

EVOLUTION AND DIVERSITY OF GLYCOMOLECULES FROM UNICELLULAR ORGANISM TO HUMANS

Richard D. Cummings*

Department of Surgery, National Center for Functional Glycomics, Harvard Medical School Center for Glycoscience, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA

Abstract

The emergence of glycomolecules in nature coincided with the dawning of the ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) world, along with the development of a colossal assortment of related monosaccharide building blocks, which vastly exceed the building blocks for any other class of macromolecules. The elaboration of glycomolecules by unicellular organisms, inside cells and on their surfaces and secretions, led to an explosion of diversity, in which unique types of glycomolecules evolved to be highly expressed in all kingdoms of life – monera, protists, fungi, plant and animal. In addition to nucleic acids, glycomolecules include glycoproteins, glycolipids, oligo- and polysaccharides, glycosylated RNA (glycoRNA), and many others. The glyco-modification of amino acids in proteins overcame the limitation inherent in the genetic code, which restricts protein structures to 20 amino acids. Thus, the virtually unlimited nature of protein glycosylation greatly expands the structures and functions of proteins. A highlight in the evolution of glycomolecules is the diversity in glycan structures, as each cell type in nature, from unicellular organisms to cells in metazoans and humans, generates incredibly complex and varied glycan structures. Thus, a central paradigm of glycomolecular evolution is that the collective glycan structures are the defining features unique to each organism. This article touches on the marvelous diversity of glycans and glycomolecules in nature that led to their dominance in the landscape of biology, as well as thoughts on evolution that likely gave rise to their elaboration, diversity, and evolutionary success.

Keywords

evolution • glycans • glycome • glycoproteins • glycosylation • monosaccharides • polysaccharides

Abbreviations

AF, aggregation factor; Arg, arginine; Asn, asparagine; bnAbs, broadly neutralizing antibodies; CBM, carbohydrate-binding module; CDGs, congenital disorders of glycosylations; CP, capsular polysaccharide; CS, chondroitin sulfate; Cys, cysteine; DNA, deoxyribonucleic acid; ER, endoplasmic reticulum; Fuc, fucose; G3P, D-glyceraldehyde 3-phosphate; Gal, galactose; GalNAc, N-acetylgalactosamine; GBP, glycan-binding protein; Glc, glucose; GlcA, glucuronic acid; GlcNAc, N-acetylglucosamine; Gln, glutamine; glycoRNA, glycosylated RNA; GPI, glycosylphosphatidylinositol; HA, hyaluronic acid; HAS, hyaluronic acid synthase; HG, homogalacturonan; HS, heparan sulfate; HydroxyLys, hydroxylysine; HydroxyPro, hydroxyproline; KDN, 2-keto-3-deoxy-D-glycero-D-galacto-2-non-ulonic acid; KDO, 3-deoxy-D-manno-octulosonic acid; LPS, lipopolysaccharides; LSDs, lysosomal storage disorders; LTA, lipoteichoic acid; Man, mannose; MOS, malto-oligosaccharides; mRNA, messenger RNA; Mtb, *Mycobacterium tuberculosis*; MurNAc, N-acetylmuramic acid; Neu5Ac, N-acetyl-neuraminic acid; Neu5Gc, N-glycolylneuraminic acid; N-glycan, asparagine-linked; O-glycan, glycan linked to OH group of various amino acids, e.g., Ser, Thr, Tyr; Ser – serine; PTM, post-

translational modification; RG-I, rhamnogalacturonan I; RG-II, rhamnogalacturonan II; RNA, ribonucleic acid; Sia, Sialic acid; Thr, threonine; tRNA, transfer RNA; Trp, tryptophan; Tyr, tyrosine; viral HA, influenza virus hemagglutinin; viral NA, influenza virus neuraminidase; XGA, xylogalacturonan.

Introduction: Glycomolecules are ubiquitous in nature

All cells of all organisms express a dazzling variety of glycomolecules on their membranes and in their secretions (1, 2). In addition, the two monosaccharides ribose and deoxyribose, comprise about one-third of all nucleic acids in ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), respectively. Glycomolecules other than nucleic acids include glycoproteins, glycolipids, oligosaccharides, polysaccharides, glycosylphosphatidylinositol (GPI)-lipid anchored glycoproteins, and the newly discovered glycosylated RNA (glycoRNA) (3, 4). These classifications are somewhat deceiving, as within any class there are many subclasses and the diversity is extraordinary. The term *glycan* refers to the *glyco* component of glycomolecules,

* Corresponding author e-mail: rcummin1@bidmc.harvard.edu

and generally denotes all kinds of saccharide classes (1). Unlike nucleic acids and proteins, glycomolecules are not generated from a template; thus, there is no information within the genomes of organisms that can accurately predict either glycan structure or diversity. The diversity of glycomolecules in nature is only partly known, as only a tiny fraction of living things has been analyzed. Nevertheless, available information supports the startling fact that all organisms make relatively unique glycans. Glycans in unicellular organisms, such as bacteria (prokaryotes), protists (protozoans), and yeast, are very different from those in plants and even the simplest invertebrates, and certainly different from those in vertebrates and humans (Figure 1).

This article highlights several concepts regarding the diversity and evolution of glycomolecules, but the complete scope of the subject would require a much larger treatise. References are provided to help guide readers to additional points of interest and further details regarding the subject, and are not meant to be comprehensive.

Unique non-template nature of glycan synthesis and glycan size

While the diversity of nucleic acids and proteins is vast, they arise from a relatively small number of building blocks. Moreover, the macromolecules themselves are made in a linear form arising from a template, either DNA or RNA, directing their specific sequences. However, glycans are created without templates, and collectively in nature they are comprised of hundreds of different monosaccharides, and perhaps even thousands of them (13, 14). A single organism may have a set of monosaccharides and derivatives that can differ greatly from that in another organism. Thus, because the building blocks can vary, and without a template guiding the synthesis, there appears to be no end to the complexity of glycans generated in nature. In contrast to nucleic acids and proteins, glycans may have multiple branches and each branch can convey unique determinants. Using just six different monosaccharides occurring within a hypothetical

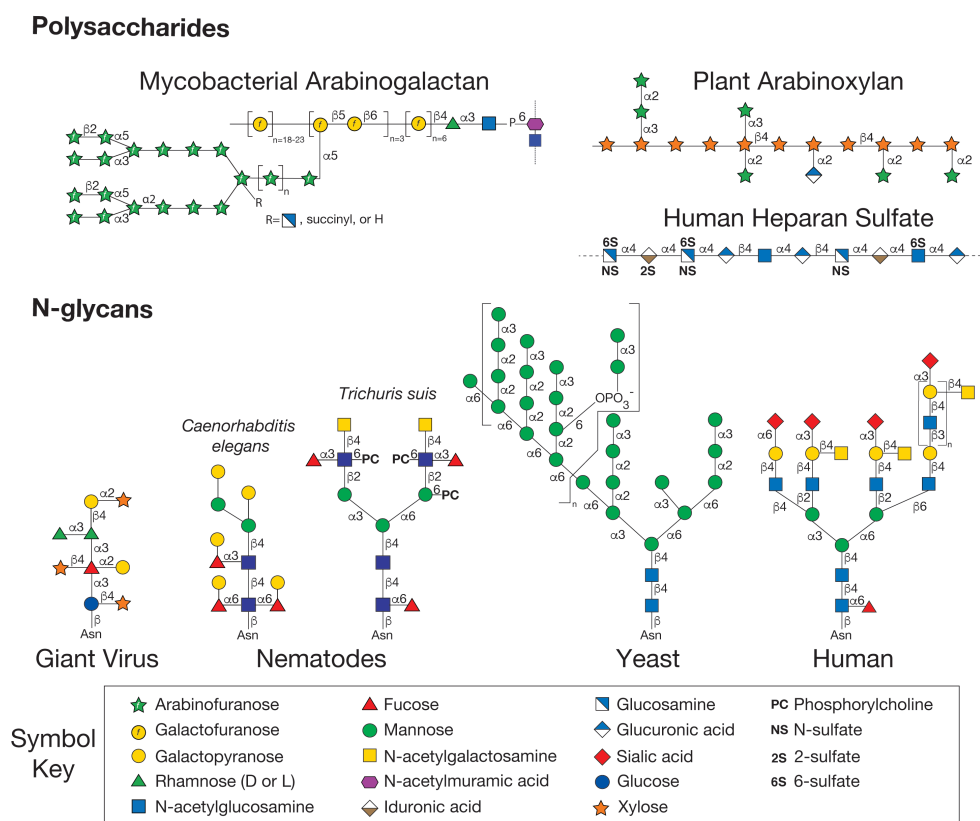


Figure 1. Here is depicted a few examples of the diversity of glycans found in glycomolecules of different organisms – mycobacteria, yeast, plants, nematodes, and humans. More details about these structures and discussions about their biosynthesis and function can be found in the following references (5–12). For more information about the symbols above used to designate monosaccharides see <https://www.ncbi.nlm.nih.gov/glycans/snfg.html>. Hundreds of different monosaccharides are known, and they can be found in multiple and unique combinations in all forms of life (13, 14). As depicted, the monosaccharide linkages are either α or β , and the linkage number, e.g., 2, 3, or 4, etc., refers to the carbon number on the monosaccharide to which the preceding monosaccharide is linked. Asn, asparagine.

hexasaccharide there are 1.05×10^{12} possible oligosaccharide isomeric products, both branched and linear, including their α and β anomers (15). By contrast, even using all 20 amino acids, there are only 6.4×10^7 possible linear hexapeptide products.

The term *glycosylation* denotes an *enzymatic* attachment of sugars to molecules, which is the focus of this article, whereas the term *glycation* denotes the *non-enzymatic*, chemical reaction of monosaccharides with amines in side chains of amino acids. This enzymatic control over glycosylation allows cells to generate specific glycan structures in an orchestrated fashion. In glycation, however, the initial chemical reaction is reversible and not specific, yet rearrangements can occur to generate irreversible adducts (16). Glycation is considered a pathological modification, as exemplified by glycation of hemoglobin and recognized as HbA1c (17), which is monitored in diabetes. Glycation might have been an early feature, however, of glycomolecular evolution. For example, pentoses, such as D-ribose, are more reactive than hexoses, such as D-glucose, but the potential of such evolutionary contributions of glycation are not well understood.

The sizes of glycomolecules in nature vary enormously; they may be tiny or massive, regardless of the organism making them. Amazingly, even the tiniest infectious and transmissible pathogen, the prion, is a single glycoprotein with multiple glycans enzymatically attached by infected cells; importantly, its glycosylation is essential for its neurotoxicity (18). Finally, transfer RNA (tRNA) (19) and non-coding RNAs can be glycosylated (3), and even DNA itself can be glycosylated, as observed for the occurrence of glucosyl-5-hydroxymethylcytosines in phage DNA, which is added by phage-encoded α - and β -glucosyltransferases (20).

In terms of size, consider the common linear polysaccharide hyaluronic acid (HA), which is synthesized by both microbes and animals. In animals HA is just one among the many glycosaminoglycans (GAGs), e.g., chondroitin sulfate (CS) and heparan sulfate (HS), that evolved early in evolution (21–24). Microbial and human HA are identical in structure, although the lengths may be different; a single HA molecule can have ~25,000 repeat units of the disaccharide *N*-acetylglucosamine-glucuronic acid (GlcNAc-GlcA), thus ~50,000 monosaccharide units, and be 0.25–25 μm in length with a size up to 10,000 kDa (25). This is larger than titin, the largest protein found in nature, which is ~3800 kDa and contains ~34,000 amino acids (26). In animals, mucins represent some of the largest glycoproteins, with among the largest being MUC16, a mucin of ~22,152 amino acids and up to 2000 kDa; >50% of its mass is carbohydrate and it can

be 1–5 μm in length (27–30). Gel forming mucins can be cross-linked to make larger gigantic polymers of 100 mDa or more, as seen for gel forming mucins (31, 32). Remarkably, such mucins are synthesized intracellularly within the tiny Golgi apparatus, and are packaged and released to the membrane by secretory vesicles; however, how the folding and organization of such large glycomolecules is coupled to their synthesis and cross-linking is only partly understood (33–36).

Microbes can also generate enormous networks of covalently-linked glycomolecules. For example, the cross-linked network of peptidoglycan with repeats of *N*-acetylglucosamine-*N*-acetylmuramic acid (GlcNAc-MurNAc) that encases bacteria, allows them to resist osmotic pressure. Its overall size can be immense, as it is isolatable as peptidoglycan sacculi, in the range of 3×10^9 Da (37). Finally, the most prevalent polysaccharide on earth, cellulose, comprised of linear, repeating units of the monosaccharide glucose, can be up to 10,000 monosaccharides in length (38). At the other extreme, the smallest glycopeptides in nature include the glycopeptide antibiotics, such as vancomycin, made by the soil bacterium *Amycolatopsis orientalis*, which contains a disaccharide L-vancosaminyl- α 1-2-D-glucopyranosyl linked to a 7 amino acid heptapeptide (39). The smallest free glycans, which are disaccharides, include the common disaccharide lactose, found only in mammalian milk, and the disaccharide sucrose, made by plants and microbes.

Thus, glycan diversity is a feature of all organisms. For any particular organism the assortment of glycans is referred to as the *glycome* (40), with the understanding that such glycans can be linked to other molecules. In higher metazoans the diversity of an organism's glycans, in which each is considered independently as a bioinformatic unit, rivals, if not exceeds, the diversity of information units of either nucleic acids or proteins, represented by individual genes and their encoded proteins.

Evolutionary considerations in glycomolecule diversity: The genetic code limitation

From these general considerations a simple question arises. What has driven the expansion and diversity of the building blocks (monosaccharides) and the resultant glycomolecules in nature? One likely explanation is that glycan diversity is driven by the same forces, e.g., mutation and natural selection, that drive evolution of other types of macromolecules, such as proteins (41). There have been many excellent reviews about aspects of glycomolecular evolution, often focusing on

specific classes of glycomolecules, certain types of glycan modifications, evolution of genes for enzymes, eukaryotic evolution, and other issues [some of them are cited here (6, 42–47)]. There are too many genetic pathways for glycomolecule generation to consider them all here, and thus, rather than go into all the interesting genetic aspects of glyco-evolution, herein is a complementary and slightly different approach with a focus more on the nature and diversity of glycomolecules in all forms of life and evolutionary challenges and opportunities associated with their generation.

The concept of evolution driving diversity of glycomolecules implies that there are pressures on organisms to evolve their glycomes in ways that enhance the possibility of their survival. The genome itself presents an interesting conundrum in this regard, which may have promoted the evolution of glycans very early. Consider the limitations inherent in the genetic code. The DNA genetic code is comprised of codons of three nitrogenous bases selected among the four – adenine (A), cytosine (C), guanine (G) and thymine (T). But this three bases codon, when transcribed into RNA with its four nitrogenous bases – adenine, cytosine, uracil, and guanine – can only accommodate 20 amino acids (plus selenocysteine and pyrrolysine), which universally make up all template-derived, complex proteins. Considering, however, all the vast roles that proteins play in biology, this relatively small diversity in side chains of amino acids limits the information content available to proteins.

One way of envisioning this is to imagine a polypeptide as a rope, linear yet very flexible, but restricted in the diversity of shapes and spatial organization it can assume. However, now imagine a rope carrying a virtually unlimited number of attached modifications that represent glycosylation, each with a potentially unique structure and composition at various *branchpoints* along its backbone; these create new physical properties of the rope. Thus, the potential advantages complementing protein structure that arise from the linkages of carbohydrates to proteins are obvious. The same protein may, when expressed in different cells, carry entirely different glycan structures are various branchpoints, and thus a single gene product, a polypeptide, may give rise to an incredible variety of modified forms.

Because of the differences between polypeptides and glycans, they evolved very different sets of functions. For example, some of the fundamental roles of proteins are to bind other molecules and to function in catalysis. Polypeptides are readily adaptable for such a function, both conformationally and structurally, to promote *microcatalytic* activities, as well as interactions with small molecules and intermediates, by which their unique amino acid chains are exploited (48). RNA is also adaptable for catalytic activity as ribozymes,

which are considered to be important in early evolution (49). However, glycans, with their large monosaccharide building blocks, cannot easily adopt relatively rigid microstructures comparable to those in proteins, and so far no catalytic glycan has been observed. Thus, glyco-components have what may be termed *macrobiological* activities. These include protecting the cell from environmental stress, protecting proteins from proteolysis, modulation of protein immunogenicity, cell-cell interactions, organization of cellular matrices and cell walls, skeletal structure, interactions with proteins, e.g., growth factors, lectins, etc., and other molecules, as well as facilitating larger aspects of protein conformation and function, and even modifying other macromolecules, such as lipids, to directly link membrane architecture to the extracellular space (50). In regard to the outer membrane of cells, a universal feature of all organisms is the asymmetric expression of glycomolecules; they are concentrated on the outer leaflet of the membrane, rather than the inner leaflet, whereas proteins can be on either side and even transmembrane in nature (51).

The modification of a protein with another molecule is termed a post-translational modification (PTM), although some can occur co-translationally (52). As the diversity and availability of monosaccharides in nature evolved, PTM-based mechanisms evolved to overcome limitations inherent in the structures and bioactivities of polypeptides (and lipids). To fully appreciate the concept of protein glycosylation as a PTM, consider first the phosphorylation of protein as a simple PTM. This is among the simplest and most common PTM of animal proteins, but it typically occurs inside the cell. It is a small, reversible modification, added and removed by enzymes (kinases and phosphatase), and regulates protein interactions. Other PTMs of proteins include acetylation, methylation, hydroxylation, and ubiquitination, among >500 known PTMs (52, 53). But all such modifications are either small in size, or specifically fixed in size, e.g., ubiquitin or SUMO proteins, and typically reversible. However, in regard to secreted and membrane glycoproteins, glycosylation is an essentially irreversible modification and the sizes can range enormously.

Of course, there are also processes of intracellular glycosylation, and unlike most PTMs involving glycosylation, some of these are reversible. An essential type of intracellular protein glycosylation in most metazoans (multicellular organisms) and some pathogenic fungi (54), and which can reside on cytoplasmic, nuclear, and mitochondrial proteins, is the addition of the monosaccharide GlcNAc in linkage to Ser/Thr residues (O-GlcNAc). This pathway uses UDP-GlcNAc as a donor for a novel O-GlcNAc transferase, which is a reversible modification, subject to removal by O-GlcNAcase (55). This modification occurs on thousands of proteins, including most nuclear proteins and nuclear pore proteins,

transcription factors, histones, and RNA polymerase II (56). O-GlcNAc addition along with phosphorylation, helps to organize and link nutrition, metabolism, cell signaling, and differentiation (57–59). O-Fucosylation by addition of fucose (Fuc) residues to nuclear pore glycoproteins has also been observed in the parasite *Toxoplasma gondii* (60), and also in nucleocytoplasmic proteins in the plant *Arabidopsis thaliana*, driven by the O-fucosyltransferase POFUT (61). Other intracellular protein glycosylation pathways also arose in protists to assist in oxygen sensing and involves modifications of hydroxylated proline (62).

Intracellular glycosylation of proteins is actually an ancient theme of glyco-evolution. The most famous and universal type of intracellular glycosylation in prokaryotes and animals involves the production of glycogen (63, 64), and in plants and algae the production of starch (65). In animal cells glycogen may originate from the protein glycogenin, which has inherent glucosyltransferase activity and auto-glucosylates a single tyrosine residue of itself using UDP-Glc to generate O-Glc; this is then extended and branched by enzymes to produce the polysaccharide glycogen (66). Interestingly, glycogen in the mammalian brain has both glucose and glucosamine residues, derived from UDP-glucosamine (67). Glycogen in animals can be a glycoprotein and reside in the cytoplasm of cells as large particles (β particles), some of which can be $\sim 10^8$ Da and ~ 300 nm in diameter (68). However, whether glycogenin is absolutely required for glycogen synthesis is unclear, as some humans and engineered mice can generate glycogen in the absence of glycogenin-1 (69, 70). Starch synthesis arises by extending malto-oligosaccharides (MOS) or existing starch primers (71).

As mentioned above, the addition of one or more monosaccharides to an amino acid allows the sizes to vary from a single monosaccharide to its elongation by addition of other monosaccharides up to many hundreds, far beyond any other type of PTM in size and diversity. This breakthrough in evolution allows even the tiniest of the amino acids, such as serine or threonine, to become giants, as seen for the O-glycans in nature (Figure 2). The addition of sugar to an amino acid may be thought of as the generation of a *new* amino acid with a *novel side chain* that can range in size from small to truly enormous. Interestingly, of the 20 canonical amino acids, 10 of them or their derivatives may be glycosylated, including asparagine (Asn), arginine (Arg), cysteine (Cys), glutamine (Gln), hydroxylysine (HydroxyLys), hydroxyproline (HydroxyPro), Ser, Thr, Trp (tryptophan), and Tyr (72–77), and all but the glycosylation of HydroxyPro have been documented in animal cells. Glycosylation most commonly occurs on –OH groups of certain amino acids (O-linked), or on amide (N-linked), but can also occur in other ways, including

C-linked, S-linked, etc. Thus, the DNA template need not encode any additional amino acids, as it can easily expand its repertoire by modifying amino acids. Glyco-modifications of proteins in this way are the only PTMs in nature that can be vastly larger collectively than the protein to which they are linked and can occur at multiple sites on a protein. Multiple modifications of a single polypeptide at different sites can generate a tremendous array of what are termed *glycoforms*, i.e., different glycosylation of a protein with the same polypeptide backbone. For example, a single glycoprotein, the recombinant human enzyme Myozyme, was recently shown to contain more than 10,000 different glycoforms (78). Also, the most infamous glycoprotein in modern times is the spike glycoprotein of SARS-CoV-2 virus, responsible for the recent worldwide COVID-19 pandemic. While this membrane glycoprotein is 17% by weight carbohydrate, its N- and O-glycans cover $\sim 42\%$ of its surface (79), and the total number of glycoforms is likely to be enormous.

