Cell

Nothing in Glycobiology Makes Sense, except in the Light of Evolution

Ajit Varki[1,*]
[1]Glycobiology Research and Training Center, Departments of Medicine and Cellular & Molecular Medicine, University of California at San Diego, La Jolla, CA 92093, USA
*Contact: a1varki@ucsd.edu
DOI 10.1016/j.cell.2006.08.022

The remarkable structural diversity of glycans in nature, and their roles in cellular processes, host-pathogen interactions, biological diversity and speciation can be explained by evolutionary processes.

The title of this Essay is a slight modification of the title of a classic article about evolution (Dobzhansky, 1973). Indeed, one can substitute many other biological terms in place of "glycobiology" and still have an appropriate title for an interesting discussion. All cells in nature are covered with a dense coating of glycans that is important not only for the biological processes of hosts but also for the binding of pathogens to them. Thus, glycans may be trapped in neverending cycles of evolutionary "Red Queen" effects in which long-lived hosts must evade the more rapidly evolving pathogens that infect them by changing their glycan expression patterns, without compromising their own survival (Van Valen, 1974; Hamilton et al., 1990). The colorful term, "Red Queen" effect, recalls the comment to Alice by the Red Queen in Lewis Carroll's *Through the Looking Glass* that "it takes all the running you can do, to keep in the same place." This may explain the remarkable structural variations of glycans in nature, which contribute to biological diversity and perhaps even to speciation.

All classes of biological molecules undergo evolution by neutral processes, natural selection, and/or sexual selection. This Essay examines how glycans—the oligosaccharide chains of sugars attached to many proteins and lipids—are prone to undergo rapid evolutionary change (Gagneux and Varki, 1999). I suggest here that rapid evolution of glycans driven by infectious agents could possibly mediate speciation (the formation of new species), that is, pathogen selection could be an even stronger force of natural selection than previously recognized.

Two Distinct Worlds of Cellular Glycosylation

Until the 1980s, it was assumed that glycosylation was only found on secreted molecules, on the exterior leaflet of the plasma membrane, or on intracellular organelle membranes with the same orientation, that is, facing away from the cytosol. This dogma was demolished by the discovery of glycosylation on many nuclear and cytosolic proteins (Hart et al., 1989). However, the inside and the outside of a cell are subject to rather different evolutionary selection forces—with internal mechanisms generally reflecting conservation and stability, and external processes characterized by dynamic and frequent evolutionary change. Indeed, the study of these two different biological spheres should perhaps be termed "endobiology" and "exobiology," respectively. Thus, it is not surprising that these two worlds of cellular glycosylation are also quite distinct, both with regard to the nature of the glycan structures and the types of functions such structures mediate. For example, although extracellular glycans can mediate cell-cell and host-pathogen interactions (Sharon, 1996; Esko and Selleck, 2002), intracellular glycans can serve as dynamic regulatory switches, often competing with protein modifications such as phosphorylation (Hart et al., 1989). This Essay focuses on the "exobiology" of glycosylation.

Patterns of Cell-Surface Glycan Expression

Francis Crick once observed that there are no (absolute) laws in biology—only gadgets (comment made at a 2002 symposium on human evolution; http://origins.ucsd.edu/Nov02). One exception may be that all cells in nature are covered with a dense and complex coat of glycans. There are probably many reasons for this including the diversity, complexity, hydrophilicity, and structural mobility of cell-surface glycans, as well as their combinatorial nature, which provides the possibility for rapid change to escape the pathogens that bind to them (Gagneux and Varki, 1999).

Most glycan structures found on the cell surface and on secreted molecules are synthesized in the endoplasmic reticulum (ER) and Golgi apparatus, in an assembly-line-like process that is not template driven and is subject to multiple sequential and competitive enzymatic pathways (Drickamer and Taylor, 1998; Esko and Selleck, 2002). Partly as a consequence of this "analog" mechanism of synthesis, the array of glycans on a given cell surface is difficult to predict, based on gene expression patterns alone. The glycans can also change dynamically, responding to small variations in the extracellular environment and intracellular events. Regardless of this, each cell type in

each organism expresses a distinct array of glycans under defined conditions, and these expression patterns tend to be conserved within the same species. Moreover, these cell-type-specific glycan expression patterns are subject to striking and stereotypic species-specific changes during development, e.g., the evolutionarily conserved switch from peanut-agglutinin negative to positive during thymic development of T cells (Lowe and Marth, 2003). These findings suggest that the expression of glycans is under strict regulatory control.

