Glycobiology vol. 17 no. 5 pp. 23R–34R, 2007
doi:10.1093/glycob/cwm005
Advance Access publication on January 19, 2007

REVIEW

Evolution of carbohydrate antigens—microbial forces shaping host glycomes?

Joseph R. Bishop[2] and Pascal Gagneux[1,2]

[2]Glycobiology Research and Training Center, Cellular and Molecular Medicine-East, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0687

Received on December 5, 2006; revised on January 10, 2007; accepted on January 10, 2007

Many glycans show remarkably discontinuous distribution across evolutionary lineages. These differences play major roles when organisms belonging to different lineages interact as host–pathogen or host–symbiont. Certain lineage-specific glycans have become important signals for multicellular host organisms, which use them as molecular signatures of their pathogens and symbionts through recognition by a toolkit of innate defense molecules. In turn, pathogens have evolved to exploit host lineage-specific glycans and are constantly shaping the glycomes of their hosts. These interactions take place in the face of numerous critical endogenous functions played by glycans within host organisms. Whether due to simple evolutionary divergence or adaptive changes under natural selection resulting from endogenous functional requirements, once different lineages elaborate on differential glycomes these mutual differences provide opportunities for host exploitation and/or pathogen defense between lineages. Such phylogenetic molecular recognition mechanisms will augment and likely contribute to the maintenance of lineage-specific differences in glycan repertoires.

Key words: glycan/co-evolution/host-pathogen/animal lectin

Introduction

Carbohydrates makeup a substantial portion of the biomass on earth, mostly in the form of the two structural polysaccharides—cellulose from plants and chitin from arthropods and fungi. All known living organisms also display an array of free or covalently attached carbohydrates collectively known as glycans (Varki et al. 1999). Some of these complex molecules decorate the surface of cells and are secreted into the surrounding environment where they function in a wide variety of processes required for life including structural support, protection, recognition, localization, and information/nutrient transfer. The precise compositions and combinations of different carbohydrates making up the glycan repertoire of each species can differ dramatically. The rapid development of glycomics methods (Raman et al. 2005) is bound to greatly increase our knowledge about natural glycan diversity, and evolutionary considerations will be crucial for interpreting glycan function within and between organisms (Varki 2006).

Despite recent advances, we are yet to have a complete inventory of naturally occurring monosaccharides used to produce the glycan portion of these molecules, as many members of the Bacteria and *Archaea* domains synthesize a number of specialized carbohydrates (Schaffer et al. 2001). In contrast, metazoan animals build most of their glycans from a very limited number of monosaccharide building blocks, allowing us to consider how these molecules might have evolved over time. Most metazoan glycoconjugates are built from six classes of monosaccharides including sialic acids, hexoses, hexosamines, deoxyhexoses, pentoses, and uronic acids (Varki et al. 1999) see Box 1. These monosaccharides can of course be modified to create greater complexity at the single monosaccharide level. Furthermore, the individual carbohydrate units can be attached via a variety of glycosidic linkages, into highly complex linear or branched structures. Thus in theory, there is virtually no limit to the number of different glycans that can be generated. In practice though metazoan animals seem to generate only a limited range of these possibilities.

It would be impossible to do justice to the overwhelming diversity of natural glycans and their functions in one review. Fortunately, a number of excellent recent reviews and texts address the biology of individual classes of glycans as well as their endogenous ligands, the glycan-binding animal lectins (Staudacher et al. 1999; Angata and Varki 2002; Esko and Selleck 2002; Spiro 2002; Lowe and Marth 2003; Varki and Angata 2006). The aim of this review is to address the taxonomic distribution of glycans and to reflect on the processes that are shaping this distribution. Our main focus will be on how interactions between multicellular animal hosts and their microbial or viral pathogens as well as symbionts may have contributed to the observed lineage-specific constellations of certain glycans, especially extracellular glycans.

Distribution of glycans within the tree of life

Glycans occur in a discontinuous and puzzling distribution across evolutionary lineages. Examples of discontinuously distributed glycans are presented in Table I. The hypothetical evolutionary relationships of living organisms can be depicted in the form of phylogenetic trees. Figure 1 shows three

[1]To whom correspondence should be addressed; e-mail: pgagneux@ucsd.edu

Table I. Some glycans with strikingly discontinuous taxonomic distribution

Peptidoglycans	Bacteria (Koch 2000)
S-layer glycoproteins	Archaeans (Kandler and Konig 1998)
Cellulose	Plants, bacteria, and tunicates (Gibeaut and Carpita 1994)
Chitin	Fungi, arthropods, and mollusks, and one teleost fish (Wagner 1994)
Hyaluronan	Vertebrates, bacteria (DeAngelis 1999)
Glycosaminoglycans	Metazoa and bacteria
Sulfated glycosaminoglycans	Metazoa
Sialic acids	Deuterostome metazoans, bacteria, few mollusks (Angata and Varki 2002)
Neu5Gc sialic acid	Vertebrates but not humans or birds (data on brids incomplete) (Varki 2001)
Gangliosides	Deuterostome metazoans and mollusks
Gal alpha 1–3 Gal	Mammals but not Catarrhines (Galili et al. 1988)
Gal alpha 1–4 Gal	Vertebrates except mammals and some bird lineages (Suzuki et al. 2004)
Lactose secretion	Mammals
Galactose beta 1–4(fucose alpha 1–4) GlcNAc	Catarrhines, Xenopus, plants, and pathogenic bacteria (*H. pylori*) (Oriol et al. 1999; Dupuy et al. 2002; Guerardel et al. 2003)

phylogenies depicting the evolutionary relationships: between the three domains of life (Figure 1A), among *Eukarya* (Figure 1B) and among the anthropoid primates (Figure 1C), respectively, along with the distribution patterns of selected glycan across different evolutionary lineages. As seen, the distribution patterns of glycans fall into four general patterns.

(1) Glycans conserved across many taxa. In contrast to ribosomal RNA that is present in all living organisms, thus allowing the reconstruction of these phylogenies, no single glycan structure has been conserved to the same extent. An example for a relatively conserved class of glycan would be *N*-glycans found in organisms of all the three primary lineages of life, albeit absent from many bacteria (Figure 1A).
(2) Glycans specific to a particular lineage, such as capsule murein peptidoglycans in bacteria (Figure 1A) or gangliosides in vertebrates (Figure 1B).
(3) Glycans similar across distant taxa, examples include glycosaminoglycans found in metazoans and bacteria (Figure 1A); cellulose in plants, bacteria and tunicates; sialic acids (long thought to be unique to metazoan animals) of the deuterostome lineage and also found in many bacteria and in cephalopod mollusks (squid and octopus); or Gal(Fuc alpha 1–4) *N*-acetylglucosamine (GlcNAc) (Lewis A) only found in primates, some other vertebrates, plants, and few pathogenic bacteria (Figure 1B), and
(4) Glycans conspicuously absent from very restricted taxa only (species, families, or higher units) within lineages that otherwise possess such glycans. Examples include Gal alpha 1–4Gal beta1–4GlcNAc present in most vertebrates but absent in mammals and some birds (Figure 1B); Gal alpha 1–3 Gal beta 1–4GlcNAc (alpha-Gal) present in most mammals, but absent in Old World monkeys, apes and humans (Catarrhines), and *N*-glycolylneuraminic acid (Neu5Gc) present in most vertebrates but absent in humans (Figure 1C).

Why do glycans evolve?

Divergence

Like that of any biological molecule, glycan evolution is likely to occur simply due to the divergence of evolutionary lineages. Phylogenies (literally: "history of lineages") come about mostly by the successive bifurcation of lineages, as populations derived from a common ancestor cease to exchange genetic information (i.e., become reproductively isolated). See Box 2 for a list of some key evolution terminology. The genetic tool kits responsible for glycan synthesis and modification of different lineages are subsequently shaped by independent mutational histories, causing the glycan repertoires (glycomes) of different lineages to diverge as well. An example would be the use of cellulose in plants but not in metazoans, with the exception of tunicates (Figure 1B). Divergence involves much historical contingency, where random changes in different lineages, such as the recruitment of certain glycan types over others for specific functions, limit the future evolution of their glycomes.

Natural selection

Selective pressures resulting from recognition processes disproportionately affect the glycans covering cell surfaces. Natural selection acts on glycans, either by favoring the maintenance of a particular glycan (stabilizing or purifying selection) or by diminishing survival and/or reproductive success of organisms carrying a certain glycan (negative selection). Maintenance of the *N*-glycan synthesis pathway in all eukaryotes is an example of stabilizing selection, since disruptions often lead to lethal consequences (Chui et al. 2001; Schachter 2002). Negative selection on glycans could occur whenever an important pathogen exploits a particular glycan as a receptor for infection. Positive selection would entail selection for rapid change in glycans e.g., to accommodate novel endogenous functions.