The ancient nature of glycosylation

While glycomolecules are ubiquitous in nature, until a few years ago it was unclear as to whether organisms in all domains of life, *Eukarya*, *Bacteria* and *Archaea* (80), exhibited commonalities in the classes or types of glycomolecule they express. For example, it was commonly believed not long ago that bacteria did not synthesize glycoproteins. It is now clear, however, that glycoproteins are found in all bacteria, and in fact collectively they express amazing varieties of glycoproteins (81–88). Examples of glycoproteins expressed by Gram negative bacteria are S-layer glycoproteins in outer membranes (82, 89, 90), the multiple and essential glycoproteins produced by a single strain of *Bacteroides fragilis* (91), and the wide range of O-glycosylated (Ser/Thr) glycoproteins in the periplasm of *Flavobacterium johnsoniae* (92). Gram positive bacteria also express S-layer glycoproteins, and other surface glycoproteins, e.g., serine-rich repeat glycoproteins (93). Thus, expression of glycoproteins is a universal feature of life. Like all organisms, bacterial glycoproteins occur in secretions and on their outer membranes. Similarly, as for animal and plant glycoproteins, protein glycosylation in microbes has highly specialized biological functions. For example, the *Neisseria gonorrhoeae* pilin glycoprotein has a critical Ser-linked disaccharide needed for adhesion (94), the AIDA-I glycoprotein is necessary for pathogenic *Escherichia coli* adhesion (95), and the many dozens of glycoproteins containing O- and N-glycans in *Mycobacterium tuberculosis* affect immune reactions to the microbes (96, 97). There are ~ 80 glycoproteins that contain an Asn-linked heptasaccharide generated by *Campylobacter jejuni*, which are required for efficient cell invasion and colonization (98, 99). Such examples

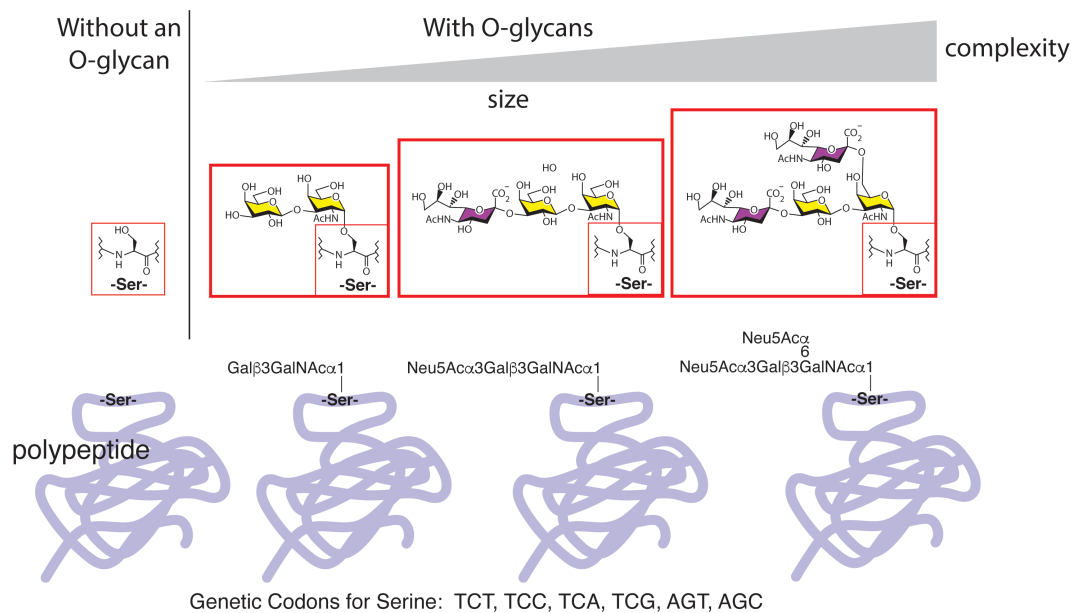


Figure 2. Modifications of proteins with carbohydrates can increase both the complexity and size, as illustrated here for simple sugar modifications of the amino acid serine, as often seen in O-glycans of animals. While the largest one shown here has a tetrasaccharide attached, some O-glycans in animals can reach enormous sizes and contain many hundreds of monosaccharide moieties. Moreover, multiple serine, threonine or tyrosine residues may also be O-glycosylated on a single protein, thus further enhancing diversity and size.

illustrate that glycoproteins are made by all organisms but their structural diversity and functions are often vastly different and specific to the organism.

A common biosynthetic concept for the elaboration of glycoproteins shared by bacteria and all other organisms regards the synthesis of lipid-linked glycans, broadly classified as glycolipids; they serve both as mature products made by cells and as precursors for generating other glycomolecules. A common type of protein glycosylation requires a lipid-linked sugar precursor. Such a *preformed* glycan can then be added to a protein or another glycolipid. For bacterial glycoproteins, this occurs such that the glycoprotein is modified within the periplasm at the membrane interface, as observed in *C. jejuni* for their Asn-linked pathway (100), and for S-layer glycoprotein generation (101). However, bacteria can also directly glycosylate proteins in some cases without preformed precursors. For example, sublancin, which is generated by *Bacillus subtilis* 168, has an S-linked (Cys-linked)-Glc, which is added by the enzyme SunS using UDP-Glc as a donor; this modification is essential for its antimicrobial activity (102). Also, all of the glycopeptide antibiotics made by bacteria are synthesized in the cytoplasm using nucleotide sugar donors and glycosyltransferases in pathways that are genetically traceable to be hundreds of millions of years old (103).

Cytoplasmic glycosylation of proteins also occurs in eukaryotes, as for the addition of O-GlcNAc and O-Glc, as discussed above. O-GlcNAc-containing glycoproteins also occur within the compartments of the mitochondria (104). This ancient nature of protein glycosylation in all organisms also includes the biosynthesis of nucleotide sugars; these are used to generate preformed lipid-linked sugar precursors or to directly glycosylate a protein (105, 106). The universality of glycosylating peptides and proteins either in the cytoplasm or elsewhere using nucleotide sugars and lipid-linked glycans has been extended and modified over time so that all types of cells, especially mammalian cells, can generate an astonishing number and variety of unique glycan modifications on an individual cell. Taken together, if we believe that longevity and ubiquity are any measures of success in molecular evolution, DNA and RNA are the most successful, but glycomolecules are a very close second.

Monosaccharide evolution

Monosaccharide evolution likely coincided with the formation of ribose and deoxyribose as these were incorporated into macromolecules essential for reproduction. Ribose is easily produced chemically by simple formose reactions involving formaldehyde and water, and proceed through glycolaldehyde

and the formation of aldoses and ketoses (107, 108) (Figure 3). Glyoxylate and C6 aldones, which can transform to ribose, are also candidates as evolutionary precursors of sugars (109, 110). Multiple lines of evidence indicate the primitive nature of formation of sugars in the prebiotic world (109, 111–113). The Murchison meteorite contained both aldonic acids and ribose (114, 115). Glycolaldehyde, the smallest possible molecule with both an aldehyde and an –OH group, has been detected in interstellar space and could be a source of pre-RNA intermediates (116). It is also interesting that sugar or alcohol alternatives to ribose and deoxyribose were probably available for inclusion in nucleic acids, e.g., threose nucleic acid or glycerol nucleic acid. However, neither their structural features, nor those of any other monosaccharide, competed effectively with ribose and deoxyribose for the formation of RNA and DNA, respectively (107, 117–119). Deoxyribose within DNA might provide greater flexibility and a higher stability to alkaline conditions compared to RNA, among other differences (120). Early in evolution, it is likely that mechanisms for the synthesis of the nitrogenous bases, select amino acids, lipids, and monosaccharides were established and conditions were such that these building blocks were at least semi-stable.

The generation of additional monosaccharides, D-fructose and D-glucose, can arise by related pathways, and also through reactions of aldoses with hydrogen cyanide (HCN), which can extend their length by a carbon through generation of cyanohydrins (121). Monosaccharides will also polymerize

chemically under the proper conditions through glycal and epoxide intermediates, which also gives them attractive structural features. These are just some among the multiple considerations regarding ribose and glucose and the intersection of their metabolism in early life forms, pertinent to the ascension of ribose (and deoxyribose) as central players in nucleic acids, and glucose as a central player in metabolism (122). Nevertheless, early in evolution glucose became the leading portable hexose building block and nutrient, and the most abundant building block on earth.

In terms of carbohydrate stereochemistry, the dominance of the stereoisomeric L-amino acids and D-sugars, the typical enantiomers of the building blocks of earth for proteins and glycomolecules, may have occurred early in chemical and biological evolution (123). The reason for chosen chiralities is unclear and debatable (124). L-amino acids and D-sugars may have dominated by random chance, or because their energy states differ slightly from their mirror images, or because differential reactivities, etc.; in any case, once established their recognition by other chiral molecules, e.g., proteins, would likely co-evolve to be specific to a restricted enantiomer. As for D-ribose, it occurs in macromolecules as β -D-ribofuranose (119, 123, 125). [It is interesting that both L-ribose and L-glucose can be easily chemically synthesized, but neither are found in organisms, and only D-ribose and D-glucose occur. D-Amino acids do not typically occur in messenger RNA (mRNA)-directed proteins, but they do occur in the peptide linkages within bacteria peptidoglycans (126)].

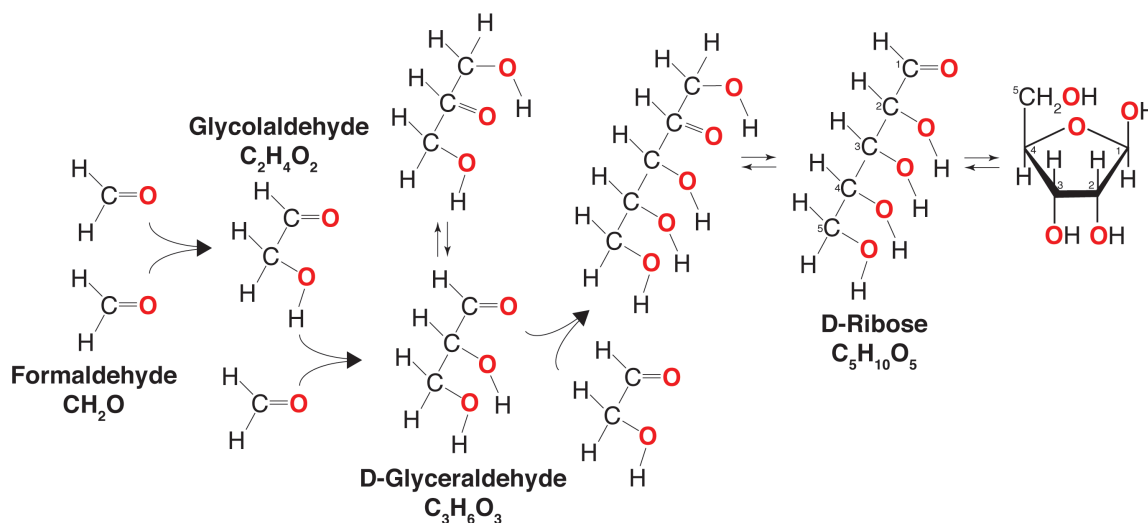


Figure 3. Examples of formose reactions leading to the generation of D-ribose from formaldehyde. The generation of glycolaldehyde, glyceraldehyde and their condensation can lead to the formation of the 5-carbon pentulose, a ketose, which can epimerize to the straight chain aldose D-ribose, which itself is in equilibrium with its ring form, as shown for β -D-ribose. See the text and references therein for a more complete discussion of the origin and interconversions of monosaccharides.

Furthermore, a monosaccharide such as D-glucose can readily be enzymatically isomerized with low energy costs into various epimers, such as the 4-epimer D-galactose, and the 2-epimer D-mannose, both of which are also incredibly common in nature. It is noteworthy that D-glucose, D-fructose, and D-mannose can slowly isomerize in certain solutions without enzymes (127). With such considerations it is easy to envision the generation of the vast array of monosaccharides found in nature (13).

The unique nature of monosaccharides drives structural diversity

Monosaccharides linked together in glycomolecules can generate linear or branched structures, there may be different stereoisomers, and they may occur in different anomeric linkages, either α or β , thus permitting a jigsaw puzzle of structural variety (Figure 1). Consider, that within starch and glycogen D-glucose occurs as an α -linked polysaccharide; it is both linear and highly branched, which promotes its compact folding. These features reduce the water content of such structures, which allows this high density nutrient to be compartmentalized within cells (128–130). However, too compact a structure creates a pathological condition associated with glycogen storage disorders, such as Lafora disease, caused by accumulation of dense, improperly branched, essentially insoluble glycogen aggregates termed Lafora bodies; these can occur in brain cells and lead to severe neurological disorders (131). By contrast, the isomeric structure β -linked D-glucose, is the basis of the linear, unbranched, giant polymer cellulose that occurs outside of cells; it is insoluble and has a profound structural role in cell wall architecture. Interestingly, the evolution of starch is thought to coincide with the endosymbionts ~1 billion years ago, in which an ancestor of cyanobacteria was internalized by a eukaryotic cell (132); both cyanobacteria and green algae can accumulate starch (133). The continued evolution of photosynthesis for the production of triose phosphates as precursors to D-glucose and its isomer D-fructose, which can be combined to produce sucrose by the ancient sucrose synthases (134), led to sucrose becoming the universal portable energy source in plants.

The ubiquity of sucrose and its unusual structure raises some interesting questions about its origin and dominance, as many other potential disaccharides could supply a similar amount of energy. Sucrose is a *non-reducing* disaccharide comprised of D-fructose and D-glucose, i.e., it lacks any aldehyde or ketone groups, and cannot generate glycation products through common Maillard reaction mechanisms, as it is not reactive with lysine residues or amino groups in nucleotides or

nucleic acids (135, 136). Although sucrose might accumulate in high concentrations in organisms, it is highly soluble but non-toxic in that regard, and thus, it could have been favored as a portable energy source over other competing disaccharides. A similar feature occurs in trehalose, in that case a disaccharide composed only of D-glucose, but also in a non-reducing form. Trehalose is produced in high amounts by many invertebrates, is a major source of energy for flying insects, and also has nutritional importance and preservative properties in cells (137, 138). By contrast, animals in general do not biosynthesize free disaccharides, except for mammals that generate high amounts of a reducing disaccharide lactose. It is comprised of D-galactose and D-glucose, and has an aldehyde group that is reactive with proteins by glycation (139). However, the restricted lactation-dependent and periodic accumulation of milk, as well as its restriction to the glandular products, may minimize the degree of potential pathology associated with its ability to glycate proteins.

Thus, while the glyco components of glycomolecules may vary in linkage, anomericity, monosaccharide units, branching, they may also be modified to become even more complex, as they may contain aglycone components, e.g., methylation, acetylation, phosphorylation, sulfation, phosphorylcholine, etc. Each glycan linkage and modification is catalyzed by discrete enzymes, but even a relatively few enzymes, acting combinatorially, but in an orchestrated fashion, can create an amazing variety of novel molecular structures. The types of structures shown in Figure 1 illustrate the terrific diversity of structures possible by linking monosaccharides to each other in such ways and further modifying the sugars. In that sense, if glycosylation is thought of as a PTM, the further modification of glycans themselves may be a post-PTM. Thus, glycomolecules present an imposing ensemble of structural diversity across all forms of life. Glycosylation of amino acids within proteins and the generation of oligo- and polysaccharides are ancient processes, and all organisms have multiple glycosylation pathways (46).

Photosynthetic systems for generating monosaccharides

The issue of sucrose raises another evolutionary consideration, which is key to the development of glycomolecules, and that is photosynthesis and the harvesting of energy to generate sugars (140, 141). This singular outcome, which is thought to have arisen in microorganisms, combines carbon dioxide and the 5-carbon sugar ribulose-1,5-bisphosphate, to generate two molecules of D-glyceraldehyde 3-phosphate (G3P), which is also a monosaccharide (142). These condense to generate

the 6-carbon D-glucose phosphate. Thus, the overall equation is simple; two relatively abundant and simple precursors, water and carbon dioxide, combine to produce glucose and oxygen. Through such processes the synthesis of trioses, riboses, and hexoses undoubtedly provided immense new possibilities for the evolution of sugars as photosynthetic organisms had a virtually unlimited supply of them. A dramatic variety of enzymatic modifications and condensations can generate additional monosaccharides, depending on the organism.

Photosynthesis evolved 3.2–3.5 billion years ago soon after life itself came into existence (143). It is utilized by cyanobacteria (blue-green algae), green algae, and plants as the most efficient way to trap solar energy and provide energy for the cell. In eukaryotes photosynthesis probably arose through endosymbiosis of a cyanobacterial-like organism (144). Amazingly, the smallest photosynthetic organism known, *Prochlorococcus*, is a marine cyanobacterium which lives in oceans and produces up to 20% of the oxygen in our biosphere, and an amazing amount of glucose and macromolecules for the marine food chain (145). An interesting Gedankenexperiment or *thought experiment* is to imagine a photosynthetic process, perhaps discarded in evolution, that might generate some molecule that is not a sugar, such as an amino acid, a lipid or a nucleotide; any of these could theoretically be energy sources and precursors for other molecules. However, the champions of evolution were the simple trioses, glucose building blocks, and specifically D-glucose. Its attributes include its incredible water solubility and multiple equatorial –OH groups; its ring form occurs as a glucopyranose, but it can also exist as a straight chain aldehyde. It can be easily degraded and further modified. Such kaleidoscopic features of D-glucose likely provided advantages over other small molecules, exploiting its manageable reactivity, and providing a stable building block and precursor for many other glycomolecules.

Glycan and glycomolecule functions

Given their incredible structural diversity, it stands to reason that glycans have a tremendous number of functions indispensable for life (50), and some have been discussed above. Some of the broad categories include the following. (i) Glycomolecules are structural components, essential for survival of a cell or organism. Examples of such functions are seen for RNA and DNA, and in polysaccharides, such as cellulose and chitin, and extracellular matrix (ECM) components, such as the polysaccharide HA and the glycoprotein collagen. Microbes that have peptidoglycans in their cell walls require them for osmotic integrity and environmental responses, and microbial

ecology. Glycolipids overall are essential in biological membranes for membrane integrity and membrane functions and biological topology. Similarly, GPI-anchored glycoproteins expressed in eukaryotes and likely in Archaea, but not bacteria, are critical for membrane organization and survival (146). (ii) Glycans in glycoproteins are commonly found to be essential to one or more features, such as protein structure, oligomerization, trafficking, turnover, resistance of proteins to proteases, free radical attack, etc., in what might be called *indirect* functions (147, 148). This is evidenced by phenotypes of cells and organisms in which protein glycosylation is defective; the constituent proteins are often unstable or poorly functional. Glycans at specific sites can modulate functions of proteins even if the glycans themselves are not directly recognized. Such is seen for chemokine interactions with their receptors, in which different structures of glycans modulate the specificity and affinity of binding (149). It is also seen in collagen, where the modifications of hydroxylysine are indirectly required for correct organization of collagen fibrils (150). O-Fucosylation of Ser/Thr residues in Notch receptors regulates interactions with their ligands, which in turn regulate expression of a tremendous variety of other genes and biological pathways (151). (iii) Glycans can also have *direct* functions, e.g., be directly recognized by proteins, glycan-binding proteins (GBPs) that include lectins, adhesins, antibodies, etc. In such cases, the glycans may be recognized rather independently. Consider the tens of thousands of GBPs and the carbohydrate-binding modules (CBMs) within them, expressed by bacteria and used to adhere to specific glycomolecules (152). Also, the formation of the lymphatic system is regulated by direct interactions of the O-glycans of the membrane glycoprotein podoplanin with the platelet GBP termed CLEC-2 (153). (iv) Glycomolecules may also be sources of energy, such as provided by starch, glycogen, lactose, sucrose, etc. (v) Finally, glycomolecules, such as the glycopeptide antibiotics and aminoglycoside antibiotics, are toxic and can allow an organism to control its environment (154, 155). There are many other categories of glycomolecule function, but these illustrate the diverse evolutionary variety of glycomolecules to provide unique structural features and biological activities.

Glycolipid biosynthesis, diversity and commonality in organisms

Glycolipids are important structural components of cellular membranes (156), but also important for sugar metabolism and glycomolecule biosynthesis. All organisms have within their membranes one or more of different types of glycolipids, which include those derived from fatty acids, glycerol, sphingolipids, phospholipids, polyisoprenols, and steroids, all

of which can be glycosylated, e.g., glucosylated cholesterol (157). In evolution, no specific glycolipid is conserved in all organisms, but each organism has essential glycolipids, and their classes are often unique to certain organisms. For example, bacteria have lipopolysaccharides (LPS), animals have glycosphingolipids, such gangliosides (158); plants and algae lack these, but have galactolipids such as mono- and digalactosyldiacylglycerol (MGDG, DGDG) and the sulfolipid sulfoquinovosyldiacylglycerol (SQDG), not found in other organisms (159). Amazingly, MGDG and DGDG are the major glycolipids in the thylakoid membranes of all photosynthetic organisms. All such glycolipids have important biological functions, too vast to consider in detail here. Glucosylceramide is one of the simplest glycolipids and is expressed by fungi, plants, and animals; it is essential for normal development in all such organisms, as shown in genetic studies in *Caenorhabditis elegans*, plants, and mice (160–162). Thus, an important theme of all cells is that they express glycolipids, but they may differ in many ways, undoubtedly with vastly different functions, depending on the organism (163). Bacteria can also synthesize an amazing variety of very complex glycolipids, but also simple glycolipids unlike those in any animal or plant (164). These include small glycans on fatty acids, fatty alcohols, sterols, hopanoids and carotenoids, that can vary in degrees of substitutions, chain lengths, etc.

