Molecular Diversity (2006) 10: 555–566
DOI: 10.1007/s11030-006-9029-1 © Springer 2006

Review

Molecular diversity and evolution of the Siglec family of cell-surface lectins

Takashi Angata
Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, JAPAN
(E-mail: takashi-angata@aist.go.jp, Tel: +81-29-861-9460, Fax: +81-29-861-2744)

Received 31 October 2005; Accepted 18 January 2006

Key words: immunity, lectins, evolution, sialic acids, Siglecs

Summary

Sialic acids are a family of acidic sugars with a 9-carbon backbone, prominently expressed in animals of deuterostome lineage. Siglecs are the largest family of vertebrate endogenous receptors that recognize glycoconjugates containing sialic acids. Although a few Siglecs are well-conserved throughout vertebrate evolution and show similar binding preference regardless of the species of origin, most others, particularly the CD33-related subfamily of Siglecs, show marked inter-species differences in repertoire, sequence, and binding preference. The diversification of CD33-related Siglecs may be driven by direct competition against pathogens, and/or by necessity to catch up with the changing landscape of endogenous glycans, which may in turn be changing to escape exploitation by other pathogens.

Abbreviations: Neu5Ac, *N*-acetylneuraminic acid; Neu5Gc, *N*-glycolylneuraminic acid; KDN, 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid; Sia, sialic acid, type not specified; CMP-Neu5Ac, cytidine 5′-monophospho-Neu5Ac; Gal, D-galactose; GalNAc, *N*-acetyl-D-galactosamine; GlcNAc, *N*-acetyl-D-glucosamine; HexNAc, *N*-acetyl-D-hexosamine, i.e., GlcNAc or GalNAc; Fuc, L-fucose; ITIM, immunoreceptor tyrosine-based inhibitory motif; Sn, sialoadhesin; MAG, myelin-associated glycoprotein; SMP, Schwann cell myelin protein; CD33rSiglec, CD33/Siglec-3–related Siglec; EST, expressed sequence tag

Introduction

Pattern of evolution of cell-surface glycans

Most animal cells (in fact, all cells of both prokaryotic and eukaryotic origins alike) are “sugar-coated”, or covered with glycan chains carried on glycoconjugates (i.e., macromolecules containing sugar residues, such as glycoproteins, glycolipids, and glycosaminoglycans). Roles of glycans are multi-fold [1], and one of the prominent ones is to interact with endogenous lectins, or glycan-recognition proteins expressed in the same organism that synthesizes the glycans. Such interactions are an essential part of some regulatory mechanisms in multicellular organisms, such as regulation of cellular traffic and activity, among others. The very same glycans, however, can be also recognized by glycan-recognition proteins of pathogens (e.g., viruses, bacteria, and protozoa) and exploited for their benefit. Therefore, multicellular organisms face a difficult question: “to change, or not to change glycan structures”. While changing glycan structures may be beneficial for the host (multicellular organism) to escape unceasing attacks by pathogens that utilize host glycans, too drastic a change may also compromise internal workings of the host [2].

The general pattern of glycan structural diversity tells us what might have happened in the course of evolution. The basic “core” structures of glycans are conserved amongst many species, probably because change of the core structures would affect so many functional glycan structures based on them, compromising many glycan-mediated functions. In fact, changes of core glycan structures, which occur in the form of naturally occurring mutations, often result in diseases [3–5]. The remaining choice was to “tinker” with the peripheral structures, just enough to escape the recognition by pathogens but not too much to compromise endogenous functions, for example, recognition by endogenous lectins [2]. Sialic acids, the sugar residues located mostly at the outermost part of glycans, are well-suited for such “tinkering”.

Molecular diversity of sialic acids – a solution to the “glycans’ dilemma”?

Sialic acids are a family of acidic sugars with 9-carbon backbone, and expressed in abundance by the animals of

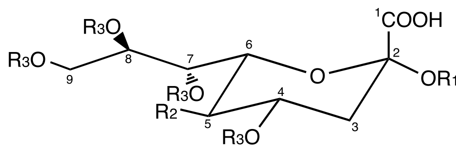

Figure 1. Structure of sialic acids. The common backbone structure of sialic acids is shown, and the carbon atoms are numbered. R_1 represents another sugar (Gal, GalNAc, Sia) to which the sialic acid is conjugated. R_2 represents substitution at C5 (= $-NH_2$ in neuraminic acid, $-NHCOCH_3$ in Neu5Ac, $-NHCOCH_2OH$ in Neu5Gc, $-OH$ in KDN). R_3 represents modification of hydroxyl groups at C4 and C7–C9 (= $-H$, $-COCH_3$, $-COCH(OH)CH_3$, $-CH_3$, and $-SO_3H$). Combination of R_2 and R_3 yields molecular diversity of sialic acids. Not all modifications are found at all positions (e.g., lactylation is found only at C9). Modifications (methylation and acetylation) of hydroxyl group in R_2, i.e., in KDN and in glycolyl group of Neu5Gc, were also reported.

