Glycobiology vol. 14 no. 4 pp. 339–346, 2004
DOI: 10.1093/glycob/cwh039
Advance Access publication on December 23, 2003

A uniquely human consequence of domain-specific functional adaptation in a sialic acid–binding receptor

Justin L. Sonnenburg[2,3], Tasha K. Altheide[3], and Ajit Varki[1,3]

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

Received on November 15, 2003; revised on November 19, 2003; accepted on November 24, 2003

Most mammalian cell surfaces display two major sialic acids (Sias), N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc). Humans lack Neu5Gc due to a mutation in CMP-Neu5Ac hydroxylase, which occurred after evolutionary divergence from great apes. We describe an apparent consequence of human Neu5Gc loss: domain-specific functional adaptation of Siglec-9, a member of the family of sialic acid–binding receptors of innate immune cells designated the CD33-related Siglecs (CD33rSiglecs). Binding studies on recombinant human Siglec-9 show recognition of both Neu5Ac and Neu5Gc. In striking contrast, chimpanzee and gorilla Siglec-9 strongly prefer binding Neu5Gc. Simultaneous probing of multiple endogenous CD33rSiglecs on circulating blood cells of human, chimp, or gorilla suggests that the binding differences observed for Siglec-9 are representative of multiple CD33rSiglecs. We conclude that Neu5Ac-binding ability of at least some human CD33rSiglecs is a derived state selected for following loss of Neu5Gc in the hominid lineage. These data also indicate that endogenous Sias (rather than surface Sias of bacterial pathogens) are the functional ligands of CD33rSiglecs and suggest that the endogenous Sia landscape is the major factor directing evolution of CD33rSiglec binding specificity. Exon-1-encoded Sia-recognizing domains of human and ape Siglec-9 share only ~93–95% amino acid identity. In contrast, the immediately adjacent intron and exon 2 have the ~98–100% identity typically observed among these species. Together, our findings suggest ongoing adaptive evolution specific to the Sia-binding domain, possibly of an episodic nature. Such domain-specific divergences should also be considered in upcoming comparisons of human and chimpanzee genomes.

Key words: comparative biology/domain-specific adaptation/human evolution/sialic acid/Siglec

[1]To whom correspondence should be addressed; e-mail: avarki@ucsd.edu
[2]Current address: Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, MO

Sequences reported here are deposited in GenBank, under accession numbers AY532661, AY532662 and AY532663.

Introduction

Sialic acids (Sias) are a family of negatively charged monosaccharides that frequently terminate sugar chains found on vertebrate cell surface glycoproteins and glycolipids. N-acetylneuraminic acid (Neu5Ac) is a major sialic acid in vertebrates (Angata and Varki, 2002; Traving and Schauer, 1998; Troy, 1992). One common natural derivative of Neu5Ac is N-glycolylneuraminic acid (Neu5Gc), which is equally abundant on the cell surfaces of many mammalian tissues, including those of the great apes (Muchmore *et al*., 1998). All humans have a unique inactivating homozygous mutation in CMP-Neu5Ac hydroxylase (*CMAH*), eliminating the enzymatic activity that generates CMP-Neu5Gc from CMP-Neu5Ac (Chou *et al*., 1998; Irie *et al*., 1998). The *CMAH* mutation was caused by an *Alu* replacement event (Hayakawa *et al*., 2001) ~2.5–3 million years ago (mya) (Chou *et al*., 2002) and resulted in the absence of Neu5Gc and a secondary increased level of Neu5Ac on human cell surfaces (Muchmore *et al*., 1998; Varki, 2002).

Mammalian cell–expressed Sias can be recognized by organism-extrinsic receptors, like viral hemagglutinins and bacterial toxins, as well as by organism-intrinsic receptors, such as the Siglecs (Sia-recognizing Ig-like lectins) and selectins (Angata and Brinkman-Van der Linden, 2002; Angata and Varki, 2002; Crocker, 2002; Crocker and Varki, 2001; Karlsson, 1998; Rosen and Bertozzi, 1994). There are some known examples wherein the single oxygen atom difference between Neu5Ac and Neu5Gc can markedly affect such recognition processes (Angata and Brinkman-Van der Linden, 2002; Angata and Varki, 2002; Blixt *et al*., 2003; Collins *et al*., 1997; Karlsson, 1998; Kelm *et al*., 1994). Thus human Sia-recognizing receptors, such as Siglecs, became candidates for evolutionary change following the human-specific loss of Neu5Gc and consequent increase in Neu5Ac expression.

Siglecs are cell surface type 1 transmembrane I–type lectins that are broadly classified into two groups—the CD33/Siglec-3-related group (CD33rSiglecs, i.e., Siglecs-3, and -5 through -11 in humans) that are clustered in one genomic region (19q13.3–4), and another group, Siglecs-1, -2, and -4, that are more distantly related (Angata and Brinkman-Van der Linden, 2002; Crocker and Varki, 2001; Vyas *et al*., 2002). The extracellular portion of the CD33rSiglecs consists of two to five immunoglobulin (Ig)-like domains, the most N-terminal one being a Sia binding V-set (antibody variable region-like) Ig-like domain (Angata and Brinkman-Van der Linden, 2002; Crocker and Varki, 2001). Conserved features of this domain are required for Sia binding (Alphey *et al*., 2003), and recognition specificity can be altered by relatively few amino acid

changes (Yamaji *et al.*, 2002). These CD33rSiglecs are of interest with regard to evolutionary changes, because comparison of the syntenic genomic regions in humans and rodents show several differences (Angata *et al.*, 2001a). Furthermore, the human-specific loss of a critical arginine residue in a Siglec-like molecule (Siglec-L1; also within the CD33rSiglec gene cluster) abolished Sia binding after the last common ancestor with apes (Angata *et al.*, 2001b).

The selective expression of most CD33rSiglecs on granulocytes, monocytes, macrophages, and NK cells (Siglecs-3, -5, and -7–11) (Angata and Brinkman-Van der Linden, 2002; Crocker and Varki, 2001) implicates them in regulating innate immune functions mediated by these blood cell types. Involvement in cellular signaling is indicated by induced phosphorylation of conserved tyrosine residues on their cytosolic tails, which in turn causes association with the phosphatases SHP-1 and/or SHP-2 (Crocker and Varki, 2001; Taylor *et al.*, 1999; Ulyanova *et al.*, 2001).