The polyisoprenoids represent a unique class of lipids for glycomolecule generation. In animals they are represented by dolichol, a polyisoprenoid of C95–C125, i.e., carbons in length (163, 165–169); the bacterial counterparts are the C55 undecaprenoids, e.g., bactoprenol (169). These are present in all organisms and serve as lipid-linked precursors that may contain monosaccharides and oligosaccharides essential for biosynthesis and sugar addition to other molecules. The polyisoprenoids are derived by condensation of the universal 5-carbon precursors of all isoprenoids, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), as well as steroids such as cholesterol (170).

It is thought that a momentous step in evolution was the formation of membranes to encapsulate molecular activities associated with life (171, 172). But compartmentalization of metabolic activities also created the problem of topology (Figure 4). Many glycomolecules are present on the outer surface or membranes of cells, which has been termed the glycocalyx (173), but the sugar precursors to generate them arise mainly in the cytoplasm. Sugars are hydrophilic and averse to the hydrophobic membrane environment and do not diffuse through lipid membranes. The polyisoprenoids

function as reusable precursors that allow the protein-controlled *flipping* of glycans across membranes as lipid-linked sugars, from the cytoplasmic side of the cell to the opposite side of the membrane, e.g., the lumen of an organelle such as the endoplasmic reticulum (ER), or the outer membrane of microbes (Figure 4). In this ancient process, a glycolipid precursor of some type is made on the cytoplasmic side of a membrane and then flipped or translocated to the outside of the membrane using protein-assisted channeling pathways (174), as seen in microbes, or the lumen of an organelle, e.g., ER (175). For example, this flipping process is used by microbes for the synthesis of LPS. While membrane-associated glycolipids are typically asymmetrically presented on the *outside* of the membrane lipid bilayer, an interesting exception is an *inside*-facing diacylglycerol-based glycolipid with a short oligosaccharide structure termed the protein integrase or MPLase. For *E. coli* it is required for general integration of proteins into the inner membrane (176). In most biosynthetic processes, following glycan transfer from the lipid-sugar intermediate, the released lipid, such as a polyisoprenoid, flips back to the opposite side of the membrane for recharging to continue the cycle.

Such biosynthetic recycling, however, is not the case for glycosphingolipid biosynthesis using ceramide; the common precursor glucosyl-ceramide, is generated from UDP-Glc and ceramide on the cytoplasmic side of the Golgi apparatus and is then flipped into the Golgi lumen (177–179), and within the Golgi apparatus it can be extended to generate many other glycolipids (Figure 4). Thus, the lipid is not immediately recycled, but of course it may be *reutilized* after glycolipid degradation in the lysosome. Oddly, the topological issue of generating the isomeric galactosyl-ceramide in animals is completely different from that for generating glucosyl-ceramide. Galactosyl-ceramide is generated within the lumen of the ER by the enzyme UDP-Gal:ceramide galactosyltransferase which uses UDP-Gal imported there as a donor and no flipping of the product is needed (180). Thus, the co-evolution of such glycolipids along with monosaccharides allowed cells to both exploit the membrane-bound and compartmentalized nature of early metabolism, as well as to preserve and target forms of biosynthesis of specific biomolecules for presentation on the inside and outside of cells. Interestingly, enzymes for synthesizing glucosyl-ceramide and its simple extensions have been reported to be present in the mitochondria-associated endoplasmic-reticulum subcompartment (MEM), which may be a site for their production and movement into the mitochondria, as this organelle has been reported to contain small quantities of such glycolipids (181).

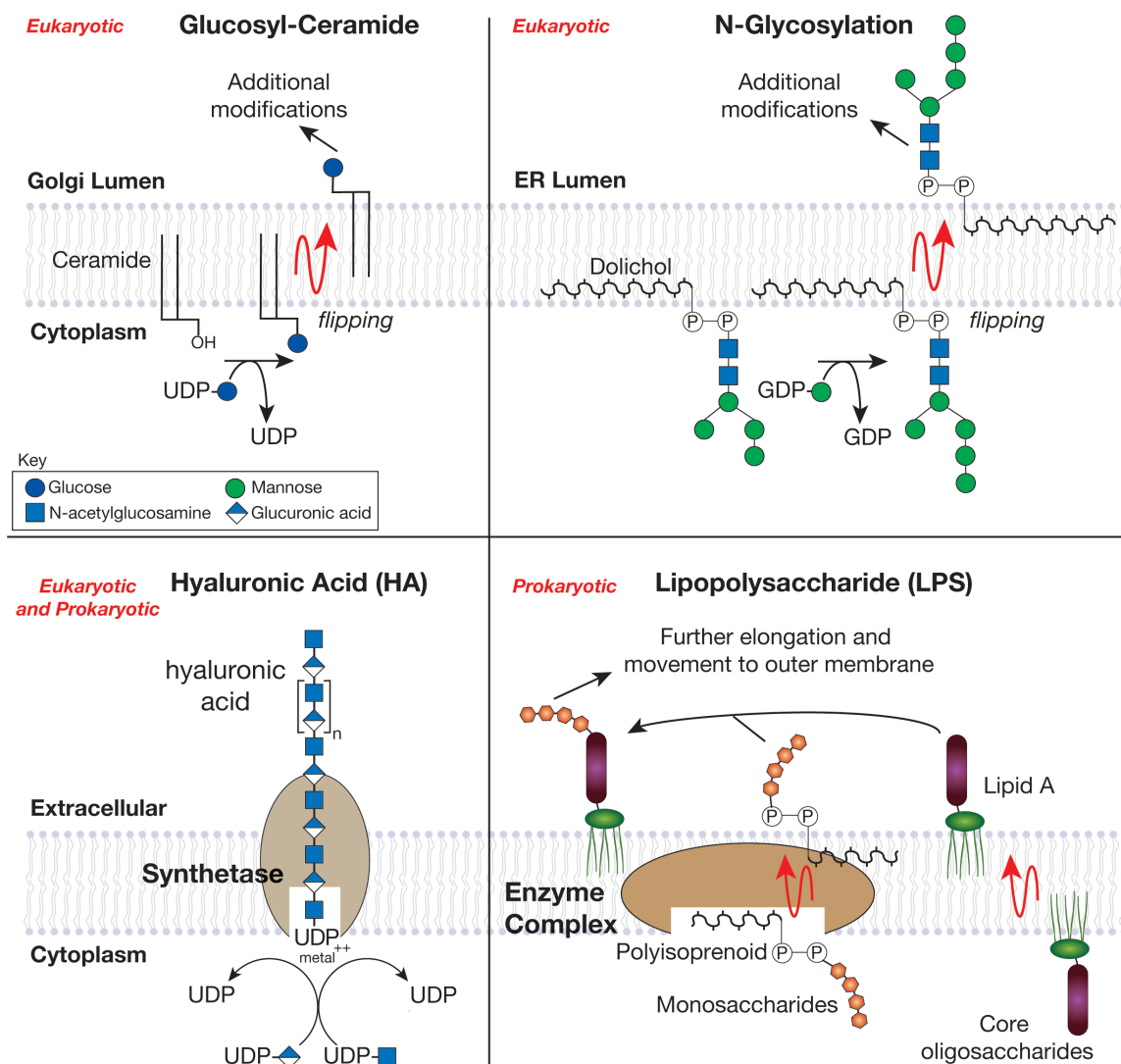


Figure 4. Cells are bounded by membranes and glycomolecules are typically on the outer membrane of cells or secreted from the cells; however, the precursors to such glycomolecules are formed in the cytoplasm. One approach found in all organisms to overcome this topological barrier is to generate unique types of lipid-linked sugars. Within the cytoplasmic side of a membrane, cells can generate a lipid-linked sugar, and it can then be flipped to the opposite side (external or luminal). In the two examples, taken from animal cells, but common to eukaryotes, glucosylceramide is synthesized on the cytoplasmic side of the Golgi apparatus using UDP-glucose as a donor, and the product is then flipped to the lumen of the Golgi by protein-assisted pathways. Similarly, for *N*-glycosylation pathways in animals, one example is shown here. The dolichol-pyrophosphoryl-GlcNAc₂Man₄ is synthesized on the cytoplasmic face of the ER, and finally modified to dolichol-pyrophosphoryl-GlcNAc₂Man₅ through the donor GDP-mannose. This particular product is then flipped to the ER lumen for further modifications there and eventual transfer of its preformed oligosaccharide to the newly synthesized protein, either co- or post-translationally. This basic concept of such topological resolutions also occurs in prokaryotes, as seen for LPS. Also shown is the direct synthesis at the plasma membrane of HA in both prokaryotes and eukaryotes using cytoplasmic precursors and extrusion of the product into the extracellular space. ER, endoplasmic reticulum; GlcNAc, *N*-acetylglucosamine; HA, hyaluronic acid; LPS, lipopolysaccharides.

Compartmentalization for intracellular glycomolecule synthesis and membrane-directed synthesis for special polysaccharides

Other pathways also evolved in eukaryotes to synthesize glycomolecules within compartments (organelles) and exploit

the topological barrier. Nucleotide sugar transporters allow the movement of nucleotide sugar precursors made in the cytoplasm into luminal compartments, e.g., ER and Golgi apparatus, through specific transporters. [An exception of this rule in eukaryotes is the nucleotide sugar CMP-sialic acid, which is generated in the cell nucleus (182).] Within those compartments glycomolecule biosynthesis can proceed in

a more regulated fashion and the nucleotide by-products are retro-transported to the cytoplasm for recycling (183). The newly synthesized glycomolecules are either soluble or membrane bound within the compartment, yet topologically outside of the cytoplasm, and may reach the outside through membrane fusion events during secretion.

A completely different pathway arose, however, for the three giant polysaccharides found in nature – cellulose, chitin, and HA. Interestingly, all of them have either Glc or the Glc derivatives GlcA or GlcNAc. For HA synthesis, as shown in Figure 4, hyaluronic acid synthase (HAS) in microbes and animals is a transmembrane enzyme that extrudes HA co-synthetically to the extracellular space; within its single polypeptide the enzyme has two catalytic activities, that can alternately use nucleotide precursors to generate the repeating HA disaccharide GlcNAc-GlcA (22, 184). Chitin and cellulose biosynthesis also proceeds through such transmembrane channel enzymes at the cellular membrane (184, 185). As for HA, the glycans have lengths that greatly exceed the dimensions of the cell synthesizing them, but the sizes are somewhat indeterminant. These processes contrast with those used for linear microbial LPS and capsular polysaccharide (CP) synthesis, where they are generated at the membrane coupled with synthesis and typically use lipid-linked sugar donors (186, 187); however, their exact lengths can also be indeterminant (Figure 4). It is interesting that polysaccharides are the only polymers in nature that can have thousands of repeats of a single building block, e.g., glucose in cellulose or starch. No such giant counterparts exist for proteins or nucleic acid comprised of a single repeating building block, although some polypeptides, relatively short by comparison to the giant polysaccharides, can be generated from a single amino acid, e.g., epsilon-poly-L-lysine (ϵ -PL) generated by non-ribosomal pathways (188).

The complex nature of glycosylating nascent polypeptides during their biosynthesis

The essential nature of protein modifications with sugars, as discussed above regarding microbial glycoproteins, also includes the common *N*-glycosylation of eukaryotic glycoproteins that occurs in the ER. There are many conceptual aspects of protein glycosylation that are shared between prokaryotes and eukaryotes. The linkage of oligomannose to Asn residues in the eukaryotic ER represents an en bloc addition of a large glycan of 14 sugar residues comprised of Glc₃Man₉GlcNAc₂ from the precursor Glc₃Man₉GlcNAc₂-P-P-dolichol, where the sugars are linked by a pyrophosphoryl linkage to the giant polyisoprenoid dolichol (8) (Figure 4). Such glycosylation often occurs at multiple Asn sites on emerging

nascent polypeptides, and thereby proteins can acquire glycans co-translationally (and also post-translationally) (189). During this process, a glycoprotein must fold upon itself, as well as interact with chaperones and other proteins in the densely packed ER. One can imagine the potentially intense intramolecular and intermolecular interplay between glycan and protein in this dense ER environment. This potential chaos in such an environment is controlled through evolution by the generation of host protein chaperones that bind proteins, and GBPs that bind glycans in the ER that act as molecular glycan chaperones, interacting with the newly transferred glycans independently of the protein, in a reversible manner, as other protein chaperones do by peptide recognition (190). Such binding, release, rebinding, and remodeling of the ER glycans through de- and re-glucosylation of the *N*-glycan, coupled with protein folding pathways, eventually allow a glycoprotein to correctly fold. There are elaborate processes in eukaryotes to deal with misfolded glycoproteins; these are forced to exit the ER and be degraded in the cytoplasm through multiple pathways involving glycan recognition and removal (190). Interestingly, such cytoplasmic *N*-glycosylated glycoproteins must be efficiently and rapidly deglycosylated; genetic defects in this process, which is universal in animals are associated with severe developmental disorders, such as NGLY1 deficiency (191). Thus, the glycosylation itself becomes not only an essential feature of the protein, but through its quality control attributes, it contributes to synthesizing functional glycoproteins (192). All eukaryotes have evolved these ER GBPs and quality control pathways to regulate their production and folding of glycoproteins in the ER. In the absence of Asn-glycosylation cells die, as they cannot efficiently produce mature glycoproteins, and the subsequent accumulation of misfolded proteins leads to ER stress, e.g., ERAD, and apoptosis. In certain protozoa, the same process occurs; however, the precursor lacks Glc residues, and Man₉GlcNAc₂ is added from the precursor Man₉GlcNAc₂-P-P-dolichol (193); importantly, in those cases glucosylation and re-glucosylation still occurs on the protein glycans to regulate glycoprotein folding (194). In other microbes, the precursor *N*-glycans can be even shorter, e.g., in *Plasmodium falciparum* the precursor is only GlcNAc₂-P-P-dolichol and lacks mannose entirely (195). Thus, this ancient pathway is fundamental to protein folding in the secretory pathway and is preserved in all its basic elements throughout eukaryotic life, but the sizes and functions of both the precursor and mature *N*-glycans in these eukaryotes have evolved dramatically.

Glycomolecules in viruses

Many viruses contain glycoproteins and other glycomolecules, such as capsid or envelope glycoproteins. Glycoproteins are

present in small capsid viruses, such as the adenoviruses, and in large enveloped viruses, such as HIV-1 (196, 197). For many viruses, their glycomolecules are synthesized by the infected host using virus-encoded proteins and host enzymes. The genomes of animal viruses lack encoded genes or machinery to dictate the glycosylation of viral molecules, but instead completely depend on the host cells they infect to add needed glycans to their viral glycoproteins (197, 198), in what may be classified as host-dependent viral glycosylation. Well studied examples of such viruses that infect animals include HIV-1, influenza viruses, and of course, coronaviruses, such as the notorious SARS-CoV-2. Envelope viruses, as they bud from cellular membranes, may also acquire glycolipids and other host glycomolecules present in membranes (199). The glycomolecules in viruses are often key to their interactions with host cells or interactions and evasion of the human immune system (197, 200, 201). It is interesting to speculate that animal viruses, which can infect animals that have immune systems, were selected to specifically not be able to glycosylate their own proteins, but use the glycosylation systems in the host cells. Such acquired host glycans, therefore, might be less likely to provoke immune responses and block virus replication, acting as a *shield* (202). Yet, the virus may not always gain the advantage in this system, as glycan presentation on the viral glycoprotein can be unique in conformation and organization and induce antibodies to the carbohydrates, as seen for broadly neutralizing antibodies (bnAbs) to HIV (203). These bnAbs bind to common oligomannose *N*-glycans, but bind uniquely because of their conformational presentation on the envelope protein surface. Interestingly, while not encoding genes to synthesize glycans, influenza virus binds to sialic acid-containing glycans and other host receptors through its virus-encoded hemagglutinin (viral HA), and then can destroy the sialic acid-containing receptor through its virus-encoded neuraminidase/sialidase (viral NA) (204, 205).

There are a few animal viruses whose genomes encode glycosyltransferase enzymes that can modify glycans. The *Bo17* gene of bovine herpesvirus 4, a bovine *Rhadinovirus*, encodes a β -1,6-*N*-acetylglucosaminyltransferase (β -1,6GnT) that is 81.1% identical on the amino acid level with the human homolog β -1,6GnT-M. It may have been acquired from its bovine host ~1.5 million years ago (206). The enzyme provided by the virus can be alternatively spliced, and participate in remodeling of the infected host cell glycome (207). The genome of the South American myxoma virus, which infects forest rabbits, encodes an α 2,3-sialyltransferase that is associated with enhanced virulence (208). The well-known baculovirus *Autographa californica* infects juvenile insects, and upon infection expresses a virus-encoded soluble glucosyltransferase that glucosylates the molting hormone,

leading to its inactivation, and a block of pupation (209, 210). Finally, in a related manner, some viruses, e.g., human T-cell leukaemia virus (HTLV-1), upon infection induce expression of the host fucosyltransferase VII, leading to elevation of the surface expression of the carbohydrate antigen sialyl-Lewis x (sLex) on host cell glycomolecules, a potential ligand for various selectins that interact with leukocytes (211).

By contrast, the genomes of many viruses that infect organisms other than animals encode enzymes for producing their own varieties of glycomolecules. The genomes of bacteriophage, which are among the smallest viruses, encode a variety of glycosyltransferase and glyco-modifying genes active on their own molecules, including their DNA (212). But other viruses also encode glycogenes, e.g., glycosyltransferases, capable of modifying their own proteins and host proteins. The genomic sequencing of giant viruses revealed extensive numbers of such glycogenes (212–215). Giant viruses are ubiquitous in all fresh water, and infect freshwater eukaryotic algae; they are known as nucleocytoplasmic large DNA viruses (NCLDVs) (216). The chlorovirus *Paramecium bursaria* chlorella virus (PBCV-1) synthesizes unusual Asn-linked glycans unlike any of those in animals. The glycans are linked to Asn through a β -glucose linkage, and are highly branched, even containing a hyperglycosylated fucose modified at all sites, and all the glycoforms contain a dimethylated rhamnose as the capping residue of the main chain (217) (Figure 1). The genome of the related *Marseillevirus marseillevirus*, a type of giant mimivirus, encodes several glycosyltransferases (214, 218), that both hydroxylate lysine residues in proteins and then add a collagen-like linkage of Glc (animals use Gal) to the hydroxylysine. The enzymes can modify their own collagen-like proteins and even animal collagens (219). These viruses are also rare non-animal organisms that generate collagen-like molecules. Amazingly, animals exposed to such mimivirus collagens produce anti-collagen antibodies that may promote arthritis (220).

Microbial polysaccharides

Many studies have documented the tremendous differences in glycomolecules between different unicellular organisms, and certainly between bacteria and archaea (9, 98, 221–223). Bacteria lack a nucleus and are bounded by a single membrane (and outer membrane in Gram negative) in which is embedded glycolipids, e.g., LPS in Gram negative and lipoteichoic acids (LTAs) in Gram positive. As noted above, glycoproteins are also produced by all types of bacteria, Gram negative and positive. In regard to polysaccharides, bacteria are amazing in their ability to generate diversity from a common theme. LPS, which is essential for bacterial

viability, is densely expressed on the surface and is a linear polysaccharide; depending on the organism, it may have repeating units of di- to nonasaccharides (two to nine), and there may be several million copies of LPS expressed on the surface of a single bacterium (224, 225). There may also be non-stoichiometric modifications of the LPS, which are partly regulated by small non-coding sRNAs (226). But the genetics and the genes encoding the enzymes responsible for such linear LPS molecules for example, are remarkably few, and each independently adds a sugar from a precursor intermediate. Thus, each microbial strain may synthesize a dominant but unique LPS or a CP of varying lengths, leading to their recognition as serotypes, based on antibodies that can distinguish the antigenic glycan components (227). The strain differences, as for *E. coli*, are usually based on variations in the LPS structure or O-antigen, which arise from variations in the linear genes in the operon responsible for LPS production (228). Over 700 different serotypes of *E. coli* are known and ~470 of these produce Shiga toxin (229). For Salmonella, there are >2400 serotypes (230). Although common bacteria, even *E. coli*, can also express other glycomolecules and glycoproteins, their glycome is typically dominated by their unique LPS. By contrast, LPS is not synthesized by cyanobacteria, and they lack the core sugars of Gram negative LPS such as 3-deoxy-D-manno-octulosonic acid (KDO) and heptoses found in typical LPS (231). Their exopolysaccharides can be assembled stepwise by enzymes in the cytoplasm and then translocated to the outside through the membrane.

Archaea lack both LPS and LTA, common components of Gram negative and Gram positive bacteria, respectively (223, 232). Archaeal glycans are much more complicated in any individual organism compared to bacteria; archaea also lack conventional murein or peptidoglycan, but instead make a tremendous variety of glycomolecules related to peptidoglycan in their cell wall. Archaea can also synthesize glycoproteins, and a major feature of the archaeal cell surface is the S-layer (233). They also generate glycoproteins containing *N*-glycans linked to asparagine, a modification found in protists, along with unusual heteropolysaccharides, methanochondroitin, and many other types of polysaccharides (223). Overall, as might be expected, therefore, genomes of archaea are much more complex and enlarged in terms of *glycogenes* responsible for elaborating their glycomolecules, compared to eubacteria.

