1

MAMMALIAN GLYCAN BIOSYNTHESIS: BUILDING A TEMPLATE FOR BIOLOGICAL RECOGNITION

M. Adam Meledeo, Venkata D. P. Paruchuri, Jian Du, Zhiyun Wang, and Kevin J. Yarema

Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, Maryland

Carbohydrate Recognition: Biological Problems, Methods, and Applications, First Edition.
Edited by Binghe Wang and Geert-Jan Boons

1.1. INTRODUCTION AND OUTLINE

The surfaces of mammalian cells are dominated by the glycocalyx, a layer of carbohydrates of incredible complexity that in large part determines how a cell interacts with its surroundings (Fig. 1.1). Glycans influence a cell at several hierarchal levels ranging from molecular and subcellular functions to the interactions a cell makes with its environment on micro-, systemic, and—referring to the ever-topical threat of influenza pandemic—sometimes quite literally, global scales. On the molecular level, the simplest manifestation of glycosylation is the *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) modification of hundreds, and most likely thousands, of nuclear and cytosolic proteins. At a higher level of complexity, the $GlcNAc_2Man_9Glc_3$ *N*-glycan structure participates in quality control during the folding of membrane-displayed and secreted glycoproteins, and once these and other classes of glycans reach the cell surface they regulate both nano- and microscale properties of the plasma membrane. An example of the former is the "glycosynapse," while one of the latter is the galectin lattice specified by *N*-glycan branching status, as discussed in more detail in Chapter 7.

Specific surface glycans, or the collective cellular complement of these molecules, reflect the internal workings and status of a cell and thus serve as accessible biomarkers for, among other conditions, cancer (Chapter 6) and stem cell status. The term "biomarker" endows these glycans with a seeming passivity that belies the numerous roles

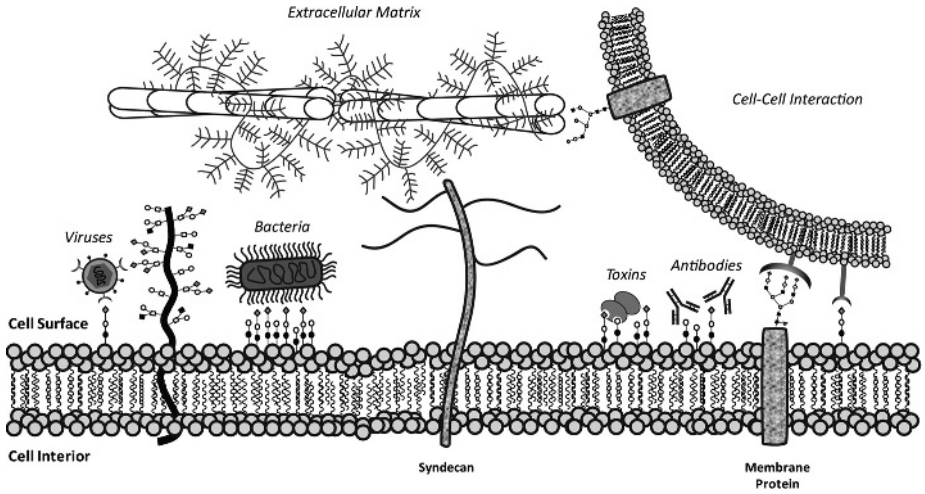

Figure 1.1. Landscape of the mammalian cell surface. A schematic view of the cell surface illustrates the breadth of biological recognition events facilitated by glycans.

and vibrant functions of cellular carbohydrates within cells (e.g., *O*-GlcNAc in signaling and *N*-glycans in protein folding) as well as in the interactions they make with their surroundings. Glycan-mediated interactions that a cell makes with its environment begin on an intimate scale as these molecules modulate adhesion to neighboring cells and the ECM (Chapter 2). Moving to the systemic and organism-wide levels, glycans are the veritable workhorses of multicellular life, allowing the complex mix of cell types to flourish in association with each other. Their contributions begin with fertilization at the very start of life (Chapter 5) and extend to choreographing the activity of certain hormones, including growth hormone during development and in the adult, as well as orchestrating the immune system (Chapters 2 and 8).

For years following pioneering images made of the glycocalyx (e.g., electron micrographs from 1965),[1] which appeared as a disordered "fuzzy halo" surrounding cells, it was baffling how so many finely tuned biological functions emanated from glycans. But slowly, culminating with an accelerating pace of discovery in the past decade, a spate of sophisticated tools has revealed that instead of being utterly disorderly and unruly, the glycocalyx can in fact be highly structured with regular features.[2] In fact, it is now well appreciated that the glycocalyx consists of cell-, tissue-, and species-specific repertoires of specific glycan structures that comprise a sophisticated "sugar code"[3] that can be evoked to explain both the myriad synergistic as well as the often counteracting and contradictory functions of these molecules.

The impetus behind this introductory chapter is the hope that that investigator or clinician who desires to exploit carbohydrate-based biological recognition events for research or medical purposes will find it instructive to have a basic knowledge of the biosynthetic processes by which a cell creates glycans. To provide slightly more context, Figure 1.2 illustrates how a pathogen—exemplified by the influenza virus that is well

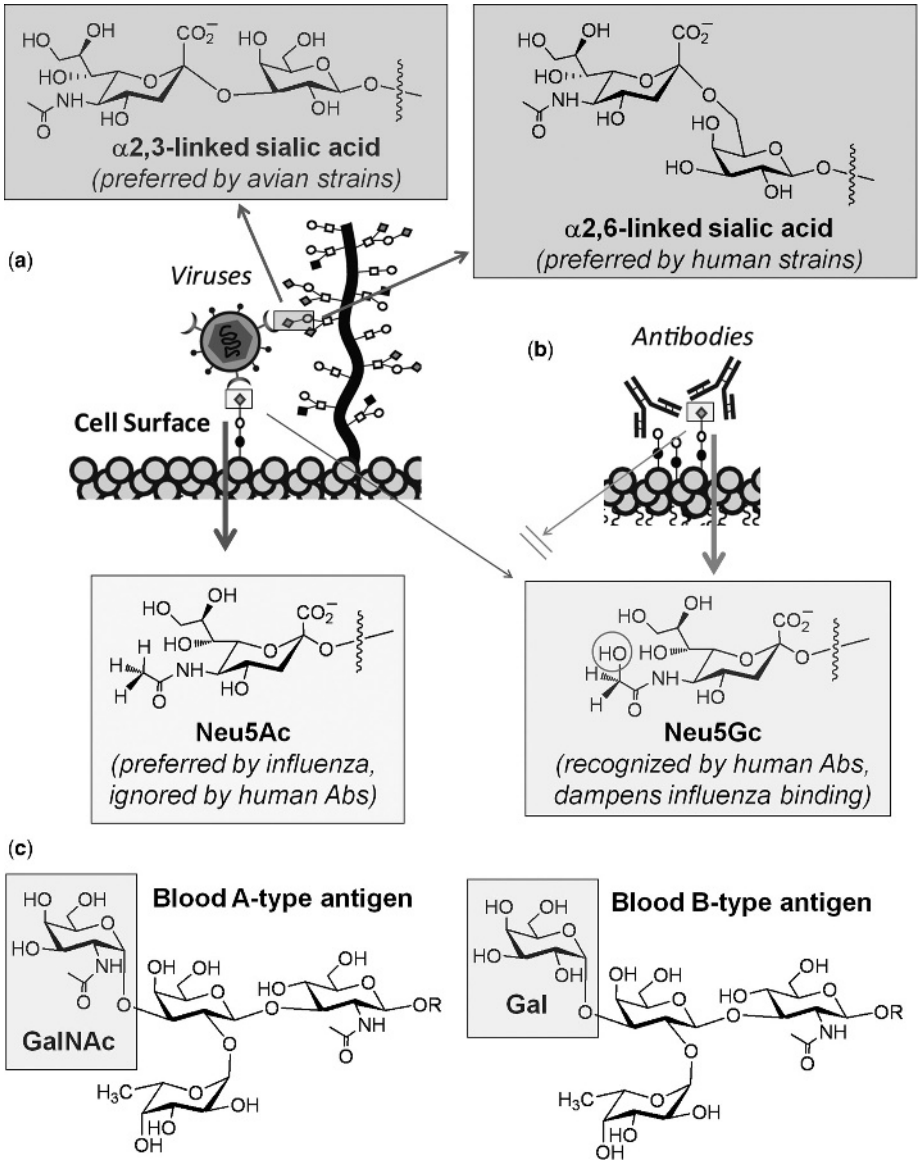

Figure 1.2. (a) The level of discrimination of molecules that "decode" the glycocalyx is illustrated by the ability of viruses such as strains of influenza to discriminate between α2,3- and α2,6-linked sialic acid and at an even more nanoscopic scale, between the Neu5Ac and Neu5Gc forms of sialic acid that vary by a single oxygen atom. (b) Glycans can also be potent antigens, with the human immune system recognizing the nonhuman "Neu5Gc" sialoside. (c) The immune system is also able to discriminate between the single added *N*-acetyl group in A-type blood compared to B-type (in O-type blood, the entire GalNAc or Gal residue is absent).

known to employ sialic acid as a binding epitope for infection of human cells—can discriminate between differences in how this sugar is attached to the underlying stratum (e.g., via an α2,3-glycosidic linkage versus an α2,6-glycosidic linkage, Fig. 1.2a) as well as respond to minute changes in the submolecular chemical structure of the monosaccharide (e.g., Neu5Ac *versus* Neu5Gc, Fig. 1.2b). The human immune system can also distinguish between glycans at this fine level of detail, with the ability to recognize Neu5Gc as "nonhuman" compared to the ubiquitous Neu5Ac. To attempt to provide a sense of scale, the extra "O" of Neu5Gc compared to Neu5Ac that can initiate a system-wide immune response is about one ten billionth the size of the person (e.g., 0.2 nm v. 2 m). By comparison, the proverbial butterfly that has been postulated to flap its wings in Tokyo, which initiates a sequence of events that causes a storm in New York City, would only have to be ~1 mm in size to initiate the fabled weather changes (and perhaps a more typical centimeter-sized insect would spawn a veritable hurricane?).

To the nonspecialist, the point just raised that the immune system can recognize minute changes in carbohydrate structure may seem surprising, especially in light of general expectations that sugars are weakly if at all immunogenic and in some cases, as exemplified by polysaccharide capsules of some pathogenic bacteria, actively block immune recognition. However, as a counterpoint, the century-old xenotransplantation problem manifest through blood-type antigens that are carbohydrate based (in fact, determined by a single acetyl group of a much larger saccharide structure that distinguish A and B blood types, Fig. 1.2c) dispels any doubt that sugars can be potent antigens. Accordingly, to sate the curious reader's desire for insight into how a cell produces glycans with α2,3- versus α2,6-linked sialic acids or blood groups with fully hydroxylated or *N*-aceylated forms of galactose, the majority of this chapter will be devoted to biosynthesis by providing an outline of the production of high energy nucleotide sugar "building blocks" and their assembly into glycoconjugates. Then a brief overview of highly sophisticated methods to analyze and manipulate carbohydrates that are now available will be given as a preview of the more in-depth discussion of these topics presented in this book.

1.2. THE MECHANICS OF MAMMALIAN GLYCOSYLATION

1.2.1. Glycosylation—A Post-Translational Modification and More

To the nonspecialist, glycosylation is perhaps best known as just one of many post-translational modifications (PTMs) that increase protein diversity and function. Although dozens of other types of PTM occur, glycosylation—broadly defined as the addition of one or more sugar residues to a protein or lipid to convey additional information, structure, or function—is arguably the most common[3] and unquestionably the most diverse. The dazzling complexity of the resulting products is not always fully appreciated; for example, the combinatorial addition of one of ~50 different carbohydrate structures to either of two sites of *N*-glycosylation found in the prion protein

converts a single chemical entity into ~2500 entities.[4,5] As evidence accumulates that different glycoforms, such as those that decorate prions, can endow their host molecule with unique biological activities, it is becoming increasingly clear that the great chemical and structural variety resident in the signaling, receptor, and structural molecules that comprise the human body is in large part due to glycans. Far from simply being entities that modulate the function of other biomolecules, however, glycans also have many important roles in and of themselves; for example, they can act as binding epitopes irrespective of their protein or lipid host and, in the case of polysaccharides, can be free-floating molecules not covalently attached to other cellular elements (or more often, assembled into complexes through noncovalent interactions).

Because of the many compelling reasons an organism employs glycosylation, it is not surprising that cells invest substantial resources in glycan biosynthesis. In the following sections of this introductory chapter, the biosynthesis of mammalian glycans is outlined by first examining the metabolic processing of basic monosaccharide building blocks (Fig. 1.3a–d) by a minimum of 220–250 genes, or roughly 1% of the human genome,[6] that act in concert to create and modify oligosaccharide and polysaccharide structures (Fig. 1.3e,f). Many of these players, such as the suites of overlapping α2,3-, α2,6-, and α2,8-sialyltransferases shown in Figure 1.3e, are seemingly redundant. In reality, although catalyzing the same biochemical reaction, these enzymes enable cells to use sugars for finely discriminate biological recognition that is tuned by the exact underlying structure of the host glycoconjugate. For example, ST6GAL1 adds an α2,6-linked sialic acid to an underlying glycoprotein whereas ST6GAL2 adds an α2,6-linked residue to free glycan chains. Showing a similar ability to distinguish the underlying biomacromolecule, ST8SIA1 and ST8SIA5 add α2,8-linked sialic acids to gangliosides while ST8SIA2 and ST8SIA4 add α2,8-linked sialic acids to glycoproteins, usually in the form of polysialic acid to the neural cell adhesion molecule (NCAM). A broader overview of various glycoprotein, glycolipid, GPI-anchored, and independently functional polysaccharide linkages is provided below with a brief description of the structure and function of the various classes of enzymes in the respective biochemical pathways.