Phenotypic Variation Despite Genetic Conservation

The regulated expression of a limited set of genes encoding glycosyltransferases and glycan-modifying enzymes appears to control most cell-type-specific glycosylation patterns. These enzymes use specific high-energy nucleotide donors to carry out glycosylation reactions within the ER-Golgi pathway (Esko and Selleck, 2002; Drickamer and Taylor, 1998) and are conserved among related taxa. For example, the sialyltransferases that mediate the transfer of terminal sialic acid residues to glycans are conserved between mice and humans, not only in terms of specific orthologs, but also in their primary sequence and in their enzymatic specificity for acceptor substrates (Harduin-Lepers et al., 2005). The same is true of the genes involved in glycosaminoglycan biosynthesis (Esko and Selleck, 2002). Thus, evolution appears to have favored conservation of these gene sequences and maintenance of distinct sets of such enzymes (Angata and Varki, 2002).

Despite the conservation of glycosylation genes, one can find major intra- and interspecies variations in glycosylation patterns (Gagneux and Varki, 1999). Variations within a given species tend to be limited but can be dramatic, such as the ABO blood groups within humans and other primates (Hakomori, 1999). Available information indicates that interspecies variations in glycosylation are more extensive. On rare occasions, a specific type of terminal glycan can be permanently eliminated in an entire evolutionary clade by inactivation of a specific glycosylation-related gene. A good example is the loss of the terminal α-galactose epitope (Galα1-3Galβ1-4GlcNAcβ1-R) in old world primates (Galili et al., 1987), and the human-specific loss of the common mammalian sialic acid Neu5Gc (N-glycolylneuraminic acid) (Angata and Varki, 2002). However, even when comparing orthologous sets of conserved glycosyltransferases in closely related species such as mice and rats, one can find dramatic differences in the expression patterns of many terminal glycan structures (N. Varki, personal communication). Furthermore, the same conserved glycoprotein can carry markedly different glycan structures in closely related species (Gagneux and Varki, 1999). Taken together with the conservation of glycosyltransferase genes, this suggests that these marked differences are mediated by differential control of the individual enzymes in different species. Indeed, limited evidence suggests that this is the case, and that regulation of gene expression and/or posttranscriptional mechanisms govern these differences.

Pathogen Recognition of Host Glycans

A wide variety of pathogens initiate infection by binding to the surface glycans of host cells (Sharon, 1996). This is not surprising as cell-surface glycans are the first molecules encountered by pathogens when they contact potential host cells or their secretions. Outer terminal glycan sequences such as those carrying sialic acid residues are even more likely to be preferred targets, as they are the first residues that pathogens encounter. Examples include influenza virus infection of the lung, erythrocyte invasion by the malaria parasite *Plasmodium falciparum*, *Helicobacter pylori* infection of the stomach, and intestinal diarrhea caused by the toxin of *Vibrio cholerae* (Mammen et al., 1998; Angata and Varki, 2002). Why are glycosylation genes and biosynthetic pathways so highly conserved if they generate glycan targets for many potentially deadly pathogens? It appears that individual cell types or whole species have committed themselves to important endogenous functions for such glycans, often in a cell-type-specific fashion, thus constraining their ability to discard glycans (Lowe and Marth, 2003). For example, inactivation of specific sialyltransferases in the mouse results in cell-type-specific defects, such as loss of certain T cell