Plant glycomolecules

There are many pathways of glycomolecule biosynthesis that plants share with all eukaryotes, but there are some that are specific to plants and even a plant species (234, 235).

For example, the initial process of modifying Asn residues in glycoproteins in plants is essentially similar to the processes in all eukaryotes, i.e., it utilizes a dolichol-linked glycan precursor to co-translationally add sugars to proteins entering the ER (236). Plants also have GPI-anchored glycoproteins using a lipid-linked precursor similar to yeast and animals (237). However, the types of mature *N*-glycans resulting in plants vs mammals, while similar in some ways, are very different in others (235, 238). Plant *N*-glycans do not contain sialic acid, which is a common terminal modification of *N*-glycans in mammals. Plant *N*-glycans often contain xylose β 1,2-linked to the core mannose, and fucose α 1,3-linked to the core chitobiosyl unit, but these are not found in mammalian *N*-glycans. In animals O-glycosylation pathways involve addition of *N*-acetylgalactosamine (GalNAc) to Ser/Thr residues, but this does not happen in plants; instead, they modify Ser, Thr, and hydroxyproline residues in their glycoproteins with other sugars, e.g., arabinose and galactose (239). In addition, keratan sulfate, which can occur in *N*-glycans of animals, and the Ser-linked GAGs within proteoglycans, such as HS on Syndecan-1, do not exist in plants but are major contributors to many biological pathways in animals (240).

Plants generate enormously complex polysaccharides with very unique structures not found in animals. A great example of a very unique set of polysaccharides synthesized by many plants is pectin, found in cell walls (241, 242). Pectin includes galacturonic acid-rich polysaccharides such as homogalacturonan (HG), rhamnogalacturonan I (RG-I), and the substituted galacturonans rhamnogalacturonan II (RG-II) and xylogalacturonan (XGA) (243). Its biosynthesis requires nearly 70 different transferases, such as glycosyltransferases and methyltransferases, which function in the Golgi to synthesize pectin that is then exported outside of cells (242). Pectin is comprised of 17 different monosaccharides in an amazing variety of linkages, at least 20 types (244). Plants also synthesize a dazzling variety of polysaccharides in general, such as xylans, mannans, galactans, arabinoxylans, and many others (Figure 1) (11). In addition, collectively plants make a nearly incomprehensible variety of GBPs (lectins) (245), which are commonly found in their seeds, roots, and bark, which are important in seed survival, plant protection, nitrogen assimilation, and many other pathways (246). They are considered to be a part of the innate immune system of plants, aiding in resistance to viruses, bacteria, fungi and animals. Amazingly, each species of plant appears to generate unique varieties of lectins. Because of their abundance and ability to bind many different monosaccharides and glycans, such plant lectins have become primary reagents to study glycosylation pathways in other organisms, as the lectins seem to recognize specific glycosylation patterns and each

lectin is unique in some way. Plants represent the principle of stringent adaption of glycan form and structure to function for each organism competing to survive.

Distinguishing feature of glycomolecules in evolution are their structural complexities

Thus, while some might consider a vertebrate polysaccharide such as HS to be incredibly complex and an example of the pinnacle of complexity, pectin easily rivals it. Furthermore, a unicellular organism may generate a relatively small repertoire of glycans, e.g., *E. coli*, or it can generate an amazing variety of small and gigantic glycans, e.g., *Mycobacterium tuberculosis* (Mtb) (247). The glycomolecules in Mtb, which uses humans as its obligate host, are incredibly complex and varied in expression and the types generated, most of which are different from anything found in an animal. Its glycome is so complex that it could perhaps rival that in any single eukaryotic cell, even a single human cell type.

The current evidence indicates that the most important aspect of the glycome of an organism is the uniqueness of its collective glycan structures. Glycomolecules in all types of organisms may overlap in size and complexity per se, i.e. made up of similar parts, but their specific structures are generally unique. Thus, if we consider existing organisms as representing the outcome of evolution, we can see that each lineage of organisms has conserved the synthesis of certain classes of glycomolecules, e.g., polysaccharides, glycolipids, glycoproteins, etc., although their structures, scale and diversity, deviate considerably and have expanded along with the evolution of increasing genetic and organismal complexity. In addition, organisms have evolved many different classes of glycomolecules, uniquely functional for their biological niches, yet closely related organisms share general pathways for glycomolecule production.

These facts also suggest that there are key forces driving evolution for each organism leading to its generation of glycomolecules uniquely suited for its development and survival. Thus, it is not surprising that genetic mutations in biological pathways that affect glycomolecule elaboration in an organism lead to disruptions in its growth and survival. In microbes and plants this is well documented, for example, for many different glycomolecules, such as glucosyl-ceramide discussed above, for cellulose (248), and also in studies of humans with Congenital Disorders of Glycosylations (CDGs), which, depending on the gene or pathway affected, results in unique and diverse pathologies (249, 250). Horvitz's pioneering work on GAGs in nematodes also showed that single gene mutations that alter pathways of glycosylation

are associated with developmental defects and infertility (251). In addition, in animals and humans any disruption in the degradation of glycans, such as seen for lysosomal storage disorders (LSDs), invariably results in pathology (183). All such genetic evidence from unicellular organisms to metazoans unequivocally validate the vast and important roles of glycomolecules, their biosynthesis and turnover, in development and survival.

Host vs invader and relationships to glycomolecule evolution

There may be multiple evolutionary pressures on organisms to evolve their glycomes. Glycomolecule diversity likely arises from biological needs, including advantages of enhanced efficiency of biological functions and generation of novel functions. A readily identifiable driver of glycan diversity regards the interactions of a host with a potential invader, e.g., a bacteriophage, virus, bacterium, parasite, or some other pathogen or even commensal organisms, e.g., the microbiome (Figure 5). A primitive example of this is the interactions of bacteria with their invaders known as bacteriophages (252, 253). The phage express GBPs which bind to glycans that are expressed on the bacterial surface. Interestingly, there is evidence that phage GBPs can also bind other glycans, perhaps even those expressed by animal hosts of microbes, as in the gastrointestinal tract mucins (254).

An interesting historical observation of the interplay of host vs invader was made decades ago by Heidelberger and Avery, who reported in 1923 that the capsule of *Streptococcus pneumoniae* contains a surface polysaccharide (255). Avery and Goebel later reported in 1933 that infected patients made antibodies to the CP (256). Follow-up studies to those of Griffiths (257), indicated that *smooth* vs *rough* appearing colonies of bacteria, differed in production of polysaccharide; the former produced surface polysaccharides giving the bacteria a smooth, shiny appearance (S), whereas the rough lack this polysaccharide (R); the R strains are also non-virulent. In other studies Hershey and Chase confirmed that DNA is the genetic material, by demonstrating that bacteriophage T2 injected their DNA into bacteria, and that the new phage emerged from the injected DNA (258). Thus, the *transforming principle* was proven to be DNA through studies of polysaccharide expression. Interestingly, this concept was generally rejected for a considerable amount of time by the scientific community, as DNA was considered to be a boring, uninteresting molecule (259). Over time all objections were overcome, and this famous discovery by Avery, MacLeod, and McCarty based partly on the nature of glycomolecule synthesis is a cornerstone of modern genetics (260).

These historical studies are important to consider in the light of evolution; bacteriophage is one of the most abundant organisms on earth, yet their existence depends on first binding to a bacterium and then replicating within it. This typically involves binding to bacterial surface molecules including cell surface polysaccharides and porins, e.g., OmpC (261–263). One of the most studied phages is the tailed bacteriophage T4, which has the classic appearance of a spaceship with its landing legs; these legs are long tail fibers that attach to the LPS of bacteria (264, 265). The ends of these fibers have a GBP essential for this interaction with the bacterial surface. Of course, bacteria can evolve over time to develop a different polysaccharide to which phage cannot bind, and of course, the phage may evolve new binding activities for the different polysaccharide, apparently *ad infinitum* (253). This is thought to be an evolutionary driving force that expanded the diversity and biological niches occupied by bacteria (253, 266).

Thus, an interesting aspect of host vs invader involves the invader expressing a GBP or lectin that may interact with host glycans in a destructive process (see glycan 1 in Figure 5). If this is pathological, there may be selective pressure for survival of cells that express a different glycan (glycan 2) that is not bound well by the original invader. The original invader may evolve to generate a GBP that can bind to the novel glycan 2, or another invader can bind to glycan 2. Such modifications may continue, leading to novel glycans and GBPs, in the process termed the *evolutionary glycan arms race* (267). This has also been referred to as the *Red Queen effect*, *ala Alice in Wonderland*, where evolution is a never ending race between hosts evolving in response to rapidly evolving invaders in their interplay, both for good and bad (43). An interesting example of the evolving nature of such interactions is also seen for the tectonins, which are lectins conserved throughout evolution. They are found in bacteria, e.g., actinomycetes, fungi, invertebrates, fish, and humans. Tectonins are proteins with protective and toxic activities, such as fungal tectonin-2, which binds to methylated monosaccharides within glycans expressed in invading microbes and nematodes (268). This concept of host vs invader is inherent in the nature of life, as all organisms compete for niches, and glycomolecules participate in these continuing challenges.

Such host vs invader interactions involving glycomolecules also likely have larger outcomes, as the evolving glycans may not only help hosts evade (or attract) invaders, but the glycans may acquire additional and important biological features independent of the invaders, and in fact, become essential components of the molecules to which they are attached. This is an interesting concept to consider in light of the divergence of glycan modifications of proteins, and the essential roles that glycans have acquired in protein functions and protein

folding. There is also the interesting rapid adaptive evolution in mammals and their immune receptors that recognize glycans of diverse microbial pathogens, including viruses and fungi (269). Thus, both glycan structures and proteins that recognize glycans may rapidly evolve and adapt in the never-ending host-invader interactions.

There is also the interesting possibility of an invader presenting glycomolecules that themselves can be pathological to the host or recognized by the host immune system. For example, human and animal parasites, such as protozoans and helminths, have unique glycomes, characterized by features that may benefit the parasite, but compromise the host. Some glycans might be recognized by the host immune system; yet overall, the parasite may survive and the host lives long enough to promote parasite survival. Some human pathogenic bacteria make CPs with sialic acid, but humans express sialic acid-binding proteins or Siglecs (270). Because the Siglecs may function to send inhibitory signals upon binding ligands, bacteria may use sialic acid as a type of glycan mimicry to evade immune attack. However, many parasites, such as the helminths or worms, do not express sialic acid and instead synthesize unusual glycans that are immunogenic (271–273), but also capable of interacting with different glycan receptors on immune cells to induce tolerance and evade immunity, a type of glycan gimmickry (274). If the titer and specificity of such anti-carbohydrate antibodies are sufficient these can result in killing of the invader, e.g., bacteria or parasite. This is the basis of anti-carbohydrate-based vaccines, such as the Hib vaccine given to children, which induces antibodies that target the surface polysaccharide of the deadly *Haemophilus influenza* type b (275). This issue of glycan expression in all organisms and their interplay with other organisms, either as commensal, parasitic and/or pathogenic, is extensive and requires a much larger discussion than can be afforded here. A currently compelling area of great interest in this regard is animal and human microbiomes and the roles of host and microbial glycans, and dietary glycans, in modulating and regulating normal homeostasis, as well as their roles in dysbiosis (276–279).

Glycomolecule evolution associated with enhanced functions

Another selective force for glycan diversity may be the enhancement and diversification of molecular interactions important in biological processes. For example, consider a common feature of protein-bound polysaccharides in vertebrates and invertebrates, which is the presence of a short 'core' structure to which is linked a longer polysaccharide, often of a repeating nature with disaccharide repeats.

This is seen in GAGs linked to soluble and membrane-bound proteoglycans (280), in the membrane bound glycoprotein α -dystroglycan (281), and in plant cell wall glycoproteins, such as the extensins and arabinogalactan proteins (AGPs) (11, 282, 283). In proteoglycans a common core containing Xyl-Ser is modified by multiple polysaccharides, e.g., CS or HS. In AGPs, which are the most highly glycosylated proteins in plants, hydroxyPro, Ser and Thr residues are modified with a galactan core and then high branched extensions are formed with galactose and arabinose, and further modified by other monosaccharides. Because the glycan component of AGP is >90% of the protein, the protein itself is considered to simply be a carrier for the glycans with no other specific function. However, for the glycoprotein α -dystroglycan it is glycosylated at multiple sites and contains a unique polysaccharide termed matriglycan (281, 284, 285). The extended glycans for both proteoglycans, plant cell wall glycoproteins, and α -dystroglycan become platforms to which multiple GBPs may recognize specific features within the glycans, thus allowing a single glycoprotein to partner with a wide range of other proteins in regulating biological activities.

While all such glycoproteins cannot be discussed here, α -dystroglycan provides an interesting example of glycan diversity associated with diverse functions. α -Dystroglycan has multiple functions in all animal tissues, and is a key component of the surface of muscle cells and many other cells for their interactions with the ECM, linking the cytoskeleton and actin via dystroglycan to the membrane and ECM. The glycan of α -dystroglycan is O-linked via mannose, and includes the core of Man, which may be phosphorylated, GlcNAc, GalNAc, and ribitol-phosphate, to which is linked a repeating polysaccharide termed matriglycan comprised of Xyl-GlcA repeats, which may vary in length due to the activity of the bifunctional glycosyltransferase LARGE1 (284). The generation of the core plus matriglycan requires at least 18 genes in vertebrates, and defects in any of these lead to a set of disorders termed α -dystroglycanopathies (286). Interestingly, specific genetic defects, recognized as congenital muscular dystrophies (CMDs), lead to different outcomes and clinical phenotypes, such as Walker-Warburg Syndrome, Muscle-eye-brain disease (MEB), Fukuyama CMD, and many others.

A main function of the glycans on α -dystroglycan is to provide recognition sites for binding of ECM proteins such as laminin and others that contain laminin G-like domains (LG domains); these include agrin, perlecan, neurexin1- α , and many others (287). Their interactions and the complexes they create are essential for formation and stability of skeletal muscle and other cellular interactions with the ECM. The vast number of proteins that bind to glycans of α -dystroglycan, and the

different features of the disorders arising from defects in glycosylation of α -dystroglycan, suggest possible evolutionary drivers for their generation. The long and varied glycans in metazoan cells provide a wide range of interactions of different affinities to regulate the complex interplay of cells with the ECM, which may differ throughout all animal tissues. Extended matriglycan may also permit multiple interactions simultaneously with different LG proteins. Interestingly, while α -dystroglycan and its core glycosylation is conserved in invertebrates, such as *Drosophila*, the extensive matriglycan elongated from the O-Man core is lacking in *Drosophila* (285). Thus, the modifications and extended lengths of such glycans may have arisen through specific interactions that increased as the complexity of the glycans increased. In that sense, glycan complexity for all the glycans discussed thus far may provide opportunities for participation in and regulation of many biological pathways.

Interestingly, the evolution of GAGs and matriglycan may also involve host-invader mechanisms. GAGs are recognized by a wide number of different viruses, bacteria, fungi and parasites (288). For matriglycan, it was recently shown that it is a receptor for Lassa fever virus (LFV) (289) and for the organism *M. leprae*, which causes leprosy (290). The combination of glycan complexity driving biological pathways and the opportunistic interactions that arise with invaders is consistent with the roles of glycomolecules in the natural competition in nature for survival.

Roles of glycomolecules in the success of metazoans

Metazoans are defined as organisms with a body comprised of differentiated cells or specialized cells. It is noteworthy that the evolution of single cells or unicellular organisms to metazoans included the *conceptual* conservation of certain pathways for glycomolecule synthesis, along with the development of new ones. A conceptual relationship for example, is in regard to generating extracellular polysaccharides of functional importance. Cyanobacteria (blue-green algae) bind together to form floating mats or aggregates through interactions with the extracellular sulfated polysaccharides they produce. Such mats can trap oxygen which the organisms produce and thereby increases their buoyancy, e.g., cyanobacterial blooms (291). Thus, the evolutionary model of using glycomolecules to enhance adhesive interactions is also a common theme in all types of bacterial biofilms (292), which have multiple functions, and protect microorganisms from challenges and promotes adhesion and colonization. Overall, all organisms make some type of extracellular polysaccharide with different biological functions.

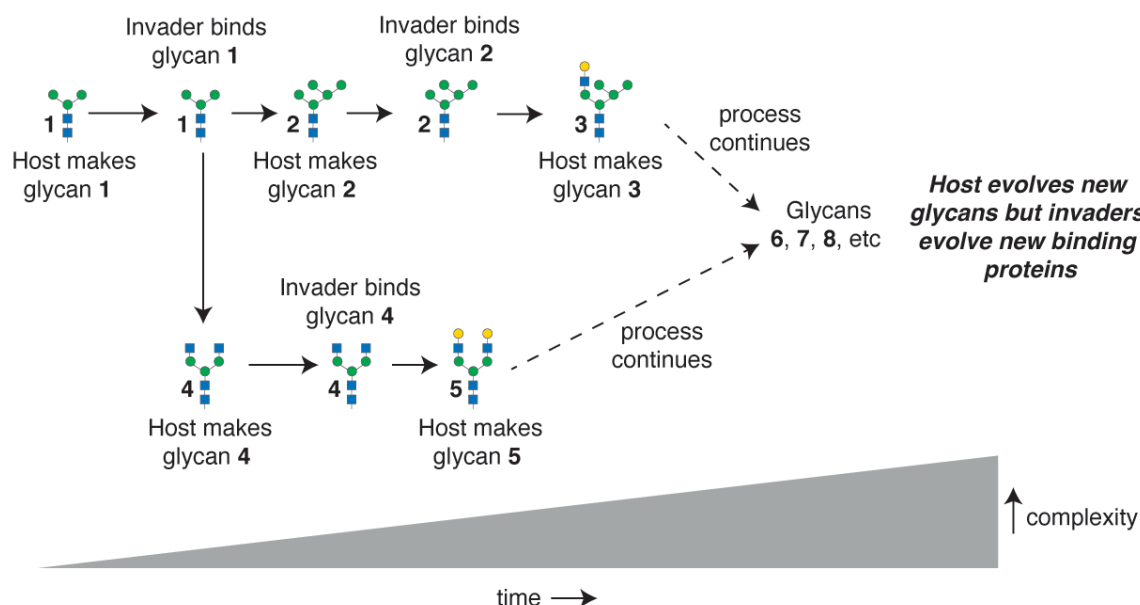


Figure 5. Evolving nature of glycan structures by the host which may be recognized by an invading organism. In this host vs invader model, the host may survive by modifying the glycans, e.g., glycan 1 to glycan 2, to which the invader may not bind well. The invader or other invaders may evolve the ability to bind the new glycan, leading the host to again respond with a modified glycan, e.g., glycan 3 or 4. This cycle may be endless, and leads to great diversity and complexity of both glycans through time on the host and multiple types of binders produced by the invaders.

This theme of using extracellular polysaccharides to enhance adhesivity is also present in metazoans cells. The sponge is the simplest multicellular organism in which glycomolecules were long ago shown to be key to their survival. There are currently ~8000 species of living sponges; they are ancient, dating back to the Precambrian age, and perhaps a similar number of species can be found in fossils (293, 294). They are in the phylum Porifera, which derives from the Latin *porus*, highlighting the *pores* throughout their bodies. Interestingly, while sponges lack organized tissues, they do have specialized cells, e.g., choanocytes, porocytes, amoebocytes, and pinacocytes. Amazingly, the visible body of a sponge is a molecular skeleton, a collection of cells encased within mineralized polysaccharides and other glycomolecules. These include spongin, which is a collagen-like glycoprotein that supports the spicules. Spongin has many similarities to human collagen, and is a member of the superfamily of collagens found in the ECM of all animals (295), but not found in prokaryotes. The commercial, flexible bath sponge is produced by processing the animal to generate an acellular, spongin-rich fibrous skeleton (296).

Spongin and its collagen cousins are the most abundant and likely most ancient glycoproteins in all metazoans. Collagens are essential for normal survival of organisms, as seen in mutations of collagen genes in humans (297). There is also conservation in evolution of the main glyco-component of spongin and all collagens, which is glucosyl-galactosyl-

hydroxylysine, a pathway and glyco-modification that may also be considered among the most ancient in metazoans (298). As noted above, even giant viruses make a form of collagen, but for them the linkage sugar to hydroxylysine is glucose. Amazingly, some bacteria also synthesize collagen-like glycoproteins capable of forming triple-helical structures; their *O*-glycosylation of the collagen-like region of BclA in *B. anthracis* is unlike that in animals and consists of short glycans containing rhamnose (299). Animal collagens also have conserved *N*-glycosylation features, with important proteostatic functions (300). Certain bacteria that inhabit sponges, such as the *Poribacteria*, have genomes encoding a wide variety of glycohydrolases and sulfatases to aid in their scavenging of the sponge ECM and carbohydrates for their unique survival there, but which may also be participated in host-invader processes (301).

Regarding the molecules holding sponge cells together, the classic experiments of Wilson in 1907 are noteworthy. Wilson demonstrated that two species of sponges, when their cells are dissociated and mixed, could reassociate over time, i.e., reaggregate, to regenerate the distinct sponges (302). This whole-body regeneration (WBR) in sponges is one of their amazing attributes, as some animals can only regenerate parts of their body (303). The factors important in specific sponge cell aggregation are termed aggregation factors (AFs), which are types of proteoglycans with sulfated polysaccharides attached, and which mediate the calcium-dependent

polyvalent interactions between glycans, i.e., homophylic interactions, and interactions with cell surface receptors (304–307). Sponges synthesize sulfated polysaccharides, but they do not synthesize GAGs of the types found in other animals (24). In regard to AFs, their genetic evolution and diversification has been well studied, and arises through high allelism, combined with domain shuffling (308). This leads to diversification that promotes allorecognition (self-non-self) in sponges, as originally discovered by Wilson. This is also an example of carbohydrate-carbohydrate interactions, which has been proposed to be a driving force of glycomolecular functions in all organisms (309). This primitive type of allorecognition involving glycomolecules and their interactions is thus carried throughout metazoan evolution.

