1.1 INTRODUCTION

This chapter takes on the ambitious task of introducing two immensely complicated biological systems. By necessity, only the briefest outline of each topic will be provided while maintaining a focus on the surprisingly robust, albeit often overlooked, overlap between the two areas. The term glycobiology was first formally used two decades ago by Rademacher, Parekh, and Dwek to describe the merging of the traditional disciplines of carbohydrate chemistry, biochemistry, and cell biology [1]. In reality, the detailed study of biologically important sugars greatly predated the formal recognition of glycobiology as a distinct field of investigation as—almost a century earlier—Emil Fischer performed an elegant series of seminal experiments that described the isomeric nature of sugars and the stereochemical configuration of common monosaccharides. In the intervening years, carbohydrates have been established as the most abundant—and arguably the most structurally diverse—organic molecules found in nature. They play major structural roles in fungi, crustaceans, and plants and are often subject to postsynthetic modifications, especially in higher animals, that greatly increase their chemical diversity and biological activities.

In mammals, carbohydrates are quantitatively less abundant when measured by bulk mass compared to many lower organisms, and they are also skewed toward a lower size distribution, often occurring as oligosaccharide structures of 20 or fewer—sometimes
only as 1 or 2—residues instead of as large polysaccharides. More than compensating for their modest size, however, mammalian oligosaccharides have critical biological functions derived from fundamental differences compared to other classes of biomacromolecules such as nucleic acids, proteins, and lipids. For example, the monomeric units of carbohydrates can connect to each other by several different linkages, resulting in branched structures that enable even relatively small oligosaccharides to exist in a profuse number of structural variations. Consequently, these molecules have immense information-carrying capacity so that, even when they are present in vanishingly small quantities, they have a profound impact on modulating the function of their host protein or lipid. With a growing arsenal of technical tools at their disposal to decipher the complexities of glycosylation with increasing precision, glycobiologists are now well positioned to tackle a formidable biological problem—unraveling the intersection of carbohydrates with immunity.

The realms of glycobiology and immunology have long overlapped in the form of the ABO(H) blood group antigens that were discovered in 1900 [2]. In the past two decades the molecular basis of these antigenic structures, which necessitate careful attention to blood-type compatibility during transfusions, have been found be an oligosaccharide structure where a small structural difference—the presence or absence of an N-acetyl group on galactose—separates the A and B epitopes. The ability of the immune system to recognize minor changes to the chemical structures of large carbohydrate structures has profound biomedical implications that go beyond the largely solved problem of blood typing in transfusions. For example, the trisaccharide α-gal antigen leads to hyperacute rejection of xenotransplanted organs and has spawned ongoing efforts to create α-gal knockout pigs [3]. Similarly, the antigenic Neu5Gc form of sialic acid has recently been found to “contaminate” human stem cell lines, raising concerns about the introduction of tissue-engineered organs into a recipient [4]. In contrast to these examples, in this book the impingement of the immune system into glycobiology is not regarded as a problem but rather as an enticing opportunity. In short, based on the century-old precedent that carbohydrates are antigenic, the logical—although not simple (hence, the need for an entire book on the topic!)—course of action is to exploit complex sugars as vaccines and for immunotherapy.

To provide a broader context for the subsequent chapters in this book, which delve into the nitty-gritty aspects of carbohydrate-based vaccine development, this chapter gives an overview of glycobiology and then presents a sampling of specific examples that exemplifies how this field connects with immunology. First, the major classes of mammalian carbohydrates, along with brief descriptions of their biosynthesis, are covered in Section 1.2. Next, in Section 1.3, an even briefer overview of immunology is provided along with a peek into several aspects of glycoimmunobiology, a field that is emerging as it becomes increasingly clear that links between glycobiology and immunology are a two-way street. Not only does the immune system reach into the realm of glycobiology by recognizing carbohydrate as antigens, but glycobiology also intrudes into immunology to the extent that glycans influence multiple levels of the immune response. Accordingly, selected specific examples of how carbohydrates tune the immune response will be given to provide a small window into the biological roles of complex sugar structures (i.e., into glycobiology). Finally, in Section 1.4 a brief
overview of antigenic carbohydrate structures found in nature that hold potential value for vaccine development will be surveyed along with a brief discussion of unique challenges faced by—if we continue the trend of dubbing sugar-related areas of investigation with the glyco prefix—the “glycovaccinist.”

1.2 GLYCOBIOLOGY

1.2.1 Glycosylation—Is It Worth the Cost?

1.2.1.1 Basic Considerations A newcomer to the field of glycobiology might ask the following question: Why do cells bother with glycosylation? This question arises from a quick accounting of the costs in energy and resources expended by a cell to produce glycans that include the hundreds of proteins that comprise the biosynthetic glycosylation machinery, the requirement for high-energy nucleotide sugars used as building blocks for multimeric carbohydrate structures, as well as the energy foregone by not using sugars for their “canonical” biological function—energy production! Even worse, after expending all of this effort, surface-displayed glycans can be co-opted by opportunistic molecular toxins, viral and microbial pathogens, and eukaryotic parasites (see Fig. 1.1 and Section 1.2.1.3); furthermore, the complexities of glycosylation provide ample opportunity for metabolic aberrations.

Figure 1.1 Landscape of the cell is dominated by carbohydrates. The surface of a mammalian cell is decorated by various classes of complex carbohydrates; these oligosaccharides are often referred to as glycans and collectively as the glyocalyx. An overview of glycan biosynthesis is provided in Figures 1.2–1.4 and more information on each class of these molecules is given in Figures 1.5–1.7 and the accompanying text. This graphic is not drawn to scale (as is evident from the relative sizes of toxins, antibodies, viruses, and bacteria, which are all pathogens that exploit surface sugars for entry into a cell). It is notable that even though the glyocalyx comprises only about 8–10% of the mass of the plasma membrane, in a typical mammalian cell it forms a continuous (albeit not uniform) layer ~8 nm thick, occupying roughly the same volume as the lipid and protein constituents of the membrane. (See color insert.)
to arise that cause or exacerbate disease (Section 1.2.1.2). Clearly, against this rather
dire backdrop, cells must have a compelling reason to produce glycans. As depicted in
Figure 1.1, and discussed in detail for specific aspects of the immune system in Section
1.3.3, surface sugars comprise the interface between a cell and its outside environment
and in a very real sense allow multicellular life to exist by enabling not only the detri-
mental pathogen binding events shown in the figure but also by facilitating cell–cell
and cell–extracellular matrix (ECM) interactions, serving as receptors for hormones
and lectins, and contributing to cell motility [5]. In the overall balance, after playing
with sugars during hundreds of millions of years of evolution, nature has decided that
they are well worth the pitfalls they create; we trust that this chapter will lead the reader
toward a similar conclusion.

1.2.1.2 Glycans in Disease Even before getting into glycosylation per se
where the complexities of the assembly of complex saccharide structures come into
play, simply having sugars around is a dicey—albeit an absolutely necessary—
proposition for a living organism. Indirectly, as evidenced by the large quantities of
antioxidants consumed by the general public as nutritional supplements, oxygen
used to liberate energy from sugars can be highly lethal to cells if reactive free radical
by-products are not tightly controlled. Less well known, but just as insidious, the
simple sugars that are the feedstock for energy-providing oxidation reactions can
damage biological macromolecules through “advanced glycation end products”
(AGE). From a chemical perspective, the nonenzymatic reaction between amino
acids and reducing sugars was discovered by L. C. Maillard a century ago. The indi-
vidual steps of the Maillard reaction have now been unraveled in detail starting with
the formation of reversible Schiff bases, which are transformed into more stable
aldoamines (Amadori products) or ketoamines (Heyns products) [6]. Following a
chain of chemical rearrangements, the latter products are converted into terminal
adducts that are irreversibly bound to the target molecules as AGEs. During the
1970s and 1980s, AGE products—present at particularly high levels in diabetes,
renal failure, and amyloidosis because of increased levels of circulating sugar [7]—
were implicated in disease complications primarily via adventitious crosslinking of
proteins. Collagen, for example, is a prime target of AGE reactions, and the results
are especially damaging to the vascular system. More recently, and appropriate to
mention in a combined discussion of glycobiology and immunology, evidence is
emerging that AGEs have antigenic properties leading to the hypothesis that AGE
structures found in vivo may elicit autoimmune responses that contribute to the athero-
genic processes associated with diabetes [8].

In addition to the hazards cells encounter by using sugars as an energy source, or
simply by having them around, the assembly of monosaccharides into complex struc-
tures leads to a whole new set of pitfalls that are manifest at various levels of severity.
At the most critical level, genes involved in the production of certain glycan structures
are absolutely essential for life—for example, aberrations involved in sugar structures
needed during fertilization would prevent formation of viable embryos from the very
start of an organism’s life span [9]. In other cases, exemplified by congenital disorders
of glycosylation that arise from defects in N-glycan biosynthesis (see Section 1.2.3.2),
disease—which is often fatal—is clinically manifest early in childhood [10]. Other
times, illustrated by mutations to glucosamine (UDP-N-Acetyl)-2-epimerase/
N-acetylmannosamine kinase (GNE) (the key biosynthetic enzyme in sialic acid
biosynthesis [11]) found in hereditary inclusion body myopathy (HIBM [12]), symp-
toms do not appear for decades and are not found in all individuals afflicted with
the disease-causing mutation. Finally, continuing to move toward the benign end of
the severity spectrum, certain congenital genetic abnormalities—demonstrated by a
lack of phenotype in knockout mice—have no impact on development and any
phenotypic effects that do occur are very subtle and are often only manifest in complex
behavioral traits [13].

Glycosylation abnormalities acquired later in life upon somatic mutation also
contribute to disease; the outstanding example is cancer, which is virtually always
accompanied by aberrant glycan production [14–17]. Nascent efforts to exploit
glycans in cancer immunotherapy are described briefly in Section 1.4 and in detail else-
where in this book (see Chapters 8–11). Another interesting example of glycosylation
in disease is provided by prions, a novel class of “protein-only” infectious pathogens
that cause a group of invariably fatal neurodegenerative diseases. Although the protein-
only description specifically refers to the surprising absence of nucleic acids found in
conventional infectious agents, it is not strictly correct insofar as the prion protein has
two highly conserved potential sites of N-glycosylation, allowing prion proteins to
exist as three classes of molecules that differ in their degree of glycosylation. The
unglycosylated (5%) and monoglycosylated (25%) isoforms are minor cell surface
components dominated by the diglycosylated form (70%). Although the roles that
N-linked glycans play in prion biology are not completely understood, a comparative
analysis of the glycans on healthy and diseased proteins reveal differences that include
the proportion of tri- and tetra-antennary structures as well as the amount of Lewis
X (LeX) and sialyl Lewis X (sLeX) epitopes [18–21]. Ultimately, differences in
glycan profiles may prove to be the critical determinant of prion disease progression
and the prefix glyco will need to be added to the protein-only descriptor.

