetical minima and maxima corresponding to intermediate conformations.⁹ In fact, a pyranose ring can explore 38 different conformations, while a furanose ring can adopt 20 distinct conformations (Figure 1.10).

Principal conformations of the furanose ring are the envelope forms (1E , E_1 , 2E , E_2 , 3E , E_3 , 4E , E_4 , 0E , E_0) and the twist forms (1T_2 , 2T_1 , 2T_3 , 3T_2 , 3T_4 , 4T_3 , 0T_1 , 1T_0 , 0T_4 , 4T_0) (Figure 1.11).

The relative stability of carbohydrate conformers is governed by multiple factors, including intramolecular van der Waals interactions (dispersion forces), hydrogen bonding, electrostatic interactions and the anomeric effect. Conformational equilibria can also vary depending on the solvent. Among six-membered rings, the chair conformation is generally the most stable, although in some cases twist forms may also be significantly populated. Recent work by Alexander Auer and colleagues has provided a deeper understanding of these conformational equilibria.¹⁰

Fundamentally, conformations of monosaccharides depend on their configuration. In glucose, the 1C_4 conformation is strongly disfavoured relative to 4C_1 due to the van der Waals repulsion between 1,3-diaxially positioned ring substituents. In contrast, altrose and idose also populate the 1C_4 and a twist conformation,

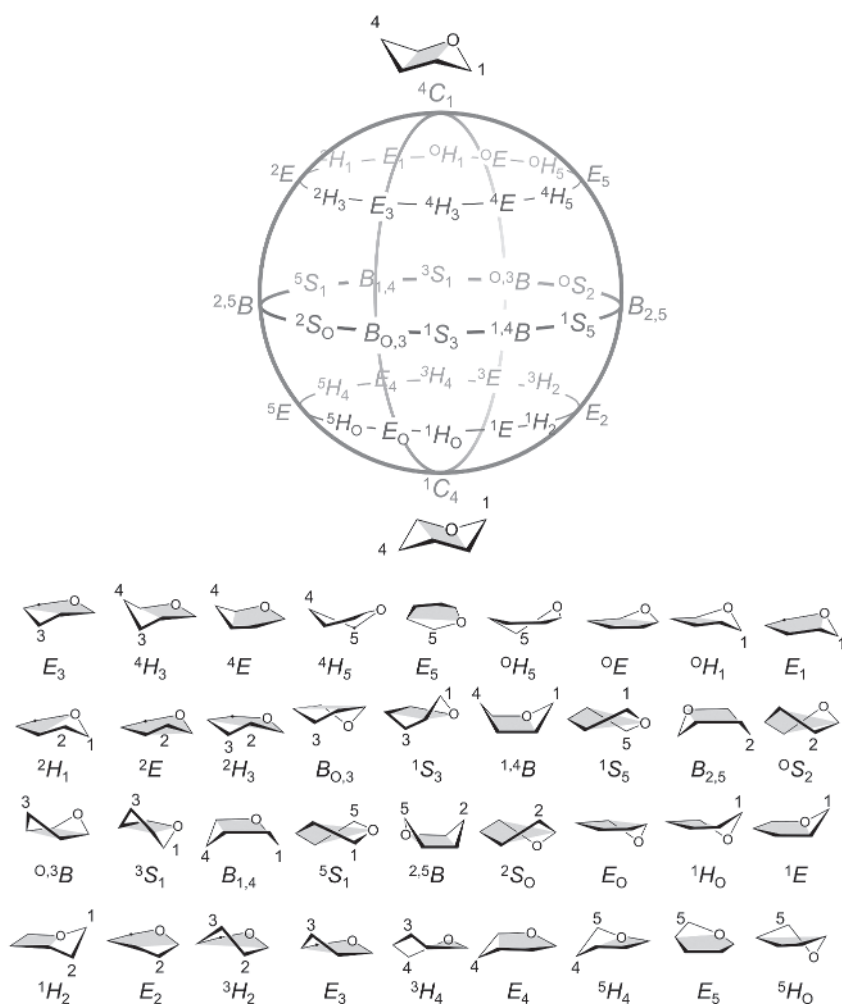


Figure 1.10 Cremer-Pople sphere of ring conformations. The poles correspond to the two chair conformations (4C_1 , 1C_4), while boat, skew, half-chair and envelope forms appear between them. Each conformer is drawn with its name and defining reference plane.

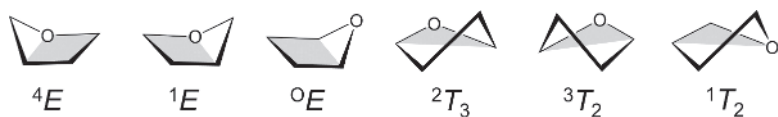


Figure 1.11 Some common furanose conformations.

Table 1.2 The free energy difference between 4C_1 and 1C_4 (in kJ mol^{-1}) and the respective chair populations (P) of all hexopyranoses.¹¹

Aldohexopyranoses	$\Delta G ({}^1C_4)$ (kJ mol^{-1})	P(4C_1) (%)	P(1C_4) (%)
α -D-Glc	17.7	99.94	0.06
β -D-Glc	27.6	99.99	0.01
α -D-Gal	14.8	99.73	0.27
β -D-Gal	25.6	99.99	0.01
α -D-Man	12.3	99.44	0.56
β -D-Man	19.7	99.97	0.03
α -D-Tal	10.3	98.50	1.50
β -D-Tal	15.9	99.88	0.12
α -D-All	6.9	94.46	5.54
β -D-All	15.8	99.84	0.16
α -D-Gul	5.6	90.54	9.46
β -D-Gul	13.6	99.57	0.43
α -D-Alt	5.6	85.6	14.4
β -D-Alt	6.6	94.10	5.90
α -D-Ido	0.4	54.4	45.6
β -D-Ido	5.7	93.58	6.42

respectively. In β -D-glucose, the free energy difference between the two chairs is approximately 28 kJ mol^{-1} ; in α -D-altrose 5.6, and in α -D-idose 0.4 kJ mol^{-1} (Table 1.2).

Because different conformers of a glycan rapidly interconvert at room temperature, the ${}^1\text{H}$ NMR spectra of such compounds reflect averaged coupling constants, since the NMR experiment is slow compared to the timescale of conformational interconversion. Determining the conformational composition in such dynamic mixtures is challenging and requires sophisticated methodologies.¹⁰ By contrast, the proton NMR spectrum of β -D-glucose and most monosaccharide derivatives reflects the coupling constants of a single, dominant conformer.

The interaction of heparin with antithrombin III is a classic example that illustrates the biological significance of conformationally dynamic sugars. Heparin is a member of the glycosaminoglycan (GAG) family, as we will discuss later. Its anticoagulant activity depends on binding to antithrombin III through a specific pentasaccharide sequence. A key constituent of this sequence is L-iduronic acid, exhibiting a dynamic equilibrium in solution between 4C_1 , 1C_4 and 2S_0 skew-boat conformations. The degree of O-sulfation of iduronic acid influences the population of these conformers. Remarkably, the overall three-dimensional shape of the heparin oligosaccharide chain is strongly determined by the conformational behaviour of its iduronic acid units. In particular, the 2S_0 conformation of iduronic

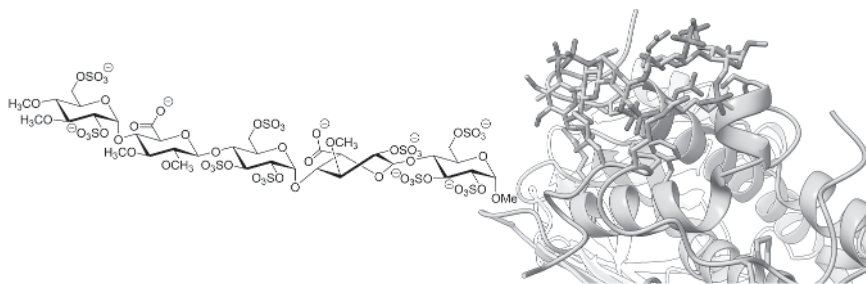
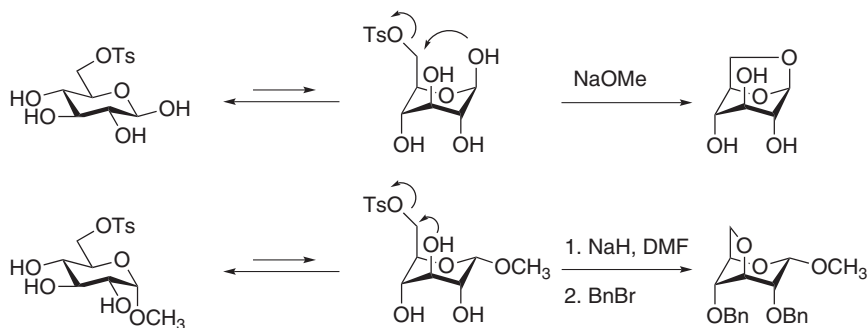


Figure 1.12 Shown is a pentasaccharidic heparin mimetic¹² in complex with antithrombin III.¹³ In this complex, the sulfated iduronic acid adopts a 2S_0 conformation. The complex is rendered with Chimera X.

acid plays a critical role in antithrombin activation. Indeed, this conformation has been observed in complexes of short therapeutic heparin mimetics bound to antithrombin III (Figure 1.12).

In carbohydrate chemistry, minor conformational isomers can be of importance, when they form the starting material in reactions which do not proceed with the major conformer. An important example is the synthesis of anhydro derivatives of monosaccharides which are versatile synthetic intermediates. Treatment of 6-*O*-tosyl-glucopyranose with sodium methoxide gives the 1,6-anhydro derivative which results from intramolecular nucleophilic attack of the anomeric hydroxyl group at the 6-position of the sugar ring in the unfavoured 1C_4 conformation. Analogously, treatment of the corresponding methyl glucoside, where the anomeric hydroxyl group is no longer available, with sodium hydride and benzyl bromide in DMF (dimethylformamide) gives the benzylated 3,6-anhydro glycoside, also in the 1C_4 conformation (Scheme 1.9).

In monosaccharides, the hydroxymethyl group ($-\text{CH}_2\text{OH}$) attached to C5 (in hexopyranoses) possesses considerable conformational freedom due to rotation about the C5—C6 bond described by the torsion angle ω . The three staggered conformations are the most commonly observed ones, referred to as *gg* (*gauche-gauche*), *gt* (*gauche-trans*) and *tg* (*trans-gauche*). They are defined according to the relative



Scheme 1.9 Products can arise via a conformer that is only a minor component at equilibrium (here, reaction through 1C_4 despite 4C_1 predominance).