1.2.2. Monosaccharides—The Building Blocks for Glycosylation

1.2.2.1. Monosaccharides Are Obtained from the Diet and Transported into Cells. All told, ~50% of all proteins[7] and about one billion glycosphingolipids, which is ~1% of the total of $\sim 10^{11}$ lipids found in a generic mammalian cell, are glycosylated. In mammals, diet is the ultimate source for monosaccharides used for the biosynthesis of these plentiful glycans, but most cell types can also scavenge sugars released into the bloodstream by other tissues and organs and salvage monosaccharides during the recycling of their own glycoconjugates. A typical mammalian diet containing polysaccharides and starches provides a rich supply of simple sugars, primarily glucose (Glc), after digestion in the gastrointestinal (GI) tract as well as several less abundant monosaccharides such as galactose (Gal), mannose (Man), or glucosamine (GlcN) (Fig. 1.3a). These hexoses are absorbed into the bloodstream and taken up by cells throughout the

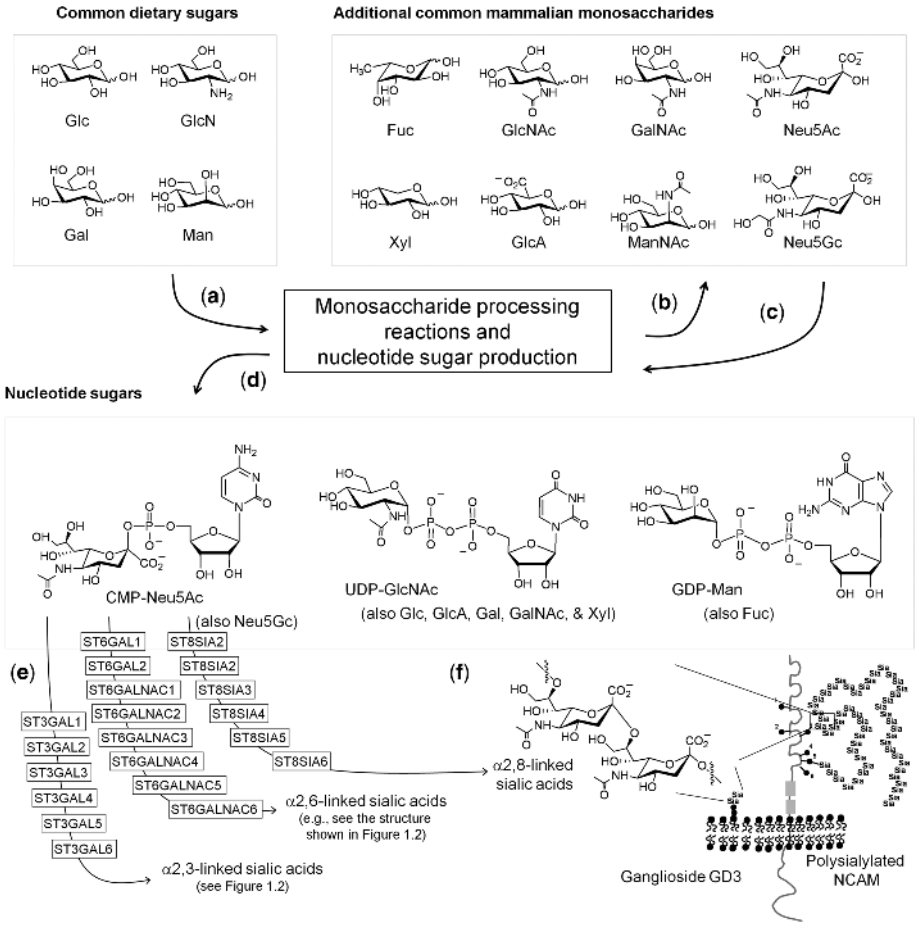

Figure 1.3. Overview of mammalian glycan biosynthesis. (**a**) Common dietary sugars such as glucose (Glc), glucosamine (GlcN), galactose (Gal), and mannose (Man) are taken into a cell by a family of transporters and diversified into (**b**) fucose (Fuc), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), xylose (Xyl), glucuronic acid (GlcA), *N*-acetylmannosamine (ManNAc), *N*-acetylneuraminic acid (Neu5Ac), and *N*-glycolylneuraminic acid (Neu5Gc). (**c**) Both sets of sugars, with the exception of GlcN and ManNAc, are then converted into nucleotide sugars represented by CMP-Neu5Ac/Gc, UDP-GlcNAc (Glc, GlcA, Gal, GalNAc, and Xyl also utilize UDP), and GDP-Man (Fuc also is linked to GDP), (**d**) which are used as building blocks for glycan assembly by families of glycosyltransferases. (**e**) An example of these enzymes is provided by the suite of sialyltransferases that construct α2,3-, α2,6-, and α2,8-linked sialosides; this latter category is represented by a single α2,8-linked residue on GD3 or the homopolymer of dozens of residues resident on the neural cell adhesion molecule (NCAM).

body via the SGLT (sodium-dependent co-transporters from the gene SLC5A) and GLUT (sodium-independent facilitative transporters from the gene SLC2A) families[8,9] of monosaccharide transporters found in the plasma membrane. Members of these two transporter families are localized to different tissue types. For example, GLUT1 is found in erythrocytes; GLUT4, GLUT5, and GLUT12 predominate in skeletal muscle tissue, although other transporters are also expressed at lower numbers in this metabolically voracious tissue[10]; and GLUT14 is specifically expressed as the predominant transporter in two alternative splice forms in human (but not mouse) testes.[11] Another feature of these proteins is that they have overlapping substrate specificity with many family members capable of transporting multiple monosaccharides, albeit with differing efficiencies.[12]

1.2.2.2. De Novo Synthesis of High-Energy Nucleotide Sugars. After monosaccharides enter cells, they undergo a series of chemical conversions including epimerization, acetylation, condensation, and phosphorylation reactions to produce the full spectrum of "building blocks" required for glycan biosynthesis (Fig. 1.3b). The intracellular metabolic network is capable of sufficient *de novo* synthesis of glucosamine (GlcN), fructose (Fru), mannose (Man), fucose (Fuc), *N*-acetylneuraminic acid (sialic acid, Neu5Ac or Sia), galactose (Gal), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), xylose (Xyl), and glucuronic acid (GlcA), all of which exist in the D-conformation except for L-fucose. Mammals other than humans also produce the *N*-glycolylneuraminic acid (Neu5Gc) form of sialic acid.[13]

In the cytosol these monosaccharides can be phosphorylated and subsequently coupled with nucleotides such as uridine diphosphate (e.g., UDP-GlcNAc), guanosine diphosphate (e.g., GDP-mannose), or cytosine monophosphate in the case of sialic acids to create a set of high energy "building blocks" for glycan assembly (Fig. 1.3d). In some cases, including initial steps in the synthesis of the dolichol-linked 14-mer used in *N*-linked glycan biosynthesis (discussed below) or for *O*-GlcNAc protein modification,[14] nucleotide sugars are used in the cytosol; more often they are transported into the endoplasmic reticulum (ER) or the lumens of the Golgi apparatus where the bulk of oligosaccharide assembly and processing occurs. In either case, the release of the monosaccharide from its bonded nucleotide phosphate provides energy that drives the formation of glycosidic bonds found in glycolipids and glycoproteins.

1.2.2.3. Transport of Nucleotide Sugars into ER/Golgi. The transport of high energy nucleotide sugars from the cytosol into the ER and the Golgi lumen occurs by highly specific membrane proteins of the SLC35 nucleotide sugar transporter family. This class of proteins has at least 17 members, some of which can accept multiple substrates; at the same time certain nucleotide sugars can be accepted by multiple transporters. These transporters are organelle specific; typically nucleotide sugars are only transported into an organelle compartment endowed with the corresponding glycosyltransferases.[15] For example, CMP-Sia, GDP-Fuc, and UDP-Gal are transported solely into the Golgi; UDP-GalNAc, UDP-GlcNAc, UDP-GlcA and UDP-Xyl are transported twice as rapidly into vesicles of Golgi as ER; conversely, UDP-Glc is transported into ER vesicles much more rapidly than into the Golgi.[12]

Regulation of the assembly of complex carbohydrates from nucleotide sugars—an extremely complex and still incompletely understood process—is accomplished by several means, including compartmentalization of glycosyltransferases, subtle differences in the activities of these enzymes, nucleotide sugar transport rates, and the available concentration of substrates.[16] By their influence on the latter two parameters, nucleotide sugar transporters play a major role in determining the outcome of glycan structure by regulating which macromolecules will undergo glycosylation.[17] These proteins are anti-porters that exchange nucleotide sugars for corresponding nucleotide monophosphates (NMPs) in an equimolar fashion.[18] Nucleotide monophosphates are produced by enzymatic dephosphorylation of nucleotide diphosphates (NDPs) generated during the glycosyltransferase-catalyzed attachment of a monosaccharide residue to a growing oligosaccharide chain. Consequently, both the specific transporter and the corresponding nucleotide diphosphatase are required within the lumen of a specific organelle for successful transport. Transport is competitively inhibited by corresponding nucleoside mono- or diphosphate in the cytosol, but not by the free sugars, and does not require an energy source such as ATP. Jumping ahead to the completion of glycan assembly, postsynthetic glycan modifications such as phosphorylation, acetylation, and sulfation also require active transport of materials into the ER and Golgi lumens; for example, PAPS (3′-phosphoadenosine 5′-phosphosulfate) that is required for sulfation is imported by the PAPST1 gene product.[19]

1.2.3. Glycoconjugate Assembly

Once the required high energy nucleotide sugars, other requisite building blocks, and biosynthetic enzymes have been localized within the appropriate cellular compartments, the glycosylation of newly synthesized proteins and lipids can begin. In the following sections we examine the production of the major classes of prevalent mammalian glycan structures (a more thorough discussion, including less prevalent glycans not discussed here, can be found in review articles[12,20]). To briefly summarize, the two major classes of glycosylated proteins, *N*-linked (Fig. 1.4) and *O*-linked (Fig. 1.5) are discussed followed by lipid-associated glycans (e.g., GSL and GPI anchors, Fig. 1.6), and, finally, polysaccharides that are primarily represented in mammals by glycosoaminoglycans (GAGs, Fig. 1.7).

1.2.4. Glycoproteins

1.2.4.1. N-Linked Glycans. *N*-Linked glycosylation, where the term "*N*-linked" simply refers to the chemical linkage of the glycan moiety to the nitrogen of the amido group of an asparagine (Asn) residue, is one of the most prevalent protein modifications, occurring on virtually all membrane-associated and secreted proteins. It serves many valuable functions, including stabilization of protein structure, enhanced solubility, immunomodulation, mediation of pathogen interaction, serum clearance rate, protein half-life, and proper folding.[21] Dysfunctional *N*-glycosylation can result in serious detriment to an organism exemplified by health issues ranging from congenital disorders of glycosylation (CDGs)[22] and adult diseases such as cancer.[23]

