1 | Structural features

As defined by the International Union of Pure and Applied Chemistry glycans are structures of multiple monosaccharides linked through glycosidic bonds. The terms sugar and saccharide are synonyms, depending on your preference for Arabic ("sukkar") or Greek ("sakkēaron"). Saccharide is the root for monosaccharides (a single carbohydrate unit), oligosaccharides (3 to 20 units) and polysaccharides (large polymers of more than 20 units).

Carbohydrates follow the basic formula $(CH_2O)N_{>2}$. Glycolaldehyde $(CH_2O)_2$ would be the simplest member of the family if molecules of two C-atoms were not excluded from the biochemical repertoire. Glycolaldehyde has been found in space in cosmic dust surrounding star-forming regions of the Milky Way galaxy. Glycolaldehyde is a precursor of several organic molecules. For example, reaction of glycolaldehyde with propenal, another interstellar molecule, yields ribose, a carbohydrate that is also the backbone of nucleic acids.
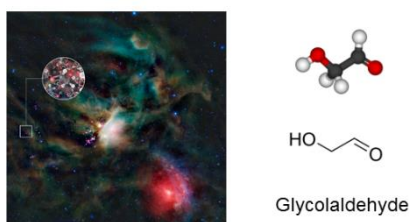

Glycolaldehyde

Figure 1 – The Rho Ophiuchi star-forming region is shown in infrared light as captured by NASA's Wide-field Infrared Explorer. Glycolaldehyde was identified in the gas surrounding the star-forming region IRAS 16293-2422, which is is the red object in the centre of the marked square. This star-forming region is 26'000 light-years away from Earth. Glycolaldehyde can react with propenal to form ribose. Image source: www.eso.org/public/images/eso1234a/

Beginning the count at three carbon atoms, glyceraldehyde and dihydroxyacetone share the common chemical formula $(CH_2O)_3$ and represent the smallest carbohydrates. As their names imply, glyceraldehyde has an aldehyde group (at C1) and dihydoxyacetone a carbonyl group (at C2). Carbohydrates with an aldehyde group at C1 like glucose are called aldoses and those with a carbonyl group

at C2 like fructose are called ketoses. Glyceraldehyde exists as two mirror image isomers (D- and L-stereoisomers), whereas dihydroxyacetone has no chiral carbon.
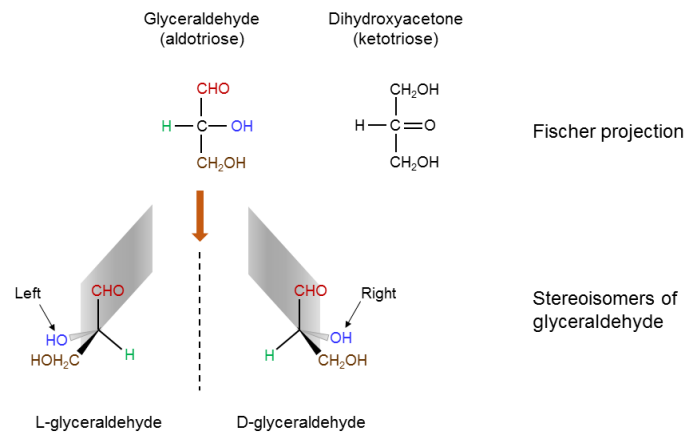

Figure 2 – Chirality of glyceraldehyde compared to the internal symmetry of the ketotriose dihydroxyacetone.