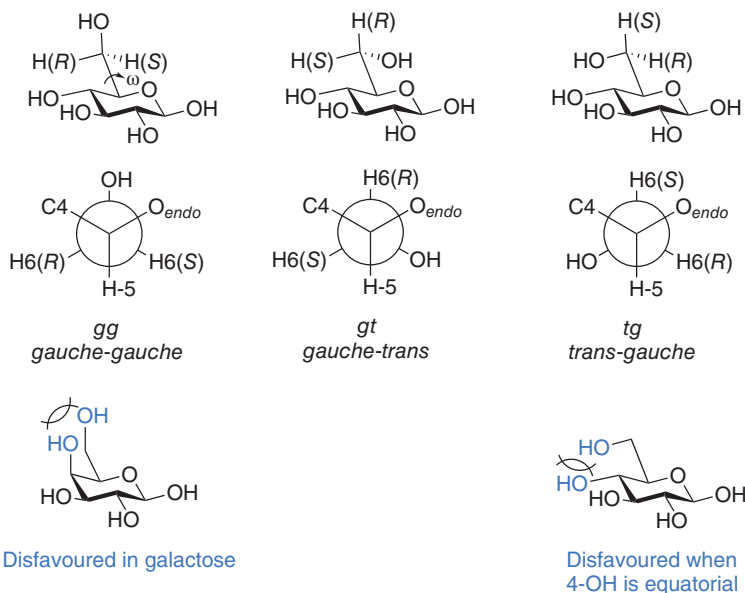


Figure 1.13 The most common staggered conformations of the 5-hydroxymethyl group in monosaccharides, obtained by rotation around C5–C6 bond. The C6-protons are denoted according to their prostereogenic configuration. Due to 1,3-diaxial interactions between the 4- and 6-hydroxyl groups, in sugars with an axial 4-OH group, the *gg* form is disfavoured, whereas in sugars with an equatorial 4-OH, the *tg* form is disfavoured.

orientation of H5 and the *pro-S* hydrogen (H6(S)), followed by the relative orientation between H5 and the *pro-R* hydrogen (H6(R)).

These conformers are conveniently depicted as Newman projections, where the *gauche* (*g*) and *trans* (*t*) relationships become evident. To study the conformations of the exocyclic hydroxymethyl groups in saccharides, NMR spectroscopy and theoretical methods are frequently combined.¹⁴ In most sugars, the 5-hydroxymethyl group adopts the *gg* or *gt* conformations, respectively. In galactosides, however, bearing an axial hydroxyl group at C4, the *gg* form is disfavoured due to 1,3-diaxial interactions between the 4- and 6-hydroxyl groups.

The various conformations of the C5-hydroxymethyl group influence multiple saccharide properties; they alter hydrogen-bonding patterns, reactivity, the molecular dipole moment and the three-dimensional structure relevant to lectin recognition, for example (Figure 1.13).

Most Relevant Monosaccharides

The monosaccharides of the sugar trees with the general formula $C_n(H_2O)_n$ described earlier are not all equally important in nature. On the other hand, many of the most relevant sugars are different from the molecular formula $C_n(H_2O)_n$. The structures of the 10 most common monosaccharides found in humans are given in Figure 1.14. In some of them, the sugar ring is modified with an acetamido group

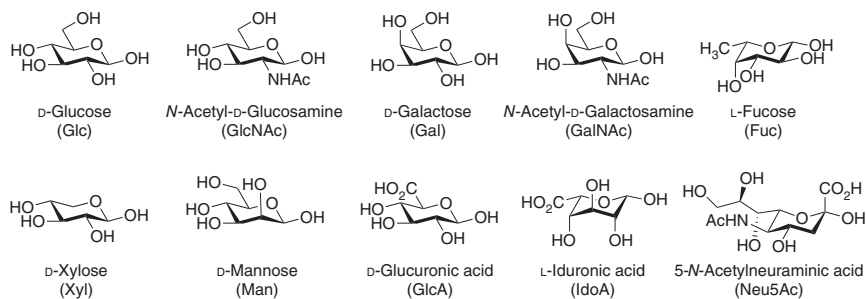


Figure 1.14 Structures and names of the 10 most common sugars found in humans. 5-N-Acetylneuraminic acid is the most frequent form of a diverse group of so-called sialic acids. The abbreviations of the monosaccharide names given are used very frequently in glycoscience.

or oxidised at the 6-position. Other than all other representatives of this group, 5-N-acetylneuraminic acid (Neu5Ac) consists of nine carbon atoms. It belongs to the family of sialic acids which are of great biological importance as will be discussed later (Figure 1.14).

There is a wide variety of monosaccharides that also occur in species other than mammals, such as microbes, yeasts and plants. Often they are deoxygenated, branched or modified by amino groups, for example. Many of them have been given trivial names as they are frequently found as constituents of different classes of biologically active compounds. For example, L-vancosamine and its *N,N*-dimethyl derivative are found as the sugar constituents of several antibiotics such as the important glycopeptide antibiotic vancomycin (Figure 1.15). The structures of the diverse saccharides compiled in this chapter are also intended to serve as a small reference work.

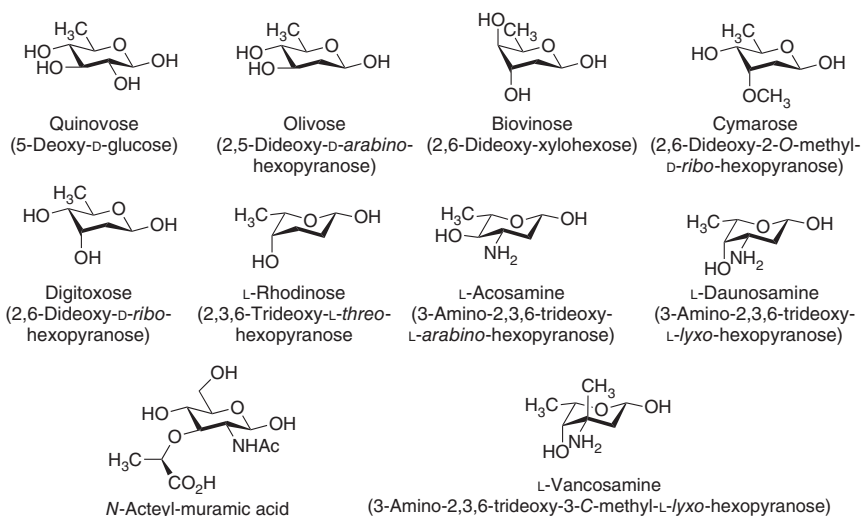


Figure 1.15 Structures and names of the most common modified monosaccharides found in nature. The IUPAC names provided may also be viewed as nomenclature exercise.

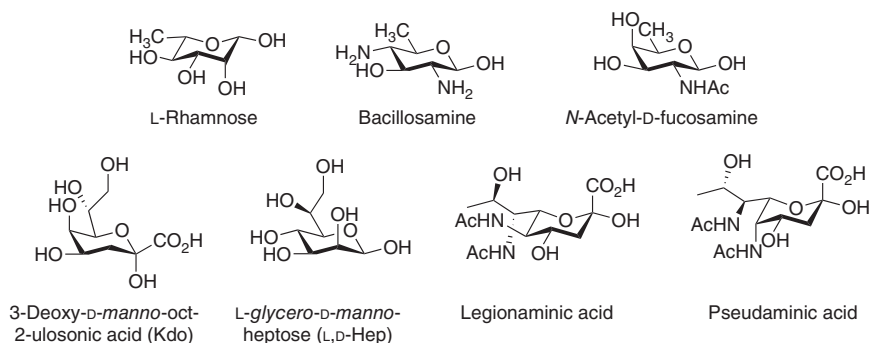


Figure 1.16 Structures and names of bacteria-specific monosaccharides.

Recently, the diversity of monosaccharides has been represented in the form of a periodic table. This is an interesting approach, since carbohydrates show both remarkable diversity and underlying similarities. In ‘A Periodic Table of Monosaccharides’, 103 monosaccharides were classified into five major groups, aldoses, ketoses, amino sugars, uronic acids and deoxy sugars, with the periods defined by the number of carbons in the backbone.¹⁵ This approach is certainly somewhat playful, but it encourages an appreciation of the remarkable structural diversity of monosaccharides.

Glycosylation is also highly relevant for the virulence of many bacterial pathogens. Various surface-associated glycoconjugates have been identified in bacteria that promote infectious behaviour and survival in the host. These glycoconjugates typically contain unique monosaccharide components which differ from mammalian monosaccharides.¹⁶ Due to their distinct differences from human carbohydrates, bacterial glycosylation systems are promising targets for antibacterial strategies. The most important monosaccharides which occur in many clinically relevant bacterial pathogens are drawn in Figure 1.16.

Cyclitols and Alditols

Cyclitols, as cyclic polyhydroxyalkanes, have a lot in common with monosaccharides and are often targets or starting materials in carbohydrate chemistry. Of the cyclitols, the hexahydroxycyclohexanes are called inositols and are best known. There are nine possible stereoisomeric inositols, of which *myo*-inositol is often simply referred to as inositol. If *myo*-inositol is phosphorylated at the 1, 4 and 5 positions on the ring, it functions as a second messenger in various signal transduction pathways (Figure 1.17).

Closely related to the inositols are the carba-sugars which were previously known as pseudo-sugars. They differ from normal sugars in the replacement of the ring oxygen atom with a methylene group. The well-known glycosidase inhibitors cyclophellitol and the synthetic anti-influenza drug oseltamivir, which will be discussed later, can also be counted among the carbohydrate sugars.

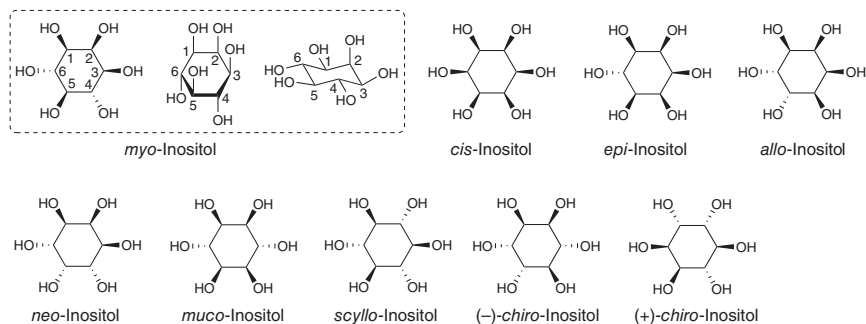


Figure 1.17 Structures of all nine stereoisomeric inositols; only the *chiro*-inositols are chiral, the other stereoisomers are *meso* forms. All inositols are numbered as indicated for *myo*-inositol, which is depicted as Mills and Haworth formula and as chair conformation.