Convergence

Still another mechanism for generating diversity occurs when organisms belonging to distantly related lineages recruit or "reinvent" similar subsets of glycan repertoires. Such parallel events may be due to particular demands of the environment or be due to random recruitment of ancestral synthetic pathways. The existence of the Lewis A antigen [Gal beta 1–4 (Fuc alpha 1–4)GlcNAc] in Catarrhines and in plants could be such an example, as the enzymes involved in its synthesis have very different genomic sequences (Palma et al. 2001; Javaud et al. 2003) (Figure 1B). Alternatively, what appears as convergent evolution could result from the differential retention of ancestral enzymatic tool kits confined to a few distantly related lineages.

Coevolution

When organisms belonging to different lineages repeatedly interact, as is the case in most natural ecological communities,

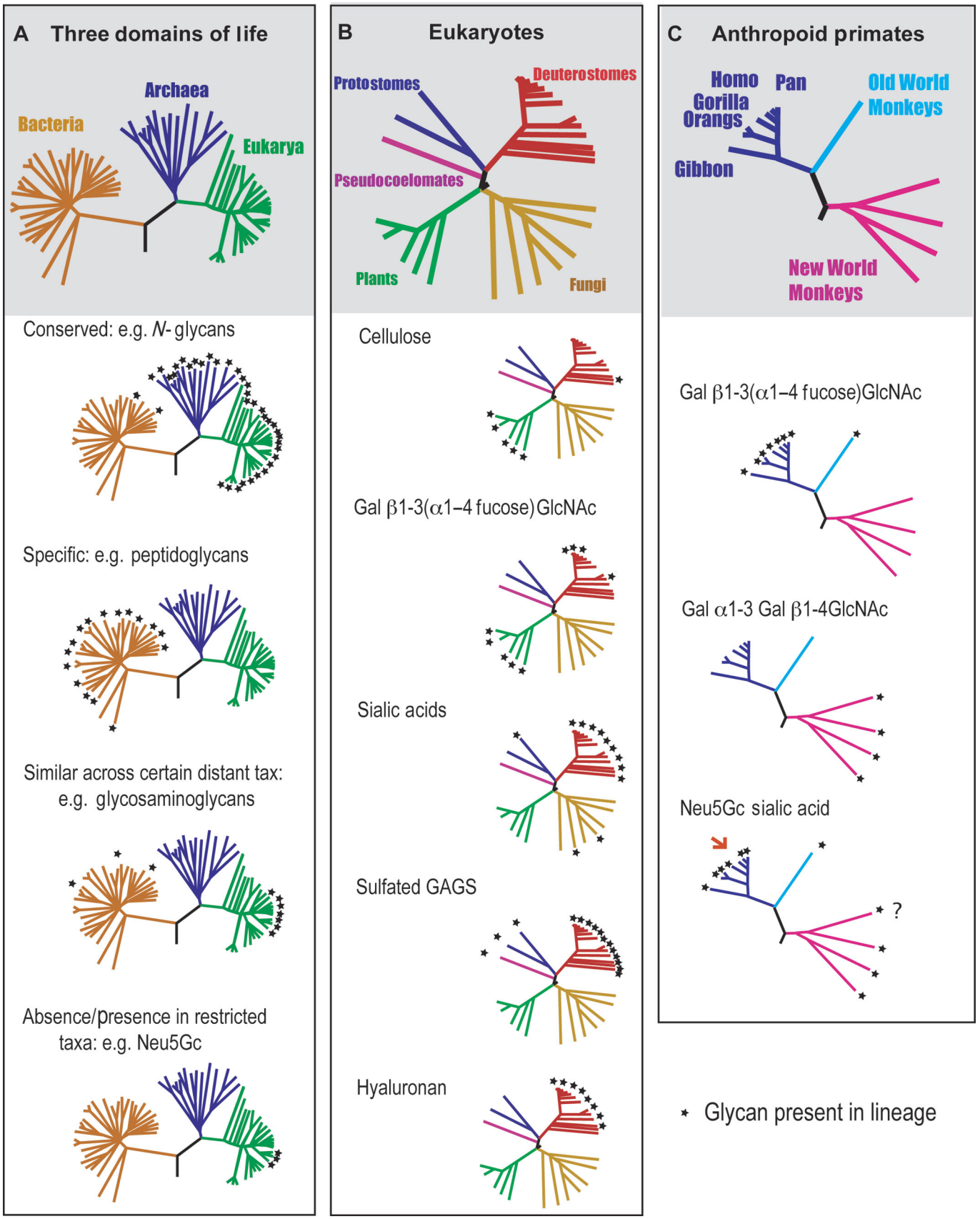

Fig. 1. Repeated phylogenies and glycan distributions for (**A**) the three domains of life, (**B**) eukaryote lineages, and (**C**) anthropoid primate lineages. Phylogenetic trees are redrawn and modified from Purvis (1995), Angata and Varki (2002), and Stearns and Hoekstra (2005).

then their glycomes can become involved in coevolutionary processes. Thus, the interactions of two distinct glycomes of the interacting lineages directly influence their mutual evolution. There is ample evidence for coevolution in glycan diversity in the interactions of microbes and their animal hosts. These cases of coevolution involve two distinct phenomena: (i) independent evolution of enzymatic tool kits for the production of identical molecules in microbes. Examples include glycans found almost exclusively in multicellular hosts and in their microbial pathogens such as glycosaminoglycans and sialic acids (Figure 1B), and (ii) synthesis of "mimic", molecules not identical but very similar to hosts glycans such as polylegionaminic acids by *Legionella* or pseudaminic acid by *Pseudomonas* (Knirel et al. 1987; Kooistra et al. 2001). The Lewis A antigen is also found in certain strains of *Helicobacter pylori*, which infect humans, likely reflecting coevolution (Monteiro et al. 1998). Coevolution could also be occurring via horizontal gene transfer between metazoans and their bacterial pathogens, as has been discussed for genes involved in sialic acid synthesis (Angata and Varki 2002).

Disclaimer about limitations of evolutionary research

While we would certainly agree with the statement that "nothing in glycobiology makes sense, except in the light of evolution" (Varki 2006), we must also realize that evolution only occurred once and that evolution does not follow well-defined rules (Lewontin 2002). This situation is somewhat alleviated by the fact that after lineages diverge, more often than not they remain separated for good and, thus provide researchers with large numbers of iterations ("pseudo samples") for which evolutionary processes have occurred independently. The study of these divergent lineages provides a good opportunity to elucidate evolutionary mechanisms.

A further limitation arises with regard to glycan changes in rapidly evolving organisms such as microbes or viruses, as it is impossible to gain information from long-extinct pathogens, which leave no fossils. The speed of evolution in pathogens means that the identity of past pathogens will never be known and that many current pathogens may be descendents of earlier innocuous microbes or even former symbionts. Rapid evolutionary rates are also associated with homoplasy, i.e., if the observed similarity between glycans is not necessarily due to recent shared ancestry but could have evolved independently in different lineages (convergence). In the era of genomics, the ability to investigate the genomic sequences of the genes coding for enzymes that assemble and modify glycans in different lineages provides a powerful means of reconstructing the evolutionary history of glycosylation by determining key events in the establishment of glycan synthesis machinery.

Glycans in metazoan animals

In metzoan animals, cell surfaces are covered with an electron dense coating of glycoconjugates known as the glycocalyx. Further, glycans are directly secreted as polymers or attached to proteins into the extracellular matrix and body fluids. This glycan landscape is often (for functional and historical reasons) characteristic of both species and particular cell types. (Paulson and Colley 1989; Roth 1996). Four basic types of glycoconjugates are present in metazoans including N-linked, O-linked, glycolipids, and proteoglycans (Varki et al. 1999). These molecules play a large array of functions required for life including support, signaling, protein folding, and protection (Table II).

Table II. Endogenous functions of glycans in metazoans that go beyond structural function

Protein folding/chaperone-assisted folding in ER
Protein subunit assembly
Cell–cell interactions
Cell-extracellular matrix interactions
Cell-complement interactions
Intra- and inter-cellular trafficking
Signaling
Protection from protease degradation
Exogenous functions of glycans in metazoans
Welcome signals for important symbionts
Long-term accommodation of symbionts
Arbitration of symbiotic microbial communities
Decoys against pathogen recognition
"Smoke-screen" against pathogen recognition
Detecting nonself based on absence of self-glycans
Detecting nonself based on conserved nonself-microbial glycans
Detecting nonself based on polymorphic nonself-glycans in the population

Why vertebrates use only such a small fraction of monosaccharide types for the assembly of their glycans remains a mystery (Box 1). For example, what is the reason why vertebrates, unlike plants do not carry terminal xylose on their *N*-glycans or incorporate any trehalose in their glycan repertoire? Absences of such structures are likely to represent cases of lineage-specific evolutionary happenstance (contingency), whereby the independent mutational history of different lineages has led to differential evolution of glycan biosynthesis enzymes along separate lineages. Paradoxically, however, even with their relatively reduced panel of monosaccharides (compared to bacteria for example), vertebrates generate a staggering amount of structural variation by combining just nine principal monosaccharides into chains of varying lengths and degrees of branching on differentially decorated proteins and lipids (Manzi et al. 2000). It also appears that despite the relative small number of different building blocks (monosaccharides),vertebrates produce much more complex branched *N*-glycans than many other lineages (Varki et al. 1999).