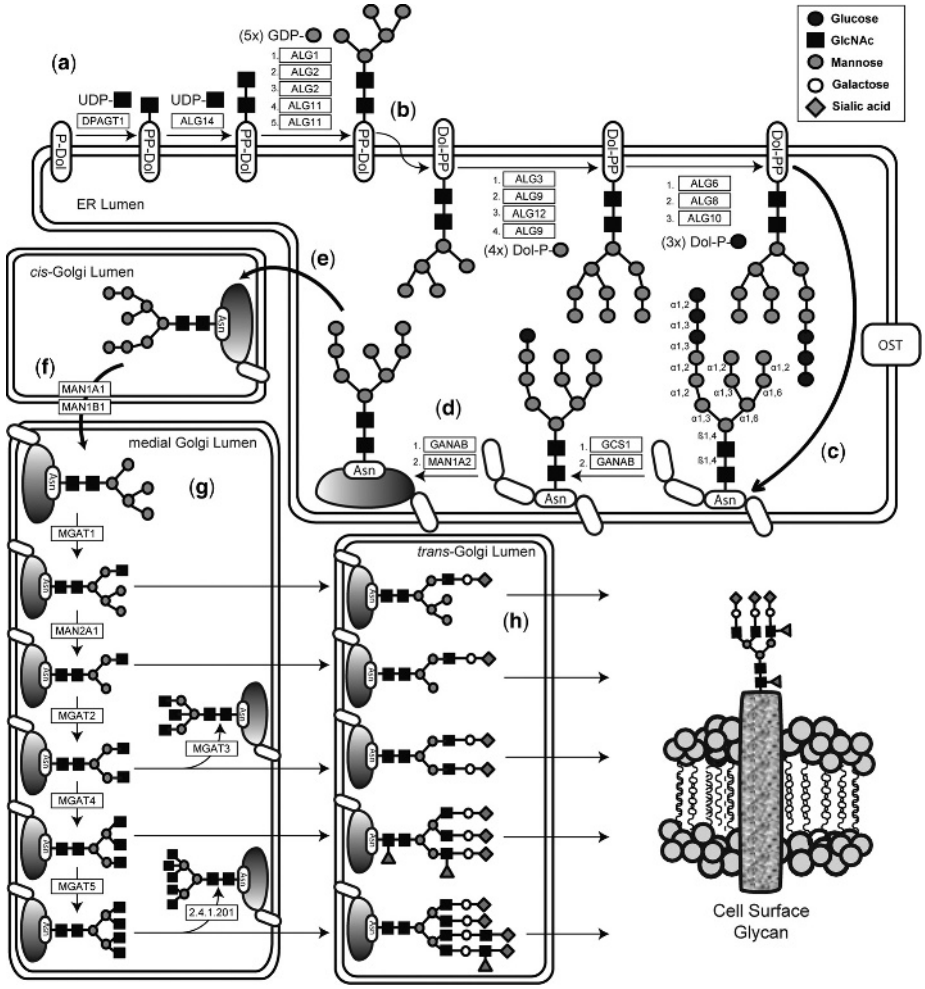

Figure 1.4. Outline of *N*-linked glycoprotein biosynthesis showing topography and major biosynthetic events. (**a**) Production of the Dol-PP-14-mer begins on the cytosolic side of the ER. (**b**) The nascent glycolipid is flipped to the luminal side after the assembly of two GlcNAc and five Man residues and an additional four Man and three Glc residues are added to create the $GlcNAc_2Man_9Glc_3$ 14-mer that is transferred *en bloc* by OST to a newly synthesized, yet unfolded peptide (**c**). (**d**) Trimming of the Glc residues controls protein folding in the calnexin/calreticulin cycle in the ER followed by (**e**) transfer to the *cis*-Golgi lumen where (**f**) removal of four Man residues produces the $GlcNAc_2Man_5$ "core" structure that is subsequently elaborated into a diverse array of high mannose, complex, and hybrid *N*-glycans. (**g**) As an example, progressive branching of N-glycans is carried out by GlcNAc transferases MGAT1 to MGAT5 and (**h**) further elaboration with fucose, galactose, and sialic acid can occur in the *trans*-Golgi compartment. Enzyme abbreviations are shown (where known), and updated information on enzymatic activity (indicated by the EC number) can be found in online data bases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) Glycan Pathway resources.[154,155]

N-Linked glycosylation is a multicompartmental affair, involving the cytosol and both the ER and the Golgi complexes (Fig. 1.4). Biosynthesis of glycoproteins begins on the cytosolic face of the ER by the formation of the polyisoprenoid lipid dolichol phosphate (Dol-P, Fig. 1.4a), which acts as a carrier for the nascent glycan structure. Assembly of the core glycan on this carrier begins by addition of a GlcNAc-P (from UDP-GlcNAc) onto Dol-P, forming GlcNAc-pyrophosphoryldolichol (GlcNAc-PP-Dol) through the action of the GlcNAc-1-phosphotransferase DPAGT1. This reaction exemplifies the use of monosaccharide transferases that occur throughout the assembly of the core structure as well as in subsequent elaboration processes. A second GlcNAc and five Man residues are added (from UDP-GlcNAc and GDP-Man, respectively) in sequence to form $Man_5GlcNAc_2$-PP-Dol, which is then flipped to the luminal side of the ER (Fig. 1.4b).[24] Additional Man and Glc residues are added in the lumen of the ER via donors Dol-P-Man and Dol-P-Glc, resulting in the primary core structure $Glc_3Man_9GlcNAc_2$-PP-Dol. The terminal α1,2-linked Glc residue is required for recognition by the oligosaccharyltransferase (OST) that attaches this core glycan structure *en bloc* to the host protein.[20]

Rigorously speaking, *N*-linked glycosylation is not a PTM but rather is a cotranslational event that occurs during translation as OST searches unfolded polypeptides emerging from the ER for a universal Asn-X-Ser/Thr consensus sequence (where X is any amino acid except proline, which is disallowed because the constraints it imposes on the flexibility of a peptide chain prevent the consensus sequence from forming a loop structure where the hydroxyl group of Ser/Thr interacts with the amido group of Asn rendering it more nucleophilic and enhancing the installation of the glycan moiety).[25] Oligosaccharyltransferase binds to $Glc_3Man_9GlcNAc_2$-PP-Dol and catalytically cleaves the phosphoglycosidic bond in the GlcNAc-P moiety thereby releasing Dol-PP during the transfer of $Glc_3Man_9GlcNAc_2$ to the targeted Asn residue (Fig. 1.4c).[26] Once the *en bloc* transfer of the core $Glc_3Man_9GlcNAc_2$ 14-mer to protein is complete, two Glc residues are removed in sequence by glucosidase I (the terminal α1,2-linked Glc) and glucosidase II (the penultimate α1,3-linked Glc). $Glc_1Man_9GlcNAc_2$ targets the nascent glycoprotein for entry into the calnexin/calreticulin cycle within the ER, a major component of the quality control system that assists glycoproteins to fold properly and to achieve their ideal conformation (Fig. 1.4d).[21,27] After folding is completed, the final Glc residue is removed by glucosidase II, and a terminal α1,2-linked Man residue is removed from either of the two other arms of the oligosaccharide by ER mannosidase I or II. The remaining $Man_8GlcNAc_2$ oligosaccharide structure is transported along with the newly formed protein to the *cis*-Golgi for further modification (Fig. 1.4e). Golgi mannosidases IA and IB subsequently remove three additional α1,2-linked Man residues to form an intermediate $Man_5GlcNAc_2$ glycan structure (Fig. 1.4f) that is subsequently built into the high-mannose, complex, and hybrid subclasses of *N*-linked glycans (Fig. 1.4g).

The formation of complex *N*-linked glycans begins in the medial-Golgi with the addition of a GlcNAc residue by mannosyl-α1,3-glycoprotein-β1,2-*N*-acetylglucosaminyltransferase (MGAT1) onto the α1,3-linked Man residue of $Man_5GlcNAc_2$.[28] Mannosidase II removes the two remaining terminal Man residues from $Man_5GlcNAc_2$, and mannosyl-α1,6-glycoprotein-β1,2-*N*-acetylglucosaminyltransferase

(MGAT2) adds a GlcNAc residue to the final remaining terminal Man residue.[29] In cases where robust flux through the hexosamine pathway occurs, sufficient UDP-GlcNAc is present to activate MGAT4 and 5 and produce tri- and tetra-antennary structures (Fig. 1.4h).[30] Further modifications, such as the addition of a Fuc to the proximal GlcNAc, addition of a β1,6-GlcNAc to the α1,6-linked Man residue that already bears β1,2-GlcNAc, or capping with terminal sialic acids, generate a wide variety of diverse structures within the complex *N*-linked glycan class.[12] Hybrid *N*-linked glycans ensue from the addition of β1,2-GlcNAc to the α1,3-linked Man residue of the intermediate $Man_5GlcNAc_2$ glycan structure; the removal of the two remaining Man residues by mannosidase II, as occurs in complex *N*-linked glycan biosynthesis, is prevented by the addition of a β1,4-GlcNAc to the proximal β1,4-Man residue.[31] This mannosidase II-protected structure is translocated to the *trans*-Golgi where additional modifications to the oligosaccharide structure occur, again generating a potentially vast array of structures.

1.2.4.2. O-Linked Glycans. *O*-Linked glycosylation is a post-translational modification where the glycan moiety is attached to the hydroxyl group of a serine or threonine amino acid residue of a protein; in extreme cases, exemplified by mucins, dense clusters of carbohydrate force the peptide chain into a highly extended, poorly folded conformation. There are several subclasses of *O*-glycans, including *O*-linked Fuc important in Notch signaling,[32] *O*-xylose that tethers GAGs to membrane proteins,[12] and *O*-glycan linkages to hydroxylysine (in the collagen sequence –Gly–X–Hyl–Gly–) as well as to hydroxyproline in plants,[33] but by far the most common form of *O*-glycosylation found on cell surfaces or secreted glycoproteins is the addition of α-*N*-acetylglucosamine to the oxygen of serine or threonine by one of a family of over 20 *O*-GalNAc transferases (24 in humans).[34] Unlike *N*-glycosylation, which begins with the *en bloc* transfer of the $Glc_3Man_9GlcNAc_2$ 14-mer core structure, mucin-type *O*-glycosylation begins in the Golgi apparatus with the addition of the single GalNAc to produce an initial *O*-linked GalNAc-Ser/Thr structure, known as the Tn-antigen. This rudimentary glycan moiety is then translocated to the *trans*-Golgi for elongation through the stepwise addition of Gal, GalNAc, or GlcNAc residues, which form the basis of eight core structures (Fig. 1.5a) that can be further modified by sialylation, sulfation, acetylation, fucosylation, or polylactosamine extension.[12]

Although there is not a clear cut consensus sequence for mucin-type *O*-linkages, statistical analysis has yielded a rule set to predict sites of *O*-glycan modification.[35] Most obviously, because *O*-glycosylation occurs on fully folded proteins, only surface-exposed Ser and Thr residues will be accessible for attachment of the initiating GalNAc. More subtly, the density patterns of *O*-linked glycans suggest that nearby glycans can influence transferase activity[36] and neighboring amino acids can also modulate *O*-glycan initiation (to date, these trends have not been well enough established to be considered legitimate consensus sequences of the type that specify *N*-glycans).[37,38] Finally, there are inherent differences in site specificity of the various GalNAc transferase family members that result in cell- and tissue-specific patterns of *O*-linked glycans.

1.2.4.3. O-Glycosylation of Nucleoplasmic Proteins. Also germane to a discussion of *O*-linked glycans is the addition of a single GlcNAc to Ser or Thr to form a

unique class of intracellular glycosylated proteins.[39] *O*-GlcNAc is a common PTM of nuclear and cytosolic proteins, including nuclear pore proteins, transcription factors, and cytoskeletal elements. *O*-GlcNAc modification is likened more to phosphorylation than to the other forms of *O*-glycosylation due to its transient nature and yin-yang status with phosphorylation at the same amino acid sites, particularly during different cell cycle stages and in development.[40] *O*-GlcNAc protein modification is essential for life in mammalian cells playing key roles that include serving as a nutrient sensor in regulating insulin signaling, the cell cycle, and calcium handling, as well as the cellular stress response.[14]

1.2.5. Glycolipids

1.2.5.1. Glycosphingolipids. A glycolipid is any compound containing one or more monosaccharide residues covalently linked to a hydrophobic moiety such as an acylglycerol, a sphingoid, or a prenyl phosphate. In mammals, the majority of glycolipids are glycosphingolipids (GSLs), a large and widely varying family of amphipathic lipids based on the ceramide *N*-acylsphingoid lipid moiety (Fig. 1.6a). Glycosphingolipids reside in cellular membranes, typically in the plasma membrane, where the glycan is almost always oriented outwards, exposed to the extracellular space. These molecules play a role in the protective glycocalyx covering of a cell and, by participating in raft assemblies such as the "glycosynapse" (Fig. 1.6b),[41] GSLs participate in cell–cell recognition, cell–matrix interactions, and cell surface reception and messaging. Glycosphingolipids are required for proper development, as evidenced by a founder effect observed in a large Old Order Amish pedigree where GM3 synthase was defective[42]; more broadly, biosynthetic or catabolic GSL defects result in pathologies ranging from liver disease to insulin-resistant diabetes, multiple sclerosis, and Tay-Sachs and Graves' diseases.