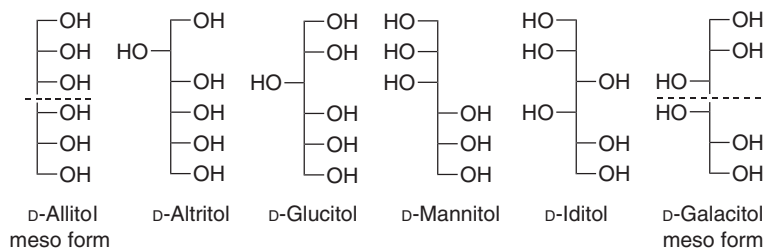


Figure 1.18 The six diastereomeric *D*-alditols derived from aldohexoses or aldoketoses, respectively. Mirror planes are indicated as dashed line.

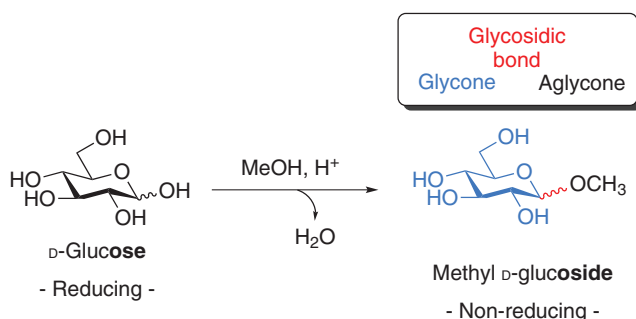
Mild reduction of aldoses and ketoses with, for example, sodium boronate leads to sugar alcohols which are called alditols. Like the corresponding monosaccharides, they are named according to the number of carbon atoms, as pentitols and hexitols, for example. Whereas in the reduction of aldoses only one product is obtained, the reduction of ketoses leads to two diastereomers.

Reduction of *D*-altrose gives the same product (*D*-altritol) as the reduction of *D*-talose. Likewise, *D*-glucose and *D*-gulose both give the same alditol, *D*-glucitol, which is also known as sorbitol. Consequently, only six different *D*-alditols are obtained from the eight stereoisomeric *D*-aldoses. From those, *D*-allitol and *D*-galactitol are *meso* forms, the other four are chiral (Figure 1.18). Alditols are, like the aldaric acids, sugar derivatives with identical end groups and therefore include interesting symmetry elements. This may be utilised to convert, through a series of chemical conversions, a sugar from the *D*-series into an *L*-sugar.

1.2 Structure of Oligosaccharides

When a free, so-called reducing monosaccharide such as *D*-glucose, resembling a cyclic hemiacetal, reacts with another alcohol under acid catalysis, a mixed acetal is formed, which is generally called a glycoside. In the simplest case, methanol

can be employed as the reacting alcohol, and then an anomeric mixture of methyl glycosides, methyl α - and β -D-glucoside, is formed from glucose. Note that a reducing monosaccharide's name ends with the suffix 'ose', whereas a glycoside is denoted with the suffix 'oside'. The moiety at the reducing end of the glycoside is called the aglycone, and the remaining saccharide portion is referred to the glycone part of the glycoside. The bond between the two components of the acetal is called the glycosidic bond. Whereas a cyclic hemiacetal can be oxidised at the anomeric centre via its open-chain form with which it is in equilibrium, this is not possible in glycosides. Pyranoses and furanoses are therefore often called 'reducing', whereas glycosides, namely pyranosides and furanosides, are non-reducing. Also, oligosaccharides with a hemiacetal 'end' are called reducing (Scheme 1.10).



Scheme 1.10 A reducing sugar such as a pyranose or a furanose, resembling a cyclic hemiacetal, can be converted into an acetal, which is generally called a glycoside.

Instead with a simple alcohol such as methanol, a cyclic hemiacetal—a reducing sugar—can also form a glycoside with other carbohydrates. For this reaction, different glycosylation methods have to be employed, rather than just boiling the sugar in alcohol under acid catalysis, and this will be discussed in Chapter 3. At this point, it is important to first note that the reaction of monosaccharides as monomers with other sugars can lead to oligomeric glycosides, which are named disaccharides in the simplest case. Larger glycosides are called tri- and tetrasaccharides and so forth. They form a group of oligosaccharides which is found in nature in unprecedented diversity. They can be linear or branched or even multiply branched. Indeed, most carbohydrates do occur as oligo- as well as polysaccharides, rather than as monosaccharides. No strict borderline is drawn with polysaccharides; however, the term 'oligosaccharide' is commonly used to refer to a defined structure as opposed to a polymer. Oligosaccharides in which structural elements, such as the interglycosidic linkages, differ from those found in nature are often called 'oligosaccharide mimetics'.

Disaccharides are the simplest oligosaccharides and a number of them belong to the least expensive carbohydrates available, as they can be easily obtained by hydrolysis of abundant polysaccharides (often regarded as 'renewable resources') or isolation from various materials, e.g., plants, microorganisms, yeast, marine sources or milk. Thus, maltose is obtained in large quantities from starch and cellobiose from

cellulose. α,α -Trehalose, in which two glucose moieties are linked via both of their anomeric centres, is a non-reducing sugar and is found in yeasts and fungi. Lactose is naturally occurring in the milk of mammals; it serves as the principal source of energy for their young. Sucrose (in German ‘Saccharose’) is chiefly found in sugar cane and sugar beet, but is also found in many other plants, especially fruits. Its particular attribute is its sweetness. Although many carbohydrates are sweet, sucrose has become the sugar of commerce, primarily because of the ease by which it is extracted in high purity from sugar cane or sugar beets. The structures of the most common disaccharides are given in Figure 1.19.

Even trisaccharides occur relatively frequently in nature, and the structures of some common examples are in Figure 1.20. In contrast, larger oligosaccharides are rarely found in their free form. Nevertheless, complex oligosaccharides with numerous structural variations are an important part in many biologically active natural products, and moreover, they are linked to proteins or lipids, forming glycoconjugates that are of major significance in glycobiology and cell biology, as we will see later.

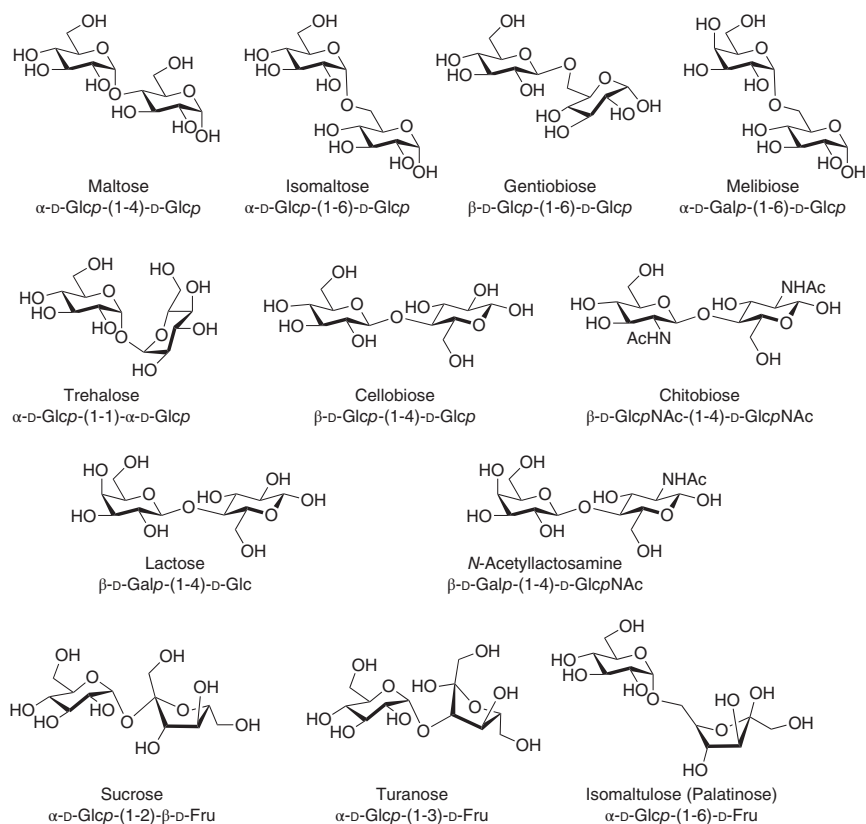


Figure 1.19 Structures and trivial names of the most common disaccharides. The abbreviated designations according to IUPAC nomenclature are also given.

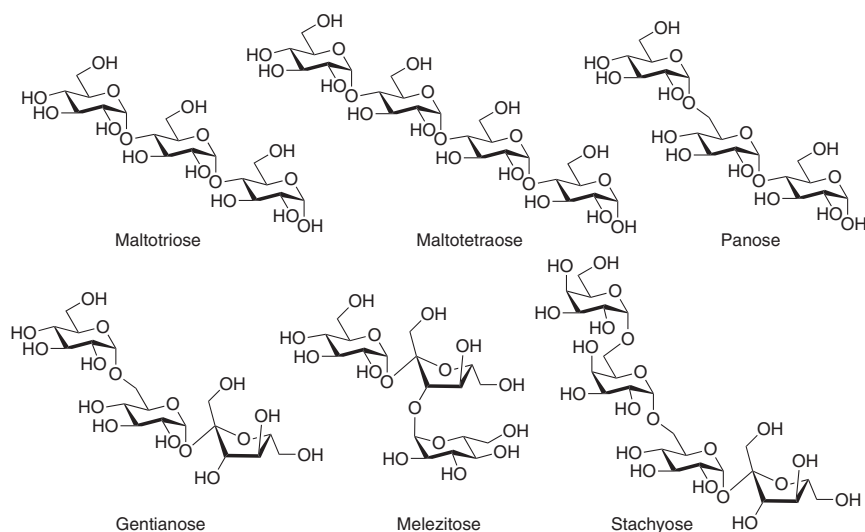


Figure 1.20 Structures and trivial names of common oligosaccharides found in nature.