The amino-terminal V-set Sia-binding sites of CD33rSiglecs on circulating innate immune cells are constitutively "masked" by interactions with Sias present on the same cell surface (Collins *et al.*, 2002; Razi and Varki, 1998, 1999). "Unmasking" Siglecs expressed on human peripheral blood occurs on cellular activation (Razi and Varki, 1999), and masking status is tied functionally to signaling events (Grobe and Powell, 2002). It is unknown whether Siglecs simply oscillate between masked and unmasked states or bind to other Sias (not expressed on the same cell surface) after unmasking.

Despite all this information, the primary biological functions of CD33rSiglecs remain unknown. One possibility is that they are self-recognition molecules that prevent inappropriate activation of innate immune cells. An alternative hypothesis is that when unmasked they serve as detectors of invasive Sia-expressing bacteria (Crocker and Varki, 2001; Jones *et al.*, 2003). Several strains of pathogenic bacteria are known to express Neu5Ac on their surfaces (Angata and Varki, 2002; Troy, 1992; Vimr and Lichtensteiger, 2002), thereby mimicking host cell surfaces and evading detection by both innate and adaptive immune systems (Jarvis, 1995; Vimr and Lichtensteiger, 2002; Wessels *et al.*, 1989). Unmasked Siglecs could thus provide a mechanism for the innate immune system cells to recognize these camouflaged bacteria. Sia-expressing pathogenic bacteria isolated from multiple mammalian species express Neu5Ac but never Neu5Gc (Angata and Varki, 2002; Troy, 1992; Vimr and Lichtensteiger, 2002). Thus the pathogen-recognition hypothesis predicts that CD33rSiglecs on cells of the innate immune system must recognize Neu5Ac.

Recent studies (Blixt *et al.*, 2003; Sonnenburg *et al.*, unpublished data) indicate that human CD33rSiglecs are relatively indiscriminate in binding both Neu5Ac and Neu5Gc. Although this competence of human CD33rSiglecs for Neu5Ac binding is consistent with their postulated role as detectors of sialylated bacteria, we provide evidence suggesting that Neu5Ac recognition is actually a derived, human-specific condition. This condition apparently resulted from recent adaptive evolution away from the ancestral strong Neu5Gc preference still observed for great apes and is a logical consequence of human loss of Neu5Gc expression.

Results and discussion

Cloned chimpanzee and gorilla Siglec-9 differ from the human ortholog in showing a strong preference for Neu5Gc-containing ligands

Given the diverse biological roles of Sias, loss of Neu5Gc in human evolution was likely accompanied by both positive and negative consequences, although eventual fixation of the causal *CMAH* mutation in humans suggests a net advantage. We reasoned that human Neu5Gc loss would have initially caused improper signaling if there was sudden unmasking of any Siglecs with a preexisting specificity for Neu5Gc. Current knowledge precludes defining the severity of such a signaling defect, but it would have necessarily been less severe than the original condition that allowed the *CMAH* mutation to rise in frequency. Subsequent selection for Neu5Ac binding capability would then have allowed restoration of human Siglecs to their constitutive masked state, relieving such a defect (and increasing the net advantage of the *CMAH* mutation).

To address the possibility that human Siglecs have evolved different Sia-binding specificity from those of great apes, we compared the Neu5Ac- and Neu5Gc-binding ability of human, chimpanzee, and gorilla Siglec-9, one member of the CD33rSiglecs that is expressed on monocytes and granulocytes. To investigate possible differences in binding specificity of Siglec-9, we cloned the N-terminal Sia-binding regions of chimpanzee and gorilla Siglec-9 and fused them to the Fc-region of human IgG, resulting in recombinant soluble chimeric proteins (Siglec-9-Fc). The Neu5Ac- versus Neu5Gc-binding properties of these molecules were compared to those of human Siglec-9-Fc (Figure 1). These binding assays clearly demonstrate that although human Siglec-9 binds both Neu5Ac and Neu5Gc, the chimp and gorilla orthologs strongly prefer Neu5Gc. These data are consistent with the hypothesis of functional adaptation in members of the human CD33rSiglecs to accommodate Neu5Ac-binding following human Neu5Gc-loss.

We also asked whether this human-specific increase in Neu5Ac-binding of Siglec-9 is evident when the recombinant molecules interact with natural cell surface glycans. Epstein-Barr virus–transformed B cells from humans (expressing >95% Neu5Ac on their surfaces) and orangutans (expressing ~85% Neu5Gc) (Brinkman-Van der Linden *et al.*, 2000) were probed with the recombinant Siglec-9-Fc molecules and binding measured by flow cytometry. Consistent with the enzyme-linked immunosorbent assay (ELISA) results, human Siglec-9-Fc bound both B cell lines well but preferred the Neu5Ac-expressing human B cells. In contrast, chimp and gorilla Siglec-9-Fcs displayed robust binding specifically to the Neu5Gc-rich cells (data not shown). Our results demonstrate that human Siglec-9 underwent significant functional adaptation after the evolutionary divergence from great apes.

