Special Invited Review

Biological roles of glycans

Ajit Varki

Departments of Medicine and Cellular & Molecular Medicine, Glycobiology Research and Training Center, University of California at San Diego, La Jolla, CA 92093-0687, USA

To whom correspondence should be addressed: Tel: +1-858-534-2214; Fax: +1-858-534-5611; e-mail: a1varki@ucsd.edu

Received 2 August 2016; Revised 15 August 2016; Accepted 16 August 2016

Abstract

Simple and complex carbohydrates (glycans) have long been known to play major metabolic, structural and physical roles in biological systems. Targeted microbial binding to host glycans has also been studied for decades. But such biological roles can only explain some of the remarkable complexity and organismal diversity of glycans in nature. Reviewing the subject about two decades ago, one could find very few clear-cut instances of glycan-recognition-specific biological roles of glycans that were of intrinsic value to the organism expressing them. In striking contrast there is now a profusion of examples, such that this updated review cannot be comprehensive. Instead, a historical overview is presented, broad principles outlined and a few examples cited, representing diverse types of roles, mediated by various glycan classes, in different evolutionary lineages. What remains unchanged is the fact that while all theories regarding biological roles of glycans are supported by compelling evidence, exceptions to each can be found. In retrospect, this is not surprising. Complex and diverse glycans appear to be ubiquitous to all cells in nature, and essential to all life forms. Thus, >3 billion years of evolution consistently generated organisms that use these molecules for many key biological roles, even while sometimes coopting them for minor functions. In this respect, glycans are no different from other major macromolecular building blocks of life (nucleic acids, proteins and lipids), simply more rapidly evolving and complex. It is time for the diverse functional roles of glycans to be fully incorporated into the mainstream of biological sciences.

Key words: biological roles, evolution, glycans

Introduction

In 1993, this journal published a review concluding that while limited evidence for all of the theories on the biological roles of glycans was available, exceptions to each could also be found (1). Some general principles were suggested. First, the reported biological consequences of experimental modification of glycosylation seemed highly variable, making it difficult to predict a priori the functions that a given glycan structure might be mediating, or its relative importance to the organism. Second, limited data suggested that the same glycan might mediate different functions at different locations within an organism, or at different times in its ontogeny. Third, the more specific intrinsic biological roles of glycans known at the time appeared to be mediated by unusual glycan sequences, unusual presentations of common sequences or further modifications of glycans. But it was also noted that such sequences were more likely to be targets for specific recognition by toxins and pathogenic microorganisms. It was therefore posited that ongoing host–pathogen interactions might contribute to the evolution of some aspects of intra- and interspecies glycan variation. Finally, some suggestions were made as to how one might elucidate more intrinsic biological functions for glycans. In particular, it was suggested that more studies of natural and induced mutations resulting in altered glycosylation within intact organisms would be required.

In the decade that followed, the author tracked several other more focused discussions of glycan functions, some examples of which are cited here (2–104). For a while it was indeed possible for an individual to track and read such reviews on biological roles of glycans, but this became increasingly difficult over time. By a decade
Later it became impossible to do so, and one had to be content with tracking a sampling of reviews in areas of ongoing personal interest (105–159). Meanwhile, some of the concepts in the original review were updated in a book chapter (160). Evolutionary and phylogenetic perspectives on the matter have also since been extensively addressed (161–176). More recently, it has been emphasized that glycans are as universal in nature as nucleic acids, proteins, lipids and metabolites (177), and as essential to the existence of all known living organisms (178). But as depicted in Figure 1, glycans are also one of the most structurally diverse and rapidly evolving major class of molecules. Taken together with much greater technical difficulties in their analysis, one can understand why the knowledge base regarding these major building blocks of life has lagged so far behind.

Despite these challenges, it is evident that information regarding the biological roles of glycans has vastly expanded in the last two decades. The present review first surveys the history of how our understanding of biological roles of glycans originally evolved, and then attempts to update the overview as of mid-2016. As a measure of how much progress has been made, any attempt at being comprehensive is now impractical, and the knowledge base of a single individual cannot do justice to this vast and complex field. Thus, one is only able to illustrate general principles with a few selected examples, and with a strong emphasis on the expertise of the author. For the same reason, the bibliography of citations cannot be comprehensive. Also, most of the broad implications of glycosylation for biotherapeutics are not addressed (179).

It is assumed that the reader is generally familiar with the major types and classes of glycans found in nature, and the conventional terminologies for describing them (180). Of course, it is important to also recognize that the full range of types and distributions of glycans in nature are still largely unexplored, and surprises continue to emerge. To cite just a few examples, the following glycans were mostly unknown when the previous version of this review was being written: functional sialylation in the fly nervous system (181); O-fucose and O-glucose glycans on Notch (182–184), and O-fucose on thrombospondin repeats (185); mucin type O-glycosylation in protists (initiated with α-GlcNAc instead of α-GalNAc) (186); O-linked N-acetylglucosamine on cell surface/extracellular proteins (187–188); the C-mannose linkage to proteins (189–190); the complexities of O-mannosyl-linked glycans in tissues such as muscle (191–194), including the novel glycosaminoglycan attached on α-dystroglycan (195–200) generated by a dual function xylosyl/glucuronosyltransferase (201–204) and attached via novel ribitol-phosphate bridge (205–208); a plant cell wall proteoglycan wherein a core arabinogalactan protein is glycosylated with cell wall matrix xylan and pectin glycans (209); identification of β-galacturonic acid in a xyloglucan involved in plant root hair tip growth (210); the expanding diversity of milk oligosaccharides (211), recognition of immunomodulatory glycans in the gut microbiome (212–213); N-linked glycans in prokaryotes (214–216) and, discovery of the large family of prokaryotic nonulosonic acids (217–219), the likely ancestors of sialic acids (220). The last two examples highlight the realization that many glycosylation types once thought to be unique to eukaryotes in fact have their origins in earlier evolved pathways in bacteria and archaea (176, 221–222). In this regard, it is notable that many major taxa of life forms such as archaea, fungi, protists and algae still remain poorly explored with regard to glycan structure and functions (many such taxa are not much addressed in this review).

**Historical background**

The first half of the 20th century saw great strides in elucidation of the structure and biochemistry of simple and complex glycans found in nature, garnering many Nobel Prizes (180). Beyond their well-known roles in energy generation and metabolism, glycans obviously had many structural and biophysical roles in many systems, including nutritional storage. Given the dense coating of complex and diverse glycans on essentially all cell surfaces (sometimes called the “glycocalyx” in animal cells) as well as on most extracellular molecules, it was also not surprising to find many examples of infectious agents or symbiotic organisms that recognized such glycans with a high degree of specificity, mediating interactions with their hosts (223). Additionally many pathogens were found to express highly specific glycans on their own surfaces, which seemed to modulate their antigenicity and/or their susceptibility to bacteriophages. Meanwhile, pathogens were also found to elaborate highly specific exo- and endoglycosidases that could degrade host glycans. In fact, many of the structural details of eukaryotic glycans were initially deduced by using such microbial glycosidases as tools (224–226).

The discovery of corresponding lysosomal glycosidases intrinsic to eukaryotic systems (227) then led to a better understanding of so-called “storage disorders”, wherein the deficiency of a single lysosomal glycosidase resulted in accumulation of the corresponding nondegraded product in lysosomes (228). Meanwhile, great strides were made in elucidating the structures of glycans in some taxa.
vascular traf
cells was shown to be due to carbohydrate
organisms that synthesized them. However, as of the end of the
provided
indirect
was still known about their speci
synthesis and metabolism of glycans in several taxa, remarkably little
was raised questions in the minds of some scientists, as to whether com-
glycans on glycoproteins. Ironically, in vitro viability of these
resulted in the correct prediction that sialylated glycan signals were involved in the trafficking of lympho-
c systems emerged experimental evidence that glyans by themselves might act as signaling molecules internal to a species. Meanwhile, studies in animals indicated that sialidase treatment could abrogate the interaction of lymphocytes with high endothelial venules in lymph nodes (286) leading to the correct prediction that sialylated glycan signals were involved in the trafficking of lympho-
cytes out of the circulation. Along with other convergent lines of evi-
dence, this eventually resulted in the definition of a family of cell adhesion molecules (287–288) that were critical for leukocyte roll-
ing on endothelium, prior to their exit from the circulation. These
molecules were called "selectins" (289), and they recognized a common
motif, consisting of sialylated fucosylated glycans (7, 13, 24, 91, 103, 128, 140, 290–309); a topic that has continued to blossom,
with implications for many fields.

While all this progress was occurring, it was generally assumed that glycosylation was only found on cell surface and secreted mole-
cules, and that the nucleus and cytoplasm were devoid of this class
of post-translational modification. The discovery of O-linked
GlcNAc (310–314) thus went unrecognized even by most other gly-
coscientists for years, until it was finally realized that this nucleocy-
toplasmic modification is the most common form of glycosylation in
eukaryotic cells (314–316), and that it mediates numerous modula-
atory functions on many proteins, including a complex interplay with
protein phosphorylation (317–320).

The increasing number of animal lectins that being discovered
and characterized was then classified based on sequence homologies
into C-type and S-type lectins (321), and the latter were eventually
designated as galectins (322–323). Discovery of the sialic acid-
binding properties of sialoadhesins (324) and of CD22 (325), fol-
lowed by the cloning of sialoadhesins (326), led to the definition of a
new family of cell-type-specific vertebrate lectins initially called "sia-
loadhesins" (15, 327), but eventually designated as a subfamily of
T-type lectins (328) and renamed as the Siglec (329). The previously
discovered phosphomannosyl receptors were now redesignated as
"P-type" Lectins (85) and some previously known plant lectins bec
came founding members of the "R-type" (Ricin-like) and
"L-type" (Legume lectin-like) families (330). The power of phylo-
genomic sequence comparisons has since revealed many additional
families of lectins such as the X-type lectins (intelec	ectins) (331–333),
Ficolins (334–335), etc. The earlier-mentioned somatic cell mutants
in pathways of N- and O-glycosylation played key roles in this pro-
gress. Along with the discovery of new human genetic disorders in
glycosylation (see next section), these and many other clues to intrinsic biological roles of glycans in the 1990s finally opened up this vast and uncharted territory of biology. Combined with the accelerating power of genomics and glycomics, we have now reached a point where numerous biological roles of glycans have been elucidated, to varying degrees of precision. For reasons of brevity, only a few examples are considered in this review.