The distinction between D- and L-stereoisomers of carbohydrates is based on the orientation of the asymmetrical C-atom the farthest away from the carbonyl terminus (i.e. C5 in hexoses, see below). If the hydroxyl group at this asymmetrical center is oriented to the right (like D-glyceraldehyde in the Fischer projection), the carbohydrate becomes a D-stereoisomer. Most monosaccharides have at least one chiral C-atom with the total number (k) being equal to the number of internal CH_2O groups, i.e. n-2 for aldoses and n-3 for ketoses with n being the number of C-atoms in the monosaccharide. Thus, the possible number of stereoisomers in a monosaccharide is equal to 2^k. Accordingly, there are eight possible stereoisomers for hexoses.
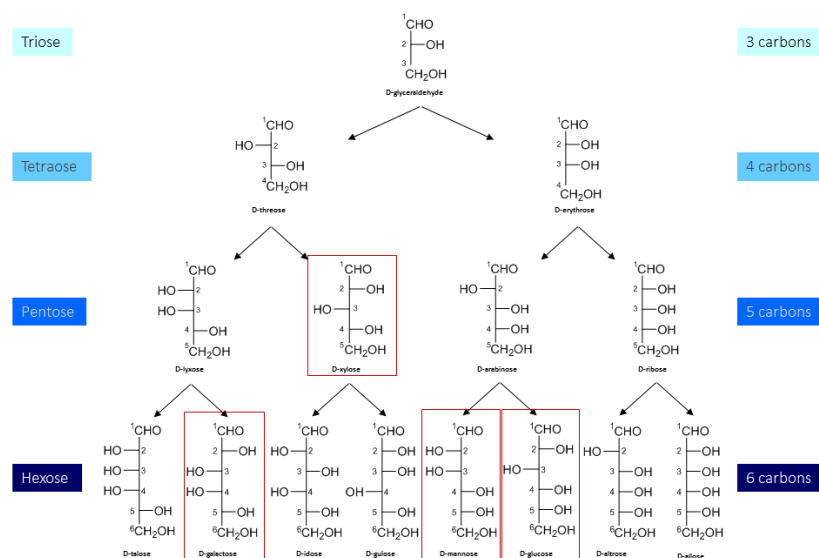

Figure 3 – Monosaccharides from triose to hexose. Monosaccharides found in mammalian glycans are outlined in red.

Two new stereoisomers are introduced with each new carbon center added. If this new center is virtually introduced between C1 (carbonyl group) and the previous C2 center, the stereoisomerism (D or L) of the new monosaccharide is maintained. Monosaccharides are grouped based on their number of C-atoms, thus yielding trioses, tetraoses, pentoses, hexoses, etc. Hexoses are the most abundant monosaccharides in biology, with glucose being at the top of the list. Two monosaccharides that differ only in the configuration of a single chiral C-atom are called epimers. Accordingly, galactose is an epimer of glucose.

Ring formation

Pentoses and hexoses do not react on tests for free aldehydes because they normally form cyclic structures. The intramolecular condensation between the aldehyde group at C1 and the hydroxyl group of in internal group (mainly C5 or less frequently C4) leads to a ring structure through the formation of a semialdehyde or cyclic hemiacetal. Ketohexoses also build ring structures by forming hemiketals.
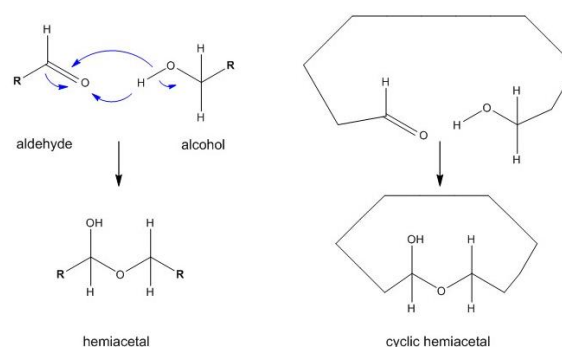

Figure 4 – Ring formation in carbohydrates by consensation of aldehyde and alcohol groups.

The formation of a ring structure adds a new level of asymmetry, thereby introducing a further isomerization. The new asymmetric carbon is called the anomeric center and α is used to denote the anomer where the absolute stereochemistry of the anomeric position and the most remote stereocenter in the sugar chain are the same. The anomer is called β when the stereochemistry is opposite. For D-monosaccharides, α and β anomers can be identified through the orientation of the hydroxyl group at C1, which is either axial (α) or equatorial (β).

Anomerism has a remarkable impact on the biological properties of carbohydrates. For example, α-D-mannose is perceived as sweet when tasted, whereas β-D-mannose is bitter (Steinhardt et al, 1962, PMID 14037773). Similarly, the α anomers of D-glucose and D-galactose are sweeter than their β anomers, indicating that taste receptors exhibit high stereoisomeric specificity. Carbohydrate anomerism is equally important in the formation of glycosidic linkages as it confers specific physical, chemical, and biological properties to the