Siglecs on circulating blood cells from human and apes show marked differences in binding Neu5Ac or Neu5Gc bearing probes

To test if endogenously expressed CD33rSiglecs from the human, chimp, and gorilla display similar binding

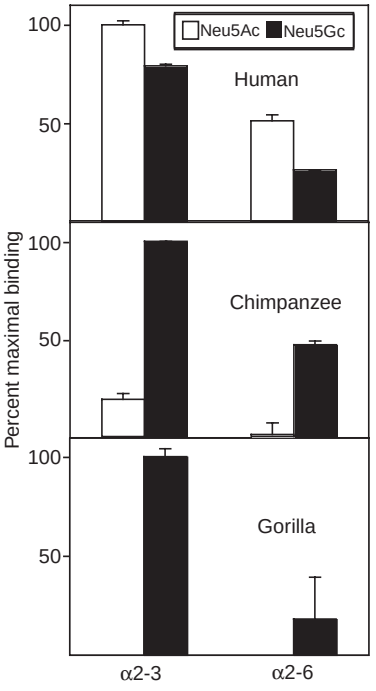

Fig. 1. Chimpanzee and gorilla Siglec-9 differ from the human ortholog in showing a strong preference for Neu5Gc-containing ligands. ELISA plate assay in which human, chimp, or gorilla Siglec-9-Fc were captured in protein A–coated wells and probed with Sia-biotinylated-PAA probes bearing Neu5Ac (white bars) or Neu5Gc (black bars) linked α2-3 or α2-6 to a Galβ1-4GlcNAc unit. Results are mean values ± SD of triplicates and are typical of three or more such experiments. Assays were performed as previously described (Angata *et al*., 2001a). The human molecule studied was cloned from the ancestral allele (see text for discussion of human polymorphism).

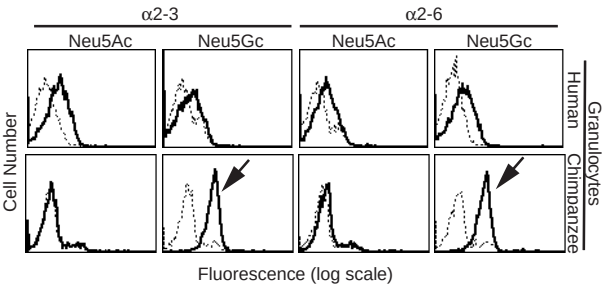

Fig. 2. Siglecs on human and chimpanzee circulating blood granulocytes differ in binding Neu5Ac- or Neu5Gc-bearing probes. Flow cytometry of human and chimp circulating blood leukocytes (gated for granulocytes) stained with one of four matching Sia-biotinylated-PAA probes differing only in the type (Neu5Ac or Neu5Gc) or linkage (α2-3 or α2-6) of Sia attached to Galβ1-4GlcNAc. Cells were either treated with mild periodate to unmask the endogenous Siglecs (heavy line) or sham-treated (dashed line) prior to staining. Sialidase treatment was not as efficient as mild periodate in unmasking Siglecs on chimp blood cells (data not shown).

differences observed for recombinant Siglec-9, we examined Sia binding of experimentally unmasked endogenous Siglecs expressed on native blood cells from these three species. Binding of soluble multivalent probes bearing Neu5Ac or Neu5Gc in either the alpha2-3 or alpha2-6 linkage (Angata *et al*., 2001b) was quantified by flow cytometry. Remarkable differences between human and great ape Siglecs were observed. Unmasked Siglecs on human monocytes (data not shown) and granulocytes bound probes bearing either Neu5Ac or Neu5Gc (Figure 2, top). In striking contrast, the corresponding chimp cells showed strong binding to Neu5Gc (see arrows in Figure 2, bottom), with no detectable binding to Neu5Ac-bearing probes, in agreement with the binding properties observed for recombinant Siglec-9. This strict Neu5Gc preference was also observed in blood samples from three other chimps and one gorilla (data not shown). The Sia binding observed by flow cytometry of human blood leukocytes represents a composite of the properties of at least five different CD33rSiglecs expressed on these cell types (Siglecs-3, -5, -7, -9, and -10) (Crocker and Varki, 2001), and we have recently found that great ape blood leukocytes also express these Siglecs (Hurtado-Ziola and Varki, unpublished data). Together these data suggest the ancestral state for the predominantly expressed ape neutrophil and monocyte-associated CD33rSiglecs was a strong preference for Neu5Gc, with human CD33rSiglecs subsequently undergoing functional adaptation to accommodate Neu5Ac binding in the absence of Neu5Gc.

The coevolution of human CD33rSiglec specificity with the human sialylation pattern, and the strong Neu5Gc-binding preference of chimp and gorilla CD33rSiglecs supports our previous findings (Razi and Varki, 1998, 1999) that the natural functional ligands of these molecules are endogenous Sias. These data also indicate that CD33rSiglecs do not function primarily to recognize invasive bacterial pathogens (which can express Neu5Ac, but never Neu5Gc). Indeed, the possibility must now be considered that some of these pathogens may be exploiting the recently evolved Neu5Ac binding ability of human Siglecs to assist in their invasion.

Domain-specific differences between the sequences of the N-terminal regions of human and ape Siglec-9

The genomic structure of the N-terminal region of Siglec-9 within all three species is identical, with the first exon encoding the signal peptide and the Sia-binding V-set domain and the second exon encoding the adjacent Ig-like C2-set domain (antibody constant domain-like). Comparison of the N-terminal sequences (Figure 3) shows multiple amino acid differences specifically in the 124 amino acid V-set domain (human-chimp identity = 94.4%; human-gorilla identity = 92.7%). In contrast, the 93 amino acids comprising the adjacent C2-set domains have the expected level of amino acid identity (~99% for human-chimp and human-gorilla) (Figure 3). This disparity in identity of the V-set domain and adjacent C2-set domain in interspecies comparisons is consistent with human Neu5Gc loss leading to selection for changes specific to the V-set domain that altered Sia binding specificity.

In further support of this, the 200-base-pair intron between the V-set and C2-set encoding exons displays 99.5–100% identity for all pairwise comparisons between the three species (see Appendix). The chimpanzee and gorilla V-set sequence also show a lower than expected amino acid identity (94.4%). However both still maintain