Many oligosaccharides and oligosaccharide analogues are known for their biological effects. For example, a large number of carbohydrates have been evaluated as inhibitors of the enzymes which are involved in polysaccharide digestion, the so-called glycosidases. Many α -glucosidase inhibitors were originally isolated from various species of *Streptomyces* bacteria. Among those, acarbose has become an important antidiabetic drug, as it serves as good α -amylase inhibitor. It is a carba-oligosaccharide, also called pseudo-oligosaccharide, consisting of two glucose moieties, a 4-amino-4,6-dideoxy-glucose derivative and a tetra-hydroxylated cyclohexene derivative.¹⁷

Streptomycin was the first oligosaccharide analogue with pharmaceutical relevance. It was discovered in 1944 as a chemotherapeutic agent and has been used to treat some penicillin-resistant strains of Gram-negative and Gram-positive bacteria, including mycobacteria. Streptomycin and its relatives were found in culture filtrates of several *Streptomyces* strains. The middle furanose part is called 'streptose' and was the first branched-chain sugar obtained from microorganisms.

The so-called aminoglycosides are composed of two to five monomers and contain amino sugars and carba-sugars as can be seen in kanamycin A. Aminoglycosides are used as broad-spectrum antibiotics targeting the bacterial ribosomes and inhibiting protein synthesis.

A remarkable variety of complex oligosaccharides are contained in human milk, many of which are fucosylated or sialylated. They are not digested by the infant but instead promote the growth of beneficial gut bacteria and block pathogen adhesion. They have an immunostimulating effect and even contribute to brain development through their sialic acid content. 2'-O-Fucosyllactose is given as a typical example.

Cardiac glycosides consist of a steroid skeleton conjugated with a lactone ring and a carbohydrate moiety. They are able to block the heart's sodium and potassium

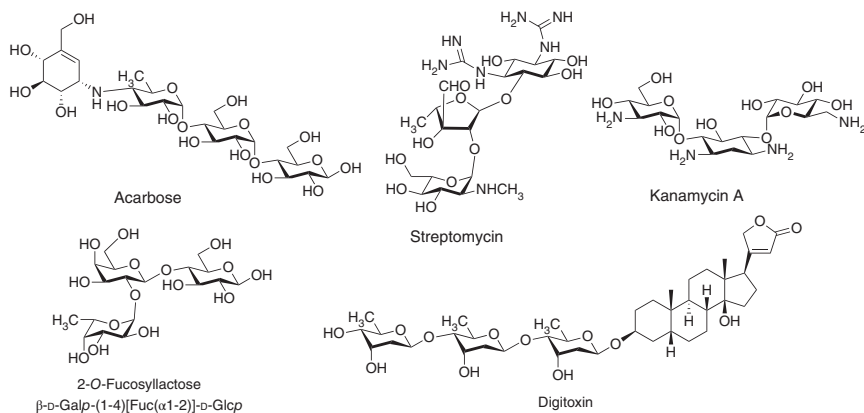


Figure 1.21 Structures and some biologically active oligosaccharides.

ATPase current. Cardiac glycosides such as digitoxin have traditionally been used in the treatment of heart failure and atrial arrhythmias (e.g. atrial fibrillation). Although their clinical use has declined with the availability of safer alternatives, they remain valuable in selected patients (Figure 1.21).

Looking at the diversity of biologically active carbohydrates highlights the structural variety of sugars and helps one appreciate the complexity of carbohydrate biosynthetic pathways. Moreover, sugars are frequently covalently bound to non-carbohydrate natural products of different kinds. When sugars are covalently linked to other biomolecules, these are called glycoconjugates, such as glycoproteins and glycolipids, which will be discussed in depth later. Many of the smaller glycoconjugates are biologically active drugs and possess, for example, antibiotic activity, as we have seen. Many pharmaceuticals belong to this category, such as cardiac glycosides, anthracyclines, macrolides, ergot alkaloids and calicheamicins, to name a few. These molecules are glycosylated with oligosaccharides of varying complexity, which are important for biological storage and transport, for pharmacokinetics and pharmacodynamics, and which influence the properties of the molecule such as solubility, efficacy and selectivity among others.

Cyclodextrins

Cyclodextrins are cyclic oligosaccharides in which monosaccharide units are linked by intersaccharidic $\alpha(1-4)$ -glycosidic bonds to form a ring structure. They are readily available from starch by large-scale preparation using cyclodextrin glucotransferases (CGTases), amylolytic enzymes produced by *Bacillus macerans* and other bacterial microorganisms. By this enzymatic procedure, a mixture of cyclodextrins is obtained which can be separated by chromatography or fractionated crystallisation. Cyclodextrins are also known as Schardinger dextrins in recognition of the fact that Schardinger first described their preparation and properties in detail. The most common and commercially available cyclodextrins are those consisting of six, seven and eight glucose moieties, which are called α -, β - and γ -cyclodextrin

according to a special nomenclature that this rather unique class of molecules has acquired.

The shapes of cyclodextrins are well established on the basis of X-ray crystallographic studies which reveal that in the solid state, cyclodextrins are conical molecules with well-defined cavities, having hydrophilic exteriors and more hydrophobic interiors. The chemistry which can be performed with cyclodextrins is largely influenced by the remarkable difference in reactivity of the three different types of hydroxyl groups present in the molecule, which decreases in the order: 6-OH > 2-OH > 3-OH.

With the ability to bind other compounds as guest molecules in close proximity to specific hydroxyl groups, cyclodextrins have been perceived as simple models of enzymes and have been found to exhibit catalytic activity in certain reactions, for example in the base-catalysed hydrolysis of carboxylic acid esters. Cyclodextrins can also be used to bring hydrophobic compounds into aqueous solution or may be developed as drug delivery systems. Furthermore, modified cyclodextrins are used as stationary phases in enantioselective gas chromatography (Figure 1.22).

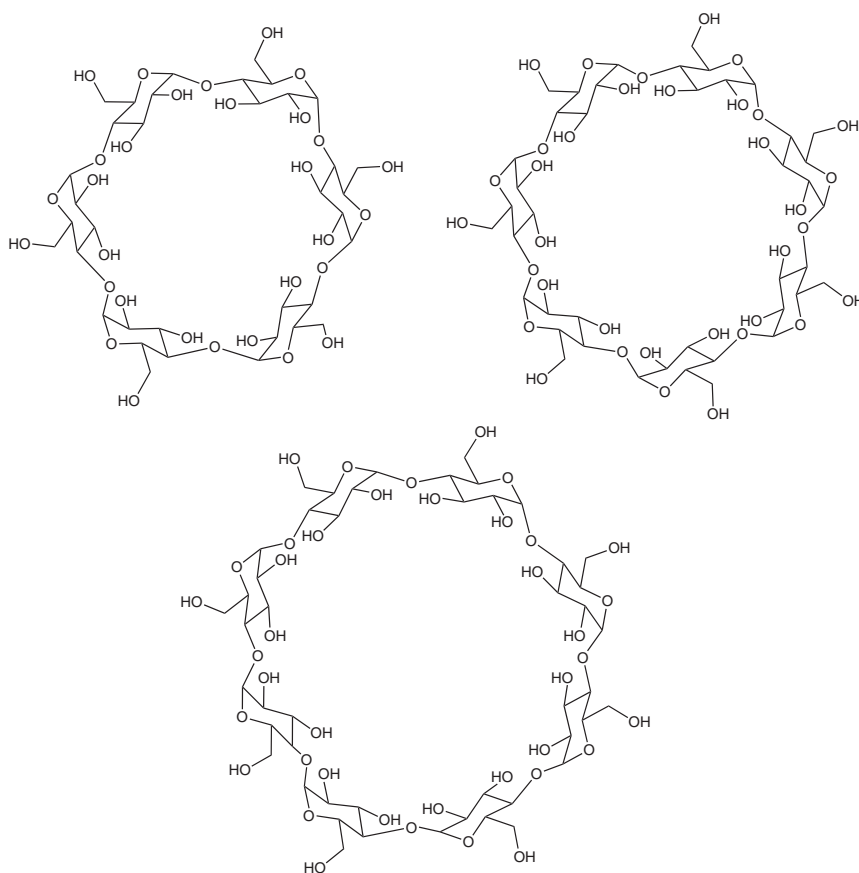


Figure 1.22 Structures of α -, β - and γ -cyclodextrin.

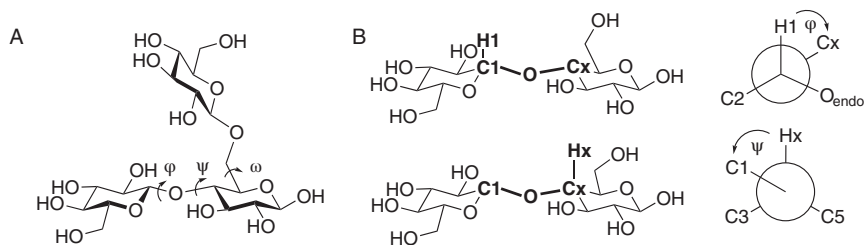


Figure 1.23 A: The conformation of an oligosaccharide is determined by the torsion angles ϕ (phi), ψ (psi) and ω (omega). B: The angle ϕ is defined by the atoms H1–C1–O–Cx, the angle ψ by C1–O–Cx–Hx and ω by O6–C6–C5–O5.

Conformations of Oligosaccharides

The three-dimensional properties of oligosaccharides are of particular importance in the biochemical context of molecular recognition and biological communication processes in which carbohydrates are involved. In addition to the conformational dynamic of the sugar ring, which we have discussed earlier, there is also a considerable conformational freedom around the glycosidic bond. Two torsion angles, ϕ (phi) and ψ (psi), define the relative orientation of two glycosidically linked monosaccharide units such as 1-1-, 1-2-, 1-3- or 1-4-linked glycosides. The angle ϕ is defined by the atoms H1–C1–Ox–Cx and the angle ψ by C1–Ox–Cx–Hx (Figure 1.23).

In the case of a 1-6-linked disaccharide, the additional degree of freedom is described by the torsion angle ω (omega), which is defined by the atoms O6–C6–C5–O5. The three most common conformations of the C5–C6 fragment are the staggered conformations *gg*, *gt* and *tg*, as we have already discussed.

The geometry of the glycosidic bond is of fundamental importance in glycobiology. Several factors influence the torsion angles defining the conformation of the glycosidic bond, including intra- and intermolecular interactions, and, of course, interactions with the solvent. In aqueous solution, oligosaccharide conformations are very different from conformations in condensed phase, gas phase or more polar solvents. The exo-anomeric effect influences ϕ , whereas the ψ dihedral angle is mostly affected by steric interactions.¹⁸

The investigation of oligosaccharide conformations, and the calculation of their energy minima in both free and receptor-bound forms, relies heavily on computational methods in combination with advanced NMR techniques. For crystalline carbohydrates, three-dimensional structures can also be determined by X-ray crystallography.

