

Review

Glycomics: revealing the dynamic ecology and evolution of sugar molecules

Stevan A. Springer^{a,b,*}, Pascal Gagneux^{a,c,**}^a Glycobiology Research and Training Center, University of California San Diego, La Jolla, CA 92039, USA^b Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92039, USA^c Department of Pathology, University of California San Diego, La Jolla, CA 92039, USA

ARTICLE INFO

Article history:

Received 5 September 2015

Received in revised form 11 November 2015

Accepted 23 November 2015

Available online 25 November 2015

Keywords:

Glycans
Glycoconjugates
Natural variation
Polymorphism
Cellular ecology

ABSTRACT

Sugars are the most functionally and structurally diverse molecules in the biological world. Glycan structures range from tiny single monosaccharide units to giant chains thousands of units long. Some glycans are branched, their monosaccharides linked together in many different combinations and orientations. Some exist as solitary molecules; others are conjugated to proteins and lipids and alter their collective functional properties. In addition to structural and storage roles, glycan molecules participate in and actively regulate physiological and developmental processes. Glycans also mediate cellular interactions within and between individuals. Their roles in ecology and evolution are pivotal, but not well studied because glycan biochemistry requires different methods than standard molecular biology practice. The properties of glycans are in some ways convenient, and in others challenging. Glycans vary on organismal timescales, and in direct response to physiological and ecological conditions. Their mature structures are physical records of both genetic and environmental influences during maturation. We describe the scope of natural glycan variation and discuss how studying glycans will allow researchers to further integrate the fields of ecology and evolution.

© 2015 Elsevier B.V. All rights reserved.

Contents

1.	The complex and labile forms of glycans	91
1.1.	No template required	91
1.2.	Synthesis	91
1.3.	Privacy	91
1.4.	Dynamics	91
1.5.	Glycomics	91
2.	Cellular ecology	91
2.1.	Parts	93
2.1.1.	Prokaryotes	93
2.1.2.	Eukaryotes	93
2.2.	Interactions	93
2.2.1.	Within individuals	93
2.2.2.	Between individuals	95
3.	Ecology and evolution: tools to understand glycans	95
3.1.	Quantifying glycan variation	95
3.2.	Predicting glycan variation	96
4.	Glycans: tools to understand ecology and evolution	96
4.1.	Sources of glycan variation	96
4.2.	Interpreting glycan variation	97
5.	Glycan phenotypes and the synthesis of ecology and evolution	97
	Conflict of interest	98
	Acknowledgments	98
	References	98

* Correspondence to: S.A. Springer, Glycobiology Research and Training Center, University of California San Diego, La Jolla, CA 92039, USA.

** Correspondence to: P. Gagneux, Department of Pathology, University of California San Diego, La Jolla, CA 92039, USA.

E-mail addresses: stevan.springer@gmail.com (S.A. Springer), pgagneux@ucsd.edu (P. Gagneux).

1. The complex and labile forms of glycans

1.1. No template required

DNA is an inert strand whose information encodes ribbons of RNA and protein that put instructions into action. For biologists, this one-to-one correspondence is convenient. It ties contemporary molecules to their history, allowing us to pinpoint the origins of molecular functions and trace their modifications through time. DNA is a template, transcribed and translated by a common set of mechanisms into just two different kinds of molecule. Because of this commonality, molecular biologists tend to use a standard set of tools to determine how information in DNA encodes the biological functions of RNA and proteins. Now that these tools exist, there is a tendency to regard template-based molecules as the most important sources of heritable variation, and other molecules as merely structural. But heritable information in the genome is also manifest at other levels; these are less bound to the barely mutable instructions in DNA, and as a consequence more sensitive to current conditions and more capable of dynamic response. A well-understood example is RNA and protein expression, which can change in response to environment and physiological condition. Less well studied are glycans — whose structures and biochemical properties have evolved to enact traits with dynamic ecology and evolution [1].

1.2. Synthesis

Glycans are sugar molecules — modular arrangements of monosaccharide units that are often attached to proteins and lipids. Glycans range in size from mono- and oligosaccharides a few nanometers long, to polysaccharide chains thousands of subunits and several micrometers in length. The monosaccharide constituents of glycans are linked together by glycosidic bonds. These can originate from several of the hydroxylated carbon atoms that form each monosaccharide ring, so glycan structures can be extensively branched. Glycans decorate the majority of extracellular proteins and many lipids (the combined molecules are called glycoconjugates) [2]. These exist primarily on the outer surface of the cell membrane, and are also secreted into extracellular spaces (Box 1). Unlike protein and RNA, glycans are not built from a template. Their synthesis is an assembly process of sequential addition, removal, and modification of sugar subunits. The form and diversity of glycan structures are guided by proteins, but also depend intrinsically on diet and physiological condition. Because the structures of mature glycans are not encoded in the genome, their evolutionary history can only be uncovered indirectly. For example, the evolution of glycan modifying proteins can reveal historical glycan changes. The gain and loss of sugar transporters, transferases, and glycosidases have a direct bearing on the suite of glycans an organism can synthesize and degrade, or use in the diet.

1.3. Privacy

Glycans are the most variable class of biological molecules, both in structural diversity and taxonomic breadth of particular molecules (Box 1) [3]. Glycans often exist at the interface of conflicts between parties with different evolutionary interests: parents and offspring, hosts and pathogens, cooperators and cheats [1]. Glycans that signal or conceal self-identity, or mediate evolutionary conflicts can change rapidly and evolve toward esoteric structures and biochemistries. Glycans also form protective structures and accumulate as stores of energy. Some storage glycans, such as glycogen and starch, can be exploited by competing organisms and must be kept sequestered inside cells. Others, like mammalian lactose, are secreted. Lactose is composed of just two simple monosaccharides, but the linkage connecting them is difficult to synthesize and equally difficult to degrade so the combined molecule can be transferred to offspring with little risk of exploitation. Structural glycans including cellulose, hemicellulose, and chitin are the most abundant biopolymers on earth. These linked monosaccharides can

accumulate to such extraordinary abundances only because they are difficult to break down for energy. The space of possible glycan structures and linkages is so large that some glycans are private, in the sense that few organisms can synthesize, degrade, or bind them. Thus, the diversity of glycans allows them to be used in contexts where less diverse molecules would be vulnerable to exploitation.

1.4. Dynamics

Glycans have a remarkable degree of spatial and temporal organization, patterns that can enact changes in development and physiology. The dynamic properties of glycans influence aspects of trait development, cellular behavior, and molecular function. For example, N-linked glycans attract chaperone proteins, guiding nascent glycoproteins into their mature fold as they are translated in the endoplasmic reticulum. Later, in the golgi, N-glycan molecules are themselves remodeled into forms that alter the function of the folded glycoprotein. During synthesis, glycans can be modified by elongation, shedding, cleaving of terminal monosaccharides, and by modifications such as: N- or O-acetylation [4], deacetylation, epimerization, or N- and O sulfation [5]. Glycan patterns change with cell type and cell stage, and influence cell migration, tissue modulation, repair and regeneration [6]. Glycans also change with cell age. The Ashwell receptor monitors deglycosylation as a sign that a cell or plasma glycoprotein needs recycling [7]. Glycans exist in organisms as clouds of variants whose properties integrate evolutionary and ecological information. The proteins that interact with glycans are robust to this ever-present variation, but can also evolve binding patterns that are sensitive to glycan changes that indicate development, diet, health, or environment.

1.5. Glycomics

Glycan distribution and pattern on cells and tissues was first probed with lectins — glycan binding proteins with well-defined specificities. An important challenge is to scale these studies up, to bridge the gap between lists of glycans identified by mass spectrometry, and their true biological abundance, topology, distribution and variability. New glycomics methods will allow us to look directly at the diversity and variability of glycans: their structures, locations, and their biochemistry (Box 2). Glycomics is revealing a world of new biochemical forms and functions: molecular mechanisms that underpin ecological and evolutionary processes at every spatial and temporal scale. The study of natural glycan variation is one of the most interesting new prospects in biology today.