In more detail, GSL synthesis begins on the cytosolic face of the ER[43] with the condensation of a serine residue and a palmityl-CoA to form 3-dehydrosphinganine, which is hydroxylated at the 4′ oxygen, *N*-acylated, and unsaturated between C4 and C5 in a *trans* fashion to form ceramide (Cer).[44] Ceramide then crosses the ER membrane and undergoes one of several modifications that lead to different classes of glycolipids; most commonly, Cer is conjugated with a Gal or Glc residue to form the simple glycolipids GalCer and GlcCer (Fig. 1.6a); these two glycolipids form the core of all mammalian GSLs. In contrast to most glycans, the GalCer core undergoes relatively few and mostly conservative modifications. In addition, GalCer-based GSLs are restricted to a few specific cell types, including myelin sheathing provided to neuronal axons by oligodendrocytes and Schwann cells, and epithelial cells of renal tubules and the GI tract[45]; this narrow distribution has been exploited insofar as "Gal-C" is a marker for oligodendrocyte differentiation.[46] Reinforcing the earlier point regarding "markers" actually being functional, transduction of GALC (the enzyme that synthesizes GalCer) into mouse forebrain improved twitcher oligodendrocyte morphology *in vivo*.[47]

Unlike GalCer-derived GSLs, the GlcCer core experiences extensive elaborations that generate hundreds of distinct structures; moreover, structures based on GlcCer are not restricted to narrow classes of cell but are almost ubiquitous. This GSL forms at

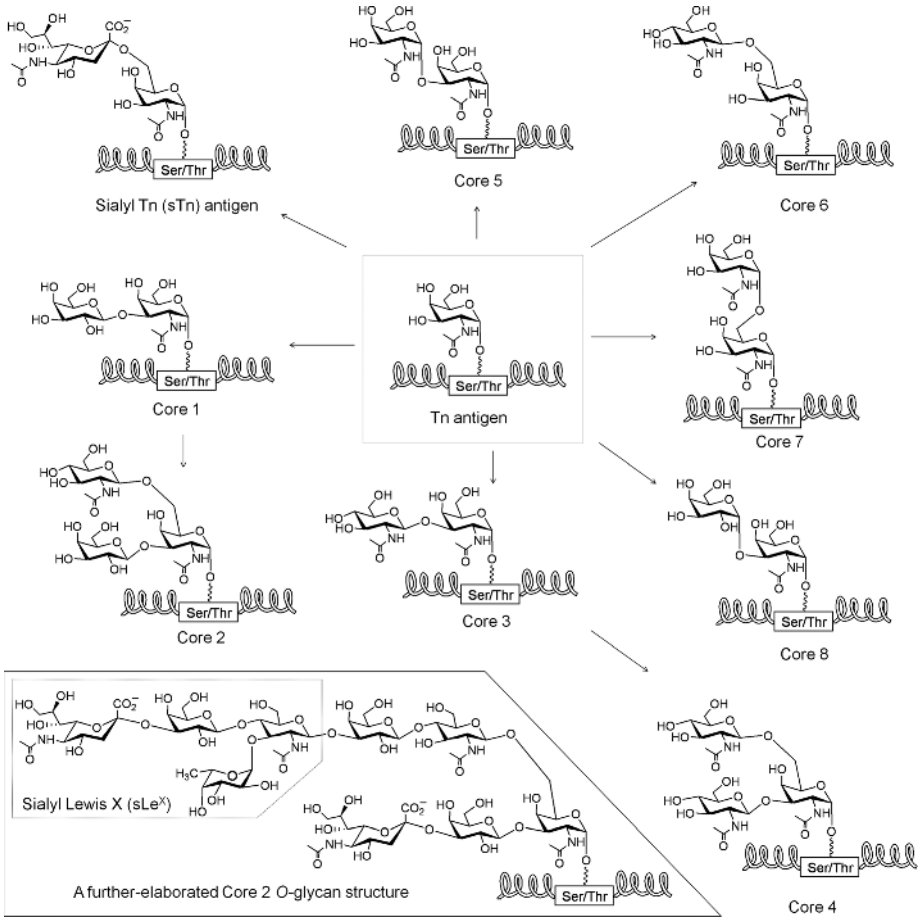

Figure 1.5. Overview of mucin-type *O*-linked glycoprotein biosynthesis. The production of the eight core structures found in the cell surface and secreted *O*-glycans is shown. An example of a further elaborated *O*-glycan bearing the sialyl Lewis X epitope[85] important in leukocyte homing and cancer metastasis is shown in the inset.

the cytosolic face of the *cis*-Golgi and is translocated to the lumen of the Golgi via the Golgi stack trafficking process[48] to become a substrate for various glycosyltransferase enzymes and complexes. The addition of a Gal residue to GlcCer results in LacCer, which is the foundation for three additional classes of GSLs. First, the (neo-)lacto-series, or blood group series, begins with the addition of a β1,3-GlcNAc. Next, the globo series is distinguished by the addition of an α1,4-Gal (also known as the Pk antigen). Third are the gangliosides, glycolipids that feature one or more sialic acid residues; membership in this group does not preclude inclusion in the lacto- or Globo series.

Gangliosides are present in nearly all animal cells, but they are particularly prevalent in the plasma membranes of cells in the central nervous system[48]; this class of GSL has

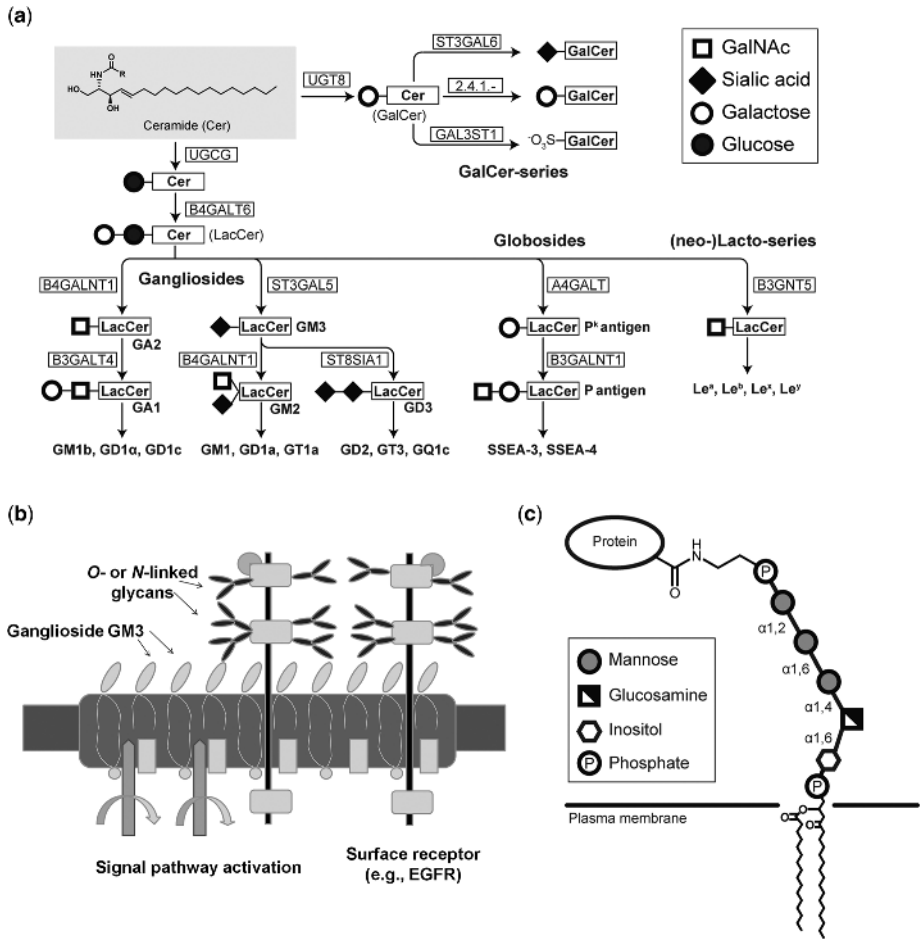

Figure 1.6. Overview of glycolipid biosynthesis. (**a**) Mammalian GSLs are synthesized from Cer after the addition of Gal, to form the small GalCer series, or after the addition of Glc and Gal to form the ubiquitous LacCer class, which is subdivided into gangliosides, globosides, and the neo-lactoseries. Additional information on the enzymes and specific GSLs shown can be found in the KEGG databases (see the legend for Fig. 1.3). (**b**) Collectively, glycolipids function as lipid raft assemblies, exemplified by the type 1 "glycosynapse" (adapted from Hakamori,[41] which shows additional examples). (**c**) Glycophosphatidylinositol anchors are important functional structures on the cell surface. The fatty acid phosphatidylinositol is embedded in the exterior leaflet of the plasma membrane and features a tether consisting of a specific series of monosaccharides and phosphoethanolamine linked to the C-terminus of a protein.

also been implicated in a variety of diseases. With the exception of GM4, all gangliosides emanate from LacCer and continue down one of two branches: the asialo pathway (also called the o-pathway) through addition of a GalNAc residue, or into the "ganglioside proper" pathways (a-, b-, and c-pathways) through the addition of one or more sialic

acid residues. The N-terminal domains of the promiscuous glycosyltransferases responsible for the construction of the gangliosides specify the distribution of these enzymes within the Golgi stacks, resulting in a differential expression pattern.[48] There also exists a salvage pathway for resynthesizing gangliosides, recycling them from their endosomal breakdown through the Golgi; this recycling pathway dominates in slowly dividing cells, while *de novo* synthesis dominates in highly mitotic cells.

1.2.5.2. GPI Anchors. The discovery that phospholipase C could release alkaline phosphatase from lipid-linked structures on cellular surfaces[49] led to the identification of the glycophosphatidylinositol (GPI) membrane component.[50] Glycophosphatidylinositol structures are a synthetic tour de force of nature, combining lipid, carbohydrates, and proteins into a single macromolecule. The basic structure of the GPI anchor (Fig. 1.6c, maintained across all species studied thus far) begins with phosphatidylinositol (PI), which spans the external ER membrane leaflet linked to an inositol via a phosphodiester. An oligosaccharide chain, attached to the inositol, consists of GlcN (donated from a rapidly de-acetylated UDP-GlcNAc) and three linear Man residues (provided by Dol-P-Man donors). Finally, phosphoethanolamine (P-EtN) is linked to the terminal Man residue (donated by phosphatidylethanolamine), resulting in the core EtN-P-Man_3-GlcN-PI structure to which proteins are covalently linked (although not all GPI anchors ultimately bear a protein).[51] In mammals, prior to the attachment of a protein, GPIs are completely assembled in the membrane of the ER by a series of enzymes which are products of the phosphatidylinositol glycan anchor biosynthesis (PIG) family of genes.[51] Although the topology of every biosynthetic step has not been elucidated, it is known that the synthetic process begins on the cytosolic face of the ER while the attachment of protein occurs on the luminal face of the ER membrane,[52] suggesting that a yet-to-be-discovered "flippase" participates in the production of GPI-anchored structures.[53]

Phosphoethanolamine provides the attachment point for a protein via an amide bond between the C-terminal residue of the protein and the N-terminal of P-EtN.[51] Proteins that are destined for GPI binding are targeted to the ER during their synthesis by an N-terminal signal and translocated to the ER lumen. They contain a C-terminal signal peptide, which upon removal exposes their acidic C-termini and allows attachment to GPI by ethanolamine through a transamidation reaction. The GPI-anchored proteins belong to the type 1 class of the GPI structures, which have Man-α1,6-Man-α1,4-GlcN-α1,6-phosphatidylinositol (PI) core linkages (other GPI structures have varying linkages between these core residues). After passing through the Golgi apparatus for further protein modification, the entire structure is translocated to the exterior leaflet of the plasma membrane. Certain proteins require GPI anchoring to be functional; for example Ly-6A/E-mediated T-cell activation is critically dependent on its GPI anchor[54] and folate uptake functions efficiently only when its receptors are GPI anchored.[55]

1.2.6. Polysaccharides

In contrast to the relatively modestly sized oligosaccharides that modify proteins, lipids, and GPI anchors, mammals also assemble carbohydrates into much larger, linear polysaccharide structures. Despite losing the inherent complexity derived from

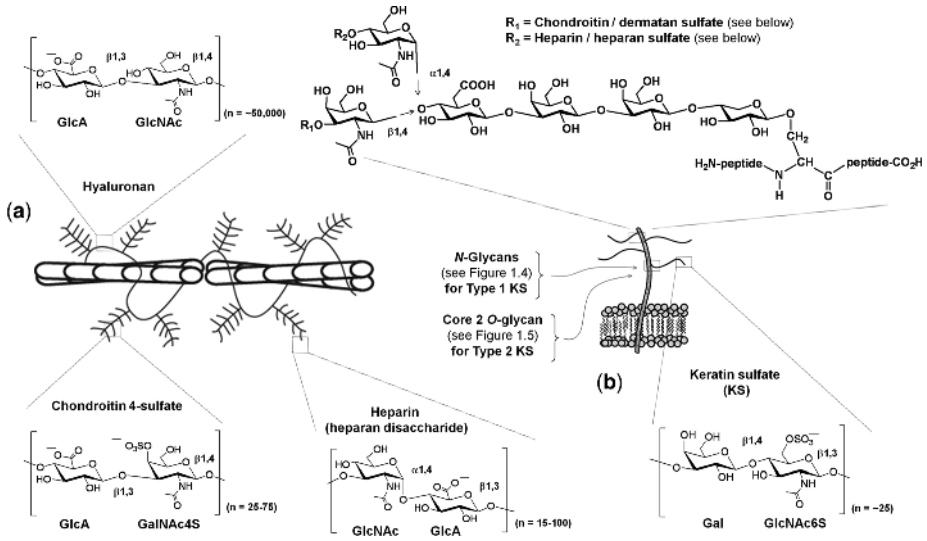

Figure 1.7. Mammalian glycosaminoglycans (GAGs). Mammalian polysaccharides are primarily GAGs (another example is polysialic acid, shown in Fig. 1.3f) that can be associated with (**a**) proteoglycans or (**b**) membrane proteins. More details are provided in the main text and other resources.[156]

branching, and being made from repeating units of only two monosaccharides, polysaccharides are nonetheless highly diverse through a series of postsynthetic modifications, primarily epimerization and sulfation reactions.[56]

Polysaccharides generally exist outside of a cell, sometimes remaining attached to surface elements to form an interface between a cell and its surroundings, and sometimes secreted freely into the extracellular matrix (ECM). These sugars possess their own inherent functionality and are of critical importance to cellular function because they modulate adhesion, migration, differentiation, and proliferation and influence angiogenesis and axonal growth. Extracellular matrix polysaccharides become highly hydrated and thus serve as hydrogels for embedded fibrous ECM proteins, such as collagen, as well as scaffolds for signaling molecules such as growth factors. In mammals, there are four main classes of structural polysaccharides that are all glycosaminoglycans (GAGs): hyaluronic acid (or hyaluronan), heparin (or heparan sulfate), keratin sulfate, and chondroitin/dermatan sulfate. Each category of GAG is now discussed briefly.