Box 1. Principal building blocks of vertebrate glycans

Sialic acids: e.rg. *N*-acetylneuraminic acid (Neu5Ac. *N*-glycolylneuraminic acid (Neu5Gc)
Hexoses: Glucose, mannose, galactose (Gal)
Hexosamines: *N*-acetylglucosamine (GlcNAc. *N*-acetylgalactoseamine GalNAc
Deoxyhexoses: Fucose (Fuc)
Pentoses: Xylose
Uronic acids: Iduronic acid, glucuronic acid

Box 2. Glossary of evolution terminology

Antagonistic coevolution: "evolutionary arms race", where changes in one lineage of a pair of host-parasite lineages are prompted by or prompt changes in the other lineage.
Convergence: similarity between taxa despite independent evolutionary histories.
Catarrhine: primates belonging to Old World monkeys, apes and humans.
Demographic bottleneck: strong reduction in population size.
Divergence: differences between taxa due to independent evolutionary histories.
Domain: one of the three radiations of life including the *Archaea, Bacteria*, and *Eukarya*.
Founder event: establishment of new populations by small numbers of founder individuals.
Frequency dependent selection: when the fitness of a genotype depends on its frequency.
Genetic drift: random variation in gene frequency from one generation to another.
Historical contingency: the effect of random event on the probability of subsequent events in a lineage.
Homoplasy: similarities in character states for reasons other than inheritance from a common ancestor. These include convergence, parallelism, and reversal.
Lineage: group of organisms sharing a common ancestor (monophyletic).
Phylogeny: hypothetical history of related lineages based on DNA sequences or any other heritable derived traits.
Purifying (stabilizing) selection: a type of selection that removes individuals from both ends of a phenotypic distribution thus maintaining the same distribution mean.
Trade-off: the balancing of different selection pressures especially when these have opposing directions.

Variation in animals glycan antigens in time and space

Transient glycan variation in animals has been documented during key processes such as pregnancy, lactation, infection, or acute phase response, whereas ontogenetic glycan variation plays key roles in the regulation of metazoan development (Haltiwanger and Lowe 2004). Glycans can also vary in space, as different compartments and adjacent tissues in many animal species carry different glycan repertoires. In a given metazoan, one could detect different glycan distribution from the outside as the secreted mucins and bound mucins of the mucous membranes, the epithelia, the basal layers, the stroma, the endothelia of blood vessels, the different types of immune cells, the cells of the peripheral and central nervous system, and the reproductive systems all usually vary with respect to their surface glycans (Ohtsubo and Marth 2006). Such variation and its distribution might reflect trade-offs between the needs for endogenous function and adaptation to external selective pressures from pathogens or accommodation of important symbionts.

Genes coding for glycan biosynthetic enzymes have undergone substantial expansion contributing to glycan diversity in metazoans

A substantial fraction (1–2%) of animal genes function in glycan biosynthesis and modification. Unlike genes coding for a single protein product, these enzymes work in an "assembly-line" like system of glycan synthesis pathways (Lowe and Marth 2003). These pathways allow organisms to generate rapid phenotypic changes based on posttranslational modification of their glycoconjugates. Glycosyltransferases, which catalyze the addition of sugars to growing glycan chains and proteins, have been subject to multiple lineage-specific expansions via gene duplication (Lespinet et al. 2002). In mammals, for example, there are 9 fucosyl transferases, compared to 4 in *Drosophila* and up to 18 in *Caenorhabdiitis elegans* (Javaud et al. 2003). In the case of fucosyl transfereases, genomic analyses have determined that the more ancient genes in mammals had multiple exons and typically encode enzymes that transfer fucose near the base of the N- or O-peptidic sequences, whereas more recent genes are monoexonic and encode transferases acting at the periphery (termini) of glycans (Oriol et al. 1999). A variety of genetic mechanisms caused this expansion, including duplication, exon shuffling, point mutations, and transposition (Javaud et al. 2003). Similar examples are seen in a host of enzymes including sialyltransferases (19 in mammals) and heparan sulfate proteoglycan modification enzymes (15 in mammals) (Esko and Lindahl 2001; Harduin-Lepers et al. 2005). As the majority of these enzymes reside in the endoplasmic reticulum (ER)–Golgi secretory system, these organelles have rightly been called the "evolvability" module of animal cells providing organisms with machinery for generating variation through combinatorial modification of expressed proteins (Kirschner and Gerhart 1998).

Genetic studies in model organisms with null mutations in biosynthesis genes have proved that many glycans are required for proper metazoan development, as these mutations produce phenotypes ranging from embryonal lethality to growth defects to impaired morphogenesis and cognitive function—but some can also have no obvious effects under laboratory conditions (Natsuka and Lowe 1994; Kotani et al. 2001; Lowe and Marth 2003; Kudo et al. 2006). It is conspicuous that the consequences of experimental abolition of many glycans are often not evident in animal cell cultures, even when these prove to be lethal as early as the embryonic stage in the whole organism from which the cells are cultured (Grobe et al. 2002). These findings point to key functions of glycans for multicellular development, but they also leave open the possibility that a certain fraction of animal glycans can be selectively neutral, i.e., these can be altered without incurring major fitness costs to the organism. Laboratory studies looking at consequences of experimental glycan alteration for individuals are unlikely to shed light on population-level effects of glycan polymorphism, such as the proposed protective effects in preventing the rapid spread of pathogens due to herd immunity-related mechanisms (Gagneux and Varki 1999). This idea remains untested in part because such effects would be based on populations rather than individuals.

Animal lectin intractions and glycan evolution

In many cases, the endogenous function of glycans requires interaction with proteins, and recent decades have seen the discovery of a growing list of animal lectins with specific carbohydrate recognition domains (CRD) (Drickamer and Taylor 1993; Gabius 1997; Probstmeier and Pesheva 1999; Rini and Lobsanov 1999). Binding is usually highly specific for glycan type, as defined by its monosaccharide composition and the nature of the glycosidic linkages by which these are connected (Drickamer and Taylor 1993; Kaltner and Stierstorfer 1998; Kilpatrick 2002). Lectin–glycan interactions can mediate a variety of cell–cell recognition events including interspecies (host–pathogen), intraspecies (fertilization and

gestation), and intercellular (development and immune regulation) interactions (Gabius et al. 2002). Much of the diversity of metazoan glycans is found at the periphery of the glycans involving terminal portions capped by sialic acids, fucose, galactose or GalNac or, in the case of proteoglycans, by directed modification of linear glycan chains (Varki 1993; Esko and Lindahl 2001). This increase in diversity towards the exterior termini of glycans on the surfaces of mammalian cells (Dennis et al. 1999; Varki 2006), combined with the fact that animal lectins often recognize specific glycan structures found on the termini, strongly suggests that recognition processes rather than simple divergence are driving this diversity. It is no surprise therefore, that microbial and viral pathogens of metazoan hosts have evolved their own sets of lectins to exploit these molecules for host recognition, attachment and tissue tropism (Karlsson 1995; Sharon 1996; Rostand and Esko 1997).

The evolutionary glycan arms race

The ubiquitous presence of species-specific glycans on host cells and secretions predispose these as convenient receptors to be exploited by microbes for host recognition, attachment, and invasion by way of a wide array of microbial and viral lectins including adhesins, pili, fimbriae, and hemagglutinins (Gilboa-Garber and Garber 1989; Wadström and Ljungh 1999). For specific examples of glycan-mediated host–pathogen interactions, we refer the reader to several excellent reviews (Sharon 1996; Rostand and Esko 1997; Hooper and Gordon 2001; Olofsson and Bergstrom 2005). Host invasion often occurs via the large epithelial layers lining the external cavities of vertebrates, which participate in gas exchange, olfaction, nutrient uptake, secretion, and reproduction. Outer epithelia are characterized by mucous covered membranes derived from glycoconjugates (mucins, proteoglycans, etc) that form an important interface between these animals and their environments replete with ubiquitous microbes. Rather than providing convenient points of invasion, mucin secretions may act as efficient decoys or smokescreens absorbing intruding pathogens before they can invade (Perrier et al. 2006). Apart from host recognition and invasion, microbes can further exploit multicellular host glycans in a variety of ways: they can (i) scavenge host glycans and use these as carbon source (Sonnenburg et al. 2005). (ii) engage in host mimicry by synthesizing glycans identical or nearly identical to those of the host (Martin et al. 1997; Bersudsky et al. 2000; Harvey et al. 2001; Vimr et al. 2004), and (iii) modulate host glycans by expressing glycosidases to destroy host decoy glycans or to expose more appropriate underlying saccharides for lectin interaction (Dwarakanath et al. 1995; Vimr et al. 2004). The intracellular parasite *Trypanosoma cruzii* even expresses an enzyme that transfers host sialic acids to its own cell surface as a type of camouflage (Colli 1993).