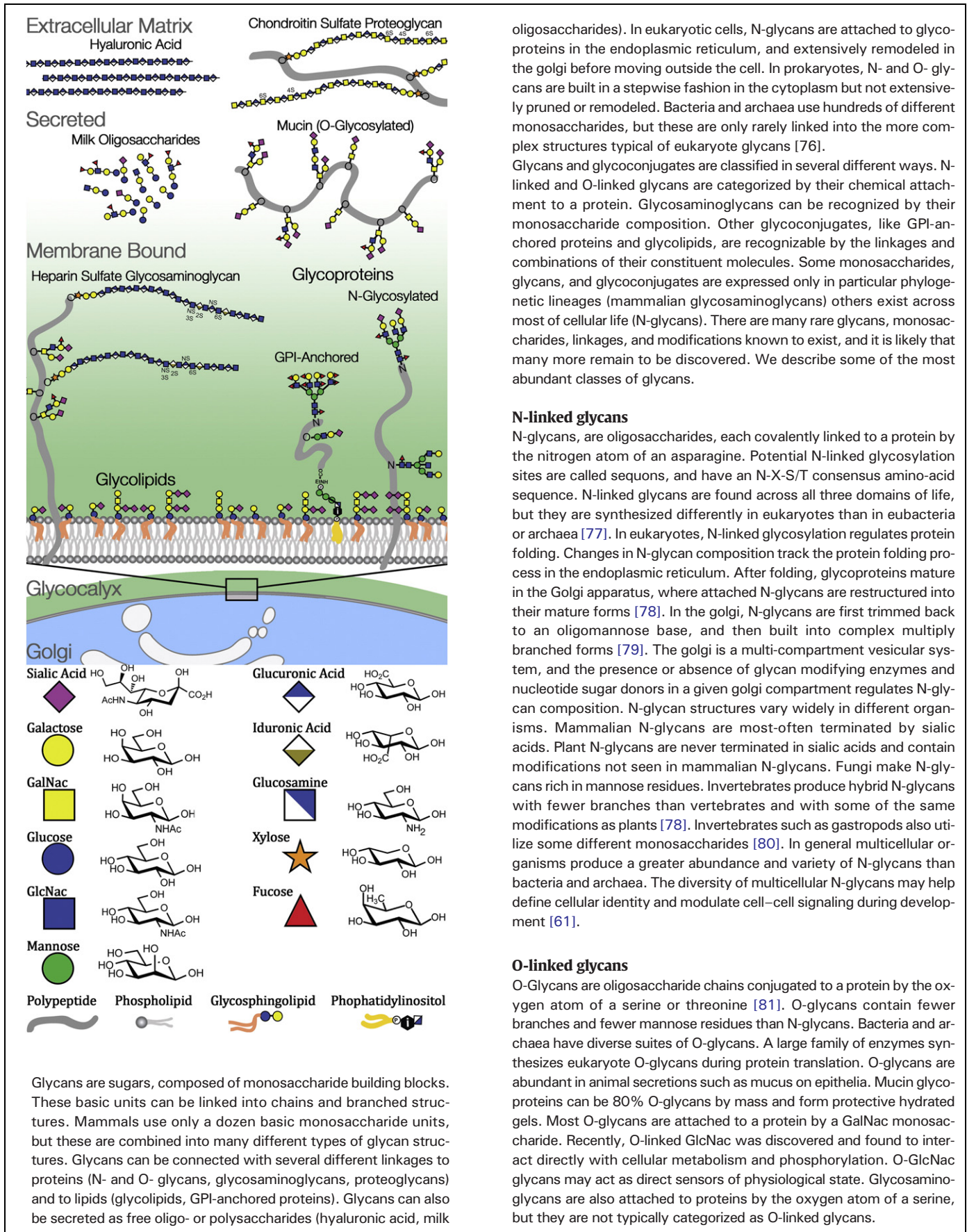
2. Cellular ecology

Classically, ecologists have studied the interactions of organisms with each other and with their environment. But ecological interactions are not limited to just one spatial scale. Cells also have ecology, with interactions as nuanced and interesting as those at larger scales (Fig. 1). But there is one very important difference between community and cellular ecology — the mechanistic complexity of the traits involved. Many of the interesting phenomena in community ecology result from patterns of organismal response, which can be developmental, physiological, or in animals neural. The responses are organismal, so the mechanisms are often too complex to be traced down to a particular molecular cause. The behavioral responses of cellular ecology are equally sophisticated in their scope, but vastly simpler mechanistically. Recognizable ecological interactions can be caused by very simple, tractable, chemical interactions [8].

The distinction is not important for a pure ecologist, but for an evolutionary ecologist mechanistically simple systems are key because their history can be more easily traced. Glycans are ecological traits. As a consequence their structures are condition dependent and their historical variation is not as easy to trace as a template based molecule.

Box 1

Types of glycans and glycoconjugates.



oligosaccharides). In eukaryotic cells, N-glycans are attached to glycoproteins in the endoplasmic reticulum, and extensively remodeled in the golgi before moving outside the cell. In prokaryotes, N- and O-glycans are built in a stepwise fashion in the cytoplasm but not extensively pruned or remodeled. Bacteria and archaea use hundreds of different monosaccharides, but these are only rarely linked into the more complex structures typical of eukaryote glycans [76].

Glycans and glycoconjugates are classified in several different ways. N-linked and O-linked glycans are categorized by their chemical attachment to a protein. Glycosaminoglycans can be recognized by their monosaccharide composition. Other glycoconjugates, like GPI-anchored proteins and glycolipids, are recognizable by the linkages and combinations of their constituent molecules. Some monosaccharides, glycans, and glycoconjugates are expressed only in particular phylogenetic lineages (mammalian glycosaminoglycans) others exist across most of cellular life (N-glycans). There are many rare glycans, monosaccharides, linkages, and modifications known to exist, and it is likely that many more remain to be discovered. We describe some of the most abundant classes of glycans.

N-linked glycans

N-glycans, are oligosaccharides, each covalently linked to a protein by the nitrogen atom of an asparagine. Potential N-linked glycosylation sites are called sequons, and have an N-X-S/T consensus amino-acid sequence. N-linked glycans are found across all three domains of life, but they are synthesized differently in eukaryotes than in eubacteria or archaea [77]. In eukaryotes, N-linked glycosylation regulates protein folding. Changes in N-glycan composition track the protein folding process in the endoplasmic reticulum. After folding, glycoproteins mature in the Golgi apparatus, where attached N-glycans are restructured into their mature forms [78]. In the golgi, N-glycans are first trimmed back to an oligomannose base, and then built into complex multiply branched forms [79]. The golgi is a multi-compartment vesicular system, and the presence or absence of glycan modifying enzymes and nucleotide sugar donors in a given golgi compartment regulates N-glycan composition. N-glycan structures vary widely in different organisms. Mammalian N-glycans are most-often terminated by sialic acids and contain modifications not seen in mammalian N-glycans. Fungi make N-glycans rich in mannose residues. Invertebrates produce hybrid N-glycans with fewer branches than vertebrates and with some of the same modifications as plants [78]. Invertebrates such as gastropods also utilize some different monosaccharides [80]. In general multicellular organisms produce a greater abundance and variety of N-glycans than bacteria and archaea. The diversity of multicellular N-glycans may help define cellular identity and modulate cell–cell signaling during development [61].

O-linked glycans

O-Glycans are oligosaccharide chains conjugated to a protein by the oxygen atom of a serine or threonine [81]. O-glycans contain fewer branches and fewer mannose residues than N-glycans. Bacteria and archaea have diverse suites of O-glycans. A large family of enzymes synthesizes eukaryote O-glycans during protein translation. O-glycans are abundant in animal secretions such as mucus on epithelia. Mucin glycoproteins can be 80% O-glycans by mass and form protective hydrated gels. Most O-glycans are attached to a protein by a GalNAc monosaccharide. Recently, O-linked GlcNAc was discovered and found to interact directly with cellular metabolism and phosphorylation. O-GlcNAc glycans may act as direct sensors of physiological state. Glycosaminoglycans are also attached to proteins by the oxygen atom of a serine, but they are not typically categorized as O-linked glycans.

Glycosaminoglycans

Glycosaminoglycans (GAGs) are among the largest molecules synthesized by animals. Their synthesis begins with an O-linked xylose, which is then elongated by two galactose molecules. Many thousands of N-Acetylglucosamine and glucuronic acid disaccharides are added to this initial trisaccharide to form gigantic chains. Importantly, GAGs, undergo further modification: glucuronic acid can be epimerized into iduronic acid, and N-Acetylglucosamines can be de-N-acetylated and N-sulfated, N-acetylated or O-sulfated. These post-translational modifications generate complex chemical patterns along the length of each GAG chain. We do not yet know how to read the syntax, but it is clear that these patterns have important signaling functions. They act as binding sites for growth factors (FGFs), anti-coagulation factors (anti-thrombin), and pathogens (HIV and HSV1) [11]. GAGs are involved in a diversity of physiological and developmental processes. They interface with metabolic state *via* lipid particles on liver cell proteoglycans [82]. They also guide morphological development [83], and even potentially behavior *via* their effects on neuronal migration [84].

Glycolipid glycoconjugates

Glycolipids are oligosaccharides conjugated to lipid molecules. Gangliosides, are sialic acid containing glycosphingolipids, which are often form patterns associated with lipid rafts on the outer cell membrane. Certain cell types can have outer membrane leaflets that consist almost entirely of glycolipid instead of phospholipid [85]. The spatial orientation of glycans is important, because their binding is often multivalent, involving multiple chemical contacts in close proximity. The clustering of glycans on the cell surface by glycolipids may thus have important functional effects.

GPI-anchored glycoconjugates

GPI-anchored glycoproteins are complex molecules that combine a lipid, a glycan, and a protein. The lipid is a phosphoinositol tail, which is linked to a short oligosaccharide glycan, itself linked to a polypeptide by a phosphoethanolamine bond. The polypeptide components of most GPI-anchored proteins are further glycosylated and can contain N- and O-glycans and even glycosaminoglycans as in glypicans. Interestingly, GPI anchored proteins can transfer information from one cell to another. They can be shed from one cell surface and incorporated into the membranes of other cells [86].

Rare glycans and glycoconjugates

There are many other glycan classes that are less abundant than the classes described above. Some have only a few known examples; others are expressed only in particular phylogenetic lineages. A huge unexplored diversity of glycans is likely to exist in bacteria and archaea. Because they lack an ER-Golgi compartment, these taxa have evolved different ways of assembling glycans than eukaryotes. The capsids and glycoconjugates of bacteria and archaea are already known to contain a bewildering diversity of complex glycans and the study of these is in its infancy [76,87]. Recent years have seen the discovery of many additional glycan classes including: O-xylose, O-fucose, C-Mannose, and S-linked glycans on cysteine residues [88] New classes of glycans, and new glycan functions are continually being discovered [89].