Learning from natural or induced genetic alterations of glycosylation in multicellular organisms

Many different approaches have been used to elucidate the biological roles of glycans. Among these, one of the most instructive has been the study of genetic alterations of glycosylation in model organisms, and in human diseases. Indeed as mentioned earlier, it was the discovery of the genetic defect in I-cell disease that conclusively proved the biological significance and importance of the mannose 6-phosphate targeting pathway in vivo. At about the same time a defect in 3′-phosphoadenosine 5′-phosphosulfate (PAPS) formation was found in brachymorphic mice with multiple sulfation defects. While PAPS has many roles, the disproportionately short stature of the mice was apparently due to undersulfation of chondroitin sulfate in epiphyseal growth-plate cartilages (336). Another decade went by (see Figure 2) before the second human biosynthetic defect specific to glycosylation was discovered, a deficiency of a glycosaminoglycan core galactosyltransferase in a progeria-like syndrome (337). Meanwhile, the concept of “Carbohydrate Deficient Glycoprotein syndromes” (CDGs) had been suggested, based on the finding that children with previously unexplained multisystem disorders showed under-glycosylation of serum transferrin (338–341)—a test originally devised to detect alcoholism via the general hypoglycosylation it causes in liver-derived serum glycoproteins (342)! The work of many investigators then led to the elucidation of the underlying enzymatic and genetic defects in these children (343–350), eventually resulting in the repurposing of the acronym CDG to denote “Congenital Disorders of Glycosylation” (66, 351–356). After a slow start in the early 1990s an international effort of many investigators has now resulted in a veritable explosion in discoveries of human genetic disorders of glycosylation (Figure 2) (61, 67, 191, 352, 357–382). These disorders continue to provide a goldmine of clues to biological roles of glycans, and this understanding has begun to benefit some patients via simple monosaccharide replacement therapies (374, 383–388). Notably, in a recent study of

Fig. 2. Accelerating progress in the discovery of human glycosylation disorders. The graph shows the cumulative number of human disorders with a major genetic defect in various glycosylation pathways and the year of their identification (2016 data for first 6 months). In early years, initial discovery was based on compelling biochemical evidence, and in later years by conclusive genetic proof. In most instances, the year indicates the occurrence of definitive proof of gene-specific mutations and correlations to biochemical results. Figure kindly provided by H. Freeze and Bobby Ng, updated from ref. 375 and reproduced with permission from Freeze HH, Chong JX, Bamshad MJ, Ng BG. 2014. Am J Hum Genet. 94:161–175. Copyright Elsevier. Reproduced with permission.
consecutively enrolled patients with unexplained intellectual developmental disorder and metabolic phenotypes, whole exome sequencing showed that >10% were attributable to genetic defects in glycosylation pathways (389), mostly hypomorphic states of genes in which complete loss would have been lethal.

Beyond the clinically obvious CDGs, there is also increasing genetic evidence for the role of glycosylation-related genes in more subtle and common diseases in the population, as observed in genome wide association studies a few examples of which are mentioned here: NDST3 (390) and ST8SIA2 (391–393) in schizophrenia and bipolar disorder; FUT2 nonsecretor status and blood group B associated with elevated serum lipase activity and risk for chronic pancreatitis (394); and type 2 diabetes susceptibility associated with ST6GAL1 (395).

Meanwhile, targeted genetic alterations of glycan biosynthetic pathways in mice (396–401) also revealed a spectrum of abnormalities, again pointing to complex and varied functions of glycans in multicellular organisms. Since then, the list of mice with genetically altered glycosylation has expanded greatly, and resulting phenotypes have been highly instructive (98, 159). It is ironic that most of the glycosylation pathways that had earlier been dismissed because genetic defects caused “limited phenotypes” in the reductionist environment of the tissue culture dish later turned out to have clear and serious consequences in the intact organism, even in the form of hypomorphic alleles in humans. On the other hand, the phenotypic outcome of gene knockouts has been rather unpredictable. For example, while the MGAT1/GnT-I null state (which prevents the processing of N-glycans) caused embryonic lethality in mice (397–398), it generated no grossly obvious phenotype in the Arabidopsis plant (402–403), and limited phenotypes in Drosophila (404). Conversely while mice lacking ST3GAL5 seem to have only moderate phenotypes (405–406), humans with similar defects suffer from severe multisystem disease (407–408). Of course, any report of a “viable and fertile mouse with no major phenotype” must be taken with a large grain of salt. For example, the consequences of altering complex ganglioside biosynthesis (409–411) or of knocking out one of the key ganglioside receptors called MAG (412–413) was mostly evident later in the life, or when the mouse was subjected to specific challenges (414–415). In contrast, complete elimination of ganglioside biosynthesis gave an early embryonic lethal phenotype (416). Further complexity has arisen from the realization that there are multiple isozymes of some glycosyltransferases (417–418), and that post-transcriptional regulation by micro-RNAs is occurring (419–420).

Summaries of all human and model organism phenotypes resulting from genetic alterations in glycosylation will not be attempted here, and are reviewed elsewhere (98, 159, 375, 378). In general, complete elimination of major classes or subclasses of glycans tends to result in embryonic lethality, while defects in outer terminal structures often give viable organisms with defects in specific functions and/or specific cell types, although these impacts are often species specific. As an example, null alleles preventing the synthesis of the core glycosaminoglycan backbone of heparan sulfate causes embryonic lethality (421–422), but the prevention of proper sulfation of this backbone can give living mice with specific defects (422–427). When embryonic lethality makes it difficult to define specific biological roles, tissue-specific targeted genetic alterations became important (428–430). Experiments of nature such as somatic mutations in X-linked genes (431) and hypomorphic alleles of essential genes (388–389, 432) have also helped our understanding of functions.

Note that the discussion above largely focused on examples from animals. While space does not allow a detailed discussion, loss of glycosylation in plants, fungi or prokaryotes can also lead to cell death. For example, a meristem-localized inducible expression of an UDP-glycosyltransferase gene is essential for growth and development in pea and alfalfa (433). Ethambutol (a traditional drug treatment for tuberculosis) is now known to target the arabinofuranosyltransferases EmbA and EmbB (434). Knockouts of these genes in mycobacteria are lethal, as is the case with some other glycosyltransferases (435). And in the fungus Aspergillus fumigatus, inhibition of cell wall β glucan synthesis is toxic (436–437).

A broad classification of the biological roles of glycans

There are several different ways to classify the biological roles of glycans, based on the glycan types in question, or the glycan-binding protein involved, etc. A simple and broad classification (160) (see Figure 3 for a conceptual organization and Table 1 for a complete listing) divides glycan functions into four somewhat distinct categories. The first is structural and modulatory roles (including nutrient sequestration). The second category involves extrinsic (interspecies) recognition. The third is intrinsic (intraspecies) recognition. Finally, there is molecular mimicry of host glycans. All of these categories can involve glycan-binding proteins (see Figure 3). The next part of this review considers these classes of biological roles and discusses one or more examples of each. Given the vastness of relevant literature, the examples and citations are rather limited and biased towards the knowledge of the author. Examples of multifunctional roles of glycans and glycan-binding proteins that cross over between these somewhat arbitrary categories will be mentioned later.

Fig. 3. General classification of the biological roles of glycans. A simplified and broad classification is presented, especially emphasizing the roles of organism-intrinsic and organism-extrinsic glycan-binding proteins in recognizing glycans. There is some overlap between the categories, e.g., some structural properties involve specific recognition of glycans. Binding shown on the left of the central “self” cell represents intrinsic recognition, and extrinsic recognition is represented by binding shown to the right of that cell. Molecular mimicry of host glycans adds further complexity to potential roles. Original drawing by R. Cummings, updated from ref. 160 with permission from the Consortium of Glycobiology Editors.
Table I. Biological roles of glycans

<table>
<thead>
<tr>
<th>Structural and modulatory roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical structure</td>
</tr>
<tr>
<td>Physical protection and tissue elasticity</td>
</tr>
<tr>
<td>Water solubility of macromolecules</td>
</tr>
<tr>
<td>Lubrication</td>
</tr>
<tr>
<td>Physical expulsion of pathogens</td>
</tr>
<tr>
<td>Diffusion barriers</td>
</tr>
<tr>
<td>Glycoprotein folding</td>
</tr>
<tr>
<td>Protection from proteases</td>
</tr>
<tr>
<td>Modulation of membrane receptor signaling</td>
</tr>
<tr>
<td>Membrane organization</td>
</tr>
<tr>
<td>Modulation of transmembrane receptor spatial organization and</td>
</tr>
<tr>
<td>function</td>
</tr>
<tr>
<td>Antiadhesive action</td>
</tr>
<tr>
<td>Depot functions</td>
</tr>
<tr>
<td>Nutritional storage</td>
</tr>
<tr>
<td>Gradient generation</td>
</tr>
<tr>
<td>Extracellular matrix organization</td>
</tr>
<tr>
<td>Protection from immune recognition</td>
</tr>
<tr>
<td>Effects of glycan branching on glycoprotein function</td>
</tr>
<tr>
<td>Cell surface glycan:lectin-based lattices</td>
</tr>
<tr>
<td>Masking or modification of ligands for glycan-binding proteins</td>
</tr>
<tr>
<td>Tuning a range of function</td>
</tr>
<tr>
<td>Molecular functional switching</td>
</tr>
<tr>
<td>Epigenetic histone modifications</td>
</tr>
<tr>
<td>Extrinsic (interspecies) recognition of glycans</td>
</tr>
<tr>
<td>Bacterial, fungal and parasite adhesins</td>
</tr>
<tr>
<td>Viral agglutinins</td>
</tr>
<tr>
<td>Bacterial and plant toxins</td>
</tr>
<tr>
<td>Soluble host proteins that recognize pathogens</td>
</tr>
<tr>
<td>Pathogen glycosidases</td>
</tr>
<tr>
<td>Host decoys</td>
</tr>
<tr>
<td>Herd immunity</td>
</tr>
<tr>
<td>Pathogen-associated molecular patterns</td>
</tr>
<tr>
<td>Immune modulation of host by symbiont/parasite</td>
</tr>
<tr>
<td>Antigen recognition, uptake and processing</td>
</tr>
<tr>
<td>Bacteriophage recognition of glycan targets</td>
</tr>
<tr>
<td>Intrinsic (intraspecies) recognition of glycans</td>
</tr>
<tr>
<td>Intracellular glycoprotein folding and degradation</td>
</tr>
<tr>
<td>Intracellular glycoprotein trafficking</td>
</tr>
<tr>
<td>Triggering of endocytosis and phagocytosis</td>
</tr>
<tr>
<td>Intercellular signaling</td>
</tr>
<tr>
<td>Intercellular adhesion</td>
</tr>
<tr>
<td>Cell–matrix interactions</td>
</tr>
<tr>
<td>Fertilization and reproduction</td>
</tr>
<tr>
<td>Clearance of damaged glycoconjugates and cells</td>
</tr>
<tr>
<td>Glycans as clearance receptors</td>
</tr>
<tr>
<td>Danger-associated molecular patterns</td>
</tr>
<tr>
<td>Self-associated molecular patterns</td>
</tr>
<tr>
<td>Antigenic epitopes</td>
</tr>
<tr>
<td>Xeno-autoantigens</td>
</tr>
<tr>
<td>Molecular mimicry of host glycans</td>
</tr>
<tr>
<td>Convergent evolution of host-like glycans</td>
</tr>
<tr>
<td>Appropriation of host glycans</td>
</tr>
</tbody>
</table>

Structural and modulatory roles

Given their ubiquitous presence and abundance in all cellular compartments, extracellular spaces and body fluids, glycans have many major biological effects mediated by their own primary structural properties, and/or by modulating functions of proteins and lipids to which they are attached. The randomly selected examples provided in each section do not focus on any particular glycan class per se.

Physical structure

β-Linked homopolymers of glucose or N-acetylgalactosamine (cellulose or chitin, respectively) are among the most abundant organic molecules on the planet, providing strength and rigidity to structures such as plant and fungal cell walls and arthropod exoskeletons (438–439). These polymers are also difficult to breakdown by physical, chemical or enzymatic means. Many other glycan polymers play major roles in fungal and plant cell wall structure and function (438–447). For example, the hemicellulose xyloglucan not only plays a key role in the loosening and tightening of cellulose microfibrils, but also enables the plant cell to change its shape during growth and differentiation, and to retain its final shape after maturation (441). Needless to say, in the absence of these and many other major glycan polymers, the diversity of macroscopic structural variations in life forms on the planet would be far more limited.