deuterostome lineage [6, 7]. As mentioned above, sialic acids are mostly located at the periphery of glycans, most suitable for recognition by endogenous lectins. Of course, this location is also well-suited for recognition by molecules of pathogenic origins. There are in fact numerous examples of sialic acid recognition by pathogens and pathogen-derived molecules [6, 7], including viral surface proteins (e.g., hemagglutinins of influenza virus [8]), bacterial toxins (e.g., cholera toxin [9]), bacterial adhesins (e.g., SabA of *Helicobacter pylori*, the causative agent of gastric ulcer [10]), and protozoan proteins (e.g., EBA175 of *Plasmodium falciparum*, the malarial parasite [11]).

Given such frequent exploitation by pathogens, it may be not surprising that deuterostomes have invented more than 50 different versions of sialic acids [6] (Figure 1), perhaps as a result of extensive "tinkering". These structural variations are generated mostly by modification of hydroxyl groups (at C4, C7–C9) of three "basic" sialic acids, namely *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc), and 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nononic acid (KDN). Neu5Gc is ultimately derived from Neu5Ac [12, 13] and can be considered a structural variation of Neu5Ac.

Sialic acids as ligands for endogenous lectins

If sialic acids serve as "pathogen receptors", why haven't deuterostomes abandoned sialic acids altogether, rather than diversify them? One explanation is that elimination of sialic acids only renders penultimate sugars terminal, which would eventually be recognized by pathogens and become "pathogen receptors". If the new terminal sugars may not be modified as easily as sialic acids and therefore not suitable for structural tinkering, then it may be better to keep sialic acids as terminal sugars, as they have additional primary and secondary alcohol groups in the C7–C9 exocyclic chain, not found in most other sugars. Another explanation, of course, is that sialic acids play essential biological roles, and are therefore indispensable. Indeed, the embryonic lethality of the knockout mice deficient in sialic acids [14], and human diseases caused by mutation of the genes involved in sialic acid biosynthesis [15, 16], seem to support this explanation. Such essential roles of sialic acids likely include interactions with endogenous lectins.

In this review, I attempt to provide a brief overview of the studies of Siglecs, a family of vertebrate lectins that recognize sialic acid-containing glycans. The molecular diversity and evolution of the family will then be discussed, as well as possible correlation between structural diversity of sialic acids and evolution of Siglecs.

Essentials of Siglecs

Definition and nomenclature

Siglecs are the largest known family of vertebrate lectins that recognize sialic acid-containing glycans [17–19]. The family name was coined based on the common characteristics of the first four family members (sialoadhesin (Sn)/Siglec-1, CD22/Siglec-2, CD33/Siglec-3, myelin-associated glycoprotein (MAG)/Siglec-4a, and Schwann cell myelin protein (SMP)/Siglec-4b), namely "sialic acid recognition", "immunoglobulin superfamily", "lectins" [20]. The same set of molecules used to be also called "sialoadhesin family". The family expanded thereafter by discovery of new members, and the number of primate Siglecs reported so far is 13, while that in rodents is 8 (Tables 1 and 2). The difference between primates and rodents in the number of Siglecs reflects differences in the subfamily called CD33/Siglec-3–related Siglecs (CD33rSiglecs) [21, 22]. In contrast to Sn/Siglec-1, CD22/Siglec-2, and MAG/Siglec-4, which are

Table 1. Siglecs in vertebrates

	Mammals	Birds	Amphibians	Fishes
Sn/Siglec-1	Yes	Yes	Not found	Not found
CD22/Siglec-2	Yes	Not found [b]	Not found	Not found
CD33/Siglec-3 -related Siglecs	Yes	Not found	Not found	Not found
MAG/Siglec-4	Yes	Yes	Yes	Yes
Other Siglec-like [a]	Yes	Yes	Yes	Yes

[a]The category "Other Siglec-like" includes genomic and EST sequences that appear to encode Siglec-like proteins but not tested for their functions.

[b]The genome and EST sequencing projects in corresponding species are still incomplete. Therefore the assignment "not found" is only tentative and may be reverted in the future. Chicken (*Gallus gallus*) genome project is considered near completion [104] but the sequence data of the chromosome equivalent of human chromosome 19 (on which most Siglec genes reside, including those for MAG/Siglec-4, CD22/Siglec-2, and CD33rSiglecs) is largely missing. Puffer fish genome projects are near completion with good coverage [105, 106]. Another fish genome project (zebrafish) is still incomplete at the time of writing. There is no amphibian genome project near completion at this moment, and the amphibian column is based solely on the search of EST sequence database of clawed frogs *Xenopus laevis* and *Xenopus tropicalis*.

