

Ajit Varki,¹ Hudson H. Freeze,² and Adriana E. Manzi¹

¹University of California San Diego, La Jolla, California

²Sanford Children's Health Research Center and Burnham Institute for Medical Research, La Jolla, California

ABSTRACT

Whereas DNA, RNA, and proteins are linear polymers that can usually be directly sequenced, glycans show substantially more complexity, having branching and anomeric configurations (α and β linkages). The biosynthesis of glycans, termed glycosylation, is extremely complex, is not template-driven, varies among different cell types, and cannot be easily predicted from simple rules. This overview discusses the stereochemistry of monosaccharides and glycans and provides diagrammatic representations of monosaccharides (Fisher projections and Haworth representations) and formulas for representation of glycan chains. A glossary of terms used in glycobiology is also provided. *Curr. Protoc. Protein Sci.* 57:12.1.1-12.1.10. © 2009 by John Wiley & Sons, Inc.

Keywords: glycan • carbohydrate • glycosylation • glycobiology • sugar • oligosaccharide

INTRODUCTION

The modern revolution in molecular biology was driven to a large extent by advances in methods for analysis and manipulation of DNA, RNA, and proteins. Although oligosaccharides (sugar chains or glycans) are also major macromolecules of the typical cell, they did not initially share in this molecular revolution. The reasons for this were to a large extent technical. Whereas DNA, RNA, and proteins are linear polymers that can usually be directly sequenced, oligosaccharides show substantially more complexity, having branching and anomeric configurations (α and β linkages). Thus, whereas three amino acids or nucleotides can be combined into six possible sequences, three hexose monosaccharides can theoretically generate 1056 possible glycans. In addition, the syntheses of DNA, RNA, and proteins are template-driven, and the sequence of one can generally be predicted from that of another. In contrast, the biosynthesis of oligosaccharides, termed glycosylation, is extremely complex, is not template-driven, varies among different cell types, and cannot be easily predicted from simple rules.

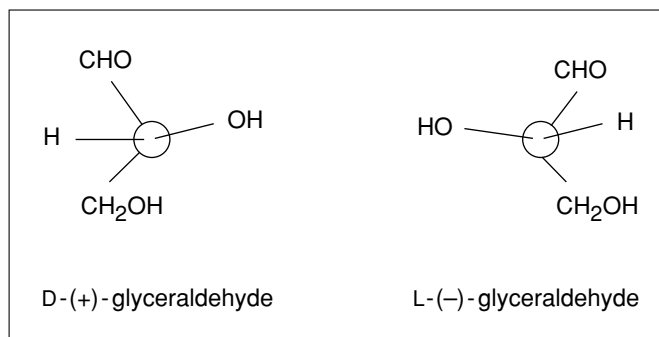
It is clear that glycans have important, albeit varied, effects upon the biosynthesis, folding, solubility, stability, subcellular trafficking, turnover, and half-life of the molecules to which they are attached. These are matters of great importance to the cell biologist, protein chemist, biotechnologist, and pharmacol-

ogist. On the other hand, the successful growth of several glycosylation mutants as permanent tissue culture cell lines indicates that the precise structure of many glycans is not critical for the growth and viability of a single cell in the protected environment of the culture dish. Thus, until recently, it was possible for many researchers working with in vitro single-cell systems to ignore the existence of glycans. However, with the increasing emphasis on studying cell-cell interactions in normal development, tissue morphogenesis, immune reactions, and pathological conditions such as cancer and inflammation, the study of glycan structure and biosynthesis has become very important.

The term "glycobiology" has found acceptance for denoting studies of the biology of glycoconjugates in both simple and complex systems. Many technical advances have occurred in the analysis of glycans, making it now feasible to study them in detail. These advances include the development of sensitive and specific assays and the availability of numerous purified enzymes (glycosidases) with high degrees of specificity.