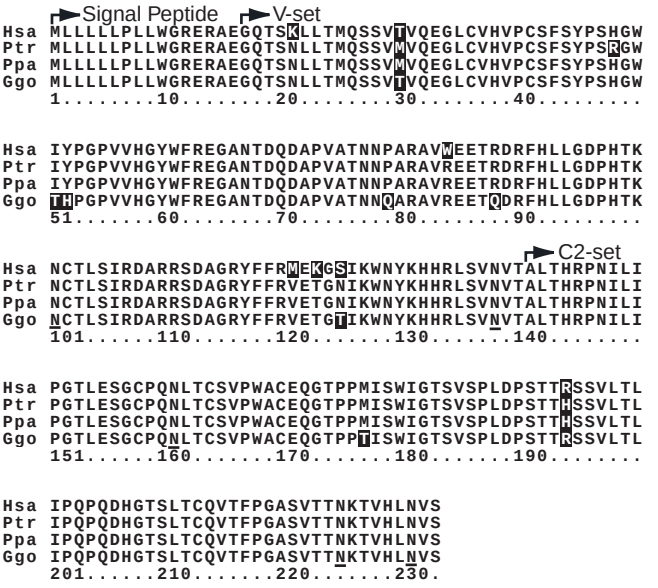

Fig. 3. Deduced amino acid sequences of N-terminal regions of human, chimpanzee, bonobo, and gorilla Siglec-9. Predicted amino acid sequences of the Siglec-9 N-terminal regions of human (Hsa), chimp (Ptr), bonobo (Ppa), and gorilla (Ggo) including the signal peptide and first two Ig-like domains (V-set and C2-set, respectively). Nonidentical residues are indicated with a white letter on black background. The presumed domain borders are indicated, and consensus N-glycosylation sites are underscored. All such potential sites between species have been conserved, indicating a possible role of N-glycans in the biology of these molecules. Complete nucleotide sequence data are presented in the Appendix and in submitted GenBank files.

the strict Neu5Gc-binding preference, indicating that selective pressures other than Neu5Gc loss may have also affected great ape Siglec-9 V-set domains.

Evidence for adaptive evolution involving the V-set domain of Siglec-9

Pairwise comparisons of Siglec-9 V-set DNA sequences amongst the three genera show that the great majority (almost 80%) of total nucleotide differences are nonsynonymous (result in amino acid changes). Indeed, seven of the eight chimp-human base pair differences in the 372 base pairs encoding the V-set domain are nonsynonymous (Table I, see also Appendix). This bias in the Siglec-9 V-set domains suggests that these amino acid changes have been selected for. To quantitate this bias, the nonsynonymous substitution rate (*dN*) and the synonymous substitution rate (*dS*) were calculated. A *dN*/*dS* >1 indicates positive directional selection, and values closer to zero suggest purifying selection (Hughes and Nei, 1988; Messier and Stewart, 1997). Table I shows the *dN*/*dS* values for human-chimp, human-gorilla, and chimp-gorilla comparisons of the Siglec-9 V-set or C2 -set domain. Although these values indicate that amino acid changes in the V-set domain are under selection in all three lineages, the *dN*/*dS* ratio of 2.28 between humans and chimps is clearly highest. In contrast to the evidence for positive selection in the V-set domains, low *dN*/*dS* ratios (0.12 between chimp and human) for the adjacent C2-set domains implicate purifying selection as the major evolutionary force acting on this domain. Although the small sizes of the regions involved preclude definitive statistical analyses of these differences, the high identity of a noncoding sequence immediately adjacent to the V-set encoding exon provides further evidence suggesting positive selection of amino acid–changing mutations specific to the Sia-binding V-set domain. In the case of humans, one of these selective forces was likely the loss of Neu5Gc and the subsequent need to allow Neu5Ac binding to restore the masked state.

We have also sequenced 1.1 kb of the N-terminus in 20 human Siglec-9 alleles representing global population diversity (20 alleles from 3 continents: 8 African (4 Mbuti, 4 Biaka); 6 Europeans; and 6 Japanese; detailed data not shown). This analysis showed only one polymorphic nonsynonymous site in the V-set domain and no synonymous changes. This polymorphic site is not one of those already mentioned as specific to the hominid lineage. No polymorphisms were found in the C2-set domain, nor in the introns. These data confirm that all other nonsynonymous changes in the V-set domain were fixed in the hominid lineage prior to the common ancestor of modern humans. The single nonsynonymous polymorphism was found in 55% of the human alleles studied, with a moderate frequency of the heterozygous genotypic state. Although the small exon sizes preclude strong statistical conclusions about the nature of inter- and intraspecific selective pressures, the data are consistent with ongoing positive or recent balancing selection on the human Siglec-9 V-set domain.

Evidence for episodic selection within the V-set of Siglec-9 in hominoids

When the *Gorilla* sequence is used as an outgroup, it is apparent that human Siglec-9 acquired approximately twice as many lineage-specific amino acid substitutions as chimpanzees since divergence from the *Homo-Pan* common ancestor, consistent with adaptive changes in human sialic acid biology (Figure 4). To explore this issue further, we sequenced the N-terminal region of Siglec-9 from a bonobo. The ancestors of bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*) were geographically separated ~3.5 mya, due to the appearance of the Zaire (Congo) river (Myers Thompson, 2003), and molecular data suggest the genetic divergence of these *Pan* species to be ~2.5 mya (Horai *et al.*, 1992). Remarkably, despite the prolonged period of evolutionary independence (which represents almost half of the evolutionary time since the common ancestor of *Homo* and *Pan*), we found only one difference between the chimp and bonobo V-set sequences (i.e., only 13% of the total divergence between chimps and humans). This conservation between the *Pan* V-set domains contrasts with the divergences among the *Pan*, human, and gorilla clades (Figure 4, left tree). On the other hand, the C2-set domain tree shows minimal divergences (Figure 4, right tree). Comparison of these domain-specific trees illustrates the disparate selective pressures on these adjacent segments of Siglec-9. The V-set domain tree also indicates that rapid evolution is not a ubiquitous feature of this domain in African hominoids, as the bonobo and chimp are tightly clustered.