Human glycomolecules and the human glycome

After the above considerations, it seems fitting to conclude with a discussion of human glycomolecules, as the themes and questions that arise in other organisms are applicable to human glycomolecules. A vast literature has documented the structures and functions of glycomolecules in mammals and human cells in terms of their involvement in basement membranes, connective tissues, cell adhesion, cell signaling, nutrition, and many other pathways (50, 147, 310–316). A preceding section on *Glycan and Glycomolecule Functions* discussed specific roles of glycans in a molecular context. The human glycome and the genome encoding the information for their production, overlap in many ways with those in other mammals (50, 317). Types of such overlapping glycomolecules include many different types of membrane-associated and soluble glycoproteins, e.g., mucins, proteoglycans/ GAGs, GPI-anchored glycoproteins, intracellular glycoproteins (O-GlcNAc), various types of glycolipids, e.g., ceramide-based, polyisoprenoid-based, and phosphatidyl-based, and an amazing variety of oligosaccharides and polysaccharides (50). Virtually all membrane proteins and most secreted proteins in mammals and human cells are glycosylated, often on multiple amino acids (318). The GAGs especially illustrate the non-coding nature of human glycomolecules. HS, which is essential for normal survival of an organism (319), is a linear polysaccharide made by most animal cells. It has a tremendous number of glycan modifications, e.g., acetylation, sulfation, and monosaccharide isomerism, which regulate its interactions with a variety of proteins, and yet these structural features are not encoded within a template and the absolute sequences cannot be predicted.

Much is being learned about the *functions* of glycans in humans, which is accelerating due to identifications of genetic mutations that alter glycomolecule production and turnover,

and engineered mutations in animal models (183, 250). However, remarkably little is known about the *structures* of glycans within the entire human glycome, as there are no systematic studies in which all glycans in all glycomolecules throughout development and in each cell type in an adult are defined. It is common to think that the glycome of the human is the most complex of all, but since all glycomes of all animals are not known, this is perhaps an overstatement. There have been attempts to estimate just the number of glycan determinants in the human glycome, defined as the linkages of 2–6 monosaccharides recognizable by a GBP or having a specific function (320); such estimates suggest >7000 determinants, but when linked multiply on branched and large glycans, the number of glycans themselves likely reaches the millions. It may be fairer to state that the human glycome, like the glycomes of other vertebrates, is extraordinarily complex, and each differs in specific ways as the animals adapted to different environmental and developmental stresses. Some core pathways are similar in humans to those in many other eukaryotes, but some are unique to mammals.

There are some modifications of mammalian glycans in other primates that are not found in human biosynthetic pathways, e.g., α -linked galactose (321) and *N*-glycolylneuraminic acid (Neu5Gc) (322). The general nature and biosynthetic pathways overall are relatively conserved in all mammals, and also highly related to those in other vertebrates. The major differences in glycomes of humans, compared to other organisms, is in the *structure* of the glycans making up their glycomes. Also of interest in light of the many types of monosaccharides found in nature, human glycans are comprised of a relatively unique set of monosaccharide building blocks (13), similar to those in many other higher animals, but different when compared to those in monera, protists, fungi, and plants. Few if any pathways of glycosylation are found only in *Homo sapiens* and missing in other mammals. But there are some clearly unique human glycans not found in all other mammals. Human milk oligosaccharides are comprised of hundreds of free glycans; however, significant structural differences are seen in human milk oligosaccharides, compared for example, to those in milk from *Canoidea*, e.g., raccoon, mink, and dog, etc. (323). Several terminal glycan modifications on glycoproteins and glycolipids are rather uniquely human or at least unique to certain primate lineages of humans. These include the ABO blood group system (324), and the absence of α -linked galactose and Neu5Gc in humans and Old-World primates.

While the genome and transcriptome of an individual cell may be deciphered in part through the analysis of the DNA and RNA, because of the ability to amplify these molecules enzymatically, the complete glycome of a single cell is not knowable at present. An adult human may contain ~37

trillion cells, not including the microbiota (325–327), and there may be 200–500 defined major cell types. Thus, while the genome is shared by all these cells, the glycomes are likely to be decidedly different. Most analyses of the human glycome have focused on select biosynthetic pathways from a genetic analysis (315). A vast number of studies on glycan structures have focused on glycomolecules that are soluble. When membrane glycoproteins have been studied, such as for the large family of integrins, cadherins, and protocadherins, the number of different *N*- and *O*-glycans and their diversity is astonishing (328, 329). Soluble glycomolecules include serum glycoproteins, where the is often a focus on *N*-glycans (330), and the free glycans in human milk (323, 331, 332). Multiple studies have also been done on cultured human cells; examples include the *N*-glycans in glycoproteins of HeLa cells, which represent cervical carcinoma cells (333), human embryonic kidney (HEK) cells (334), and human colorectal cell lines (335). There have also been studies on *N*- and *O*-glycomes in a few human organs or their components, e.g., brain (336–340), liver (341), and intestine (342).

While limited and not really systematically performed, these studies all indicate that the glycomes of human cells and organs are varied and reveal unique differences and some similarities, both quantitatively and qualitatively. Moreover, the glycoproteomes of cells and organs, i.e. the glycoproteins carrying the glycans, are also significantly different in all cell lines and organs examined. While important strides in this area are developing (343, 344), most studies are generally incomplete, due to the limitations of the analytical tools available, and the typical focus of most studies on the most abundant, identifiable glycans. There is also a limited diversity of standards available for glycans, and certainly those with unique modifications, e.g., sulfation or phosphorylation, which complicates the rigor of analyses. Studies also seldom compare all the glycans in all classes in one individual or human donor tissue to another. Furthermore, glycomics is often focused on glycans obtainable by enzyme release, e.g., PNGase F, but enzymes to release all types of glycans, e.g., *O*-linked GalNAc, *O*-linked Man, *O*-linked Fuc, and unusual modified amino acids mentioned above, are not available. It has been estimated that the human glycome may comprise many thousands of terminal sequences recognizable as determinants by antibodies and lectins (320). The human genome also encodes many types of GBPs that recognize microbial glycans and can distinguish human from microbial glycans (345–347). Such limitations of studies on the human glycome are not unique, however, as they apply to any animal glycome, and thus it is fair to say we know little overall about the total glycome of any animal.

Individual variations in the human glycome

There are many challenges to identifying the total human glycome, including technical and accessibility issues (147). Even with our limited knowledge, however, it is clear that there are age-, gender-, and disease-specific differences in the glycome, as seen, for example, for the serum *N*-glycome (348, 349), as well as significant differences in the human milk glycome from one individual to another (350), and from one mammalian species to another (323, 351). Such major differences in milk oligosaccharides between mammals highlights the importance of glycan *structure* as being a key feature of glycan evolution.

Variations in the human glycome may arise from the differential expression of so-called *glycogenes*, for which there may be >700 (147, 352). These are genes encoding enzymes and other proteins that are involved in biosynthesis, turnover, and elaboration of glycans of all classes; this would also include their polymorphic alleles and splice forms (353). However, there are likely to be many additional genes encoding modifiers, i.e., pathways altering the glycan structure or expression independently of creating a glycosidic bond, e.g., unusual acetyltransferases, methylases, sulfotransferases, kinases, isomerases, etc. (354). These lists do not include the many hundreds of genes encoding proteins that bind to specific glycans, which could also influence the diversity and expression of glycans. The constellation of genetic factors that dictate the eventual glycomolecule production both qualitatively and quantitatively in humans is being studied, as has been done for the genetic regulation of plasma protein glycosylation (355); however, it is difficult to imagine in the near future gaining comprehensive knowledge about all the structures of glycomolecules in all cells and tissues. Additionally, there is strong evidence that microRNAs (miRNA, miR), which are non-coding RNAs, play important roles in regulating protein glycosylation and the overall glycome (356, 357). There are also interesting aspects of monosaccharide metabolism relative glycomolecule production, as seen for the unusual sugar 2-keto-3-deoxy-D-glycero-D-galacto-2-nonulosonic acid (KDN) (358). KDN is a component of glycomolecules in cold-blooded vertebrates but not yet found in those of humans, although it is found in human plasma and tissues, and may serve to help regulate mannose metabolism. By contrast, Neu5Gc, which humans cannot synthesize, is found in some human glycomolecules through consumption of Neu5Gc in the diet and uptake by cells, and its presence induces anti-Neu5Gc antibodies (359, 360). Thus, there are continuing uncertainties about even the entire monosaccharide constellation of human glycans.

It may be concluded that studies on human glycans reveal that many glycans have structures clearly different even from other primates, and certainly different between humans compared to those in non-primates (273, 361) (Figure 1). The results suggest that human evolution is associated with discreet changes and diversity in the glycome. However, while humans have a highly diverse set of pathways for generating glycomolecules (362), most of these are conceptually similar to those in other primates and even most higher vertebrates.

Human glycome and ABO blood groups

Among the hundreds of human glycosyltransferase genes are also many polymorphic alleles (363). One of the best examples of polymorphism is the most ancient and best studied genetic feature of human glycosylation, which is the ABO blood group system. It represents carbohydrate antigens on red blood cells (RBCs) and some other cells, including intestinal cells. These antigens are comprised of Gal, GalNAc, GlcNAc, and Fuc monosaccharides in unique linkages that define each antigen (364). While the ABO blood group is relatively common among primates, there are tremendous variations in expression, and there are many hundreds of polymorphic alleles involved in ABO generation. There is also evidence that diversity in the ABO system is driven by the microbiome (324), which may be a version of host-invader interactions. This is interesting, as there is growing evidence that microbial expression of ABO-like antigens led to the generation of anti-ABO antibodies in cognate individuals; these are restricted, however, to recognize the non-expressed antigens in an individual (365). In addition, an ancient class of GBPs in metazoans is galectins, which interestingly can recognize blood group antigens and kill microbes expressing them in a type of host vs invader response (366). Thus, it is possible that the microbiome of an individual may not only shape the antibody repertoire in that person, but also shape the types of glycans expressed in RBCs, and the intestinal tract and other associated cells.

Even in regard to this simple issue of the ABO-containing glycans, we know little about the ABO glycome of humans or of any particular individual regarding the exact structures, either quantitatively or qualitatively, made by that individual compared to another, despite strong evidence of multiple disease susceptibilities associated with ABO blood group status (367). Only recently have studies provided insight into the *N*-glycans on RBC glycoproteins (368), and remarkably individual differences were observed between donors with different blood types. Humans also vary in many other blood types, often only thought of in regard to transfusion medicine, and these include the Lewis, P blood group, and the just

recently described Forssman blood group (317, 369). One could easily imagine that no two individuals would express exactly the same amounts or structures of all their blood group-related glycans and especially the constellation of all blood group antigens. Thus, in some ways the total glycome, if it could be known, might be as unique to an individual as the genome or fingerprints.

Therefore, many questions remain about the human glycome. We can group these types of questions using the ABO system, but they can apply to any specific aspect of the human glycome. As there are antibodies to many glycan determinants in the blood group systems, such tools could be used to aid in answering these questions. What glycans express ABO determinants in the human body? Where are they located? Which glycolipids, GAGs, and glycoproteins carry these? How do they vary between people? Do they change over time and with age? How does the ABO glycome of a person impact health and disease? Such simple questions could be asked about any cell in the body and any class of glycans. Unfortunately, the availability of multiple human materials for structural analyses is limited and the technologies do not yet scale sufficiently to adequately and perhaps reliably enough to address the complete human glycome (370), at least compared to the technologies useful for studies on the human genome and proteome. For example, consider a model organism that is well studied and exploited in immunology and in all types of questions in mammalian biology, the laboratory mouse *Mus musculus*. While we know its genome well, we know very little about its total glycome, though some systematic studies are underway (371–374), which, not surprisingly, are revealing major differences between the glycomes of each mouse organ, and mouse glycans compared to human glycans. Even different strains of mice can differ in certain aspects of their *N*-glycosylation of Fc domain of their IgG (375).

Constellation of evolutionary factors combining to generate glycomolecule diversity

The above considerations infer that there are many forces at play in the evolution of glycomolecules and that there are clearly major differences in glycomes in all unicellular organisms and all metazoans. Some of the major roles of glycomolecules in these ancient and ongoing processes leading to evolutionary change among organisms are summarized below.

Some molecular considerations driving the evolution of glycomolecules

- Glycosylation and sugar production is universal and appear at the earliest stages of macromolecule evolution.

- Glycosylation enhances diversity of protein modifications and functions independently of the genetic code.
- Carbohydrates with their multiple $-OH$ and $-NH_2$ groups and their nature as aldoses and ketoses permit multiple modifications and branching to generate large spatial structures.
- Glycomolecules coat the surfaces of all cells, providing a plethora of biological capabilities and potential interactions, in which changes in the glycocalyx, independently of surface protein expression per se, can lead to changes in cellular interactions with other cells and their microenvironment.
- Glycosylation of surface molecules participate in host-invader dynamics and provide survival advantages to host, as glycans are recognized by GBPs, toxins, and antibodies.
- Glycomolecules can function as intercellular adhesion molecules and comprise the ECM of multicellular organisms.
- Acquired functions for glycans is likely associated with changes in their structures and increasing complexity.
- For microbes, glycomolecules make up their biofilms and provide both protection and promote microbial ecology.
- Glycans are ancient forms of nutrient storage.
- *Glyco* components on a single species of protein or lipid can vary tremendously, giving rise to glycoforms that greatly multiplies the information content of a single encoded gene product.
- Glycomolecules, such as chitin and cellulose, are essentially insoluble and provide a structural component to organisms that is both easy to generate without a template and modular, i.e. size distributions can be enormous.
- The anomeric nature of monosaccharide linkages, α or β , provides a novel secondary structural feature of glycans that change the biological functions of monosaccharides.
- Many monosaccharides, being already relatively oxidized, are rather stable to heat and light and even wide pH ranges, comparable to other building blocks.
- Carbohydrates readily participate in H-bonds and can both promote intramolecular structure, extramolecular associations, and promote solubility of glycomolecules that might otherwise be relatively insoluble.
- The compartmentalization of glycomolecules and their biosynthesis within eukaryotic cells promotes novel biosynthetic pathways and allows control over assembly, storage, and release.
- Many enzymes and proteins involved in glycomolecule metabolism engage different metals as key cofactors, and potential variations in availability of biometals regulates glycomolecule generation and enzyme activities (376).

Conclusion

Understanding evolution is unimaginable without considering the emergence of carbohydrates and their key roles in structure and metabolism. Carbohydrates became essential components of all types of glycomolecules beyond the nucleic acids, as well as being central players in metabolism. The elaboration of glycomolecules beyond simple sugars and monosaccharides exploited the many advantages of carbohydrates over other types of building blocks, including their inherent multivalency due to multiple functional groups, solubility, ease of biosynthesis and metabolism, the low energy considerations of the glycosidic linkage compared to other bonds, e.g., peptide bonds, and the non-template dependence on glycan generation and modulation. The evolutionary factors involving competition between host-invader interactions and the emergence of the diverse functions of glycans associated with unique biological pathways are among several drivers of diversity of glycomolecules in viruses and all living cells. Emerging information about the glycogenomes of organisms, the genetic regulation of glycosylation, and the unique glycomes of each organism, is contributing to our enlightenment about the complex evolution of glycan structure and function. However, as the glycome cannot yet be directly predicted from the genome, new tools and analytical technologies are likely to further illuminate how we think about glyco-evolution and the dynamic landscape of glycan expression, recognition, and function in all life forms.

Acknowledgements

The author thanks Jamie Heimbürg-Molinaro, Sandra F. Cummings, Rajindra P. Aryal, Maxence Noel, and Akul Y. Mehta for reading and commenting on the manuscript and for their valuable suggestions. This work was supported by NIH Grant R24GM137763 to RDC.

References

1. Varki A, Kornfeld S. Historical background and overview. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 1–20.
2. Fraser-Reid BO, Tatsuta K, Thiem J. *Glycoscience: chemistry and chemical biology*, 2nd Ed. Berlin; New York: Springer; 2008.
3. Flynn RA, Pedram K, Malaker SA, Batista PJ, Smith BAH, Johnson AG, et al. Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell*. 2021;184: 3109–3124.e22.

4. Zhang N, Tang W, Torres L, Wang X, Ajaj Y, Zhu L, et al. Cell surface RNAs control neutrophil recruitment. *Cell*. 2024;187: 846–860.e17.
5. Wilson IBH, Paschinger K, Cummings RD, Aebi M. Nematoda. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 335–348.
6. Gagneux P, Panin V, Hennet T, Aebi M, Varki A. Evolution of glycan diversity. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 265–278.
7. Merry CLR, Lindahl U, Couchman J, Esko JD. Proteoglycans and sulfated glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 217–232.
8. Stanley P, Moremen KW, Lewis NE, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 103–116.
9. Whitfield C, Szymanski CM, Lewis AL, Aebi M. Eubacteria. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 279–296.
10. Speciale I, Di Lorenzo F, Notaro A, Noel E, Agarkova I, Molinaro A, et al. N-glycans from *Paramecium bursaria* chlorella virus MA-1D: re-evaluation of the oligosaccharide common core structure. *Glycobiology*. 2022;32: 260–273.
11. O'Neill MA, Darvill AG, Etzler ME, Mohnen D, Perez S, Mortimer JC, et al. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 319–334.
12. Paschinger K, Wols F, Yan S, Jin C, Vanbeselaere J, Dutkiewicz Z, et al. N-glycan antennal modifications are altered in *Caenorhabditis elegans* lacking the HEX-4 N-acetylgalactosamine-specific hexosaminidase. *The Journal of Biological Chemistry*. 2023;299: 103053.
13. Cummings RD. A periodic table of monosaccharides. *Glycobiology*. 2023: cwad088.
14. Imperiali B. Bacterial carbohydrate diversity – a Brave New World. *Current Opinion in Chemical Biology*. 2019;53: 1–8.
15. Laine RA. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*. 1994;4: 759–767.
16. Sackstein R, Hoffmeister KM, Stowell SR, Kinoshita T, Varki A, Freeze HH. Glycans in acquired human diseases. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 615–630.
17. Trujillo MN, Galligan JJ. Reconsidering the role of protein glycation in disease. *Nature Chemical Biology*. 2023;19: 922–927.
18. Schilling KM, Jorwal P, Ubilla-Rodriguez NC, Assafa TE, Gatdula JRP, Vultaggio JS, et al. N-glycosylation is a potent regulator of prion protein neurotoxicity. *The Journal of Biological Chemistry*. 2023;299: 105101.
19. Zhao X, Ma D, Ishiguro K, Saito H, Akichika S, Matsuzawa I, et al. Glycosylated queuosines in tRNAs optimize translational rate and post-embryonic growth. *Cell*. 2023;186: 5517–5535.e24.
20. Wang S, Sun E, Liu Y, Yin B, Zhang X, Li M, et al. Landscape of new nuclease-containing antiphage systems in *Escherichia coli* and the counterdefense roles of bacteriophage T4 genome modifications. *Journal of Virology*. 2023;97: e0059923.
21. Csoka AB, Stern R. Hypotheses on the evolution of hyaluronan: a highly ironic acid. *Glycobiology*. 2013;23: 398–411.
22. DeAngelis PL. Microbial glycosaminoglycan glycosyltransferases. *Glycobiology*. 2002;12: 9R–16R.
23. DeAngelis PL. Evolution of glycosaminoglycans and their glycosyltransferases: implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *The Anatomical Record*. 2002;268: 317–326.
24. Yamada S, Sugahara K, Ozbek S. Evolution of glycosaminoglycans: comparative biochemical study. *Communicative and Integrative Biology*. 2011;4: 150–158.
25. Dubrovin EV, Barinov NA, Ivanov DA, Klinov DV. Single-molecule AFM study of hyaluronic acid softening in electrolyte solutions. *Carbohydrate Polymers*. 2023;303: 120472.
26. LeWinter MM, Wu Y, Labelit S, Granzier H. Cardiac titin: structure, functions and role in disease. *Clinica Chimica Acta*. 2007;375: 1–9.
27. Lloyd KO, Yin BW, Kudryashov V. Isolation and characterization of ovarian cancer antigen CA 125 using a new monoclonal antibody (VK-8): identification as a mucin-type molecule. *International Journal of Cancer*. 1997;71: 842–850.
28. Das S, Batra SK. Understanding the unique attributes of MUC16 (CA125): potential implications in targeted therapy. *Cancer Research*. 2015;75: 4669–4674.
29. Higuchi T, Orita T, Nakanishi S, Katsuya K, Watanabe H, Yamasaki Y, et al. Molecular cloning, genomic structure, and expression analysis of MUC20, a novel mucin protein, up-regulated in injured kidney. *The Journal of Biological Chemistry*. 2004;279: 1968–1979.
30. Felder M, Kapur A, Gonzalez-Bosquet J, Horibata S, Heintz J, Albrecht R, et al. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Molecular Cancer*. 2014;13: 129.
31. Johansson ME, Ambort D, Pelaseyed T, Schutte A, Gustafsson JK, Ermund A, et al. Composition and functional role of the mucus layers in the intestine. *Cellular and Molecular Life Sciences*. 2011;68: 3635–3641.
32. Radicioni G, Cao R, Carpenter J, Ford AA, Wang T, Li L, et al. The innate immune properties of airway mucosal surfaces are