ensuing products. For example starch and cellulose are both linear polymers of glucose, which only differ by the anomerism of the glycosidic linkage. Cellulose as a β1-4 polymer adopts a stiff linear conformation and forms a near crystalline structure, whereas starch as an α1-4 polymer arranges in a more relaxed structure. Animals can break down starch to Glc since they possess an α1-4 glucosidase activity (amylase) in the saliva and in the intestine. By contrast, animals lack β1-4 glucosidases and thus cannot digest cellulose. If humans could do so, a leaf of salad would provide more calories than a chocolate bar!
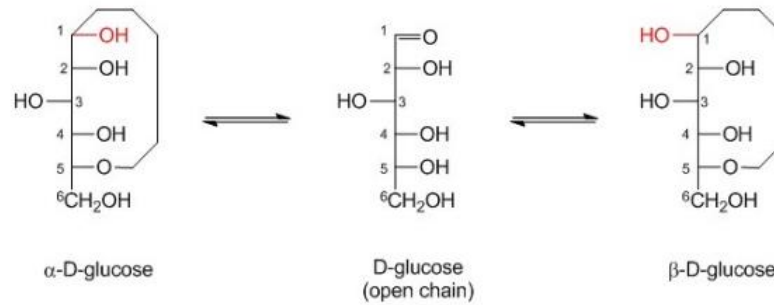

Figure 5 – Anomerism of D-glucose.

Most hexoses form a 6-atom ring, similar to a pyran ring, hence the term pyranose for such structures. Strictly speaking, the suffix *p* should be added to the abbreviation of the monosaccharide in question to designate this fact, i.e. Glc*p*. By analogy, the formation of a 5-atom ring yields furanose based on the 5-atom furan ring.
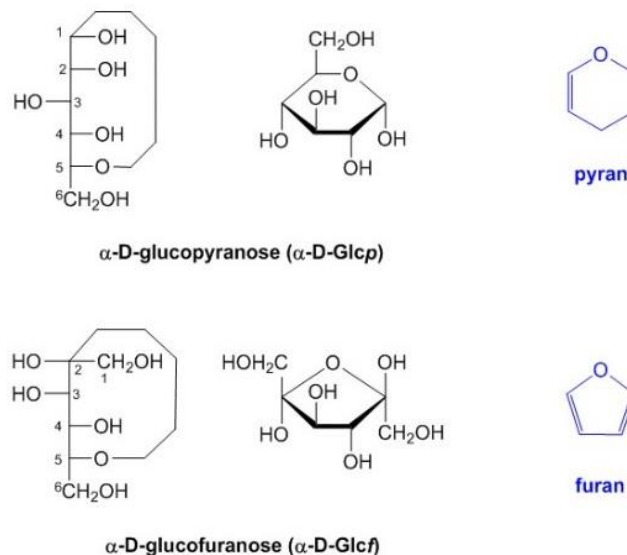

Figure 6 – Pyranose and furanose ring structures of α-D-glucose.

Since pentoses and hexoses form rings, the linear Fischer projection is best replaced by the Haworth representation, which also takes in account the anomerism of carbohydrates. In the conversion from Fischer to Haworth, the groups oriented to the right in Fischer representations are set downwards in

Haworth diagrams. The CH_2OH group at C6 is shown upwards. The Haworth representation suggests that monosaccharides are flat, but neither furanose nor pyranose rings are actually planar in their lowest energy confirmations. Pyranoses like the cyclohexane ring adopt a low-energy conformation that looks like a chair. The C2-C3-C5-O square forms a plane whereas the C1 and C4 atoms take up a higher or lower position, thus yielding the two possible chairs $^1C_4$ and $^4C_1$.
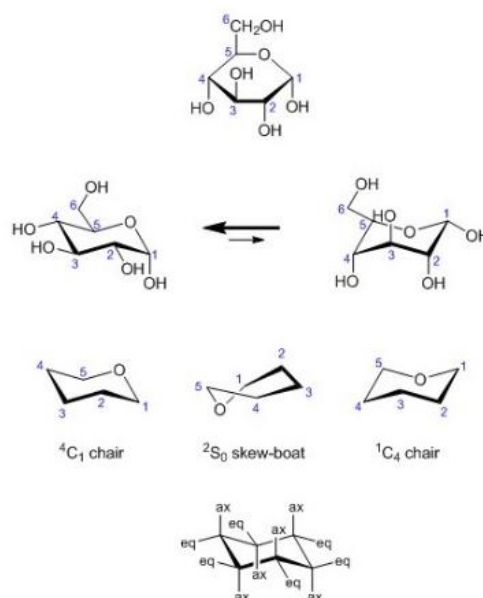

Figure 7 – Chair conformations of α-D-glucose.