In spite of all these advances, many analytical techniques in glycobiology remained in the domain of the few laboratories that specialized in the study of glycans. Likewise, published compendia of carbohydrate methods were designed mainly for use by experts. Chapter 12 attempts to place some of this

Figure 12.1.2 Enantiomers of glyceraldehyde.



only with the major forms of glycosylation found in “higher” animal glycoconjugates—*N*-acetylglucosamine (GlcNAc)-*N*-Asn-linked, *N*-acetylgalactosamine (GalNAc)-*O*-Ser/Thr-linked glycans on glycoproteins, xylose-*O*-Ser-linked glycosaminoglycans on proteoglycans, ceramide-linked glycosphingolipids, phosphatidylinositol-linked glycopospholipid anchors, and *O*-linked *N*-acetylglucosamine (GlcNAc-*O*-Ser). These structures are depicted in Figure 12.1.1. Other less common forms of glycosylation may need to be considered, especially if prior literature suggests their existence in a given situation. Examples of rarer sugar chains include (1) *O*-linked glucose, mannose, and fucose; (2) *N*-linked glucose; (3) glucosyl-hydroxylysine; and (4) GlcNAc-P-Ser (phosphoglycosylation).

Likewise, Chapter 12 deals only with the most common forms of shared outer sequences (e.g., sialylated and fucosylated lactosamines, polylactosamines, *O*-glycosaminoglycan chains, and blood group sequences), and not with rarer sequences (e.g., bisecting xylose residues, *N*-linked glycosaminoglycans, and β -linked GalNAc residues on *N*-linked glycans). For further information, the reader is directed to the Key References section at the end of this unit.

STEREOCHEMISTRY AND DIAGRAMMATIC REPRESENTATION OF MONOSACCHARIDES AND OLIGOSACCHARIDES

Basic Stereochemistry of Monosaccharides

In addition to this discussion, the reader is referred to Allen and Kisalius (1992) for a more detailed discussion of the principles of carbohydrate structure. The italicized terms in the discussion below are defined in the glossary at the end of this unit.

Glyceraldehyde, the simplest monosaccharide, has a single chiral (asymmetric) carbon (C-2). Therefore, it is a chiral molecule that shows optical isomerism; it can exist in the form of two nonsuperimposable mirror images called *enantiomers* (see Fig. 12.1.2). These enantiomers have identical physical properties, except for the direction of rotation of the plane of polarized light (–, left hand; +, right hand). Historically, the (+)-glyceraldehyde was arbitrarily assigned the prefix **D** (for dextrorotatory), and the (–)-glyceraldehyde, the prefix **L** (for levorotatory). The pair of enantiomers also have identical chemical properties, except toward optically active reagents. This fact is particularly important in biological systems, because most enzymes and the compounds they work on are optically active.

The configuration of the highest-numbered asymmetric carbon atom in the carbon chain of a higher monosaccharide is determined by comparison with the configuration of the chiral (or asymmetric) carbon of glyceraldehyde. Thus, a prefix **D**- is added to the name of each monosaccharide having the configuration of D-glyceraldehyde at the highest-numbered asymmetric carbon, and the prefix **L**- to those having the configuration of L-glyceraldehyde at the highest-numbered asymmetric carbon. The direction in which the compound rotates the plane of polarized light needs to be determined experimentally, and is indicated by a – or + sign, in parentheses, immediately after the prefix **D** or **L** [e.g., D-(+)-glucose]. In biological systems, where stereochemical specificity is the rule, **D** molecules may be completely active and **L** molecules completely inert, or vice versa. All known naturally occurring monosaccharides in animal cells are in the **D** configuration, except for fucose and iduronic acid, which are in the **L** configuration.

Each *aldose* sugar is usually present in a cyclic structure, because the *hemiacetal* produced by reaction of the aldehyde group at C-1 with the hydroxyl group at C-5 gives