Table 2. CD33-related Siglecs in primates, rodents, and dog

	Human	Chimpanzee	Baboon	Mouse/Rat	Dog[a]
CD33/Siglec-3	Yes	Yes	Yes	Yes	Yes (3r2)
Siglec-5	Yes	Yes[b]	Yes	No	Yes (3r7)
Siglec-6	Yes	Yes	Yes[b]	Yes (Siglec-F)	Yes (3r6)[b]
Siglec-7	Yes	Yes	No	No	No
Siglec-8	Yes	Yes	Yes	No	No
Siglec-9	Yes	Yes	Yes	Yes (Siglec-E)	No
Siglec-10	Yes	Yes	Yes	Yes (Siglec-G)	Yes (3r4)
Siglec-11	Yes	Yes	? [c]	No	No
Siglec-12	Yes[b]	Yes	No	No	No
Siglec-13	No	Yes	Yes	No	No
Siglec-H	No	No	? [c]	Yes	No

This table is based on previous publications [21, 22, 89] and available genomic and EST sequence data. Of the remaining 4 dog CD33rSiglecs, #1 and #8 may be equivalent to human *SIGLECP16* [89, 107] and human *SIGLEC5** [21], respectively. There appear to be no equivalents of dog CD33rSiglecs #3 and #5 in primate or rodent genomes.
[a] Angata et al., unpublished.
[b] Essential arginine-mutated.
[c] Not enough data available at present.

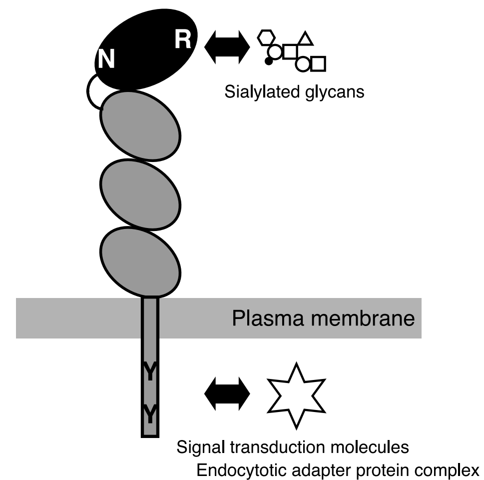

Figure 2. Schematic representation of Siglec protein. Each oval represents individual immunoglobulin-like domain. The amino-terminal Ig-like domain belongs to V-set, while others belong to C2-set. The first and the second Ig-like domains are assumed to be tethered by an inter-domain disulfide bond. The first Ig-like domain contains several amino acid residues essential for glycan recognition, including the "essential arginine" residue (*R*) required for optimal recognition of sialylated glycans. Most Siglecs have an *N*-glycosylation site (*N*) in the first Ig-like domain, on the side opposite to the one involved in glycan recognition. Cytoplasmic domain typically contains tyrosine-based signaling motif(s) (*Y*). Transmembrane domain in some Siglecs contains a positively charged residue (not indicated), which may be involved in interaction with activating adapter proteins, such as DAP12 and FcεRγ.

conserved between mouse and human, CD33/Siglec-3 and its closely related paralogs (i.e., CD33rSiglecs) are generally less conserved, and it is difficult to assign orthologous correspondence between human and mouse CD33rSiglecs. Therefore, human CD33rSiglecs were numbered in the order of discovery, while mouse CD33rSiglecs (except for CD33/Siglec-3) were given alphabetical designations (i.e., Siglec-E–H; see Table 2).

Most Siglecs are expressed on the cells involved in immunity, such as B-cells, natural killer cells, monocytes, macrophages, neutrophils, and eosinophils [17, 19, 23]. A notable exception is MAG/Siglec-4, which is expressed exclusively in nervous system. Also, while human Siglec-6 is prominently expressed on decidual cells of placenta, it is also expressed on B-cells [24].

Protein structure and function of each domain

Siglecs are type-I transmembrane proteins, with the extracellular domain exclusively composed of Ig-like domains, followed by transmembrane and short cytoplasmic domains (see Figure 2 for a schematic representation). The number of Ig-like domains in each Siglec varies, ranging from 2 (in CD33/Siglec-3) to 17 (in Sn/Siglec-1). The amino-terminal Ig-like domain, belonging to the V-set Ig-like folds, contains several amino acid residues critically important for glycan recognition, including an arginine residue that makes a charge interaction with the carboxyl group of sialic acids [25, 26]. Other Ig-like domains of Siglecs are smaller (belonging to the C2-set Ig-like folds), and whose functions not clear, except for the second Ig-like domain, which is apparently involved in glycan recognition by some Siglecs. The first two Ig-like domains of MAG/Siglec-4 are tethered by an inter-domain disulfide bond [27], and the conservation of 3 cysteine residues in each of the first and second Ig-like domains of most other Siglecs suggests this is the case for others as well.