- regulated by dynamic interactions between mucins and interacting proteins: the mucin interactome. *Mucosal Immunology*. 2016;9: 1442–1454.
33. Tan S, Cheng PW. Mucin biosynthesis: identification of the cis-regulatory elements of human C2GnT-M gene. *American Journal of Respiratory Cell and Molecular Biology*. 2007;36: 737–745.
 34. Syed ZA, Zhang L, Ten Hagen KG. In vivo models of mucin biosynthesis and function. *Advanced Drug Delivery Reviews*. 2022;184: 114182.
 35. Johansson ME, Larsson JM, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(Suppl 1): 4659–4665.
 36. Tarp MA, Clausen H. Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochimica et Biophysica Acta*. 2008;1780: 546–563.
 37. Garde S, Chodiseti PK, Reddy M. Peptidoglycan: structure, synthesis, and regulation. *EcoSal Plus*. 2021;9.
 38. Klemm D, Heublein B, Fink HP, Bohn A. Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie (International ed. in English)*. 2005;44: 3358–3393.
 39. Jia Z, O'Mara ML, Zuegg J, Cooper MA, Mark AE. Vancomycin: ligand recognition, dimerization and super-complex formation. *The FEBS Journal*. 2013;280: 1294–1307.
 40. Rudd PM, Karlsson NG, Khoo KH, Thaysen-Andersen M, Wells L, Packer NH. Glycomics and glycoproteomics. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 689–704.
 41. Jayaraman V, Toledo-Patino S, Noda-Garcia L, Laurino P. Mechanisms of protein evolution. *Protein Science*. 2022;31: e4362.
 42. Lauc G, Kristic J, Zoldos V. Glycans – the third revolution in evolution. *Frontiers in Genetics*. 2014;5: 145.
 43. Varki A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell*. 2006;126: 841–845.
 44. Suzuki N. Glycan diversity in the course of vertebrate evolution. *Glycobiology*. 2019;29: 625–644.
 45. Springer SA, Gagneux P. Glycan evolution in response to collaboration, conflict, and constraint. *The Journal of Biological Chemistry*. 2013;288: 6904–6911.
 46. Varki A. Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harbor Perspectives in Biology*. 2011;3: a005462.
 47. West CM, Malzl D, Hykollari A, Wilson IBH. Glycomics, glycoproteomics, and glycogenomics: an inter-taxa evolutionary perspective. *Molecular and Cellular Proteomics*. 2021;20: 100024.
 48. Alberts B, Heald R, Johnson A, Morgan D, Raff M. Protein function. In: *Molecular biology of the cell*, 7th Ed. New York: WW. Norton & Company, Inc.; 2022. p. 195–222.
 49. Wilson TJ, Lilley DM. RNA catalysis – is that it? *RNA*. 2015;21: 534–537.
 50. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor Laboratory Press: Cold Spring Harbor; 2022.
 51. Murate M, Kobayashi T. Distribution of glycolipids in the plasma membrane monitored by specific probes in combination with sodium dodecyl sulfate-digested freeze-fracture replica labeling (SDS-FRL). *Methods in Molecular Biology*. 2023;2613: 189–202.
 52. Keenan EK, Zachman DK, Hirschey MD. Discovering the landscape of protein modifications. *Molecular Cell*. 2021;81: 1868–1878.
 53. Ramazi S, Allahverdi A, Zahiri J. Evaluation of post-translational modifications in histone proteins: a review on histone modification defects in developmental and neurological disorders. *Journal of Biosciences*. 2020;45: 135.
 54. Zhang X, Li D, Zhu J, Zheng J, Li H, He Q, et al. RNAPII degradation factor Def1 is required for development, stress response, and full virulence of *magnaporthe oryzae*. *Journal of Fungi (Basel)*. 2023;9: 467.
 55. Hart GW, Housley MP, Slawson C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature*. 2007;446: 1017–1022.
 56. Zachara NE, Akimoto Y, Boyce M, Hart GW. The O-GlcNAc modification. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 251–264.
 57. Akan I, Olivier-Van Stichelen S, Bond MR, Hanover JA. Nutrient-driven O-GlcNAc in proteostasis and neurodegeneration. *Journal of Neurochemistry*. 2018;144: 7–34.
 58. Mannino MP, Hart GW. The beginner's guide to O-GlcNAc: from nutrient sensitive pathway regulation to its impact on the immune system. *Frontiers in Immunology*. 2022;13: 828648.
 59. Krause MW, Love DC, Ghosh SK, Wang P, Yun S, Fukushige T, et al. Nutrient-driven O-GlcNAcylation at promoters impacts genome-Wide RNA Pol II distribution. *Front Endocrinology (Lausanne)*. 2018;9: 521.
 60. Bandini G, Haserick JR, Motari E, Ouologuem DT, Lourido S, Roos DS, et al. O-fucosylated glycoproteins form assemblies in close proximity to the nuclear pore complexes of *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113: 11567–11572.
 61. Sun TP. Novel nucleocytoplasmic protein O-fucosylation by SPINDLY regulates diverse developmental processes in plants. *Current Opinion in Structural Biology*. 2021;68: 113–121.
 62. West CM, Blader IJ. Oxygen sensing by protozoans: how they catch their breath. *Current Opinion in Microbiology*. 2015;26: 41–47.
 63. Sekar K, Linker SM, Nguyen J, Grunhagen A, Stocker R, Sauer U. Bacterial glycogen provides short-term benefits in changing

- environments. *Applied and Environmental Microbiology*. 2020;86: e00049–20.
64. Liu QH, Tang JW, Wen PB, Wang MM, Zhang X, Wang L. From prokaryotes to eukaryotes: insights into the molecular structure of glycogen particles. *Frontiers in Molecular Biosciences*. 2021;8: 673315.
 65. Pfister B, Zeeman SC. Formation of starch in plant cells. *Cellular and Molecular Life Sciences*. 2016;73: 2781–2807.
 66. Lomako J, Lomako WM, Whelan WJ. Glycogenin: the primer for mammalian and yeast glycogen synthesis. *Biochimica et Biophysica Acta*. 2004;1673: 45–55.
 67. Sun RC, Young LEA, Bruntz RC, Markussen KH, Zhou Z, Conroy LR, et al. Brain glycogen serves as a critical glucosamine cache required for protein glycosylation. *Cell Metabolism*. 2021;33: 1404–1417.e9.
 68. Li C, Hu Z. Is liver glycogen fragility a possible drug target for diabetes? *FASEB Journal*. 2020;34: 3–15.
 69. Visuttijai K, Hedberg-Oldfors C, Thomsen C, Glamuzina E, Kornblum C, Tasca G, et al. Glycogenin is dispensable for glycogen synthesis in human muscle, and glycogenin deficiency causes polyglucosan storage. *The Journal of Clinical Endocrinology and Metabolism*. 2020;105: 557–566.
 70. Testoni G, Duran J, Garcia-Rocha M, Vilaplana F, Serrano AL, Sebastian D, et al. Lack of glycogenin causes glycogen accumulation and muscle function impairment. *Cell Metabolism*. 2017;26: 256–266.e4.
 71. Tetlow IJ, Bertoft E. A review of starch biosynthesis in relation to the building block-backbone model. *International Journal of Molecular Sciences*. 2020;21: 7011.
 72. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 2002;12: 43R–56R.
 73. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. *Nature Reviews Molecular Cell Biology*. 2020;21: 729–749.
 74. Stepper J, Shastri S, Loo TS, Preston JC, Novak P, Man P, et al. Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Letters*. 2011;585: 645–650.
 75. Olsen EH, Rahbek-Nielsen H, Thogersen IB, Roepstorff P, Engild JJ. Posttranslational modifications of human inter-alpha-inhibitor: identification of glycans and disulfide bridges in heavy chains 1 and 2. *Biochemistry*. 1998;37: 408–416.
 76. Maynard JC, Burlingame AL, Medzihradszky KF. Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), a new post-translational modification in mammals. *Molecular and Cellular Proteomics*. 2016;15: 3405–3411.
 77. Xiao H, Wu R. Global and site-specific analysis revealing unexpected and extensive protein S-GlcNAcylation in human cells. *Analytical Chemistry*. 2017;89: 3656–3663.
 78. Di Marco F, Blochl C, Esser-Skala W, Schapertons V, Zhang T, Wuhler M, et al. Glycoproteomics of a single protein: revealing tens of thousands of myozyme glycoforms by hybrid HPLC-MS approaches. *Molecular and Cellular Proteomics*. 2023;22: 100622.
 79. Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports*. 2020;10: 14991.
 80. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*. 1990;87: 4576–4579.
 81. Messner P. Bacterial glycoproteins. *Glycoconjugate Journal*. 1997;14: 3–11.
 82. Schaffer C, Messner P. Emerging facets of prokaryotic glycosylation. *FEMS Microbiology Reviews*. 2017;41: 49–91.
 83. Upreti RK, Kumar M, Shankar V. Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics*. 2003;3: 363–379.
 84. Barrett K, Dube DH. Chemical tools to study bacterial glycans: a tale from discovery of glycoproteins to disruption of their function. *Israel Journal of Chemistry*. 2023;63: e202200050.
 85. Janhvi S, Saini S, Singh V, Sharma T, Rao A. ProGlycProt V3.0: updated insights into prokaryotic glycoproteins and their glycosyltransferases. *Glycobiology*. 2023: cwad103.
 86. Mescher MF, Strominger JL. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarum*. *The Journal of Biological Chemistry*. 1976;251: 2005–2014.
 87. Abu-Qarn M, Eichler J, Sharon N. Not just for *Eukarya* anymore: protein glycosylation in Bacteria and Archaea. *Current Opinion in Structural Biology*. 2008;18: 544–550.
 88. Benz I, Schmidt MA. Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology*. 2002;45: 267–276.
 89. Sleytr UB, Schuster B, Egelseer EM, Pum D. S-layers: principles and applications. *FEMS Microbiology Reviews*. 2014;38: 823–864.
 90. Fagan RP, Fairweather NF. Biogenesis and functions of bacterial S-layers. *Nature Reviews Microbiology*. 2014;12: 211–222.
 91. Comstock LE. Importance of glycans to the host-bacteroides mutualism in the mammalian intestine. *Cell Host and Microbe*. 2009;5: 522–526.
 92. Veith PD, Gorasia DG, Reynolds EC. Characterization of the O-Glycoproteome of *Flavobacterium johnsoniae*. *Journal of Bacteriology*. 2023;205: e0009323.
 93. Seepersaud R, Anderson AC, Bensing BA, Choudhury BP, Clarke AJ, Sullam PM. O-acetylation controls the glycosylation of bacterial serine-rich repeat glycoproteins. *The Journal of Biological Chemistry*. 2021;296: 100249.
 94. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA. Structure of the fibre-forming protein pilin at 2.6 Å resolution. *Nature*. 1995;378: 32–38.

95. Benz I, Schmidt MA. Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Molecular Microbiology*. 2001;40: 1403–1413.
96. Yari S, Afrough P, Yari F, Ghazanfari Jajin M, Fateh A, Hadizadeh Tasbiti A. A potent subset of *Mycobacterium tuberculosis* glycoproteins as relevant candidates for vaccine and therapeutic target. *Scientific Reports*. 2023;13: 22194.
97. Tucci P, Portela M, Chetto CR, Gonzalez-Sapienza G, Marin M. Integrative proteomic and glycoproteomic profiling of *Mycobacterium tuberculosis* culture filtrate. *PLoS One*. 2020;15: e0221837.
98. Szymanski CM. Bacterial glycosylation, it's complicated. *Frontiers in Molecular Biosciences*. 2022;9: 1015771.
99. Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, Lanthier PH, et al. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *The Journal of Biological Chemistry*. 2002;277: 42530–42539.
100. Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology*. 1999;32: 1022–1030.
101. Ristl R, Steiner K, Zarschler K, Zayni S, Messner P, Schaffer C. The s-layer glycome-adding to the sugar coat of bacteria. *International Journal of Microbiology*. 2011;2011: 127870.
102. Oman TJ, Boettcher JM, Wang H, Okalibe XN, van der Donk WA. Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nature Chemical Biology*. 2011;7: 78–80.
103. Waglechner N, McArthur AG, Wright GD. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nature Microbiology*. 2019;4: 1862–1871.
104. Sacoman JL, Dagda RY, Burnham-Marusich AR, Dagda RK, Berninson PM. Mitochondrial O-GlcNAc transferase (mOGT) regulates mitochondrial structure, function, survival in HeLa cells. *The Journal of Biological Chemistry*. 2017;292: 4499–4518.
105. Mikkola S. Nucleotide sugars in chemistry and biology. *Molecules*. 2020;25: 5755.
106. Surmacz L, Swiezewska E. Polyisoprenoids – secondary metabolites or physiologically important superlipids? *Biochemical and Biophysical Research Communications*. 2011;407: 627–632.
107. Banfalvi G. Prebiotic pathway from ribose to RNA formation. *International Journal of Molecular Sciences*. 2021;22: 3857.
108. Delidovich IV, Simonov AN, Taran OP, Parmon VN. Catalytic formation of monosaccharides: from the formose reaction towards selective synthesis. *ChemSusChem*. 2014;7: 1833–1846.
109. Krisnamurthy R, Liotta CL. The potential of glyoxylate as a prebiotic source molecule and a reactant in protometabolic pathways—The glyoxylose reaction. *Chem Biology*. 2023;9.
110. Yi R, Mojica M, Fahrenbach AC, James Cleaves H, 2nd, Krishnamurthy R, Liotta CL. Carbonyl migration in uronates affords a potential prebiotic pathway for pentose production. *Journal of the American Chemical Society*. 2023;145: 2522–2535.
111. Higgs PG, Lehman N. The RNA World: molecular cooperation at the origins of life. *Nature Reviews Genetics*. 2015;16: 7–17.
112. Omran A, Menor-Salvan C, Springsteen G, Pasek M. The messy alkaline formose reaction and its link to metabolism. *Life (Basel)*. 2020;10: 125.
113. Benner SA, Kim HJ, Kim MJ, Ricardo A. Planetary organic chemistry and the origins of biomolecules. *Cold Spring Harbor Perspectives in Biology*. 2010;2: a003467.
114. Cooper G, Kimmich N, Belisle W, Sarinana J, Brabham K, Garrel L. Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth. *Nature*. 2001;414: 879–883.
115. Furukawa Y, Chikaraishi Y, Ohkouchi N, Ogawa NO, Glavin DP, Dworkin JP, et al. Extraterrestrial ribose and other sugars in primitive meteorites. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116: 24440–24445.
116. Zellner NEB, McCaffrey VP, Butler JHE. Cometary glycolaldehyde as a source of pre-RNA molecules. *Astrobiology*. 2020;20: 1377–1388.
117. Tsai CH, Chen J, Szostak JW. Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104: 14598–14603.
118. Banfalvi G. Why ribose was selected as the sugar component of nucleic acids. *DNA and Cell Biology*. 2006;25: 189–196.
119. Banfalvi G. Ribose selected as precursor to life. *DNA and Cell Biology*. 2020;39: 177–186.
120. Bernhardt HS, Tate WP. Primordial soup or vinaigrette: did the RNA world evolve at acidic pH? *Biology Direct*. 2012;7: 4.
121. Yi R, Tran QP, Ali S, Yoda I, Adam ZR, Cleaves HJ, 2nd, et al. A continuous reaction network that produces RNA precursors. *Proceedings of the National Academy of Sciences of the United States of America*. 2020;117: 13267–13274.
122. Rivas M, Becerra A, Lazcano A. On the early evolution of catabolic pathways: a comparative genomics Approach. I. The cases of glucose, ribose, and the nucleobases catabolic routes. *Journal of Molecular Evolution*. 2018;86: 27–46.
123. Blackmond DG. The origin of biological homochirality. *Cold Spring Harbor Perspectives in Biology*. 2019;11: a032540.
124. Breslow R, Cheng ZL. L-amino acids catalyze the formation of an excess of D-glyceraldehyde, and thus of other D sugars, under credible prebiotic conditions. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107: 5723–5725.
125. Wiercigroch E, Szafranec E, Czamara K, Pacia MZ, Majzner K, Kochan K, et al. Raman and infrared spectroscopy of carbohydrates: a review. *Spectrochimica Acta Part A, Biomolecular Spectroscopy*. 2017;185: 317–335.

126. Abdulbagi M, Wang L, Siddiq O, Di B, Li B. D-amino acids and D-amino acid-containing peptides: potential disease biomarkers and therapeutic targets? *Biomolecules*. 2021;11: 1716.
127. Gao DM, Kobayashi T, Adachi S. Kinetic analysis for the isomerization of glucose, fructose, and mannose in subcritical aqueous ethanol. *Bioscience, Biotechnology, and Biochemistry*. 2015;79: 1005–1010.
128. Apriyanto A, Compart J, Fettke J. A review of starch, a unique biopolymer – Structure, metabolism and in planta modifications. *Plant Science*. 2022;318: 111223.
129. Pfister B, Zeeman SC, Rugen MD, Field RA, Ebenhoeh O, Raguin A. Theoretical and experimental approaches to understand the biosynthesis of starch granules in a physiological context. *Photosynthesis Research*. 2020;145: 55–70.
130. Damager I, Engelsen SB, Blennow A, Moller BL, Motawia MS. First principles insight into the alpha-glucan structures of starch: their synthesis, conformation, and hydration. *Chemical Reviews*. 2010;110: 2049–2080.
131. Gentry MS, Guinovart JJ, Minassian BA, Roach PJ, Serratos JM. Lafora disease offers a unique window into neuronal glycogen metabolism. *The Journal of Biological Chemistry*. 2018;293: 7117–7125.
132. Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *Journal of Experimental Botany*. 2011;62: 1775–1801.
133. Mollers KB, Cannella D, Jorgensen H, Frigaard NU. Cyanobacterial biomass as carbohydrate and nutrient feedstock for bioethanol production by yeast fermentation. *Biotechnology for Biofuels*. 2014;7: 64.
134. Xu X, Yang Y, Liu C, Sun Y, Zhang T, Hou M, et al. The evolutionary history of the sucrose synthase gene family in higher plants. *BMC Plant Biology*. 2019;19: 566.
135. Rabbani N, Thornalley PJ. Glycation research in amino acids: a place to call home. *Amino Acids*. 2012;42: 1087–1096.
136. Twarda-Clapa A, Olczak A, Bialkowska AM, Koziolkiewicz M. Advanced glycation end-products (AGEs): formation, chemistry, classification, receptors, and diseases related to AGEs. *Cells*. 2022;11: 1312.
137. Richards AB, Krakowka S, Dexter LB, Schmid H, Wolterbeek AP, Waalkens-Berendsen DH, et al. Trehalose: a review of properties, history of use and human tolerance, and results of multiple safety studies. *Food and Chemical Toxicology*. 2002;40: 871–898.
138. Elbein AD, Pan YT, Pastuszak I, Carroll D. New insights on trehalose: a multifunctional molecule. *Glycobiology*. 2003;13: 17R–27R.
139. Xiang J, Liu F, Wang B, Chen L, Liu W, Tan S. A literature review on maillard reaction based on milk proteins and carbohydrates in food and pharmaceutical products: advantages, disadvantages, and avoidance strategies. *Foods*. 2021;10: 1998.
140. Cardona T. Origin of bacteriochlorophyll A and the early diversification of photosynthesis. *PLoS One*. 2016;11: e0151250.
141. Rothschild LJ. The evolution of photosynthesis...again? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2008;363: 2787–2801.
142. Stirbet A, Lazar D, Guo Y, Govindjee G. Photosynthesis: basics, history and modelling. *Annals of Botany*. 2020;126: 511–537.
143. Blankenship RE. Early evolution of photosynthesis. *Plant Physiology*. 2010;154: 434–438.
144. Margulis L. *Symbiosis in cell evolution*. San Francisco: W.H. Freeman; 1992.
145. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EM, Chisholm SW. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*. 2006;311: 1737–1740.
146. Kinoshita T. Glycosylphosphatidylinositol (GPI) anchors: biochemistry and cell biology: introduction to a thematic review series. *Journal of Lipid Research*. 2016;57: 4–5.
147. Cummings RD, Pierce JM. The challenge and promise of glycomics. *Chem Biology*. 2014;21: 1–15.
148. Goettig P. Effects of glycosylation on the enzymatic activity and mechanisms of proteases. *International Journal of Molecular Sciences*. 2016;17: 1969.
149. Goth CK, Mehta AY, McQuillan AM, Baker KJ, Hanes MS, Park SS, et al. Chemokine binding to PSGL-1 is controlled by O-glycosylation and tyrosine sulfation. *Cell Chemical Biology*. 2023;30: 893–905.e7.
150. Yamauchi M, Sricholpech M. Lysine post-translational modifications of collagen. *Essays in Biochemistry*. 2012;52: 113–133.
151. Tanwar A, Stanley P. Synergistic regulation of Notch signaling by different O-glycans promotes hematopoiesis. *Frontiers in Immunology*. 2023;14: 1097332.
152. Bonnardel F, Haslam SM, Dell A, Feizi T, Liu Y, Tajadura-Ortega V, et al. Proteome-wide prediction of bacterial carbohydrate-binding proteins as a tool for understanding commensal and pathogen colonisation of the vaginal microbiome. *Nature Partner Journals Biofilms and Microbiomes*. 2021;7: 49.
153. Pan Y, Xia L. Emerging roles of podoplanin in vascular development and homeostasis. *Frontiers of Medicine*. 2015;9: 421–430.
154. Serio AW, Keepers T, Andrews L, Krause KM. Aminoglycoside revival: review of a historically important class of antimicrobials undergoing rejuvenation. *EcoSal Plus*. 2018;8.
155. Blaskovich MAT, Hansford KA, Butler MS, Jia Z, Mark AE, Cooper MA. Developments in glycopeptide antibiotics. *American Chemical Society Infectious Diseases*. 2018;4: 715–735.
156. Hanafusa K, Hotta T, Iwabuchi K. Glycolipids: linchpins in the organization and function of membrane microdomains. *Frontiers in Cell and Developmental Biology*. 2020;8: 589799.
157. Marques AR, Mirzaian M, Akiyama H, Wisse P, Ferraz MJ, Gaspar P, et al. Glucosylated cholesterol in mammalian cells and tissues: formation and degradation by multiple cellular beta-glucosidases. *Journal of Lipid Research*. 2016;57: 451–463.
158. Guo Z. The structural diversity of natural glycosphingolipids (GSLs). *Journal of Carbohydrate Chemistry*. 2022;41: 63–154.