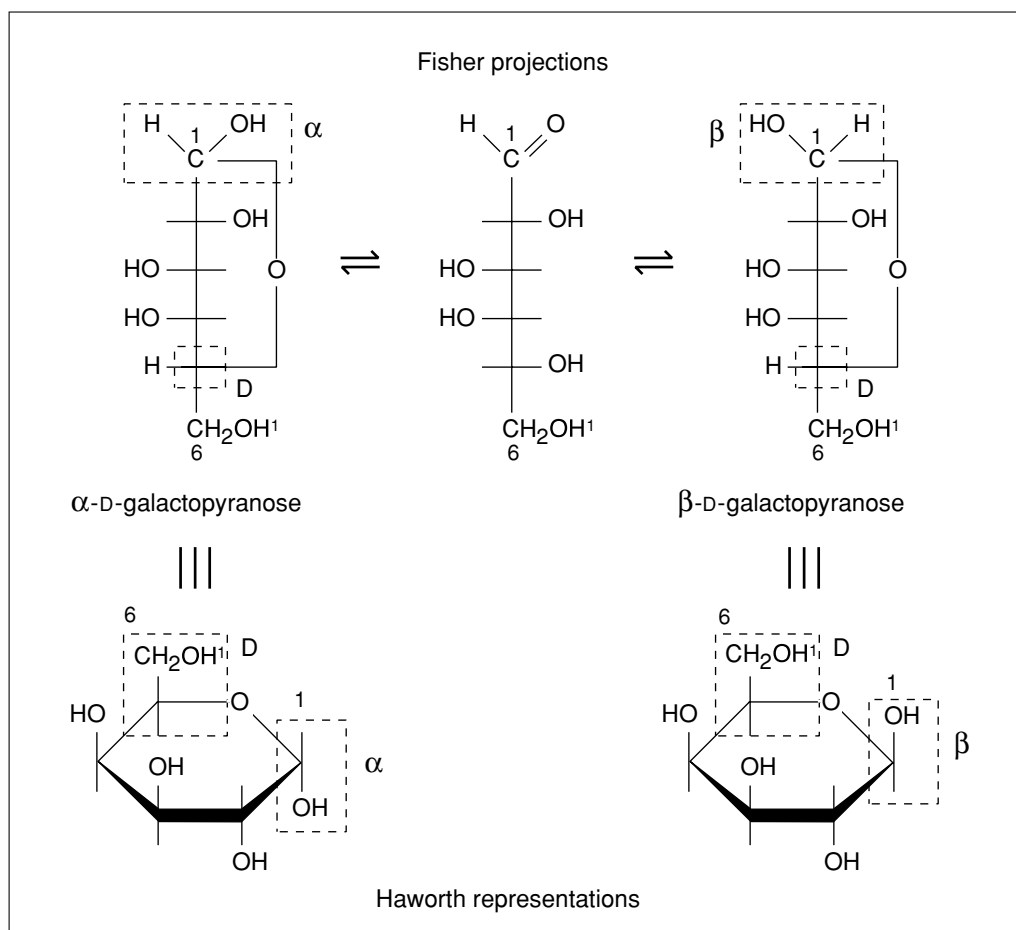


Figure 12.1.3 Formation of cyclic structures: open and ring forms of galactose.

a six-membered ring, called a *pyranose* (see Fig. 12.1.3). When five-membered rings are formed by reaction of the C-1 aldehyde with the C-4 hydroxyl, they are called *furanoses*. Hexoses commonly form pyranosidic rings, and pentoses form furanosidic rings, although the reverse is possible. Similarly, ketoses (e.g., sialic acids) can form *hemiketals* by reaction of the keto group at C-2 with the hydroxyl group at C-6.

These ring forms are the rule in intact oligosaccharides. The formation of the ring produces an additional chiral center at C-1 (or C-2 for keto sugars). Thus, two additional isomers are possible, which are called *anomers*. These anomers are designated α and β . Because glycosidic linkages (see above) occur via anomeric centers, they are called α linkages and β linkages.

Diagrammatic Representations of Monosaccharides

Monosaccharides are conventionally presented using the Fisher projection or the Haworth representation. Some examples are shown in Figure 12.1.4.

The Fisher projection

The carbon chain of a monosaccharide is written vertically with carbon atom number 1 at the top. The numbering of carbon atoms follows the rules of organic chemistry, so the aldehyde carbon is C-1. The horizontal lines represent bonds projecting out from the plane of the paper, whereas the vertical lines represent bonds projecting behind the plane of the paper. When the hydroxyl group at the highest-numbered asymmetric carbon is on the right, the monosaccharide belongs to the D series. When the hydroxyl group at the highest-numbered asymmetric carbon is on the left, the monosaccharide belongs to the L series.