Possible functions of Ig-like domains other than the first two may include the following: (1) optimal presentation of glycan-recognition site, as suggested for Sn/Siglec-1 [28] and Siglec-5 [29]; (2) involvement in glycan-independent interaction with some ligands, as suggested for MAG/Siglec-4 – Nogo66 receptor interaction [30, 31] and CD22/Siglec-2 – surface IgM interaction [32]; (3) display of glycans that are recognized by other lectins, as is the case with Sn/Siglec-1 serving as a ligand for other macrophage lectins [33, 34].

Glycan recognition by Siglecs

Glycan recognition by Siglecs requires sialic acids, including the C7–C9 exocyclic side chain (although there are some exceptions). The mere presence of sialic acid is not enough for recognition by Siglecs, however. Basic structures recognized by Siglecs contain the Sia−Gal–HexNAc sequence (although Neu5Acα2-6GalNAc and Neu5Acα2-8Neu5Ac are also recognized by some Siglecs). The type of sialic acid (e.g., Neu5Ac or Neu5Gc), its linkage to Gal (α2–3 or α2–6), and the structure of underlying

glycans (e.g., presence of Fuc on GlcNAc, presence of sulfate on Gal or GlcNAc, etc.) can all affect recognition by Siglecs. Human and mouse Sn/Siglec-1 preferentially binds Neu5Acα2-3Galβ1-4GlcNAc (neither tolerate substitution of Neu5Ac with Neu5Gc [35, 36]), human and mouse CD22/Siglec-2 binds Siaα2-6Galβ1-4GlcNAc (while human one tolerates both Neu5Ac and Neu5Gc, mouse one prefers Neu5Gc [35–38]), and rat MAG/Siglec-4 and quail SMP/Siglec-4 preferentially bind to Neu5Acα2-3Galβ1-3[HSO_3-6 or Neu5Acα2-6]GalNAc, again not tolerating Neu5Gc [36, 39–41]. As for CD33rSiglecs, each member shows different preferences, and the overall preference seems to be more promiscuous than with the others. However, recent development of glycan arrays has brought a new dimension to the analysis of fine specificity of glycan-binding proteins [42]. For example, human Siglec-8 and mouse Siglec-F were shown to specifically interact with Neu5Acα2-3[HSO_3-6]Galβ1-4 [Fucα1-3]GlcNAc (6′-sulfo-sialyl-Lewis^x structure) [43, 44]. Also, using a more classical approach, Siglec-7 has been shown to preferentially bind to Neu5Acα2-3Galβ1-3[Neu5Acα2-6]HexNAc [45, 46]. Taken together, it is possible that glycan recognition by CD33rSiglecs (or by some of them, at least) is far more specific than previously thought, requiring not only terminal sialic acids but also another acidic moiety in the vicinity, along with some other structural elements.

Regardless, it should be kept in mind that different assay formats have yielded somewhat different results from each other, and even the same assay format gives different results depending on some parameters, especially the concentration/amount of glycan-containing probes [36]. Also, extensive study on Siglec preference of various naturally occurring modified sialic acids has not been undertaken, because of the instability of these modifications at C4 and/or C7–C9 hydroxyl groups and the lack of defined synthetic routes.

Signal transduction by Siglecs

The cytoplasmic domain of most Siglecs contains sequence motif(s) involved in signal transduction. CD22 and most CD33rSiglecs contain an immunoreceptor tyrosine-based inhibitory motif (ITIM) [47], which can recruit protein tyrosine phosphatases SHP-1 and SHP-2 upon phosphorylation of the tyrosine residue. Based on *in vitro* studies, the negative regulatory signaling of these Siglecs is attributed to the ITIM motif and its interaction with SHP-1 [48–55].

Most CD33rSiglecs and MAG/Siglec-4 has another tyrosine-based conserved sequence motif closer to the C-terminal, although its involvement in signal transduction remains speculative [56]. Both the ITIM and the putative motif also conform to the consensus binding motif of adapter protein 2 (AP2) complex (YxxØ), where Ø stands for amino acid residue with bulky hydrophobic side chain [57]. Therefore it is possible that these motifs are involved in rapid internalization of Siglecs on antibody cross-linking [32, 58].

Some Siglecs (e.g., mouse CD33/Siglec-3 and Siglec-H) lack one or both of these motifs but instead have a positively charged amino acid residue in the transmembrane domain. They might interact with activating adapter molecules such as DAP12 and FcεRγ, which have a negatively charged amino acid residue in the transmembrane domain and interact with transmembrane receptors with a positively charged amino acid residue in the transmembrane domain [59]. In fact, it was recently reported that mouse Siglec-H associates with DAP12 [60].