159. Kalisch B, Dormann P, Holz G. DGDG and Glycolipids in plants and Algae. *Sub Cellular Biochemistry*. 2016;86: 51–83.
160. Xatse MA, Olsen CP. Defining the glucosylceramide population of *C. elegans*. *Frontiers in Physiology*. 2023;14: 1244158.
161. Jennemann R, Sandhoff R, Wang S, Kiss E, Gretz N, Zuliani C, et al. Cell-specific deletion of glucosylceramide synthase in brain leads to severe neural defects after birth. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102: 12459–12464.
162. Msanne J, Chen M, Luttgeharm KD, Bradley AM, Mays ES, Paper JM, et al. Glucosylceramides are critical for cell-type differentiation and organogenesis, but not for cell viability in *Arabidopsis*. *The Plant Journal*. 2015;84: 188–201.
163. Allen KN, Imperiali B. Structural and mechanistic themes in glycoconjugate biosynthesis at membrane interfaces. *Current Opinion in Structural Biology*. 2019;59: 81–90.
164. Abdel-Mawgoud AM, Stephanopoulos G. Simple glycolipids of microbes: chemistry, biological activity and metabolic engineering. *Synthetic and Systems Biotechnology*. 2018;3: 3–19.
165. Kelleher DJ, Banerjee S, Cura AJ, Samuelson J, Gilmore R. Dolichol-linked oligosaccharide selection by the oligosaccharyl-transferase in protist and fungal organisms. *The Journal of Cell Biology*. 2007;177: 29–37.
166. Jones MB, Rosenberg JN, Betenbaugh MJ, Krag SS. Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. *Biochimica et Biophysica Acta*. 2009;1790: 485–494.
167. Oriol R, Martinez-Duncker I, Chantret I, Mollicone R, Codogno P. Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. *Molecular Biology and Evolution*. 2002;19: 1451–1463.
168. Swiezewska E, Danikiewicz W. Polyisoprenoids: structure, biosynthesis and function. *Progress in Lipid Research*. 2005;44: 235–258.
169. Hartley MD, Imperiali B. At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. *Archives of Biochemistry and Biophysics*. 2012;517: 83–97.
170. Hunter WN. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *The Journal of Biological Chemistry*. 2007;282: 21573–21577.
171. Gould SB. Membranes and evolution. *Current Biology*. 2018;28: R381–R385.
172. Mulikjanian AY, Galperin MY, Koonin EV. Co-evolution of primordial membranes and membrane proteins. *Trends in Biochemical Sciences*. 2009;34: 206–215.
173. Richter JR, Sanderson RD. The glycocalyx: pathobiology and repair. *Matrix Biology Plus*. 2023;17: 100128.
174. Ruiz N. Lipid flippases for bacterial peptidoglycan biosynthesis. *Lipid Insights*. 2015;8: 21–31.
175. Sanyal S, Menon AK. Flipping lipids: why an' what's the reason for? *ACS Chemical Biology*. 2009;4: 895–909.
176. Fujikawa K, Han Y, Osawa T, Mori S, Nomura K, Muramoto M, et al. Structural requirements of a glycolipid MPLase for membrane protein integration. *Chemistry*. 2023;29: e202300437.
177. Chalal M, Menon I, Turan Z, Menon AK. Reconstitution of glucosylceramide flip-flop across endoplasmic reticulum: implications for mechanism of glycosphingolipid biosynthesis. *The Journal of Biological Chemistry*. 2012;287: 15523–15532.
178. Reza S, Ugorski M, Suchanski J. Glucosylceramide and galactosylceramide, small glycosphingolipids with significant impact on health and disease. *Glycobiology*. 2021;31: 1416–1434.
179. Zelnik ID, Ventura AE, Kim JL, Silva LC, Futerman AH. The role of ceramide in regulating endoplasmic reticulum function. *Biochimica et Biophysica Acta Mol Cell Biol Lipids*. 2020;1865: 158489.
180. Sprong H, Kruithof B, Leijendekker R, Slot JW, van Meer G, van der Sluijs P. UDP-galactose:ceramide galactosyltransferase is a class I integral membrane protein of the endoplasmic reticulum. *The Journal of Biological Chemistry*. 1998;273: 25880–25888.
181. Ardail D, Popa I, Bodennec J, Louisot P, Schmitt D, Portoukalian J. The mitochondria-associated endoplasmic-reticulum subcompartment (MAM fraction) of rat liver contains highly active sphingolipid-specific glycosyltransferases. *The Biochemical Journal*. 2003;371: 1013–1019.
182. Munster-Kuhnel AK, Tiralongo J, Krapp S, Weinhold B, Ritz-Sedlacek V, Jacob U, et al. Structure and function of vertebrate CMP-sialic acid synthetases. *Glycobiology*. 2004;14: 43R–51R.
183. Freeze HH, Steet R, Suzuki T, Kinoshita T, Schnaar RL. Genetic disorders of glycan degradation. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 583–598.
184. Zimmer J. Structural features underlying recognition and translocation of extracellular polysaccharides. *Interface Focus*. 2019;9: 20180060.
185. Chen DD, Wang ZB, Wang LX, Zhao P, Yun CH, Bai L. Structure, catalysis, chitin transport, and selective inhibition of chitin synthase. *Nature Communications*. 2023;14: 4776.
186. Greenfield LK, Whitfield C. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydrate Research*. 2012;356: 12–24.
187. Raetz CR, Whitfield C. Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*. 2002;71: 635–700.
188. Yamanaka K, Maruyama C, Takagi H, Hamano Y. Epsilon-poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. *Nature Chemical Biology*. 2008;4: 766–772.
189. Shriml S, Cherepanova NA, Gilmore R. Cotranslational and posttranslational N-glycosylation of proteins in the endoplasmic reticulum. *Seminars in Cell and Developmental Biology*. 2015;41: 71–78.
190. Suzuki T, Cummings RD, Aebi M, Parodi A. Glycans in glycoprotein quality control. In: Varki A, Cummings RD, Esko JD, Stanley

- P, Hart GW, Aebersold M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 529–538.
191. Suzuki T, Yoshida Y. Ever-expanding NGLY1 biology. *Journal of Biochemistry*. 2022;171: 141–143.
 192. Harada Y, Ohkawa Y, Maeda K, Taniguchi N. Glycan quality control in and out of the endoplasmic reticulum of mammalian cells. *The FEBS Journal*. 2022;289: 7147–7162.
 193. Parodi AJ, Lederkremer GZ, Mendelzon DH. Protein glycosylation in *Trypanosoma cruzi*. The mechanism of glycosylation and structure of protein-bound oligosaccharides. *The Journal of Biological Chemistry*. 1983;258: 5589–5595.
 194. Parodi AJ. Protein glucosylation and its role in protein folding. *Annual Review of Biochemistry*. 2000;69: 69–93.
 195. Banerjee S, Vishwanath P, Cui J, Kelleher DJ, Gilmore R, Robbins PW, et al. The evolution of N-glycan-dependent endoplasmic reticulum quality control factors for glycoprotein folding and degradation. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104: 11676–11681.
 196. Mary B, Maurya S, Arumugam S, Kumar V, Jayandharan GR. Post-translational modifications in capsid proteins of recombinant adeno-associated virus (AAV) 1-rh10 serotypes. *The FEBS Journal*. 2019;286: 4964–4981.
 197. Li Y, Liu D, Wang Y, Su W, Liu G, Dong W. The importance of glycans of viral and host proteins in enveloped virus infection. *Frontiers in Immunology*. 2021;12: 638573.
 198. Feng T, Zhang J, Chen Z, Pan W, Chen Z, Yan Y, et al. Glycosylation of viral proteins: implication in virus-host interaction and virulence. *Virulence*. 2022;13: 670–683.
 199. Akiyama H, Miller C, Patel HV, Hatch SC, Archer J, Ramirez NG, et al. Virus particle release from glycosphingolipid-enriched microdomains is essential for dendritic cell-mediated capture and transfer of HIV-1 and henipavirus. *Journal of Virology*. 2014;88: 8813–8825.
 200. Banerjee N, Mukhopadhyay S. Viral glycoproteins: biological role and application in diagnosis. *Virusdisease*. 2016;27: 1–11.
 201. Checkley MA, Luttge BG, Freed EO. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology*. 2011;410: 582–608.
 202. Miller NL, Clark T, Raman R, Sasisekharan R. Glycans in virus-host interactions: a structural perspective. *Frontiers in Molecular Biosciences*. 2021;8: 666756.
 203. Wagh K, Hahn BH, Korber B. Hitting the sweet spot: exploiting HIV-1 glycan shield for induction of broadly neutralizing antibodies. *Current Opinion in HIV and AIDS*. 2020;15: 267–274.
 204. Chauhan RP, Gordon ML. An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates. *Virus Genes*. 2022;58: 255–269.
 205. Thompson AJ, Paulson JC. Adaptation of influenza viruses to human airway receptors. *The Journal of Biological Chemistry*. 2021;296: 100017.
 206. Zimmermann W, Broll H, Ehlers B, Buhk HJ, Rosenthal A, Goltz M. Genome sequence of bovine herpesvirus 4, a bovine *Rhadinovirus*, and identification of an origin of DNA replication. *Journal of Virology*. 2001;75: 1186–1194.
 207. Lete C, Markine-Goriaynoff N, Machiels B, Pang PC, Xiao X, Canis K, et al. Bovine herpesvirus 4 modulates its beta-1,6-N-acetylglucosaminyltransferase activity through alternative splicing. *Journal of Virology*. 2016;90: 2039–2051.
 208. Jackson RJ, Hall DF, Kerr PJ. Myxoma virus encodes an alpha2,3-sialyltransferase that enhances virulence. *Journal of Virology*. 1999;73: 2376–2384.
 209. O'Reilly DR, Miller LK. A baculovirus blocks insect molting by producing ecdysteroid UDP-glucosyl transferase. *Science*. 1989;245: 1110–1112.
 210. O'Reilly DR. Baculovirus-encoded ecdysteroid UDP-glucosyltransferases. *Insect Biochemistry and Molecular Biology*. 1995;25: 541–550.
 211. Hiraiwa N, Yabuta T, Yoritomi K, Hiraiwa M, Tanaka Y, Suzuki T, et al. Transactivation of the fucosyltransferase VII gene by human T-cell leukemia virus type 1 Tax through a variant cAMP-responsive element. *Blood*. 2003;101: 3615–3621.
 212. Markine-Goriaynoff N, Gillet L, Van Etten JL, Korres H, Verma N, Vanderplasschen A. Glycosyltransferases encoded by viruses. *The Journal of General Virology*. 2004;85: 2741–2754.
 213. Speciale I, Notaro A, Abergel C, Lanzetta R, Lowary TL, Molinaro A, et al. The astounding world of glycans from giant viruses. *Chemical Reviews*. 2022;122: 15717–15766.
 214. Piacente F, De Castro C, Jeudy S, Molinaro A, Salis A, Damonte G, et al. Giant virus *Megavirus chilensis* encodes the biosynthetic pathway for uncommon acetamido sugars. *The Journal of Biological Chemistry*. 2014;289: 24428–24439.
 215. Piacente F, Gaglianone M, Laugieri ME, Tonetti MG. The autonomous glycosylation of large DNA viruses. *International Journal of Molecular Sciences*. 2015;16: 29315–29328.
 216. Van Etten JL, Agarkova IV, Dunigan DD. Chloroviruses. *Viruses*. 2019;12: 20.
 217. De Castro C, Duncan GA, Garozzo D, Molinaro A, Sturiale L, Tonetti M, et al. Biophysical approaches to solve the structures of the complex glycan shield of chloroviruses. *Advances in Experimental Medicine and Biology*. 2018;1104: 237–257.
 218. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, et al. Giant *Marseillevirus* highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106: 21848–21853.
 219. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennot T. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. *The Journal of Biological Chemistry*. 2011;286: 43701–43709.

220. Shah N, Hulsmeier AJ, Hochhold N, Neidhart M, Gay S, Hennen T. Exposure to mimivirus collagen promotes arthritis. *Journal of Virology*. 2014;88: 838–845.
221. Messner P, Schaffer C, Kosma P. Bacterial cell-envelope glycoconjugates. *Advances in Carbohydrate Chemistry and Biochemistry*. 2013;69: 209–272.
222. Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology*. 2010;2: a000414.
223. Meyer BH, Albers SV, Eichler J, Aebi M. Archaea. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 297–306.
224. Gorska-Fraczek S, Sandstrom C, Kenne L, Rybka J, Strus M, Heczko P, et al. Structural studies of the exopolysaccharide consisting of a nonasaccharide repeating unit isolated from *Lactobacillus rhamnosus* KL37B. *Carbohydrate Research*. 2011;346: 2926–2932.
225. Rietschel ET, Kirikae T, Schade FU, Mamat U, Schmidt G, Loppnow H, et al. Bacterial endotoxin: molecular relationships of structure to activity and function. *FASEB Journal*. 1994;8: 217–225.
226. Klein G, Raina S. Regulated control of the assembly and diversity of LPS by noncoding sRNAs. *BioMed Research International*. 2015;2015: 153561.
227. Poxton IR. Antibodies to lipopolysaccharide. *Journal of Immunological Methods*. 1995;186: 1–15.
228. Kalynych S, Morona R, Cygler M. Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiology Reviews*. 2014;38: 1048–1065.
229. Zhang X, Payne M, Kaur S, Lan R. Improved genomic identification, clustering, and serotyping of shiga toxin-producing *Escherichia coli* using cluster/serotype-specific gene markers. *Frontiers in Cellular and Infection Microbiology*. 2021;11: 772574.
230. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. Salmonella nomenclature. *Journal of Clinical Microbiology*. 2000;38: 2465–2467.
231. Kehr JC, Dittmann E. Biosynthesis and function of extracellular glycans in cyanobacteria. *Life (Basel)*. 2015;5: 164–180.
232. Klingl A, Pickl C, Flechsler J. Archaeal cell walls. *Sub Cellular Biochemistry*. 2019;92: 471–493.
233. Assandri MH, Malamud M, Trejo FM, Serradell MLA. S-layer proteins as immune players: tales from pathogenic and non-pathogenic bacteria. *Current Research in Microbial Sciences*. 2023;4: 100187.
234. Seifert GJ, Strasser R, Van Damme EJM. Editorial: Plant glycobiology – a sweet world of glycans, glycoproteins, glycolipids, and carbohydrate-binding proteins. *Frontiers in Plant Science*. 2021;12: 751923.
235. Fogarty CA, Harbison AM, Dugdale AR, Fadda E. How and why plants and human N-glycans are different: insight from molecular dynamics into the “glycoblocks” architecture of complex carbohydrates. *Beilstein Journal of Organic Chemistry*. 2020;16: 2046–2056.
236. Strasser R. Plant protein glycosylation. *Glycobiology*. 2016;26: 926–939.
237. Beihammer G, Maresch D, Altmann F, Strasser R. Glycosylphosphatidylinositol-anchor synthesis in plants: a glycobiology perspective. *Frontiers in Plant Science*. 2020;11: 611188.
238. Strasser R. Biological significance of complex N-glycans in plants and their impact on plant physiology. *Frontiers in Plant Science*. 2014;5: 363.
239. Xu J, Du H, Shi H, Song J, Yu J, Zhou Y. Protein O-glycosylation regulates diverse developmental and defense processes in plants. *Journal of Experimental Botany*. 2023;74: 6119–6130.
240. Sarrazin S, Lamanna WC, Esko JD. Heparan sulfate proteoglycans. *Cold Spring Harbor Perspectives in Biology*. 2011;3: a004952.
241. Lei L. Polysaccharide structure: a hint from gut bacteria. *Nature Plants*. 2017;3: 17062.
242. Mohnen D. Pectin structure and biosynthesis. *Current Opinion in Plant Biology*. 2008;11: 266–277.
243. Atmodjo MA, Hao Z, Mohnen D. Evolving views of pectin biosynthesis. *Annual Review of Plant Biology*. 2013;64: 747–779.
244. Le Normand M, Rietzler B, Vilaplana F, Ek M. Macromolecular model of the pectic polysaccharides isolated from the bark of Norway Spruce (*Picea abies*). *Polymers (Basel)*. 2021;13: 1106.
245. Bonnardel F, Perez S, Lisacek F, Imberty A. Structural database for lectins and the unilectin web platform. *Methods in Molecular Biology*. 2020;2132: 1–14.
246. Ingale AG, Hivrale AU. Plant as a plenteous reserve of lectin. *Plant Signaling and Behavior*. 2013;8: e26595.
247. Alderwick LJ, Harrison J, Lloyd GS, Birch HL. The mycobacterial cell wall – peptidoglycan and arabinogalactan. *Cold Spring Harbor Perspectives in Medicine*. 2015;5: a021113.
248. Endler A, Persson S. Cellulose synthases and synthesis in *Arabidopsis*. *Molecular Plant*. 2011;4: 199–211.
249. Francisco R, Brasil S, Poejo J, Jaeken J, Pascoal C, Videira PA, et al. Congenital disorders of glycosylation (CDG): state of the art in 2022. *Orphanet Journal of Rare Diseases*. 2023;18: 329.
250. Lefebvre DJ, Freeze HH, Steet R, Kinoshita T. Congenital disorders of glycosylation. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 599–614.
251. Berninsone P, Hwang HY, Zemtseva I, Horvitz HR, Hirschberg CB. SQV-7, a protein involved in *Caenorhabditis elegans* epithelial invagination and early embryogenesis, transports UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98: 3738–3743.
252. Koderi Valappil S, Shetty P, Deim Z, Terhes G, Urban E, Vaczi S, et al. Survival comes at a cost: a coevolution of phage and its host leads to phage resistance and antibiotic sensitivity of *Pseudomonas aeruginosa* multidrug resistant strains. *Frontiers in Microbiology*. 2021;12: 783722.

253. Koskella B, Brockhurst MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*. 2014;38: 916–931.
254. Rothschild-Rodriguez D, Hedges M, Kaplan M, Karav S, Nobrega FL. Phage-encoded carbohydrate-interacting proteins in the human gut. *Frontiers in Microbiology*. 2022;13: 1083208.
255. Heidelberger M, Avery OT. The soluble specific substance of pneumococcus. *The Journal of Experimental Medicine*. 1923;38: 73–79.
256. Avery OT, Goebel WF. Chemoimmunological studies on the soluble specific substance of pneumococcus: I. The isolation and properties of the acetyl polysaccharide of pneumococcus type I. *The Journal of Experimental Medicine*. 1933;58: 731–755.
257. Griffith F. The significance of pneumococcal types. *The Journal of Hygiene*. 1928;27: 113–159.
258. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology*. 1952;36: 39–56.
259. Cobb M. Oswald avery, DNA, and the transformation of biology. *Current Biology*. 2014;24: R55–R60.
260. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type Iii. *The Journal of Experimental Medicine*. 1944;79: 137–158.
261. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters*. 2016;363: fnw002.
262. Dunstan RA, Bamert RS, Tan KS, Imbulgoda U, Barlow CK, Taiaroa G, et al. Epitopes in the capsular polysaccharide and the porin OmpK36 receptors are required for bacteriophage infection of *Klebsiella pneumoniae*. *Cell Reports*. 2023;42: 112551.
263. Filik K, Szermer-Olearnik B, Oleksy S, Brykala J, Brzozowska E. Bacteriophage tail proteins as a tool for bacterial pathogen recognition—a literature review. *Antibiotics (Basel)*. 2022;11: 555.
264. Islam MZ, Fokine A, Mahalingam M, Zhang Z, Garcia-Doval C, van Raaij MJ, et al. Molecular anatomy of the receptor binding module of a bacteriophage long tail fiber. *PLoS Pathogens*. 2019;15: e1008193.
265. Wilson JH, Luftig RB, Wood WB. Interaction of bacteriophage T4 tail fiber components with a lipopolysaccharide fraction from *Escherichia coli*. *Journal of Molecular Biology*. 1970;51:423–434.
266. Borin JM, Lee JJ, Lucia-Sanz A, Gerbino KR, Weitz JS, Meyer JR. Rapid bacteria-phage coevolution drives the emergence of multiscale networks. *Science*. 2023;382: 674–678.
267. Bishop JR, Gagneux P. Evolution of carbohydrate antigens – microbial forces shaping host glycomes? *Glycobiology*. 2007;17: 23R–34R.
268. Sommer R, Makshakova ON, Wohlschlager T, Hutin S, Marsh M, Titz A, et al. Crystal structures of fungal tectonin in complex with O-methylated glycans suggest key role in innate immune defense. *Structure*. 2018;26: 391–402.e4.
269. Hilbert ZA, Haffener PE, Young HJ, Schwiesow MJ, W, Leffler EM, Elde NC. Rapid evolution of glycan recognition receptors reveals an axis of host-microbe arms races beyond canonical protein-protein interfaces. *Genome Biology and Evolution*. 2023;15: evad119.
270. Chang YC, Nizet V. Siglecs at the host-pathogen interface. *Advances in Experimental Medicine and Biology*. 2020;1204: 197–214.
271. Bunte MJM, Schots A, Kammenga JE, Wilbers RHP. Helminth glycans at the host-parasite interface and their potential for developing novel therapeutics. *Frontiers in Molecular Biosciences*. 2021;8: 807821.
272. Hokke CH, van Diepen A. Helminth glycomics – glycan repertoires and host-parasite interactions. *Molecular and Biochemical Parasitology*. 2017;215: 47–57.
273. Eckmair B, Gao C, Mehta AY, Dutkiewicz Z, Vanbeselaere J, Cummings RD, et al. Recognition of highly branched N-glycans of the porcine whipworm by the immune system. *Molecular and Cellular Proteomics*. 2024;23: 100711.
274. van Die I, Cummings RD. Glycan gimmickry by parasitic helminths: a strategy for modulating the host immune response? *Glycobiology*. 2010;20: 2–12.
275. Rappuoli R, De Gregorio E, Costantino P. On the mechanisms of conjugate vaccines. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116: 14–16.
276. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology*. 2012;10: 323–335.
277. Kudelka MR, Stowell SR, Cummings RD, Neish AS. Intestinal epithelial glycosylation in homeostasis and gut microbiota interactions in IBD. *Nature Reviews Gastroenterology and Hepatology*. 2020;17: 597–617.
278. Fekete E, Buret AG. The role of mucin O-glycans in microbiota dysbiosis, intestinal homeostasis, and host-pathogen interactions. *American Journal of Physiology Gastrointestinal and Liver Physiology*. 2023;324: G452–G465.
279. Kuziel GA, Rakoff-Nahoum S. The gut microbiome. *Current Biology*. 2022;32: R257–R264.
280. Couchman JR, Pataki CA. An introduction to proteoglycans and their localization. *The Journal of Histochemistry and Cytochemistry*. 2012;60: 885–897.
281. Mirouse V. Evolution and developmental functions of the dystrophin-associated protein complex: beyond the idea of a muscle-specific cell adhesion complex. *Frontiers in Cell and Developmental Biology*. 2023;11: 1182524.
282. Moller SR, Yi X, Velasquez SM, Gille S, Hansen PLM, Poulsen CP, et al. Identification and evolution of a plant cell wall specific glycoprotein glycosyl transferase, ExAD. *Scientific Reports*. 2017;7: 45341.