It may seem that rapid glycan structural change in response to pathogenic microbes would be the best route to evade infection, however, change of glycans has the potential of negatively affecting critical endogenous functions or jeopardizing successful interaction with symbionts (see *Hostsymbiont coevolution*). Given this, we speculate that microbe driven alteration in host glycan structure is more likely if the change minimally affects endogenous function(s) or if the selection is strong enough to outweigh the impaired endogenous function. Similarly, pathogens can evolve to counter-adapt to changes in host glycan structure by altering ligand/receptor specificity. Such antagonistic coevolution (also called "evolutionary arms race") is known to lead to rapid evolutionary change (Buckling and Rainey 2002). It appears that the ongoing arms race between microbes and their animal hosts is constantly shaping the makeup of glycans of both sides, and such glycan changes must be considered against the background of "normal" glycan variation.

Owing to the observed glycome differences in distant lineages, it is tempting to speculate that glycome differences often represent insurmountable barriers for pathogens of one distant lineage for infecting hosts of another lineage. For example, plants and animals share few terminal glycans and, with one possible exception (Gibbs and Weiller 1999), there seem to be no plant pathogens that also infect animals or vice versa.

Glycans as innate markers of nonself

Many microbes seem to be affected by lineage-dependent constraints such as the glycan composition of their cell walls. The highly conserved capsule glycans in pathogenic microbes can be exploited by multicellular hosts as "pathogen associated molecular patterns" (PAMPs) or "microbial motifs" and used as target molecules for (Kawabata and Tsuda 2002) pathogen recognition receptors of the innate immune system (Weis et al. 1998). Typical structures associated with bacterial lineages and used as PAMPs by multicellular host lectins are lipopolysaccharides of the outer membrane of gram-negative bacteria, lipoteichoic acid of gram-positive bacteria, high mannose glycan and betaglucans of fungi. Multicellular hosts have been able to exploit such PAMPs to the extent where their recognition is encoded in the germ line of the host in the form of innate immune receptors such as toll-like receptors, DC-SIGN, or mannose-binding lectins of dendritic cells (Cherayil et al. 1990; Akira et al. 2001; Appelmelk et al. 2003). These glycan recognition molecules are essential for survival and some are even shared by metazoa and plants (Toll-like receptors). Innate immune systems of invertebrates seem to compensate for the absence of an adaptive immune system by having special lectins with divergent ligand specificities for recognizing different polysaccharides of pathogen membranes as well (Zhu et al. 2006). Given the multitude of lectin-based innate immune recognition mechanisms, it appears that glycans have formed a substantial part of the basis for lineage–specific recognition of prevailing pathogens via innate immune systems (Janeway and Medzhitov 2002).

Further, by expanding terminal glycan structures which are absent from pathogen lineages, metazoan hosts can recruit these same structures as innate determinants of self. Mammals have evolved 19 sialic acid glycosyltranseferases and utilize the absence of this terminal glycan for the detection of nonself. Lack of sialic acid on any cell surface perturbs factor H binding and allows complement molecules to be deposited on the surface leading to an immune attack (Pangburn et al. 2000). Simultaneously, a family of endogenous mammalian lectins called Siglecs mediate immune cell functions based on the presence of sialic acid (Crocker and Varki 2001). Thus, host innate immune systems directly

target microbe glycans and readily detect the absence of self-glycans as well (Janeway and Medzhitov 2002).

Nonself glycan and adaptive immunity

Jawed vertebrates have the capacity to generate virtually unlimited variation of receptors with their adaptive immune systems. This important evolutionary innovation provides these animals with a flexible system, capable of learning (affinity maturation) and experienced-based memory. This innovation is also double edged, as antibodies targeting foreign peptides may cross-react with host epitopes including glycans (Hedrick 2004), such as infection with *Campylobacter*, which can result in the autoimmune disease Guillain–Barré syndrome (Ang et al. 2004). Ironically, the adaptive immune system itself became possible by recruitment of recombination-associated genes (RAG), which are themselves of viral origin (Du Pasquier 2004). Some glycans have the capacity to elicit immune responses when introduced into animals, which do not possess the same structure as part of their glycoprofile/glycan portfolio (Schauer 1988). More importantly, perhaps, glycans often provide crucial parts of antigenic epitopes found on glycolipids or glycoproteins. Recent studies have shown, for example, that lineage-specific glycans on plant glycoproteins are major antigens and are responsible for human allergies to plants (Bardor et al. 2003). Humans also produce anti-Neu5Gc antibodies against this otherwise very common mammalian sialic acid (Nguyen et al. 2005). Also, vaccines such as the *Haemophilus influnezae* b (Hib) vaccine take advantage of the fact that when conjugated to bacterial proteins (such as toxins) glycan antigens generate a strong T-cell dependent immune response (Kelly et al. 2004). Finally, some of the most potent adjuvants in mammals are glycans from very distantly related taxa such as the mollusk keyhole limpet or horseshoe crab (Jennemann et al. 1994). Thus, it seems likely that vertebrates can generate specific antibodies to pathogen glycoproteins, however, these must be limited to those that fail to recognize self-glycans.

Host symbiont coevolution

Metazoans must tolerate huge numbers of microbial (nonself) symbionts. Thus, host immune systems must accommodate "a vast consortium of symbiotic bacteria" and all their surface glycans, while distinguishing them from pathogens (Cash et al. 2006) (Ironically, one of the reasons such microbes are essential is that vertebrates can extract valuable nutrients from the abundant, but biochemically inaccessible, plant structural polysaccharides only with microbial enzymatic help.). Host glycans appear to play crucial roles for providing symbiotic microbes with attractive niches ("welcome mats") while discriminating against pathogens. For example, mammalian hosts have microbe-binding lectins lacking complement recruitment domains for gut symbionts (Cash et al. 2006). Gut micro flora can specifically modulate the gut glycosylation pattern (Freitas et al. 2002), and in mammals some of these effects are important for the establishment of proper host glycosylation after weaning (Bry et al. 1996). In addition, hosts need to have mechanisms for monitoring symbiotic microbe communities that are capable of turning into "pathologic communities", given the wrong circumstances (Ley et al. 2006). Successful sequestration of important microbes is a prerequisite for successful symbiosis and avoidance of invasion/infection. As such, symbiont management by hosts could be considered a stepping-stone for control of pathogenic microbes e.g., via secretion of antimicrobial peptides. In setting the tolerance appropriately from zero against pathogens and to substantial for well-sequestered symbiotic microbes, one role of glycans has been termed "legislators of host–microbial interactions" and may have played a role in the distribution of glycans among divergent lineages (Hooper and Gordon 2001).

Adaptation by glycan loss

A drastic mechanism for hosts to alter the glycan composition of their cell surfaces is to abolish the expression of a terminal glycan structure in order to curtail pathogen interaction. The complete loss of a particular glycan usually involves inactivating mutations of one or more genes involved in assembly followed by the fixation of the inactive allele across the population. Fixation of such mutations can come about due to selection for absence of the glycan or by genetic drift due to small population size (founder events or demographic bottlenecks). The complete loss a glycan modification, which is otherwise very common in many closely related lineages (e.g., alpha-Gal in Catarrhines or Neu5Gc in humans) has at least two advantages: (i) the loss quickly prevents recognition by pathogens using structure as a receptor and (ii) it opens the possibility of adding the abolished glycan to the panel of nonself-glycans recognized by adaptive immunity. For example, in the human and other primate blood groups, the absence of a glycan type is also accompanied by the presence of antibody against the missing glycan (Clausen and Hakomori 1989).

Of course, there is a potential cost to such an adaptive glycan loss. If the nonfunctional allele responsible for the loss becomes fixed in the population, the lost glycan will likely be lost forever, as random mutations are much more likely to further incapacitate a gene rather than to revive its function. A further cost will result when a glycan with important endogenous functions is lost (e.g., due to very strong negative selection by a pathogen), as this will require subsequent compensatory changes in the endogenous lectins. It follows that the set of endogenous lectins of each lineage can be expected to closely mirror that lineage's glycan repertoire as far as endogenous function is concerned. The human specific changes in several siglec genes might be an example for such compensatory changes, as humans have lost the ability to make Neu5Gc and some of their sialic-acid-binding siglecs have shifted from binding both Neu5Gc and Neu5Ac to a strong preference for binding Neu5Ac (Brinkman-Van der Linden et al. 2000). The potential costs associated with such radical glycan remodeling are illustrated by the many different forms of congenital disorders of glycosylation involving deficiencies in *N*-glycan synthesis (even if each particular form is rare) (Aebi and Hennet 2001). It has been suggested that selection for altered levels of *N*-glycan synthesis could be linked to an inhibitory effect on viral replication (Freeze and Westphal 2001). Most animal populations are likely in a permanent process of striking a trade-off between glycan

changes in response to pathogens and conservation of endogenous functions based on the same glycans.