Coupling of glycan recognition and signal transduction

Glycan binding at the extracellular domain of Siglecs and signaling at the cytosolic domain appear to be coupled. While extracellular recognition of ligands affects intracellular signaling [61, 62], phosphorylation of intracellular amino acid residues can also alter extracellular ligand binding [63–66]. Although the exact mechanism for this coupling between extra- and intracellular events is not yet understood, it may involve clustering of Siglecs on cell surface [67], which will affect both avidity of ligand recognition and signaling function. An *N*-glycosylation site conserved in the first Ig-like domain of most Siglecs, right opposite of the glycan-binding face [26], may serve as a ligand for another Siglec molecule and facilitate such homophilic clustering [68]. It has been reported that the *N*-glycosylation of this site is essential for CD22/Siglec-2 binding to sialylated glycans [69], while a mutation of equivalent site does not alter glycan recognition by the MAG/Siglec-4, Siglec-5 or Siglec-7 [70, 71], or enhanced glycan recognition by CD33/Siglec-3 [69]. However, these studies have been done using COS or CHO cells, and it should be interesting to test if point-mutation of the *N*-glycosylation site in Siglecs results in altered glycan-binding and/or signaling properties when expressed in the cells more similar to the ones that naturally carry them.

Biological functions of Siglecs

Biological functions of some Siglecs have been analyzed *in vivo*, using genetically modified mice as animal models. Deficiency of mouse MAG/Siglec-4 results in reduced stability of myelin sheaths [72, 73]. Genetic deletion of mouse CD22/Siglec-2 results in exaggerated calcium response by B-cell receptor cross-linking and various changes in B-cell biology (e.g., life span, subpopulations, antibody production, etc.), although the phenotypes of different CD22-deficient strains appear to be inconsistent, and to depend on the genetic background of mice and/or the experimental conditions used [74–77]. Mice deficient for CD33/Siglec-3 have been also generated, although their phenotypes appear relatively mild [78]. Sn/Siglec-1-deficient mice have been also generated [79], although full characterization of these mice is yet to be reported.

As for CD33rSiglecs, appropriate animal models for functional *in vivo* studies of human Siglecs cannot be

generated in many cases, because of the difference between humans and mice in this subfamily. Therefore, functional studies of human Siglecs have mostly utilized cell lines, primary cell cultures and freshly isolated cells from patients and healthy donors. The method of choice has been antibody cross-linking of Siglecs and analysis of subsequent cellular events. Antibodies against some Siglecs (e.g., CD33/Siglec-3, Siglec-8, and Siglec-9) have been reported to induce apoptosis [80–82], and the one against Siglec-7 was reported to induce arrest of myeloid differentiation [83] and inhibition of NK cell activation [52]. These reports are consistent with the view that CD33rSigelcs in general (the ones with ITIM, to be exact) are negative regulatory receptors.

Diversity and genomic evolution of Siglecs

Distribution of Siglecs amongst species

Siglecs are found in all branches of vertebrates (Table 1). There is thus far no definitive evidence of any Siglec-like molecule in other branches of deuterostomes, including *Ciona intestinalis*, a sea squirt species whose genome has been sequenced [84], and *Strongylocentrotus purpuratus*, a sea urchin species whose EST and genome sequencing projects are under way [85, 86].

The most highly conserved Siglec is MAG/Siglec-4 (SMP appears to be its bird ortholog), which is found even in fishes. It was reported that fish MAG/Siglec-4 showed similar glycan binding preference to mammalian counterparts [87]. Other Siglecs show varying degree of conservation. Sn/Siglec-1 is found in birds but not in amphibians or fishes, and CD22/Siglec-2 is not even found in birds so far (it should be noted, however, that the sequencing of the chicken chromosome syntenic to human chromosome 19, on which most Siglec genes lie, is incomplete at the time of writing). The fact that CD22/Siglec-2 is not found in non-mammalian vertebrates is worth mentioning, because the genes for MAG/Siglec-4 and CD22/Siglec-2 are adjacent to each other in mammalian genome. Thus, CD22/Siglec-2 could be derived from MAG/Siglec-4 by gene duplication in the mammalian ancestor. CD33rSiglecs, even as a group, could not be spotted in the genomes or in EST sequences of vertebrates other than those of mammals. There appear to be some potential Siglecs found only in fishes and amphibians (Angata et al., unpublished observations), although their relationships to mammalian Siglecs are not clear.

MAG/Siglec-4 might be the founding member of this family, giving rise to other members (Sn/Siglec-1, CD22 and CD33rSiglecs) over the course of vertebrate evolution. However, other scenarios are also possible, e.g., the real founding member of the family was a Siglec expressed in immune system (where molecular evolution tend to be very fast) that has evolved beyond recognition, and MAG/Siglec-4 is highly conserved (enough to be recognized to have been conserved throughout vertebrate evolution) only because it is exclusively expressed in nervous system (where molecular evolution tend to be relatively slow). Many factors, including the rapid evolution of the members expressed in immune system and uneven pace of molecular evolution amongst family members, make it difficult to delineate the evolutionary history of the whole Siglec family with confidence.