But glycans are a middle ground – their evolutionary history can be inferred from the evolution of their synthetic enzymes. Thus, the application of phylogenetic tools to glycans with recognizable cellular ecology is a bridge that links the fields of ecology and evolution. Figs. 2, 3, and 4 present case studies from primates. These examples were developed as extensions of human medical studies, but there will be countless other examples. Cellular ecology pervades biology. Glycomic methods

will allow researchers to enumerate the parts of each system, and from there cellular ecologists can begin to unravel the interactions.

2.1. Parts

2.1.1. Prokaryotes

All living cells are coated by a glycocalyx, a complex layer of protein and lipid glycoconjugates with protective and signaling functions (Box 1). Prokaryote glycans act as defensive barriers, provide support, store resources, and can also act as signaling molecules [9] [10]. Eubacteria and Archaea use over 700 different monosaccharides to produce their bacterial capsid polysaccharides and archaeal cell envelopes. Unicellular organisms use the glycocalyx to shield themselves from environmental influence, and these extracellular glycoconjugates also mediate cellular interactions such as niche construction and the formation of biofilms.

2.1.2. Eukaryotes

In Metazoans the glycocalyx must additionally hold information about cell lineage identity, and mediate and cell–cell communication during development and cell migration. Animal glycans are built from only about a dozen different monosaccharides, in many different linkages. The extracellular matrix (ECM) of multicellular animals is rich in glycosaminoglycans, long secreted polymers such as hyaluronan, or proteoglycans such as heparan sulfate, chondroitin sulfate, and dermatan sulfate [11]. In plants, cell wall polysaccharides are key structural molecules [12]. Plants make complex glycans consisting of many more individual monosaccharide types than found in animals.

2.2. Interactions

2.2.1. Within individuals

Glycans act at all life history stages, their composition and patterns on cell surfaces changing during the course of development. Sperm-egg interactions of invertebrates and vertebrates require glycan mediated binding [13–15]. Developing tissues can be identified by Stage Specific Embryonic Antigens (SSEAs), some of which are glycans [16]. Different surfaces of the same cell can even bear different glycans. For example, heparan sulfate proteoglycans give polarity to fly germ line cells [17,18]. Glycans pattern the extracellular matrix. Their structural and informational roles allow signaling gradients to form: a critical requirement of proper development [19,20].

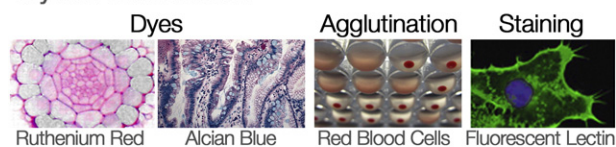
Physiological and dietary changes also influence glycan synthesis [21]. The abundant O-GlcNAc modification (of transcription factors, kinases, histones and other proteins) directly connects nutrient levels with glycosylation, gene regulation, and epigenetic marks on histone tails [22]. Glycan receptors, such as siglecs and other lectins, are expressed on immune cells. These bind self-glycan patterns and regulate the innate immune response [23]. Other glycan receptors induce inflammatory responses when they bind fragments of large structural glycans like cellulose or hyaluronan: these fragments presumably indicate a lysed cell, a pathogen intrusion, or some other danger [24,25].

Glycans are often used in more than one context within an organism. The asialoglycoprotein (Ashwell) receptor binds aging platelets, RBCs and glycoproteins so they can be removed from circulation [26] [7]. This same receptor mitigates bacterial infection [27]. Polysialic acids mediate neuronal development, and are also expressed by all three germ layer precursors [28,29]. Galectins form intercellular lattices by cross-linking N-glycans terminating in galactose [30], and also aid in cytoplasmic pathogen defense [31]. Selectins on leukocytes, platelets and endothelia mediate both leukocyte trafficking [32] and blastocyst implantation [33]. Within individuals, glycans take on a diversity of roles, and the synthesis and regulation of individual glycan molecules is tailored to suit each particular circumstance.

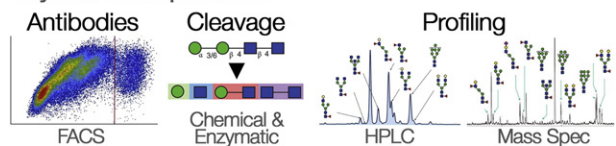
Box 2

Methods of glycan discovery.

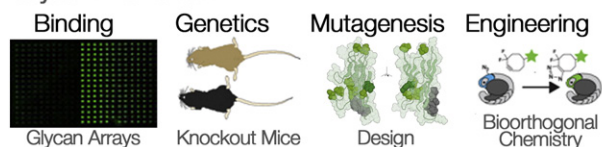
Glycan Distribution



Glycan Composition



Glycan Function

**Glycan distribution**

Dyes. Cationic dyes stain the glycocalyx and extracellular matrix on epithelial cells because these structures have a negative charge overall. Several cationic dyes have been traditionally used to stain anionic glycans — such as Alcian Blue for mucins in epithelial secretions or hyaluronic acid in cartilage, and Ruthenium Red to stain the glycocalyx [90]. Cationic dyes do stain glycans, but they are generally not specific to particular molecules, so they don't provide the resolution required for detailed analysis.

Agglutination. Blood cells often clump together when exposed to lectins. Plant lectins differentially agglutinate the red blood cells from different individuals and from different mammalian species. This is caused by the specificity of a multivalent lectin for polymorphic glycans on the red blood cell surface. The understanding that glycans caused blood cell agglutination allowed the development of lectin histology — the characterization of specific glycan patterns on the cell surface.

Staining. There is now a large collection of natural lectins from plants, fungi and invertebrates with a variety of binding specificities for different glycans [91]. Animal cells or tissue sections can be stained with lectins to reveal the abundance and distributions of glycans. Lectins with a particular specificity can be attached to fluorescent markers to reveal cellular structures and patterns of glycan distribution on cells. The use of glycan modifying enzymes to directly remodel cell surface glycans allows the target of a particular lectin stain to be verified. Lectin staining even permits simple analyses of the composition of glycans on the cell surface.

Glycan composition

Antibodies. The presence or absence of intact glycans on glycoconjugates can be assessed with antibodies. For glycoproteins, antibodies can be used to specifically stain Western blots to identify proteins bearing particular glycans. Fluorescence activated cell sorting (FACS) can be combined with antibody or lectin staining to build sophisticated analyses of the presence and composition of glycans on cell surfaces. Anti-glycan antibodies were added to the tool kit of molecular probes for glycans and so were recombinant glycan binding proteins expressed in microbial, plant or animal cells [92].

Cleavage. A variety of chemical, biochemical and biophysical tools have been developed to characterize the structures of glycans: their monosaccharide composition, glycosidic linkage types and branching patterns [93]. Compounds such as 4-methyl umbelliferyl (4-MU), which fluoresce once a monosaccharide has been cleaved, can monitor the presence or specificity of degradation enzymes. Chemical and enzymatic methods sequentially degrade glycans or release them from their glycoconjugates. The structures of free glycans can then be inferred by chromatography and mass spectrometry.

Profiling. Glycans released from the underlying protein or lipid and are then chemically derivatized prior to separation by chromatography. Glycans are identified on an HPLC column by comparing their characteristic retention with standards. Chromatography can be coupled with mass spectrometry such as MALDI TOF (matrix assisted laser desorption/ionization time of flight) mass spectrometry to identify the individual glycan species. Even intact glycan structures, can be identified by the characteristic masses of their fragments [94].

Glycan function

Binding. Glycan arrays can assess binding specificity by presenting many glycans each with a slightly different structure. Lectin arrays can be coupled with surface plasmon resonance to directly measure the force of interaction between glycan binding proteins and glycans. In rare cases, lectins can be co-crystallized with their glycan ligand to characterize the glycan-protein interaction at the atomic level [95]. Glycan arrays require careful interpretation since binding on an array may not perfectly reflect natural binding. Recently synthetic glycans arranged in multivalent assemblages on scaffold molecules have allowed researchers to probe for glycan binding proteins in more naturalistic circumstances [96].

Genetics. Genomic comparisons of genes encoding glycan synthesis, modification, or degradation enzymes can indicate historical glycan presence and absence. Phylogenetic comparisons of particular glycan binding proteins can test whether their evolution is driven by selection. Genetic tools have uncovered the ER and Golgi enzymes that act during glycan synthesis and remodeling [97]. Genetic manipulation in model organisms can give clues to natural glycan functions *in vivo*. Findings from medicine inform us about the role of glycans in disease [21]. For example, congenital diseases of glycosylation can have extraordinarily simple genetic causes, and equally simple solutions. Many are caused by the loss of a single glycan-modifying enzyme, and some can be completely reversed by supplementing the missing glycan in the diet [98].

Mutagenesis. Artificial mutations of glycan binding proteins can be used in conjunction with glycan arrays to determine which amino-acid residues promote glycan specificity, or to design new specificities. Cloning animal glycan binding proteins and their recombinant expression provides additional molecular probes for glycans. It also allows functional studies of lectin-glycan interactions in animals [99].

Engineering. Bio-orthogonal compounds are molecules that cannot be synthesized or broken down by organisms. Bioorthogonal glycans combined with click chemistry artificial molecules can be used to image changes in the distribution and metabolism of natural glycans in real time [100]. Synthesis of glycans using non-natural monosaccharides also allows metabolic labeling and functional manipulation *in vivo*.

Challenges

Glycans are chemically diverse, so that even superficially similar classes of molecules like O- and N-glycans cannot be fully analyzed by any single set of methods. For some glycans, like GAGs, even their basic physical structures cannot yet be fully resolved.