Invariably, before a glycan can be lost from a particular linage, the loss will only occur in certain individuals within the populations of a given species (the carriers of the inactivating mutation). If the loss conveys selective advantages due to frequency-dependent selection (under which the absence of the glycan is only selected for while it is relatively rare within the population, but becomes selected against when too common), this will produce polymorphisms within the species. Theoretically, any polymorphic system involving the presence or absence of a particular glycan should be viewed as a candidate for eventual loss of the glycan due to negative selection, genetic drift, or a combination of both. Because of the irreversible nature of loss, populations that maintain polymorphisms retain more plasticity for future evolution than those that completely abolish a glycan.

Adaptation by glycan gain

A second way to evade pathogens is to create a new glycan structure either by way of synthesis or modification. There seem to exist few documented cases of "neo-glycans" or glycan "inventions" in vertebrates. However, one apparent example is Neu5Gc with its narrow distribution exclusively among vertebrates and "higher" invertebrates (Angata and Varki 2002). While several nonhuman pathogens exploit Neu5Gc as a receptor (Kyogashima et al. 1989; Suzuki et al. 2000), the lack of reports on microbes carrying Neu5Gc would indicate that this vertebrate-specific sialic acid may have provided vertebrate hosts with some freedom from microbial mimicry (see *Appreciating complex microbial glycan strategies*). Another example is found in the glycosaminoglycan (GAG) chain modifications of vertebrates. Vertebrates, particularly mammals, have evolved elaborate enzyme sets for the modification of sugars on the GAG chains of proteoglycans (Grobe et al. 2002). In this system, modifications, which include *N*-sulfation of GlcNAc residues, *O*-sulfation of both GlcNAc and uronic acids, and epimerization uronic acids, have played a major role in the creation of ligand-binding sites, while they also distinguish mammalian GAGs from "mimics" produced by bacteria that are not modified (Esko and Lindahl 2001). In turn, pathogens have evolved modified GAG-specific lectins to recognize permissive cell types for invasion (Spear et al. 1992).

Is there neutral variation in glycans?

As an adaptation to a world filled with pathogens, multicellular host populations would ideally maintain large and flexible glycan repertoires, which include numerous neutral or near-neutral glycan structures encoded by genes with intermediate frequencies in functional and nonfunctional alleles. Polymorphisms in humans involving "natural knock outs" with no discernable disadvantages do exist. Examples include those lacking A and B blood-group glycans, the Bombay blood group, and Lewis A negative individuals. While of no immediate adaptive value to an individual under normal conditions, some of these polymorphisms seem to persist in populations with surprisingly stable frequencies, an indication that these are maintained by natural selection effectively making them "non-neutral" (Marionneau et al. 2001). These histo-blood groups also include variation in tissue/cell type-specific glycan expression, as exemplified by the secretor/nonsecretor status for ABO and Lewis antigens (Marionneau et al. 2001). How much of such tissue-specific microheterogeneity and variation of glycans on adjacent tissues and cell types could be attributed to neutral variation is an equally important yet unanswered question. Unfortunately, the current protocols for glycomic studies run the risk of missing much of this type of variation, as these generally rely on extractions of glycans from homogenized tissues, where ever cell types cannot be easily isolated (such as blood cells).

Glycans as markers of viral "micro transplantation"—protective host glycan diversity?

Viruses are of unknown evolutionary origin and still fall outside the three domains of life. They are characterized by minute genomes even when compared to bacteria and (with the notable exception of glycosidases such as neuraminidases) they do not have genes for glycan synthesis or modification. Because glycosylation of infected host cells is carried out by host enzymatic machinery in the ER–Golgi of the host, viruses will inherit host cell glycans after each round of replication in a new host, a process especially noteworthy for enveloped viruses, which also inherit entire host membrane glycoconjugates such as major histocompatibility complex molecules (Liedtke et al. 1994). This means that the host cell in which the virus last replicated generated the glycans on viral glycoproteins. Tissue graft rejection (a human invention) is frequently mentioned as one of the costs of the adaptive immune system in the medical literature, as if this was a failure of the adaptive immune system. Enveloped virus infection can actually be considered as the oldest form of "micro" graft, as most enveloped viruses infecting a new host arrive carrying glycoconjugates derived from the cells of the previous host from which they last emanated. We had suggested a few years ago that host cell surface glycan diversity could act as protective diversity with respect to enveloped virus infection (Gagneux and Varki 1999). Two studies have since found evidence for such a protective effect due to the presence of mismatched ABO histo-blood group glycans on enveloped viruses. One involved in vitro studies with measles virus (Preece et al. 2002) and the other epidemiology of human immunodeficiency virus-1 (Neil et al. 2005). The alpha-Gal epitope is present in millions of epitopes per cell in all mammals except Catarrhines. Anti-alpha-Gal antibodies are the most abundant natural antibodies in human serum. It has been suggested that the loss of alpha-Gal has allowed Catarrhines to severely reduce infection risks from enveloped viruses emanating from other mammalian species in their ecosystem by abolishing an abundant surface glycan used by viruses as a receptor, and by carrying a preformed immune reaction in the form of circulating anti-alpha-Gal antibodies (Repik et al. 1994; Rother and Squinto 1996). Several authors have commented on the importance of keeping such protective mechanisms against cross-species viral transmission in mind when considering the multiple efforts to rid pigs of glycan antigens in order to provide sources of xeno-transplantable organs (Weiss 2000; Yoo and Giulivi 2000; Magre et al. 2003). A recent study using modeling has proposed a selective

mechanism for the maintenance of ABO polymorphisms based on the selection pressure from intracellular enveloped viruses incorporating host glycans versus those of extracellular bacteria exploiting glycans as receptors (Seymour et al. 2004). In this regard, it is interesting that several primate lineages have independently evolved the O blood group by different loss of function mutations in their respective (orthologous) transferase gene (Doxiadis et al. 1998; Kermarrec et al. 1999).

Appreciating complex microbial glycan strategies

In the course of the evolutionary arms race between microbes and their hosts, some microbes have evolved to exploit the germ line encoded (and thus relatively inflexible) pathogen recognition receptor (PRR) by letting themselves be recognized only to then infect the dendritic and other immune cells (effectively using their own PAMPs as Trojan horses in order to gain entry by exploiting host PRR). *Neisseria gonorrhoeae* uses the asialo lipooligosaccharide receptor on sperm to hitch a ride into the female reproductive tract (Harvey et al. 2000). DC–SIGN are dendritic cell (DC) specific intercellular adhesion molecule-grabbing nonintegrins (CD209). DC-SIGNs are C-type lectins that recognize mannose and fucose containing glycans on microbes such as *Mycobacterium tuberculosis, Helicobacter pylori*, and *Leishmania* (Appelmelk et al. 2003). These dendritic cell lectins recognize helminth and microbial glycans but also certain similar plant glycans, e.g., the major peanut antigen (PNAg) peptide nucleic acids glycoprotein in a glycan-dependent fashion (Shreffler et al. 2006). This illustrates the dangers of innate pathogen receptors depending on a type of molecule, which may also be encountered in innocuous sources such as diet. Microbial glycans are also used as first steps in "antigen trapping", for example, by macrophage mannose receptor (Stahl and Ezekowitz 1998). The human macrophage mannose receptor binds mannose, fucose, and GlcNAc via multiple CRD's. Certain macrophages carry galactose-recognizing lectins, which bind to terminal glycans lacking sialic acid (Cherayil et al. 1990). Evading these innate immune receptors of the host forms part of the glycan strategies of numerous pathogens.

Several important pathogens have independently evolved synthetic pathways for host terminal glycan analogs such as *Pseudomonas'* pseudaminic acid and *Legionella'*s legionaminic acid, both 9-carbon sugar mimics of sialic acid (Knirel et al. 1986; Luneberg et al. 2000). Interestingly, these analog monosaccharides are added to the same underlying sugars as sialic acid is attached to in the host, effectively resulting in the mimicry of the entire terminal branch of highly abundant host glycans. Certain pathogens are capable of modulating these terminal glycans further to produce strains with particularly virulent and/or invasive properties (Lewis et al. 2006). Another important escape mechanism for microbial pathogens is "phase variation", where bacterial populations, including highly clonal populations can undergo radical changes in surface glycans based on the high mutation rate of repeat elements in their genomes (Appelmelk et al. 1998). As we discover novel aspects of endogenous glycan function, one of the obvious questions will have to concern the undiscovered ways by which pathogenic microbes are exploiting these host processes.