Extensive diversity of mammalian CD33rSiglecs

A draft dog genome sequence was completed recently [88], allowing us for the first time to do comparative sequence analysis of mammalian genomes beyond those of primates and rodents, whose genomic sequences have been available previously. In the following section I describe some patterns of mammalian Siglec diversity deduced by such an analysis, although it may be premature to generalize without analyzing Siglec sequences of other branches of mammals (especially of those species that have split out early in the mammalian evolution, e.g., monotremes and marsupials). Dog Siglec genes were sought and predicted based on sequence similarity to known Siglec genes, and CD33rSiglecs were named in their locus order on the chromosome 1. The nomenclature for dog CD33rSiglecs in this review is *ad hoc* for the convenience of description, and not authorized by any nomenclature committee.

As mentioned earlier, Sn/Siglec-1, CD22/Siglec-2, and MAG/Siglec-4 are conserved between primates and rodents, but CD33rSiglecs are generally less conserved. Molecular phylogenetic analysis of all Siglecs in human, mouse, and dog (Figure 3) indicates that this pattern is applicable to carnivores as well. Thus, mammalian Siglecs can be classified into two subgroups: CD33rSiglecs, which are poorly conserved amongst species, and the others (Sn/Siglec-1, CD22/Siglec-2, and MAG/Siglec-4), which are better conserved.

In mammalian genome, most CD33rSiglec genes are clustered in a defined chromosomal region called the Siglec gene cluster [21, 22] (Figure 4). This fact suggests that CD33rSiglec subfamily has expanded via repeated gene duplications. Duplicated genes in a gene cluster tend to undergo complex pattern of evolution. Some of the duplicated genes become pseudogenes, some are recombined to form chimeric genes, while some are deleted by chromosomal rearrangements. It appears that all of these processes have taken place in the Siglec gene cluster [22], explaining dynamics behind the great diversity of this subfamily.

The CD33rSiglecs can now be further classified into some subgroups (e.g., Siglec-3-like, Siglec-9/E-like, Siglec-5-like, and Siglec-10/G-like groups) based on sequence similarity, as shown in Figure 3. Combined with the map positions and structures of these genes, orthologous correspondence may be tentatively assigned amongst some of the CD33rSiglecs in these species. For example, the following triads may be considered orthologs: [dog CD33rSiglec #4, human Siglec-10, and mouse Siglec-G], [dog CD33rSiglec #2, human CD33/Siglec-3, and mouse CD33/Siglec-3], and

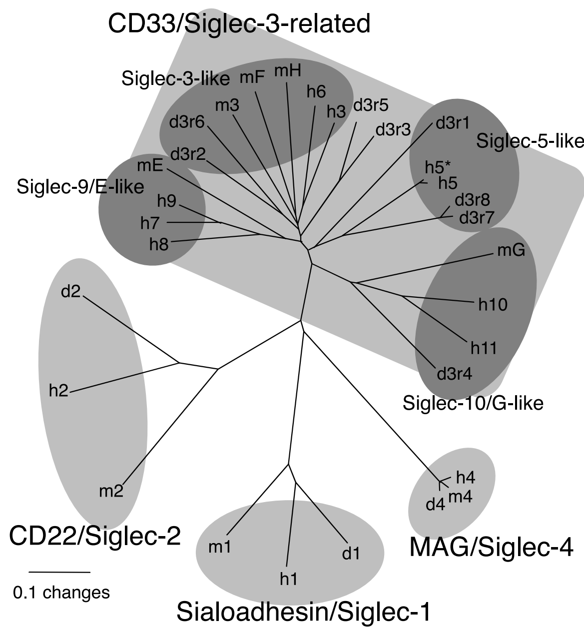

Figure 3. Molecular phylogenetic tree of mammalian Siglecs. Amino acid sequences of the N-terminal part (signal peptide + first two Ig-like domains) of all Siglecs in human, mouse, and dog were aligned, and used for reconstruction of the molecular phylogenetic tree by the neighbor-joining method. Dog Siglec-like sequences were sought in the draft dog genome sequence [88] based on similarity to known human and mouse Siglecs. The letters d, h, and m represent dog, human, and mouse, and numbers 1, 2, and 4 represent Sn/Siglec-1, CD22/Siglec-2, and MAG/Siglec-4, respectively (e.g., d4 stands for dog MAG/Siglec-4). Human CD33rSiglecs are indicated with numbers (e.g., h6 stands for human Siglec-6), while mouse CD33rSiglecs with alphabets (e.g., mE stands for mouse Siglec-E). Dog CD33rSiglecs were named in the order of their gene loci on the chromosome 1 (e.g., d3r4 stands for dog CD33rSiglec, #4). Human Siglec-12 was not included, because it has two V-set domains and cannot be aligned properly with all other Siglecs with only one V-set domain. Humans lack Siglec-13 [22].