1.2.6.1. Hyaluronan. A human is estimated to contain about 15 g of hyaluronic acid (HA), of which a remarkably high proportion—about one third—is turned over each day[57]; if a person had to replenish this HA through commercial sources, it would cost him or her tens of thousands of dollars each day. Fortuitously, HA is synthesized endogenously, conveniently enough at the plasma membrane (rather than in the ER or Golgi apparatus) allowing it to be easily secreted directly to the ECM by one of three distinct hyaluronan synthases.[58] Hyaluronan has the simplest chemical composition of the

GAGs, consisting of the repeating unit [-GlcA-β1,3-GlcNAc-β1,4-] but is by far the largest in size, ranging up to 10,000 disaccharide units resulting in a molecular weight of ~3.7 million Da. Unlike most other GAGs, HA forgoes postsynthetic modification and remains unbound to surface proteins; in solution at physiological salt conditions, a single molecule of HA condenses into a highly hydrated entity with a radius of gyration of ~180 nm, on the same size scale as small organelles. Under appropriate conditions, such as in the presence of multivalent cations or assembling proteins, HA can exist in extended conformation and line up side by side in a brushlike manner that meshes the glyocalyx with a pericellular layer up that is up to 4 μ thick in chondrocytes.[59]

1.2.6.2. Heparin/Heparan Sulfate and Chondroitin/Dermatan Sulfate. Heparin/heparan sulfate GAGs (HSGAGs) and chondroitin/dermatan sulfate GAGs (CSGAGs) share a common synthetic origin, both being linked to a core protein through a specific *O*-linked sequence (GlcA-β1,3-Gal-β1,3-Gal-β1,4-Xyl-β1–) at the same consensus sequence (–Ser–Gly/Ala–X–Gly–). The assembly of the tetrasaccharide linker begins in the ER where Xyl (from UDP-Xyl) is transferred to the hydroxyl group of the serine in the consensus sequence by a xylosyltransferase. The nascent glycoprotein then moves into the *cis*-Golgi where two Gal (by galactosyl transferase I and II) and a GlcA (by glucoronic acid transferase I) are attached sequentially to complete the tetramer.

Synthesis of the polysaccharide portion of these GAGs begins with the addition of GalNAc (or GlcNAc) and GlcA residues to the *O*-linked tetramer in an alternating fashion by multidomain glycosyltransferases.[60] The addition of the first GalNAc or GlcNAc residue determines whether the GAG will belong to the heparan sulfate or chondroitin sulfate family, respectively. The HSGAGs consist of the repeating unit –GlcNAc-α1,4GlcA-α/β1,4–, constructed by enzymes from the EXT gene family glycosyltransferases.[61] The CSGAGs have a different basic disaccharide repeat unit (–GalNAc-β1, 4-GlcA-α/β1,3–), containing GalNAc rather than GlcNAc and employing 1,3- rather than 1,4-glycosidic linkages between the repeating disaccharides; despite the differences in the monosaccharide building blocks used, CSGAGs are also constructed from genes in the EXT family.[60] When the HSGAG or CSGAG chain has grown to an appropriate length, additional enzymes impart structural uniqueness: one or more of a set of 2-*O*-, 3-*O*-, and 6-*O*-sulfotransferases add sulfate groups at appropriate locations,[62] *N*-deacetylase *N*-sulfotransferase can expose the amine groups of GalNAc, and C5 epimerase converts a portion of GlcA residues to IdoA. It is noteworthy that the epimerization of GlcA to IdoA results distinguishes condroitin sulfate from dermatan sulfate.[63]

1.2.6.3. Keratan Sulfate. Keratan sulfate differs from other GAGs in two major respects. First, it can be either *N*- or *O*-linked to the core protein.[64] Second, its repeating disaccharide unit contains a galactose rather than one of the uronic acids (GlcA or IdoA) in its disaccharide repeat. The basic repeating unit is –Gal-β1,4-GlcNAc-β1,3–, assembled by β1,4-galactosyl transferase (B4GALT1) and a β1,3-GlcNAc transferase (B3GNT1 or B3GNT2). There are three classes of keratan sulfate, which are distinct in their protein linkages. KSI members are *N*-linked to an Asn of the protein; they are found primarily in the cornea and can be terminated with sialic acids, Gal, or GlcNAc.

KSII members are *O*-linked to a Ser/Thr residue of the core protein; they are primarily found in cartilage, are highly sulfated, and are terminated by sialic acids. KSIII are found in brain tissue and have a unique serine-*O*-mannose linker between the keratan sulfate chain and the protein.

1.3. METHODOLOGY—NEW TECHNOLOGIES MESH WITH "TRIED AND TRUE" APPROACHES

1.3.1. Carbohydrate Complexity Requires Specialized and Highly Sophisticated Methods

Although mammalian glycosylation has now been elucidated in sufficient detail to provide a basic understanding of glycan biosynthesis, structure, and function, many aspects of these molecules remain mysterious and constitute stumbling blocks as modern medical research seeks sugar-based therapeutics.[65] To illustrate this point, a seeming disparity exists between the incredible complexity theoretically possible in glycan structures (e.g., a set of six nucleotides can be assembled into 4,096 different oligomers, six amino acids into 64,000,000 peptides, and six monosaccharides into 192,780,943,360 oligosaccharides).[66] Clearly, not all of these glycoforms can be found in nature because enzymes do not exist to make every type of glycosydic linkage (e.g., α2,4-linked sialosides are not found in mammals), nevertheless computer models predict from thousands[67] up to millions of biologically possible *N*-glycans.[68] By contrast to these robust numbers, actual studies where glycan profiles have been characterized result in modest sets of glycans ranging from 50 to 60 *N*-glycans characterized from prions using methods available a decade ago,[5] to ~80 isolated from cancer cells more recently.[69]

An argument can be made that the myriad of theoretical or model-predicted glycoforms that may exist in cells (but are undetectable by current methods) have little or no biological relevance or medical importance. Indeed, it logically follows from this school of thought—if taken to an extreme—that low abundance glycoforms need not be characterized in exhaustive detail because glycan function is based on broad classes of these molecules rather than infrequently occurring, individual molecules acting on their own. The galectin lattice, which depends on copious production of the highly branched glycans produced by MGAT4/5 (Fig. 1.4g), provides a specific example of this concept. The glycosynapse (Fig. 1.6b) provides another, where the collective chemical properties of large numbers of GM3 molecules can dominate the biophysical properties of the lipid raft assemblies and the activities of the embedded proteins. Although bulk conversion of GM3 to LacCer (through loss of sialic acid) or to GD3 (through gain of sialic acid) can affect glycosynapse function, it is unlikely that minor changes in equilibrium (e.g., one molecule in a thousand) or the appearance of a few copies of gangliosides of increased complexity, have a measurable impact.

Conceptually, the general reliance of glycan-based biological recognition events on the cluster glycoside effect[70,71] and multivalency[72] provides theoretical backup for the notion that a low abundance glycoform—necessarily acting on its own because it is

statistically unlikely to be in close association with a structural sibling—is unlikely to have any significant impact. Notwithstanding these considerations, the growing use of single molecule detection systems for nucleic acids and proteins continues to pique interest into whether rare "one in a million" glycans have unexpected biological consequences. A scenario where this may be the case is if a particular glycan structure, perhaps a blood group antigen (Fig. 1.2c), is a particularly potent antigen. Historically, an obstacle to experimentally resolving this matter has been the lack of a template for carbohydrate structures (akin to the DNA sequence that specifies the primary amino acid sequences of proteins) that precludes PCR-type amplification of low copy number glycans. Recently, however, a wave of new technologies, often combined with classic methods, have dramatically accelerated progress, as outlined next.

1.3.2. Advances in Bioinformatics, Analytical Methods, and High Throughput Technologies

1.3.2.1. Glycomics—Combining Bioinformatics with Analytical Tools and High Throughput Methods. Despite daunting challenges, modern methods of mass spectrometry (Chapter 11), chromatography, nuclear magnetic resonance, and capillary electrophoresis have identified in aggregate a vast number of carbohydrate structures.[73–75] These techniques, along with high throughput arrays consisting either of immobilized glycans or, reciprocally, of lectins and other glycan-binding proteins[76] (Chapter 10), have resulted in a significant amount of information that is now available in databases, such as those available online from the German Cancer Research Center (http://www.glycosciences.de), the Consortium for Functional Glycomics (http://www.functionalglycomics.org), and the AFMB-CNRS of University of Provence and the University of the Mediterranée (http://www.cazy.org).

In theory, these resources drive "glycomics" efforts to globally characterize the "glycome" (i.e., all carbohydrate structures) in a cell, tissue, or organism. In reality they have not been immune from the common pitfall of genomic and proteomic efforts that often generate vast amounts of data that cannot be interpreted meaningfully. Therefore, automated methods for predicting function, structure, and localization of newly discovered glycans and glycan-related enzymes alongside their arrays of references, composition, and spatial structures, and gathered NMR shift data[6] will be critical both for the glycobiology specialist undertaking further study of the intricacies of glycosylation and well as for the nonspecialist such as a physician who seeks to apply glycobiology-based technologies in the clinic. A few of these computer-based tools are listed in the next section, with a more detailed discussion provided in Chapter 16.

1.3.2.2. Computational Tools and Bioinformatics. In order to render meaningful the increasingly copious "glycomics" data that is being generated, computational methods capable of processing large amount of information are sorely needed and have been under development for over a decade. One example of an information handling technology is the automated conversion of mass spectrometry data into plausible glycan structures.[68,77] Some computer tools and modeling approaches focus on, or at least include, metabolic flux considerations required for nucleotide sugar production.[30,78–80] Other

approaches relate glycosylation enzyme activities, or their gene expression levels, to glycan structure based on a statistical analysis of the relationships between types of bonds each enzyme creates or cleaves with the abundances of the various bond types in each member of a large database of glycans.[81,82] For a quantitative understanding of how changes in enzyme activities affect the profile of glycan structures produced, models that include details of the cellular processes that govern glycosylation have been constructed.[67,68] To date, modeling attempts have progressed from small subsets of the glycan structures,[83] to specific enzymes such as those for sialylation,[84] *N*-linked glycans made by Chinese hamster ovary cells,[67] *O*-linked glycans associated with selectin ligands,[85] and a comprehensive model of mammalian *N*-linked glycans.[68] In the future, it is anticipated that fully integrated computer tools capable of simultaneously modeling all types of glycans will be developed.

1.3.3. Chemistry—Renewing Classic Techniques

1.3.3.1. Chemistry—A Valuable Contributor to Glycobiology. Chemistry has made critical contributions to the unraveling of the biology of sugars for over a century since Emil Fischer performed an elegant series of seminal experiments that described the isomeric nature of sugars and the stereochemical configuration of common monosaccharides. The contributions of chemistry continue today, as instrumentation derived from the chemical sciences including mass spectrometry (Chapter 11) and NMR has been invaluable in unraveling glycan structural complexity. Recently, synthetic approaches have caught up to analytical methods, with combinatorial biosynthesis (Chapter 10), synthetic carbohydrates and glycoconjugates (Chapter 12), and carbohydrate-based vaccines (Chapter 15) discussed in detail in this book with a brief overview of a few synthetic highlights provided here.

1.3.3.2. Fully Synthetic Glycans. An obvious attraction of a fully synthetic strategy is that chemically distinct glycan structures can be produced allowing evaluation of the biological response of an individual glycoform, rather than an averaged response obtained when testing a mixture of the profusion of glycoforms found in nature. Moreover, this ability in theory can be applied to vary rare structures predicted by a computer model but impossible to isolate in quantity from a natural source. To hypothetically illustrate how the ability to synthesize glycans could be applied to solve a vexing biomedical problem, prion proteins from diseased and healthy cells have different glycan profiles that have been implicated in disease progression.[5] This premise, however, has proved difficult to verify rigorously without the synthetic ability to produce the oligosaccharide chains, link them to the host protein, and finally assemble the entire GPI-anchored construct to thereby produce testable quantities of individual prion glycoforms; each of these steps is now possible in theory.

In practice, modern synthetic chemistry has been able to reproduce several glycan structures of considerable complexity and biomedical relevance. The pioneering example is the use of synthetic sialyl Lewis X for the treatment of reperfusion injury.[86] For the past decade much effort, facilitated by automated synthesis,[87] has focused on the creation of carbohydrate-based vaccines. It is possible to use synthetic carbohydrate analogs of viral

and microbial surface polysaccharides as vaccines to elicit an immune response against the microorganism. In fact, because an "artificial" polysaccharide can be carefully designed through precise synthesis, this type of vaccine may be both safer and more effective at lower dosage (i.e., through multivalency)[88,89] than a naturally derived vaccine, such as that of a live or killed microbe that contains a mixture of glycoforms, some of which may not be immunogenic.[90] Synthetic polysaccharide vaccines have been recently developed for several targets including *Haemophilus influenza* type b,[91] human immunodeficiency virus,[92] and various cancers.[93,94] In the future, as synthetic strategies are streamlined to become both technically effective and cost-effective, the possibilities of using carbohydrates to positively impact human health are numerous. For example, human breast milk contains a multitude of oligosaccharides that are distinct from other species such as the cow; human-specific milk sugars are both developmentally important and have activity against pathogens,[95] and the ability to supplement infant formula with these sugars would be valuable especially in the third world nations where malnutrition is endemic and infectious diseases are prevalent.