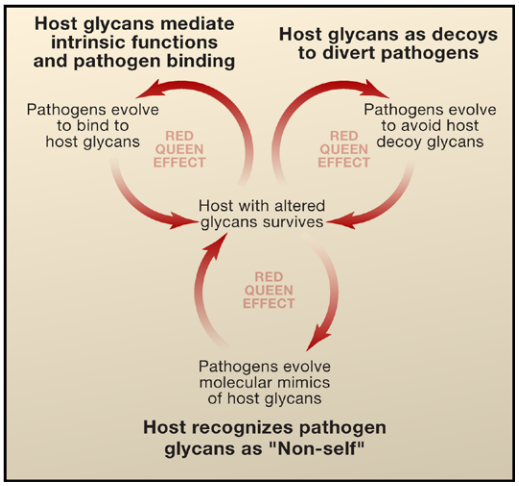

Figure 1. Evolutionary Diversification of Glycans

Each arrowed circle represents a potential vicious cycle, driven by a Red Queen effect, in which hosts are constantly trying to evade the more rapidly evolving pathogens that infect them (Gagneux and Varki, 1999). Hosts require glycans for critical cellular functions but must constantly change them to evade glycan binding pathogens yet without impairing their own survival. Hosts also produce soluble glycans such as mucins, which act as decoys to divert pathogens from cell surfaces; but pathogens are constantly adjusting to these defenses. Hosts recognize pathogen-specific glycans as markers of "nonself," but pathogens can modify their glycans to more closely mimic host glycans. There are also possible secondary Red Queen effects involving host glycan binding proteins that recognize "self" (Varki and Angata, 2006). In each of these cycles, hosts with altered glycans that can still carry out adequate cellular functions are most likely to survive.

subsets or blood coagulation abnormalities (Lowe and Marth, 2003). In contrast, complete elimination of any major glycan class results in death of the mouse embryo (Schwarzkopf et al., 2002; Esko and Selleck, 2002; Lowe and Marth, 2003). Meanwhile, glycan binding pathogens are evolving much more rapidly than their hosts, by virtue of their short generation times and high mutation rates. Thus, the glycans of complex multicellular organisms with long life cycles may be subject to evolutionary "Red Queen" effects (Figure 1), in which such organisms must evolve rapidly to survive the onslaught of microbial pathogens that can replicate (and thus evolve) even faster (Van Valen, 1974; Hamilton et al., 1990).

Complex multicellular organisms may change their glycan profiles frequently to escape the pathogens that are tracking them (Figure 1). Indeed, the same scenario is apparently being played out even in single-celled prokaryotes, in which the bacteria's own pathogens (phages) often mediate attachment by recognition of bacterial surface polysaccharides. Further complexity arises from the fact that soluble secreted glycan-bearing molecules (such as the mucin molecules in mucus or serum glycoproteins) can potentially act as decoys, diverting the binding of pathogens away from target cell surfaces (Gagneux and Varki, 1999; Perrier et al., 2006). With viruses, we have suggested that even non-nucleated cells such as mammalian erythrocytes could potentially act as glycan decoys. Of course, pathogens can also evolve rapidly away from such decoys, or even start to exploit them, fueling additional bouts of such arms races or cycles of "Red Queen" effects (Figure 1).

Other Evolutionary Forces Driving Diversification of Glycan Expression

There are many other evolutionary forces potentially and simultaneously driving the diversification of glycan expression. For example, cell-type-specific expression of certain glycans can mediate specific biological roles within an organism (Lowe and Marth, 2003), which thus may be under positive selection. Also, the sudden elimination of a glycan in an entire evolutionary lineage (clade)—such as the loss of α-galactose residues in old world primates—can be accompanied by the spontaneous expression of high levels of complement-fixing antibodies directed against the same structure (Galili et al., 1987). Such antibodies have been proposed to restrict lateral transmission of enveloped viruses, as these usually carry along the surface glycan structures of the organisms from which they originate (Takeuchi et al., 1996). For example, a new world monkey virus could have its envelope glycoproteins extensively modified with α-galactose residues, resulting in killing of the virus upon first contact with the blood of an old world primate such as a human. This may help to explain the contrast between the many zoonotic old world primate viruses that cause human diseases and the rarity of human diseases caused by new world monkey viruses (P. Gagneux, personal communication). Additional Red Queen effects may arise from the fact that many pathogens use a successful form of molecular mimicry, disguising themselves with cell-surface glycans similar to those of their hosts (Figure 1). Interestingly, most of these instances do not represent lateral gene transfer from vertebrates but rather involve recruitment of pre-existing homologous genes or possibly convergent evolution (reinventions). These and other forces are likely to have further driven the diversification of species-specific glycosylation patterns during evolution. Meanwhile, host glycan binding proteins such as siglecs (sialic-acid binding immunoglobulin-like lectins) dedicated to recognizing self also need to evolve rapidly in order to keep up with the ever evolving host "glycome" (Crocker, 2005; Varki and Angata, 2006). (There are many possible definitions of the term "glycome," ranging from a static and simple parts list of types of glycans found in a given organism to a dynamic array of presentations of glycans that vary with individual cell types and change with time, space, and environmental conditions. I use the term glycome to mean the latter.)