Physical protection and tissue elasticity

There are many instances where thick layers of glycans provide an important physical protective role. In addition to the polymers mentioned above, the dense layer of mucins that coats many epithelial surfaces such as the inner lining of airways and intestines provides critical barrier functions, including protection against the invasion by microorganisms that live within the lumen (139, 418, 448–452). Disruption of this layer by genetically altering mucin backbones, O-linked glycosyltransferases (or key chaperones like Cosmc) can have very serious consequences, including inflammation and carcinogenesis associated with microbial invasion (453–456). Likewise the thick and biochemically robust cell walls make it difficult for invading fungi and bacteria to reach the membrane of the plant cells (438, 441, 444, 446–447). In other instances, the thick layer of glycans also provides tissue strength. Nature is rife with many more such examples including fungal and bacterial cell walls and polysaccharide coats, and the glycosaminoglycans of vertebrate cartilage, which are partly responsible for its elasticity, resiliency and compressibility (457).

Water solubility of macromolecules

It is interesting that many vertebrate internal body fluids such as blood plasma are rich in heavily glycosylated proteins. Apart from specific functional reasons for glycosylation, hydrophilic and acidic glycans also contribute significantly to the water solubility of these macromolecules. Indeed the remarkably high concentration of proteins in the blood plasma (~50–70 mg/mL in humans, carrying ~2 mM of bound sialic acids) would probably be impossible without this glycosylation. The antifreeze glycoproteins of certain fish alter the structure of bulk water itself, preventing nucleation of ice crystals in body fluids (458–461). Antifreeze functions can also be mediated by certain polysaccharides with a lipid component (460).

Lubrication

The remarkably efficient lubrication provided by soluble and membrane-bound mucins on the lining of hollow organs may seem like a trivial “function”—until one realizes that deficiencies in oral salivary mucins caused by radiation damage to salivary glands (a side effect of head-and-neck cancer treatment) (462–463) or by autoimmune disease (Sjogren’s) (464) can be life-threatening, especially by limiting the ability to swallow food. Another example is the critical lubricating role of hyaluronan in body fluids (465–467),
such as the synovial fluid in joint cavities and tear fluid in the eyes, wherein deficiencies can be supplemented therapeutically (467–471).

Physical expulsion of pathogens

Heavily glycosylated secretions produced in large amounts can serve as a response to physically expel intruders. For example, expulsion of N. brasilensis worms from the rat intestine is associated not only with quantitative, but also with qualitative changes in the composition of mucins in goblet cells (472). On the microscopic level, a recent study shows that a “sentinel” goblet cell localized to the mouse colonic crypt entrance recognizes bacterial products, activating the Nlrp6 inflammasome, eventually inducing mucin secretion from adjacent goblet cells in the upper crypt, which expels bacterial intruders that have penetrated the protective inner mucus layer (473).

Diffusion barriers

 Extracellular matrix glycosaminoglycans and/or heavily sialylated glycoproteins can comprise critical diffusion barriers. For example, the heavily sialylated protein podocalyxin on glomerular podocyte foot processes (474–478) and heparan sulfate glycosaminoglycans within the glomerular basement membrane (479–484) seem to play important roles in maintaining the integrity of blood plasma filtration by the kidney. Pathological or experimental damage to such glycans causes large molecules like albumin to escape into the urine, and is associated with glomerular diseases (478, 485–491).

Glycoprotein folding

Protein molecules that are synthesized and secreted via the ER-Golgi pathway can be subjected to ER modifications such as O-fucosylation (492) and O-mannosylation (493), with important effects on facilitating proper folding in the ER lumen. A major fraction of such ER-synthesized proteins are also modified by N-linked glycans at Asn-X-Ser/Thr sequences (494), and it is reasonable to think that the large, generally hydrophilic sugar chains contribute to proper folding of nascent polypeptides emerging into the lumen of the ER. Indeed, it has long been known that preventing N-linked glycosylation using the inhibitor tunicamycin can have negative effects on the initial folding of such proteins (495–496). In keeping with the highly conserved structure of the initial glycan added to asparagine residues, we now know that such N-glycans play a much more precise role in actually directing the folding, via specific recognition of certain features of N-glycans (see further discussion on quality control below). Even at the level of initial protein folding the exact context of the sequon can dictate the outcome. For example, experimentally placing a phenylalanine residue two or three positions before a glycosylated asparagine in distinct reverse turns facilitates stabilizing interactions between the aromatic side chain and the first GlcNAc residue of the glycan, while increasing glycosylation efficiency (497).

In this context, it is worth noting that the vast majority of published crystal structures of naturally occurring glycoproteins are derived from proteins that either had their glycosylation sites mutated, or had their glycans partially or completely degraded, prior to crystallization. The practical reason for making such a drastic change is that glycans are often heterogeneous and have a high range of motion, making it difficult to obtain an ordered crystal. Even if crystallization is possible, the glycans are typically disordered within the resulting image. In instances where glycans are left intact, the glycoproteins are often expressed in heterologous cells, resulting in nonspecies-specific glycosylation. This major technical artifact is rarely addressed in prominent protein crystallography papers. The bottom line is that when glycosylation sites are mutated or glycosylation is modified, there is a significant possibility that the folded form defined by crystallography may not be the native state. Solving this technical problem remains a major challenge for the future (498–500), one in which new techniques such as cryo-electron microscopy (501–502) may help. Meanwhile, it is definitely a worthwhile exercise to model the glycans back into the crystal structure to a best approximation (501, 503). Exceptions to the general lack of glycans in crystal structures can occur when the glycan is buried and partially immobilized within the folds of the protein, such as in the case of the Immunoglobulin-G (IgG) Fc-region (504–506), or tightly packed on the surface, such as in the HIV virion (507–509).

Protection from proteases

Heavily glycosylated proteins are protected from protease cleavage by glycans, likely by steric hindrance or negative charge. This effect is particularly prominent for mucins carrying densely packed O-glycans (451, 510). Indeed, extended segments of some mucins are even resistant to broad-spectrum proteases like proteinase K (511–512). This property can actually be taken advantage of, to isolate mucin segments away from other proteins that could be more easily proteolyzed into smaller fragments, or even from whole tissues (511–512). In a prokaryotic example, N-glycosylation in Campylobacter improves fitness, by providing protection against proteases in the gut (513). Conversely, there is evidence that glycosylation at single sites can regulate specific cleavage events with large impacts on protein activity (514), e.g., the protection of Tango1 by O-glycosylation is critical to apical secretion in Drosophila (515).

Modulation of membrane receptor signaling

Classic studies have shown that glycolipids can alter the signaling properties of protein receptors present within the same cell membrane (516). For example, subtly different forms of the sialylated ganglioside GM3 can have differential effects on tyrosine kinase signaling of the EGF receptor (517–522) and elimination of GM3 affects insulin receptor action (405, 523–524). Another classic example is the co-receptor activity of heparan sulfate in FGF signaling (525–526). Glycosylation can also affect the signaling properties of the proteins to which it is attached. For example, α1-6 core fucosylation of N-glicans affects transforming growth factor (TGF) signaling (527). Dysregulation of TGF-β1 receptor activation leads to abnormal lung development. While most core fucose-deficient mice die 3 d after birth, the survivors develop emphysematous changes of the lung. The underlying mechanism appears to be dysregulation of downstream TGF signaling, causing MMP gene activation, which eventually degrades alveolar membranes to give emphysema. In a similar vein, both sialylation and fucosylation modulate epidermal growth factor receptor-mediated intracellular signaling (528–530).

An entirely new field opened up with the discovery that the Fringe molecule is a glycosyltransferase that modifies the important signaling protein Notch and thus modulates Notch-Delta interactions (182, 184). Before Fringe can act, Notch must first be glycosylated with an O-fucose, and the protein O-fucosyltransferase 1 is thus an essential component of Notch signaling pathways (531–532). It was later discovered that the O-glucose modification on Notch added by a glycosyltransferase encoded by Rumi is also essential for Notch signaling and embryonic development (533–534). Thus there are many roles of glycosylation in Notch signaling (535–536). The structural basis of glycosylation-mediated
Membrane organization
Glycans can have profound effects on the organization of cell membranes. For example, GPI-anchored proteins are mainly associated with glycolipid-enriched membrane microdomains (538) and are organized in submicron domains at the cell surface (539). Cell surface lectins may also participate. Galectin-4 appears to be the major organizing factor of such “lipid rafts” on gastric epithelial cells (540). GPI-anchored proteins are selectively targeted to the apical surface in fully polarized epithelial cells (541). It also stands to reason that the glycans on cell surface glycoproteins can modulate membrane domain organization by their bulk and charge (542–543). It now appears that glycans on one class of glycoproteins can even modulate the organization of other classes of glycans on other proteins present on the same cell surface, perhaps forming “clustered carbohydrate patches” (146, 302, 544). The formation of galectin-mediated lattices in the glycocalyx is discussed below.

Modulation of transmembrane receptor spatial organization and function
In addition to the role of heparan sulfate proteoglycans in modulating transmembrane receptor spatial organization discussed above, bulky glycoproteins in the cell surface glycocalyx can indirectly promote cell adhesion and signaling, facilitating integrin clustering by funneling active integrins into adhesions and altering their state, by applying tension to these matrix-bound molecules (545). This in turn promotes focal adhesion assembly and facilitates integrin-dependent growth factor signaling to support cell growth and survival. Since a bulky glycocalyx is a feature of malignant cells, it is suggested that these features could foster the spread of cancer by mechanically enhancing cell surface receptor function (545). However such mechanisms are also likely to operate in normal cells, which presumably exist in a continuum of biophysical states of the glycocalyx.

Antiadhesive action
Large acidic polymers such as hyaluronan and polysialic acid can inhibit cell–cell and cell–matrix interactions by virtue of both bulk and negative charge. These antiadhesive functions are particularly prominent during phases of development when cell migration is very active. The “plasticity” resulting from polysialic acid expression appears to be important for neuronal migration as well as reorganization following injury (546–554).

Depot functions
Hydrophilic glycans on cell surfaces and extracellular matrices are capable of attracting and ordering water molecules (555). Beyond retaining water and cations (for unknown reasons, positively charged glycans are uncommon in nature), extracellular matrix glycosaminoglycans and polysialic acid can act as depots for growth factors and other bioactive molecules, which can be stored locally and released when needed, e.g., during injury and wound healing (86, 556–561).

Nutritional storage
Polymeric glycans like glycogen in animal cells and starch in plants serve obvious roles in the long-term storage of glucose as an energy source, and marathon runners must build up liver glycogen stores before the big race. The earlier comment about O-linked GlcNAc being the first known form of cytosolic glycosylation is not strictly true, as the glycogenin protein was also known to self-glucosylate itself on a tyrosine residue with a short 8–12 glucose residue polymer in order to serve as the primer for glycogen synthesis (562–567). In contrast, the mechanism of potato starch biosynthesis appeared to involve de novo synthesis, not an amylogenin primer (568).

Gradient generation
Gradients of growth factors can be generated by binding to extracellular matrix glycosaminoglycans such as heparan sulfate, especially in embryonic development (86). This organization of growth factors by glycosaminoglycans may contribute the morphogen gradients that are critical during development (569–572).

Extracellular matrix organization
Many components of the extracellular matrix in vertebrates are large glycan polymers such as sulfated glycosaminoglycans and hyaluronan, that self-organize along with specific proteins into larger aggregates to generate structures such as basement membranes (573–574) and cartilage (575–577). Cartilage also acts as a template for primary and secondary ossification centers, development of the growth plates and the end of long bones, and the laying down of bone (578). Organizational roles are also obvious for glycans in the extracellular matrices of plants (see discussion above), and new roles are emerging for glycans in the biofilms surrounding bacteria, enabling them to form discrete multicellular communities (579–583).