Fundamental asymmetries in evolvability between microbes and metazoans with respect to glycans

There is a fundamental asymmetry between the many endogenous constraints existing for glycan evolution in multicellular hosts and the relative lack thereof for most of their unicellular microbial pathogens. Naturally, microbial glycomes are also subject to certain constraints, such as the function of their capsule, the capacity to form biofilms or their susceptibility to infection by phages. Relatively speaking, however, bacteria probably can afford to undergo more rapid and dramatic glycan changes than their hosts. This asymmetry adds to the uneven playing field between multicellular hosts and their microbial pathogens with respect to their respective evolutionary rates. Microbes can evolve thousand to millions of times faster than their hosts. Furthermore, microbes can exchange genetic material across very distant lineages (Ochman et al. 2005), while such exchange in the form of sexual reproduction is limited to within species for multicellular host organisms. These long-distance genetic exchanges also mean that microbes, unlike their vertebrate hosts, are more likely to re-acquire glycan biosynthetic enzymatic capabilities that have previously been lost. The reoccurring distribution pattern of glycans restricted to vertebrates and their bacterial pathogens (Figure 1) (sialic acids, GAGs, and hyaluronan) could well result from such bacterial evolutionary plasticity. Sexual reproduction also incurs the cost of loss of optimal gene combination in each generation (Maynard Smith 1978) and multicellular hosts are capable of transiently altering glycan patterns, such as during the acute phase reaction (Brinkman-Van der Linden et al. 1996) or during lactation (Chaturvedi et al. 2001). Systemic changes in the glycome, however, are based on changes in allele frequencies of glycan biosynthesis genes and such changes take much longer for host populations than for their rapidly evolving microbial pathogens. Given these disadvantages of the hosts, one wonders why microbial pathogens have not already eliminated their hosts. Among the potential answers must be that different microbes compete for the same hosts, pathogens are forced to specialize for different hosts or cell types within a given host, and that levels of virulence often diminish over time as over-virulent pathogens run out of susceptible hosts.

Future directions and conclusions

The discontinuous distribution of glycan antigens across taxa of the living world is likely due to a combination of processes. Random events in the mutational history of each lineage may remove, add, or expand the capacity for the biosynthesis of particular glycans and sometimes, entire glycan classes. Subsequent interactions between organisms from different lineages in the form of host–pathogen or host–symbiont interactions can exert powerful selective pressure on the composition of lineage-specific glycomes. These lineage-specific glycomes can then be exploited for host recognition, pathogen recognition, and self-recognition. Their involvement in evolutionary arms races has the potential to both dramatically accelerate the rate of glycan evolution and to contribute to the maintenance of lineage-specific glycomes.

We would like to end this review by highlighting some key questions and proposing some testable hypothesis. The

pressing questions are as follows: (i) How much of the observed glycan variation in hosts is selectively neutral? (ii) What are the evolutionary time periods necessary for glycans to become targets of innate immunity? (iii) How much does the need to accommodate crucial symbionts constrain the adaptation of metazoan hosts away from pathogen exploitation? and (iv) Why are there not more pathogens pretending to be symbionts?

The following testable hypotheses come to mind: (i) there are likely many more cases of loss of particular glycan antigens in only isolated taxa of any given lineage (analogous to alpha-Gal and Neu5Gc loss); (ii) there are probably lineages, which have given up certain modified proteoglycans for similar reasons; (iii) loss of a particular glycan antigen is more likely when fewer important endogenous functions depend on such a glycan; (iv) following the loss of a particular glycan antigen, there are likely to be a series of compensatory changes in the endogenous lectins of the host species to accommodate for the loss of endogenous functions; (v) few pathogens can exploit many different glycans, i.e., most microbial pathogen are "specialists" in their "glycoecology", and (vi) species suffering from higher loads of enveloped viruses should maintain more glycan-based polymorphism.

Finally, we hope this review will motivate much needed research into the role of glycan diversity during host–pathogen/symbiont interactions at the individual level and at the level of entire host populations.

Acknowledgments

We thank Ajit Varki and Sébastien Gagneux for their many helpful comments and suggestions.

Conflict of interest statement

None declared.

Abbreviations

CRD, carbohydrate recognition domains; ER, endoplasmic reticulum; GAGs, glycosaminoglycans; GlcNAc, *N*-acetylglucosamine; Hib, *Haemophilus influenzae* b; PAMPs, pathogen associated molecular patterns; PRR, pathogen recognition receptor.

References

Aebi M, Hennet T. 2001. Congenital disorders of glycosylation: genetic model systems lead the way. Trends Cell Biol. 11:136–141.

Akira S, Takeda K, Kaisho T. 2001. Toll-like receptors: critical proteins linking innate and acquired immunity. Nat Immunol. 2:675–680.

Ang CW, Jacobs BC, Laman JD. 2004. The Guillain–Barre syndrome: a true case of molecular mimicry. Trends Immunol. 25:61–66.

Angata T, Varki A. 2002. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev. 102:439–470.

Appelmelk BJ, et al. 1998. Phase variation in *Helicobacter pylori* lipopolysaccharide. Infect Immun. 66:70–76.

Appelmelk BJ, van Die I, van Vliet SJ, Vandenbroucke-Grauls CM, Geijtenbeek TB, van Kooyk Y. 2003. Cutting edge: carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells. J Immunol. 170:1635–1639.

Bardor M, Faveeuw C, Fitchette AC, Gilbert D, Galas L, Trottein F, Faye L, Lerouge P. 2003. Immunoreactivity in mammals of two typical plant glyco-epitopes, core α(1,3)-fucose and core xylose. Glycobiology. 13: 427–434.

Bersudsky M, Rosenberg P, Rudensky B, Wirguin I. 2000. Lipopolysaccharides of a *Campylobacter coli* isolate from a patient with Guillain-Barre syndrome display ganglioside mimicry. Neuromuscul Disord. 10:182–186.

Brinkman-Van der Linden ECM, Van OECR, Van DW. 1996. Glycosylation of (alpha)1-acid glycoprotein in septic shock: changes in degree of branching and in expression of sialyl Lewis x groups. Glycoconj J. 13:27–31.

Brinkman-Van der Linden ECM, Sjoberg ER, Juneja LR, Crocker PR, Varki N, Varki A. 2000. Loss of *N*-glycolylneuraminic acid in human evolution—implications for sialic acid recognition by siglecs. J Biol Chem. 275: 8633–8640.

Bry L, Falk PG, Midtvedt T, Gordon JI. 1996. A model of host-microbial interactions in an open mammalian ecosystem. Science. 273:1380–1383.

Buckling A, Rainey PB. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc Biol Sci. 269:931–936.

Cash HL, Whitham CV, Behrendt CL, Hooper LV. 2006. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. Science. 313: 1126–1130.

Chaturvedi P, Warren CD, Altaye M, Morrow AL, Ruiz-Palacios G, Pickering LK, Newburg DS. 2001. Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. Glycobiology. 11:365–372.

Cherayil BJ, Chaitovitz S, Wong C, Pillai S. 1990. Molecular cloning of a human macrophage lectin specific for galactose. Proc Natl Acad Sci USA. 87:7324–7328.

Chui D, Sellakumar G, Green RS, Sutton-Smith M, McQuistan T, Marek KW, Morris HR, Dell A, Marth JD. 2001. Genetic remodeling of protein glycosylation in vivo induces autoimmune disease. Proc Natl Acad Sci USA. 98: 1142–1147.

Clausen H, Hakomori S. 1989. ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. Vox Sang. 56:1–20.

Colli W. 1993. Trans-sialidase: a unique enzyme activity discovered in the protozoan *Trypanosoma cruzi*. FASEB J. 7:1257–1264.

Crocker PR, Varki A. 2001. Siglecs, sialic acids and innate immunity. Trends Immunol. 22:337–342.

De Angelis PL. 1999. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. Cell Mol Life Sci. 56:670–682.

Dennis JW, Granovsky M, Warren CE. 1999. Protein glycosylation in development and disease. BioEssays. 21:412–421.

Doxiadis GG, Otting N, Antunes SG, de GNG, Harvey M, Doxiadis II, Jonker M, Bontrop RE. 1998. Characterization of the ABO blood group genes in macaques: evidence for convergent evolution. Tissue Antigens. 51:321–326.

Drickamer K, Taylor ME. 1993. Biology of animal lectins. Annu Rev Cell Biol. 9:237–264.

Du Pasquier L. 2004. Speculations on the origin of the vertebrate immune system. Immunol Lett. 92:3–9.