1.3.3.3. Synthesis—Toward Diversified Technologies. Steps toward solving a common limitation of conventional synthetic strategies—the insufficiently small amount of material obtained—are being taken by combining "one pot" synthetic strategies reported by the Wong laboratory[96,97] with automated synthesis being pioneered by the Seeberger group.[98,99] Although not capable of producing any glycan structure on demand as automated DNA synthesizers have long been able to do and protein synthesizers can do fairly adequately, automation provides a major boost toward several endpoints of major medical significance, including glycans associated with malaria and leishmaniasis that can potentially be exploited as potent vaccines.[87]

Regardless of whether a fully synthetic or automated strategy is used, certain glycosidic linkages that enzymes make with relative ease remain refractory to synthetic efforts; such recalcitrance has spurred efforts to combine more conventional methods with emerging chemoenzymatic transformations that use the suite of enzymes cells employ for glycosylation.[100] The much studied sialyl Lewis X tetrasaccharide provides a prime, and potentially biomedically important, example of combining enzymology with conventional synthesis.[101,102] Overall, hybrid approaches that combine chemoselective ligation methodology[103] with biological tools that facilitate programmable one-pot strategies[96] are proving to have remarkable versatility in the production of not just a carbohydrate moiety but rather an entire glycoprotein,[104,105] glycolipid,[106] or even GPI-anchored protein.[107]

1.3.4. Biological Approaches

1.3.4.1. Manipulating Glycans in Living Cells and Animals. In the chemoenzymatic approaches just mentioned, glycosylation enzymes are removed from their native milieu and exploited for cell-free synthesis. Taking this concept to the next level as the components of the glycosylation machinery have been elucidated, classic molecular biology techniques to overexpress or knockdown different enzymes and transporters are being used to manipulate glycan biosynthesis within living cells and even

in vivo. In many cases genetic approaches have led to predictable and versatile outcomes, such as the manipulation of α2,6-linked sialic acids or *N*-linked glycans and the resulting modulation integrin-mediated adhesion *in vitro*.[108–112] Moving to animal experiments, the situation is often much more complicated as the knockout of a single biosynthetic enzyme can result in manifold and severe effects such as early lethality[113] well beyond the effects expected from loss of the gene and ramifications for the targeted glycans. Other times, genetic approaches are vexing in the opposite direction, where the loss of a "critical" gene has no immediate consequence, as exemplified by the "αGal" knockout pig created to supply organs for xenotransplantation.[114] In this case, the removal of the key glycosyltransferase thought to be responsible for the antigenic αGal trisaccharide did not abolish production of the targeted oligosaccharide epitope.[115]

1.3.4.2. Retooling the Glycosylation Machinery in Cells. Not deterred by the often perplexing results of genetic regulation through the overexpression or knockdown of biosynthetic elements, researchers have begun to exploit modern molecular biology techniques for more subtle manipulation of the glycosylation process. For example, realizing that the biosynthesis of glycan structures relies on precisely localized enzymes within the ER and Golgi cisternae for proper assembly,[116] efforts have been made to relocalize glycosyltransferases and thereby alter the repertoire of oligosaccharide structures produced by a cell. One way that this localization is achieved is through the thickness of the membranes, which increase from the ER to the *cis*-, medial-, and *trans*-Golgi compartments; glycosyltransferase enzymes possess transmembrane domains of a length optimal to anchor them to a specific location in a cell's secretory organelles.[117] It is therefore possible to relocate an enzyme involved in glycosylation by swapping that enzyme's native transmembrane region with a transmembrane domain of a different length and thereby change substrate preference.[118] Similarly, the stem region located between the transmembrane and catalytic domains can also be swapped to tune the activity of a glycosyltransferase.[119] Although currently at a very early stage of development, these nascent laboratory tools point the way to a future where fine control over glycosylation may be possible by mixing and matching the membrane, stem, and catalytic domains of glycan-processing enzymes.

1.3.4.3. Lectins—An Example of Merging Biology and Technology. Lectins, a term that generically refers to proteins that recognize and bind carbohydrates, albeit with a broad range of affinities, avidities, and specificities,[120,121] provide an outstanding example of the use of biological tools in glycoscience and vividly illustrate how improvements in technology iteratively lead to a deepening biological understanding of glycans. As a specific example, lectins (with the term used broadly to include sugar-binding antibodies) have become valuable tools in array technologies.[76,122–128] To fully exploit lectins in an array context and properly interpret binding results, understanding the structural basis of lectin binding is helpful (Chapter 13) and conversely, results from the lectin array (as well as from glycan arrays probed by lectins[73]) can shed light on multivalency and carbohydrate binding[129] (Chapter 14). Ultimately, information gleaned from these experiments will be valuable for endeavors such as the creation of synthetic lectin mimics for artificial carbohydrate receptors (Chapter 12).

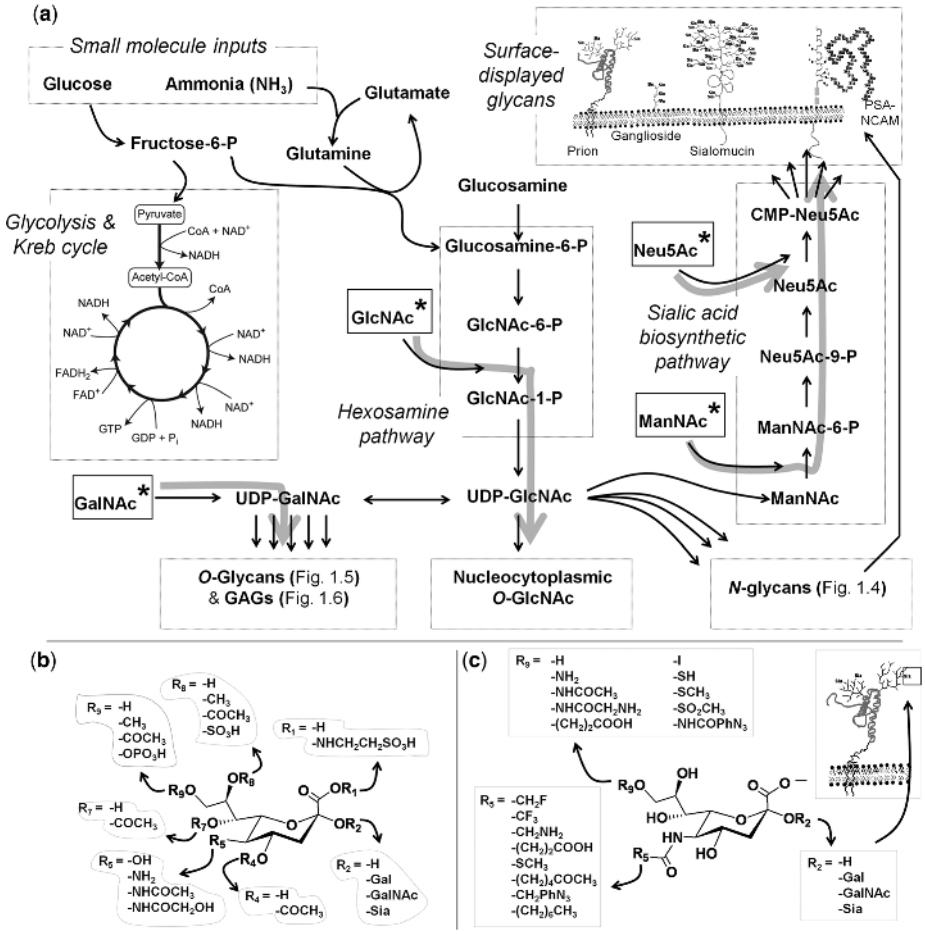

Figure 1.8. Metabolic flux and substrate-based glycoengineering. (**a**) Small molecules, primarily glucose, enter the carbohydrate-processing pathways of a mammalian cell and are primarily used for energy production via glycolysis and the Krebs cycle. A small proportion of hexose (e.g., fructose-6-phosphate) is diverted into the hexosamine pathway that supplies activated nucleotide sugars (such as UDP-GlcNAc) for glycan biosynthesis; this diversion of flux is promoted by exogenous ammonia. UDP-GlcNAc is a versatile intermediate that can be used for *O*-GlcNAcylation of nucleocytoplasmic proteins; participate in several stages of *N*-glycan production; be converted into UDP-GalNAc for use in *O*-glycan and GAG biosynthesis; and feed flux into the sialic acid pathway. Connections between intermediates represent one or more enzymes or transporters as indicated in the thin arrows; the majority of the estimated 225–250 mammalian glycosylation enzymes are not shown in this illustrative diagram (more detail is provided in online resources such as the KEGG database). Although not common in the diet, the amino sugars denoted by an asterisk (*) can be intercept glycosylation pathways through salvage or recycling mechanisms; this ability has been exploited in metabolic glycoengineering experiments. Interestingly, the non-natural analogs

1.3.5. Metabolites—An "Easy" Way to Manipulate Glycosylation

1.3.5.1. Glycosylation can be Altered Through Metabolic Intermediates. In theory, a very straightforward method to manipulate glycosylation is through the use of small molecules that alter metabolic flux; an interesting example of this concept is provided by reports that ammonia alters polysialic acid production.[130] Mechanistically, as outlined in Figure 1.8a, ammonia increases glutamine, the rate limiting supplier of amine for diversion of glucose (via fructose-6-phosphate), from energy production to the hexosamine pathway. Thus, ammonia enhances flux into the hexosamine pathway and increases *N*-glycan branching,[131] a result consistent with elevated UDP-GlcNAc and the requirement for progressively higher millimolar concentrations of this substrate for MGAT4 and 5 (the enzymes that initiate highly-branched *N*-glycans, see Fig. 1.4).[132] Interestingly, even though UDP-GlcNAc is the key supplier of flux into the sialic acid pathway via GNE,[133] polysialic acid levels actually decreased in ammonia-treated cells.[134] An explanation for this unexpected decline may lie in feedback inhibition loops that more than offset increased flux (e.g., UDP-GlcNAc inhibits glucosamine-6-phosphate conversion from hexose, and CMP-Neu5Ac inhibits ManNAc production from UDP-GlcNAc) resulting in inhibition of the final product.

In short, although incorporation of metabolites into glycosylation pathways can be employed toward defined ends (e.g., introduction of GlcNAc via salvage mechanisms into the hexosamine pathway reliably increases *N*-glycan branching), achieving a desired result is by no means assured with today's understanding. For every experiment that shows an expected or favorable result, such as the ability of oral fucose to ameliorate leukocyte adhesion deficiency (LAD) type II,[135] counterexamples arise, such as other manifestations of LAD that are refractory to fucose supplementation.[136] Complicating matters further, outcomes often vary dramatically from cell type to cell type or from species to species. On a hopeful note, in the future the computational methods now under development (mentioned above) may be valuable in rationally deciphering (and even predicting) the consequences of metabolic flux perturbation to glycosylation pathways and in designing supplementation strategies with greater precision than possible with today's largely trial-and-error methods.

1.3.5.2. Metabolic Glycoengineering—Biosynthetic Incorporation of Non-Natural Monosaccharide Analogs. "Metabolic glycoengineering" builds on the efforts discussed in the previous section to manipulate glycosylation with

Figure 1.8. (*Continued*) of GalNAc, GlcNAc, ManNAc, and Neu5Ac used in metabolic engineering do not ubiquitously partition throughout the "glycosylation machinery" but instead gain restricted entry into particular types of glycans as indicated by the broad gray arrows. (**b**) An example of naturally occurring metabolic glycoengineering is provided by sialic acid where over 50 natural variants of Neu5Ac—a sampling shown here—are used to modulate biological activity. (**c**) A sampling of abiotic sialic acids used in laboratory glycoengineering experiments are shown[137,148,149,157–161]; these analogs intercept the sialic acid pathway and are incorporated into cellular glycans in the place of natural sialosides.

exogenously supplied metabolites but does so with an interesting twist. This methodology, pioneered by the Reutter laboratory for sialic acid[137] and now extended to GalNAc[138,139] and GlcNAc,[140] is based on the remarkable ability of certain non-natural monosaccharide analogs to be metabolically incorporated into glycosylation pathways and replace the corresponding sugar residue in cell surface-displayed glycans. One aspect of this technology is that classes of analogs can be targeted into selected glycans with considerable precision; this is somewhat unexpected based on the aforementioned unruly effects of metabolites and the fact that labeled metabolites can broadly partition into numerous biomolecules. For example, GalNAc analogs can be converted to UDP-GalNAc by salvage or recycling mechanisms and then used by GalNAc transferases for incorporation into mucin-type *O*-glycans and possibly GAGs (Fig. 1.8a). However, UDP-GalNAc analogs are not converted to UDP-GlcNAc, or vice versa, even though the natural forms of these nucleotide sugars are readily epimerized to each other. Perhaps even more remarkably, salvaged GlcNAc analogs eschew all but one route into glycans, bypassing opportunities to enter *N*-linked glycans or be converted to ManNAc for sialylation or, returning the favor, not becoming UDP-GalNAc; instead selective incorporation in nucleocytosolic *O*-GlcNAc-modified proteins occurs.