The Haworth representation

The six-membered ring is represented with the oxygen at the upper right corner, approximately perpendicular to the plane of the paper, and with the groups attached to the carbons above or below the ring. All groups that appear to the right in a Fisher projection appear below the plane of the ring in a Haworth representation. The D-series hexoses have the carboxymethyl group (C-6) above the ring. The

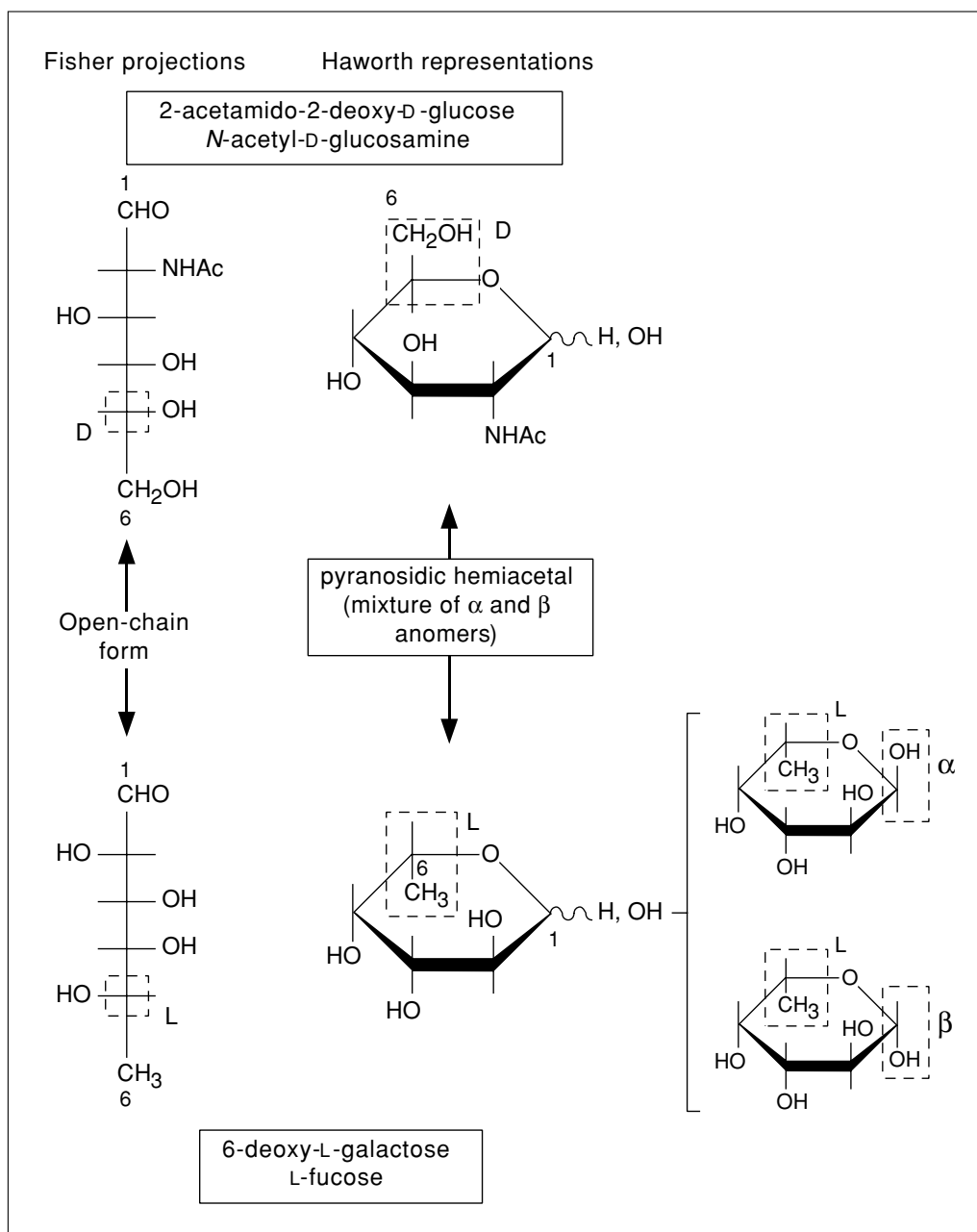


Figure 12.1.4 Fisher projections and Haworth representations of monosaccharides.

reverse is true for the L series. Therefore, the only difference between the α and β anomers of a given hexose would be the relative position of the hydroxyl and the hydrogen at C-1. In the α anomer, the hydrogen is below the ring for the D series (above the ring for the L series). In the β anomer, the hydrogen is above the ring for the D series (below it for the L series).

Conformation of Monosaccharides in Solution

The Haworth representation does not depict the real conformation of these molecules.

The preferred conformation in solution of the six-membered ring monosaccharides is what is known as a chair conformation. In this conformation, each different group can adopt either an *equatorial* or an *axial* position.