Protection from immune recognition
The adaptive immune system of vertebrate organisms functions largely by recognition of foreign peptide sequences, which are directly recognized by the B cell surface Ig receptor (584–585), and are also loaded into the grooves of the major histocompatibility receptors to be presented to specific T-cell receptors (586). If the peptide carries a very small glycan, this moiety can contribute novel specificity to recognition (587–591). However larger glycans typically disrupt peptide loading and/or T-cell receptor recognition, and often eliminate it altogether. This explains a common immune escape strategy of enveloped viruses, whose surface glycoproteins tend to be very heavily glycosylated (592–593). Sometimes, such protective glycosylation can become so dense that it generates unique clustered epitopes recognized by specific antibodies, such as that seen on the surface of the HIV virion (507–509, 594). In other instances, one type of glycans can block immune recognition of another, such as the Cryptococcus neoformans yeast cell wall, which is required for virulence (595–596), apparently by protecting the deeper structures of the organism from recognition and attack by the host immune system.

Effects of glycan branching on glycoprotein function
The N-linked glycans on cell surface glycoproteins can have varying degrees of branching (597), and glycan branching is specifically upregulated in T-cell activation (598), and in malignant transformation (599–604). Beyond their effects on protein structure per se, certain branched glycans can affect a variety of biological functions. Thus, there are reports of regulation of cytokine receptors by modulation of endocytosis rates by the type of glycan structure and branching (605), and N-glycan number and degree of branching can cooperate to regulate cell proliferation and differentiation (606) as
well as thymocyte positive selection (607). The degree of branching of N-glycans is primarily dependent on the addition of β-linked GlcNAc residues donated by UDP-GlcNAc. Given that glucosamine and GlcNAc are major metabolic intermediates in most cells, the level of UDP-GlcNAc provides a likely connection between cellular metabolism, cell surface organization and disease (608). In keeping with this concept, the cell surface residency time of glucose transporter 2 is regulated by branching of its N-glycans, and alters insulin secretion as well (609). This provides a connection between diabetes, pancreatic β cell glycosylation and glucose transport (610). A different kind of N-glycan branching (so-called bisecting GlcNAc) inhibits growth factor signaling and retards mammary tumor progression (611) and E-cadherin may be a target molecule for this glycan modulating effect (612–613).

Cell surface glycan:lectin-based lattices
The glyocalyx on the surface of vertebrate cells is often likened to a semi-randomly organized tropical rain forest (146), or to a sea-floor kelp bed (the latter analogy by P. Gagnex adds water and motion to the image, making it even more realistic). But the glyocalyx is also suggested to include self-organizing ordered lattices of glycans and lectins (83, 614–615). An intriguing connection has been established between the glycan branching phenomena mentioned above and the formation of such cell surface lattices involving galectin recognition of polylactosaminoglycans, which tend to be enriched on highly branched glycans (117, 606, 608, 616). The concept is that an ordered lattice forms within the glyocalyx that alters interactions between cell surface molecules, also affecting their rates of clearance from the cell surface by endocytosis. Thus, evolutionary selection is suggested to have modulated the number of glycans of inhibitory versus activating growth factor receptors, such that branching (controlled by UDP-GlcNAc and GlcNAc transferases) can differentially affect their relative ratio on the cell surface, by altering cell surface residence times (606, 608, 616, 616–617). Complexity arises because capping of polylactosaminoglycans by sialic acids can modulate galectin recognition by its presence and or linkage type (618–622).

Masking or modification of ligands for glycan-binding proteins
In some cases, modifications of monosaccharides and/or specific monosaccharides themselves can act as biological masks that prevent the recognition of the underlying glycan by specific glycan-binding proteins (623). Classic examples can be found in the case of terminal sialic acid wherein O-acetyl modifications can block the binding of some influenza viruses (22, 623–624), and the removal of sialic acid itself can unmask binding sites for receptors or antibodies that recognize subterminal β-galactose residues (22, 623). In another example, certain enzymes called Sulfs mediate extracellular removal of binding sites for heparin sulfate ligands (625), which can then signal through other receptors, e.g., Wnt/frizzled or IFN-βIFNAR (626–628).

Tuning a range of function
The size, number, branching and degree of sialylation of N-glycans can generate numerous glycoforms of a single polypeptide such as erythropoietin (629–640) or GM-CSF (641–644). It turns out that the nature of the glycosylation, extent of branching and level of sialylation modulate the activity of such cytokines over a range of function, by affecting its interaction with its cognate receptor, and also by altering the rate of clearance from the circulation. In passing, it is worth mentioning that differences in the sulfation and sialylation of the N-glycans expressed on endogenous versus exogenous erythropoietin are used by the Anti-Doping Agency to detect illicit administration, and has led to rescinding of many major sporting trophies (645–646).

Molecular functional switching
The once obscure (310), but now well-known and widespread, O-GlcNAc modification of nuclear and cytoplasmic proteins, has been shown to be a multifunctional molecular switch, which can work with, or in competition against, Ser/Thr phosphorylation, altering the functions of a wide variety of modified proteins and affecting numerous physiological and pathological processes. This remarkable system and its numerous ramifications have been well reviewed elsewhere (21, 70, 647–654), and will not be addressed in detail here. Other forms of nucleoeroplasmic glycosylation have since been discovered and characterized functionally in many organisms. For example, oxygen-sensing in diverse protozoa depends on prolyl-4-hydroxylation of the E3(SCF)ubiquitin ligase family subunit Skp1, and modification of the resulting hydroxyproline with a series of sugars. In the social amoeba Dictyostelium, O2 availability is rate limiting for hydroxylation of newly synthesized Skp1. Knockout mutants of the Skp1 prolyl hydroxylase and each of the Skp1 glycosyltransferases confirmed that O2-dependent post-translational glycosylation of Skp1 promotes association with F-box proteins and their engagement in functional E3(SCF)Ub ligases, which in turn regulate O2-dependent developmental progression (653–659).

Returning to the extracellular compartment, another classic example is the modulation of IgG effector functions by the structural features of the N-glycans in the IgG-Fc region (660–663). Incomplete galactosylation of these glycans has been associated with chronic inflammatory diseases (664–666), and there are clear-cut effects of IgG-Fc glycan core fucosylation on antibody-dependent cellular cytotoxicity (667–671) that are relevant to biotechnology (672–673). Sialylation of a minor fraction of the IgG-Fc N-glycans also appears to convert the IgG molecule into an inhibitor of inflammation, and is thought to underlie the anti-inflammatory properties of therapeutically delivered intravenous immunoglobulin in humans. While many papers have been written about this effect, there is still some controversy about the extent of the effects, and the details of mechanisms in different models and species (674–701). One possible explanation for the confusing results is that the immune responses are subject to “hormesis”. This is poorly appreciated but common biological phenomenon wherein low and high doses of the same stimulus can result in opposite biological responses and outcomes (702–705). Regardless of controversies, it is clear that subtle changes in the glycosylation state of IgG-Fc regions of N-glycans can profoundly influence not only circulating half-life, but also the effector function of antibodies (706). Meanwhile, contrary to 40 years of X-ray crystallography suggesting immobility of the Fc-region N-glycan, recent NMR studies suggest that this glycan is actually mobile and dynamic in solution (707). Thus, it is possibly the range of motion of the glycan that is being altered by the various modifications, secondarily affecting interactions with Fc-receptors of various types via an allosteric mechanism (708). Finally, while much of what is stated above refers to IgGs prepared for therapeutic use, it appears that modulation of IgG-Fc glycosylation occurs naturally in vivo, in various inflammatory and infectious conditions (687).
although the mechanisms of modulation are largely unknown. More recent evidence suggests that the Fc-region of other Ig classes may also alter effector function (709–710).

Epigenetic histone modifications
It is now clear that the addition of O-GlcNAc residues to histone proteins surrounding chromosomal DNA is a key component of the histone code that regulates gene expression. O-GlcNAcylation targets key transcriptional and epigenetic regulators including RNA polymerase II, histones, histone deacetylase complexes and members of the Polycomb and Trithorax groups. Given its dependence on cytosolic UDP-GlcNAc levels, O-GlcNAc cycling is thought to serve as a homeostatic mechanism linking nutrient availability to higher-order chromatin organization. Evidence also suggests that this “simple” glycosylation mechanism can also influence X chromosome inactivation and genetic imprinting (650, 652–653, 711), which may be related to the fact that the O-GlcNAc transferase is encoded on the X chromosome.

Extrinsic (interspecies) recognition of glycans
As mentioned earlier, it is not at all surprising that numerous pathogens and symbionts have evolved highly specific ways to recognize aspects of the dense and complex forest of cell surface glycans they encounter in host species. These interactions often involve glycan-binding proteins (see Figure 3), and can result in symbiosis, commensalism or disease, depending on the interaction in question and on the biological circumstances. A few examples from this vast field of knowledge are mentioned.

Bacterial, fungal and parasite adhesins
Among the numerous examples that can be cited for bacterial adhesins (223, 712), the example of Helicobacter recognition of gastric sialoglycans is particularly interesting, given its role in pathogenesis of gastric ulcers and cancers (713–719). The F-pilus mediated glycan-dependent binding of uropathogenic Escherichia coli accounts for millions of urinary tract infections a year (720–721) and small molecule inhibitors are being explored as therapies or prophylactics (722). With regard to parasites, a well-known example is the merozoite stage of Plasmodium falciparum, which initiates malaria via recognition of densely sialylated glycoporphins on target erythrocytes (723–730), with the types of sialic acids presented affecting species specificity (731–733). At a later stage in malaria, heparan sulfate on endothelial cells mediates the binding of P. falciparum-infected erythrocytes via the DBL1α domain of PfEMP1 (734), likely accounting for some of the most serious complications of the disease. Specificity for host glycans also plays a role in the binding of a pathogenic fungus (Candida glabrata) to various target tissues (735).

It is notable that in many instances expert researchers eventually find “glycan-independent” mechanisms of pathogen interaction with target cells and sometimes assume that the glycan-dependent process is therefore unimportant. However, such studies are often done in static conditions with long contact times, making the initial glycan “handshake” less critical. The situation is quite different in real life, where opportunities for contact and infection may be transient and difficult. Of course, with increasing evolutionary time a highly successful pathogen may come to rely more on glycan-independent mechanisms, as appears to be the case with endemic P. falciparum infections (727, 736–738).

Viral agglutinins
By tradition, viral glycan-binding proteins are called hemagglutinins, because many were originally discovered by virtue of their ability to agglutinate erythrocytes (which ironically are noninfectable, because they do not have the machinery for viral replication). Of these the best known is probably influenza hemagglutinin (the “H” in “H1N1”), which plays a key role in the infection process of this highly successful group of viruses. Much has been written about the specificity of the binding of these pathogens in relation to the sialic acid ligand, particularly the specific linkage to the underlying sugar chain, which determines the preference of the virus for avian versus human hosts (739–749). The evolution of the avian influenza viruses towards infecting humans involves selection for a change in binding specificity, which can be replicated experimentally (744). Interestingly, even our closest evolutionary cousins (chimpanzees) do not have a high density of human sialic acid composition or linkage on their airway epithelium (750), explaining the lack of nonhuman primate models and the unlikely choice of the ferret as a model for human influenza—because it happens to express the human-like linkage on its airway epithelium (751–752), and also because it turns out to, like humans, be missing the nonhuman sialic acid Neu5Gc (753). There are other examples of even more exquisite sialic acid specificity of viruses, based on the presence of O-acetyl esters at specific positions: a 4-O-acetyl ester on sialic acid targets is required for mouse hepatitis virus infection (754–756), a 9-O-acetyl ester on the sialic acid side chain is required for the binding of certain other coronaviruses and influenza C and D viruses (624, 757–761). The difference between these two specificities is dictated by only a few key amino acid changes in the viral receptors (762–763).