Decoding the patterns of chemical modification that occur in glycans will be difficult, and determining the function of these patterns more difficult still. The study of glycan function relies heavily on decomposition analysis and genetic manipulation of their synthesis: both tools could benefit from refinement. For example, many different glycans can have identical signatures by mass spectrometry, and it is often difficult to know which branch of a complex N-glycan contained which monosaccharide. Not all of the approaches used to characterize glycans are yet amenable to high throughput methods. Additionally, the involvement of glycans in many biological processes can be missed because their mode of chemical interaction is different from proteins or antibodies. Glycan interactions, with each other and with the proteins that bind them, are typically multivalent — it requires several molecular contacts in close physical proximity to form a stable attachment [101]. This complicates the assessment of their function, because the molecular orientations that allow binding in natural conditions cannot always be replicated in an artificial assay. There is a critical need for binding assays that take into account multivalency, and the spatial orientations of glycans and their binding partners.

Finally, the functional analysis of glycans is potentially more complicated than that of proteins or RNA, there are many potential modifications, and it is not always possible to make them independently at each site in the molecule. It is also difficult to know which changes are realistic, in the sense that they can be achieved by mutations of a glycan synthesis protein. Genetic studies, and analysis of glycan variants produced by mutation in model organisms will yield important insights into glycan function and variation *in vivo* [102]. Ecological glycan functions may be restricted to certain interactions, phases of development, or to specific parts of an organism. Studying these traits will require new chemical methods of uncovering glycan variation, new statistical methods of describing it, and new theory to describe the patterns of glycan variation expected in natural circumstances.

Standards

The structural diversity of glycans and glycoconjugates has been an impediment to their systematic analysis. This is an underdeveloped area of glycobiology that deserves research effort. The first step is simply cataloging the variation of glycans, glycoconjugates and glycan-modifying enzymes as it exists in nature. There is a rapidly growing repository of glycan active enzymes: CAZy (<http://www.cazy.org/>). There are also databases that attempt to record the diversity of glycan structures themselves: (<http://www.glycome-db.org/>), (<http://expasy.org/resources/search/keywords:glycomics>), (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>). Resources for glycobiology methods are available from several organizations (<http://glycosciences.de/>), (<http://www.functionalglycomics.org/static/consortium/resources.shtml>). The open glycoworkbench is a resource for interpreting and annotating mass spectrometry data (<http://www.glycoworkbench.org/>). Glycopedia is a general reference for glycoscience related information (<http://www.glycopedia.eu/>). Glycobiologists should aim to develop standards of analysis that allow comparison between studies, to the extent that this will be possible for such diverse molecules. Journals and funding agencies should mandate the deposition of glycan data into public databases, where it can be accessed and subject to synthetic analysis.

2.2.2. Between individuals

Glycans are also a nexus for interactions between individuals: in symbiosis and in pathogenesis. Animals recruit microbes to expand the range of glycans available to them in the diet. Gut microbe genomes are typically enriched with polysaccharide utilizing loci that allow symbionts to digest glycans otherwise unavailable to the host [34]. Animals do not have

cellulose-digesting enzymes, but many (termites, ruminants, koalas, leaf eating monkeys) have evolved symbioses with microbes that can digest cellulose for energy [35]. In mammals [36], the establishment of a specific gut microbiota (and with it proper immune maturation) is aided by oligosaccharides secreted in mother's milk [37]. Some symbiotic microbes evade host immunity by molecular mimicry: camouflaging their surfaces by synthesizing glycans that resemble the host [38]. Other symbionts incorporate glycans synthesized by the host to decorate their surfaces and increase host tolerance [39]. Oligosaccharides such as polysaccharide A, secreted by the mammalian gut symbiont *Bacteroides fragilis*, can directly induce tolerance by interacting with T regulatory cells in the gut [40].

The potential for glycans to induce immune tolerance can also be exploited, and this conflict results in evolutionary arms races between hosts and their pathogens. Some pathogenic bacteria attach to host cells using adhesins with high specificity for host glycans (*E.coli*, *P. falciparum*). Other pathogens secrete toxins that target host glycans (*S. Typhi* and *Clostridium tetani*), or modify host glycans into appropriate ligands for binding (*V. cholerae* sialidase) [41]. Viruses also use host glycans for attachment (Paramyxoviruses for example) [3,42]. In response, hosts have evolved lectins: including high mannose receptor, DC-SIGN, dectin and intelectin-1, which are expressed on a variety of immune cells. These innate immune receptors allow plants and animals to rapidly detect foreign antigens [24], and some innate lectins like RegIII even have direct bactericidal properties [43].

Hosts also use glycans to defend themselves against pathogens. Defensive mucins are secreted on most animal epithelia. These heavily O-glycosylated glycoproteins resist proteolytic degradation and form hydrated gels, forming a physical barrier against incoming microbes [44]. Mucins can also act as decoys, mimicking the structures of cell surface receptors. Many viruses express sialidases (neuraminidases) so that they can bypass sialic acid rich mucin decoys. Bacteria are also infected by viruses, and bacteriophages are known to target glycans on the surface of their bacterial hosts [45]. Most viruses do not synthesize their own glycans, but viral proteins can be glycosylated by the host cell, allowing them to evade host immunity (HIV1 GP120) [46] and/or inhibition (Filovirus mucin-like glycoprotein) [47]. Parasites with complex host cycles are glyco-engineers, negotiating and manipulating the different glycan environments in each of their vertebrate, insect and mollusk hosts [48]. Glycans mediate aspects of ecology that cause them to evolve rapidly. The tools developed by ecologists and evolutionary biologists to study diverse and rapidly evolving traits will help us better understand the extraordinary diversity of glycans.

3. Ecology and evolution: tools to understand glycans

Ecology and evolutionary biology study the distribution of biological variation in space and in time. Each field has developed statistical methods for quantifying variation, and each uses null models to predict the patterns of variation expected under specific processes [49–51]. We are only now gaining the technology to observe most natural glycan variants. Applying existing methods from ecology and evolution will expand our understanding of the determinants and consequences of glycan variation. The first step is developing biochemical tools and statistical metrics to quantify glycan variation in nature: as it responds to ecology, and as it evolves. Soon, system-level models will predict the patterns of glycan variation expected by specific genetic or regulatory changes. Similarly, appropriate null ecological or evolutionary models could be used as means of inferring the processes that shape glycan variation in space and time. The features of these models will be specific to glycans, but it is certain that they will resemble, and likely extend, statistical methods of analysis that already exist in ecology and evolution.

3.1. Quantifying glycan variation

One of the main uses of statistics in ecology and evolution is to separate phenotypic variation into components, so that differences

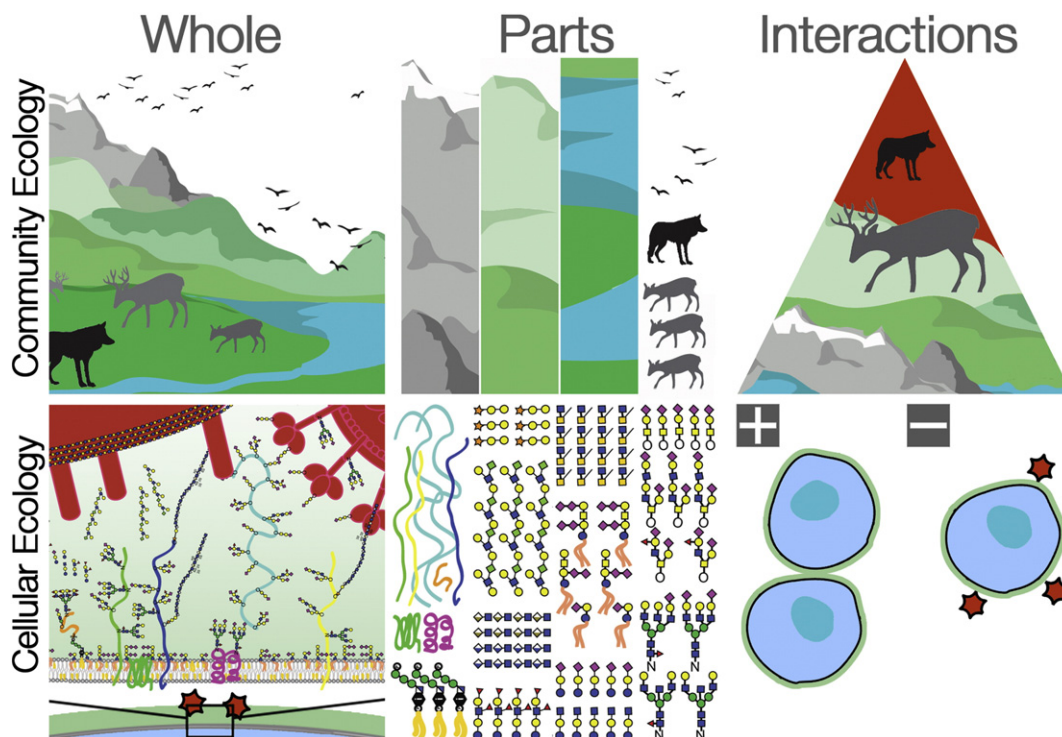


Fig. 1. Community and cellular ecology. Classically, ecological studies consider interactions at the organismal level and above. The interacting parts are whole organisms and their biotic and abiotic environments. The variability in these interactions results from organismal behaviors and physiologies whose mechanistic bases are complex. Cellular ecology is concerned with the interaction of cells, whether within or between individuals. As with traditional ecology, the first task is to survey the parts of the ecosystem. Glycomic methods now permit the enumeration of parts lists for a given interaction, and often these interacting parts are glycans. Unraveling the molecular details and ecological consequences of these mechanistically simple interactions will be the domain of cellular ecology.

caused by genetics and environment can be isolated and studied independent of measurement error [49,52]. Ecologists have developed many different metrics to quantify patterns of variation in space, and these are flexible enough that most can be applied directly to glycan variation. Statistics from ecology can be used to quantify glycan distribution and abundance, to compare diversity and patterns of association, or to study community structure and dynamics [51]. Experimental methods developed by ecologists also translate directly to glycobiology, and are thus already widely used [53,54]. Evolutionary methods are less straightforward to apply. Many rely on estimates of history, which are difficult to infer because traces of historical change are only partially evident in the structure of non-template based molecules. Quantitative genetic methods that draw inference solely from phenotypic variation are directly applicable, although the discrete nature of some glycan variation may violate the assumption that phenotypic variation arises by mutations of tiny effect [55]. However, with methods to quantify natural glycan variation, it should be possible to dissect genetic and environmental effects, estimate heritability, response to selection, rate of evolution, and other metrics applicable to quantitative trait evolution [52,56].