1.3 Structure of Polysaccharides

It is estimated that approximately 4×10^{11} tons of carbohydrates are biosynthesised each year on Earth by plants and photosynthesising bacteria. The majority of these carbohydrates are produced as polysaccharides. Polysaccharides are macromolecules consisting of a large number of monosaccharide residues. For

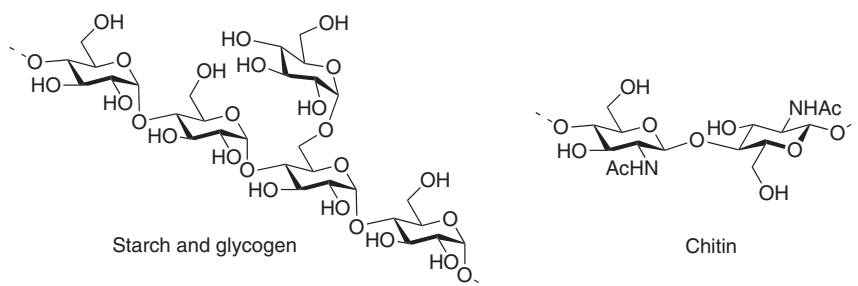


Figure 1.24 Monosaccharide connections found in starch, glycogen and chitin. In starch and glycogen, *D*-glucopyranosyl moieties are α -(1-4) and α -(1-6)-glycosidically linked. In chitin, GlcNAc residues are β -(1-4)-linked.

polysaccharides which contain a substantial proportion of amino sugar residues, the term glycosaminoglycan is a common one. Polysaccharides which consist of only one kind of monosaccharide are called homopolysaccharides (homoglycans); when they are built up of two or more different monomeric units, they are named heteropolysaccharides (heteroglycans). In the latter type, the monosaccharide units are usually linked to each other in a definite pattern, rather than randomly. Certain sequences of monomeric building blocks are often found to be regularly repeated as so-called repeating units. Homo- as well as heteropolysaccharides can be linear or branched (Figure 1.24).

For polysaccharides, as for every polymer, it is not possible to attribute one distinct molecular weight as they are polydisperse molecules, which are characterised by an average molecular weight. The number of monosaccharide units in a polysaccharide is termed degree of polymerisation or DP. The size of a polysaccharide varies between approximately 16,000 and 16,000,000 Da.

Polysaccharides exist in an enormous structural diversity as they are produced by a great variety of species, including microbes, algae, plants and animals. Among these are fructans, xanthans, fucans, bacterial gel polysaccharides, capsule polysaccharides of bacteria, and agar, which is a mixture of two polysaccharides and is obtained from red-purple seaweeds.

In some cases, polysaccharides are conjugated to peptide or protein structures, respectively, such as in peptidoglycans and proteoglycans which serve important biological roles. Their structures together with other most well-known polysaccharide structures are briefly reviewed in the following pages.

Starch and Dextrins

Starch is a mixture of two glucans (polysaccharides built from glucose), which are called α -amylose and amylopectin. It is synthesised by plants as their principal food reserve and deposited in the plant cell cytoplasm as insoluble granules.

α -Amylose is a linear polymer of several thousand glucose residues, α -(1-4)-glycosidically linked. This polymer adopts an irregularly aggregating, helically coiled conformation containing regular left-handed helix regions. Amylopectin, on

the other hand, carries α -(1-6)-linked branches every 24–30 glucose residues of the α -(1-4)-linked chain, resulting in a tree- or brush-like structure. It contains up to a million glucose residues which makes it among the largest molecules occurring in nature.

Starch is degraded by enzymes called amylases, which randomly hydrolyse the α -(1-4)-glycosidic bonds digesting the polysaccharide into oligosaccharide fragments, such as maltose and maltotriose as well as oligosaccharides containing α -(1-6) branches. The oligosaccharides produced by amylase digestion can be further hydrolysed to glucose by specific glucosidases and by debranching enzymes, which remove the α -(1-6) branches.

The low-molecular-weight polymers produced by the enzymatic degradation of starch (and glycogen) are called dextrins. Dextrins are mixtures of glucans containing α -(1-4)- and α -(1-6)-glycosidic bonds. Dextrins are white, yellow, or brown powders that are partially or fully water-soluble. White and yellow dextrins from starch roasted with little or no acid are called British gum which is used in industry as a paper adhesive.

Glycogen

Glycogen is the storage polysaccharide of animals and is present in all cells but most prevalently in skeletal muscle and liver, where it occurs in cytoplasmic granules. The primary structure of glycogen only differs from that of amylopectin in that it is more highly branched, with branching points occurring every 8–12 glucose residues of the α -(1-4)-linked glucan chain. The degree of polymerisation of glycogen is similar to that of amylopectin. In the cell, glycogen is degraded for metabolic use by glycogen phosphorylase, which cleaves the α -(1-4)-linkages sequentially inwards from its non-reducing ends to release glucose-1-phosphate which can be fed into the citric acid cycle. The branching points are hydrolysed by glycogen debranching enzyme. Glycogen contains about 1% covalently linked protein.

Cellulose

Cellulose is an abundant carbohydrate of commercial and biological importance, found in all plants as the major structural component of the cell walls. Cellulose in wood is mixed with many other polymers such as hemicelluloses and lignin. It has to be split from these components to be used for paper production. The fluffy fibre found in the cotton ball is the purest naturally occurring form of cellulose.

Cellulose is the β -isomer of amylose consisting of β -(1-4)-linked glucose residues. The different stereochemistry of the glycosidic linkage compared to amylose gives cellulose totally different properties. In contrast to amylose, the β -linkages in cellulose allow the polymer to fold in a fully extended conformation to form a sheet-like secondary structure. The tertiary structure of cellulose is characterised by parallel-running intermolecular hydrogen-bonded cellulose chains further associated by hydrogen bonds and van der Waals forces to produce three-dimensional microfibrils. This gives cellulose fibres exceptional strength and makes them water

insoluble despite their hydrophilicity. The cellulose microfibrils give an X-ray diffraction pattern that indicates regular, repeating microcrystalline structures interspersed by less-ordered paracrystalline regions.

As a consequence of its three-dimensional structure, cellulose cannot be hydrolysed by starch-degrading enzymes. Enzymes that degrade cellulose, known as cellulases, are produced by microorganisms.

Chitin and Chitosan

Chitin is the principal structural component of the exoskeleton of invertebrates such as crustaceans, insects and spiders and is also present in the cell walls of most fungi and many algae, thus representing the widespread occurrence of carbohydrates on Earth, their abundance being almost equivalent to that of cellulose. The structure of chitin is related to that of cellulose. It is a homopolymer of β -(1-4)-linked *N*-acetyl-D-glucosamine residues. X-ray analysis indicates that chitin and cellulose have similar secondary and tertiary structures. De-*N*-acetylated chitin is called chitosan. Chitosan is of interest as a material for numerous applications such as in green synthesis and biocatalysis as well as for its antimicrobial activity.

Pectins

Pectins are heteropolysaccharides occurring in all plants primarily in their cell walls. They act as intracellular cementing material that gives body to fruits and helps them keep their shape. When fruit becomes overripe, the pectin is broken down into its monosaccharide constituents. As a result, the fruit loses its firmness and becomes soft. One of the most prominent characteristics of pectins is their ability to form gels. Pectins contain a high proportion of D-galacturonic acid units, which are α -(1-4)-linked in repeating units. Additionally, pectins contain α -L-rhamnose moieties that carry oligosaccharide side chains comprising arabinose, galactose or xylose. The galactopyranosyl uronic acid units can be esterified and acetylated, modifications that significantly influence the properties of pectins. The molecular size of pectins greatly varies with a typical average molecular size of 100 kDa.

Dextrans

It has long been known that sucrose solutions convert into viscous solutions, gels or flocculent precipitates. The material that produced such changes in sucrose solutions was isolated and found to be a polysaccharide that was called dextran. Dextrans are α -(1-6)-linked, mostly branched α -D-glucopyranosyl polysaccharides synthesised from sucrose by lactic acid bacteria. All known dextrans are branched at different branching points and to different extents. The unique properties of dextrans, their biocompatibility, biodegradability and three-dimensional structures are due to the percentage and manner in which the branches are arranged. Also

the molecular weight of dextrans determines their physicochemical properties and consequently, dextrans of different molecular weights are used for different applications. Quite prominent are clinical-type dextrans with molecular weights of 40 or 70 kDa. They are used as blood plasma extenders because of their volume-expanding effects.

Peptidoglycan or Murein

Peptidoglycan, also known as murein, is a complex large macromolecule consisting of sugars and amino acids. Peptidoglycan surrounds the bacterial cytoplasmic membrane, where it serves a structural role, conferring cell shape and contributing to the ability of bacteria to resist the effects of internal osmotic pressure. The amount and location of peptidoglycan surrounding the bacterial cell determines the outcome of the so-called Gram staining. In Gram-positive bacteria (surrounded by only one membrane), the peptidoglycan layer is substantially thicker than in Gram-negative bacteria, where the cell wall consists of two membranes which are separated by the periplasm in which thin layers of peptidoglycan reside. The average chain length of the glycan strands in peptidoglycans varies between Gram-negative and Gram-positive bacteria. The saccharide chains consist of alternating residues of β -(1-4)-linked *N*-acetyl-D-glucosamine and *N*-acetyl-D-muramic acid (MurNAc). Attached to MurNAc is an oligopeptide chain of three to five amino acids, which cross-link the polysaccharide chains of different strands forming a mesh-like layer. Biosynthesis of peptidoglycan is inhibited by the antibiotic penicillin.