Bacterial and plant toxins
Many soluble plant and bacterial toxins mediate their effects by binding to target glycans on cells of another species. Typically, a bacterial glycan-binding B subunit is multimeric and serves to bring the toxic A subunit close to the membrane, whereby the latter then crosses over to mediate its toxic actions in the cytosol (with or without endocytosis). Classic examples include cholera toxin, which binds GM1 ganglioside (256, 764), the infamous ricin toxin that binds to terminal β-linked galactose residues (765–766), and the enteroto-hemorrhagic E. coli/Shiga verotoxin that recognizes globotriaosylceramide (Gb3Cer) and globotetraosylceramide (Gb4Cer) glycosphingolipids (767–769). The precise spacing of target ligands can be very important to the optimal binding of the pentameric lectins to the target (770–771). The single oxygen atom difference between the Neu5Ac and Neu5Gc forms of sialic acids can also determine the specificity of toxin binding, such as in the cases of the typhoid (772) and SubAB (773) toxins. In some cases, there is also evidence of dual specificity, e.g., fucosylated blood group structures on glycoproteins may contribute to cholera toxin binding (774), via an independent binding site (775).

Soluble host proteins that recognize pathogens
Vertebrates also express toxic glycan-recognizing peptides that can attack bacteria. For example, the small intestinal mucus layer is rich in RegIIgamma, a secreted host antibacterial lectin, which is essential for maintaining partial sterility of a ∼50-μm zone that physically separates the luminal microbes from the intestinal epithelial surface (776). Also, host galectins have been found to have unexpected
toxicity towards bacteria via recognition of their surface glycans (777–778). Killing occurs rapidly and independently of complement and is accompanied by disruption of membrane integrity. Galectin-3 may also play an important role in innate immunity against infection and colonization of Helicobacter pylori (779). Galectin-1 can have dual and opposing effects on virus infection of human endothelial cells (780). In other instances, circulating soluble multimeric (typically pentameric) glycanc-binding proteins recognize surface glycans of foreign pathogens but do not directly kill the pathogen. Instead they provide a signpost to attract other active components of the immune system such as complement and macrophages. Examples including collectins like the mannann-binding lectin, and the ficolins (334, 781). Indeed this kind of triggering of innate immune reactions via multivalent recognition of foreign glycans represents some of the most ancient and effective forms of immunity. For example, the hemolymph of horseshoe crabs recognizes invaders through a combinatorial approach, using lectins with different specificities against glycans exposed on pathogens (782), allowing these organisms to survive almost unchanged for >100 million years, without the benefit of adaptive immunity.

Pathogen glycosidases
Numerous pathogens generate a diverse array of cell surface and secreted glycosidases that serve to remodel or destroy the host glycoalyx, sometimes then utilizing the released monosaccharides as food sources and/or providing a nutritional resource for other microorganisms in the same milieu (783–785). Some mammals also rely on symbiotic microorganisms within their digestive tract to gain energy from plant biomass that is resistant to mammalian digestive enzymes (786). In other instances, the glycosidase acts in a balance with the binding activity of the same pathogen (787). For example, the sialic acid-binding (“hemagglutinating”), H activity of the influenza viruses is balanced by the activity of its sialic acid releasing enzyme (neuraminidase, N), the latter working both to allow the virus to gain access to cell surfaces by cutting through interfering molecules (788–789), and also to allow release from cells after replication (790). The elegant structure-based design of a modified version of the previously known sialidase inhibitor Neu5Ac2en (791) gave rise to the potent and specific inhibitor zanamivir (Relenza) (792–793), and later to the structurally related agent oseltamivir (Tamiflu) (794–795). It is worth noting that oseltamivir is not a sialidase, since the absence of sialic acid on cellular surfaces is designed to mimic glycans. In yet other cases, microbial glycosidases remodel host glycans to generate the optimal receptor for subsequent binding. For example the secreted neuraminidase of Vibrio cholerae removes all but one specific sialic acid residue from host gangliosides, leaving behind the GM1 monosialoganglioside, which is the specific receptor for the B subunit of the AB5 exotoxin secreted by the same organism (796).

Host decoys
It has been suggested that circulating erythrocytes might act as non-infectable decoy receptors for glycan-recognizing viruses that gain access to the bloodstream (164). The thick layer of mucin glycans on the surface of epithelial cells lining hollow organs (139, 418, 448–452) also plays a critical role by providing decoy binding sites for pathogens, diverting them from their intended targets on the cells. Of course commensals may take advantage of such mucin binding to remain within their preferred ecological niche (797–800) and to favor dental biofilm development (801). But on the rare occasions when such bacteria accidentally find their way into the bloodstream, these same commensalism-favoring adhesins become virulence factors, mediating interactions with platelets, which act as carriers of the organisms to eventual infection of damaged heart valves (798, 802–807). This is a recurring theme at sites of interspecies interactions, wherein factors favoring routine commensalism turn into potent “virulence factors”, on the occasions when physical barriers are breached and/or host immunity is compromised.

Herd immunity
As discussed previously in the context of glycan evolution (164), herd immunity refers to a form of indirect protection from infectious disease, which occurs when a large percentage of a population is resistant to an infectious agent, effectively providing protection to individuals who are not immune. Since glycans are often the targets for many infectious agents, intra- and interspecies polymorphisms in the expression of such targets can provide herd immunity, and restrict the spread of disease. As an example the ABO(H) blood group system can affect the spread of a highly infectious noroviruses that selectively binds one blood group structure and not another (808–811). This is likely why not everyone is sick by the time a cruise ship suffering a norovirus epidemic makes it back to port. ABO blood group polymorphisms also appear to affect susceptibility to cholera, as the cholera toxin has a secondary binding site for such glycans (775). The high levels of competitive oligosaccharides in human milk likely provide protection to breast-fed infants against some viruses and toxins (812–814).

Pathogen-associated molecular patterns
It is now well recognized that innate immune cells also detect pathogen-associated molecular patterns (PAMPs) using Pattern Recognition Receptors (815–816), particularly Toll-like receptors (TLRs) (817–818), NOD-like receptors (NLRs) (819–822) and C-type lectins (823–826). Many PAMPs are glycoconjugates, e.g., bacterial lipo-oligosaccharides or glycan-based polymers, e.g., lipopolysaccharides and bacterial peptidoglycans, including bacterial DNA or viral RNA (which are (deoxy)ribose-based polymers) (827–828). Glucan and oligochitin oligosaccharides released from fungal cell walls can also function as elictors of plant defense (829).

Immune modulation of host by symbiont/parasite
In some instances, glycans molecules mediate symbiont or parasite modulation of host immune responses. For example, glycans such as polysaccharide A (an unusual pentasaccharide repeat), derived from important mammalian gut microbiome members, helps to modulate the host immune system into a more tolerant state (via T-reg engagement) (213). Similarly, glycans derived from parasitic worms alter the immune status of their long-term hosts (830), a process dubbed as “glycan gimmickry” (831).

Antigen recognition, uptake and processing
Antigenic proteins must first be taken up by antigen presenting cells (macrophages and especially dendritic cells), which process them into peptides, to be presented by MHC Class II molecules, for recognition by T lymphocytes. This process can be facilitated by glycans on the target protein. For example, the presence of high densities of terminal Man or GlcNAc residues on foreign proteins or microbes can trigger phagocytosis via C-type lectins on antigen presenting cells, with resulting delivery of the antigenic proteins to processing
compartments (832–834). As another example, nonhuman α-Gal (835) or Neu5Gc (836) residues carried on injected glycoproteins can result in immune reactions, and in the formation of immune complexes, which in turn can enhance immune reactivity against the peptide backbone. An alternative form of self/nonself recognition is exemplified by foreign glycolipid presentation by CD1 molecules, which are detected by restricted or invariant TCRs of NKT cells (837–842).

Bacteriophage recognition of glycan targets
The complexity and diversity of surface polysaccharides found on strains of a single bacterial species (>100 in the case of the pneumococcus) (843) might be explained not only by selection for evasion of the vertebrate antibody response (844), but also by the need to evade attack by environmental bacteriophages (843, 845), which often use bacteria surface glycans as targets for recognition, and sometimes subsequent cleavage. While studies are continuing (846–849), this remains a poorly explored area. Given the very high density of bacteriophages in nature (10 million viruses per milliliter of surface seawater!) (850), there are likely a huge number of such as yet undiscovered viruses with exquisite specificities for diverse glycan structures. Indeed, it is possible that a cognate bacteriophage exists for every variant of every bacterial surface polysaccharide that occurs in nature. Thus bacteriophages are effectively a massive reservoir of glycanc-binding and glycan-hydrolyzing proteins still waiting to be exploited for glycan analysis and bacterial diagnostics as well as therapeutics (851), e.g., a potential new source of therapeutic “enzymbiotics” (852) or disrupters of biofilms (853). Early steps in this kind of systematic search are promising (854).

Intrinsic (intraspecies) recognition of glycans
As mentioned earlier, numerous pathogens and symbionts have evolved highly specific glycan-binding proteins that can recognize aspects of the cell surface glycans they encounter in host species. For a long time, examples of glycan-binding proteins with clear-cut functions intrinsic to the same species (see Figure 3) proved elusive. Even when candidates such as the asialoglycoprotein receptor were found, their intrinsic functions were not obvious. Beginning with the discovery of the specific functions of P-type lectins in mediating lysosomal enzyme trafficking (discussed earlier), many examples of glycan-binding proteins with intrinsic functions are now well known, and participate in a wide variety of functions. Only a few examples are mentioned below.

Intracellular glycoprotein folding and degradation
In addition to the biophysical effects of attached glycans on nascent glycoprotein folding discussed above, specific recognition of certain glycan residues plays a key role regulating the process of ER-associated degradation (ERAD). After the unusual Glc₃Man₃GlcNAc₂-P-P dolichol structure of the lipid-linked oligosaccharide donor for N-glycosylation was first fully defined (855), it turned out to be identical in almost all eukaryotes studied. Conservation of this structure for more than a billion years of evolution strongly suggested that it serves a very important purpose. But while many features of the structure were clearly needed to ensure optimal N-glycan transfer (856–859), variations were possible in some parasites (860–861) and mutant cell lines (235), and the exquisite conservation of the native structure remained largely unexplained. A clue finally emerged when it was discovered that the third glucose residue on N-glycans is repeatedly removed and then put back on again during glycoprotein folding in the ER (862–863). This in turn led to the discovery that this terminal glucose residue is recognized by certain ER chaperones, calnexin and calreticulin (864–868). The key role of this glucosylation/deglucosylation cycle in protein folding is now well established (63, 71, 869–878). However, even after the last glucose residue has been permanently removed, there are further steps of recognition of the oligomannose type N-glycans that have been partially processed by ER mannosidases (879–884). These recognition events are mediated in part by mannos 6-phosphate receptor homology domains in several chaperone proteins, as well as additional mannosidase-like proteins and recognition complexes (130, 879, 881–888). Effectively, a byzantine array of glycan-modifying and glycan-recognizing proteins determines the final fate of a glycoprotein molecule in the ER—whether it will be allowed to go forward into the Golgi pathway towards its final destination, or be consigned for ERAD. And since most proteins that enter the ER are glycosylated, this system has a huge impact in normal and diseased states, as well as unfolded protein stress responses (883, 889). As mentioned earlier, O-mannosylation and O-fucosylation can also play a role in monitoring the folding of newly synthesized proteins. Proteins that fail to fold are eventually removed from harmful futile protein folding cycles and prepared for disposal, via reverse translocation into the cytosol. There is even a sophisticated cytosolic pathway for removing and recycling the N-glycans from misfolded proteins prior to the action of proteasomes, beginning with the action of a previously mysterious cytosolic Peptide: N-glycanase (890). Notably O-GlcNAcylated nucleocytoplasmic proteins can also occur cotranslationally, protecting nascent polypeptide chains from premature degradation by decreasing cotranslational ubiquitylation (891).