3.2. Predicting glycan variation

Many ecological and evolutionary processes occur on timescales that are too long to observe directly. We must infer the action of a given process by observing its outcomes in nature and comparing them to an expectation. Ecology and evolution use models to precisely specify assumptions that describe a process, and to predict which outcomes we should observe if that process is operating. For example, models of molecular evolution allow us to generate null predictions of evolution in the absence of selection and thus to separate patterns expected by neutral and adaptive evolution [50]. Glycans are often found in phenotypes that mediate evolutionary

conflicts, so they are likely to evolve rapidly. There is a genuine need for theory that can provide null expectations for the evolution of glycan phenotypes. This will be perhaps the most difficult synthesis of glycobiology and evolution because it requires substantial innovation in both fields. In particular, theorists should attempt to develop metrics that can infer deviations from neutral glycan distribution in time and space so that glycobiologists can develop methods to measure glycan variation appropriately.

4. Glycans: tools to understand ecology and evolution

4.1. Sources of glycan variation

Evolutionary biology is now reconciling the modern synthesis of selection and genetics with the reality that phenotypic traits are not entirely genetically determined [57]. Development, diet, environment, and many other non-genetic processes connect information encoded in genes and the phenotypes that are realized in nature. Non-genetic processes could even create additional sources of heritable variation on which selection can act, or impose constraints on evolution that are not obvious from studying genetic variation alone [58]. For example changes in nutrient availability can directly alter available UDP-GlcNAc levels, which can change epigenetic marks on histone tails and alter the regulation of gene expression [59]. These mechanisms directly link environment and gene regulation, and could underpin phenotypically plastic traits where environmental cues create regulatory changes that cause the development of appropriate phenotypes.

Attempts at a broader evolutionary synthesis face the difficulty of uncovering processes that connect genotype and phenotype [60]. Glycans are useful in this respect, because their structures are physical records of the genetic and environmental processes that created them. Synthesis proteins encoded directly in the genome control some

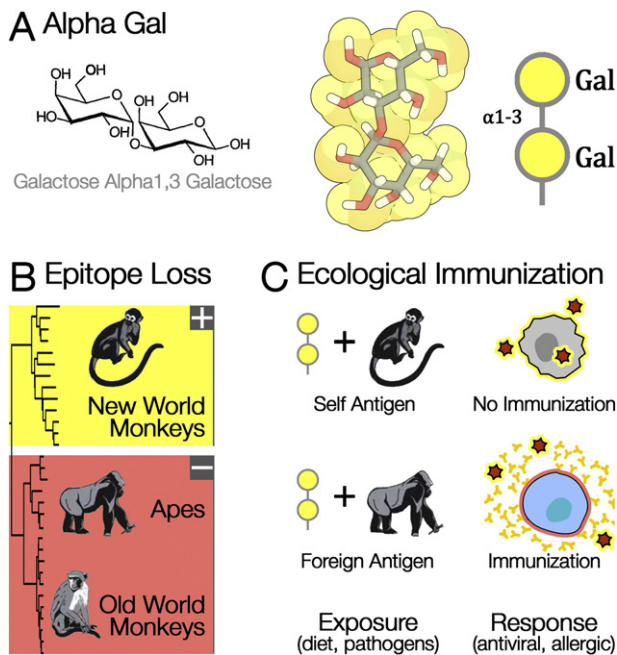


Fig. 2. Alpha gal: a recently evolved, highly reactive, foreign antigen. A) Alpha gal, a common structure on animal glycans, is a terminal galactose alpha 1,3 linked to an underlying galactose. B) α 1,3galactosyltransferase (α 1,3GT), the enzyme that synthesizes alpha gal, was inactivated at the common ancestor of Old World Primates [63]. Apes, humans, and Old World monkeys cannot synthesize the terminal alpha gal linkage. Most mammals, including the New World primates, do synthesize the alpha gal epitope. C) Individuals can be immunized against foreign glycans by exposure to them on pathogens, or by their presence in the diet. The loss of alpha gal may have made Old World primates resistant to infection, because anti alpha gal antibodies target this foreign epitope on enveloped viruses coming from other mammalian species [64]. Anti alpha gal antibodies can cause severe allergic reactions in humans that eat red meat after immunization by a tick bite [65]. Alpha gal antibodies also prevent organ transplantation from other mammal species – pig organs are rejected partly because they are rich in alpha gal [66]. Anti alpha gal antibodies have potential medical applications if they could be directed against pathogens or cancerous cells [67].

synthesis processes, and others depend intrinsically on organismal context. Sugars integrate both kinds of information, and thus will offer a unique opportunity to study the roles of non-genetic information in evolution. Additionally, the spectrum of phenotypic variation created by mutations of glycan modifying proteins varies enormously. Some synthesis steps have specific effects temporally and spatially; others have broad organismal effects. For example, genetic manipulation of N-glycan synthesis in mammals shows that the early steps of glycan synthesis are more conserved than the late steps [61]. As systems-level information accumulates, it may be possible to examine these outcomes and look at relations between pleiotropy, evolutionary rate, and patterns of gene family diversification.

4.2. Interpreting glycan variation

Ecology has yet to undergo a full integration with molecular biology, in part because the ecological relevance of most molecular traits is not yet easily recognizable [8]. Most current molecular ecology studies uncover history using molecular markers but do not elucidate molecular mechanism. Here too, studying glycans will offer new opportunities. Glycans are variable on ecological timescales and the interactions mediated by glycans occur at small spatial scales. Glycans are involved in traits whose ecological importance is recognizable, so glycans will allow experimental tests of ecological theory that take molecular mechanisms of interaction explicitly into account [1]. Glycan phenotypes will also aid on the ground conservation efforts because their variants are sensors of organismal condition as it responds to local ecology. Some preserve aspects of each individual's ecological history

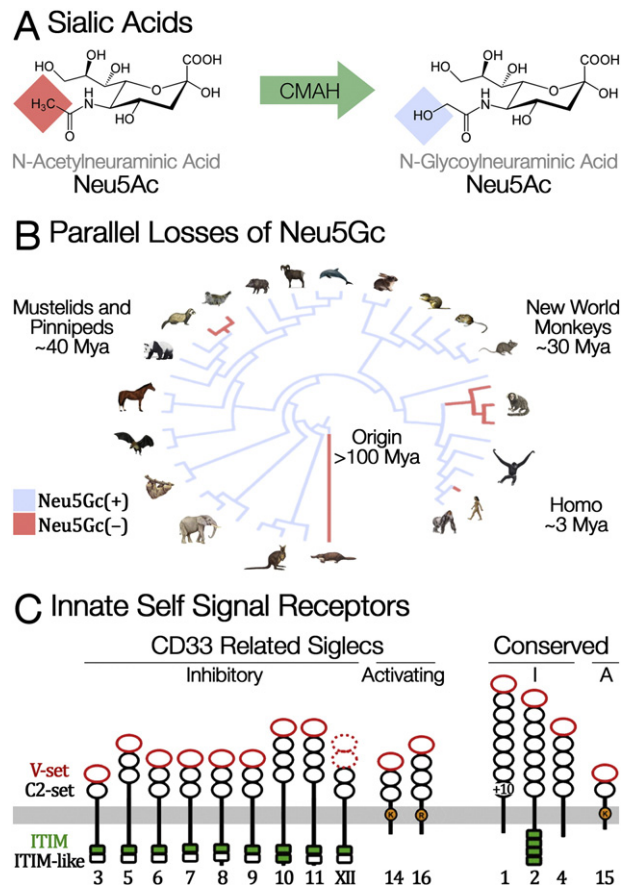


Fig. 3. Sialic acids: parallel evolution of innate self-signals. A) Sialic acids are nine-carbon monosaccharides found at the terminal tips of N- and O-glycans, on glycolipids, and as long chains of polysialic acid. N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc) are two of the most common forms: each a family of molecules with various modifications of the canonical monosaccharide. Neu5Gc is synthesized from Neu5Ac by the CMAH protein (cytidine monophosphate-N-acetylneuraminic acid hydroxylase), which hydroxylates a terminal methyl group. Neu5Gc is a private glycan in the sense that no pathogen has yet been found to synthesize this monosaccharide. B) Humans cannot synthesize Neu5Gc, because human CMAH was inactivated over two million years ago [68]. The inactivating mutation fixed rapidly after originating, which suggests that the loss could have been adaptive – driven by pathogen avoidance, sexual conflict, or a combination of the two [69]. Independent losses of CMAH have recently been found in New World Primates [62] and in Mustelids [70]. C) Sialic acids are bound by a family of Siglec receptor proteins expressed on most immune cells [71]. Binding generally inhibits inflammation, thus sialic acids are innate markers of self. Many pathogens bind Siglecs to exploit these immunosuppressive effects, and in response several immune-activating Siglecs have evolved. Other pathogens use Neu5Gc itself as a receptor on host cells. Several human pathogens have modified their receptor affinity to recognize Neu5Ac following the loss of Neu5Gc in humans.

such as dietary changes or health; others are indicative of disease susceptibility [62].