1.2.1.3 Toxins and Infectious Agents Bind to Surface Glycans

Because they are easily accessible, sugars displayed on the surfaces of mammalian cells
represent enticing binding opportunities for opportunistic pathogens ranging from
molecular toxins to viruses and from primitive bacteria to sophisticated eukaryotic
parasites. To give representative examples, at the molecular level ricin—a versatile
and durable toxin (allegedly) used by KGB assassins during the cold war and by
terrorists today—consists of two peptide chains crosslinked by a disulfide bond.
One of the protein chains is a lectin (lectin is a generic name for a protein that
binds to a carbohydrate) that recognizes terminal galactose on cell surface glycans
and serves to ferry the other peptide chain into the lumen of secretory vesicles
where a single molecule, upon translocation into the cytosol, can kill the cell via cat-
alytic deactivation of ribosomes [22, 23]. Viruses also ubiquitously exploit surface
glycans as binding epitopes; in a well-known example, influenza virus binds to
sialic acid and does so with remarkable discrimination by distinguishing between
α2,3-linked and α2,6-linked sialosides as well as eschewing the Neu5Gc form of
sialic acid in preference to Neu5Ac [24–27]. Moving to microbes, many pathogenic bacteria bind to host cells through glycan recognition. A particularly interesting example is the adherence of *Escherichia coli* to epithelial cells of the gastrointestinal or urinary tract mediated by bacterial lectins present on fimbriae. These lectins preferentially bind to cell surface glycoproteins containing mannose and have counterintuitive—but very useful from the bug’s point of view—catch bond behavior characterized by tighter binding as shear force increases [28, 29]. Among other benefits, this clever binding mode prevents elimination of the infecting bacterium during urination while allowing it to swim free in the bladder at other times.

In addition to comprising easily accessible binding epitopes, the information-rich “sugar code” provided by glycan diversity [30] contributes to host–pathogen specificity. For example, the gonorrhea organism *Neisseria gonorrhoeae* has long been known to adhere to human cells of the genital and oral epithelia but not to cells from other organs or other animal species, explaining why only humans are prone to gonorrhea. The molecular basis for host specificity of *N. gonorrhoeae* is in part explained by glycosylation patterns of CD66 [31] and reaches even greater importance in the exquisite sensitivity of influenza viral strains for different sialic acids, as mentioned above. For the nonspecialist to make sense of jargon such as α2,3-linked and α2,6-linked sialosides, to comprehend the subtle but important differences between similar sugars such as Neu5Ac and Neu5Gc that many pathogens can distinguish with precision, or to understand why a pathogenic bacteria would opt for mannose as a binding epitope, it is helpful to take a closer look at the biosynthetic processes for mammalian glycans and understand the chemical basis of the “building blocks” and the structural ramifications of how monosaccharides are assembled into complex structures. An investment in time to understand these issues (discussed in Sections 1.2.2–1.2.4) will benefit any researcher working with carbohydrate–based vaccines because the mammalian immune system is equally attuned to the subtleties of carbohydrate structure as the sampling of pathogens just mentioned.

### 1.2.2 Glycan Biosynthesis—A Dauntingly Complex Process

As a consequence of their structural complexity and ubiquitous nature, carbohydrates play important roles in many physiological and pathological cellular functions. Glycobiologists study these biological functions but are also keenly interested in the biosynthetic processes through which complex sugars are assembled. A molecular-level understanding of the production of glycans has greatly lagged nucleic acid and proteins because, unlike these other biopolymers, carbohydrate structures are not template based. An estimated 1–3% of the human genome as well as dozens of metabolic intermediates are involved in the biosynthetic process; furthermore, oligosaccharide synthesis and modification is spatially located across several subcellular compartments; these complexities have rendered the biosynthesis of glycans a substantial challenge to unravel.

Despite formidable challenges posed by the structural complexity and less than straightforward biosynthetic routes of glycans, substantial progress in several disciplines has converged over the past few years and has contributed to a greatly increased understanding of the assembly and identification of oligosaccharides [38, 39].
A critical advance was the sequencing of the human genome, which allowed all of the genes involved in glycosylation—that code for various enzymes and membrane transporters—to be at least tentatively identified. In tandem, increasingly sensitive analytic techniques are now readily available that allow intracellular metabolites and mature glycan structures to be characterized even in minute quantities \[40, 41\]. Efforts are also underway to develop computational models that use intracellular information (i.e., the genomic and small-molecule metabolite compositions) to predict surface glycan and to furthermore connect these parameters with development status, a disease state, or environmental insult \[42–45\].

The production of complex carbohydrates starts with basic building blocks—a set of monosaccharides—that are analogous to the nucleotides or amino acids used for nucleic acids or proteins biosynthesis, respectively. The sugary feedstock for glycan assembly is primarily glucose, which can be obtained from the diet, transported into a cell, converted into other monosaccharides through epimerization (and other) reactions, phosphorylated, and ultimately converted into a high-energy or “activated” nucleotide sugar donor. A summary of the reactions that convert monosaccharides obtained from the extracellular milieu into nucleotide sugar donors is shown in Figure 1.2; chemical structures of common mammalian monosaccharides and nucleotide sugars are provided in Figure 1.3.

The monosaccharide processing reactions shown in Figure 1.2 primarily occur in the cytosol, setting the stage for oligosaccharide assembly that can begin on the cytosolic face of the endoplasmic reticulum (ER), but more commonly the nucleotide sugars are transported into the lumen of the ER or a Golgi compartment where the majority of glycosyltransferases are localized. Glycosyltransferase reactions that assemble nucleotide sugars into multimeric macromolecules—exemplified by the suite of sialyltransferases shown in Figure 1.4—add another substantial level of complexity to the glycosylation machinery. In humans there are 20 different sialyltransferases that create 6 distinct types of glycosidic linkages (i.e., $\alpha 2,3$, $\alpha 2,6$, and $\alpha 2,8$ on either protein or lipid-attached glycans), thus providing redundancy that depends on factors such as developmental stage or tissue type. Once the mature glycan is synthesized, a process that involves considerable nuance depending on whether the sugar structure is attached to a protein, lipid, glycosylphosphatidylinositol (GPI) anchor, or exists as a “free” polysaccharide (as described in the next sections), the glycosylated macromolecule is almost always secreted or moved to the cell surface; the major exception are $O$-GlcNAc-modified cytosolic and nuclear proteins. Finally, scavenging of glycans and recycling of their molecular components occurs through the action of glycosidases that degrade oligo- and polysaccharides and thereby liberate monosaccharides for reuse.

### 1.2.3 Glycoproteins

#### 1.2.3.1 Glycosylation Is a Ubiquitous and Diverse Co- and Posttranslational Protein Modification

In terms of diversity of structures as well as the sheer proportion of molecules affected, glycosylation is the most significant co-translational and posttranslational modification of proteins. Like all complex
Figure 1.2 Overview of glycan biosynthesis. (a) Monosaccharide uptake and processing (see Fig. 1.3 for structures). Sugars obtained exogenously by cells such as Gal, Glc, and GlcN (full names and chemical structures of common mammalian monosaccharides are shown in Fig. 1.3a) are taken up by families of membrane transporters [32, 33] and converted to nucleotide sugar donors such as UDP-GlcNAc and CMP-Neu5Ac (Fig. 1.3c) by mostly cytosolic enzymatic reactions. Additional information on the enzymes is provided in other review articles [34] or from online search engines or databases such as Pubmed [35], KEGG [36], or HUGO [37]. (b) Glycan assembly (see Fig. 1.4 for details of sialylation). The nucleotide sugars are assembled by a suite of glycosyltransferases (illustrated in detail for sialyltransferases in Fig. 1.4) into structurally complex surface-displayed glycans. (c) Representative glycans include GPI-anchored prions (see Fig. 1.7b, which in turn bear N-linked glycans, Fig. 1.5), the glycosphingolipid ganglioside GM3 (see Figs. 1.7a and 1.10e), and CD34, a mucin-type glycoprotein that bears numerous O-linked glycans that often include TACAs (see Figs. 1.6a and 1.6b). (See color insert.)
Figure 1.3 Chemical structures of monosaccharides, nucleotide sugars, and an explanation of \( \alpha/\beta \) glycosidic linkages. (a) The 10 monosaccharides found in human glycans are shown; GlcA and IdoA are found exclusively in GAGs (see Fig. 1.6e), and the Neu5Gc form of sialic acid is found in animals other than humans and chickens. (b) Monosaccharides are converted to one of three types of nucleotide sugars, based either on UDP, GDP, or CMP. (c) Glucosamine, a sugar not found in human glycans without further modification (i.e., conversion to GlcNAc; see Figs. 1.2 and 1.4), is used as an example to show ring numbering and \( \alpha/\beta \) linkage (in general \( \beta \) linkages are equatorial and \( \alpha \) linkages are axial, except for sialic acid, which only has equatorial glycosidic linkages that are designated \( \alpha \); see Fig. 1.4). (d) Specific examples of glycosidic linkages are illustrated by the oligosaccharide structures comprising the ABO blood group antigens discussed in the Introduction.
Figure 1.4 Details of sialylation. Sialic acid is one of the ~10 sugars found in mammalian glycans and can be produced from glucose or glucosamine (GlcN; see Figs. 1.2 and 1.3) obtained from exogenous sources that enter the hexosamine pathway [46]. (a) Passage of Glc through the hexosamine pathway results in the production of UDP-GlcNAc, which can be used directly for glycan assembly (or converted to UDP-GalNAc and used in a likewise manner) or for O-GlcNAc protein modification [47, 48]. (b) Another fate for UDP-GlcNAc is conversion to N-acetyl-D-mannosamine (ManNAc) by the bifunctional enzyme GNE [11]; this sugar does not appear in mammalian glycans but is instead converted to Neu5Ac [49] and installed into glycans by a set of 20 sialyltransferases (in humans) that have various overlapping linkage and substrate specificities (top). (c) In the predominant example of “metabolic glycoengineering,” metabolic flux is supplied into the sialic acid pathway by nonnatural ManNAc analogs bearing abiotic “R” groups at the N-acyl position. These modifications include extended alkyl chains [50] and various chemical functional groups not usually found in sugars such as thiols, ketones, azides, or alkynes [51]. These analogs transit the biosynthetic pathway and appear in mature glycans in place of the Neu5Ac form of sialic acid most commonly found in humans.
carbohydrates, the glycan chains of glycoproteins are biosynthesized by the concerted action of a set of enzymes, rather than on the basis of a template akin to the way nucleotide sequence specifies primary amino acid structure. Instead, the final composition of a glycan is determined by multiple factors such as peptide sequence of the protein undergoing glycosylation, the availability of substrates in various subcellular locales, which in turn is determined by the expression and activity of membrane transporters, the localization of the particular glycosyltransferases to certain regions of the ER or Golgi, and competition between glycosyltransferases and glycosidases. This complex and indirect biosynthetic process results in heterogeneous glycosylation at two levels. First, any potential site of glycosylation—that is, the side chain of a candidate amino acid—may or may not be occupied. Different proteins vary considerably in the number of potential glycosylation sites—ranging from two sites where \( N \)-linked glycans can be attached to the prion protein to 26 possible sites for \( \alpha_3\beta_1 \) or \( \alpha_5\beta_1 \) integrin dimers [52]. Mucin-type proteins heavily invested with \( O \)-linked glycans—such as CD34 (see Fig. 1.2)—can have dozens of sites where sugars are attached to the peptide backbone [53]. It is clear that—even if the same oligosaccharide structure was attached, or not attached, to each site—a greatly diverse pool of glycoproteins would exist for each primary gene sequence.

The situation becomes exponentially more complex because each site of glycosylation can potentially be endowed with any one of dozens of different glycan structures. These range from a single monosaccharide to complex branching oligosaccharides of 20 or more residues in size to—more rarely because they are generally not covalently linked to surface elements—polysaccharides hundreds of residues in length. The diversity of glycans that can occur at each glycosylation site is referred to as microheterogeneity [54] and a “simple” case of microheterogeneity is provided by the prion protein that has two sites for \( N \)-linked glycan attachment. With 50 different glycan structures available for display at each site [21, 55], a prion can exist as \( \sim 50^2 \) or \( \sim 2500 \) distinct chemical entities. Clearly, this number is markedly greater than the three classes (un-, mono-, and di-glycosylated) of prions referred to earlier, making it transparent how sufficient diversity exists to allow significantly different profiles of glycans to be present on healthy and scrapie prion proteins [18].