There are two possible chair conformations called *conformers* that exist in equilibrium. The position of this equilibrium differs from one monosaccharide to another depending on the relative positions of hydroxyl groups or other substituents. The preferred conformation is that with the lowest number of bulky groups in an axial position (Fig. 12.1.5).

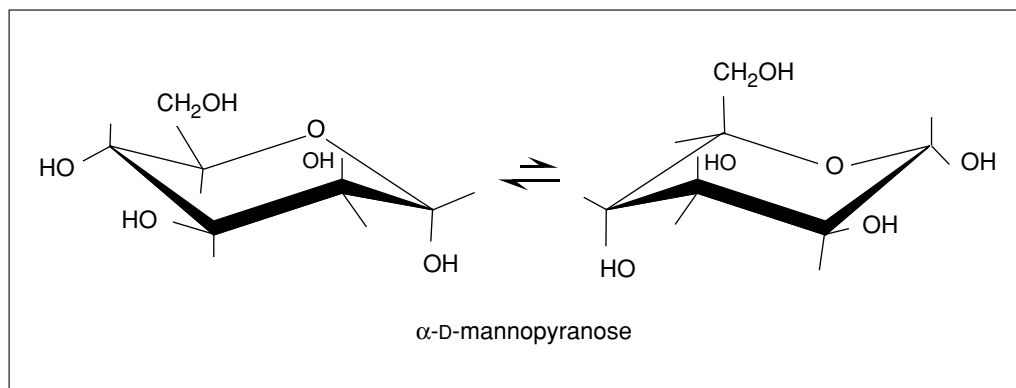


Figure 12.1.5 D-Mannose chair conformation.

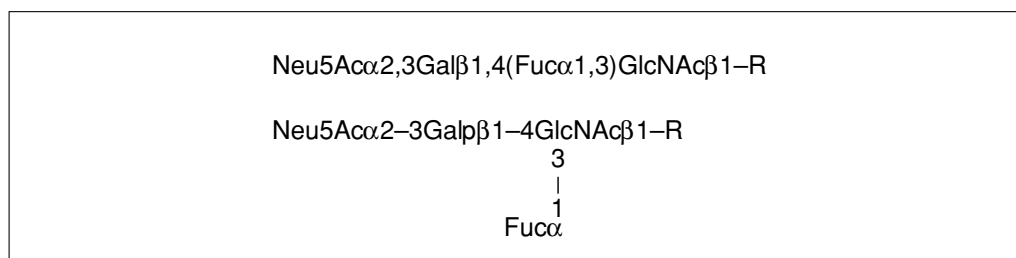
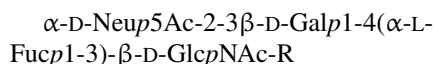


Figure 12.1.6 Two common short forms for denoting oligosaccharide structures.

Figure 12.1.7 (on following page) Recommended symbols and conventions for drawing glycan structures. **(A)** Each monosaccharide class (e.g., hexose) has the same shape, and isomers are differentiated by color/black/white/shading. The same shading/color is used for different monosaccharides of the same stereochemical designation, e.g., Gal, GalNAc, and GalA. To minimize variations, sialic acids and uronic acids are in the same shape, and only the major uronic and sialic acid types are represented. When the type of sialic acid is uncertain, the abbreviation Sia can be used instead. Only common monosaccharides in vertebrate systems are assigned specific symbols. All other monosaccharides are represented by an open hexagon or defined in the figure legend. If there is more than one type of undesigned monosaccharide in a figure, a letter designation can be included to differentiate between them. Unless otherwise indicated, all of these vertebrate monosaccharides are assumed to be in the D configuration (except for fucose and iduronic acid, which are in the L configuration), all glycosidically linked monosaccharides are assumed to be in the pyranose form, and all glycosidic linkages are assumed to originate from the 1-position (except for the sialic acids, which are linked from the 2-position). Anomeric notation and destination linkages can be indicated without spacing/dashes. Although color is useful, these representations will survive black-and-white printing or photocopying with the colors represented in different shades (the color values in the figure are the RGB color settings obtained within PowerPoint). Modifications of monosaccharides are indicated by lowercase letters, with numbers indicating linkage positions, if known (e.g., 9Ac for the 9-O-acetyl group, 3S for the 3-O-sulfate group, 6P for a 6-O-phosphate group, 8Me for the 8-O-methyl group, 9Acy for the 9-O-acyl group, and 9Lt for the 9-O-lactyl group). Esters and ethers are shown attached to the symbol with a number. For N-substituted groups, it is assumed that only one amino group is on the monosaccharide with an already known position (e.g., NS for an N-sulfate group on glucosamine, assumed to be at the 2-position). **(B)** Typical branched “biantennary” N-glycan with two types of outer termini, depicted at different levels of structural details. **(C)** Some typical glycosaminoglycan (GAG) chains (reproduced with permission from figure 1.5 in Varki et al., 2009).