Dupuy F, Germot A, Marenda M, Oriol R, Blancher A, Julien R, Maftah A. 2002. 1,4-Fucosyltransferase activity: a significant function in the primate lineage has appeared twice independently. Mol Biol Evol. 19: 815–824.

Dwarakanath AD, Tsai HH, Sunderland D, Hart CA, Figura N, Crabtree JE, Rhodes JM. 1995. The production of neuraminidase and fucosidase by *Helicobacter pylori*: their possible relationship to pathogenicity. FEMS Immunol Med Microbiol. 12:213–216.

Esko JD, Lindahl U. 2001. Molecular diversity of heparan sulfate. J Clin Invest. 108:169–173.

Esko JD, Selleck SB. 2002. Order out of chaos: assembly of ligand binding sites in heparan sulfate. Annu Rev Biochem. 71:435–471.

Freeze HH, Westphal V. 2001. Balancing N-linked glycosylation to avoid disease. Biochimie. 83:791–799.

Freitas M, Axelsson LG, Cayuela C, Midtvedt T, Trugnan G. 2002. Microbial-host interactions specifically control the glycosylation pattern in intestinal mouse mucosa. Histochem Cell Biol. 118:149–161.

Gabius HJ. 1997. Animal lectins. Eur J Biochem. 243:543–576.

Gabius HJ, André S, Kaltner H, Siebert HC. 2002. The sugar code: functional lectinomics. Biochim Biophys Acta Gen Subj. 1572:165–177.

Gagneux P, Varki A. 1999. Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology. 9:747–755.

Galili U, Shohet SB, Kobrin E, Stults CL, Macher BA. 1988. Man, apes, and old world monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. J Biol Chem. 263:17755–17762.

Guerardel Y, et al. 2003. Identification of the blood group Lewis (a) determinant in the oviducal mucins of *Xenopus tropicalis*. FEBS Lett. 554(3): 330–336.

Gibbs MJ, Weiller GF. 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. Proc Natl Acad Sci USA. 96:8022–8027.

Gibeaut DM, Carpita NC. 1994. Biosynthesis of plant cell wall polysaccharides. FASEB J. 8:904–915.

Gilboa-Garber N, Garber N. 1989. Microbial lectin cofunction with lytic activities as a model for a general basic lectin role. FEMS Microbiol Rev. 5:211–221.

Grobe K, Ledin J, Ringvall M, Holmborn K, Forsberg E, Esko JD, Kjellén L. 2002. Heparan sulfate and development: differential roles of the *N*-acetylglucosamine *N*-deacetylase/*N*-sulfotransferase isozymes. Biochim Biophys Acta Gen Subj. 1573:209–215.

Haltiwanger RS, Lowe JB. 2004. Role of glycosylation in development. Annu Rev Biochem. 73:491–537.

Harduin-Lepers A, Mollicone R, Delannoy P, Oriol R. 2005. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. Glycobiology. 15:805–817.

Harvey HA, Porat N, Campbell CA, Jennings M, Gibson BW, Phillips NJ, Apicella MA, Blake MS. 2000. Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm. Mol Microbiol. 36:1059–1070.

Harvey HA, Swords WE, Apicella MA. 2001. The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus. J Autoimmun. 16:257–262.

Hedrick SM. 2004. The acquired immune system: a vantage from beneath. Immunity. 21:607–615.

Hooper LV, Gordon JI. 2001. Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. Glycobiology. 11:1R–10R.

Janeway CA Jr, Medzhitov R. 2002. Innate immune recognition. Annu Rev Immunol. 20:197–216.

Javaud C, Dupuy F, Maftah A, Julien R, Petit JM. 2003. The fucosyltransferase gene family: an amazing summary of the underlying mechanisms of gene evolution. Genetica. 118:157–170.

Jennemann R, Gnewuch C, Bosslet S, Bauer BL, Wiegandt H. 1994. Specific immunization using keyhole limpet hemocyanin-ganglioside conjugates. J Biochem (Tokyo). 115:1047–1052.

Kaltner H, Stierstorfer B. 1998. Animal lectins as cell adhesion molecules. Acta Anat (Basel). 161:162–179.

Kandler O, Konig H. 1998. Cell wall polymers in Archaea (Archaebacteria). Cell Mol Life Sci. 54:305–308.

Karlsson KA. 1995. Microbial recognition of target-cell glycoconjugates. Curr Opin Struct Biol. 5:622–635.

Kawabata S, Tsuda R. 2002. Molecular basis of non-self recognition by the horseshoe crab tachylectins. Biochim Biophys Acta Gen Subj. 1572: 414–421.

Kelly DF, Moxon ER, Pollard AJ. 2004. *Haemophilus influenzae* type b conjugate vaccines. Immunology. 113:163–174.

Kermarrec N, Roubinet F, Apoil PA, Blancher A. 1999. Comparison of allele O sequences of the human and non-human primate ABO system. Immunogenetics. 49:517–526.

Kilpatrick DC. 2002. Animal lectins: a historical introduction and overview. Biochim Biophys Acta Gen Subj. 1572:187–197.

Kirschner M, Gerhart J. 1998. Evolvability. Proc Natl Acad Sci USA. 95: 8420–8427.

Knirel YA, Vinogradov EV, Shashkov AS, Dmitriev BA, Kochetkov NK, Stanislavsky ES, Mashilova GM. 1986. Somatic antigens of *Pseudomonas aeruginosa*. The structure of O-specific polysaccharide chains of *P. aeruginosa* O10 (Lanyi) lipopolysaccharides. Eur J Biochem. 157:129–138.

Knirel YA, Kocharova NA, Shashkov AS, Dmitriev BA, Kochetkov NK, Stanislavsky ES, Mashilova GM. 1987. Somatic antigens of *Pseudomonas aeruginosa*. The structure of O-specific polysaccharide chains of the lipopolysaccharides from *P. aeruginosa* O5 (Lanyi) and immunotype 6 (Fisher). Eur J Biochem. 163:639–652.

Koch AL. 2000. The exoskeleton of bacterial cells (the sacculus): still a highly attractive target for antibacterial agents that will last for a long time. Crit Rev Microbiol. 26:1–35.

Kooistra O, Lüneberg E, Lindner B, Knirel YA, Frosch M, Zähringer U. 2001. Complex *O*-acetylation in *Legionella pneumophila* serogroup 1 lipopolysaccharide. Evidence for two genes involved in 8-*O*-acetylation of legionaminic acid. Biochemistry. 40:7630–7640.

Kotani N, Asano M, Iwakura Y, Takasaki S. 2001. Knockout of mouse beta 1,4-galactosyltransferase-1 gene results in a dramatic shift of outer chain moieties of *N*-glycans from type 2 to type 1 chains in hepatic membrane and plasma glycoproteins. Biochem J. 357:827–834.

Kudo T, et al. 2006. Mice lacking α1,3-fucosyltransferase IX demonstrate disappearance of Lewis x structure in brain and increased anxiety-like behaviors. Glycobiology. 17:1–9.

Kyogashima M, Ginsburg V, Krivan HC. 1989. *Escherichia coli* K99 binds to *N*-glycolylsialoparagloboside and *N*-glycolyl-GM3 found in piglet small intestine. Arch Biochem Biophys. 270:391–397.

Lespinet O, Wolf YI, Koonin EV, Aravind L. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12:1048–1059.

Lewis AL, Hensler ME, Varki A, Nizet V. 2006. The group B Streptococcal sialic acid *O*-acetyltransferase is encoded by neuD, a conserved component of bacterial sialic acid biosynthetic gene clusters. J Biol Chem. 281: 11186–11192.

Lewontin RC. 2002. Directions in evolutionary biology. Annu Rev Genet. 36: 1–18.

Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 124:837–848.

Liedtke S, Adamski M, Geyer R, Pfützner A, Rübsamen-Waigmann H, Geyer H. 1994. Oligosaccharide profiles of HIV-2 external envelope glycoprotein: dependence on host cells and virus isolates. Glycobiology. 4:477–484.

Lowe JB, Marth JD. 2003. A genetic approach to mammalian glycan function. Annu Rev Biochem. 72:643–691.

Luneberg E, Zetzmann N, Alber D, Knirel YA, Kooistra O, Zahringer U, Frosch M. 2000. Cloning and functional characterization of a 30 kb gene locus required for lipopolysaccharide biosynthesis in *Legionella pneumophila*. Int J Med Microbiol. 290:37–49.

Magre S, Takeuchi Y, Bartosch B. 2003. Xenotransplantation and pig endogenous retroviruses. Rev Med Virol. 13:311–329.

Manzi AE, Norgard-Sumnicht K, Argade S, Marth JD, van HH, Varki A. 2000. Exploring the glycan repertoire of genetically modified mice by isolation and profiling of the major glycan classes and nano-NMR analysis of glycan mixtures. Glycobiology. 10:669–689.