1.4 Nomenclature of Saccharides

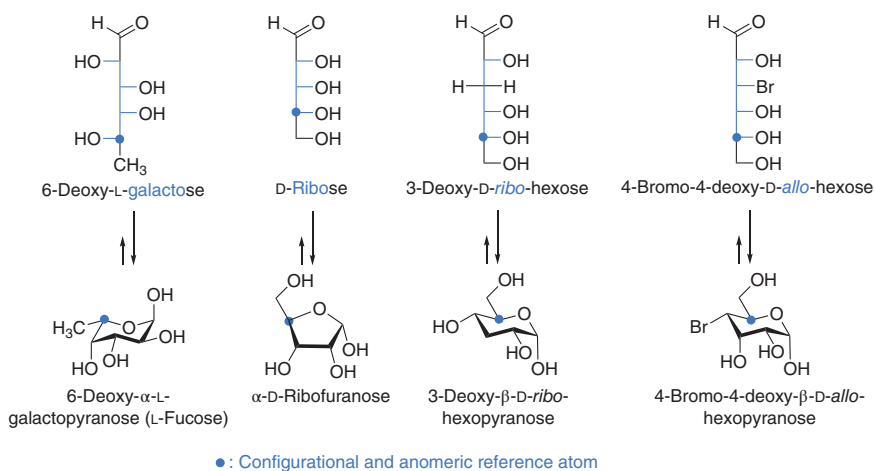
Nomenclature of Monosaccharides

Many sugars are named with trivial names, but derivatisation of carbohydrates requires systematic naming. Therefore, it is important to understand at least the most basic principles of carbohydrate nomenclature.^{19, 20} The correct naming of carbohydrate derivatives is dependent on stem names, which are derived from the trivial names, which have been given to the common aldoses consisting of three, four, five and six carbon atoms. Thus, these aldose stem names correspond to a specific sequence of one, two, three and four stereocentres, respectively. This sequence is used to deduce a 'configurational prefix' for a specific carbohydrate derivative, which is typically written in italics in the sugar's name, for example, in a *ribo*-hexose or *erythro*-pentose derivative.

When a monosaccharide contains a deoxy function, this chain position is ignored for the purpose of assigning the configurational prefix. Thus, the configurational prefix covers the entire sequence of stereocentres even when a CH₂ group divides the sequence of stereocentres into two sets. The prefix 'deoxy' is used preceded by the position. For certain deoxy sugars, trivial names are also commonly used, especially when the deoxygenation is not at a stereocentre. 6-Deoxy-L-galactose for example is called L-fucose, and 6-deoxy-L-mannose is known as L-rhamnose

Depending on the size of the formed ring, monosaccharides are named as furanoses when the ring is five-membered, and six-membered rings are named as pyranoses. Often, the ring sizes of a six-membered sugar is not specified; for example, 'glucose' is used instead of 'glucopyranose', but this has no particular implication. When a five-membered hexose is named, it is always specified as 'hexofuranose'.

The assignment of the D- or L-prefix, respectively, is required, for which the stereochemistry at a so-called 'configurational reference atom' is essential. As we have discussed already earlier, the highest-numbered stereocentre of the carbohydrate chain is used as the configurational reference atom. For the assignment of the anomeric configuration, which is not reflected in the configurational prefix, the stereochemistry at the 'anomeric reference atom' has to be considered, which is the same as the configurational reference atom in simple sugars (Scheme 1.11).



Scheme 1.11 Examples of carbohydrate nomenclature: Systematic names are based on stem names corresponding to the stereochemical sequence of aldoses in the Fischer projection. Deoxy substitutions are ignored when assigning the configurational prefix.

When the hydroxyl group is substituted, the substituted position is indicated as 'deoxy-substituent-prefix' in the name of the derivative preceded by the position of the modification, for example 3-deoxy-3-methyl or 4-bromo-4-deoxy. The stereochemistry at the carbon atom carrying the new substituent is expressed as before with the substituent regarded as equivalent to OH. Substitution of the hydrogen atom of a hydroxyl group is specified as O-substitution, e.g. '4-O-methyl-' or '2,3-di-O-acetyl-'. Likewise, substitution of one (or two) hydrogen(s) of an amino group are named as N-substitution, for example, 'N-acetyl-'. Monosaccharides carrying a halogen atom at the anomeric centre are named 'glycosyl halides' such as the classical glycosyl donors. Acyl substituents on the anomeric hydroxyl group are designated as O-substitution just as with all other hydroxyl groups of the sugar ring.

Note, however, that anomeric *O*-alkyl (and aryl) substitution requires the carbohydrate to be named as a glycoside, in which the alkyl/aryl residue, -the aglycone portion-, precedes the glycoside's name as a separate word. Hence, glycosides are cyclic acetals in which the anomeric hydroxy group is replaced by an OR group. Glycosides are named as furanosides and pyranosides, depending on the ring size. When a heteroatom other than oxygen is attached to the anomeric centre, the terms *N*-glycosides and *S*- and *C*-glycosides are commonly used, though they are not strictly correct.

Monosaccharide substituents are given the suffix '-yl'. For example, in the disaccharide *N*-acetylglucosamine (LacNAc), the non-reducing galactose residue is called 'galactosyl' and thus the full name of LacNAc is β -D-galactopyranosyl-(1-4)-2-acetamido-2-deoxy- β -D-glucopyranose.

In addition, the glycone part of a glycoside can be named as 'glycosyloxy-' and used as prefix, such as in the plant toxin daphnin which is officially named a glucopyranosyloxy-coumarin as shown in Figure 1.25.

When sugars contain more than four stereocentres, more than one configurational prefix is required. For example, monosaccharides with more than six carbons—such as aldohexoses, aldooctoses or aldononoses—use multiple configurational prefixes in their systematic names. For assigning the configurational prefixes, the stereocentres are divided into groups beginning with a group of four sequential stereocentres nearest to C1. The 'rest' of the chain forms the second group of stereocentres, which may include one or up to four sequential stereocentres. Then, the two configurational prefixes are ordered such that the group of stereocentres (or just one stereocentre, respectively) farthest from C1 is placed first.

Each of the two groups of asymmetrically substituted atoms represented by a given prefix has its own configurational symbol and hence its own configurational reference atom. This reference atom is the highest-numbered stereocentre in the group and determines its *D*- or *L*-configuration.

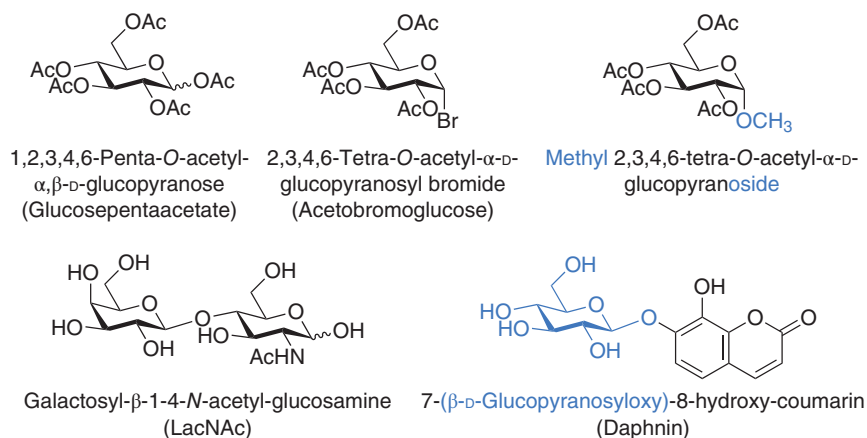


Figure 1.25 Examples of the nomenclature of substituted carbohydrates and simple glycosides.

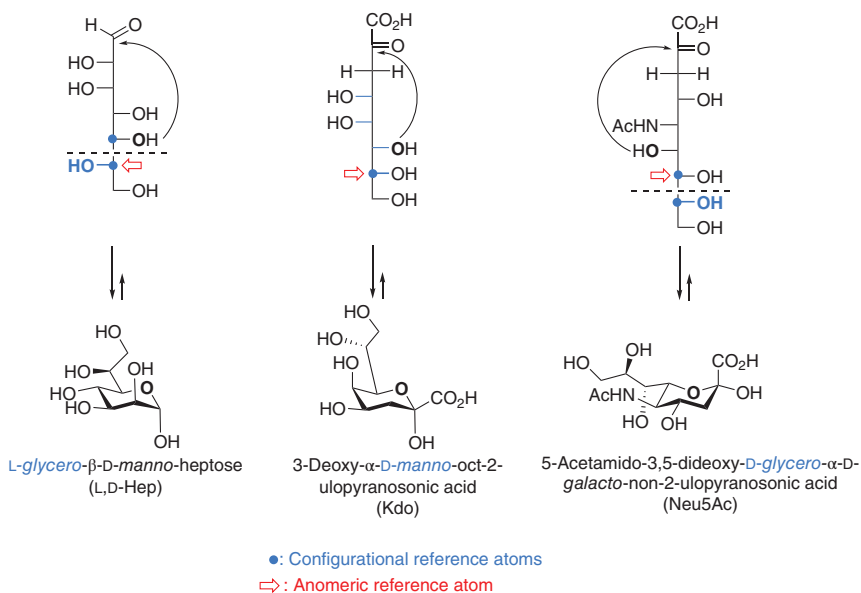
In cases where two configurational reference atoms must be considered for naming a monosaccharide, the anomeric reference atom cannot simply be equated with either of them. Instead, the anomeric reference atom is the stereocentre farthest from C1 *and* nearest to the hydroxy group involved in ring formation. As in pentoses and hexose, the position of the hydroxy group at *this* carbon atom in the Fischer projection determines the anomeric prefix, as explained earlier.

The anomeric symbol, α or β , is placed before the configurational prefix denoting the group of stereocentres that includes the hydroxyl group, which becomes the endocyclic oxygen upon cyclisation.

There are quite a number of cases where more than an aldehydic or carbonyl group is present in a monosaccharide, as in ketoaldonic acids. Names for ketoaldonic acids are formed by replacing the ending ‘-ulose’ of the corresponding ketose by ‘-ulosonic acid’, preceded by the location of the ketonic carbonyl group. Numbering starts at the carboxy group.

This may all seem somewhat esoteric, but correct nomenclature examples are surprisingly difficult to find in the literature for certain sugars. The examples provided here for biologically relevant sugars (L,D-Hep, Kdo and Neu5Ac) are therefore intended to be especially helpful for the reader. Important examples for naming ketoaldonic acids are Kdo and Neu5Ac, whose systematic IUPAC names are quite lengthy, as shown (Scheme 1.12).

Thus, naming of carbohydrates can become really sophisticated. In avoiding the complicated IUPAC nomenclature in such cases, trivial names are frequently used.