To date—superseding GalNAc and GlcNAc (and fucose, which is not shown in Fig. 1.8)—the sialic acid pathway has come to exemplify metabolic glycoengineering. A bevy of ManNAc and sialic acid (Neu5Ac) analogs can intercept the pathway at earlier and later entry points and gain access to surface-displayed sialosides. To put laboratory efforts into context, there are over 50 naturally occurring chemical variants of sialic acid (a sampling of which are shown in Fig. 1.8b) that modulate biological activity as an innate form of metabolic glycoengineering.[141] Perhaps presumptuously, the glycochemist endeavors to improve on the natural forms by introducing his or her own repertoire of sialosides (a sampling of which are indicated in Fig. 1.8c for Neu5Ac; comparable panels of analogs for fucose, GalNAc, GlcNAc, and ManNAc are reviewed elsewhere[142–144]). By endowing glycans with novel chemical features, these analogs provide novel antiviral properties,[142] enhance immunogenicity,[145,146] modulate cell adhesion,[147] or control stem cell fate.[148] A subset of analogs bear chemical functional groups unique to the glycocalyx, such as ketones,[149] azides,[150] thiols,[148] or alkynes[151]; such sugars can act as "tags" for the delivery of genes,[152] toxins,[149] or imaging agents[153] by exploiting chemoselective ligation chemistry that has been developed to be compatible with physiological conditions.[103]

ACKNOWLEDGEMENTS

The authors would like to thank the National Institutes of Health for financial support (EB005692-03 and CA112314-04).

REFERENCES

1. Ito, S. *J. Cell Biol.* **1965**, *27*, 475–491.
2. Weinbaum, S.; Zhang, X.; Han, Y.; Vink, H.; Cowin, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7988–7995.

3. Gabius, H.-J.; Siebert, H.-C.; André, S.; Jiménez-Barbero, J.; Rüdiger, H. *ChemBioChem* **2004**, *5*, 740–764.
4. Endo, T.; Groth, D.; Prusiner, S. B.; Kobata, A. *Biochemistry* **1989**, *28*, 8380–8388.
5. Rudd, P. M.; Endo, T.; Colominas, C.; Groth, D.; Wheeler, S. F.; Harvey, D. J.; Wormald, M. R.; Serban, H.; Prusiner, S. B.; Kobata, A.; Dwek, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13044–13049.
6. von der Lieth, C. W.; Lutteke, T.; Frank, M. *Biochim. Biophys. Acta* **2006**, *1760*, 568–577.
7. Apweiler, R.; Hermjakob, H.; Sharon, N. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8.
8. Bell, G. I.; Burant, C. F.; Takeda, J.; Gould, G. W. *J. Biol. Chem.* **1993**, *268*, 19161–19164.
9. Joost, H.-G.; Bell, G. I.; Best, J. D.; Birnbaum, M. J.; Charron, M. J.; Chen, Y. T.; Doege, H.; James, D. E.; Lodish, H. F.; Moley, K. H.; Moley, J. F.; Mueckler, M.; Rogers, S.; Schürmann, A.; Seino, S.; Thorens, B. *Am. J. Physiol. Endocrinol. Metab.* **2002**, *282*, E974–E976.
10. Stuart, C. A.; Yin, D.; Howell, M. E.; Dykes, R. J.; Laffan, J. J.; Ferrando, A. A. *Am. J. Physiol. Endocrinol. Metab.* **2006**, *291*, E1067–E1073.
11. Wu, X.; Freeze, H. H. *Genomics* **2002**, *80*, 553–557.
12. Chen, H.; Wang, Z.; Sun, Z.; Kim, E. J.; Yarema, K. J. In *Handbook of Carbohydrate Engineering*; Yarema, K. J., Ed.; Taylor & Francis/CRC: Boca Raton, FL, **2005**; pp. 1–48.
13. Varki, A. *Am. J. Phys. Anthropol.* **2002**, *116*, 54–69.
14. Zachara, N. E.; Hart, G. W. *Biochim. Biophys. Acta* **2006**, *1761*, 599–617.
15. Berninsone, P. M.; Hirschberg, C. B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 542–547.
16. Pels Rijcken, W. R.; Hooghwinkel, G. J.; Ferwerda, W. *Biochem. J.* **1990**, *266*, 777–783.
17. Hirschberg, C. B. *J. Clin. Invest.* **2001**, *108*, 3–6.
18. Capasso, J. M.; Hirschberg, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 7051–7055.
19. Kamiyama, S.; Suda, T.; Ueda, R.; Suzuki, M.; Okubo, R.; Kikuchi, N.; Chiba, Y.; Goto, S.; Toyoda, H.; Saigo, K.; Watanabe, M.; Narimatsu, H.; Jigami, Y.; Nishihara, S. *J. Biol. Chem.* **2003**, *278*, 25958–25963.
20. Spiro, R. G. *Glycobiology* **2002**, *12*, 43R–56R.
21. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364–2369.
22. Freeze, H. H.; Aebi, M. *Curr. Opin. Struct. Biol.* **2005**, *15*, 490–498.
23. Dennis, J. W.; Granovsky, M.; Warren, C. E. *Biochim. Biophys. Acta* **1999**, *1473*, 21–34.
24. Helenius, J.; Aebi, M. *Semin. Cell. Dev. Biol.* **2002**, *13*, 171–178.
25. Bause, E. *Biochem. J.* **1983**, *209*, 331–336.
26. Burda, P.; Aebi, M. *Biochim. Biophys. Acta* **1999**, *1426*, 239–257.
27. Parodi, A. J. *Biochem. J.* **2000**, *348*, 1–13.
28. Kumar, R.; Yang, J.; Larsen, R. D.; Stanley, P. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9948–9952.
29. Tan, J.; D'Agostaro, G. A. F.; Bendiak, B.; Reck, F.; Sarkar, M.; Squire, J. A.; Leong, P.; Schachter, H. *Eur. J. Biochem.* **1995**, *231*, 317–328.
30. Lau, K. S.; Partridge, E. A.; Grigorian, A.; Silvescu, C. I.; Reinhold, V. N.; Demetriou, M.; Dennis, J. W. *Cell* **2007**, *129*, 123–134.
31. Brockhausen, I.; Narasimhan, S.; Schachter, H. *Biochimie* **1988**, *70*, 1521–1533.
32. Stanley, P. *Curr. Opin. Struct. Biol.* **2007**, *17*, 530–535.

33. Shpak, E.; Leykam, J. F.; Kieliszewski, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14736–14741.
34. Ten Hagen, K. G.; Fritz, T. A.; Tabak, L. A. *Glycobiology* **2003**, *13*, 1R–16R.
35. Hang, H. C.; Bertozzi, C. R. *Bioorg. Med. Chem.* **2005**, *13*, 5021–5034.
36. Gerken, T. A. *Biochemistry* **2004**, *43*, 4137–4142.
37. Perrine, C.; Ju, T.; Cummings, R. D.; Gerken, T. A. *Glycobiology* **2009**, *19*, 321–328.
38. Gerken, T. A.; Ten Hagen, K. G.; Jamison, O. *Glycobiology* **2008**, *18*, 861–870.
39. Torres, C. R.; Hart, G. W. *J. Biol. Chem.* **1984**, *259*, 3308–3317.
40. Zachara, N. E.; Hart, G. W. *Chem. Rev.* **2002**, *102*, 431–438.
41. Hakomori, S.-I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 225–232.
42. Simpson, M. A.; Cross, H.; Proukakis, C.; Priestman, D. A.; Neville, D. C. A.; Reinkensmeier, G.; Wang, H.; Wiznitzer, M.; Gurtz, K.; Verganelaki, A.; Pryde, A.; Patton, M. A.; Dwek, R. A.; Butters, T. D.; Platt, F. M.; Crosby, A. H. *Nat. Genet.* **2004**, *36*, 1225–1229
43. Michel, C.; van Echten-Deckert, G. *FEBS Lett.* **1997**, *416*, 153–155.
44. van Echten, G.; Sandhoff, K. *J. Biol. Chem.* **1993**, *268*, 5341–5344.
45. Degroote, S.; Wolthoorn, J.; van Meer, G. *Semin. Cell. Dev. Biol.* **2004**, *15*, 375–387.
46. Satoh, J. I.; Tai, T.; Kim, S. U. *Brain Res. Dev. Brain Res.* **1996**, *31*, 172–181.
47. Meng, X.-L.; Shen, J.-S.; Watabe, K.; Ohashi, T.; Eto, Y. *Mol. Genet. Metab.* **2005**, *84*, 332–343.
48. Maccioni, H. J. F.; Giraudo, C. G.; Daniotti, J. L. *Neurochem. Res.* **2002**, *27*, 629–636.
49. Low, M. G.; Finean, J. B. *Biochem. J.* **1977**, *167*, 281–284.
50. Futerman, A. H.; Low, M. G.; Ackermann, K. E.; Sherman, W. R.; Silman, I. *Biochem. Biophys. Res. Commun.* **1985**, *129*, 312–317.
51. McConville, M. J.; Menon, A. K. *Mol. Membr. Biol.* **2000**, *17*, 1–16.
52. Vidugiriene, J.; Menon, A. K. *J. Cell Biol.* **1994**, *127*, 333–341.
53. Chatterjee, S.; Mayor, S. *Cell. Mol. Life Sci.* **2001**, *58*, 1969–1987.
54. Su, B.; Waneck, G. L.; Flavell, R. A.; Bothwell, A. L. M. *J. Cell Biol.* **1991**, *112*, 377–384.
55. Ritter, T. E.; Fajardo, O.; Matsue, H.; Anderson, R. G.; Lacey, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3824–3828.
56. Esko, J. D.; Lindahl, U. *J. Clin. Invest.* **2001**, *108*, 169–173.
57. Hascall, V.; Esko, J. D. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Press: Woodbury, New York, **2009**; pp. 219–228.
58. Itano, N.; Kimata, K. *IUBMB Life* **2002**, *54*, 195–199.
59. Cohen, M.; Joester, D.; Geiger, B.; Addadi, L. *ChemBioChem* **2004**, *5*, 1393–1399.
60. Sasisekharan, R.; Venkataraman, G. *Curr. Opin. Chem. Biol.* **2000**, *4*, 626–631.
61. Kim, B.-T.; Kitagawa, H.; Tamura, J.-i.; Saito, T.; Kusche-Gullberg, M.; Lindahl, U.; Sugahara, K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7176–7681.
62. Sugahara, K.; Kitagawa, H. *Curr. Opin. Struct. Biol.* **2000**, *10*, 518–527.
63. Silbert, J. E.; Sugumaran, G. *IUBMB Life* **2002**, *54*, 177–186.
64. Funderbergh, J. L. *IUBMB Life* **2002**, *54*, 187–194.
65. Dove, A. *Nat. Biotechnol.* **2001**, *19*, 913–917.