Intracellular glycoprotein trafficking
As discussed earlier, the classic example of glycan roles in intracellular trafficking is that of the mannose 6-phosphate recognition system for the targeting of lysosomal enzymes to lysosomes. There is now evidence for other lectin-like molecules within the ER-Golgi pathway, which likely modulate the trafficking of specific classes of glycoproteins. For example, the LMAN1 gene product ERGIC-53 in the ER-Golgi intermediate compartment is a mannose-selective and calcium-dependent human homolog of leguminous L-type lectins (892) and acts as a critical chaperone for the coagulation factors V and VIII during their biosynthesis in hepatocytes (893) and endothelial cells (894) respectively, also potentially affecting the biosynthesis of some other glycoproteins (895). Other examples are VIPL and VIP36 (896–901). Overall, it is reasonable to predict the existence of more such glycan-recognizing proteins in the ER-Golgi pathway, potentially involved in trafficking and/or chaperone functions.

Triggering of endocytosis and phagocytosis
A variety of cell surface receptors that recognize terminal glycans can trigger uptake of molecules (endocytosis), particles (phagocytosis) or even intact cells. The classic examples are the asialoglycoprotein receptor of hepatocytes and the mannose receptor of macrophages, mentioned in the introduction. A large variety of lectins are known to carry out endocytosis in macrophages and dendritic cells. Such recognition processes may be critical not only for providing antigens to process and present to T cells, but also for clearing away damaged cells or glycoproteins, such as occurs when microbial sialidases enter the circulation during sepsis and cause...
desialylation of platelets (902–903), or when cancers secrete incompletely glycosylated mucins (512).

**Intercellular signaling**

In addition to the plant oligosaccharins already mentioned in the introduction, oligogalacturonides released from pectins can also act as regulators in plants (141, 904–906). Another well-established example of intercellular signaling is represented by bacterial nod factors (820) (not to be confused with the nLRs of vertebrate inflammasomes), which communicate signals between rhizobacteria and the roots of their leguminous plant hosts (907–911), initiating the symbiosis that is eventually responsible for the bulk of the natural nitrogen fixation on the planet—a process key to the survival of many organisms that benefit from the resulting food chain. The chito-oligosaccharides that transmit the signal show structural specificity for organism and host (907–908) and appear to be detected by a specific lectin (911). Transferring this capability into nonlegume crop species is obviously an exciting prospect. However, this may not be easy, since nonlegumes recognize the nod factor via a mechanism that results in strong suppression of responses (912). In vertebrate systems hyaluronan fragments released during injury can be detected by TLRs, thus triggering host immune responses (467, 913–915).

**Intercellular adhesion**

The mechanism of species-specific recognition of disaggregated sponge cells has already been mentioned in the introductory sections. The selectin-based system for cell–cell interactions of leukocytes, platelets and endothelial cells has also been discussed. The fact that oral fucose feeding results in correction of leukocyte adhesion deficiency-II by restoring selectin ligands provides the genetic proof of concept of this system in humans (385). The role of selectin interactions in a variety of normal and pathological conditions like inflammation and cancer is now understood (309, 916–921), and therapeutic approaches are in evaluation (921–923). One particularly promising therapeutic outcome is based on the finding that selectins interact with sickled red cells and leukocytes in the circulation to facilitate endothelial adhesion and other interactions (924–927) ultimately contributing to vascular occlusion and “sickle cell crisis” (928). The effectiveness of the pan-selectin inhibitor GMI-1070 in reducing selectin-mediated cell adhesion and abrogating crisis shows much promise in early clinical trials (922, 929). Another classic example is the role of Myelin-associated glycoprotein (MAG, Siglec-4) in mediating key interactions between neurons and glia (930–932), a process critical for maintaining the stability of the myelin sheath that insulates axons (933–934).

**Cell–matrix interactions**

Evidence for critical matrix interactions with cell surface glycans can be found in the variety of muscular dystrophies resulting from altered glycosylation of the α-dystroglycan ligand for major matrix proteins such as laminin, described in the introduction (191–204). In another example, hyaluronan matrices synthesized by stressed cells recruit inflammatory cells are early events in many pathological processes (935–936). Interestingly, this is a process that most if not all cells undergo, when dividing in a hyperglycemic environment. This phenomenon likely impacts experiments of many investigators who use “standard” commercial tissue culture media, which actually have unphysiologically high amounts of glucose (937), at levels that might even cause diabetic coma in a patient.

**Fertilization and reproduction**

Many early studies suggested that glycan-recognition processes were a critical part of many sperm–egg interactions (75, 938–939). This field lagged behind for a while, partly because many researchers were looking for a single overarching glycan-recognition mechanism—until the realization that species-specific variations were in fact, exactly what one would expect! In a few instances such as in humans, specific glycans have now been identified as binding targets (940–941). Glycans also appear to be involved at many steps in the reproductive process, and in the processes of sperm migration to the site of fertilization (942). During the latter process, there is even evidence that circulating antibodies can enter the uterine fluid and destroy sperm carrying nonspecies-specific glycan antigens (943). After fertilization, there is evidence that glycans and glycan-binding proteins are involved in the processes of implantation (944) and placental functions (945–946) in mammals.

**Clearance of damaged glycoconjugates and cells**

Terminal sialic acids on circulating glycoproteins can be removed by endogenous sialidases during natural aging of the proteins (947), or suffer an attack by a pathogen expressing a sialidase. In either case, there would be exposure of underlying glycans recognized by specific receptors, such as the hepatocyte asialoglycoprotein receptor mentioned earlier. Data indicate that this kind of “eat me” signal may even mitigate the lethal coagulopathy of sepsis by clearing away damaged platelets (902–903). There also appears to be a very high-capacity system for clearance of incompletely glycosylated mucins by the liver (512). Such molecules are released in large amounts by cancer cells, but could also potentially appear during damage to otherwise healthy organs. The subset of cancer-derived molecules (e.g., CA125, Sialyl-Tn and CA19-9) that survive such clearance then become useful markers of disease progression (948–951). The value of such markers for early detection remains unclear (952), a problem known to plague many predictive serum markers (953).

**Glycans as clearance receptors**

Glycans themselves can act as clearance receptors for other molecules. For example, hepan sulfate proteoglycans in the liver space of Disse mediate clearance of triglyceride-rich lipoproteins independently of the well-known LDL receptor family members (954–955).

**Danger-associated molecular patterns**

Innate immune cells also recognize glycans released from tissue damage in vertebrates such as hyaluronan fragments (467, 914–915) and some matrix proteoglycans (936–957) as danger-associated molecular patterns (DAMPs) or “alarmins”, triggering responses similar to those generated by exogenous PAMPs (see earlier discussion). While fungal glycans can act as PAMPs to activate the host immune response, they can also instead mask other glycoconjugates to prevent such activation. Examples include *Candida albicans* (443, 958–959) and *Histoplasma capsulatum* (960).
Self-associated molecular patterns
As mentioned above, signals initiated by DAMPs and PAMPs are transduced via similar pathways, activating innate immune inflammatory responses. It was recently proposed that glycans could also act as self-associated molecular patterns (SAMPs) (961), being recognized by intrinsic inhibitory receptors to maintain the baseline nonactivated state of innate immune cells, and to dampen their reactivity following an immune response. A clear example of glycan-based SAMPs has been reported in the form of inhibitory Siglec recognition of cell surface sialoglycans (962–964), which may also provide a mechanism for the host to discriminate between infectious nonself and noninfectious self (965). Recent work (157, 966–968) has also affirmed prior evidence that sialoglycan recognition by factor H can blunt immune responses by inhibiting the alternate pathway of complement activation (966, 969). Siglec-9 recognition of hyaluronan may be another example of a SAMP system (970). Not surprisingly, these very same self-glycans are also common candidates for molecular mimicry by commensals or pathogens that engage these inhibitory receptors (see below).

Antigenic epitopes
In addition to the blood groups already mentioned, intra- and interspecies variations in glycosylation can result in strongly antigenic epitopes. Indeed a significant fraction of circulating Ig found in normal humans may be directed against foreign glycan antigens (971–973). Certain types of modifications of N-glycans found on plant and invertebrate glycoproteins can trigger immune reactions in humans (974–977), including therapeutic glycoproteins (835–836). In a more complex scenario, individuals exposed to Lone Star tick bites seem to develop IgE antibodies against α-Gal epitopes (humans do not have these epitopes). Upon subsequent exposure to mammalian foods that are rich in α-Gal motifs (such as red meats), individuals react (sometimes severely) in an apparent “red meat allergy” (978–979). On another practical note, glycans like α-Gal and the nonhuman sialic acid Neu5Gc represent the major xenoantigens that must be bypassed, to pursue the goal of xeno- (pig organ) transplantation into humans. In pursuit of this difficult goal, α-Gal- and Neu5Gc-double null pigs have recently been generated (980–983).

Xeno-autoantigens
It has recently been found that the nonhuman sialic acid Neu5Gc can become metabolically incorporated from dietary sources (particularly red meat) into certain cell types in the body, appearing on the surfaces of human cells as if it was synthesized by the individual (984–985). These “xeno-autoantigens” are recognized by pre-existing circulating “xeno-autoantibodies”, and the resulting “xenosialitis” is suggested as one mechanism for the epidemiological association between red meat consumption and the exacerbation of some common disease states, such as carcinomas and complications of atherosclerosis (984–985). It would not be surprising if other examples exist. One can imagine for example that bacterial nonulosonic acids or plant monosaccharides that are structurally related to host monosaccharides might be activated to their corresponding nucleotide sugars and then get transferred onto endogenous glycans at a low rate. While the efficiency of such a process would likely be lower than that of Neu5Gc, the resulting immune responses might be even stronger.

Molecular mimicry of host glycans
Given that the host immune system recognizes typical glycans found on many pathogens are PAMPs and that endogenous glycans function as SAMPs, it is not surprising that microorganisms have evolved ways to achieve molecular mimicry of host glycans. What is remarkable is the striking extent to which such mimicry has been achieved, via all imaginable mechanisms. Just a few examples are cited here, with an emphasis on sialoglycan mimicry by vertebrate pathogens.