5. Glycan phenotypes and the synthesis of ecology and evolution

Almost every biological process involves sugars. Their ubiquity reflects an unmatched capacity for structural and functional variability. Glycans can respond to demands that change on physiological, developmental, ecological, and evolutionary timescales. As such, they are often recruited as mechanisms to enact dynamic traits. Glycans are molecules that record the outcome of entire systems of protein interaction – their non-template based synthesis integrates environmental and genetic information. The specificity of glycosylation (position, timing, and variability) is a physical representation of how tightly the synthesis process is regulated, and the effects of dysregulation. This mode of

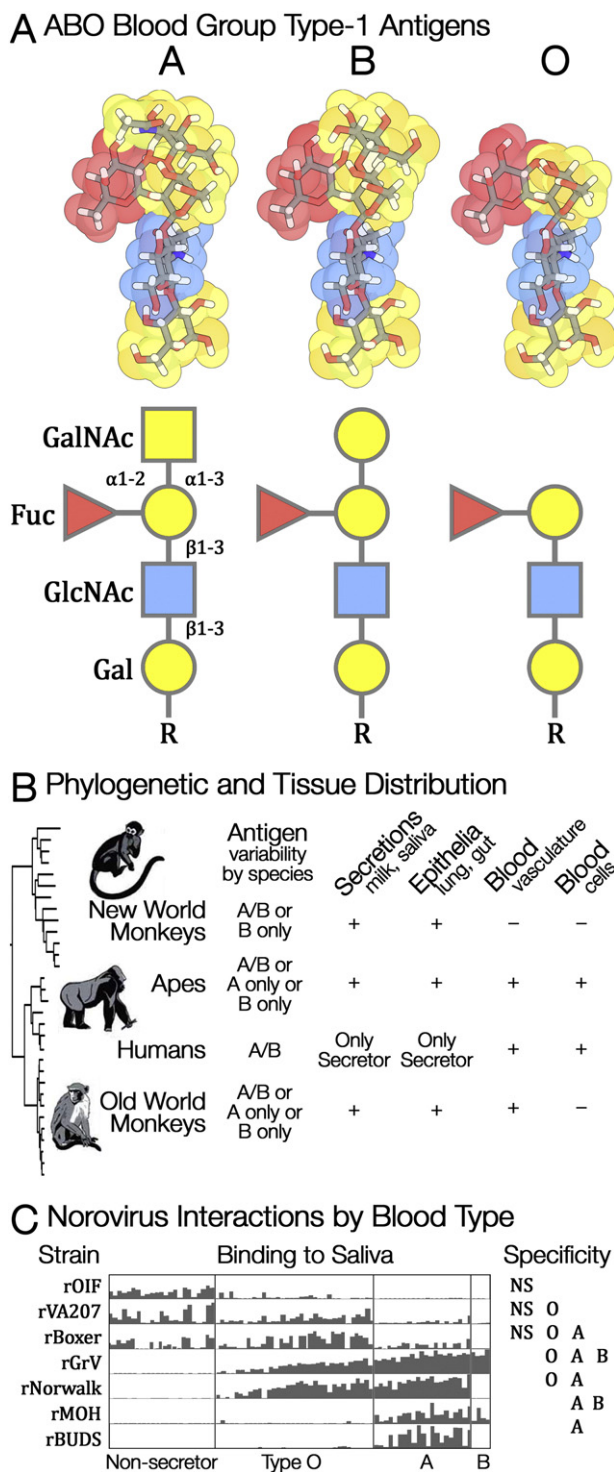


Fig. 4. ABO blood group: pathogens and polymorphism. A) The first molecular polymorphism discovered in humans was the ABO blood group [72]. Blood types were discovered because certain lectins agglutinate the red blood cells of some individuals and not others. These lectins bind glycans that are polymorphic within human populations. The glycans that define A, B, and O blood types are created by the activity of a GalNAc/Gal transferase. An enzyme that transfers N-Acetylgalactosamine results in blood type A; one that transfers Galactose results in type B. An inactive enzyme does not modify the underlying glycan and results in type O. B) The ABO polymorphism originated at the common ancestor of primates [73]. Various patterns of ABO polymorphism are found among primate species. Additionally, a fucosyl transferase (FUT2) causes variation in blood group antigen expression in different tissues. Individual humans are either secretors (ABO glycans are found on epithelia and secretions) or non-secretors (ABO glycans are found only in the blood and vasculature). Other primates have various different secretion patterns. C) Many pathogens exploit specific blood type glycans for host recognition and attachment including noroviruses and rotaviruses [74]. Anti-AB antibodies may help maintain ABO polymorphisms by targeting incoming enveloped viruses bearing mismatched epitopes [75].

phenotype production will allow progress on fronts in ecology and evolution that are difficult to study with proteins or RNA, because the structures, distribution and variability of glycans can vary on organismal timescales. The variability of glycan molecules is caused both by information in DNA and by the context of their synthesis. Mature glycans represent both factors, and so they are the phenotypic records that we need to recognize the ecological relevance of molecular traits.

It has been argued that the disciplines and methods of ecology and evolution still do not fully interface [57]. Difficulties in uniting scientific fields arise either because the information required for synthesis is difficult to collect, or because the importance of existing information is difficult to recognize. Studying glycans can help bridge these divides. Glycans are tractable molecular traits whose genetic and environmental determinants can be recognized and whose ecological relevance can be understood. The non-template based synthesis of glycans creates many technical and theoretical challenges. Glycans do not vary in ways that are fully compatible with existing experimental methods, statistical methods, or theory. But new glycomics methods are emerging that will collect information on natural glycan variation at full biological resolution and scale. Quantifying and predicting the synthesis, regulation, and function of natural glycan variation will allow us to study aspects of ecology and evolution that are not accessible to studies of proteins and RNA alone. Because glycans vary on organismal timescales, they promise an opportunity to further integrate ecology and evolution themselves: two closely aligned but still separate fields of biology.

Conflict of interest

The authors declare that there are no competing interests, financial or otherwise.

Acknowledgments

We thank HS Reynoso, A Lizcano, J Okerblom, and TK Altheide for discussion. This work was supported by the G. Harold and Leila Y. Mathers Charitable Foundation, Grant CSD081 (to P.G.) and by the National Institutes of Health, Grant R01 GM095882 (to P.G.)

References

- [1] S.A. Springer, P. Gagneux, Glycan evolution in response to collaboration, conflict, and constraint, *J. Biol. Chem.* 288 (2013) 6904–6911.
- [2] J.D. Marth, A unified vision of the building blocks of life, *Nat. Cell Biol.* 10 (2008) 1015–1016.
- [3] J.R. Bishop, P. Gagneux, Evolution of carbohydrate antigens—microbial forces shaping host glycomes? *Glycobiology* 17 (2007) 23R–34R.
- [4] M.R. Kasaai, Various methods for determination of the degree of N-acetylation of chitin and chitosan: a review, *J. Agric. Food Chem.* 57 (2009) 1667–1676.
- [5] K. Honke, N. Taniguchi, Sulfotransferases and sulfated oligosaccharides, *Med. Res. Rev.* 22 (2002) 637–654.
- [6] T. Muramatsu, H. Muramatsu, Carbohydrate antigens expressed on stem cells and early embryonic cells, *Glycoconj. J.* 21 (2004) 41–45.
- [7] W.H. Yang, P.V. Aziz, D.M. Heithoff, M.J. Mahan, J.W. Smith, J.D. Marth, An intrinsic mechanism of secreted protein aging and turnover, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 13657–13662.
- [8] S.A. Springer, B.J. Crespi, W.J. Swanson, Beyond the phenotypic gambit: molecular behavioural ecology and the evolution of genetic architecture, *Mol. Ecol.* 20 (2011) 2240–2257.
- [9] B.L. Ridley, M.A. O'Neill, D. Mohnen, Pectins: structure, biosynthesis, and oligogalacturonide-related signaling, *Phytochemistry* 57 (2001) 929–967.
- [10] N.C. Carpita, D.M. Gibeaut, Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth, *Plant J.* 3 (1993) 1–30.
- [11] J.D. Esko, K. Kimata, U. Lindahl, Proteoglycans and sulfated glycosaminoglycans, *Essentials of Glycobiology*, 2009.
- [12] I. Sorensen, D. Domozych, W.G. Willats, How have plant cell walls evolved? *Plant Physiol.* 153 (2010) 366–372.
- [13] C.H. Biermann, J.A. Marks, A.C. Vilela-Silva, M.O. Castro, P.A. Mourao, Carbohydrate-based species recognition in sea urchin fertilization: another avenue for speciation? *Evol. Dev.* 6 (2004) 353–361.
- [14] P.C. Pang, P.C. Chiu, C.L. Lee, L.Y. Chang, M. Panico, H.R. Morris, S.M. Haslam, K.H. Khoo, G.F. Clark, W.S. Yeung, A. Dell, Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida, *Science* 333 (2011) 1761–1764.