66. Werz, D. B.; Ranzinger, R.; Herget, S.; Adibekian, A.; von der Lieth, C.-W.; Seeberger, P. H. *ACS Chem. Biol.* **2007**, *2*, 685–691.
67. Krambeck, F. J.; Betenbaugh, M. J. *Biotechnol. Bioeng.* **2005**, *92*, 711–728.
68. Krambeck, F. J.; Bennun, S. V.; Narang, S.; Choi, S. S.; Yarema, K. J.; Betenbaugh, M. J. *Glycobiology* **2009**, *19*, 1163–1175.
69. Naka, R.; Kamoda, S.; Ishizuka, A.; Kinoshita, M.; Kakehi, K. *J. Proteome Res.* **2006**, *5*, 88–97.
70. Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002**, *102*, 555–578.
71. Brewer, C. F.; Miceli, M. C.; Baum, L. G. *Curr. Opin. Struct. Biol.* **2002**, *12*, 616–623.
72. Kiessling, L. L.; Pohl, S. *Chem. Biol.* **1996**, *3*, 71–77.
73. Campbell, C. T.; Yarema, K. J. *Gen. Biol.* **2005**, *6*, Article 236.
74. von der Lieth, C. W.; Bohne-Lang, A.; Lohmann, K. K.; Frank, M. *Brief. Bioinform.* **2004**, *5*, 164–178.
75. Morelle, W.; Michalski, J.-C. *Curr. Pharm. Des.* **2005**, *11*, 2615–2645.
76. Pilobello, K. T.; Krishnamoorthy, L.; Slawek, D.; Mahal, L. K. *ChemBioChem* **2005**, *6*, 985–989.
77. Cooper, C. A.; Gasteiger, E.; Packer, N. H. *Proteomics* **2001**, *1*, 340–349.
78. Lambeth, M. J.; Kushmerick, M. J. *Ann. Biomed. Eng.* **2002**, *30*, 808–827.
79. Murrell, M. P.; Yarema, K. J.; Levchenko, A. *ChemBioChem* **2004**, *5*, 1334–1447.
80. Murrell, M. P.; Yarema, K. J.; Levchenko, A. In *Handbook of Carbohydrate Engineering*; Yarema, K., Ed.; Taylor & Francis/CRC: Boca Raton, FL, **2005**; Chapter 8.
81. Kawano, S.; Hashimoto, K.; Miyama, T.; Goto, S.; Kanehisa, M. *Bioinformatics* **2005**, *21*, 3976–3982.
82. Suga, A.; Yamanishi, Y.; Hashimoto, K.; Goto, S.; Kanehisa, M. *Genome Informatics* **2007**, *18*, 237–246.
83. Umana, P.; Bailey, J. E. *Biotechnol. Bioeng.* **1997**, *55*, 890–908.
84. Monica, T. J.; Andersen, D. C.; Goochee, C. F. *Glycobiology* **1997**, *7*, 515–521.
85. Liu, G.; Marathe, D. D.; Matta, K. L.; Neelamegham, S. *Bioinformatics* **2008**, *24*, 2740–2747.
86. Schermerhorn, M. L.; Nelson, D. P.; Blume, E. D.; Phillips, L.; Mayer, J. E.Jr *Ann. Thorac. Surg.* **2000**, *70*, 890–894.
87. Seeberger, P. H.; Werz, D. B. *Nat. Rev. Drug Discov.* **2005**, *4*, 751–763.
88. Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. *Curr. Opin. Chem. Biol.* **2000**, *4*, 696–703.
89. Ragupathi, G.; Koide, F.; Livingston, P. O.; Cho, Y. S.; Endo, A.; Wan, Q.; Spassova, M. K.; Keding, S. J.; Allen, J.; Ouerfelli, O.; Wilson, R. M.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2006**, *128*, 2715–2725.
90. Werz, D. B.; Seeberger, P. H. *Chem. Eur. J.* **2005**, *11*, 3194–3206.
91. Verez-Bencomo, V.; Fernández-Santana, V.; Hardy, E.; Toledo, M. E.; Rodríguez, M. C.; Heynngnezz, L.; Rodriguez, A.; Baly, A.; Herrera, L.; Izquierdo, M.; Villar, A.; Valdés, Y.; Cosme, K.; Deler, M. L.; Montane, M.; Garcia, E.; Ramos, A.; Aguilar, A.; Medina, E.; Toraño, G.; Sosa, I.; Hernandez, I.; Martínez, R.; Muzachio, A.; Carmenates, A.; Costa, L.; Cardoso, F.; Campa, C.; Diaz, M.; Roy, R. *Science* **2004**, *305*, 522–525.
92. Pashov, A. D.; Plaxco, J.; Kaveri, S. V.; Monzavi-Karbassi, B.; Harn, D.; Kieber-Emmons, T. *J. Biol. Chem.* **2006**, *281*, 29675–29683.

93. Deshpande, P. P.; Danishefsky, S. J. *Nature* **1997**, *387*, 164–166.
94. Keding, S. J.; Danishefsky, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11937–11942.
95. Hueso, P.; Martín-Sosa, S.; Martín, M.-J. In *Handbook of Carbohydrate Engineering*; Yarema, K. J., Ed.; Taylor & Francis/CRC: Boca Raton, FL, **2005**; pp. 141–175.
96. Koeller, K. M.; Wong, C. H. *Glycobiology* **2000**, *10*, 1157–1169.
97. Koeller, K. M.; Wong, C.-H. *Chem. Rev.* **2000**, *100*, 4465–4494.
98. Seeberger, P. H. *Chem. Commun.* **2003**, *21*, 1115–1121.
99. Seeberger, P. H.; Werz, D. B. *Nature* **2007**, *446*, 1046–1051.
100. Meldal, M.; St. Hilaire, P. M. *Curr. Opin. Chem. Biol.* **1997**, *1*, 552–563.
101. Cao, H.; Huang, S.; Cheng, J.; Li, Y.; Muthana, S.; Son, B.; Chen, X. *Carbohydr. Res.* **2008**, *343*, 2863–2869.
102. Bintein, F.; Auge, C.; Lubineau, A. *Carbohydr. Res.* **2003**, *338*, 1163–1173.
103. Lemieux, G. A.; Bertozzi, C. R. *Trends Biotechnol.* **1998**, *16*, 506–513.
104. Pratt, M. R.; Bertozzi, C. R. *Chem. Soc. Rev.* **2005**, *34*, 58–68.
105. Bernardes, G. J. L.; Castagner, B.; Seeberger, P. H. *ACS Chem. Biol.* **2009**, *4*, 703–713.
106. Pan, Y.; Chefalo, P.; Nagy, N.; Harding, C.; Guo, Z. *J. Med. Chem.* **2005**, *48*, 875–883.
107. Becker, C. F. W.; Liu, X.; Olschewski, D.; Castelli, R.; Seidel, R.; Seeberger, P. H. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 8215–8219.
108. Guo, P.; Zhang, Y.; Zhao, J.-h.; Wang, L.-y.; Guo, H.-b.; Zhang, X.-y.; Chen, H.-I. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 619–626.
109. Seales, E. C.; Jurado, G. A.; Brunson, B. A.; Bellis, S. L. *Glycobiology* **2003**, *13*, 860–861.
110. Gu, J.; Taniguchi, N. *Glycoconjug. J.* **2004**, *21*, 9–15.
111. Lin, S.; Kemmner, W.; Grigull, S.; Schlag, P. M. *Exp. Cell Res.* **2002**, *276*, 101–110.
112. Zhao, Y.; Sato, Y.; Isaji, T.; Fukuda, T.; Matsumoto, A.; Miyoshi, E.; Gu, J.; Taniguchi, N. *FEBS Journal* **2008**, *275*, 1939–1948.
113. Furukawa, K.; Sato, T. *Biochim. Biophys. Acta* **1999**, *1473*, 54–66.
114. Dai, Y.; Vaught, T. D.; Boone, J.; Chen, S.-H.; Phelps, C. J.; Ball, S.; Monahan, J. A.; Jobst, P. M.; McCreath, K. J.; Lamborn, A. E.; Cowell-Lucero, J. L.; Wells, K. D.; Colman, A.; Polejaeva, I. A.; Ayares, D. L. *Nat. Biotechnol.* **2002**, *20*, 251–255.
115. Milland, J.; Christiansen, D.; Sandrin, M. S. *Immunol. Cell Biol.* **2005**, *83*, 687–693.
116. Colley, K. J. *Glycobiology* **1997**, *7*, 1–13.
117. Sasai, K.; Ikeda, Y.; Tsuda, T.; Ihara, H.; Korekane, H.; Shiota, K.; Taniguchi, N. *J. Biol. Chem.* **2001**, *276*, 759–765.
118. de Graffenried, C. L.; Bertozzi, C. R. *J. Biol. Chem.* **2003**, *278*, 40282–40295.
119. de Graffenried, C. L.; Bertozzi, C. R. *J. Biol. Chem.* **2004**, *279*, 40035–40043.
120. Dam, T. K.; Brewer, C. F. *Chem. Rev.* **2002**, *102*, 387–429.
121. Laederach, A.; Reilly, P. J. *J. Comput. Chem.* **2003**, *24*, 1748–1757.
122. Zheng, T.; Peelen, D.; Smith, L. M. *J. Am. Chem. Soc.* **2005**, *12728*, 9982–9983.
123. Hirabayashi, J. *Glycoconjug. J.* **2004**, *21*, 35–40.
124. Hsu, K.-L.; Mahal, L. K. *Nat. Protoc.* **2006**, *1*, 543–549.
125. Hsu, K.-L.; Pilobello, K. T.; Mahal, L. K. *Nat. Chem. Biol.* **2006**, *2*, 153–157.
126. Tao, S. C.; Li, Y.; Zhou, J.; Qian, J.; Schnaar, R. L.; Zhang, Y.; Goldstein, I. J.; Zhu, H.; Schneck, J. P. *Glycobiology* **2008**, *18*, 761–769.

127. Hsu, K.-L.; Gildersleeve, J. C.; Mahal, L. K. *Mol. Biosyst.* **2008**, *4*, 654–662.
128. Hirabayashi, J. *J.Biochem.* **2008**, *144*, 139–147.
129. Manning, J. C.; Seyrek, K.; Kaltner, H.; André, S.; Sinowatz, F.; Gabius, H. J. *Histol. Histopathol.* **2004**, *19*, 1043–1060.
130. Zanghi, J. A.; Mendoza, T. P.; Schmelzer, A. E.; Knop, R. H.; Miller, W. M. *Biotechnol. Prog.* **1998**, *14*, 834–844.
131. Gawlitzek, M.; Valley, U.; Wagner, R. *Biotechnol. Bioeng.* **1998**, *57*, 518–528.
132. Lau, K. S.; Dennis, J. W. *Glycobiology* **2008**, *18*, 750–760.
133. Keppler, O. T.; Hinderlich, S.; Langner, J.; Schwartz-Albiez, R.; Reutter, W.; Pawlita, M. *Science* **1999**, *284*, 1372–1376.
134. Zanghi, J. A.; Mendoza, T. P.; Knop, R. H.; Miller, W. M. *J. Cell. Physiol.* **1998**, *177*, 248–263.
135. Marquardt, T.; John, K.; Freeze, H. H.; Harns, E.; Vesteweber, D. *Blood* **1999**, *94*, 3976–3985.
136. Sturla, L.; Puglielli, L.; Tonetti, M.; Berninsone, P. M.; Hirschberg, C. B.; De Flora, A.; Etzioni, A. *Pediatr. Res.* **2001**, *49*, 537–542.
137. Kayser, H.; Zeitler, R.; Kannicht, C.; Grunow, D.; Nuck, R.; Reutter, W. *J. Biol. Chem.* **1992**, *267*, 16934–16938.
138. Hang, H. C.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 1242–1243.
139. Dube, D. H.; Prescher, J. A.; Quang, C. N.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4819–4824.
140. Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13132–13137.
141. Angata, T.; Varki, A. *Chem. Rev.* **2002**, *102*, 439–469.
142. Keppler, O. T.; Horstkorte, R.; Pawlita, M.; Schmidt, C.; Reutter, W. *Glycobiology* **2001**, *11*, 11R–18R.
143. Campbell, C. T.; Sampathkumar, S.-G.; Weier, C.; Yarema, K. J. *Mol. Biosyst.* **2007**, *3*, 187–194.
144. Aich, U.; Yarema, K. J. In *Glycosciences*; 2 ed.; Fraser-Reid, B., Tatsuta, K., Thiem, J., Eds.; Springer-Verlag: Berlin, **2008**; pp. 2136–2190.
145. Liu, T.; Guo, Z.; Yang, Q.; Sad, S.; Jennings, H. J. *J. Biol. Chem.* **2000**, *275*, 32832–32836.
146. Wang, Q.; Zhang, J.; Guo, Z. *Bioorg. Med. Chem.* **2007**, *15*, 7561–7567
147. Villavicencio-Lorini, P.; Laabs, S.; Danker, K.; Reutter, W.; Horstkorte, R. *J. Mol. Med.* **2002**, *80*, 671–677.
148. Sampathkumar, S.-G.; Li, A. V.; Jones, M. B.; Sun, Z.; Yarema, K. J. *Nat. Chem. Biol.* **2006**, *2*, 149–152.
149. Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. *Science* **1997**, *276*, 1125–1128.
150. Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007–2010.
151. Sawa, M.; Hsu, T.-L.; Itoh, T.; Sugiyama, M.; Hanson, S. R.; Vogt, P. K.; Wong, C.-H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12371–12376.
152. Lee, J. H.; Baker, T. J.; Mahal, L. K.; Zabner, J.; Bertozzi, C. R.; Wiemar, D. F.; Welsh, M. J. *J. Biol. Chem.* **1999**, *274*, 21878–21884.
153. Lemieux, G. A.; Yarema, K. J.; Jacobs, C. L.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 4278–4279.

154. Hashimoto, K.; Goto, S.; Kawano, S.; Aoki-Kinoshita, K. F.; Ueda, N.; Hamajima, M.; Kawasaki, T.; Kanehisa, M. *Glycobiology* **2006**, *16*, 63R–70R.

155. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K. F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. *Nucleic Acids Res.* **2006**, *34*, D354–357.

156. Esko, J. D.; Kimata, K.; Lindahl, U. In *Essentials of Glycobiology*; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Press: Woodbury, New York, **2009**; pp. 229–248.

157. Gross, H. J. *Eur. J. Biochem.* **1992**, *203*, 269–275.

158. Jacobs, C. L.; Goon, S.; Yarema, K. J.; Hinderlich, S.; Hang, H. C.; Chai, D. H.; Bertozzi, C. R. *Biochemistry* **2001**, *40*, 12864–12874.

159. Oetke, C.; Brossmer, R.; Mantey, L. R.; Hinderlich, S.; Isecke, R.; Reutter, W.; Keppler, O. T.; Pawlita, M. *J. Biol. Chem.* **2002**, *277*, 6688–6695.

160. Kim, E. J.; Sampathkumar, S.-G.; Jones, M. B.; Rhee, J. K.; Baskaran, G.; Yarema, K. J. *J. Biol. Chem.* **2004**, *279*, 18342–18352.

161. Han, S.; Collins, B. E.; Bengtson, P.; Paulson, J. C. *Nat. Chem. Biol.* **2005**, *1*, 93–97.