Marionneau S, Cailleau-Thomas A, Rocher J, Le M-VB, Ruvoën N, Clément M, Le PJ. 2001. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. Biochimie. 83:565–573.

Martin SL, Edbrooke MR, Hodgman TC, Van dEDH, Bird MI. 1997. Lewis X biosynthesis in *Helicobacter pylori*—molecular cloning of an alpha (1,3)-fucosyltransferase gene. J Biol Chem. 272:21349–21356.

Maynard Smith J. 1978. The evolution of sex. Cambridge University Press, Cambridge, UK.

Monteiro MA, et al. 1998. Simultaneous expression of type 1 and type 2 Lewis blood group antigens by *Helicobacter pylori* lipopolysaccharides. J Biol Chem. 273:11533–11543.

Natsuka S, Lowe JB. 1994. Enzymes involved in mammalian oligosaccharide biosynthesis. Curr Opin Struct Biol. 4:683–691.

Neil SJ, McKnight A, Gustafsson K, Weiss RA. 2005. HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation. Blood. 105:4693–4699.

Nguyen DH, Tangvoranuntakul P, Varki A. 2005. Effects of natural human antibodies against a nonhuman sialic acid that metabolically incorporates into activated and malignant immune cells. J Immunol. 175:228–236.

Ochman H, Lerat E, Daubin V. 2005. Examining bacterial species under the specter of gene transfer and exchange. Proc Natl Acad Sci USA. 102 Suppl 1:6595–6599.

Ohtsubo K, Marth JD. 2006. Glycosylation in cellular mechanisms of health and disease. Cell. 126:855–867.

Olofsson S, Bergstrom T. 2005. Glycoconjugate glycans as viral receptors. Ann Med. 37:154–172.

Oriol R, Mollicone R, Cailleau A, Balanzino L, Breton C. 1999. Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. Glycobiology. 9:323–334.

Palma AS, Vila-Verde C, Pires AS, Fevereiro PS, Costa J. 2001. A novel plant alpha 1,4-fucosyltransferase (*Vaccinium myrtillus* L.) synthesises the Lewis(a) adhesion determinant. FEBS Lett. 499:235–238.

Pangburn MK, Pangburn KL, Koistinen V, Meri S, Sharma AK. 2000. Molecular mechanisms of target recognition in an innate immune system: interactions among factor H, C3b, and target in the alternative pathway of human complement. J Immunol. 164:4742–4751.

Paulson JC, Colley KJ. 1989. Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. J Biol Chem. 264: 17615–17618.

Perrier C, Sprenger N, Corthesy B. 2006. Glycans on secretory component participate in innate protection against mucosal pathogens. J Biol Chem. 281:14280–14287.

Preece AF, Strahan KM, Devitt J, Yamamoto F, Gustafsson K. 2002. Expression of ABO or related antigenic carbohydrates on viral envelopes leads to neutralization in the presence of serum containing specific natural antibodies and complement. Blood. 99:2477–2482.

Probstmeier R, Pesheva P. 1999. I-type lectins in the nervous system. Prog Neurobiol. 58:163–184.

Purvis A. 1995. A composite estimate of primate phylogeny. Philos Trans R Soc Lond B Biol Sci. 348:405–421.

Raman R, Raguram S, Venkataraman G, Paulson JC, Sasisekharan R. 2005. Glycomics: an integrated systems approach to structure-function relationships of glycans. Nat Methods. 2:817–824.

Repik PM, Strizki JM, Galili U. 1994. Differential host-dependent expression of alpha-galactosyl epitopes on viral glycoproteins: a study of eastern equine encephalitis virus as a model. J Gen Virol. 75:1177–1181.

Rini JM, Lobsanov YD. 1999. New animal lectin structures. Curr Opin Struct Biol. 9:578–584.

Rostand KS, Esko JD. 1997. Microbial adherence to and invasion through proteoglycans. Infect Immun. 65:1–8.

Roth J. 1996. Protein glycosylation in the endoplasmic reticulum and the Golgi apparatus and cell type specificity of cell surface glycoconjugate expression: analysis by the protein A-gold and lectin-gold techniques. Histochem Cell Biol. 106:79–92.

Rother RP, Squinto SP. 1996. The alpha-galactosyl epitope: a sugar coating that makes viruses and cells unpalatable. Cell. 86:185–188.

Schachter H. 2002. The role of the GlcNAcbeta1,2Manalpha- moiety in mammalian development. Null mutations of the genes encoding UDP-*N*-acetylglucosamine: alpha-3-D-mannoside beta-1,2-*N*-acetylglucosaminyltransferase I and UDP-*N*-acetylglucosamine: alpha-D-mannoside beta-1,2-*N*-acetylglucosaminyltransferase I.2 cause embryonic lethality and congenital muscular dystrophy in mice and men, respectively. Biochim Biophys Acta Gen Subj. 1573:292–300.

Schaffer C, Graninger M, Messner P. 2001. Prokaryotic glycosylation. Proteomics. 1:248–261.

Schauer R. 1988. Sialic acids as antigenic determinants of complex carbohydrates. Adv Exp Med Biol. 228:47–72.

Seymour RM, Allan MJ, Pomiankowski A, Gustafsson K. 2004. Evolution of the human ABO polymorphism by two complementary selective pressures. Proc Biol Sci. 271:1065–1072.

Sharon N. 1996. Carbohydrate-lectin interactions in infectious disease. Adv Exp Med Biol. 408:1–8.

Shreffler WG, Castro RR, Kucuk ZY, Charlop-Powers Z, Grishina G, Yoo S, Burks AW, Sampson HA. 2006. The major glycoprotein allergen from *Arachis hypogaea*, Ara h 1, is a ligand of dendritic cell-specific ICAM-grabbing nonintegrin and acts as a Th2 adjuvant in vitro. J Immunol. 177:3677–3685.

Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI. 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science. 307:1955–1959.

Spear PG, Shieh MT, Herold BC, WuDunn D, Koshy TI. 1992. Heparin sulfate glycosaminoglycans as primary cell surface receptors for herpes simplex virus. Adv Exp Med Biol. 313:341–353.

Spiro RG. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology. 12:43R–56R.

Stahl PD, Ezekowitz RA. 1998. The mannose receptor is a pattern recognition receptor involved in host defense. Curr Opin Immunol. 10:50–55.

Staudacher E, Altmann F, Wilson IBH, März L. 1999. Fucose in *N*-glycans: from plant to man. Biochim Biophys Acta Gen Subj. 1473:216–236.

Stearns SC, Hoekstra R. 2005. Evolution, an introduction. 2nd ed. Oxford University Press, Oxford, UK.

Suzuki N, Laskowski M Jr, Lee YC. 2004. Phylogenetic expression of Gal alpha1–4Gal on avian glycoproteins: glycan differentiation inscribed in the early history of modern birds. Proc Natl Acad Sci USA. 101: 9023–9028.

Suzuki Y, Ito T, Suzuki T, Holland REJ, Chambers TM, Kiso M, Ishida H, Kawaoka Y. 2000. Sialic acid species as a determinant of the host range of influenza A viruses. J Virol. 74:11825–11831.

Varki A. 1993. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 3:97–130.

Varki A, Cummings R, Esko JD, Freeze HF, Hart GW, Marth JD. 1999. Essentials of glycobiology. Cold Spring Harbor Laboratory Press, Plainview, NY.

Varki A. 2001. Loss of *N*-glycolylneuraminic acid in humans: mechanisms, consequences and implications for hominid evolution. Am J Phys Anthropol. Suppl 33:54–69.

Varki A, Angata T. 2006. Siglecs—the major subfamily of I-type lectins. Glycobiology. 16:1R–27R.

Varki A. 2006. Nothing in glycobiology makes sense, except in the light of evolution. Cell 126:841–845.

Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. 2004. Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev. 68:132–153.

Wadström T, Ljungh A. 1999. Glycosaminoglycan-binding microbial proteins in tissue adhesion and invasion: key events in microbial pathogenicity. J Med Microbiol. 48:223–233.

Wagner GP. 1994. Evolution and multi-functionality of the chitin system. EXS. 69:559–577.

Weis WI, Taylor ME, Drickamer K. 1998. The C-type lectin superfamily in the immune system. Immunol Rev. 163:19–34.

Weiss RA. 2000. Certain promise and uncertain peril. The debate on xenotransplantation. EMBO Rep. 1:2–4.

Yoo D, Giulivi A. 2000. Xenotransplantation and the potential risk of xenogeneic transmission of porcine viruses. Can J Vet Res. 64: 193–203.

Zhu Y, Ng PM, Wang L, Ho B, Ding JL. 2006. Diversity in lectins enables immune recognition and differentiation of wide spectrum of pathogens. Int Immunol. 18:1671–1680.