Table I. Pairwise comparisons of N-terminal regions of Siglec-9

	Nucleotide				Protein (predicted)	
		Substitutions				
	Identity (%)	Nonsyn.	Syn.	*dN*/*dS*[a]	Identity (%)	Substitutions
V-set domain (372 bp, 124 aa)						
Human:chimp	97.9	7	1	2.28	94.4	7
Human:bonobo	98.1	6	1	1.95	95.2	6
Human:gorilla	96.8	9	3	0.96	92.7	9
Chimp:gorilla	97.6	7	2	1.15	94.4	7
Bonobo:gorilla	97.9	6	2	0.98	95.2	6
Chimp:bonobo	99.7	1	0	*	99.2	1
C2-set domain (279 bp, 93 aa)						
Human:chimp	98.6	1	3	0.12	98.9	1
Human:bonobo	98.9	1	2	0.18	98.9	1
Human:gorilla	98.9	1	2	0.18	98.9	1
Chimp:gorilla	98.2	2	3	0.23	97.8	2
Bonobo:gorilla	98.6	2	2	0.35	97.8	2
Chimp:bonobo	99.6	0	1	0	100.0	0

*Ratio could not be calculated.
[a]Estimated with DnaSP 3.51 (Rozas and Rozas, 1999), following the method of Nei and Gojobori (1986).

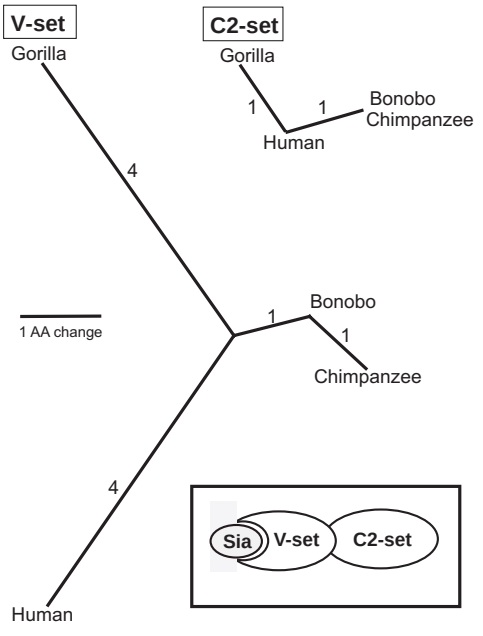

Fig. 4. Phylogenetic analysis of the amino acids of the V-set and adjacent C2-set domains of Siglec-9 in humans and African apes. A single most parsimonious branch and bound phylogram for the amino acid sequence of each domain was inferred using PAUP* (version 4.0b10). Trees were rooted with the *Gorilla* sequence to determine polarity of the *Homo-Pan* mutations. The number of amino acid substitutions is indicated on the branches, and branches are drawn to scale. Amino acid position 125 was removed in the V-set domain analysis due to unique states in each genus that prevented establishment of an ancestral state. Because most nucleotide changes are nonsynonymous, a nucleotide-based tree (not shown) looks very similar. The inset shows schematically the amino-terminal V-set domain involved in binding sialic acids and the following C2-set domain.

In interpreting the unusual pattern of V-set domain evolution, we cannot assign the ancestral state for any of the differences between *Gorilla* and the other species. Thus three explanations must be considered (all of which suggest intermittent or episodic selection among African hominoids). First, this domain may have been subjected to strong selection during the early period of the African hominoid divergences about 6–8 mya, and once again in the hominid lineage, after the divergence from *Pan*. Second, recent independent selection events may have occurred in the hominid and gorilla lineages, but not in the *Pan* species. Third, there may have been continual positive selection on the Siglec-9 V-set domain in African hominoids, which decreased specifically in the *Pan* lineage. Regardless of which of these possibilities is true, the *dN*/*dS* data in Table I indicate that the greatest degree of evolution has occurred in the hominid lineage, and our functional data suggest that loss of Neu5Gc was the major force driving this selection (Figure 5). What selective pressures caused the other divergences, whether extant interspecies differences reflect ancient selective sweeps, and whether there are other functional consequences, are questions that remain to be answered. One possibility we are exploring is that other species-specific changes in sialylation might have driven some of the selection in the other lineages.

Indeed, very recent comparisons of the entire CD33rSiglec gene cluster among human, chimp, baboon, mouse, and rat have revealed a general increase of nonsynonymous mutations within the V-set domain of these molecules (Angata *et al*., unpublished data). Because the endogenous glycan landscape appears to be the major determinant guiding the specificity of these receptors during evolution, it is

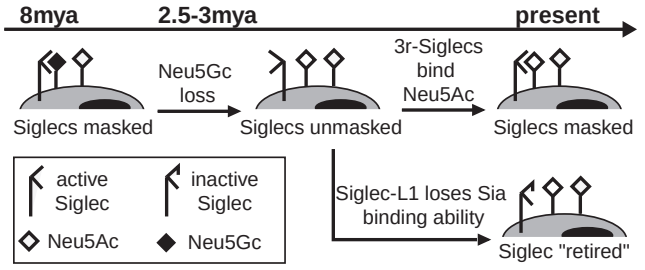

Fig. 5. Model depicting the impact of Neu5Gc loss on human Siglec evolution. Based on the data from chimps and gorillas, the ancestral state of hominoid and hominid leukocytes must have been dominated by Siglecs that recognized and were masked primarily by cell surface Neu5Gc. Approximately 2.5–3 million years ago, multiple hominid Siglecs that bound Neu5Gc were suddenly unmasked by the loss of Neu5Gc. These Siglecs could either undergo adaptive accommodation of Neu5Ac binding to restore their masked state (e.g., Siglec-9 and multiple other CD33rSiglecs), or they could be retired by losing their ability to bind Sias altogether (Siglec-L1).

not surprising that unique aspects of Sia expression within each species may select for binding specificity-altering mutations in the Sia-binding V-set domain. In certain cases (e.g., humans) drastic changes in CD33rSiglec specificity would be required to accommodate drastic changes in Sia expression, such as the loss of Neu5Gc. In other species, fine-tuning of CD33rSiglec specificity to subtle changes in Sia expression would be sufficient. Taken together, these speculations raise testable hypotheses regarding the differing specificity of CD33rSiglecs from different species.