Formulas for Representation of Oligosaccharide Chains

Full and correct representation of the formula of an oligosaccharide chain requires notation describing the complete stereochemistry of the component monosaccharides. Thus, the simple tetrasaccharide commonly called sialyl-Lewis^x (attached to an underlying oligosaccharide, R) would be written as follows:



In more common practice, the D and L configurations and the nature of ring structures are assumed, and the formula is written in one of the ways shown in Figure 12.1.6. There is also increasing use of a simplified symbol set to conveniently represent the common monosaccharides of vertebrate glycans (Fig. 12.1.7).

Higher-Order Structures in Glycans

The presence of monosaccharides with different types of glycosidic linkages in oligo- and polysaccharides creates further complexity, causing the development of secondary and tertiary structures in these molecules. The approximate shape adopted in solution by a given carbohydrate chain can be predicted according to the type of glycosidic linkages involved. These higher-order structures may be critical for the biological roles of glycoconjugates.

GLOSSARY

The following definitions are presented for the terms most commonly used throughout Chapter 12.

aldose: monosaccharide with a carbonyl group at the end of the carbon chain (aldehyde group); the carbonyl group is assigned the lowest possible number (i.e., carbon 1).

carbohydrates: polyhydroxyaldehydes or polyhydroxyketones, or compounds that can be hydrolyzed to them.

diastereoisomers: compounds with identical formulas that have a different spatial distribution of atoms (e.g., galactose and mannose).

enantiomers: nonsuperimposable mirror images of any compound (e.g., D- and L-glucose).

epimers: two monosaccharides differing only in the configuration of a single chiral carbon.

Fuc (L-fucose): type of deoxyhexose (see also *types of monosaccharides* below).

Gal (D-galactose): type of hexose (see also *types of monosaccharides* below).

GalNAc (N-acetyl-D-galactosamine): type of hexosamine (see also *types of monosaccharides* below).

ganglioside: anionic glycolipid containing one or more units of sialic acid.

Glc (D-glucose): type of hexose (see also *types of monosaccharides* below).

GlcNAc (N-acetyl-D-glucosamine): type of hexosamine (see also *types of monosaccharides* below).

glyceraldehyde: simplest monosaccharide (an aldotriose; i.e., containing three carbon atoms).

glycoconjugate: natural compound in which one or more monosaccharide or oligosaccharide units are covalently linked to a noncarbohydrate moiety.

GIUA or GlA (D-glucuronic acid): type of uronic acid (see also *types of monosaccharides* below).

glycolipid or glycosphingolipid: glycan attached via glucose or galactose to the terminal primary hydroxyl group of ceramide, which is composed of a long-chain base (i.e., sphingosine) and a fatty acid. Glycolipids can be neutral or anionic (negatively charged).

glycophospholipid anchor: glycan bridge between phosphatidylinositol and a phosphoethanolamine in amide linkage to the C terminus of a protein; constitutes the sole membrane anchor for such proteins.

glycoprotein: glycoconjugate in which a protein carries one or more glycan chains covalently attached to a polypeptide backbone via N-GlcNAc- or O-GalNAc-linkages.

glycosaminoglycan: linear copolymers of disaccharide repeating units, each composed of a hexosamine and a hexose or hexuronic acid; these are the glycan chains that define the proteoglycans. The type of disaccharide unit can define the glycosaminoglycans as chondroitin or dermatan sulfate, heparan or heparin sulfate, hyaluronic acid, and keratan sulfate. The glycosaminoglycans (except hyaluronic acid) also contain sulfate esters substituting either hydroxyl or amino groups (N- or O-sulfate groups).

glycosidic linkage: linkage of a monosaccharide to another residue via the anomeric hydroxyl group.