- [15] E. Teclé, P. Gagneux, Sugar coated sperm: unravelling the function of the mammalian sperm glycocalyx, *Mol. Reprod. Dev.* 82 (2015) 635–650.
- [16] E. Hennen, T. Czopka, A. Faissner, Structurally distinct LewisX glycans distinguish subpopulations of neural stem/progenitor cells, *J. Biol. Chem.* 286 (2011) 16321–16331.
- [17] V. Mirouse, C.P. Christoforou, C. Fritsch, D.S. Johnston, R.P. Ray, Dystroglycan and perlecan provide a basal cue required for epithelial polarity during energetic stress, *Dev. Cell* 16 (2009) 83–92.
- [18] E. Ikonen, K. Simons, Protein and lipid sorting from the trans-Golgi network to the plasma membrane in polarized cells, *Semin. Cell Dev. Biol.* 9 (1998) 503–509.
- [19] I. Matsuo, C. Kimura-Yoshida, Extracellular distribution of diffusible growth factors controlled by heparan sulfate proteoglycans during mammalian embryogenesis, *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* (2014) 369.
- [20] B.A. Fenderson, I. Stamenkovic, A. Aruffo, Localization of hyaluronan in mouse embryos during implantation, gastrulation and organogenesis, *Differentiation* 54 (1993) 85–98.
- [21] K. Ohtsubo, J.D. Marth, Glycosylation in cellular mechanisms of health and disease, *Cell* 126 (2006) 855–867.
- [22] G.W. Hart, Three Decades of research on O-GlcNAcylation — a major nutrient sensor that regulates signaling, transcription and cellular metabolism, *Front. Endocrinol. (Lausanne)* 5 (2014) 183.
- [23] A. Varki, Since there are PAMPs and DAMPs, there must be SAMPs? Glycan “self-associated molecular patterns” dampen innate immunity, but pathogens can mimic them, *Glycobiology* 21 (2011) 1121–1124.
- [24] T. Boller, G. Felix, A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors, *Annu. Rev. Plant Biol.* 60 (2009) 379–406.
- [25] H. Kono, K.L. Rock, How dying cells alert the immune system to danger, *Nat. Rev. Immunol.* 8 (2008) 279–289.
- [26] G. Ashwell, J. Harford, Carbohydrate-specific receptors of the liver, *Annu. Rev. Biochem.* 51 (1982) 531–554.
- [27] P.K. Grewal, S. Uchiyama, D. Ditto, N. Varki, D.T. Le, V. Nizet, J.D. Marth, The Ashwell receptor mitigates the lethal coagulopathy of sepsis, *Nat. Med.* 14 (2008) 648–655.
- [28] R.L. Schnaar, R. Gerardy-Schahn, H. Hildebrandt, Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration, *Physiol. Rev.* 94 (2014) 461–518.
- [29] P.M. Lackie, C. Zuber, J. Roth, Polysialic acid of the neural cell adhesion molecule (N-CAM) is widely expressed during organogenesis in mesodermal and endodermal derivatives, *Differentiation* 57 (1994) 119–131.
- [30] D. Compagno, F.M. Jaworski, L. Gentilini, G. Contrufo, I. Gonzalez Perez, M.T. Elola, N. Pregi, G.A. Rabinovich, D.J. Laderach, Galectins: major signaling modulators inside and outside the cell, *Curr. Mol. Med.* 14 (2014) 630–651.
- [31] H.Y. Chen, I.C. Weng, M.H. Hong, F.T. Liu, Galectins as bacterial sensors in the host innate response, *Curr. Opin. Microbiol.* 17 (2014) 75–81.
- [32] R.P. McEver, K.L. Moore, R.D. Cummings, Leukocyte trafficking mediated by selectin-carbohydrate interactions, *J. Biol. Chem.* 270 (1995) 11025–11028.
- [33] O.D. Genbacev, A. Prakobphol, R.A. Foulk, A.R. Krtolica, D. Ilic, M.S. Singer, Z.Q. Yang, L.L. Kiessling, S.D. Rosen, S.J. Fisher, Trophoblast L-selectin-mediated adhesion at the maternal-fetal interface, *Science* 299 (2003) 405–408.
- [34] H.J. Flint, E.A. Bayer, M.T. Rincon, R. Lamed, B.A. White, Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis, *Nat. Rev. Microbiol.* 6 (2008) 121–131.
- [35] G. Smant, J.P. Stokkermans, Y. Yan, J.M. de Boer, T.J. Baum, X. Wang, R.S. Hussey, F.J. Gommers, B. Henriks, E.L. Davis, J. Helder, A. Schots, J. Bakker, Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 4906–4911.
- [36] F. Backhed, R.E. Ley, J.L. Sonnenburg, D.A. Peterson, J.I. Gordon, Host-bacterial mutualism in the human intestine, *Science* 307 (2005) 1915–1920.
- [37] A.M. Zivkovic, J.B. German, C.B. Lebrilla, D.A. Mills, Human milk glycobiome and its impact on the infant gastrointestinal microbiota, *Proc. Natl. Acad. Sci. U. S. A.* 108 (Suppl. 1) (2011) 4653–4658.
- [38] K. Ilg, E. Yavuz, C. Maffioli, B. Priem, M. Aebi, Glycomimicry: display of the GM3 sugar epitope on *Escherichia coli* and *Salmonella enterica* sv Typhimurium, *Glycobiology* 20 (2010) 1289–1297.
- [39] E.R. Vimr, K.A. Kalivoda, E.L. Deszo, S.M. Steenbergen, Diversity of microbial sialic acid metabolism, *Microbiol. Mol. Biol. Rev.* 68 (2004) 132–153.
- [40] N. Surana, K., and D. L. Kasper. The yin yang of bacterial polysaccharides: lessons learned from *B. fragilis* PSA, *Immunol. Rev.* 245 (2012) 13–26.
- [41] L.V. Hooper, J.I. Gordon, Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity, *Glycobiology* 11 (2001) 1R–10R.
- [42] P. Gagneux, A. Varki, Evolutionary considerations in relating oligosaccharide diversity to biological function, *Glycobiology* 9 (1999) 747–755.
- [43] R.E. Lehotzky, C.L. Partch, S. Mukherjee, H.L. Cash, W.E. Goldman, K.H. Gardner, L.V. Hooper, Molecular basis for peptidoglycan recognition by a bactericidal lectin, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 7722–7727.
- [44] M. Cohen, X.Q. Zhang, H.P. Senaati, H.W. Chen, N.M. Varki, R.T. Schooley, P. Gagneux, Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase, *Virology* 10 (2013) 321.
- [45] A.A. Lindberg, Bacteriophage receptors, *Annu. Rev. Microbiol.* 27 (1973) 205–241.
- [46] J.N. Reitter, R.E. Means, R.C. Desrosiers, A role for carbohydrates in immune evasion in AIDS, *Nat. Med.* 4 (1998) 679–684.
- [47] J.D. Cook, J.E. Lee, The secret life of viral entry glycoproteins: moonlighting in immune evasion, *PLoS Pathog.* 9 (2013), e1003258.
- [48] C.H. Smit, A. van Diepen, D.L. Nguyen, M. Wührer, K.F. Hoffmann, A.M. Deelder, C.H. Hokke, Glycomic analysis of life stages of the human parasite *Schistosoma mansoni* reveals developmental expression profiles of functional and antigenic glycan motifs, *Mol. Cell. Proteomics* 14 (2015) 1750–1769.
- [49] R.A. King, Statistical ecology, *Annual Reviews of Statistics and its Application*, 1 2014, pp. 401–426.
- [50] Z. Yang, J.P. Bielawski, Statistical methods for detecting molecular adaptation, *Trends Ecol. Evol.* 15 (2000) 496–503.
- [51] Q. Zeng, J. Sukumaran, S. Wu, A. Rodrigo, Neutral models of microbiome evolution, *PLoS Comput. Biol.* 11 (2015), e1004365.
- [52] N.H. Barton, P.D. Keightley, Understanding quantitative genetic variation, *Nat. Rev. Genet.* 3 (2002) 11–21.
- [53] J.E. Koenig, A. Spor, N. Scalfone, A.D. Fricker, J. Stombaugh, R. Knight, L.T. Angenent, R.E. Ley, Succession of microbial consortia in the developing infant gut microbiome, *Proc. Natl. Acad. Sci. U. S. A.* 108 (Suppl 1) (2011) 4578–4585.
- [54] P.C. Kashyap, A. Marcobal, L.K. Ursell, S.A. Smit, E.D. Sonnenburg, E.K. Costello, S.K. Higginbottom, S.E. Domino, S.P. Holmes, D.A. Relman, R. Knight, J.I. Gordon, J.L. Sonnenburg, Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 17059–17064.
- [55] A. Grafen, Natural selection, kin selection and group selection, in: J.R. Krebs, N.B. Davies (Eds.), *Behavioural Ecology an Evolutionary Approach*, Blackwell Scientific Publications, Oxford 1984, pp. 62–84.
- [56] A.P. Hendry, M.T. Kinnison, An introduction to microevolution: rate, pattern, process, *Genetica* 112–113 (2001) 1–8.
- [57] M. Pigliucci, An extended synthesis for evolutionary biology, *Ann. N. Y. Acad. Sci.* 1168 (2009) 218–228.
- [58] T. Day, R. Bonduriansky, A unified approach to the evolutionary consequences of genetic and nongenetic inheritance, *Am. Nat.* 178 (2011) E18–E36.
- [59] S. Hardiville, G.W. Hart, Nutrient regulation of signaling, transcription, and cell physiology by O-GlcNAcylation, *Cell Metab.* 20 (2014) 208–213.
- [60] R.D. Barrett, H.E. Hoekstra, Molecular spandrels: tests of adaptation at the genetic level, *Nat. Rev. Genet.* 12 (2011) 767–780.
- [61] J.B. Lowe, J.D. Marth, A genetic approach to Mammalian glycan function, *Annu. Rev. Biochem.* 72 (2003) 643–691.
- [62] S.A. Springer, S.L. Diaz, P. Gagneux, Parallel evolution of a self-signal: humans and new world monkeys independently lost the cell surface sugar Neu5Gc, *Immunogenetics* 66 (2014) 671–674.
- [63] U. Galili, S.B. Shohet, E. Kobrin, C.L. Stults, B.A. Macher, Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells, *J. Biol. Chem.* 263 (1988) 17755–17762.
- [64] Y. Takeuchi, C.D. Porter, K.M. Strahan, A.F. Preece, K. Gustafsson, F.L. Cosset, R.A. Weiss, M.K. Collins, Sensitization of cells and retroviruses to human serum by (alpha 1-3) galactosyltransferase, *Nature* 379 (1996) 85–88.
- [65] S.E. Wolver, D.R. Sun, S.P. Commins, L.B. Schwartz, A peculiar cause of anaphylaxis: no more steak? The journey to discovery of a newly recognized allergy to galactose-alpha-1,3-galactose found in mammalian meat, *J. Gen. Intern. Med.* 28 (2013) 322–325.
- [66] D.K. Cooper, Identification of alpha Gal as the major target for human anti-pig antibodies, *Xenotransplantation* 16 (2009) 47–49.
- [67] S.A. Kristian, J.H. Hwang, B. Hall, E. Leire, J. Iacomini, R. Old, U. Galili, C. Roberts, K.B. Mullis, M. Westby, V. Nizet, Retargeting pre-existing human antibodies to a bacterial pathogen with an alpha-Gal conjugated aptamer, *J. Mol. Med. (Berl.)* 93 (2015) 619–631.
- [68] H.H. Chou, T. Hayakawa, S. Diaz, M. Krings, E. Indriati, M. Leakey, S. Paabo, Y. Satta, N. Takahata, A. Varki, Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 11736–11741.
- [69] D. Ghaderi, S.A. Springer, F. Ma, M. Cohen, P. Secrest, R.E. Taylor, A. Varki, P. Gagneux, Sexual selection by female immunity against paternal antigens can fix loss of function alleles, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 17743–17748.
- [70] P.S. Ng, R. Bohm, L.E. Hartley-Tassell, J.A. Steen, H. Wang, S.W. Lukowski, P.L. Hawthorne, A.E. Trezise, P.J. Coloe, S.M. Grimmond, T. Haselhorst, M. von Itzstein, A.W. Paton, J.C. Paton, M.P. Jennings, Ferrets exclusively synthesize Neu5Ac and express naturally humanized influenza A virus receptors, *Nat. Commun.* 5 (2014) 5750.
- [71] F. Schwarz, J.J. Fong, A. Varki, Human-specific evolutionary changes in the biology of siglecs, *Adv. Exp. Med. Biol.* 842 (2015) 1–16.
- [72] J.P. Aymard, [Karl Landsteiner (1868–1943) and the discovery of blood groups], *Transfus. Clin. Biol.* 19 (2012) 244–248.
- [73] F. Yamamoto, E. Cid, M. Yamamoto, N. Saitou, J. Bertranpetit, A. Blancher, An integrative evolution theory of histo-blood group ABO and related genes, *Sci. Rep.* 4 (2014) 6601.
- [74] S. Marionneau, N. Ruvoen, B. Le Moullac-Vaidye, M. Clement, A. Cailleau-Thomas, G. Ruiz-Palacois, P. Huang, X. Jiang, J. Le Pendu, Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals, *Gastroenterology* 122 (2002) 1967–1977.
- [75] S.J. Neil, A. McKnight, K. Gustafsson, R.A. Weiss, HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation, *Blood* 105 (2005) 4693–4699.
- [76] Adibekian, A., P. Stallforth, M. Hecht, D. B. Werz, P. Gagneux, and P. H. Seeberger. Comparative bioinformatics analysis of the mammalian and bacterial glycomics. *Chem. Sci.* 2: 337–344.
- [77] F. Schwarz, M. Aebi, Mechanisms and principles of N-linked protein glycosylation, *Curr. Opin. Struct. Biol.* 21 (2011) 576–582.