[dog CD33rSiglec #6, human Siglec-6, and mouse Siglec-F]. If we compare Siglecs in only two species, we find more putative orthologous pairs: [human Siglec-9 and mouse Siglec-E] and [dog CD33rSiglecs #7 and human Siglec-5]. However, the validity of such "orthologous correspondence" assignment based on sequence similarity, map position and gene structure may be limited, because some genes appear to be chimeras of ancestral genes, generated either by recombination of two genes or partial gene conversion by another gene [22, 89]. Thus, different parts of a gene may have different evolutionary history, and only a part of gene X in one species may be orthologous to gene X' in another species.

In addition to these genome-level diversification discussed above, CD33rSiglec subfamily shows signs of rapid nucleotide-level diversification that affects amino acids (those of the first Ig-like domain in particular), which can be detected by sequence comparison of orthologous Siglecs in closely related species [22]. This issue will be discussed in detail later.

Loss of Neu5Gc in humans and changes in Siglecs

How the diversification of sialic acid structures has affected endogenous lectin functions is a question yet to be fully answered, but we now have some clues to answer this question: the loss of Neu5Gc in humans and concomitant adaptive functional changes of Siglecs.

Neu5Gc is one of the sialic acids forming the basis of structural variations, and abundantly expressed in most mammalian tissues (except for brains, where Neu5Gc is almost absent). However, the lineage leading to modern humans has lost its expression, as a result of the inactivating mutation in the gene encoding for CMP-Neu5Ac hydroxylase (gene symbol: *CMAH*) [90–92]. In contrast, chimpanzees and other great apes, the closest living evolutionary relatives of humans, still have the active enzyme and express Neu5Gc [93]. It has been suggested, but not proven that this human-specific loss of Neu5Gc is related to escape from pathogens. Regardless of the original reason, this situation provides a clear-cut opportunity to look into what have happened to sialic acid-binding proteins after such a drastic change in the glycan landscape.

First, what happened to the Siglecs that had preference to Neu5Gc over Neu5Ac? Chimpanzee Siglec-12 (originally named Siglec-L1 [94] or S2V [95]), also found in other great apes, has strong preference to Neu5Gc [94]. In humans, however, the gene is mutated at the essential arginine codon required for optimal sialic acid recognition, diminishing glycan binding of the protein almost completely. Introduction of a "reverse-mutation", which restores the arginine residue, revived the glycan-binding activity of human Siglec-12, which also strongly preferred Neu5Gc. Therefore, the mutation (which is shared among all humans tested so far) could be considered "functional retirement of a Siglec" due to the loss of preferred ligand. Some other chimpanzee CD33rSiglecs (e.g., Siglec-7 and Siglec-9) also have strong preference to Neu5Gc, while the orthologs in humans bind both Neu5Ac- and Neu5Gc-containing glycans equally well [96]. In these cases, the loss of Neu5Gc in humans has presumably led to the change of sialic acid preference of these Siglecs from "strong Neu5Gc preference" to "accommodation of Neu5Ac". Siglec-13, which is apparently missing in humans but present in other primates [22], might represent another example of Siglec affected by the loss of Neu5Gc, perhaps in this case leading to "elimination of a Siglec", although there is no experimental data regarding its sialic acid preference.

The loss of Neu5Gc also results in an increase in the amount of its precursor, Neu5Ac. What, then, have happened to the Siglecs that prefer Neu5Ac? In the case of Sn/Siglec-1, human cells have an excess of binding sites, and the localization of macrophages expressing Sn/Siglec-1 has changed [35]. The distribution of Sn/Siglec-1-positive macrophages in chimpanzee spleen is more similar to that in rat spleen, in that only a subset of macrophages expresses Sn/Siglec-1 and populates mainly marginal zone and surrounding periarteriolar lymphoid sheaths. In contrast, most macrophages