Convergent evolution of host-like glycans
It was originally thought that pathogen molecular mimicry was being achieved via vertebrate to bacterial gene transfer. While there is continued controversy about the extent of horizontal gene transfer between prokaryotes and eukaryotes (986), most instances of glycan molecular mimicry by pathogens seem to involve convergent evolution of pre-existing pathogen biosynthetic pathways, or de novo generation of functional genes. Demonstrating the power of natural selection at the host-pathogen interface, Group B Streptococcus polysaccharides (987) display identity to specific host glycan structure such as the Neu5Gcα2-3Galβ1-4GlcNAcβ1- (which perfectly matches the structure of N-glycan antennae on many human glycoproteins), and Campylobacter species carry out near-perfect mimicking of complex brain ganglioside glycans (988–989). In the former case, it is evident that this mimicry allows the organism to imitate endogenous SAMPs and down-regulate innate immune responses by engaging the inhibitory Siglecs (962). In the latter case rare human immune responses against the ganglioside-like structures can even result in serious illness, with the complement-fixing antibodies damaging peripheral nerves (Guillain–Barré syndrome) (990–991). While Campylobacter sialylation may also modulate immune responses via Siglecs during sporadic contacts with humans (992–995), it is unclear why the organism (which normally lives in the chicken intestine) (996) has evolved this remarkable degree of molecular mimicry of vertebrate ganglioside. Perhaps there are Siglec-like inhibitory pathways in the chicken that have yet to be discovered?

Appropriation of host glycans
Continuing with the example of sialic acids as host molecular mimics, microorganisms seem to have evolved every other conceivable mechanism to achieve this goal. These mechanisms range from the simple acquisition of host sialylglycans (997) to the direct transfer of host sialic acids by trans-sialidases (998), to the highly efficient uptake of the small amounts of environmental free sialic acids (999) or even the direct utilization of trace amounts of CMP sialic acid present in host body fluids (968, 1000). In addition to acting as SAMPs recognized by Siglecs or limiting complement activation via factor H recruitment, such terminal sialic acids also serve to mask antibody recognition of underlying structures. The fact that numerous organisms have independently evolved so many different ways to decorate themselves with host-like sialoglycans (1001) speaks to the strong selection pressure for this mimicry.

As with examples mentioned earlier, these “virulence factors” may actually represent attempts at commensalism and symbiosis, which become pathological in some circumstances. Many other examples of host glycan mimicry can be cited (1002), such as the bacterial re-invention of hyaluronan (163, 1003), heparosan (the backbone of heparan sulfate) and chondroitin (the backbone of
chondroitin sulfate) (1004–1005). Interestingly, there appear to be limits to the “inventiveness” (constraints to convergent evolution) of microorganisms. Despite hundreds of millions of years of selection, no prokaryotes seem to have reinvented sulfation of glycosaminoglycan backbones, nor recreated the difficult biosynthesis of the non-human sialic acid Neu5Gc. While not exactly full-blown “mimicry”, most viruses simply take over and use the host glycosylation machinery to install glycans that mask and protect them from immune destruction. Examples of molecular mimicry by pathogens of plants or invertebrates need to be further investigated.

**Multifunctional roles of the single type of glycan**

The above classification of biological roles falls apart when one considers certain glycan molecules that can mediate many different types of roles, depending on the circumstance. An example is the lipophosphoglycan of Leishmania species, which is needed for establishment of initial infections in vertebrate hosts, and not for persistence or pathology (1006)—but is later needed for binding to galectins located on the surface of midguts of their invertebrate vectors (1007). A more striking example is the myriad functions of heparan sulfate proteoglycans, with different roles being mediated by slight modifications of the molecule (155, 425, 427, 572, 628, 1008–1011). This is also exemplified by the widely disparate phenotypes arising from genetic modifications in various steps involved in biosynthesis of the molecule. Thus for example, mice deficient in heparan sulfate 6-O-sulfotransferase-1 exhibit defective heparan sulfate biosynthesis, abnormal placentation and late embryonic lethality (1012) and autism-like socio-communicative deficits and stereotypes appear in mice lacking heparan sulfate only in the brain (1013).

Even a structurally simple molecule like polysialic acid (an a2-8-linked homopolymer of N-acetylneuraminic acid) can have a remarkable range of endogenous functions. For example, polySia has been implicated in numerous normal and pathological processes and phenotypes, including cell migration (1014–1016); cell differentiation (1014, 1016); neurite outgrowth (1017–1019); blockade of myelination (1020–1022); binding and modulation of neurotrophin function (559, 1023–1025); alteration of synaptic plasticity (1026–1028); effects on learning and memory (1029–1032); facilitation of repair following injury (546–547, 1033–1036); schizophrenia pathogenesis (391, 1037–1042); major depression (553, 1043–1045); bipolar disorder (391–392, 1044); alcoholism (1046); epilepsy (1047–1048) and social interaction (1049).

**Some questions and issues arising**

Some readers will likely feel that the examples of biological functions discussed are not the most striking ones, and others will doubtless complain that numerous additional functions have not been mentioned. Such deficiencies and limitations are simply an indication of how far the field has come in the last 20+ years since the last review in 1993. Let us conclude this incomplete attempt by discussing some questions and issues arising, and some future prospects.

**Why did glycans become the preferred cell surface covering during evolution?**

With the possible exception of transient bloodstream phase of certain parasites, there appears to be no exception thus far to the “rule” that the surfaces of all cells in nature are covered with a dense and complex coating of glycans, which is taxon-, species- and cell-type specific (178). If it had been biologically possible to evolve a living cell devoid of such a coating, such a cell would have no doubt emerged from >3 billion years of evolutionary selection. There is no single best explanation for this ubiquity of cell surface glycans, and several mutually nonexclusive ones can be considered. In addition to providing a physical barrier to protect the plasma membrane, glycans tend to be hydrophilic, often do not have rigid structures, and instead have significant freedom of motion in aqueous solution. These are the optimal properties for a class of molecules that interact at the interface with an aqueous environment. Also, it is difficult for cells coated only with proteins to evolve and escape mechanism from a pathogen that binds to a specific cell surface protein. Most amino acid changes are not usually well tolerated by proteins, impacting folding and/or stability or even rendering the molecule dysfunctional. In contrast, most intrinsic glycan functions are mediated not by a single absolutely required sequence, but by an ensemble of structures, spanning a continuum. Even the apparently “lock-and-key” example of Man-6-P recognition of lysosomal enzyme N-glycans discussed earlier actually involves a spectrum of Man-6-P bearing structures with a range of binding properties to two different M6PRs. In other words, many glycan functions are “analog” not “digital”. Thus, it is easier for a host to escape pathogens by subtly changing glycosylation (i.e., glycans may convey more robustness to the organism) without drastically altering intrinsic functions. Last but not least, a vastly greater number of structural variations can be generated via monosaccharide oligomerization and branching in comparison with nucleic acids or amino acids (1050). This increases the odds of evolutionary selection to escape from a cell surface interacting pathogen or toxin.

**Red Queen effects in glycan evolution?**

Given the above considerations, it is reasonable to suggest that glycans are particularly prone to Red Queen effects (running to stay in one place) (164). As illustrated in Figure 4, one can envisage several such effects involving glycan interactions, leading to a delicate balance between preserving endogenous function and evading pathogen attack (170). A more nuanced and sophisticated view that takes into account additional evolutionary considerations can be found in Figure 5 (173).

**What is the significance of lineage-specific deletions or additions of specific glycans?**

In contrast to the genetic code, there are many more species-specific variations in glycans, ranging from entire classes of glycoconjugates, e.g., sulfated glycosaminoglycans not found in prokaryotes, to specific glycans, e.g., the absence of α-Gal epitopes in old world monkeys (1051), and the independent losses of the sialic acid Neu5Gc in humans (1052), new world monkeys (1053), mustelids and related taxa (753), and apparently in sauropsids (the ancestors of birds and reptiles) (1054). In every instance of apparent lineage-specific loss, a careful phylogenetic analysis is needed to ascertain if the differences are due to gain or loss of a specific gene or pathway and/or due to convergent evolution (or particularly in the case of prokaryotes, horizontal gene transfer). Regardless of the underlying mechanisms, more studies are needed to understand not only the functions of taxon-specific glycans but also the biological significance of their loss in some lineages. Some data suggest that taxon-specific glycan losses may have played a role in protection from parasites like malaria (732, 1055), and even in speciation events, such as the origin of the genus Homo (943).
Why did evolution select O-GlcNAc as the dominant form of intracellular eukaryotic glycosylation?

In striking contrast to the bewildering diversity of extracellular glycosylation, there seems to be a limited number of forms of intracellular glycosylation, with a single modification (O-GlcNAc) numerically dominating the scene (647–648). One possible explanation is that this intracellular environment is not subject to selection pressures by myriad pathogens that express diverse and specific glycan-binding proteins. More specifically with regard to O-GlcNAc it has been suggested that the donor molecule UDP-GlcNac acts as an optimal metabolic sensor for multiple pathways, i.e., uridine, phosphate, glucose, nitrogen and acetate (649, 651–653).

Do free oligosaccharides in the cytosol and serum have specific functions?

During the N-glycosylation of glycoproteins in the ER, considerable amounts of unconjugated polymannose-type glycans are generated from breakdown of the lipid-linked precursor (1036). Later, unfolded glycoproteins that are returned to the cytosol for proteasomal degradation are first subject to a cytosolic PNGase enzyme that releases free oligosaccharides (890). Such free oligosaccharides are then subject to either further cytosolic catabolism or pumped back into lysosomes for degradation (1056). Even free complex N-glycans bearing sialic acids can also be found in the cytosol (1057). The question arises as to whether such glycans mediate any specific functions in the nucleocytoplasmic compartment, before they are degraded, such affecting transcription. Meanwhile, free sialyloligosaccharides related to N-glycans have recently been found in serum (1058), and may also have novel functions yet to be discovered.

Can a glycan-binding protein recognize more than one class of glycan?

Glycan-binding proteins are often discovered based on their recognition properties, and given names related to their initially discovered binding targets, e.g., Galectins bind β-galactosides (323), and Siglecs recognize sialic acids (329). However, many examples have emerged wherein well-known glycan-binding proteins are discovered to also bind to unrelated glycan class, sometimes not even obviously similar in structure. In some instances, this is simply because the protein in
question has two distinct binding modules, e.g., the L-type lectin mannose receptor can also have a separate R-type lectin module that recognizes sulfated GalNAc residues on putative glycoprotein hormones (1059). However, in many other cases, the binding region is shared or very close by. Thus for example, selectins that were originally defined by their binding to siaiylated fucosylated glycans can bind quite well to certain subsets of heparan sulfate glycans (1060–1061). Likewise, Fibroblast Growth Factor-2 can bind both to polylactic acid and heparan sulfate (1025), and Siglec-9 binds both sialic acids and hyaluronan (970). Recently, it has even been shown that some galectins bind efficiently to as yet undefined motifs on bacterial surfaces (777) as well as to some sulfated glycans (1062), in a manner still inheritable by its canonical ligand lactose. In most such instances, it is unclear what the shared glycan motif is. Given great dissimilarities in primary structure, cross-recognition perhaps arises from “clustered patch” combinations (146, 302) of components of more than one monosaccharide, such as hydroxyl, carboxyl, sulfate, acetyl groups, etc. Given existing difficulties in efficiently incorporating even small, defined cognate ligands into the binding pockets of crystal structures of glycan-binding proteins, it will be challenging to compare such disparate ligands and define the shared recognition components. Regardless, it is clear biological functions of a glycan-binding protein should not be assumed to be mediated by the canonical glycan ligand class that originally defined the name of the protein.

How many more glycan-binding proteins are there yet to be discovered?

Regarding extrinsic (interspecies) binding proteins, it is reasonable to predict that for every specific glycan found on the cell surfaces of a host there is somewhere, a pathogen or symbiont that has developed an exquisitely specific binding protein for the glycan. Indeed, if this vast array of binding proteins could be isolated and converted into useful probes, one could have a new approach to glycomics, which actually studies the intact (naturalistic) glycome, in a manner that is exactly as it is “seen” by binding proteins in nature. The first steps in this direction have already been taken, and the results are very promising (1063–1066).