IdUA or IdA (L-iduronic acid): type of uronic acid (see also *types of monosaccharides* below).

ketose: monosaccharide with a carbonyl group in an inner carbon; the carbonyl group is assigned the lowest possible number (i.e., carbon 2).

Man (D-mannose): type of hexose (see also *types of monosaccharides* below).

monosaccharide: carbohydrate that cannot be hydrolyzed into simpler units (see also *types of monosaccharides* below).

mucin: large glycoproteins that contain many (up to several hundred) *O*-GalNAc-linked glycan chains that are often closely spaced.

N-linked oligosaccharide: glycan covalently linked to an asparagine residue of a polypeptide chain in the consensus sequence -Asn-X-Ser/Thr. Although there are many different kinds of N-linked glycans, they have certain common features: (1) a common core pentasaccharide containing two mannosyl residues α -linked to a third mannosyl unit, which is in turn β -linked to a chitobiosyl group and (2) a chitobiosyl group that is β -linked to the asparagine amide nitrogen. The N-linked glycans can be divided into three main classes: high-mannose type, complex type, and hybrid type.

Neu5Ac (N-acetyl-D-neuraminic acid): type of sialic acid (Sia; see also *types of monosaccharides* below).

O-linked oligosaccharide: glycan linked to the polypeptide via *N*-acetylgalactosamine (GalNAc) to serine or threonine. Note that other types of O-linked glycans also exist (e.g., *O*-GlcNAc) but that the *O*-GalNAc linkage, being the most well known, is often described by this generic term.

oligosaccharide: branched or linear chain of monosaccharides attached to one another via glycosidic linkages (see also *polysaccharides* below).

polylactosaminoglycan: long chain of repeating units of the disaccharide β -Gal(1-4)-GlcNAc. These chains may be modified by sialylation, fucosylation, or branching. When sulfated, they are called keratan sulfate (see *glycosaminoglycans* above).

polysaccharides: branched or linear chain of monosaccharides attached to each other via glycosidic linkages that usually contain repetitive sequences. The number of monosaccharide residues that represents the limit between an oligosaccharide and a polysaccharide is not defined. A tetraantennary carbohydrate moiety containing eighteen monosaccharide units, present in a glycoprotein, is considered an oligosaccharide. On the other hand, a carbohydrate composed of eighteen glucose residues linked β -(1-4) present in a plant extract is considered a polysaccharide.

polysialic acid: homopolymer of sialic acid selectively expressed on a few vertebrate proteins and on the capsular polysaccharides of certain pathogenic bacteria.

proteoglycan: glycoconjugate having one or more *O*-xylose-linked glycosaminoglycan chains (rather than *N*-GlcNAc- or *O*-GalNAc-linked glycans) linked to protein. The distinction from a glycoprotein is otherwise arbitrary, because some proteoglycans can have both glycosaminoglycan chains and N- or O-linked oligosaccharides attached to them.

reducing sugar: sugar that undergoes typical reactions of aldehydes (e.g., is able to reduce Ag^+ or Cu^{++}). Mono-, oligo-, or polysaccharides can be reducing sugars when the aldehyde group in the terminal monosaccharide residue is not involved in a glycosidic linkage.

saccharide modifications: hydroxyl groups of different monosaccharides can be subject to phosphorylation, acetylation, sulfation, methylation, or fatty acylation. Amino groups can be free, *N*-acetylated, or *N*-sulfated. Carboxyl groups are occasionally subject to lactonization to nearby hydroxyl groups.

Sia (sialic acid): generic name for a family of acidic nine-carbon monosaccharides (e.g., *N*-acetylneuraminic acid, Neu5Ac; and *N*-glycolylneuraminic acid, Neu5Gc).

types of monosaccharides: monosaccharides may have a carbonyl group at the end of the carbon chain (aldehyde group) or in an inner carbon (ketone group). The carbonyl group is assigned the lowest possible number, e.g., carbon 1 (C-1) for the aldehyde group; carbon 2 (C-2) for the most common ketone groups. These two types are named aldoses and ketoses accordingly. The simplest monosaccharide is glyceraldehyde (see Fig. 12.1.2), an aldotriose (i.e., containing three carbon atoms). Natural aldoses with different number of carbon atoms in their chain are named accordingly (e.g., aldohexoses, containing six carbon atoms). Two monosaccharides differing only in the configuration of a single chiral carbon are called epimers. For example, glucose and galactose are epimers of each other at C-4. The common monosaccharides present in animal glycoconjugates are (1) deoxyhexoses (e.g., L-Fuc); (2) hexosamines, usually *N*-acetylated (e.g., D-GalNAc and D-GlcNAc); (3) hexoses (e.g., D-Glc, D-Gal, and D-Man); (4) pentoses (e.g., D-Xyl); (5) sialic acids (Sia; e.g., Neu5Ac); and (6) uronic acids (e.g., D-GlA and L-IdA).