Although D-glucose has a strong preference for the $^4C_1$ chair conformation, this is not true for all monosaccharides. For example, L-iduronic acid (an epimer of glucuronic acid) is more stable in the $^1C_4$ configuration. In fact, L-iduronic acid is a special case, because the $^1C_4$ configuration is in equilibrium with the so-called skew-boat shape (^2S_o), which favors the formation of hinges in linear glycan chains.

Monosaccharides in solution undergo spontaneous mutarotation that changes the axial (α) and equatorial (β) position of the hydroxyl group at C1. Mutarotation is catalyzed by mild acidic conditions. Although the linear aldehyde form is a negligible fraction of the monosaccharide structure (<0.01%), it represents the obligate intermediate in the mutarotation reaction. The aldehyde group of linear glucose can for example react with amino groups of amino acids like lysine and form an N-substituted glycosylamine. This reaction, which is called Maillard reaction, is the first step in the process of non-enzymatic glycation, which results in the formation of advanced glycation endproducts (AGE). Glycation is obviously not to be confused with glycosylation!

Whereas glucose and galactose are predominantly found in the β anomeric form, mannose is most frequently found as α anomer. Why is it so? Look at the orientation of the hydroxyl group at C2.
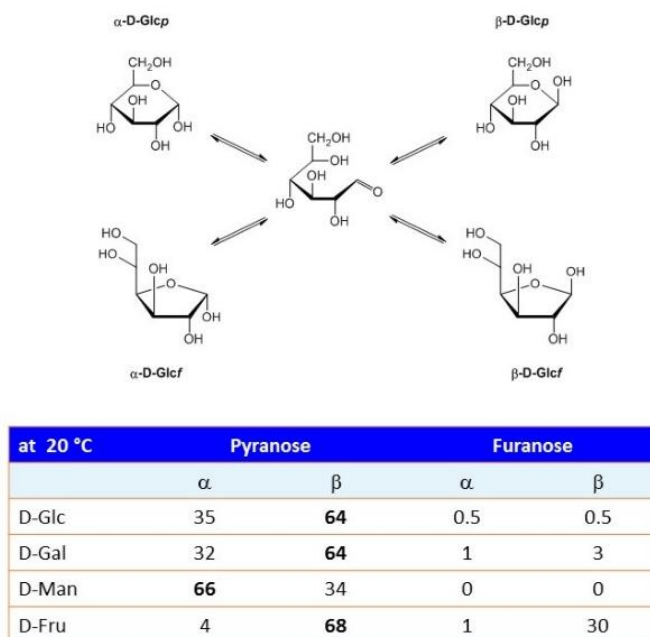
at 20 °C	Pyranose		Furanose	
	α	β	α	β
D-Glc	35	**64**	0.5	0.5
D-Gal	32	**64**	1	3
D-Man	**66**	34	0	0
D-Fru	4	**68**	1	30

Figure 8 – Mutarotation of D-glucose and prefered conformations of common monosaccharides in aqueous solution at 20 °C.

Monosaccharides in biology

Glycosylation is ubiquitous on earth as it occurs in all living organisms and even in viruses. The monosaccharide alphabets applied in the respective domains of life are though quite different from each other. Paradoxically, bacteria and archaea as the smallest organisms show the highest diversity with alphabets including close to hundred distinct monosaccharides. Among eukaryotes, plants demonstrate the broadest diversity with about 25 monosaccharides. By comparison, only 11 different monosaccharides are found in animals. These monosaccharides are glucose (Glc), galactose (Gal), mannose (Man), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), fucose (Fuc), xylose (Xyl), glucuronic acid (GlcA), iduronic acid (IdUA), N-acetylneuraminic acid (NeuAc), and N-glycolylneuraminic acid (NeuGc). NeuAc and NeuGc are members of the sialic acid family. Since NeuGc is the dominant form of sialic acid found in vertebrates, with the exception of primates and Old-World monkeys, the general term Sia is usually used to refer to this class of monosaccharides. More than 50 different types of Sia have been identified in living organisms.