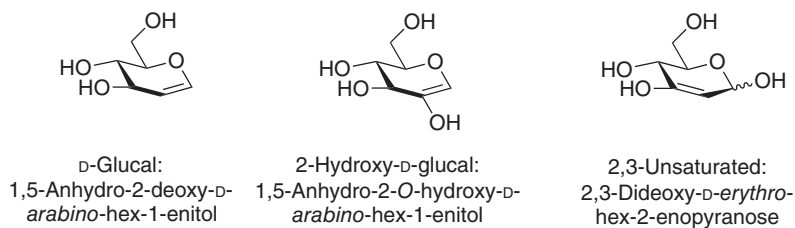


Figure 1.26 IUPAC nomenclature of unsaturated sugars.

One of the typical examples is the naming of unsaturated monosaccharides. The group of 1,2-unsaturated monosaccharides, commonly referred to as ‘glycals’, is named 1,5-anhydro-2-deoxy-1-enitols according to IUPAC, and a 2,3-unsaturated derivative is called a 2,3-dideoxy-2-eno-pyranose (Figure 1.26).

Furthermore, it is worthwhile to note that monosaccharides in which an alcoholic hydroxyl group has been replaced by an amino group are called ‘amino sugars’, and monosaccharides carrying the amino group at the anomeric centre are named ‘glycosyl amines’.

Monocarboxylic acids formally derived from aldoses by replacement of the aldehydic group by a carboxy group are called ‘aldonic acids’, while those derived from aldoses by replacement of the terminal CH_2OH group with a carboxy group are named ‘uronic acids’. The dicarboxylic acids formed from aldoses by replacement of both terminal groups (CHO and CH_2OH) with carboxy groups are called ‘aldaric acids’.

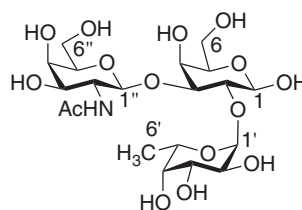
Nomenclature of Oligosaccharides

For the nomenclature of oligosaccharides, several systems are in use, ranging from traditional abbreviations to specialised codes. The most common are the three-letter monosaccharide codes (e.g., Glc, Man, Gal) with linkage positions and anomeric configuration indicated as in $\text{Gal}\beta 1\text{-4Glc}$ (lactose). For more complex glycans, shorthand notations such as the Symbol Nomenclature for Glycans (SNFG) or condensed text formats are used.

Branched oligosaccharides are named as glycosyl[glycosyl]glycosides. For complicated oligosaccharides, a linear backbone is identified and then the branches are added in parentheses right after the residue they are linked to, with multiple branches distinguished by different types of brackets. Glycosidic linkages can be indicated with an arrow, with or without a dash, but with no commas or spacing. The originating carbon number may be omitted since the starting point of a glycosidic linkage is always the anomeric centre: C1 in aldoses and C2 in ketosides.

The blood group A antigenic trisaccharide serves as a brief example. It is named as 2-acetamido-2-deoxy- α -D-galactopyranosyl-(1-3)-[α -L-fucopyranosyl-(1-2)]- β -D-galactopyranose, in short $\text{GalpNAc}(\alpha 1\text{-3})[\text{Fucp}(\alpha 1\text{-2})]\text{Gal}(\beta 1)$. Square brackets indicate an additional Fuc branch at the same acceptor Gal. The italic ‘p’ denotes

Figure 1.27 Structure of the blood group A antigenic trisaccharide.



the pyranose form; likewise, an italic '*f*' specifies the furanose form. These specifications are often omitted for practical reasons (Figure 1.27).

Each monosaccharide residue has its own numbering of carbons, starting at the anomeric centre (C1). The sugar ring at the reducing end receives unprimed numbers, the first branch residue receives primed numbers, the second branch residue receives double-primed numbers and so on.

Note that for bioinformatics and glycomics, machine-readable notations are required to handle large datasets and branched structures.²¹

Excursus: Carbohydrate NMR Spectroscopy

Due to their structural diversity and pronounced conformational flexibility, carbohydrates pose unique analytical challenges. To investigate their structural features, a wide range of complementary methods has been developed, with NMR spectroscopy playing a central role. In the following, the basic information obtainable from NMR spectra and the most commonly applied NMR techniques are highlighted and briefly discussed.

NMR spectroscopy of simpler carbohydrate derivatives is primarily concerned with the hydrogen atoms of the sugar ring(s). There are three fundamental sets of information which can be obtained from the ^1H NMR spectrum of organic compounds: (i) the chemical shift of the signals, δ ; (ii) their integration and (iii) the coupling constants J deduced from the detected multiplets.

The chemical shift of a particular ring hydrogen atom is strongly dependent on whether it is an anomeric or a non-anomeric proton and, furthermore, on the substitution pattern of the adjacent groups, especially of the functional group attached to the same carbon atom as the resonating proton. The anomeric proton is typically characterised by a pronounced downfield shift and usually appears as a doublet due to coupling with H2. It is often the most readily assigned hydrogen atom of the carbohydrate ring.

For the other ring protons, characteristic chemical shift differences are also observed in the ^1H NMR spectrum as a consequence of their different positions within the sugar ring. In *O*-peracetylated sugars, H2, H3 and H4 exhibit very similar chemical shifts, while the two H6 protons—commonly denoted H6 and H6' (or H6a and H6b)—are shifted upfield. H5 usually resonates at the smallest chemical shift (lowest ppm value) of all ring protons. The signals for H2, H3, H4, H6 and H6' typically appear as double doublets (dd), as each couples with

(Continued)

Excursus: (Continued)

two neighbouring protons (e.g. H2 couples with H1 and H3). The protons H2, H3 and H4 of a pyranose ring may also appear as pseudo-triplets when both coupling constants are nearly equal. H5 generally gives rise to a double doublet of doublets (ddd), coupling to H4, H6 and H6'.

O-Acetylation can be easily followed in the ^1H NMR spectrum because the acetyl- CH_3 protons of acetylated derivatives appear as singlets at around 2 ppm and are useful reporter groups in a molecule. A similarly useful landmark signal is the singlet obtained for the aglycone group of methyl glycosides, resonating around 3.5 ppm

Based on the different chemical shifts and observed coupling patterns of sugar protons, the hydrogen framework of a given compound can be reconstructed. In most cases, with the knowledge of the experimental details of the chemical reaction, the structure and conformation of the reaction product can be fully determined by ^1H NMR analysis. If this is not sufficient, a variety of additional NMR experiments are available.

Anomers can be clearly distinguished in the ^1H NMR spectrum since differences in the chemical shift of the anomeric H1 and the size of the coupling constant $^3J_{1,2}$ are significant. As a rule, an equatorial hydrogen substituent will give rise to a signal at lower field than an axial hydrogen when they are in a similar environment and the molecules are in the same conformation. Consequently, the H1 of an α -D-configured sugar derivative has higher ppm values than the H1 of the β -D-configured analogue, with chemical shift differences around 0.5 ppm

The difference in $^3J_{1,2}$ is caused by different dihedral angles ϕ between H1 and H2. The dihedral angle is 60° in the α -analogue and 180° in the β -analogue, respectively. The dependence of the size of coupling constants 3J (the superscript '3' indicating coupling over three bonds) on the dihedral angle ϕ of vicinal hydrogen atoms was first recognised by Karplus. According to Karplus, *trans*-diaxial hydrogens, with $\phi = 180^\circ$, have a large coupling constant 3J , whereas axial-equatorial and equatorial-equatorial orientations, both with $\phi = 60^\circ$, have much smaller coupling constants. Karplus identified the relationship between J and ϕ by showing that the values for J are approximately proportionate to ϕ and can be plotted against $\cos 2\phi$ to produce a characteristic curve.^{22,23} 3J couplings occurring in regular six-membered rings in chair conformation fit the Karplus curve well. However, the size of coupling constants is also dependent upon the presence of strongly electron-withdrawing groups. With increasing electronegativity of a substituent, a decrease in the value of the coupling constants is observed (Figure 1.28).

The Karplus relationship is an especially valuable tool for the determination of the anomeric configuration of a sugar derivative. Moreover, many synthetic derivatisations of the sugar ring include a change of the configuration at one or more stereocentres, and in such cases the 3J proton-proton coupling constants provide an excellent method to verify the outcome of epimerisation.

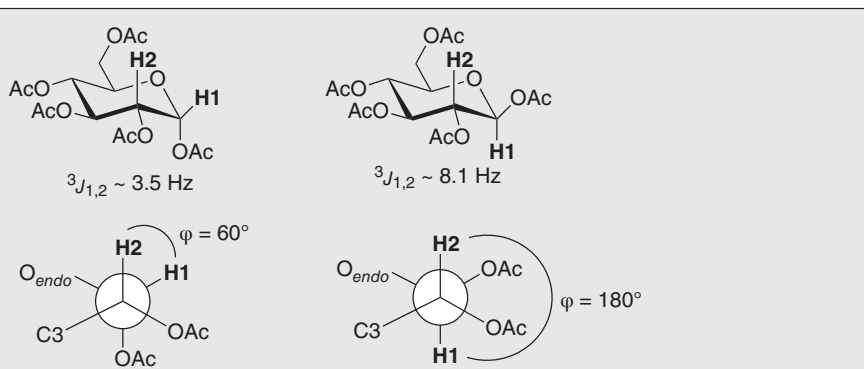


Figure 1.28 The anomeric forms of glucose pentaacetate can be unequivocally distinguished on the basis of their $^3J_{1,2}$ couplings constants, which differ owing to the different dihedral angles between H1 and H2.

In mannosides, carrying an axial 2-OH group, the size of $^3J_{1,2}$ is not indicative of the anomeric configuration. The coupling constant between H1 and H2 is small for both α - and β -mannosides because the H1-C1-C2-H2 dihedral angle is approximately 60° in either case. Therefore, to determine the anomeric configuration of mannosides by NMR, a different criterion must be used. This is provided by the one-bond heteronuclear constant $^1J_{C1,H1}$, which is sensitive to the torsion angle and reflects the anomeric configuration. Glycosides with different anomeric configurations typically show a difference of approximately 10 Hz in the size of $^1J_{C1,H1}$. The anomer with an equatorially disposed H1 has the larger coupling constant, generally around ~ 170 Hz in pyranosides, whereas anomers with an axially oriented H1 exhibit $^1J_{C1,H1}$ values of ~ 160 Hz. In mannosides, both of these values are often ~ 5 Hz smaller.

$^1J_{C1,H1}$ values can be determined in gated-decoupling ^{13}C NMR experiments, where heteronuclear C-H coupling is observable.²⁴ More recently, heteronuclear coupling constants are routinely measured using 2D NMR methods such as *J*-resolved HSQC (Heteronuclear Single Quantum Coherence), HSQC-TOCSY (Total Correlation Spectroscopy) or HMBC (Heteronuclear Multiple-Bond Correlation), where the $^1J_{C,H}$, $^2J_{C,H}$ and $^3J_{C,H}$ can be explicitly quantified.