Glycans Are Intimately Involved in Reproductive Biology

There is compelling evidence that cell-surface and secreted glycans are involved in many aspects of reproductive biology such as the coating of germ cells, the passage of sperm through the female reproductive tract, maturation of sperm, attachment of sperm to female reproductive surfaces, fertilization, and implantation (Mengerink and Vacquier, 2001; Diekman, 2003; Lapid and Sharon, 2006). Some biologists remain unconvinced that glycosylation plays a key role in reproductive biology, perhaps because of the diverse and confusing findings reported to date and the apparent lack of conservation of expression patterns of most of the glycans involved. Also, some genetic mutations have not produced the predicted loss of fertility (Lowe and Marth, 2003). Due to its key role in evolution, reproduction is a rapidly evolving biological process for which one should actually expect to find marked diversification and redundancy of function. Such diversification and redundancy may be driven by sexual selection and sexual conflict, and by the need to isolate the fertilization event to prevent unwanted crossfertilization by other related species, while limiting the opportunity for infection by sexually transmitted pathogens. Unwanted crossfertilization is of special importance to species with external fertilization (such as echinoderms and fish), and sexually transmitted pathogens are of particular importance to species with internal fertilization (such as mammals).

Could Selection by Glycan Binding Pathogens Drive Speciation?

Much is written about the role of rapid evolutionary processes in the diversification of symbionts and pathogens and their relationships to hosts (Lederberg, 1999). It is assumed that infectious diseases can have remarkable selective effects in gen-

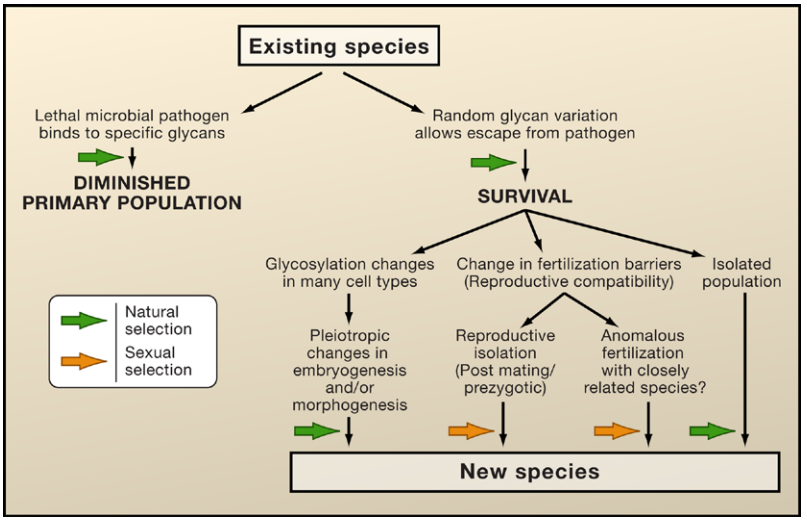

Figure 2. Glycan Binding Pathogens Might Facilitate Speciation
Exposure of a population to a lethal glycan binding pathogen could initiate one or more of the mechanisms shown. A lethal pathogen that binds to glycans of a certain species to initiate infection might markedly reduce the primary population, leaving geographically isolated subpopulations that have the opportunity to evolve into new species. Some of the survivors are likely to have been selected because of random glycan variations that allowed them to escape from infection with the pathogen. Such survivors may have secondary alterations in glycan biology that are permissive for pleiotropic changes in embryogenesis and/or morphogenesis. Glycan changes also might alter fertilization barriers, causing either reproductive isolation or anomalous fertilization by closely related species. Any or all of these mechanisms could support the formation of new species. Many of these speculative ideas are testable by observational studies and possibly by long-term experiments.

erating intraspecies polymorphisms, such as sickle cell disease resulting from selection by malaria (Lederberg, 1999). Despite this, little attention has been given to the notion that such pathogen-mediated selection might contribute to driving the speciation process itself. The reason for this is that infectious diseases usually do not directly affect the germline, nor do they affect specific primary steps of reproduction, such as mating and fertilization.