Conclusions and perspectives

For human Siglec-9, we have directly shown functional adaptation towards Neu5Ac-binding, evidence of Sia-binding domain-specific divergence from the African great apes and a bias of nonsynonymous substitutions within this domain consistent with positive selection for mutations affecting receptor functionality. Experimental probing of multiple CD33rSiglecs present on human, chimp, and gorilla blood cells (Figure 2) indicates that many of these human receptors may have undergone a similar human-specific functional change, away from a strong preference for Neu5Gc. Indeed, *in vitro* binding studies do show significant binding of other recombinant human CD33rSiglecs (-3, -5, -7, and -10) to both Neu5Ac and Neu5Gc (Blixt *et al.*, 2003; Sonnenburg *et al.*, unpublished data). Additionally, initial analyses of recombinant human and chimp Siglec-7 indicate multiple amino acid changes in the V-set domain, a strong chimp Siglec-7 preference for Neu5Gc, and a human Siglec-7 accommodation of Neu5Ac (Sonnenburg *et al.*, unpublished data). Overall, our data also indicate that endogenous Sias (rather than surface Neu5Ac molecules of bacterial pathogens) are the functional ligands of CD33rSiglecs. This in turn suggests that the endogenous Sia landscape is the major factor directing evolution of CD33rSiglec binding specificity. Meanwhile, it is possible that some of Neu5Ac-expressing pathogens are actually exploiting the recently evolved Neu5Ac binding ability of human Siglecs.

We have also mapped the amino acid differences between human and chimp for Siglec-7 and -9 onto the available Siglec-7 V-set structure (Alphey *et al.*, 2003) and found no obvious changes that would rationally explain the observed difference in binding specificity (Sonnenburg *et al.*, unpublished data). Although one or two amino acid changes could potentially change Siglec specificity, it is also possible that the observed differences are the cumulative result of several mutations that altered the Sia-binding pocket. This is supported by the observation that only 1 of the 13 sites of amino acid differences in human and chimp Siglec-7 and Siglec-9 V-set domains is in common, and this site does not appear to directly interact with the sialic acid (Sonnenburg *et al.*, unpublished data). Overall, the human change can be best characterized not as a specific switch from Neu5Gc to Neu5Ac preference but as a relaxation of specificity to accommodate Neu5Ac binding, without loss of Neu5Gc binding. Thus human Siglec-7 and -9 have accommodated Neu5Ac binding by distinct molecular strategies, suggesting that many of the V-set amino-acid differences that distinguish the human and chimp orthologs are required for the differences in observed binding properties. Together, these facts make it rather difficult to predict which amino acid changes are responsible and design rational mutagenesis studies.

Few human proteins are known to have undergone rapid selection during anthropoid primate evolution (Enard *et al.*, 2002; Goldberg *et al.*, 2003; Johnson *et al.*, 2001; Messier and Stewart, 1997; Nadezhdin *et al.*, 2001; Wyckoff *et al.*, 2000). To our knowledge, none of these reports directly demonstrated biochemically human-specific functional adaptation, as we have done here. It remains to be seen if changes in CD33r-Siglec-binding specificity are the residual signature of a past selective sweep involving a pathogen that affected Sia biology, whether they are relevant to present human resistance or susceptibility to disease or if they had any secondary consequences for the evolution of other human-specific traits.

We also note that the evidence for rapid evolution in Siglec-9 is much less apparent when the human-chimp comparisons include both the highly conserved exon 2 and the divergent exon 1. The Siglec-9 *dN*/*dS* decreases from 2.28 to 0.67 when the C2-set domain is included with the V-set domain in the analysis. Indeed, analysis of the entire Siglec coding regions from human and chimp (an additional three exons) would likely yield a value that does not noticeably deviate from that expected under purifying selection. For example, the average *dN*/*dS* ratio for 19 other immune-related genes of humans and chimpanzees is 0.45 (Chen *et al.*, 2001). Thus the evidence suggesting domain-specific accelerated evolution would have been overlooked in conventional comparisons using entire open reading frames or full-length cDNAs. With chimp genome sequencing now well under way (Fujiyama *et al.*, 2002; Olson and Varki, 2003), large-scale efforts to identify additional chimp-human genetic differences are anticipated. Because most genes will likely show only a few amino acid differences between the two species, our data caution that functionally significant genetic differences will be missed unless an exon-by-exon comparative approach is taken, with an emphasis on the functionality of the protein domain(s) encoded by each exon.

Materials and methods

Reagents and cells

CMP-Neu5Gc was kindly produced by J. Kawakami and B. Hayes (supported by grant GM61894) and was synthesized as described (Angata *et al.*, 2001b). Epstein-Barr virus–transformed B cells from orangutan and human were a gift from Peter Parham (Stanford University) and were maintained in RPMI-1640 (Invitrogen, Carlsbad, CA), 10% heat-inactivated fetal calf serum. See acknowledgments for the sources of the great ape blood samples.

Genomic PCR and cloning of chimp, gorilla, and bonobo Siglecs

The following primers were used for amplification of Siglec-9 from chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), or bonobo (*Pan paniscus*) genomic DNA: chimp Siglec-9 and bonobo Siglec-9 (5′-GTTCTGAGAGAAGAACC-3′ and 5′-GCCTTCTCCTTGGAAGACAG-3′) and gorilla Siglec-9 (5′-CTCGGATCCCTGGCACCTCTAACCC-3′ and 5′-ATCTCCTTGGAAGACAGTCATGG-3′). Polymerase chain reaction (PCR) was conducted using Expand-High Fidelity polymerase (Roche, Mannheim, Germany). PCR fragments were cloned into pCR2.1-TOPO (Invitrogen) and sequenced or sequenced directly from the genomic PCR products on an ABI Prism 310 Genetic Analyzer. Sequences were verified on multiple independent PCR products.

Sequence analysis

Alignments were prepared using CLUSTAL W. DnaSP (version 3.51) was used to calculate *dN* and *dS* for given pairs of sequences (Rozas and Rozas, 1999).

Construction of recombinant Siglec-Fc expression vectors

Vectors encoding recombinant fusion proteins consisting of the chimp or gorilla Siglec-9 N-terminus and the Fc region of human IgG (Siglec-Fc) were prepared as follows. The great ape Siglecs were subcloned by PCR using the following primers: 5′-CCCTCTAGAGCCACCATGCTGCTGC TGCTGCTGCCCCTGC-3′, 5′-ATCTCCTTGGAAGACAGTCATGG-3′. PCR fragments were isolated, digested with XbaI, and cloned into Ek-Fc/pcDNA3.1(–) (Angata *et al.*, 2001a), digested with XbaI and EcoRV. Siglec-Fcs were produced as previously described (Angata *et al.*, 2001a).