Applying a similar analysis to CD34, each copy of this molecule in the human body could be chemically distinct; in fact, “ballpark” calculations indicate that each human CD34 that has ever existed could be unique. To elaborate briefly, if all of a person’s approximately one trillion leukocytes bear an upper estimate of \( \sim 100,000 \) copies of CD34, a human body has \( \sim 10^{17} \) of this molecule. Over a lifetime, assuming a turnover of once a day, a person would produce \( \sim 2.5 \times 10^{21} \) copies of CD34. Next, extrapolating to the \( \sim 100 \) billion people estimated to have ever lived [56], nature has produced \( \sim 2.5 \times 10^{32} \) human CD34 molecules. By comparison, based on an average CD34 molecule having \( \sim 50 \) sites where \( O \)-linked glycans can be attached, with a conservative estimate of 20 different glycans occurring at each site, the resulting number of chemically distinct forms of CD34 is \( 20^{50} \) or \( \sim 10^{65} \). While this calculation is presented mainly for entertainment purposes in the vein of “every snowflake is different,” it nonetheless vividly demonstrates how thoroughly nature has obliterated the “one gene—one protein” hypothesis and turned it into a “one gene—innumerable glycoproteins” reality through the clever use of sugars.
1.2.3.2 N-Linked Glycans Although many types of peptide–sugar linkages exist in nature [57], two classes of protein glycosylation—N and O linked—are dominant in mammals. We will discuss N-linked glycans, which are more abundant than O-linked glycans in most cells, first. This class of sugars is covalently attached to the peptide backbone of a protein through a 2-acetamido-2-deoxy-β-D-glucopyranosyl (GlcNAc) β linked to the amide nitrogen (hence, N-linked) of an asparagine (Asn) side chain (Fig. 1.5, inset). N-linked glycan structures occur at the consensus sequence Asn-Xaa-Ser/Thr, where Xaa is any amino acid other than Pro; despite the consensus sequence, any particular site may or may not be occupied by a glycan for reasons that remain obscure. The GlcNAc-β-Asn linkage is widely observed in glycoproteins isolated from eukaryotes and is also widely distributed through phyla ranging from archaea and eubacteria [58].

N-glycans can be subdivided into three distinct groups that include high-mannose-type, hybrid-type, and complex-type structures (Fig. 1.5). The groups share a common biosynthetic route where the first step is the co-translational transfer of the dolichol-linked oligosaccharide Glc3Man9GlcNAc2 by oligosaccharyl transferase to an Asn residue of the growing polypeptide chain in the endoplasmic reticulum. As an aside, the production of this 14-mer is an interesting story in itself, beginning on

**Figure 1.5** Biosynthesis and major classes of N-linked glycans. The Dol-GlcNAc2Man9Glc3 14-mer is assembled on the cytosolic face of the ER, flipped to the lumen, and transferred to a consensus sequence during translation of a nascent protein (the chemical linkage is shown in the inset along with the symbols used to depict the constituent monosaccharides). Subsequent trimming of glucose and mannose produces the high mannose-type glycan structure, which can be elaborated to form hundreds of hybrid or complex-type structures.
the cytosolic face of the ER where the sugars are linked to the lipid dolichol phosphate and ultimately involves the improbable translocation of the hydrophilic glycan structure across the membrane via a “flippase” [59, 60]. The importance of this preparatory process, reviewed in ample detail elsewhere by us [61] and many others [10, 62–65], is evidenced in a bevy of congenital disorders of glycosylation (CDG) that arise from biosynthetic missteps and that are usually fatal early in childhood (although a few can be overcome with fairly straightforward remedies such as dietary supplementation of rare sugars such as fucose [66]).

Once the Glc3Man9GlcNAc2 structure is transferred to an Asn in nascent peptide chain, thereby creating the eponymous N-linked structure, a series of trimming events are set into motion. Initial trimming by the glucoside hydrolases I and II—in reasonably fast reactions that occur over a time span of a few minutes—remove the three glucose residues and yield the high-mannose-type glycoprotein. While it may seem odd that a cell is undoing a biosynthetic process in which it just invested a significant amount of energy, in reality the trimming process plays a vital role in protein folding and quality control along the secretory pathway [67–70]. The importance of this process is evident as a comparable processing mechanism for high mannose N-glycan chains is present in ancient pathways found in almost all eukaryotes including unicellular organisms such as yeast.

In mammals, further trimming and additions to N-linked glycans occur in the Golgi apparatus where sialic acid residues are added to yield hybrid-type or complex-type glycans (see Fig. 1.5) depending on the three-dimensional structure of the proteins, the cell type, and the organism [71]. From an evolutionary point of view, it is interesting to note that the primitive unicellular organisms do not have the mechanisms needed for the synthesis of hybrid and complex N-glycans. These organisms—which rely on early steps of N-glycan metabolism to ensure protein quality control—apparently do not require the complex intercellular recognition events needed to orchestrate multicellular life, which are in large part enabled by the sugar code [30] and thus could dispense with the effort needed to evolve the complete range of complexity found in the N-linked glycans of higher organisms.

1.2.3.3 O-Linked Glycans A second major class of glycoproteins is known as O-linked because the sugar moiety is attached to the oxygen of the hydroxyl group of either a serine (Ser) or a threonine (Thr) residue of the polypeptide chain [72]. The most prevalent type of O-glycan arises from mucin-type glycosylation, where N-acetyl-d-galactosamine (GalNAc) is attached to a Ser or Thr through an α linkage [73] (Fig. 1.6a). In general, O-linked glycans are an unruly bunch compared to their N-linked siblings; first, they are not confined to a consensus sequence but can seemingly occur on any Ser or Thr. Second, they are not limited to a single type of monosaccharide used to link the sugar to the peptide [57] but commonly use GalNAc, GlcNAc, and xylose-linked O-glycans (Fig. 1.6; additional O-linkages are provided in comprehensive reviews [57]). Nor are O-glycans based on an en bloc structure common to the entire class; rather, the mucin-type O-glycans are assembled into “core” structures by elaboration of an α-linked GalNAc (Fig. 1.6b). O-linked glycans have several more notable differences compared to N-linked structures, a
major one being that they are added post- instead of co-translationally; consequently, they do not play a role in quality control during folding. Thus, instead of simply being an artifact of quality control during protein folding, which is postulated to be the case for a subset of N-linked glycoproteins, a cell presumably installs O-linked glycans on proteins only when they are critical modulators of biological activity.

Biosynthesis of the ubiquitous mucin-type O-linked glycoproteins (Fig. 1.6a) found in mammals and other eukaryotes is initiated by the family of GalNAc transferases (ppGalNAcTs) that utilize UDP-GalNAc as the nucleotide donor substrate to modify protein substrates [74]. There are ~24 unique ppGaNTase human genes that display tissue-specific expression in adult mammals as well as unique spatial and temporal patterns of expression during development. One explanation for why there are so many genes coding for similar biochemical activity is that this redundancy provides protection against defects in any one particular gene. An emerging picture is

**Figure 1.6** Structures of O-linked glycans. (a) The predominant surface displayed O-linked glycans in humans are the mucin-type characterized by an α-linked GalNAc attached to a serine or threonine that can be further elaborated at either the C4 or C6 hydroxyl group to give the “core” structures and tumor-associated carbohydrate antigens (TACAs) shown in (b). (c) The O-GlcNAc protein modification, remaining as a single monosaccharide residue, is a unique example of glycosylation found on nuclear and cytosolic proteins. (d) A xylose-originated tetrasaccharide is used to covalently link a subset of glycosaminoglycans (GAGs) to cell surface proteins such as syndecans or glypicans; the major classes of GAGs (the majority of which are not covalently linked to cell surface elements) are shown in (e).
much more complex, however, as a subset of the ppGaNTases have overlapping substrate specificities and certain ppGaNTases require the prior addition of GalNAc to a peptide before they can catalyze sugar transfer to the substrate. Moreover, site-specific O-glycosylation by several ppGaNTases is influenced by the position and structure of previously added O-glycans [74]. The product of any ppGaNTase reaction, α-GalNAc on Ser/Thr residues, is termed the “Tn antigen” and is further elaborated by downstream glycosyltransferases to generate a series of core O-linked glycans [75] (Fig. 1.6b). These core structures are then further modified by other Golgi-resident glycosyltransferases to generate complex O-linked glycans that are involved in a variety of biological processes in health and disease [76]; for example, O-linked glycan abnormalities occur in cancer when relatively subtle changes to the normal core structures convert them into tumor-associated carbohydrate antigens (TACAs, Fig. 1.6b).

1.2.3.4 O-GlcNAc Modification of Nuclear and Cytosolic Proteins

For decades it was accepted wisdom that only cell-surface-displayed and secreted proteins were glycosylated in mammals. In 1984, however, Torres and Hart described a posttranslational modification where Ser and Thr residues found in nuclear and cytoplasmic proteins were O-linked to GlcNAc through a β linkage (Fig. 1.6c) [47, 77]. O-GlcNAc had two novel aspects: First, as mentioned, it had a nucleocytoplasmic distribution, whereas “traditional” glycoproteins were localized to the cell surface and topologically equivalent intracellular compartments, such as the lumens of the endoplasmic reticulum and Golgi apparatus [78]. Second, with the possible exception of plant nuclear pore proteins [79], O-GlcNAc is not elongated into more complex structures but rather remains as a single monosaccharide on the peptide backbone. As such, the biosynthetic “machinery” for O-GlcNAc is extremely simple—with an important caveat mentioned below regarding the dynamic nature of this modification—compared to other types of glycosylation. For example, in marked contrast to dozens of enzymes involved in sialylation (Fig. 1.4), O-GlcNAc metabolism involves one enzyme that adds the sugar to the protein [uridine diphospho-N-acetyl-D-glucosamine: peptide β-N-acetylglucosaminyl transferase (OGT)] and another [O-β-N-acetylglucosaminidase hexosaminidase (O-GlcNAcase)] that removes it [80]. Another contrast is that these two enzymes work in concert in a highly dynamic fashion, remodeling the “O-GlcNAcome” in a matter of minutes rather than over the hours-to-days turnover rates of most glycans.

The dynamic nature of O-GlcNAc protein modification is one of several similarities that this form of glycosylation shares with phosphorylation in cellular regulation. In addition to the rapid cycling of O-GlcNAc in response to metabolic factors, extracellular signals, stress, or stages in cell cycle progress, sugar attachment often occurs at the very same amino acid side chains on the protein backbone that are modified by protein kinases [48]. Unlike protein phosphorylation, however, where over 650 genetically distinct enzymes regulate the addition and removal of phosphate, as just mentioned only two catalytic polypeptides catalyze the turnover of O-GlcNAc. Consequently, O-GlcNAc is much simpler than both of its comparable biochemical systems of glycosylation and phosphorylation. Having down played the complexity of O-GlcNAc, we emphasize that the apparent simplicity of O-GlcNAc protein modification is countered
by links through the metabolic substrate UDP-GlcNAc—an indicator of flux through the hexosamine pathway [46]—to cell nutritional status and the greater complexity of sialylation and glycan biosynthesis (see Figs. 1.2 and 1.4).