^{13}C NMR measurements are an important supplement to the NMR analysis of an unknown compound. Compared to ^1H spectra, carbon signals in ^{13}C NMR are dispersed over a wider chemical shift range, making signal overlap less likely and allowing individual carbons, such as the anomeric carbon, to be assigned and linkages identified. However, due to the low natural abundance of ^{13}C , ^{13}C NMR is less sensitive than proton NMR and requires larger amounts of material. In standard ^{13}C NMR, spectra are usually recorded with broadband proton decoupling to increase sensitivity, but this comes at the cost of losing structural information from C-H coupling constants. Gated decoupling restores these couplings while still providing higher sensitivity than a fully proton-coupled ^{13}C spectrum.

(Continued)

Excursus: (Continued)

To determine the structure of more complicated carbohydrate derivatives, simple ^1H and ^{13}C NMR experiments may not be sufficient. Many sophisticated pulse sequences have been introduced, which record two-dimensional NMR spectra and can therefore provide more information than the one-dimensional spectra. These techniques facilitate the structural elucidation of very large and complicated molecules, such as complex oligosaccharides or glycoconjugates. The most common two-dimensional spectra are ^1H - ^1H -COSY (**Correlated Spectroscopy**) and ^1H - ^{13}C -COSY spectra such as HSQC and HMBC.²⁵

Carbohydrate-protein interactions can be studied using Saturation Transfer Difference (STD) NMR. In this experiment, protein resonances are selectively saturated, and the resulting spectrum is subtracted from a reference spectrum recorded without protein saturation. The difference spectrum displays only the signals of the ligand that received saturation transfer from the protein, allowing epitope mapping of the carbohydrate-protein interactions.²⁶

References

- 1 S.-I. Nakakita, J. Hirabayashi, "Transforming monosaccharides: recent advances in rare sugar production and future exploration" *BBA Adv.* 2025, 7, 100143.
- 2 F. W. Lichtenthaler, "Emil Fischer, his personality, his achievements, and his scientific progeny" *Eur. J. Org. Chem.* 2002, 4095–4122.
- 3 F. W. Lichtenthaler, "Emil Fischer's proof of the configuration of sugars: a centennial tribute" *Angew. Chem. Int. Ed.* 1992, 31, 1541.
- 4 Z. J. Witczak, "Monosaccharides: occurrence, significance, and properties" in *Glycoscience* (Eds: B. O. Fraser-Reid, K. Tatsuta, J. Thiem), Springer, Berlin, Heidelberg 2008, pp. 815–840.
- 5 R. U. Lemieux, "Effects of unshared pairs of electrons and their solvation on conformational equilibria" *Pure Appl. Chem.* 1971, 25, 527.
- 6 E. Matamoros, E. M. S. Pérez, M. E. Light, P. Cintas, R. F. Martínez, J. C. Palacios, "A true reverse anomeric effect does exist after all: a hydrogen bonding stereocontrolling effect in 2-iminoaldoses" *J. Org. Chem.* 2024, 89, 7877.
- 7 A. D. McNaught, "Nomenclature of carbohydrates (IUPAC recommendations 1996)" *Pure Appl. Chem.* 1996, 68, 1919.
- 8 D. Cremer, J. A. Pople, "General definition of ring puckering coordinates" *J. Am. Chem. Soc.* 1975, 97, 1354–1358.
- 9 H. B. Mayes, L. J. Broadbelt, G. T. Beckham, "How sugars pucker: electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for enzymatic catalysis" *J. Am. Chem. Soc.* 2014, 136, 1008.
- 10 D. Loru, C. Lütjohann, C. Näther, T. K. Lindhorst, M. Schnell, A. A. Auer, "Synergy of theory, NMR, and rotational spectroscopy to unravel structural details of D-altroside puckering and side chain orientation" *Chem. Eur. J.* 2025, 31, e02358.

- 11 E. Autieri, M. Sega, F. Pederiva, G. Guella, “Puckering free energy of pyranoses: a NMR and metadynamics-umbrella sampling investigation” *J. Chem. Phys.* 2010, 133, 095104; Erratum: E. Autieri, M. Sega, F. Pederiva, G. Guella, *J. Chem. Phys.* 2011, 134, 149901.
- 12 J. Basten, G. Jaurand, B. Olde-Hanter, P. Duchaussoy, M. Petitou, C. A. A. van Boeckel, “Biologically active heparin-like fragments with a “non-glycosamino”glycan structure. Part 3: O-alkylated-O-sulphated pentasaccharides” *Bioorg. Med. Chem. Lett.* 1992, 2, 905.
- 13 L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike, R. W. Carrell, “The anticoagulant activation of antithrombin by heparin” *Proc. Natl. Acad. Sci. USA* 1997, 94, 14683.
- 14 R. Stenutz, I. Carmichael, G. Widmalm, A. S. Serianni, “Hydroxymethyl group conformation in saccharides: structural dependencies of $^2J_{\text{HH}}$, $^3J_{\text{HH}}$, and $^1J_{\text{CH}}$ Spin-Spin coupling constants” *J. Org. Chem.* 2002, 67, 949.
- 15 R. D. Cummings, “A periodic table of monosaccharides” *Glycobiology* 2024, 34, cwad088.
- 16 B. Imperiali, “Bacterial carbohydrate diversity — a Brave New World” *Curr. Opin. Chem. Biol.* 2019, 53, 1.
- 17 P. H. Seeberger, D. B. Werz, “Automated synthesis of oligosaccharides as a basis for drug discovery” *Nat. Rev. Drug Discov.* 2005, 4, 751.
- 18 Y. Yu, M. Delbianco, “Conformational studies of oligosaccharides” *Chem. Eur. J.* 2020, 26, 9814.
- 19 IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN), “Nomenclature of Carbohydrates (Recommendations 1996)” can be found under <https://iupac.qmul.ac.uk/2carb/> (accessed 11 September 2025).
- 20 D. Horton (Ed.), “Nomenclature of Carbohydrates, (Recommendations 1996)” in *Advances in Carbohydrate Chemistry and Biochemistry*, Vol. 52, Academic Press, 1997, pp. 44–177, ISSN 0065-2318, ISBN 9780120072521. [https://doi.org/10.1016/S0065-2318\(08\)60090-6](https://doi.org/10.1016/S0065-2318(08)60090-6).
- 21 R. D. Cummings, J. M. Pierce, “The challenge and promise of glycomics” *Chem. Biol.* 2014, 21, 1–15.
- 22 M. Karplus, “Vicinal proton coupling in nuclear magnetic resonance” *J. Am. Chem. Soc.* 1963, 85, 2870.
- 23 K. Bock, C. Pedersen, “Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides” *Adv. Carbohydr. Chem. Biochem.* 1983, 41, 27.
- 24 C. Fontana, G. Widmalm, “Primary structure of Glycans by NMR spectroscopy” *Chem. Rev.* 2023, 123, 1040.
- 25 G. Widmalm, “Glycan shape, motions, and interactions explored by NMR spectroscopy” *JACS Au* 2024, 4, 20.
- 26 A. Gimeno, P. Valverde, A. Ardá, J. Jiménez-Barbero, “Glycan structures and their interactions with proteins. A NMR view” *Curr. Opin. Struct. Biol.* 2020, 62, 22.

In a Nutshell

- Carbohydrates are the most abundant natural products on Earth.
- They are polyhydroxycarbonyl compounds, classified as either aldoses or ketoses.
- Emil Fischer elucidated the stereochemical interrelationships of the D- and L-sugars, with optical rotation measurements playing a key role.
- The names given to aldoses and ketoses in the 'sugar trees' form the basis for configurational prefixes used in carbohydrate nomenclature.
- Carbohydrates exist as cyclic hemiacetals with C1 designated as anomeric centre. Its configuration is either α or β . In glycosides, the anomeric configuration is influenced by the anomeric effect.
- The *endo*-anomeric effect stabilises the axial anomer of a glycoside, while the *exo*-anomeric effect stabilises the anti-periplanar orientation of the exocyclic alkoxy group.
- Monosaccharide rings can be five-membered or six-membered and are then called furanoses or pyranoses, respectively.
- Glycosides are cyclic acetals. Glycosidic bonds link monosaccharides to form disaccharides, oligosaccharides and polysaccharides, which can be linear or branched. Oligosaccharides and polysaccharides differ in composition and the regio- and stereochemistry of their glycosidic linkages.
- Carbohydrates are frequently found covalently linked to other natural products. Many pharmacologically active drugs are glycosylated, and, most importantly, proteins and lipids of the cell are glycosylated, forming glycoproteins and glycolipids essential for cell biology.
- Sugar ring conformations are not all equally stable. The chair conformation is the most stable and most common one; but depending on the carbohydrate structure, other conformers may also be populated.
- Oligosaccharides adopt multiple conformations due to the flexibility around glycosidic linkages, and their 3D structures are essential for biological function.
- The structural diversity of carbohydrates is vast and varies across the kingdoms of life.

Key Terms

Master these terms, and you possess the essential concepts of this chapter.

Aldoses and ketoses, reducing sugar ☺ Fischer projection ☺ Newman projection ☺ Mills depiction ☺ Haworth representation ☺ chair conformation ☺ D,L-nomenclature, ☺ R,S-nomenclature ☺ CIP rules ☺ configurational reference atom ☺ stereo(genic) centre ☺ optical activity ☺ specific rotation value ☺ enantiomer ☺ diastereomer ☺ anomer ☺ racemate ☺ furanose and pyranose ☺ α - and β -anomeric configuration ☺ mutarotation ☺ glycoside ☺ glycone and aglycone ☺ anomeric centre ☺ anomeric reference atom ☺ configurational reference atom ☺ anomeric effect ☺ skew-boat conformation ☺ Cremer Pople sphere ☺ torsion angles ϕ , Ψ and ω ☺ cyanohydrin ☺ alditol ☺ uronic acid ☺ aldonic acids ☺ aldonic acids ☺ cyclodextrins, ☺ anhydro sugar ☺ cyclitols (inositols) ☺ amino sugar ☺ carba-sugar ☺ ketoaldonic acids ☺ glycoconjugates ☺ Glc, GlcNAc, Gal, GalNAc, Fuc, Xyl, Man, GlcA, Ido and Neu5Ac ☺ sialic acids.