- [78] P. Stanley, H. Schachter, N. Taniguchi, N-Glycans, *Essentials of Glycobiology*, 2009.
- [79] I. Hang, C.W. Lin, O.C. Grant, S. Fleurkens, T.K. Villiger, M. Soos, M. Morbidelli, R.J. Woods, R. Gauss, M. Aebi, Analysis of site-specific N-glycan remodelling in the ER and the Golgi, *Glycobiology* (2015).
- [80] H. Stepan, E. Staudacher, Optimization of monosaccharide determination using anthranilic acid and 1-phenyl-3-methyl-5-pyrazolone for gastropod analysis, *Anal. Biochem.* 418 (2011) 24–29.
- [81] I. Brockhausen, H. Schachter, P. Stanley, O-GalNAc Glycans, *Essentials of Glycobiology*, 2009.
- [82] K.I. Stanford, J.R. Bishop, E.M. Foley, J.C. Gonzales, I.R. Niesman, J.L. Witztum, J.D. Esko, Syndecan-1 is the primary heparan sulfate proteoglycan mediating hepatic clearance of triglyceride-rich lipoproteins in mice, *J. Clin. Invest.* 119 (2009) 3236–3245.
- [83] B.E. Crawford, O.B. Garner, J.R. Bishop, D.Y. Zhang, K.T. Bush, S.K. Nigam, J.D. Esko, Loss of the heparan sulfate sulfotransferase, Ndst1, in mammary epithelial cells selectively blocks lobuloalveolar development in mice, *PLoS One* 5 (2010), e10691.
- [84] G.M. Miller, L.C. Hsieh-Wilson, Sugar-dependent modulation of neuronal development, regeneration, and plasticity by chondroitin sulfate proteoglycans, *Exp. Neurol.* (2015).
- [85] R.L. Schnaar, A. Suzuki, P. Stanley, Glycosphingolipids, *Essentials of Glycobiology*, 2009.
- [86] M.A.J. Ferguson, T. Kinoshita, G.W. Hart, Glycosylphosphatidylinositol Anchors, *Essentials of Glycobiology*, 2009.
- [87] C. Schaffer, M. Graninger, P. Messner, Prokaryotic glycosylation, *Proteomics* 1 (2001) 248–261.
- [88] H.H. Freeze, R.S. Haltiwanger, Other Classes of ER/Golgi-derived Glycans, *Essentials of Glycobiology*, 2009.
- [89] S. Defaus, P. Gupta, D. Andreu, R. Gutierrez-Gallego, Mammalian protein glycosylation—structure versus function, *Analyst* 139 (2014) 2944–2967.
- [90] S. Ito, Form and function of the glycocalyx on free cell surfaces, *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 268 (1974) 55–66.
- [91] N. Sharon, H. Lis, History of lectins: from hemagglutinins to biological recognition molecules, *Glycobiology* 14 (2004) 53R–62R.
- [92] M.K. O'Reilly, J.C. Paulson, Multivalent ligands for siglecs, *Methods Enzymol.* 478 (2010) 343–363.
- [93] J.C. Paulson, O. Blixt, B.E. Collins, Sweet spots in functional glycomics, *Nat. Chem. Biol.* 2 (2006) 238–248.
- [94] B. Tissot, S.J. North, A. Ceroni, P.C. Pang, M. Panico, F. Rosati, A. Capone, S.M. Haslam, A. Dell, H.R. Morris, Glycoproteomics: past, present and future, *FEBS Lett.* 583 (2009) 1728–1735.
- [95] D.N. Moothoo, J.H. Naismith, A general method for co-crystallization of concanavalin A with carbohydrates, *Acta Crystallogr. D Biol. Crystallogr.* 55 (1999) 353–355.
- [96] M.L. Huang, M. Cohen, C.J. Fisher, R.T. Schooley, P. Gagneux, K. Godula, Determination of receptor specificities for whole influenza viruses using multivalent glycan arrays, *Chem. Commun. (Camb.)* 51 (2015) 5326–5329.
- [97] S.L. Orr, D. Le, J.M. Long, P. Sobieszczuk, B. Ma, H. Tian, X. Fang, J.C. Paulson, J.D. Marth, N. Varki, A phenotype survey of 36 mutant mouse strains with gene-targeted defects in glycosyltransferases or glycan-binding proteins, *Glycobiology* 23 (2013) 363–380.
- [98] H.H. Freeze, E.A. Eklund, B.G. Ng, M.C. Patterson, Neurological aspects of human glycosylation disorders, *Annu. Rev. Neurosci.* 38 (2015) 105–125.
- [99] P.R. Crocker, J.C. Paulson, A. Varki, Siglecs and their roles in the immune system, *Nat. Rev. Immunol.* 7 (2007) 255–266.
- [100] P.V. Chang, J.A. Prescher, M.J. Hangauer, C.R. Bertozzi, Imaging cell surface glycans with bioorthogonal chemical reporters, *J. Am. Chem. Soc.* 129 (2007) 8400–8401.
- [101] M. Cohen, A. Varki, Modulation of glycan recognition by clustered saccharide patches, *Int. Rev. Cell Mol. Biol.* 308 (2014) 75–125.
- [102] K.W. Moremen, M. Tiemeyer, A.V. Nairn, Vertebrate protein glycosylation: diversity, synthesis and function, *Nat. Rev. Mol. Cell Biol.* 13 (2012) 448–462.