1.2.4 Lipid-Based Glycans

1.2.4.1 Glycosphingolipids Glycosphingolipids (GSLs) are components of the plasma membranes of all eukaryotic cells. Roughly speaking, approximately 1%—or about one billion—of the lipids found in the plasma membrane of a typical 20-μm mammalian cell is glycosylated [81]. Approximately 300 different GSL structures have been identified [82] where at least one monosaccharide residue is glycosidically linked to a hydrophobic ceramide or sphingoid long-chain aliphatic amino alcohol that is imbedded in the lipid bilayer. The presence of these molecules at the plasma membrane enriches the outer surface in a layer of carbohydrate that helps to protect the cell membrane from chemical and mechanical damage. Despite the relatively small overall contribution of glycosphingolipids (≤8%) to the aggregate

![Figure 1.7 Lipid-based glycans. (a) Glycosphingolipid (GSL) biosynthesis begins with ceramide, which is most commonly elaborated with a Glc and a Gal to form LacCer, which in turn forms the “core” for three major classes of GSL [the lacto(neo) and globo series as well as gangliosides; top], each of which contain dozens (or hundreds) of structures. Alternately, and less commonly, a Gal instead of a Glc is added to Cer to form GalCer, a marker of autoimmune disease [89]. GalCer can be further elaborated to form a limited number of structures, including ganglioside GM4, sulfatide, and digalactosylceramidesulfate. (b) Glycosylphosphatidylinositol (GPI) membrane anchors, showing structural differences between various species.](image-url)
mass of the plasma membrane, they play several critical functions including cell adhesion, cell growth regulation, and differentiation [83]. The critical importance of GSLs in development has been demonstrated by the embryonic lethality in the mouse resulting from disruption of the gene-encoding ceramide-specific glucosyltransferase, an enzyme that initiates the synthesis of all glycosphingolipids [84]. Continuing to mature organisms, GSL “lubricate” signaling pathways [85] through “the glycosynapse” [86] and play ongoing roles in the maintenance of health.

Glycosphingolipids are derived from a common biosynthetic pathway that starts with the condensation reaction between palmitoyl-CoA and serine that leads to the formation of ceramide, which is the basic lipid structure of GSL (Fig. 1.7a). One class of GSL results from the addition of a galactose residue via a galactosyltransferase-catalyzed reaction to form galactosylceramide (GalCer), while glucosylceramide (GlcCer) is a product of a glucosyltransferase-catalyzed reaction. The glycosyltransferases responsible for these two reactions do not reside in the same subcellular compartment or have similar structural features, which is surprising since both use the same acceptor (ceramide) and similar nucleotide sugar donors. The ceramide-specific galactosyltransferase is a type I transmembrane protein whose catalytic domain is localized to the lumen of the endoplasmic reticulum [87]. By contrast, the catalytic domain of the ceramide-specific glucosyltransferase faces the cytosol with the enzyme restricted to the Golgi membrane [88].

In order to accommodate the various luminal and cytosolic orientations of the processing enzymes, newly synthesized ceramide is able to translocate across the membrane during bulk flow to the plasma membrane due to the rapid and spontaneous interbilayer transfer (flip flop) [90]. In turn, glucosylceramide translocates across the Golgi membrane where it becomes the substrate for Golgi resident glycosyltransferases [91]. The addition of galactose to form a lactose unit on ceramide is mediated by a glucosylceramide-specific galactosyltransferase [92]. At this point in the biosynthetic pathway, there is considerable competition for common substrates because lactosylceramide (LacCer) is the acceptor for various transferases that generate distinct groups of complex GSLs that consist of hundreds of already known [82]—and likely many more yet-to-be discovered—structures.

Not surprisingly considering their important roles in cellular physiology, GSLs play many roles in pathological processes when their complicated biosynthetic processes go awry. For example, when catabolism of glycosphingolipids is impaired, several severe pathological conditions are manifest [93], often as glycolipid lysosomal storage disease typified by the well-known example of Tay Sachs disease [94]. Collectively, congenital glycolipid disease is estimated to occur at a frequency of ~1 in 18,000 live births worldwide and is the most common cause of pediatric neurodegenerative disease. Aberrations in GSL also play significant roles in acquired disease—for example, in cancer, where changes in the relative expression of these molecules virtually always accompany oncogenic transformation [83, 95–98]. The close proximity of glycosphingolipids to the lipid bilayer—unlike farther outlying protein-associated glycans, in particular, polysaccharides associated with the extracellular matrix, that can more easily slough off and act as ineffectual binding decoys—is exploited by a number of viral and bacterial pathogens that have adapted...
to adhere selectively to these carbohydrate residues as a prelude to internalization and pathogenesis [99].

1.2.4.2 Glycosylphosphatidylinositol Membrane Anchors Proteins were discovered to be anchored to membranes through a covalently attached glycosylphosphatidylinositol (GPI) moiety in the 1980s [100, 101]. Subsequent structural determination of these GPI “anchors” that have been found in protozoa, yeast, plants, and mammals has revealed the common core structure: ethanolamine-PO₄-6-Man-α-1,2-Man-α-1,4-GlcNH₂-α-1,6-Mann-α-1,2-Man-α-1,6-Man-α-1,4-GlcNH₂-α-1,6-myoinositol-1-PO₄-lipid (Fig. 1.7B) where the ethanolamine is amide bonded to the α-carboxyl group of the C-terminal amino acid of the mature protein. The conserved core structure may possess a variety of side-chain modifications (additional phosphoethanolamines and sugars such as GalNAc, galactose, mannose) that are protein, tissue, and species specific [102].

The biosynthesis of GPI membrane anchors has been investigated in a broad range of eukaryotes ranging from protozoa and yeast to humans. Other than in mammals, GPI biosynthesis has been most extensively investigated in trypanosomes and many similarities—as well as a few differences (as described in detail elsewhere [105, 106])—have been found between mammals and bloodstream forms of the parasite Trypanosoma brucei [107]. In mammals, GPI anchors provide an alternative to hydrophobic transmembrane polypeptide anchors and also participate in intracellular sorting, in the endocytic process of potocytosis, in transmembrane signaling. They also facilitate the embedding of proteins into glycosphingolipid-rich lipid rafts and allow host proteins to be selectively released from the cell surface by the action of phospholipases. In trypanosomes, GPI anchors are present at up to 10–20 million copies (~100 times more than found on a mammalian cell) and dominate the cell surface molecular architecture of these organisms. The GPI anchors of trypanosomes ties in with the second major topic discussed in this chapter—immunity—because these molecules are well known for their ability to help the parasite avoid immune elimination by switching the immunodominant variant surface glycoprotein (VSG) coat during infection. The fact that trypanosomes have up to 1000 different VSG genes, combined with the ability of the parasite to rapidly release existing VSG by virtue of phospholipase C cleavage of proteins from the GPI moiety, affords the parasite extensive opportunity to escape host B- and T-cell responses by displaying new coat antigens [108].

1.2.5 Polysaccharides: Glycosaminoglycans and Bacterial Capsular Components

Although highly complex and collectively the dominant feature of the cell surface landscape, the majority of the N-, O- and lipid-linked glycans are relatively
modest in size, consisting of 20 or fewer monosaccharide residues. By contrast, polysaccharides—primarily the glycosaminoglycans (GAGs; see Fig. 1.6e for examples) but also including specialized structures such as α2,8-linked homopolymeric polysialic acids (see Fig. 1.4b, top)—are much larger with sizes of 100 monosaccharide residues or greater being commonplace. In general, although some GAGs are linked covalently to the cell surface—for example, to glypican or syndecans [16, 109] via the xylose-linked O-glycan structure shown in Figure 1.6d—in the majority of cases GAGs are free of explicit surface entanglements and instead exist as part of the proteoglycan component of the extracellular matrix (ECM). The major GAGs of physiological significance are hyaluronic acid, dermatan sulfate, chondroitin sulfate, heparin, heparin sulfate, and keratan sulfate. While each has a distinctive molecular composition, all GAGs are based on disaccharide units that contain either GalNAc or GlcNAc combined with one of two uronic acids (glucuronate or iduronate). Structurally, GAGs are long unbranched polysaccharides and are the most abundant heteropolysaccharides in the human body. GAGs are highly negatively charged molecules, highly hydrated, and exist in extended conformations that give high viscosity to the ECM. The “hydrogel” properties of these molecules are evident by considering that while most tissues have between 2 and 50 μg/mg dry weight of GAG [110], when hydrated these polysaccharides can constitute 70% or more of the volume of the ECM.

Mammalian GAGs are being implicated as an increasingly broad repertoire of biological functions besides the strictly structural. To illustrate with hyaluronic acid (HA), this very large glycosaminoglycan (with molecular weights of 100,000–10,000,000) is expressed in virtually all tissues and has long been known to be a critical structural component in the tissue interstitium by providing the core backbone of proteoglycans. HA is unique among the GAGs in that it does not contain any sulfate and is not found covalently attached to proteins but rather solely forms noncovalent proteoglycan complexes in the ECM. The discovery of HA-binding proteins led to the hypothesis that HA also serves as an adhesive substrate for cellular trafficking [111] and, most recently, the finding that HA fragments can deliver maturational signals to dendritic cells (DCs) and high-molecular-weight HA polymers can deliver co-stimulatory signals to T cells has established HA as an important immunomodulatory molecule [112]. Immune complications witnessed from the use of low-molecular-weight heparin as an anticoagulant [113] established that the ability of GAGs to engage an immune response is by no means unique to HA.

Although mammalian GAGs are relatively conserved in their basic structures, considerable structural diversity is obtained by postsynthetic modifications such as sulfation and epimerization of GlcA to IdoA. The chemical diversity nonetheless lags bacterial polysaccharides used either in the cell wall or in capsules where numerous structures not found in mammals abound. The ability of bacteria to employ an expanded repertoire of immunogenic monosaccharides in their glycans has led to the glycovaccinist exploiting these molecules to combat pathogenic microbes. Indeed, some of the earliest and most successful examples of carbohydrate vaccines are targeted against these molecules (as discussed in Section 1.4 and in detail in Chapters 2 and 4).
1.3 THE IMMUNE SYSTEM

1.3.1 Introductory Comments

Now that glycosylation has been briefly outlined, we will next provide an even less thorough, but hopefully helpful for the nonspecialist, overview of the immune system. Clearly, immunity is a huge topic—with well over one million articles available through searches of computer databases such as PubMed [35]—therefore, at the outset we emphasize that we provide a “bare-bones” discussion just sufficient to place into context some of the intriguing connections between glycosylation and immunity, and we trust that the valued reader will not feel slighted if his or her favorite aspect of the immune system is omitted in this chapter.

1.3.2 Overview of the Immune System

1.3.2.1 Immune System Provides Protection

The term immune system generically refers to a collection of mechanisms within an organism that protects against disease by identifying and killing invading pathogens and, in higher organisms, even providing protection against tumor cells. Virtually all multicellular organisms—including mollusks, worms, and insects—have primitive but effective immune systems; here we will primarily limit discussion to mammalian immunity. Of course, many features of immunity are broadly shared across phyla; on the other hand substantial differences separate humans from even mice, for example, complicating the biomedical researcher’s efforts to investigate human disease in this widely used animal model. The immune system of a mammal has the daunting task of not only recognizing a wide variety of agents, from viruses to parasitic worms, but also needs to distinguish them from the organism’s own healthy cells and tissues. In immunology, self molecules are those components of an organism’s body that can be distinguished from foreign substances by the immune system. Conversely, nonself molecules are those recognized as foreign molecules. The class of nonself molecules that are the smallest unit that the immune system responds to by binding to specific immune receptors to elicit an immune response are called antigens, short for antibody generators.