But the regulation of glycan production can potentially provide a mechanism for linking the processes of natural selection and sexual selection (Figure 2). Thus, infection of a population by a deadly pathogen that binds to specific host glycans could result in survival of only a few individuals whose glycosylation patterns happen to be altered due to genomic or transcriptional changes. This type of altered glycosylation pattern might in turn also be manifest in the reproductive tract or on germ cells, leading to significant changes in fertility and the fertilization process, for example by blocking fertilization by the original population allowing sympatric speciation (speciation without geographic separation from the ancestor), or perhaps contributing to anomalous fertilization by closely related species. Meanwhile, if such glycosylation changes eventually become fixed in other cell types, this could result in pleiotropic changes affecting embryogenesis and morphogenesis, further contributing to speciation. Marked reductions in population size by deadly pathogens could also leave isolated pockets of survivors, helping to drive allopatric speciation (speciation after geographic separation from the ancestor) via such founder events (Figure 2). Scenarios can even be envisioned in which female antibodies against sperm bound glycans could help to drive speciation by generating postmating prezygotic isolation, such as selective elimination of sperm after they have entered the female reproductive tract (P. Gagneux, personal communication). Overall, one can speculate that some episodes of speciation might be mediated by pathogens that bind to specific forms of cell-surface glycans.

Biology Is a Snapshot in Evolutionary Time

Whatever the reader might think of the speculations in this Essay, it is safe to suggest that approaches to understanding glycan biology must fully take into account the role of multiple and often simultaneous evolutionary processes, an aspect that has received limited attention. Indeed, almost 150 years after Darwin's revelation and more than 50 years after the Neo-Darwinian synthesis, some biologists still assume that natural selection through "survival of the fittest" (a term never actually used by Darwin) has already honed most biological processes to near perfection. In reality, we biologists are simply studying a brief window of biological time, which represents the present status of trade-offs reached by currently living organisms subject to a number of evolutionary forces: positive selection, neutral drift, purifying selection, and sexual selection, all taking place under ever-changing biotic and abiotic environments. These forces are all acting jointly but nondirectionally, diversifying and complicating biological processes in unexpected ways, including the process of speciation.

ACKNOWLEDGMENTS

I thank P. Gagneux for enlightening discussions and N. Varki and T. Angata for helpful comments.

REFERENCES

Angata, T., and Varki, A. (2002). Chem. Rev. *102*, 439–470.

Crocker, P.R. (2005). Curr. Opin. Pharmacol. *5*, 431–437.

Diekman, A.B. (2003). Cell. Mol. Life Sci. *60*, 298–308.

Dobzhansky, T. (1973). Am. Biol. Teach. *35*, 125–129.

Drickamer, K., and Taylor, M.E. (1998). Trends Biochem. Sci. *23*, 321–324.

Esko, J.D., and Selleck, S.B. (2002). Annu. Rev. Biochem. *71*, 435–471.

Gagneux, P., and Varki, A. (1999). Glycobiology *9*, 747–755.

Galili, U., Clark, M.R., Shohet, S.B., Buehler, J., and Macher, B.A. (1987). Proc. Natl. Acad. Sci. USA *84*, 1369–1373.

Hakomori, S. (1999). Biochim. Biophys. Acta Gen. Subj. *1473*, 247–266.

Hamilton, W.D., Axelrod, R., and Tanese, R. (1990). Proc. Natl. Acad. Sci. USA *87*, 3566–3573.

Harduin-Lepers, A., Mollicone, R., Delannoy, P., and Oriol, R. (2005). Glycobiology *15*, 805–817.

Hart, G.W., Haltiwanger, R.S., Holt, G.D., and Kelly, W.G. (1989). Annu. Rev. Biochem. *58*, 841–874.

Lapid, K., and Sharon, N. (2006). Glycobiology *16*, 39R–45R.

Lederberg, J. (1999). Genetics *153*, 1–3.

Lowe, J.B., and Marth, J.D. (2003). Annu. Rev. Biochem. *72*, 643–691.

Mammen, M., Choi, S.-K., and Whitesides, G.M. (1998). Angew. Chem. Int. Ed. Engl. *37*, 2754–2794.

Mengerink, K.J., and Vacquier, V.D. (2001). Glycobiology *11*, 37R–43R.

Perrier, C., Sprenger, N., and Corthesy, B. (2006). J. Biol. Chem. *281*, 14280–14287.

Schwarzkopf, M., Knobeloch, K.P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002). Proc. Natl. Acad. Sci. USA *99*, 5267–5270.

Sharon, N. (1996). Adv. Exp. Med. Biol. *408*, 1–8.

Takeuchi, Y., Porter, C.D., Strahan, K.M., Preece, A.F., Gustafsson, K., Cosset, F.L., Weiss, R.A., and Collins, M.K.L. (1996). Nature *379*, 85–88.

Van Valen, L. (1974). Nature *252*, 298–300.

Varki, A., and Angata, T. (2006). Glycobiology *16*, 1R–27R.