Sialylation of polyacrylamide-biotin probes

Fifty micrograms of a biotinylated-PAA probe (Glycotech, Rockville, MD) bearing multiple copies of Galβ1-4GlcNAc was sialylated in a mixture of 10 μl 0.5 M 4-morpholine propane sulfonic acid, pH 7.4, 0.5 μl 10% sodium azide, 1 μl 1 U/μl calf-intestine alkaline phosphatase, 25 μl 10 mM CMP-Neu5Ac or CMP-Neu5Gc, and 50 μl water, to which was added 120 mU of the sialyltransferases ST6Gal-I (Calbiochem, San Diego, CA) or ST3Gal-III (gift from Eric Sjoberg; formerly of Cytel). Reactions were allowed to proceed at 37°C overnight. Probes were recovered by ultrafiltration using a Microcon-10 (Millipore, Bedford, MA) and resuspension in water. Sialylation was quantitated by acid release in 2 M acetic acid, 80°C, for 3 h; derivitization by 1,2-diamino-4,5-methylenedioxybenzene; and high-pressure liquid chromatography analysis as previously described (Sonnenburg *et al.*, 2002).

Flow cytometry of peripheral blood leukocytes

Samples (10–20 ml) of blood were collected in Vacutainers (Becton Dickinson, Piscataway, NJ) containing ethylenediamine tetra-acetic acid. Phosphate buffered saline (PBS)-washed cells were treated twice for 5 min in 5 volumes ACK lysing buffer (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM ethylenediamine tetra-acetic acid, pH 7.2) to lyse red blood cells. Leukocytes were pelleted, washed, and used at 0.5–1 million/sample for flow cytometry. Samples were either incubated in 100 μl freshly prepared 2 mM sodium periodate in PBS for 30 min, on ice, in the dark, or were sham-treated in PBS. Periodate specifically truncates the side chain of Sias resulting in CD33rSiglecs unmasking on the cell surface (Razi and Varki, 1999), allowing them to be probed for the binding of Neu5Ac- or Neu5Gc-bearing probes (Razi and Varki, 1999). Forward- and side-scatter dot plots were used to identify granulocyte, monocyte, and peripheral lymphocyte populations.

Acknowledgments

We gratefully acknowledge Takashi Angata, Pascal Gagneux, and Erica Dutil Sonnenburg for helpful discussions and Els Brinkman van der Linden for some of the initial observations leading to this work. Chimpanzee, bonobo, and gorilla samples were kindly provided by the Yerkes Primate Center (Atlanta, GA), the San Diego Zoo, and Lincoln Park Zoo (Chicago, IL), respectively. This work was supported by grants from the NIH and fellowship from the ACS to TKA.

Abbreviations

C2-set domain, antibody constant region-like domain; ELISA, enzyme-linked immunosorbent assay; PBS, phosphate buffered saline; PCR, polymerase chain reaction; Sia, sialic acid; V-set domain, antibody variable region-like domain.

References

Alphey, M.S., Attrill, H., Crocker, P.R., and Van, A.D.M. (2003) High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family. *J. Biol. Chem.*, **278**, 3372–3377.

Angata, T. and Brinkman-Van der Linden, E. (2002) I-type lectins. *Biochim. Biophys. Acta*, **1572**, 294.

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

Angata, T., Hingorani, R., Varki, N.M., and Varki, A. (2001a) Cloning and characterization of a novel mouse Siglec, mSiglec-F: differential evolution of the mouse and human (CD33) Siglec-3-related gene clusters. *J. Biol. Chem.*, **276**, 45128–45136.

Angata, T., Varki, N.M., and Varki, A. (2001b) A second uniquely human mutation affecting sialic acid biology. *J. Biol. Chem.*, **276**, 40282–40287.

Blixt, O., Collins, B.E., Van Den Nieuwenhof, I.M., Crocker, P.R., and Paulson, J.C. (2003) Sialoside specificity of the Siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin associated glycoprotein. *J. Biol. Chem.*, **278**, 31007–31019.

Brinkman-Van der Linden, E.C.M., Sjoberg, E.R., Juneja, L.R., Crocker, P.R., Varki, N., and Varki, A. (2000) Loss of N-glycolylneuraminic acid in human evolution—implications for sialic acid recognition by siglecs. *J. Biol. Chem.*, **275**, 8633–8640.

Chen, F.C., Vallender, E.J., Wang, H., Tzeng, C.S., and Li, W.H. (2001) Genomic divergence between human and chimpanzee estimated from large-scale alignments of genomic sequences. *J. Hered.*, **92**, 481–489.

Chou, H.H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K.L., Muchmore, E.A., Nelson, D.L., Warren, S.T., and Varki, A. (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc. Natl Acad. Sci. USA*, **95**, 11751–11756.

Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Pääbo, S., Satta, Y., Takahata, N., and Varki, A. (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl Acad. Sci. USA*, **99**, 11736–11741.

Collins, B.E., Kiso, M., Hasegawa, A., Tropak, M.B., Roder, J.C., Crocker, P.R., and Schnaar, R.L. (1997) Binding specificities of the sialoadhesin family of I-type lectins—sialic acid linkage and substructure requirements for binding of myelin-associated glycoprotein, Schwann cell myelin protein, and sialoadhesin. *J. Biol. Chem.*, **272**, 16889–16895.

Collins, B.E., Blixt, O., Bovin, N.V., Danzer, C.P., Chui, D., Marth, J.D., Nitschke, L., and Paulson, J.C. (2002) Constitutively unmasked CD22 on B cells of ST6Gal I knockout mice: novel sialoside probe for murine CD22. *Glycobiology*, **12**, 563–571.

Crocker, P.R. (2002) Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell–cell interactions and signalling. *Curr. Opin. Struct. Biol.*, **12**, 609–615.

Crocker, P.R. and Varki, A. (2001) Siglecs, sialic acids and innate immunity. *Trends Immunol.*, **22**, 337–342.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P., and Pääbo, S. (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, **418**, 869–872.

Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T.D., Itoh, T., Tsai, S.F., Park, H.S., Yaspo, M.L., Lehrach, H., Chen, Z., and others. (2002) Construction and analysis of a human-chimpanzee comparative clone map. *Science*, **295**, 131–134.

Goldberg, A., Wildman, D.E., Schmidt, T.R., Huttemann, M., Goodman, M., Weiss, M.L., and Grossman, L.I. (2003) Adaptive evolution of cytochrome c oxidase subunit VIII in anthropoid primates. *Proc. Natl Acad. Sci. USA*, **100**, 5873–5878.

Grobe, K. and Powell, L.D. (2002) Role of protein kinase C in the phosphorylation of CD33 (Siglec-3) and its effect on lectin activity. *Blood*, **99**, 3188–3196.

Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata, N. (2001) Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proc. Natl Acad. Sci. USA*, **98**, 11399–11404.

Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S., and Takahata, N. (1992) Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.*, **35**, 32–43. Erratum. *J. Mol. Evol.,* (1993) **37**(1), 89.

Hughes, A.L. and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.

Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T., and Suzuki, A. (1998) The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J. Biol. Chem.*, **273**, 15866–15871.

Jarvis, G.A. (1995) Recognition and control of neisserial infection by antibody and complement. *Trends Microbiol.*, **3**, 198–201.

Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., and Eichler, E. E. (2001) Positive selection of a gene family during the emergence of humans and African apes. *Nature*, **413**, 514–599.

Jones, C., Virji, M., and Crocker, P.R. (2003) Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake. *Mol. Microbiol.*, **49**, 1213–1225.

Karlsson, K.A. (1998) Meaning and therapeutic potential of microbial recognition of host glycoconjugates. *Mol. Microbiol.*, **29**, 1–11.

Kelm, S., Pelz, A., Schauer, R., Filbin, M.T., Tang, S., De, B.M.-E., Schnaar, R.L., Mahoney, J.A., Hartnell, A., Bradfield, P., and Crocker, P.R. (1994) Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Curr. Biol.*, **4**, 965–972.

Messier, W. and Stewart, C.B. (1997) Episodic adaptive evolution of primate lysozymes. *Nature*, **385**, 151–154.

Muchmore, E.A., Diaz, S., and Varki, A. (1998) A structural difference between the cell surfaces of humans and the great apes. *Am. J. Phys. Anthropol.*, **107**, 187–198.

Myers Thompson, J.A. (2003) A model of the biogeographical journey from Proto-pan to Pan paniscus. *Primates*, **44**, 191–197.

Nadezhdin, E.V., Vinogradova, T.V., and Sverdlov, E.D. (2001) Interspecies subtractive hybridization of cDNA from human and chimpanzee brains. *Dokl. Biochem. Biophys.*, **381**, 415–418.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.

Olson, M.V. and Varki, A. (2003) Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.*, **4**, 20–28.

Razi, N. and Varki, A. (1998) Masking and unmasking of the sialic acid-binding lectin activity of CD22 (Siglec-2) on B lymphocytes. *Proc. Natl Acad. Sci. USA*, **95**, 7469–7474.

Razi, N. and Varki, A. (1999) Cryptic sialic acid binding lectins on human blood leukocytes can be unmasked by sialidase treatment or cellular activation. *Glycobiology*, **9**, 1225–1234.

Rosen, S.D. and Bertozzi, C.R. (1994) The selectins and their ligands. *Curr. Opin. Cell Biol.*, **6**, 663–673.

Rozas, J. and Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Sonnenburg, J.L., Van Halbeek, H., and Varki, A. (2002) Characterization of the acid stability of glycosidically linked neuraminic acid—use in detecting de-N-acetyl-gangliosides in human melanoma. *J. Biol. Chem.*, **277**, 17502–17510.

Taylor, V.C., Buckley, C.D., Douglas, M., Cody, A.J., Simmons, D.L., and Freeman, S.D. (1999) The myeloid-specific sialic acid-binding receptor, CD33, associates with the protein-tyrosine phosphatases, SHP-1 and SHP-2. *J. Biol. Chem.*, **274**, 11505–11512.

Traving, C. and Schauer, R. (1998) Structure, function and metabolism of sialic acids. *Cell Mol. Life Sci.*, **54**, 1330–1349.

Troy, F.A. (1992) Polysialylation: from bacteria to brains. *Glycobiology*, **2**, 5–23.

Ulyanova, T., Shah, D.D., and Thomas, M.L. (2001) Molecular cloning of MIS, a myeloid inhibitory siglec, that binds protein-tyrosine phosphatases SHP-1 and SHP-2. *J. Biol. Chem.*, **276**, 14451–14458.

Varki, A. (2002) Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences and implications for hominid evolution. *Yearbook Phys. Anthropol.*, **44**, 54–69.

Vimr, E. and Lichtensteiger, C. (2002) To sialylate, or not to sialylate: that is the question. *Trends Microbiol.*, **10**, 254–257.

Vyas, A.A., Patel, H.V., Fromholt, S.E., Heffer-Lauc, M., Vyas, K.A., Dang, J.Y., Schachner, M., and Schnaar, R.L. (2002) Gangliosides are functional nerve cell ligands for myelin-associated glycoprotein (MAG), an inhibitor of nerve regeneration. *Proc. Natl Acad. Sci. USA*, **99**, 8412–8417.

Wessels, M.R., Rubens, C.E., Benedi, V.-J., and Kasper, D.L. (1989) Definition of a bacterial virulence factor: sialylation of the group B streptococcal capsule. *Proc. Natl Acad. Sci. USA*, **86**, 8983–8987.

Wyckoff, G.J., Wang, W., and Wu, C.I. (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature*, **403**, 304–309.

Yamaji, T., Teranishi, T., Alphey, M.S., Crocker, P.R., and Hashimoto, Y. (2002) A small region of the natural killer cell receptor, Siglec-7, is responsible for its preferred binding to alpha 2,8-disialyl and branched alpha 2,6-sialyl residues. A comparison with Siglec-9. *J. Biol. Chem.*, **277**, 6324–6332.