The immune system protects an organism from infection through a multilayer blanket comprised of several different steps. First, physical or mechanical barriers—such as skin in humans—prevent pathogens from entering the body [114]. If a pathogen crosses a physical barrier, the innate immune system—which is found in all plants and animals—acts quickly responding to challenges in a few minutes but in a nonspecific manner [115]. If a pathogen thwarts this second layer, a third level of protection found in most vertebrates—the adaptive immune system—can be activated. In adaptive (also known as acquired or learned) immunity, whose existence in invertebrates has been postulated but remains controversial [116], the immune response improves its ability to deal with an infectious agent by retaining a “memory” of the pathogen. Immunological memory allows the adaptive immune system to work faster and stronger each time a particular pathogen is encountered; for primary infections an effective response can require up to 5–7 days, whereas responses to subsequent insults
are mounted within 1–3 days [114]. Both innate and adaptive immunity are highly complex biological systems and each will now be discussed briefly (the innate immunity of carbohydrates is discussed throughout this book in greater detail, in particular in Chapters 3, 9, and 11).

1.3.2.2 Innate Immunity The mammalian innate immune system has two arms that are capable of mounting the complement and the inflammatory responses. First, the complement system consists of over 20 different proteins capable of mounting a biochemical cascade that attacks the surfaces of foreign cells. It is named for its ability to “complement” the killing of pathogens by antibodies and functions most effectively as a front line of defense when preexisting, circulating antibodies are present. Less helpful, complement-mediated cell killing complicates medical intervention; for example, immune response directed at the α-gal epitope is the source of hyperacute rejection of xenotransplanted organs and tissues. Either way, the existence, or new production of these antibodies, in turn relies on the larger functioning of the immune system, as described below.

Inflammation, one of the first “active” responses of the immune system to infection, results when cytokines or eicosanoids are released by injured or infected cells. Common cytokines include interleukins that are responsible for communication between white blood cells, chemokines that promote chemotaxis, and interferons that have antiviral effects, such as shutting down protein synthesis in the host cell [117, 118]. Eicosanoids include prostaglandins that produce fever and the dilation of blood vessels associated with inflammation, and leukotrienes serve as chemoattractants for certain types of white blood cells. Despite overall coordination by the immune system, white blood cells—or the leukocytes—behave as independent, single-celled organisms with many different duties and are the functional workhorses of the innate immune system.

Leukocytes that actively participate in innate immunity include mast cells, eosinophils, basophils, natural killer cells, and the phagocytes (macrophages, neutrophils, and dendritic cells). Collectively, these cells identify and eliminate pathogens, either by engulfing and digesting smaller microorganisms or attacking larger pathogens through contact. To briefly describe the roles of specific cells, mast cells reside in connective tissues and mucous membranes, regulate the inflammatory response, and contribute to allergies and anaphylaxis [119]. Basophils and eosinophils secrete chemicals involved in defending against parasites and play a role in allergic reactions, such as asthma [120]. Natural killer (NK) cells are leukocytes that attack and destroy tumor cells or cells that have been infected by viruses [121, 122].

Phagocytosis performed by cells called phagocytes that engulf and eat pathogens or other threatening particles is an important part of innate immunity. Of the three main types of phagocytic cells, neutrophils and macrophages constantly patrol the body searching for pathogens but can also be summoned to specific locations by cytokines. During the acute phase of inflammation, during bacterial infection, for example, neutrophils migrate toward the site of insult in a process called chemotaxis and are usually the first cells to arrive at the scene. Once a pathogen has been engulfed by a phagocyte, it becomes trapped in an intracellular vesicle called a phagosome, which subsequently
fuses with a lysosome to form a phagolysosome. The pathogen is killed by the activity of digestive enzymes or by a respiratory burst that releases free radicals into the phagolysosome [123]. Dendritic cells (DC) are phagocytes that reside in tissues that come into contact with the external environment such as the skin, nose, lungs, stomach, and intestines. These cells resemble—and derive their name from—neuronal dendrites, as both have many spinelike projections, but dendritic cells are not functionally connected to the nervous system. Instead, they function as a link between the innate and adaptive immune systems by presenting antigen to T cells, one of the key cell types of the adaptive immune system [124, 125].

1.3.2.3 Adaptive Immunity

The adaptive immune system arose early in vertebrate evolution and provided a stronger immune response as well as immunological memory, where each pathogen is “remembered” by a signature antigen through the coordinated action of B and T lymphocytes. B cells identify pathogens when surface antibodies bind to foreign antigens [126] and in turn their chief function is to secrete antibodies into bodily fluids in what is known as the humoral response. This sequence of events occurs because after a B cell encounters its triggering antigen, it gives rise to many large cells known as plasma cells. Every plasma cell is essentially a factory for producing a specific antibody, for example, one produces antibody against this year’s strain of influenza virus while another might produce antibody against the bacterium that causes pneumonia. In relatively short order, each plasma cell manufactures millions of identical antibody molecules and pours them into the bloodstream or lymph fluid.

An antibody matches an antigen much like a key matches a lock; some pairs form exact matches and bind with high affinity while others fit more loosely like a skeleton key. Whenever antigen and antibody successful form complexes, however, the antibody marks the antigen for destruction. Antibodies belong to a family of large molecules known as immunoglobulins and different family members play distinct—but sometimes overlapping roles—in host defense. Immunoglobulin G, or IgG, efficiently coats microbes speeding their uptake by phagocytic immune cells, whereas IgM is more effective at killing bacteria. Immunoglobulin A, or IgA, is concentrated in bodily fluids—tears, saliva, the secretions of the respiratory tract, and the digestive tract—and guards the entrances to the body. Immunoglobulin E, or IgE, has the beneficial natural job of protecting against parasitic infections but is also the villain responsible for the symptoms of allergy. Finally, immunoglobulin D, or IgD, remains attached to B cells and play a key role in initiating early B-cell response. Collectively, the five classes constitute about 25% of all serum proteins and, in any one individual, have a diversity of about $10^7$ different binding specificities. Interestingly, a diversity of $10^8 – 10^9$ binding specificities has been estimated to be found collectively in all humans but up to $10^{11}$ are theoretically possible as shown by combinatorial human antibody libraries [114]; these numbers may be of interest to the vaccine developer who will note that the human immune system has the (at least in theory) “extra” capacity to respond to novel antigens such as glycoconjugates.

The second component of adaptive immunity that complements the humoral response, known as the cell-mediated response, involves specialized white blood
cells called thymus-derived cells, T lymphocytes, or T cells. Unlike B cells, T cells do not recognize free-floating antigens. Rather, their surfaces contain specialized antibody-like receptors—the well-known T-cell receptor (TCR)—designed to recognize fragments of antigens on the surfaces of damaged, invading, infected, or even cancerous cells. T cells contribute to immune defenses in two major ways: some direct and regulate immune responses; others directly attack diseased cells. Helper T cells regulate both the innate and adaptive immune responses and help determine which types of immune responses the body will make to a particular pathogen [127, 128]. These T cells are not cytotoxic and thus do not kill infected cells or clear pathogens directly. They instead control the immune response by directing other cells to perform these tasks. By contrast, killer T cells, which are also called cytotoxic T lymphocytes, or CTLs, are a subgroup of T cells that kill cells infected with viruses (and other pathogens) or are otherwise damaged or dysfunctional. CTLs directly attack other cells carrying foreign or abnormal molecules on their surfaces and are particularly valuable for their antiviral action because viruses often hide from other parts of the immune system while they reproduce inside infected cells. CTLs, however, can recognize small fragments of these viruses on the membranes of infected cells and can eradicate the diseased cell [129].

1.3.3 Glycoimmunobiology

Now that we have provided a rudimentary introduction to immunology, we will delve into three specific areas in more detail to provide a small window into just how complicated the immune system is and also to merge the two areas of discussion—glycosylation and immunology—into a brief discussion of glycoimmunobiology. These examples touch base on three integral functions of the immune system—antibody function, cell movement, and cell activation—and illustrate how carbohydrates play a critical role in various facets of immunity. The intent once again is not to cover these areas comprehensively but to provide a sampling of concrete examples (selected out of many) to hopefully whet the interest of the reader in the manifold roles that sugar enjoys in the functioning of virtually all complex biological systems.

1.3.3.1 Antibody Glycosylation

Recombinant monoclonal antibodies (mAbs) are arguably the most important protein-based therapeutic agents. At the end of 2006, 18 mAb had been approved by the U.S. Food and Drug Administration (FDA) to treat a wide range of human diseases [130]; mAbs are also invaluable research tools for the biomedical community, which increases their biomedical and biotechnological importance yet further. For therapeutic purposes, mAbs are usually produced using mammalian cell lines, purified, concentrated, and subject to appropriate formulation for in vivo administration. In order to enhance biological activity and optimize pharmacologic properties, as well as to avoid deleterious off-target effects, increasing attention has been paid to the posttranslational modifications (PTM) of recombinant antibodies and, surely not surprising to a reader of this chapter, glycosylation is emerging as a critical PTM determinant of antibody structure and function [131–134].
The presence of various particular glycan structures on antibodies is crucial for antibody structure [135], interaction with Fc receptors [136, 137], binding to the complement component C1q, and the lack of particular sugar attachments has been implicated in autoimmune disease [136]. N-Glycosylation of immunoglobulin G (IgG) has been studied in detail and 32 different IgG glycoforms have been identified in human serum [138]. The N-linked oligosaccharides attached to antibodies produced by mammalian cells are mostly of the complex biantennary type, containing a mannosyl-chitobiose core and two N-acetylglucosamine (GlcNAc) residues, with variable additions of fucose, galactose, sialic acids, and bisecting GlcNAc (a bisecting GlcNAc is shown on the hybrid-type N-glycan in Fig. 1.5) [136]. At the submolecular scale, individual monosaccharides that comprise the glycans found on antibodies have now received intense scrutiny with galactose [139, 140], bisecting GlcNAc [141, 142], fucose [143], and sialic acid [144] all studied and found to have distinct contributions to antibody structure and function.

1.3.3.2 Leukocyte Extravasation

Inflammation is a process in which the body’s white blood cells and chemical agents protect one from infection by foreign substances including pathogens such as bacteria, parasites, and viruses. But, consider the situation where a person—maybe you—steps on a rusty, tetanus-laden nail; hopefully, leukocytes will rush to the site of insult to your foot. However, if these protective cells—which actually spend only ~2% of their time in the blood—were to wend their way through tissue or the ECM from where they are likely to be stationed in your body, which could be a meter (or more) away, at the top speed of 2.0 μm/h for this mode of locomotion [145], they would not reach the site of injury for 500,000 h—or about 57 years later. Clearly, such a situation is untenable because you would have long since died from the infection, or possibly old age. Leukocytes solve this problem by exploiting the blood as a rapid transit system capable of reaching any point in the body within a minute or so; but to do so in a way that successfully fights the infection, immune cells must have a way to exit the swiftly flowing blood at the site of injury. Carbohydrates play an absolutely critical role in the extravasation of leukocytes from the bloodstream into the underlying tissue at the site of insult. Later, when the crisis is over, cells with memory of the pathogen return to their home in lymph nodes—also far away—and use the same vascular transport and homing mechanisms.