Monosaccharides must be activated to act as substrates in glycosylation reactions. The biosynthesis of activated monosaccharides will be outlined in a subsequent chapter (see donor substrates). From the 11 monosaccharides found in animals, IdUA is the only one that is not transferred by itself to glycan chains, but first added as GlcA and then epimerized on glycans to IdUA.
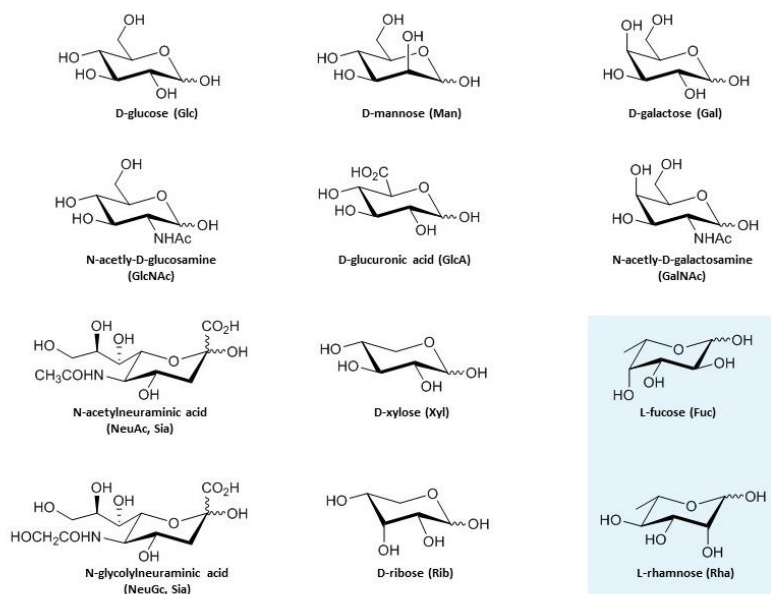
Figure 9 – Common monosaccharides. The deoxy-hexoses fucose and rhamnose mainly occur as L-stereoisomers. All monosaccharides shown here, with the exception fo rhamnose, are found in animals.

Glycosidic linkages

Monosaccharides are bound to each other and to amino acids and lipids via glycosidic linkages. The glycosidic linkage is defined as the link between the anomeric center of a monosaccharide and another residue. To describe the glycosidic linkage between two carbohydrates, there are basically two things to consider: a) the anomeric configuration of the linkage itself and b) the identity of the carbon on the next monosaccharide that shares the bridging oxygen. Carbohydrate chains are named by starting at the last non-reducing monosaccharide and progressing towards the reducing end. The "reducing carbohydrate" is the carbohydrate that still contains a free anomeric carbon. Between each unit, the glycosidic bond is abbreviated by indicating the anomerism and the respective number of the C-atoms involved in the linkage. Based on this system, the disaccharide maltose is written as D-Glc-α(1-4)-D-Glc.

There are multiple isomers of maltose since there are multiple ways to connect two Glc units. These consist of kojibiose (α1-2), nigerose (α1-3), isomaltose (α1-6), sophorose (β1-2), laminaribiose (β1-3), cellobiose (β1-4), and gentiobiose (β1-6). By contrast, disaccharides, like sucrose and trehalose, are connected through their reducing ends. These non-reducing disaccharides do not react positively in the Fehling test for free aldehydes. Note that the α or β configurations are no longer "free" in glycosidic linkages. The anomerism is in fact very important in conferring specific physical and chemical properties to oligosaccharides and polysaccharides, as outlined for starch and cellulose here above.
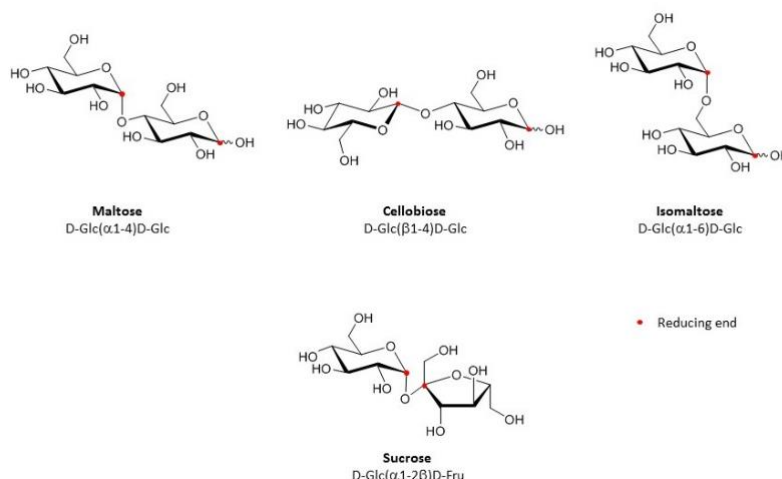
Figure 10 – Chair configurations of the three glucose disaccharides maltose, cellobiose, and isomaltose. In sucrose, glucose and fructore are linked through their reducing ends (marked with red dots).