Xyl (D-xylose): type of pentose (see also *types of monosaccharides* above).

For further detailed information on the analysis of glycans, the reader is referred to the specific literature cited in the individual protocol units. In addition, the following sources can be used for general information, for details on specific methods, and for many additional methods that are not included in Chapter 12.

LITERATURE CITED

Allen, H.J. and Kisailus, E.C. (eds.) 1992. Glycoconjugates: Composition, Structure, and Function. Marcel Dekker, New York.

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., and Etzler, M.E. (eds.) 2009. Essentials of Glycobiology. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

KEY REFERENCES

Allen and Kisailus, 1992. See above.

Bertozzi, C. and Rabuka, D. 2009. Structural basis of glycan diversity. *In* Essentials of Glycobiology (A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart and M.E. Etzler, eds.) pp. 601-616. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Bill, M.R., Revers, L., and Wilson, I.B.H. 1998. Protein Glycosylation. Kluwer Academic Publishers, Boston.

Boons, G.-J. (ed.) 1998. Carbohydrate Chemistry. Blackie Academic & Professional, London.

Chaplin, M.F. and Kennedy, J.F. (eds.) 1987. Carbohydrate Analysis: A Practical Approach. IRL Press, Oxford.

Fleischer, S. and Fleischer, B. (eds.) 1983. Biomembranes, part L. *Methods Enzymol.* 98:1-688.

Ginsburg, V. (ed.). 1972. Complex Carbohydrates, Part B. *Methods Enzymol.* Vol. 28.

Ginsburg, V. (ed.). 1978. Complex Carbohydrates, Part C. *Methods Enzymol.* Vol. 50.

Ginsburg, V. (ed.). 1982. Complex Carbohydrates, Part D. *Methods Enzymol.* Vol. 83.

Ginsburg, V. (ed.). 1987. Complex Carbohydrates, Part E. *Methods Enzymol.* Vol. 138.

Ginsburg, V. (ed.). 1989. Complex Carbohydrates, Part F. *Methods Enzymol.* Vol. 179.

Hart, G. and Lennarz, W. 1993. Guide to techniques in glycobiology. *Methods Enzymol.* Vol. 230.

Haslam, S.M., North, S.J., and Dell, A. 2006. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr. Opin. Struct. Biol.* 16:584-591.

Kamerling, J., Boons, G.-J., Lee, Y., Suzuki, A., Taniguchi, N., and Voragen, A.G.J. 2007. Comprehensive glycoscience. 1-4. Elsevier Science, London.

Manzi, A.E. and van Halbeek, H. 1999. Glycan structure and nomenclature. *In* Essentials of Glycobiology (A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, and J. Marth, eds.) pp. 17-29. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Manzi, A.E. and van Halbeek, H. General principles for the analysis and sequencing of glycans. *In* Essentials of Glycobiology (A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, and J. Marth, eds.) pp. 581-598. 1999. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

McNaught, A.D. 1997. Nomenclature of carbohydrates. *Carbohydr. Res.* 297:1-92.

McCloskey, J.A. (ed.) 1990. Mass spectrometry. *Methods Enzymol.* Vol. 193.

Stick, R.V. 2001. Carbohydrates: The Sweet Molecules of Life. Academic Press, New York.

Varki, A. 1991. Radioactive tracer techniques in the sequencing of glycoprotein oligosaccharides. *FASEB J.* 5:226-235.

Vliegthart, J.F.G., Dorland, L., and van Halbeek, H. 1983. NMR spectroscopy of carbohydrates. *Adv. Carbohydr. Chem. Biochem.* 41:209-378.

Whistler, R. and Wolfram, F. (eds.) 1968-1980. Methods in Carbohydrate Chemistry, Vols I-VIII. Academic Press, San Diego.