Glycans play three distinct roles in leukocyte extravasation (Fig. 1.8). The first step—involving the endothelial glycoalyx layer (EGL)—is probably the least understood. What is known is that the EGL lines the lumens of blood vessels and, at up to half a micron in thickness, has an antiadhesive effect by preventing interaction between adhesion molecules on passing leukocytes (e.g., L-selectin, Fig. 1.8b) and their binding partners on the endothelium (e.g., CD34, Fig. 1.8c). The EGL has mechanical properties and structural rigidity such that the microvilli of leukocytes “tip-toe across it much like a Jesus Christ lizard can run across water” [146, 147]. Thus, based on the height that selectins and their binding partners extend above the lipid bilayer—a maximum of ~50 μm—the EGL must collapse by close to 90% of its usual thickness of >400 μm to allow these molecules to interact and mediate the well-known “tethering and rolling” behavior characteristic of leukocyte extravasation. In this second step (Fig. 1.8a),
Role of glycans in leukocyte extravasation. (a) Free-flowing leukocyte is shown in a cross section of a venule (not to scale; the leukocytes are typically 6–12 μm in diameter, the vessels where extravasation takes place have a diameter of 30–80 μm, and the EGL is approximately 0.2–0.5 μm thick) where extravasation to sites of injury or homing in the lymph node occurs. (step 1) The EGL collapses to a height of less than 50 μm allowing (step 2) selectin-mediated tethering and rolling to take place followed (step 3) by integrin-mediated firm adhesion and extravasation. (b) Graphic of selectin structures where CR represents the cysteine-rich consensus repeat domains, EGF represents the EGF-like domain, and Lec represents the carbohydrate recognition (lectin) domain. (c) CD34 exemplifies mucin-type counter-receptor for selectins where the peptide forms a scaffold for multivalent glycan display; up to 80% of the mass of CD34 can be carbohydrate [53]. (d) Chemical modification of sLe X determines physiological binding specificity between various selectins and counter-receptors. Of note, the reverse situation, shown where selectins are found on the epithelial substrate (e.g., P-selectin) and interact with ligands (e.g., PSGL-1) on the incoming cell, can also occur. (See color insert.)
selectins—a family of three structurally related lectins with affinity for the sialyl Lewis tetrasaccharide (sLeX) [148] (Fig. 1.8b)—interact with mucin-type glycoprotein ligands (Fig. 1.8c). Selectin–mucin interactions lead to unique flow-dependent rolling behavior mediated by unusual catch bond-to-slip bond characteristics [149, 150] of these binding partners that allow rapidly moving cells to slow down and sample the local environment near the site of infection [151, 152]. If the threat is legitimate, evidenced by the presence of the appropriate chemokines, a transition to firm adhesion and extravasation into the surrounding tissue will take place, primarily mediated through integrin activation and protein–protein binding interactions that are nonetheless also tuned by carbohydrates.

Of the three steps involved in leukocyte extravasation—ESL collapse, selectin-mediated tethering and rolling, and integrin-facilitated firm adhesion—by far the most is known about the contributions of carbohydrates to the middle step. Selectins vividly demonstrate the exquisite control that nature has achieved through subtle variation in the structure of the sLeX tetrasaccharide epitope. The binding partners for all selectins—L-selectin (e.g., GlyCAM-1 or CD34), P-selectin (e.g., PSGL-1), and E-selectin (e.g., CD44) all require the multivalent presentation of sLeX but have little cross-reactivity under physiological conditions [153]. It is now clear that in vivo binding specificity for each selectin depends on chemical modifications to their counter receptor such as sulfation; for L-selectin binding partners sLeX itself needs to be sulfated while for P-selectin ligands this modification occurs on the peptide backbone of PSGL-1 [154, 155]. Finally, nature has come up with a way to mask sLeX—to essentially sequester it in an inactive form that can be rapidly mobilized—by cyclization and decyclization [156] (Fig. 1.8d). In summary, submolecular microsurgery of carbohydrate epitopes—exemplified by sLeX—enables cells to move to where they are needed; of course, if they were not properly activated upon arrival, the entire process would be rather futile. Thus, in the next section we will discuss Siglecs, a family of molecules that participates in the activation of various leukocytes.

1.3.3.3 Cell- to Systems-Level Control Is Regulated through Siglecs

In addition to providing cells with a precise homing mechanism as they move rapidly to specific locations throughout the body, sugars also assist immune cells with activation through Siglecs (sialic acid-binding Ig-like lectins). Siglecs belong to an immunoglobulin superfamily (IgSF) of about a dozen—in humans—of cell surface receptors that recognize sugar ligands and play a smattering of roles in coordinating the myriad activities of the immune system [157, 158]. The first Siglec discovered was sialoadhesin (Siglec-1/CD169), a lectin-like adhesion molecule found on macrophages [159]. Other members of the Siglec family subsequently described include CD22 (Siglec-2), which is restricted to B cells and has an important role in regulating their adhesion and activation [160], CD33 (Siglec-3) [161], and myelin-associated glycoprotein (MAG/Siglec-4) [162]. Several additional Siglecs (Siglecs 5–12) have been identified in humans that are highly similar in structure to CD33 and collectively referred to as CD33-related Siglecs [163, 164]. CD33-related Siglecs have two conserved immunoreceptor tyrosine-based inhibitory (ITIM)-like motifs in their cytoplasmic tails implicating their involvement in cellular activation [165].
The mammalian glycome contains numerous sialylated glycans that are potential ligands for Siglecs and are therefore candidates as modulators of these receptors as they regulate certain aspects of adhesion, cell signaling, and endocytosis in the immune system. A decade of painstaking work is now reaching fruition in deciphering the relative affinities of individual Siglecs for $\alpha$$2,3$-, $\alpha$$2,6$-, or $\alpha$$2,8$-linked sialic acids [166–170], as well as their preference for the Neu5Ac or Neu5Gc forms of sialic acids [171, 172]. Moreover, the requirement of certain Siglec family members for sulfated carbohydrate ligands [173–175] is reminiscent of the binding habits of selectins involved in leukocyte extravasation.

The local concentration of sialic acids on surfaces of immune cells is very high; for example, on B cells it has been estimated to exceed 100 mM [176]. As a consequence, Siglec binding sites are typically “masked” by cis interactions with other glycan ligands expressed on the same cell [177–179]. In fact, Siglec-2 (CD22) has recently been shown to prefer itself as a binding partner [180]. In general, interactions with cis ligands dominate interactions with trans ligands in modulating the biological activities of Siglecs and these interactions tend to keep the cell quiescent. Important exceptions to the dominance of cis interactions in Siglec biology are provided by several examples of trans interactions that have the potential to regulate distant elements of the immune system. Sialoadhesin, for example, has an extended structure that projects its sialic-acid-binding site away from the plasma membrane, which reduces cis interactions. In other examples, trans interactions have been implicated in the activity of B cells [181] and in the suppression of Siglec-7-dependent natural killer cell activation [182, 183] in tissues such as those of the nervous system in which the inhibitory Neu5Ac-$\alpha$$2,8$-Neu5Ac-containing glycan ligands for this lectin are abundantly expressed [184, 185]. All in all, Siglecs comprise a versatile regulatory mechanism for the immune response.

1.3.4 Interplay between Glycosylation and Sugars: a Two-Way Street

Up to this point, we have covered briefly the biosynthesis of glycans, learning how cells invest substantial resources at considerable peril in the production of complex sugars. This effort does not go to waste, or impudently put an organism at risk of infection or metabolic disease because these glycans play innumerable roles in all aspects of the life a multicellular organism, with several specific examples—antibody structure and function, leukocyte homing, and Siglec regulation—of their critical roles in the functioning of the immune system described above. The reader who values fair play will find it heartening that glycans do not make all of these contributions without proper recognition from the immune system as it—quite literally but perhaps in an underappreciated manner—in turn recognizes carbohydrates as antigens. The fact that sugar structures are immunogenic runs counter to several generally held premises and leads to a number of questions. One of these is “how do carbohydrates fit into the conventional peptide processing system for protein-based antigens?” Another issue is protein–carbohydrate binding interactions typically conform to the cluster glycoside effect where multivalent carbohydrate presentation and multiple simultaneous binding
interactions are needed to ensure high affinity and avidity as well as binding specificity [186, 187]. These requirements are unlike the highly specific and high-affinity interactions generally thought to occur between a single antigen and matched antibody. Despite these, and several other puzzles in various stages of being solved, it is now undisputable that carbohydrates comprise antigens important to human health, both from the perspective of human disease (discussed next in Section 1.4.1) and from interaction with pathogens lurking within the environment (Section 1.4.2).

1.4 CARBOHYDRATE ANTIGENS

1.4.1 Carbohydrate Antigens in Humans

1.4.1.1 Historically, the Antigenicity of Carbohydrate Structures Has Been a Biomedical Problem Carbohydrate-based antigens have long posed problems for the biomedical community. The outstanding historical example—dating back to 1900—is the ABO(H) blood group system [2]. Over the past two decades, the biochemical basis of these antigens has been unraveled with the discovery of a common structure—the “H antigen” found in individuals with the O blood type—that is elaborated with a GalNAc or galactose residue to produce the structures that specify the A and B blood types, respectively (see Fig. 1.3d) [188]. More recently, and firmly establishing that blood type antigens are not an outlier but rather that carbohydrate immunogenicity is a central issue in transplantation [189], a similar problem has arisen from efforts toward the xenotransplantation of nonprimate organs that bear the “α-gal trisaccharide” (190–192). When α-gal-specific natural antibodies bind to the endothelium of vascularized xenografts, the complement system is activated, which leads to the activation of the coagulation cascade and rapid (within minutes to hours) graft rejection. This hyperacute immune rejection has spawned efforts to create α-gal knockout pigs because their organs are similar to human organs in many respects and if not for immune incompatibility would be attractive [193, 194]. Interestingly—and somewhat distressingly—despite knocking out the α-gal transferase gene, the first go-round of engineered α-gal knockout pigs still express this trisaccharide epitope [195].

Recently, another example of sugar-based transplant antigenicity has arisen for stem cell research. Based on the use of murine feeder layers or animal products such as fetal bovine serum, human stem cells scavenge the nonhuman Neu5Gc form of sialic acid and present it on the cell surface [4]. Because humans have circulating antibodies to Neu5Gc [196], concerns arise that the implantation of Neu5Gc-bearing engineered tissues would be subject to immune rejection. While regenerative medicine is a little outside the scope of this chapter, this information does raise two points that are worth emphasizing. First, the sensitivity of the immune system to minute changes in the chemical structure of carbohydrates is highlighted where a single hydroxyl of Neu5Gc—a difference in mass of only ~1% when these sugars are in a decasaccharide—is sufficient to elicit an immune response (a similar response where discrimination between a hydroxyl and N-acetyl group of A and B blood
type occurs). Second, the overexpression of Neu5Gc—presumably obtained from the diet—on human cancer cells [197] raises intriguing possibilities for new forms of cancer treatment. For example, the ability of cancer cells to scavenge nonhuman sugars and preferentially display them in their surface glycans provides impetus toward the development of metabolically glycoengineering strategies to develop cancer vaccines (see Section 1.4.3.4 and Chapter 10).

1.4.1.2 Abnormal Glycosylation and Cancer—An Opportunity to Exploit TACAs Therapeutically? Immunosurveillance mechanisms exist such that only a small fraction—perhaps as low as one in a million—of nascent tumor cells actually develop into full-fledged malignant disease [198–200]. The ability of the immune system to detect and eradicate cancer cells is in part due to the sometimes subtle and at other times quite dramatic changes to cell surface carbohydrates that take place during transformation [16, 201]. Since the 1980s numerous tumor-associated carbohydrate antigens (TACAs) expressed on both N- and O-linked glycoproteins as well as glycolipids have been cataloged [14, 201–203]. Common alterations—resulting from incomplete glycosylation or neoglycan production—occur in both N- and O-linked glycan structures and include increased size due to GlcNAc branching of core sequences close to the protein and variation in the terminal sequences [15]. Less commonly, but with significant repercussions particularly in breast cancers, mucins—such as MUC1—often have truncated glycans (such as the Tn antigen, Fig. 1.6b) that can be highly sialylated (e.g., sTn) and less sulfated than their normal counterparts [204–207]. TACAs expressed on glycolipids often involve changes to ganglio- or globo-series structures—such as glycolipid displayed fucosyl GM1—that are abundantly present in specific types of human cancers such as melanoma, Burkitt’s lymphoma, neuroblastoma, and small-cell lung carcinoma [208–211]. TACAs are further discussed in Chapter 8.