Nomenclature

The huge structural diversity of glycans mainly derives from the multiple linkages and branching points between monosaccharide units. Unlike linear polymers such as nucleic acids and proteins, variably-linked and branched glycan sequences cannot be represented by lines of texts. A standard nomenclature systems based on graphical symbols has been devised by the Consortium for Functional Glycomics. A document describing this nomenclature is available here. To simplify the presentation and discussion of glycan structures, I will usually omit the specification of anomerism and exact linkage points.

Figure 11 – Standard nomenclature system for common monosaccharides based on symbols established by the Consortium for Functional Glycomics.

8

Classes of glycosylation

Classes of glycosylation are defined according to their respective invariant core structures and linkage to carrier molecules, which are amino acids and lipidic compounds. These invariant cores are further elongated, with the exception of C-linked Man and O-linked GlcNAc. The diversity of glycosylation classes encountered in prokaryotes is immense and probably only partially documented. Even in eukaryotes, some classes of glycosylation have only been discovered very recently and further classes likely remain to be described. In any case, the claim that 13 classes of glycosylation are found in vertebrates is certainly very close to the actual count. These 13 classes represent 7 major groups, featuring: N-linked glycans, O-linked glycans, C-linked mannose, glycosaminoglycans, the glycosylphosphatidylinositol anchor, glycosphingolipids, and cytoplasmic/nuclear O-GlcNAc.

In addition to the major classes mentioned here, further core structures are found in specific groups of organisms. For example, lipopolysaccharide, peptidoglycan, lipoarabinomannan are glycoconjugates encountered in bacteria, and Basidiolipids are glycolipids found in higher mushrooms. Some other glycans are not conjugated to lipids or proteins and therefore are referred to as oligosaccharides and polysaccharides. The first group include structures like colanic acid in Enterobacteriaceae and human milk oligosaccharides. Typical polysaccharides include hyaluronan, pectin, hemicellulose, and cellulose.
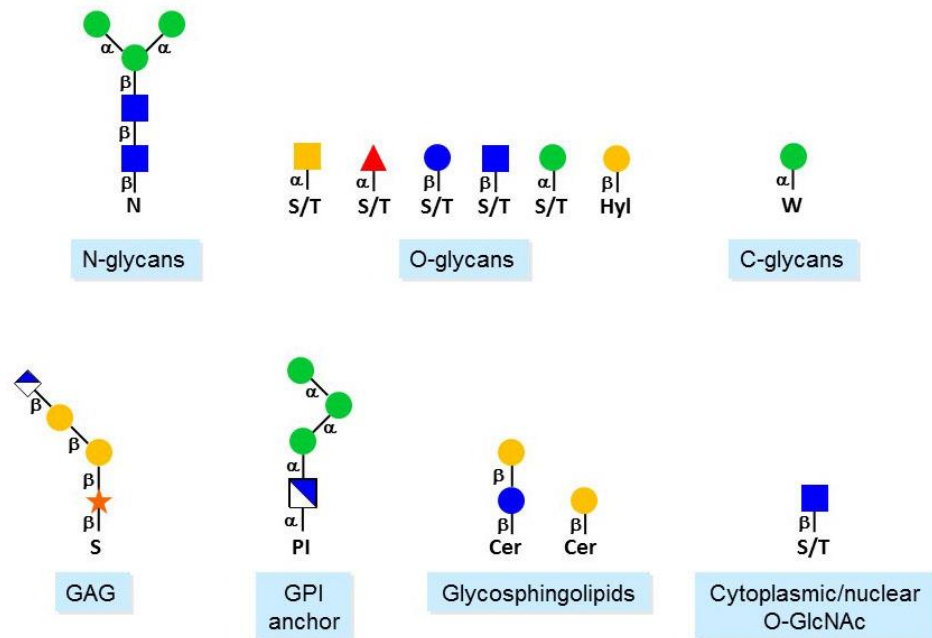

Figure 12 – Invariant core structures of the glycosylation classes found in vertebrates.