283. Leszczuk A, Kalaitzis P, Kulik J, Zdunek A. Review: structure and modifications of arabinogalactan proteins (AGPs). *BMC Plant Biology*. 2023;23:45.
284. Okuma H, Hord JM, Chandel I, Venzke D, Anderson ME, Walimbe AS, et al. N-terminal domain on dystroglycan enables LARGE1 to extend matriglycan on alpha-dystroglycan and prevents muscular dystrophy. *Elife*. 2023;12: e82811.
285. Koff M, Monagas-Valentin P, Novikov B, Chandel I, Panin V. Protein O-mannosylation: one sugar, several pathways, many functions. *Glycobiology*. 2023;33: 911–926.
286. Endo T. Glycobiology of alpha-dystroglycan and muscular dystrophy. *Journal of Biochemistry*. 2015;157: 1–12.
287. Dempsey CE, Bigotti MG, Adams JC, Brancaccio A. Analysis of alpha-dystroglycan/LG domain binding modes: investigating protein motifs that regulate the affinity of isolated LG domains. *Frontiers in Molecular Biosciences*. 2019;6: 18.
288. Jinno A, Park PW. Role of glycosaminoglycans in infectious disease. *Methods in Molecular Biology*. 2015;1229: 567–585.
289. Joseph S, Campbell KP. Lassa fever virus binds matriglycan—a polymer of alternating xylose and glucuronate-on alpha-dystroglycan. *Viruses*. 2021;13: 1679.
290. Rambukkana A, Yamada H, Zanazzi G, Mathus T, Salzer JL, Yurchenko PD, et al. Role of alpha-dystroglycan as a Schwann cell receptor for *Mycobacterium leprae*. *Science*. 1998;282: 2076–2079.
291. Maeda K, Okuda Y, Enomoto G, Watanabe S, Ikeuchi M. Biosynthesis of a sulfated exopolysaccharide, synechan, and bloom formation in the model cyanobacterium *Synechocystis* sp. strain PCC 6803. *Elife*. 2021;10: e66538.
292. Limoli DH, Jones CJ, Wozniak DJ. Bacterial extracellular polysaccharides in biofilm formation and function. *Microbiology Spectrum*. 2015;3.
293. Riesgo A, Farrar N, Windsor PJ, Giribet G, Leys SP. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Molecular Biology and Evolution*. 2014;31: 1102–1120.
294. Becerro MA, Uriz MJ, Maldonado M, Turon X. Advances in sponge science: phylogeny, systematics, ecology. Preface. *Advances in Marine Biology*. 2012;61: ix–x.
295. Ehrlich H, Wysokowski M, Zoltowska-Aksamitowska S, Petrenko I, Jesionowski T. Collagens of poriferan origin. *Marine Drugs*. 2018;16: 79.
296. Jesionowski T, Norman M, Zoltowska-Aksamitowska S, Petrenko I, Joseph Y, Ehrlich H. Marine spongin: naturally prefabricated 3d scaffold-based biomaterial. *Marine Drugs*. 2018;16: 88.
297. Bateman JF, Shoulders MD, Lamande SR. Collagen misfolding mutations: the contribution of the unfolded protein response to the molecular pathology. *Connective Tissue Research*. 2022;63: 210–227.
298. Katzman RL, Halford MH, Reinhold VN, Jeanloz RW. Isolation and structure determination of glucosylgalactosylhydroxylysine from sponge and sea anemone collagen. *Biochemistry*. 1972;11: 1161–1167.
299. Daubenspeck JM, Zeng H, Chen P, Dong S, Steichen CT, Krishna NR, et al. Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis* exosporium. *The Journal of Biological Chemistry*. 2004;279: 30945–30953.
300. Li RC, Wong MY, DiChiara AS, Hosseini AS, Shoulders MD. Collagen's enigmatic, highly conserved N-glycan has an essential proteostatic function. *Proceedings of the National Academy of Sciences of the United States of America*. 2021;118: e2026608118.
301. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *The ISME Journal*. 2013;7: 2287–2300.
302. Wilson HV. On some phenomena of coalescence and regeneration in sponges. *Journal of Experimental Zoology*. 1907;5: 245–258.
303. Ereskovsky A, Borisenko IE, Bolshakov FV, Lavrov AI. Whole-body regeneration in sponges: diversity, fine mechanisms, and future prospects. *Genes (Basel)*. 2021;12: 506.
304. Jarchow J, Burger MM. Species-specific association of the cell-aggregation molecule mediates recognition in marine sponges. *Cell Adhesion and Communication*. 1998;6: 405–414.
305. Vilanova E, Ciodaro PJ, Bezerra FF, Santos GRC, Valle-Delgado JJ, Anselmetti D, et al. Adhesion of freshwater sponge cells mediated by carbohydrate-carbohydrate interactions requires low environmental calcium. *Glycobiology*. 2020;30: 710–721.
306. Vilanova E, Coutinho CC, Mourao PA. Sulfated polysaccharides from marine sponges (Porifera): an ancestor cell-cell adhesion event based on the carbohydrate-carbohydrate interaction. *Glycobiology*. 2009;19: 860–867.
307. Vilanova E, Coutinho C, Maia G, Mourao PA. Sulfated polysaccharides from marine sponges: conspicuous distribution among different cell types and involvement on formation of in vitro cell aggregates. *Cell and Tissue Research*. 2010;340: 523–531.
308. Grice LF, Gauthier MEA, Roper KE, Fernandez-Busquets X, Degnan SM, Degnan BM. Origin and evolution of the sponge aggregation factor gene family. *Molecular Biology and Evolution*. 2017;34: 1083–1099.
309. Bucior I, Scheuring S, Engel A, Burger MM. Carbohydrate-carbohydrate interaction provides adhesion force and specificity for cellular recognition. *The Journal of Cell Biology*. 2004;165: 529–537.
310. Gagneux P, Henne T, Varki A. Biological functions of glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebersold M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 79–92.
311. Dennis JW, Granovsky M, Warren CE. Protein glycosylation in development and disease. *Bioessays*. 1999;21: 412–421.

312. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. *Nature Reviews Nephrology*. 2019;15: 346–366.
313. Colley KJ, Varki A, Haltiwanger RS, Kinoshita T. Cellular organization of glycosylation. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 43–52.
314. Wandall HH, Nielsen MAI, King-Smith S, de Haan N, Bagdonaite I. Global functions of O-glycosylation: promises and challenges in O-glycobiology. *The FEBS Journal*. 2021;288: 7183–7212.
315. Narimatsu Y, Joshi HJ, Nason R, Van Coillie J, Karlsson R, Sun L, et al. An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Molecular Cell*. 2019;75: 394–407.e5.
316. Cummings RD. Stuck on sugars – how carbohydrates regulate cell adhesion, recognition, and signaling. *Glycoconjugate Journal*. 2019;36: 241–257.
317. Terrapon N, Henrissat B, Aoki-Kinoshita KF, Suroli A, Stanley P. A genomic view of glycobiology. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 93–102.
318. An HJ, Froehlich JW, Lebrilla CB. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Current Opinion in Chemical Biology*. 2009;13: 421–426.
319. Bishop JR, Schuksz M, Esko JD. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature*. 2007;446: 1030–1037.
320. Cummings RD. The repertoire of glycan determinants in the human glycome. *Molecular Biosystems*. 2009;5: 1087–1104.
321. Galili U. Biosynthesis of alpha-Gal Epitopes (Galalpha1-3Gal-beta1-4GlcNAc-R) and their unique potential in future alpha-Gal therapies. *Frontiers in Molecular Biosciences*. 2021;8: 746883.
322. Altman MO, Gagneux P. Absence of Neu5Gc and presence of Anti-Neu5Gc antibodies in humans-an evolutionary perspective. *Frontiers in Immunology*. 2019;10: 789.
323. Durham SD, Wei Z, Lemay DG, Lange MC, Barile D. Creation of a milk oligosaccharide database, MilkOligoDB, reveals common structural motifs and extensive diversity across mammals. *Scientific Reports*. 2023;13: 10345.
324. Segurel L, Gao Z, Przeworski M. Ancestry runs deeper than blood: the evolutionary history of ABO points to cryptic variation of functional importance. *Bioessays*. 2013;35: 862–867.
325. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. *Annals of Human Biology*. 2013;40: 463–471.
326. Sender R, Fuchs S, Milo R. Are we really vastly outnumbered? revisiting the ratio of bacterial to host cells in humans. *Cell*. 2016;164: 337–340.
327. Roy AL, Conroy RS. Toward mapping the human body at a cellular resolution. *Molecular Biology of the Cell*. 2018;29: 1779–1785.
328. Marsico G, Russo L, Quondamatteo F, Pandit A. Glycosylation and integrin regulation in cancer. *Trends in Cancer*. 2018;4: 537–552.
329. Larsen ISB, Narimatsu Y, Joshi HJ, Siukstaite L, Harrison OJ, Brasch J, et al. Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114: 11163–11168.
330. Merleev AA, Park D, Xie Y, Kailemia MJ, Xu G, Ruhaak LR, et al. A site-specific map of the human plasma glycome and its age and gender-associated alterations. *Scientific Reports*. 2020;10: 17505.
331. Thomes L, Karlsson V, Lundstrom J, Bojar D. Mammalian milk glycomes: connecting the dots between evolutionary conservation and biosynthetic pathways. *Cell Reports*. 2023;42: 112710.
332. de Weerth C, Aatsinki AK, Azad MB, Bartol FF, Bode L, Collado MC, et al. Human milk: from complex tailored nutrition to bioactive impact on child cognition and behavior. *Critical Reviews in Food Science and Nutrition*. 2023;63: 7945–7982.
333. Turiak L, Sugar S, Acs A, Toth G, Gomory A, Telekes A, et al. Site-specific N-glycosylation of HeLa cell glycoproteins. *Scientific Reports*. 2019;9: 14822.
334. Uhler R, Popa-Wagner R, Kroning M, Brehm A, Rennert P, Seifried A, et al. Glyco-engineered HEK 293-F cell lines for the production of therapeutic glycoproteins with human N-glycosylation and improved pharmacokinetics. *Glycobiology*. 2021;31: 859–872.
335. Wang D, Kuzyk V, Madunic K, Zhang T, Mayboroda OA, Wuhler M, et al. In-depth analysis of the N-glycome of colorectal cancer cell lines. *International Journal of Molecular Sciences*. 2023;24: 4842.
336. Klaric TS, Gudelj I, Santpere G, Novokmet M, Vuckovic F, Ma S, et al. Human-specific features and developmental dynamics of the brain N-glycome. *Science Advances*. 2023;9: eadg2615.
337. Suttapitugsakul S, Stavenhagen K, Donskaya S, Bennett DA, Mealer RG, Seyfried NT, et al. Glycoproteomics landscape of asymptomatic and symptomatic human Alzheimer's disease brain. *Molecular and Cellular Proteomics*. 2022;21: 100433.
338. Lee J, Ha S, Kim M, Kim SW, Yun J, Ozcan S, et al. Spatial and temporal diversity of glycome expression in mammalian brain. *Proceedings of the National Academy of Sciences of the United States of America*. 2020;117: 28743–28753.
339. Williams SE, Noel M, Lehoux S, Cetinbas M, Xavier RJ, Sadreyev RI, et al. Mammalian brain glycoproteins exhibit diminished glycan complexity compared to other tissues. *Nature Communications*. 2022;13: 275.
340. Helm J, Hirtler L, Altmann F. Towards mapping of the human brain N-glycome with standardized graphitic carbon chromatography. *Biomolecules*. 2022;12: 85.
341. Xiao K, Han Y, Tian Z. Large-scale identification and visualization of human liver N-glycome enriched from LO2 cells. *Anal BioAnalytical Chemistry*. 2018;410: 4195–4202.

342. Kavanaugh D, O'Callaghan J, Kilcoyne M, Kane M, Joshi L, Hickey RM. The intestinal glycome and its modulation by diet and nutrition. *Nutrition Reviews*. 2015;73: 359–375.
343. Ashwood C, Pratt B, MacLean BX, Gundry RL, Packer NH. Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping. *The Analyst*. 2019;144: 3601–3612.
344. Riley NM, Bertozzi CR, Pitteri SJ. A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Molecular and Cellular Proteomics*. 2021;20: 100029.
345. Wesener DA, Dugan A, Kiessling LL. Recognition of microbial glycans by soluble human lectins. *Current Opinion in Structural Biology*. 2017;44: 168–178.
346. Liu FT, Stowell SR. The role of galectins in immunity and infection. *Nature Reviews Immunology*. 2023;23: 479–494.
347. Drickamer K, Taylor ME. Evolving views of protein glycosylation. *Trends in Biochemical Sciences*. 1998;23: 321–324.
348. Kristic J, Vuckovic F, Menni C, Klaric L, Keser T, Beccheli I, et al. Glycans are a novel biomarker of chronological and biological ages. *The Journals of Gerontology Series A, Biological Sciences and Medical Sciences*. 2014;69: 779–789.
349. Lado-Baleato O, Torre J, O'Flaherty R, Alonso-Sampedro M, Carballo I, Fernández-Merino C, et al. Age-Related changes in serum N-glycome in men and women—clusters associated with comorbidity. *Biomolecules*. 2024;14: 17.
350. Azad MB, Robertson B, Atakora F, Becker AB, Subbarao P, Moraes TJ, et al. Human milk oligosaccharide concentrations are associated with multiple fixed and modifiable maternal characteristics, environmental factors, and feeding practices. *The Journal of Nutrition*. 2018;148: 1733–1742.
351. Shi Y, Han B, Zhang L, Zhou P. Comprehensive identification and absolute quantification of milk oligosaccharides in different species. *Journal of Agricultural and Food Chemistry*. 2021;69: 15585–15597.
352. Moremen KW, Ramiah A, Stuart M, Steel J, Meng L, Forouhar F, et al. Expression system for structural and functional studies of human glycosylation enzymes. *Nature Chemical Biology*. 2018;14: 156–162.
353. Aoki-Kinoshita KF, Campbell MP, Lisacek F, Neelamegham S, York WS, Packer NH. Glycoinformatics. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. (eds.) *Essentials of glycobiology*, 4th Ed. Cold Spring Harbor (NY); 2022. p. 705–718.
354. Muthana SM, Campbell CT, Gildersleeve JC. Modifications of glycans: biological significance and therapeutic opportunities. *ACS Chemical Biology*. 2012;7: 31–43.
355. Sharapov SZ, Timoshchuk AN, Aulchenko YS. Genetic control of N-glycosylation of human blood plasma proteins. *Vavilovskii Zhurnal Genet Selektsii*. 2023;27: 224–239.
356. Jame-Chenarboo F, Ng HH, Macdonald D, Mahal LK. High-throughput analysis reveals miRNA upregulating alpha-2,6-sialic acid through direct miRNA-mRNA interactions. *American Chemical Society Central Science*. 2022;8: 1527–1536.
357. Thu CT, Mahal LK. Sweet control: microRNA regulation of the glycome. *Biochemistry*. 2020;59: 3098–3110.
358. Kawanishi K, Saha S, Diaz S, Vaill M, Sasmal A, Siddiqui SS, et al. Evolutionary conservation of human ketodeoxynonulosonic acid production is independent of sialoglycan biosynthesis. *The Journal of Clinical Investigation*. 2021;131: e137681.
359. Bashir S, Fezeu LK, Leviatan Ben-Arye S, Yehuda S, Reuven EM, Szabo de Edelenyi F, et al. Association between Neu5Gc carbohydrate and serum antibodies against it provides the molecular link to cancer: French NutriNet-Sante study. *BMC Medicine*. 2020;18: 262.
360. Diaz SL, Padler-Karavani V, Ghaderi D, Hurtado-Ziola N, Yu H, Chen X, et al. Sensitive and specific detection of the non-human sialic Acid N-glycolylneuraminic acid in human tissues and biotherapeutic products. *PLoS One*. 2009;4: e4241.
361. Schiller B, Hykollari A, Yan S, Paschinger K, Wilson IB. Complicated N-linked glycans in simple organisms. *Biological Chemistry*. 2012;393: 661–673.
362. Huang YF, Aoki K, Akase S, Ishihara M, Liu YS, Yang G, et al. Global mapping of glycosylation pathways in human-derived cells. *Developmental Cell*. 2021;56: 1195–1209.e7.
363. Joud M, Moller M, Olsson ML. Identification of human glycosyltransferase genes expressed in erythroid cells predicts potential carbohydrate blood group loci. *Scientific Reports*. 2018;8: 6040.
364. Jajosky RP, Wu SC, Zheng L, Jajosky AN, Jajosky PG, Josephson CD, et al. ABO blood group antigens and differential glycan expression: perspective on the evolution of common human enzyme deficiencies. *IScience*. 2023;26: 105798.
365. Arthur CM, Stowell SR. The development and consequences of red blood cell alloimmunization. *Annual Review of Pathology*. 2023;18: 537–564.
366. Stowell SR, Arthur CM, Dias-Baruffi M, Rodrigues LC, Gouridine JP, Heimburg-Molinari J, et al. Innate immune lectins kill bacteria expressing blood group antigen. *Nature Medicine*. 2010;16: 295–301.
367. Abegaz SB. Human ABO blood groups and their associations with different diseases. *BioMed Research International*. 2021;2021: 6629060.
368. Bua RO, Messina A, Sturiale L, Barone R, Garozzo D, Palmigiano A. N-glycomics of human erythrocytes. *International Journal of Molecular Sciences*. 2021;22: 8063.
369. Svensson L, Hult AK, Stamps R, Angstrom J, Teneberg S, Storry JR, et al. Forssman expression on human erythrocytes: biochemical and genetic evidence of a new histo-blood group system. *Blood*. 2013;121: 1459–1468.
370. De Leoz MLA, Duewer DL, Fung A, Liu L, Yau HK, Potter O, et al. NIST interlaboratory study on glycosylation analysis of monoclonal antibodies: comparison of results from diverse ana-

- lytical methods. *Molecular and Cellular Proteomics*. 2020;19: 11–30.
371. Otaki M, Hirane N, Natsume-Kitatani Y, Nogami Itoh M, Shindo M, Kurebayashi Y, et al. Mouse tissue glycome atlas 2022 highlights inter-organ variation in major N-glycan profiles. *Scientific Reports*. 2022;12: 17804.
372. Suttapitugsakul S, Matsumoto Y, Aryal RP, Cummings RD. Large-scale and site-specific mapping of the murine brain O-glycoproteome with IMPa. *Analytical Chemistry*. 2023;95: 13423–13430.
373. Khidekel N, Ficarro SB, Clark PM, Bryan MC, Swaney DL, Rexach JE, et al. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nature Chemical Biology*. 2007;3: 339–348.
374. Noel M, Cummings RD, Mealer RG. N-glycans show distinct spatial distribution in mouse brain. *Glycobiology*. 2023;33: 935–942.
375. Zaytseva OO, Seeling M, Kristic J, Lauc G, Pezer M, Nimmerjahn F. Fc-linked IgG N-glycosylation in FcγR knock-out mice. *Frontiers in Cell and Developmental Biology*. 2020;8: 67.
376. Durin Z, Houdou M, Legrand D, Foulquier F. Metalloglycobiology: the power of metals in regulating glycosylation. *Biochimica et Biophysica Acta Gen Subj*. 2023;1867: 130412.