Now, taking a step back to combine the two ideas presented above—that is, that the human immune system can eradicate cells based on surface carbohydrate antigens such as α-gal during xenotransplantation or TACAs in cancer—the logical route of deliberately targeting TACA for immunotherapy becomes attractive. This course of action is intended to assist nature in eradicating the occasional but potentially devastating cancer cell that was otherwise overlook or allowed to survive. Indeed, tumors that develop into full-fledged malignant disease appear to be immunosculpted specifically for immune tolerance and sometimes even suppression [199]. One approach to help the immune system identify these disease-causing outliers, interestingly enough, directly exploits circulating antibodies against carbohydrate antigens by seeking to use gene therapy to selectively express α-gal on cancer cells, thus restoring their immunicity and targeting them for eradication [3, 212]. Complementary to such biology-based approaches, elegant—but exhausting—efforts to use synthetic chemistry to produce complex TACAs are underway (see Section 1.4.3.3). Finally, as already mentioned, innovative approaches that combine the power of biology and chemistry into “chemical biology” approaches using metabolic glycoengineering (Fig. 1.4c) to install neo-TACA onto cancer cells are underway (Section 1.4.3.4).
Clearly, much effort—evidenced by several chapters in this book alone—is going into TACA vaccine development, which raises a somewhat philosophical but nonetheless relevant question of “if a cancer cell can ‘trick’ the immune system into leaving it alone once, couldn’t its capacity for hypermutability come back into play and thwart a carbohydrate-based vaccine?” This concern is allayed by accumulating evidence that TACAs are not merely cancer “markers” (as an aside, carbohydrates “markers” are not meant to be diminished by this statement because of their value in immunodiagnosis; see Chapter 12) but play an active role in diseases progression. Consequently, if a cancer cell attempted to adapt to a vaccine by expressing less of a vaccine-targeted TACA, it may indeed be able to escape eradication by the therapeutic agent but in the process would become less virulent by virtue of no longer displaying the offending glycan. A specific example is provided by Lewis antigens, such as sLeX that facilitates leukocyte extravasation (Fig. 1.8) but is also relevant to cancer because the homing mechanism used by leukocytes to travel through the body has been co-opted by invasive cancer cells during metastasis [213]. Hence, a vaccine targeted against sLeX would be expected to reduce the invasive potential of low expressing cancer cell populations selected for survival by this therapeutic agent because any potentially metastatic cells would no longer be able to efficiently exit the bloodstream and establish secondary tumors.

1.4.2 Carbohydrates and Pathogens

1.4.2.1 Pathogens and Glycosylation—Inexorably Connected As discussed earlier, many pathogens exploit host glycans as binding epitopes during infection. Pathogens, immunology, and glycobiology, however, are interrelated in many additional ways as well. For example, on a fundamental level using the very broad definition of immunity to include protection achieved by physical barriers [114], bacteria that live in nutrient-rich but hostile environments—such as sewage or the human body—often utilize a carbohydrate capsule to protect themselves from bacteriophage and human immunity, respectively. As described in Section 1.4.2.2, bacterial glycans, which include these capsular polysaccharides as well as highly immunogenic lipopolysaccharides (LPSs), have already been used as vaccines to elicit protective immune responses. In a functionally analogous manner, eukaryotic parasites such as malaria and Leishmania (Section 1.4.2.3) also use surface carbohydrates to thwart host immunity and have likewise spawned rapidly maturing efforts to develop practical vaccines. Finally, viruses, which exploit of host glycans with exquisite binding specificity during infection, are ubiquitously glycosylated themselves, and viral glycans may someday be profitably exploited in vaccine development as well, as discussed briefly in Section 1.4.2.4.

1.4.2.2 Bacterial Carbohydrate-Based Antigens By contrast to TACAs whose glycoforms are co-opted from healthy cells, bacterial glycosylation presents a relatively easy target for vaccine development because of distinctive differences between microbial and human glycans. Even without outside intervention by a vaccine developer, humans and animals mount massive humoral responses against the LPSs of
gram-negative bacteria. Features that contribute to the high inherent antigenicity of LPS include the bacterial O-antigen glycan structures that is attached to lipid A. The immunogenicity of the O antigen is determined by outer core glycan structures that vary both within and between strains and species (differences between *Yersinia enterocolitica* O:3 and O:8 strains [215] are shown here). The O antigen polysaccharides consist of monosaccharides that are common to the mammalian host (see Fig. 1.3 for structures) as well as those unique to the bacterium [representative bacterial monosaccharides that occur in *Y. enterocolitica* O:3 and O:8 are shown in (b)].

**Figure 1.9** Bacterial O antigens and unique monosaccharide structures. (a) Lipopolysaccharide (LPS) structures are based on an inner core oligosaccharide that is attached to lipid A. The immunogenicity of the O antigen is determined by outer core glycan structures that vary both within and between strains and species (differences between *Yersinia enterocolitica* O:3 and O:8 strains [215] are shown here). The O antigen polysaccharides consist of monosaccharides that are common to the mammalian host (see Fig. 1.3 for structures) as well as those unique to the bacterium [representative bacterial monosaccharides that occur in *Y. enterocolitica* O:3 and O:8 are shown in (b)].

Despite the pronounced “non-self” features of bacterial LPS, producing an effective vaccine composed of nontoxic, immunogenic polysaccharides found in O
antigens has been very challenging, as illustrated by pioneering efforts against *Pseudomonas aeruginosa* [216]. One difficulty arises from the chemical diversity found among the different O antigens representative of the 20 major serotypes of this pathogen (additional diversity comes from variant subtype O antigens) that translates into a large degree of serologic variability. Accordingly, a broad acting O-antigen targeted vaccine by necessity must consist of a highly complex mixture of often poorly characterized glycans. Further complications originate from the poor immunogenicity of the major protective epitope expressed by some O antigens, and a large degree of diversity in animal responses that preclude predicting the optimal vaccine formulation from in vitro experiments or studies with model organisms.

In a complementary approach that does not depend on the LPS of gram-negative bacteria, immunogenic capsular polysaccharides from various pathogenic species that include *Streptococcus pneumoniae*, *Neisseria meningitidis*, *Haemophilus influenzae*, *Salmonella typhi*, *Shigella dysenteriae*, and *Klebsiella pneumoniae* are sufficiently abundant to be isolated from large-scale fermentation cultures and used as vaccines [217]. Similar to the problem encountered with the O antigen of LPS, however, carbohydrate diversity within species and between strains hampers vaccine development because once again complex mixtures of glycan structures must be dealt with. The polysaccharide vaccine PPV23, for example, contains 23 antigenically distinct polysaccharides found on the surface capsules of *S. pneumoniae*. These 23 serotypes were selected for inclusion in the vaccine because at least one of the polysaccharides occurs in most clinical cases of pneumococcal infections [218]. Unfortunately, for the impoverished developing countries where the majority of the hundreds of thousands of annual deaths from bacterial infections occur, the complex process of bacterial fermentation and isolation from many strains make “conventional” carbohydrate-based vaccines prohibitively expensive for widespread public health programs.

### 1.4.2.3 Malaria and Leishmaniasis—Parasites
Malaria afflicts about 300 million people annually worldwide causing up to 2 million deaths, predominantly in children. Among the four different *Plasmodium* species that cause this disease, *P. falciparum* is the most common and the most virulent. In recent years, malaria has spread at an alarming rate owing to the increased resistance of the parasite to drugs and of carrier mosquitoes to insecticides, and new approaches to combat malaria are urgently needed. As evidence pointing to the importance of GPI anchor structures in this disease’s morbidity and mortality mounts (the parasite’s GPI anchors can activate PTK- and PKC-dependent signaling pathways to regulate REL-A, C-Rel, and NF-κB/rel-dependent expression of cytokines, cell adhesion molecules, and iNOS resulting in erythrocyte sequestration and immune dysregulation characteristic of malaria pathogenesis), carbohydrate-based vaccines offer an attractive approach toward the amelioration of this pathogen. Toward this end, the Seeberger group recently showed that a synthetic malaria vaccine candidate (Fig. 1.9a) dramatically increased survival in infected mice (from 0–9% to ~70% [219]). In a similar approach, this group explored using a portion of the GPI structure (Fig. 1.7b) as a vaccine candidate for leishmaniasis, which as discussed earlier, is caused by parasites
that have surfaces dominated by these anchoring structures [219]. Additional perspective and recent advances in antiparasitic vaccines are described in Chapter 6 of this book and a topic not otherwise covered in this introductory chapter—carbohydrate-based fungal vaccines—is covered in Chapter 7.

1.4.2.4 Viral Glycosylation

Viruses co-opt host biosynthetic pathways to generate their genetic and structural material and use host glycosylation pathways to modify viral proteins with N-linked glycans. As occurs with the host’s surface and secreted proteins, N-glycosylation of viral envelope proteins promotes proper folding through interactions with the host’s cellular chaperones and facilitates proper trafficking through the secretory apparatus. In addition to these “quality control” functions, changes in glycosylation can reduce the ability of a virus to be recognized by the host’s immune system; for example, HIV (human immunodeficiency virus) and influenza, two clear threats to human health, rely on expression of specific oligosaccharides to evade detection by the host immune system. In addition, N-glycosylation plays important roles in a diverse set of vital biological functions of viruses that are specific to various classes of these pathogens. A few examples include the heavy influence of glycosylation over infectivity and intracellular transport in the hepatitis C virus [220]. Similarly, the Ebola, Hantaan, Newcastle, Hendra, Nipah, metapneumovirus, and SARS-CoV viruses all have N-linked glycans that make vital contributions to infectivity, protein folding, tropism, proteolytic processing, and immune evasion [221–229]. Finally, the glycosylation status of the West Nile virus has recently been linked to neuroinvasiveness and replication efficiency in several strains [230–233] in a manner reminiscent to the role glycosylation plays in modulating the conformational changes to influenza HA protein during cellular uptake [234–237].

Because sugars borne on viruses are produced by host cells, they are similar to endogenous glycans and a clearly defined repertoire of viral glycan epitopes—akin to the TACA counterparts that accompany cancer—that can be targeted by vaccines is challenging to identify. Consequently, although vaccines that directly target viral glycans lag in development compared to bacterial or parasitic efforts, there are nevertheless compelling reasons to pursue this line of investigation (as is detailed in Chapter 5). To briefly summarize here, natural variability in the glycosylation status of many viruses exists that is exemplified by the human immunodeficiency virus-1 (HIV-1). This pathogen causes AIDS (acquired immunodeficiency syndrome) by recognizing host cells though the interaction of the viral glycoprotein, gp120 [238, 239] with CD4 present on the surface of human T-lymphocytes and a second “co-receptor” molecule on the host cell surface. A survey of global HIV gp120 showed that this glycoprotein had a range of N-linked glycosylation site occupancy of between 18 and 33 with a mean of 25 [240]. This variability is thought to be influenced by competing pressures on the virus, similar to those experienced by influenza [61], where the presence of glycans is driven by their indispensable contributions to viral infectivity and protection against neutralizing antibodies [241, 242]. On the other hand, excessive glycosylation masks the necessary receptor ligand binding contacts through steric hindrance and the nonspecifically antiadhesive nature of the glycocalyx, thereby supplying selective pressure that limits the upper range of
glycosylation of the viral particle. Thus, in a manner similar to TACA vaccines used in cancer therapy as discussed above, although the production of a vaccine against a viral carbohydrate may be far from a panacea due to rapid mutation away from the targeted epitope, disruption of optimized glycosylation patterns through selective pressure imposed by the vaccine would render any surviving virus a less effective pathogen.