With regard to intrinsic (intraspecies) binding proteins, the situation is less clear. However, the serendipitous mechanism by which many of them have been discovered suggests that a systematic approach to future discovery may be useful. Consider the case of sialic acid-binding proteins. As late as the 1970s, it was thought that sialic acids were just biological masks, and that there were no binding sites. In these and many other such instances, the ligand is actually a specific glycoform of the named polypeptide, and is only synthesized by the cell surface as nothing is present in only one copy. Thus effective affinity in nature is actually quite high. Regardless, in reality most glycans in aqueous solution are in constant motion and constitute an ensemble of many different shapes generated by many mobile bond angles, which are constantly interchanging (1074–1075). In order for the more rigid and ordered binding pockets of glycan-binding proteins to bind such “shape-shifting” glycans, they must actually “trap” one of the numerous possible solution conformations of the cognate glycan into the pocket, where the immobilized glycan can be seen in a small structure. Strictly speaking then the effective concentration of the true cognate glycan is far lower than that of the total glycan concentration, likely in the nanomolar range. Exceptions may arise when the glycan targets are restricted in their motion, forming “clustered saccharide patches”, such as on the surface of densely glycosylated cells, mucins or viruses (146, 544).

Can we better define and name specific glycoform ligands for glycan-binding proteins?

In some cases, the natural ligands for glycan-binding proteins can be defined by the primary sequence of the cognate glycan, e.g., the Sambucus nigra agglutinin binds the motif Siaα2-6Gal[NAc]. However, in other instances the ligand is a specific glycoform of a particular glycoprotein, which is difficult to define in terms of a cognate glycan sequence. This results in the inadvertently erroneous statements, implying that the polypeptide is the ligand, e.g., “PSGL-1 is the ligand for P-selectin” (1076); or “CD24 is the ligand for Siglec-10” (965) (PSGL-1 and CD24 are actually the designated names of the core polypeptides). It also leads to incorrect assumptions, e.g., that a glycoform of CD24 must automatically be the ligand for mouse Siglec-G, the mouse ortholog of human Siglec-10 (965). In these and many other such instances, the ligand is actually a specific glycoform of the named polypeptide, and is only synthesized by certain cell types with the right kind of glycosylation machinery, i.e., the same carrier polypeptide does not serve as a ligand in other situations. In the case of a CD44 glycoform from hematopoietic stem cells that is a specific ligand for E- and L-selectin (1077–1078), the authors reasonably chose to rename the molecule altogether as HCELL (hematopoietic cell E- and L-selectin ligand) (921). However, this leaves out the useful information that the underlying polypeptide is CD44. A compromise may be to list the name of the polypeptide and use the superscript to indicate that it is a specific glycoform that generated the ligand in question, e.g., CD44HCELL or HCELLCD44. As with most nomenclature issues, it may be hard to find consensus on this matter. Suffice it to say that
there is need for a resolution, to make it easier to understand the literature on ligands for glycan-binding proteins.

Why has it taken so long to elucidate biological roles of glycans?

It is clear that glycans got “left out” of the initial phase of the molecular biology revolution of the 1980s, not only because they were more complex and difficult to study, but also because they were not part of the original “central dogma” (1079). This resulted in a peculiar distortion of the bioscience community, in which an entire generation of biologists (beginning in the 1980s) has been trained without much knowledge or appreciation about the structure, biosynthesis and roles of glycans in nature (1080). The relative lack of interest in glycans can also be partly traced to the early lack of understanding of their biological roles. So, why was it so difficult to elucidate biological roles of glycans? Some of the reasons are obvious, such as the technical difficulties in detection, analysis and manipulation in biological systems. Some additional considerations are outlined in Figure 6. Because of the information embedded in the template-driven biosynthesis of nucleic acids and proteins, it has been relatively easy to go from one to the other, using sophisticated yet facile experimental methods, and via bioinformatic predictions. Also as shown in the upper panel of Figure 6, the path to defining a specific function as being mediated by a specific protein has been relatively straightforward. In striking contrast the field of glycosciences originated in “descriptive” carbohydrate chemistry and biochemistry and remained in these domains for a long time. New glycans were discovered by a variety of means (such as those shown in lower panel of Figure 6) and their structure and biosynthesis were elucidated. Studies of changes in development and disease were almost guaranteed to show interesting findings, justifying further funding and research. It was also necessary to decipher the biosynthetic enzymes and mechanisms involved in generating each glycan. Thus there was plenty of interesting work to do, other than take on the most difficult task of elucidating function. Also, many of the functions of glycans tend to be “analog” and not “digital”, and many glycans have more than one disparate function. Finally, the rapid evolution of glycans has generated a lot of species-specific differences, making it difficult to find common themes applicable to all major model systems studies in biology. With the power of modern glycomics and the move to integrate glycosylation data into metatranscriptomic studies, it now possible to get past these difficulties and study the functions of glycans like never before. But we still then need to define the glycome.

What is the glycome?

It is clear that the glycome of an organism is far, far more complex than that of its genome, transcriptome or proteome, and it is only recently that “glycomics” has become practically feasible (1065, 1081–1089). Daunting and sophisticated as it is, most of what is called glycomics in 2016 still amounts to generating a “parts list” of all the glycans one can find in an given cell type or tissue at a particular point in time and space, i.e., similar to a peptide map of a mixture of proteins. In addition, current methods partially or completely destroy or miss labile modifications like acetylation, sulfation, phosphorylation, lactylation, pyruvylation, etc. More efforts are needed to discover all the glycan attachment sites on proteins and lipids, in a cell type in question (1089). Eventually, we need not only to define all of the above, but also to understand and visualize the conformation and organization of glycans on individual cell types and surfaces, in the form of “clustered saccharide patches” (146, 302), or glycosynapses (87). In the final analysis, full understanding of the biology of glycans will require this comprehensive type of view for which analytical techniques are yet to be defined. But great advances are being made by many investigators in all of the above levels of glycomics, and the future looks bright. Moreover, we can take advantage of the fact that pathogens and commensals have already spent million of years adapting to interact with the glycans of their hosts, and have already evolved highly specific binding proteins for recognition. Thus as mentioned earlier, an entire array of probes for defining the glycome is already available in nature, waiting to be isolated, characterized and eventually converted into practical tools, if necessary with further mutations. Initial steps in this direction are also very promising (1090).

What biological roles do glycans not mediate?

This rhetorical question seeks to emphasize that the biological roles of glycans are highly varied, and span the spectrum of possibilities. So the exceptions are few. So far there does not seem to be an example of multigenerational information transfer directly mediated by glycans, such as that mediated by DNA or RNA. But it has recently become evident that O-GlcNAc can modify RNA polymerase II, histones, histone deacetylase complexes and members of the Polycomb and Trithorax groups (1091–1092). Thus, it is suggested...
that O-GlcNAc cycling serves as a mechanism linking nutrient availability to chromatin organization, histone modification and epigenetics (711). It remains to be seen if such epigenetic effects can mediate intergenerational transfer, in a manner similar to other epigenetic marks. There are also no clear-cut examples of glycans acting as enzymes [if one excludes RNA from being considered as a polysaccharide, and intramolecular self-cleavage of PolySia (1093) as a chemical anomaly].

**Why the persisting lack of attention to this fundamental component of biology?**

Glycans are a major and integral part of all biological systems, and >3 billion years of biological evolution has failed to generate any life form on the planet that is not absolutely dependent on glycan chains for its existence. Yet the current situation is comparable to that in cosmology, with a standard model based on extant knowledge— that functioned well until it was realized that the bulk of the universe consists of dark energy and dark matter, which had been previously ignored. In effect, glycans have become the “dark matter” of the biological universe (1094), important yet poorly understood and therefore deserving special attention. However the levels of funding, the number of scientists involved and the scientific popularity of Glycosciences remain low. Many of the reasons are evident from the foregoing discussion. As mentioned earlier, a major issue is the fact that an entire generation of scientists has been trained with a limited knowledge of this class of molecules, and they are unlikely

---

**Fig. 7.** Approaches towards elucidating biological roles of glycans. The figure assumes that a specific biological role is being mediated by recognition of a certain glycan structure by a specific glycan-binding protein. Clues about biological roles could be obtained by a variety of different approaches. For detailed discussion of each approach, see the original reference. Not shown are newer methods taking advantage of the power of chemoenzymatic synthesis and the introduction of modified sugars with bioorthogonal reporter groups into biological systems. Drawing by R. Cummings, updated from ref. 160 with permission from the Consortium of Glycobiology Editors.
to now turn to studying them. Thus, a new generation of young minds needs to be educated, in this aspect of biology. The other major factor is the lack of easily available technologies for the synthesis and analysis of glycans. In this regard, the 2012 US National Academies/National Research Council report on the future of glycoscience concluded by recommending: “...transforming Glycoscience from a field dominated by specialists to a widely studied and integrated discipline, which could lead to a more complete understanding of glycans and help solve key challenges in diverse fields”, and emphasized the need to invest in education and technology development (1095). A more recent NIH working group report further emphasizes the need for training in Glycoscience (1080).

**Future prospects**

Many functions of glycans will continue to be discovered by the conventional processes of scientific investigation that will serendipitously come upon such functions. However, there are several potential systematic approaches to uncovering these functions that are depicted in Figure 7. Each approach has its pros and cons, which are discussed in detail elsewhere (160). Also not fully shown in this figure are newer methods taking advantage of the power of chemoenzymatic synthesis (62, 127, 158) and introduction of modified sugars with bioorthogonal reporter groups into biological systems (111). As with any biological questions, there are pros and cons of studying isolated cells versus intact organisms. And one must always ask how species- or taxon-specific a given function might be. All of approaches depicted in Figure 7 are rendered difficult in the case of glycan types that have numerous nonoverlapping functions in the same biological system. As the late Philip Majerus once put it, trying to decipher the roles of such molecules by preventing their synthesis or by destroying them after the fact is like “sifting through the ashes to find out how dynamite works”. Of course this problem is not unique to glycans. Complexity and pleiotropy are inherent in all of biology, and the same could be said of other post-translational modifications.

One would have to go back very many decades to find reviews about “biological roles of nucleic acids” or “biological roles of proteins”. The fact that such a review on roles of glycans was necessary in 1993 indicates how far behind we were in our understanding of their biology. As this update after 23 years shows, we have come a very long way, and one author now can barely scratch the surface of the topic in a single review. The time has come for the biology of glycans to be “mainstreamed” with that of the other major macromolecules that are universal to all life forms. But this requires a concerted effort on the part of all biologists and naturalists, to fully integrate the roles of glycans into their thinking about living systems. Once that happens, there will no longer be any need for writing another review like this one.

**Acknowledgements**

The author apologizes to the numerous excellent scientists whose work is not adequately discussed or cited in this review, and thanks Linda Baum, Tamara Doering, Hudson Freeze, Pascal Gagneux, Rita Gerardy-Schahn, Robert Halinwager, Herbert Hildebrandt, Laura Kiessling, Stuart Kornfeld, Lara Mahal, Mike Pierce, Nancy Schwartz, Christine Szymanski, Mukund Thattai, Naoyuki Taniuchi and Chris West for valuable comments and suggestions. Much of the inspiration for this review has come from ongoing discussions with these and other colleagues, especially current and former members of my lab and editors of *Essentials of Glycobiology*. Long-term support from the National Institutes of Health and the Mathers Foundation of New York is gratefully acknowledged.

**Conflict of interest statement**

None declared.

**References**