1.4.2.5 Translating Carbohydrate Antigens into Viable Vaccines This chapter does not provide a comprehensive list of potential carbohydrate-based vaccines; instead its purpose is to make a compelling case for the development of these vaccines. A first necessary prerequisite—the fact that glycans are immunogenic—was met by discussing the downsides of longstanding challenges facing transplantation efforts and surveying carbohydrate aberrations characteristic of selected human disease and pathogens. The bottom line is that, now that it has been established that the immune system can detect detrimental glycan antigens ranging from the capsule polysaccharides of a pathogenic bacterium to the TACA of a cancer cell, the possibility exists that a vaccine can be developed to eradicate the offending entity. Then, because the development of carbohydrate-based vaccines is not a trivial undertaking, it is worth emphasizing one more time that these agents are sorely needed because they meet urgent health problems both in rich (cancer) and poor (infectious disease) nations and are well worth expending the effort needed to bring them to fruition.

As will be discussed in more detail below, unique challenges accompany carbohydrate-based vaccine development; for example, with the exception of recent discovery of zwitterionic bacterial polysaccharides that can elicit a T-cell response [243], glycans activate B cells in a thymus-independent type 2 (TI-2) manner [244]. Because carbohydrates typically engage antibodies B cells without the help of helper T cells, IgM is the predominant isotype produced, and there is negligible class switching, no affinity maturation, and little development of memory cells. Consequently, vaccines composed entirely of carbohydrate typically are only effective in children over the age of 18 months to 2 years, and their response in adults generally lasts for only 3–5 years [245]. These problems can be overcome by employing appropriate glycan conjugation and adjuvant strategies, which are a major topic of this book (covered in Chapters 2–7, 9, and 10).

1.4.3 Carbohydrate-Based Vaccines

1.4.3.1 Brief History of Vaccination Although reports of people purposely inoculating themselves with other types of infections to protect themselves from disease date back to reports from 200 B.C. from China and India, the modern era of human vaccination is generally credited to Edward Jenner’s efforts in 1796 to use the cowpox virus to prevent smallpox [246]. Almost 75 years later, Louis Pasteur first used the terms immune and immunity in the scientific sense but acknowledged Jenner’s pioneering research by retaining the word vaccination (from the Latin vacca for cow) to describe his own accomplishments in the prevention of rabies and anthrax. Since then, great success has been realized in the development of vaccines to
manage and reverse infections caused by bacteria, viruses, and parasites, notably for
diphtheria (von Behring), polio (Salk), and smallpox, which have been largely
(or wholly) eliminated as threats to human health.

The field of carbohydrate vaccines, although blossoming tremendously of late, also
has a venerable past dating back to the 1920s and 1930s when Landsteiner, Avery, and
Goebel demonstrated that nonimmunogenic carbohydrates could become antigenic
when covalently attached to proteins [247]. This early work reached clinical practice
in ~1980 when Jennings and Roy derived polysialic acid from meningitis bacterial
capsules, coupled this polysaccharide to a carrier protein to render it immunogenic,
and ultimately produced a commercial vaccine [248]. Despite this (and now other
examples of) initial success, immense challenges remain in bringing carbohydrates
into the mainstream of vaccine development and immune therapies, which will be
outlined in Section 1.4.3.3 after briefly describing general requirements that must
be met during vaccine development.

1.4.3.2 General Requirements for Vaccines During the more than
200 years since successful vaccination was demonstrated by Jenner, a general set of
conditions has become evident for the design and development of any vaccine.
First and foremost, the identification and—to the extent possible—structural charac-
terization of the antigen must be done. In the case of nonpeptide antigens, structural
knowledge of the epitope is helpful in chemical synthesis of the antigen or a suitable
mimetic. Once an appropriate antigen has been synthesized, or isolated from natural
sources, a linker or spacer unit needs to be introduced for attachment to an immuno-
genic carrier protein or other immunostimulant while maintaining the immunological
integrity of the antigen in order to produce a sufficiently potent vaccine. Once an
appropriate conjugate has been obtained, immunological studies in animal models
must be done to evaluate the vaccine’s efficacy, and the antibodies elicited by the
vaccine should be isolated for detailed study of their interaction with the target antigen.
This latter step is particularly important for passive immunization strategies such as
those being developed for cancer therapy against TACA. Finally, evaluation in
human clinical trials, as discussed in Chapter 11, must be done before widespread
use of the vaccine can begin.

1.4.3.3 Wrinkles Thrown at the Glycovaccinist Many aspects of
carbohydrate-based vaccine construction, including the need to identify an appropriate
antigen, conjugation to a suitable immunogenic carrier, and evaluation of various
immunological adjuvants for co-administration are shared with general vaccine
development efforts. By contrast, one uniquely difficult challenge that confronts the
use of carbohydrate antigens in vaccine development is heterogeneity of naturally
occurring glycans that renders the isolation and purification of these molecules to hom-
ogeneity a daunting task. Consequently, the development of fully synthetic antigens
has become a large part of carbohydrate vaccine development efforts over the past
decade. In addition to gaining homogeneous material, synthetic strategies allow lin-
kers to be built into the carbohydrate structure appropriate for conjugation to a carrier.
This is particularly important due to the unique processing of carbohydrate-only
antigens that, among other challenges, render sugars incapable of raising an immune response in infants (as mentioned in Section 1.4.2.5 and Chapter 2).

The complete synthesis of oligosaccharides in sufficiently large quantities for practical use in vaccine development has remained difficult despite the elegant and exhaustive efforts pioneered by the Danishefsky group who made Globo-H [251], KH-1 [252], and the Le^Y and Tn [253] TACAs (Figs. 1.10c and 1.10d). Fully synthetic carbohydrate vaccines have important advantages over those isolated from natural sources because synthetic glycans can, in theory, be produced as homogeneous compounds in a controlled manner with little or no batch-to-batch variability, thus making heroic synthetic efforts worthwhile. The safety of completely synthetic antigens is also higher than vaccines derived from live cultures where the danger of contaminating immunogens, or disease-causing microbes, is small but real. In addition, medicinal chemistry techniques can potentially be used to derivatize and modify synthetic carbohydrates to make vaccines that are more immunogenic than those based on natural carbohydrates. The present status of synthetic glycoconjugates used in vaccine development is provided in Chapters 2, 4–7, 9, and 10.

Steps toward solving a major limitation of conventional synthetic strategies—the insufficiently small amount of material obtained—are being taken by automated
synthesizers being pioneered by the Seeberger group [254, 255] and “one pot” synthetic strategies reported by the Wong laboratory [256, 257]. Although not capable of producing any glycan structure on demand as automated DNA (deoxyribonucleic acid) synthesizers have long been able to do, this methodology provides a major boost toward several endpoints of major medical significance, including malaria (Fig. 1.10a) and leishmaniasis (Fig. 1.10b) [219]. Another important nuance of carbohydrate binding is to bring the cluster glycoside effect into play—demonstrated by the leishmaniasis vaccine candidate in Figure 1.10b and the Tn TACA in Figure 1.10c. A refinement of this technique is exemplified by multimeric antigenic constructs that target prostate and breast cancers with multiple TACA on the same molecular construct (Fig. 1.10d) [249, 258].

1.4.3.4 Metabolic Glycoengineering—Enhancing Immunogenicity and the Therapeutic Window  Despite notable examples of potent immunogenicity, coupled with a growing number of successes at exploiting this for vaccine development, the immunogenicity of carbohydrate antigens is far from universally adequate and remains problematically weak in some of the most lucrative applications such as cancer. In cancer, two related problems arise. First, immune tolerance to TACAs—many of which are fetal-oncogenic markers—limits a robust immune response. Second, because many or most TACAs are expressed to some degree even on healthy cells, the “therapeutic window” remains a potential obstacle. An interesting approach to overcoming these problems is to use a metabolic glycoengineering strategy (see Fig. 1.4d) that was inspired by observations that human cancer cells displayed relatively high levels of the Neu5Gc form of sialic acid compared to normal cells [197]. The source of the high levels of “nonhuman” sialic acid on tumor cells was traced to their highly efficient ability to scavenge this sugar from a carnivorous diet and replace the commonly occurring Neu5Ac with the modified sialoside [259, 260].

Based on the ability of the sialic acid biosynthetic pathway to accept nonnatural metabolic substrates, primarily ManNAc analogs, and process them into the corresponding nonnatural cell surface displayed sialosides, the Jennings group demonstrated almost a decade ago that PSA-displaying cancer cells could be selectively killed by a passive immunity approach [261]. In this landmark study, ManNProp was used to incorporate Sia5Prop into polysialic acid and an antibody to the Prop form of PSA—which had been made by chemical methods [262]—was co-injected into animals providing a potent anticancer effect. Since then, the Guo group has expanded this novel approach by increasing the repertoire of immunogenic sugar analogs available to include those with highly immunogenic N-phenylacetyl groups and expanding from the relatively uncommon TACA polysialic acid to more broadly distributed markers such as GM3 found in melanomas [250, 263, 264]. Interestingly, pathogenic bacteria such as Hemophilus ducyrei can also display nonnatural sialic acids via metabolic glycoengineering [265, 266]. Thus, because these pathogens use natural sialic acids to fool the immune system into believing they have humanlike qualities, the analogs may function as Trojan horses and provide a vehicle for the microbe’s demise by replacing their humanized sugars with abiotic, immunogenic counterparts.
1.4.4 Concluding Comments: Building on Success

The marriage of synthetic chemistry with established technologies is facilitating rapid progress in the development of a new generation of carbohydrate-based vaccines. Already, a synthetic vaccine that targets the bacterium *H. influenzae* type B was developed in 2004 in Cuba and is now part of that country’s national vaccination program [267]; this project is a prime example of how a chemical approach is superior to fermenting pathogenic bacteria in giant vats, which is “messy, expensive, and inexact” [268]. A “chemical biology” approach is also paying off in the development of fully synthetic vaccines against TACAs capable of eliciting robust immune responses by combining a TLR2 agonist, a promiscuous petide T-helper epitope, and a tumor-associated glycopeptide into a single construct [269]. Increasingly sophisticated approaches of this kind, described in more depth in Chapter 9, portend a bright future for carbohydrate-based strategies to modulate immunity toward solving many of today’s urgent health challenges.

ACKNOWLEDGMENT

The authors thank the National Institutes of Health for financial support (NCI CA112314-01A1).

REFERENCES


37. The HUGO (Human Genome Organisation) database is available at: http://www.hugo-international.org/.


56. Ramsey T (1999) The number of people estimated to have ever lived was 96,100,000,000. Available at: http://www.math.hawaii.edu/~ramsey/People.html.


REFERENCES


244. González-Fernández M, Carrasco-Marín E, Alvarez-Domínguez C, Outschoorn IM, Leyva-Cobián F (1997) Inhibitory effects of thymus-independent type 2 antigens on...


REFERENCES