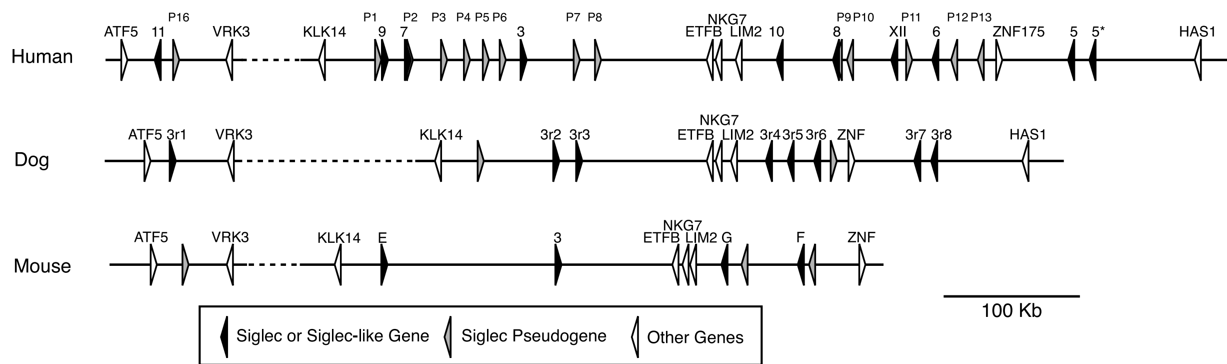

Figure 4. Siglec gene cluster in human, mouse, and dog genomes. Genes and pseudogenes for CD33rSiglecs and adjacent genes are represented with *filled*, *open* and *gray arrowheads*, respectively. The direction of each arrowhead represents the direction of transcription. Human Siglec genes were indicated with numbers (e.g., 5 stands for *SIGLEC5,* and P3 stands for pseudogene *SIGLECP3*), while mouse CD33rSiglec genes were denoted with alphabets (e.g., E stands for *SiglecE*). Dog CD33rSiglec genes (predicted by sequence similarity to known Siglecs) were named in their locus order on the chromosome 1. The nomenclature for dog CD33rSiglecs in this figure is *ad hoc* for the convenience of description in this review and not authorized by any nomenclature committee. The region containing *SIGLEC11* in human genome and its syntenic region in mouse and dog genomes was included in the figure, although *SIGLEC11* is usually not considered a part of the cluster for its great distance from the "main" part of the cluster (about 1 Mb apart from the main cluster in human genome).

in human spleen are positive for Sn/Siglec-1 and populate perifollicullar zone. Thus, the change of Sn/Siglec-1 expression and localization appears to be a derived state found only in humans. Although there is no study on MAG/Siglec-4 in this respect, the loss of Neu5Gc in humans should not have affected the microenvironment surrounding MAG/Siglec-4 much from that in chimpanzees, because the central nervous system is a tissue extremely poor in Neu5Gc in any mammals [97]. Although there is no study on human MAG/Siglec-4 glycan binding specificity, it is expected to be the same as rodent MAG/Siglec-4 which strongly prefer Neu5Ac [36, 38, 41], for these show almost complete (97%) amino acid sequence identity at the first two Ig-like domains involved in glycan recognition.

Finally, what have happened to the Siglecs that do not have strong preference to Neu5Gc or Neu5Ac? Nothing, perhaps. Great ape CD22/Siglec-2 recognize glycans containing both Neu5Ac and Neu5Gc equally well, and there seems to be no qualitative change in the preference of human CD22/Siglec-2 compared with its great ape orthologs [35].

In conclusion, the loss of Neu5Gc in humans has likely led to multiple changes regarding Siglec biology. The adjustment of Siglecs to a new environment regarding sialic acids (or "sialome") led to some losses, but overall they seem to have adjusted fairly well.

Accelerated evolution of CD33-related Siglecs and the dynamics behind it

Given that the loss of Neu5Gc in humans resulted in multiple changes in biology of Siglecs as listed above, particularly those related to the CD33rSiglecs, one would expect that it has left recognizable marks on their sequences. The CD33rSiglec subfamily is one of the ten fastest-evolving gene family deduced by human-chimpanzee genomic sequence comparison [98], and we previously reported that CD33rSiglecs are undergoing accelerated evolution at the first Ig-like domain, based on the sequence comparison of Siglec gene cluster in human, chimpanzee, and baboon [22]. Does this fact mean that the loss of Neu5Gc in humans has driven the rapid sequence change in the first Ig-like domain of CD33rSiglecs? The answer may not be simple. If the loss of Neu5Gc in humans has driven the accelerated evolution of the first Ig-like domain of CD33rSiglecs, the number of changes accumulated in the affected human genes should exceed those accumulated in the orthologs in other species (i.e., baboon and chimpanzee), and branch length (reflecting the number of changes in each lineage) for the affected human genes in a molecular phylogenetic tree should be longer than those for the orthologs in others. However, we did not observe such human gene-specific excessive accumulation of changes compared with orthologs in chimpanzee or baboon. We also noticed that some Siglecs in chimpanzee and baboon have the essential arginine residue mutated, just like human Siglec-12, and that some baboon Siglecs are missing altogether, just like human Siglec-13 (Table 2 and ref. [22]). Thus, not only the accelerated evolution of CD33rSiglecs, but also the drastic changes in Siglec gene cluster, are common amongst primates. This fact seems to force us to consider alternative explanations. It is possible, however, that changes in the sialome in each species, as drastic as the loss of Neu5Gc in humans, are taking place in other primates as well (and in rodents, which showed lesser but significant degree of the first Ig-like domain-specific accelerated evolution), and driving the evolution of CD33rSiglecs in these species. Given that the sialic acids are constantly exploited by pathogens, it may not be such an outlandish idea that pathogen-driven change in the sialome is actually going on in each lineage, which may not be as visible as the complete loss of a class of sialic acids but nevertheless has an equally large cumulative effect on the recognition by Siglecs over the course of

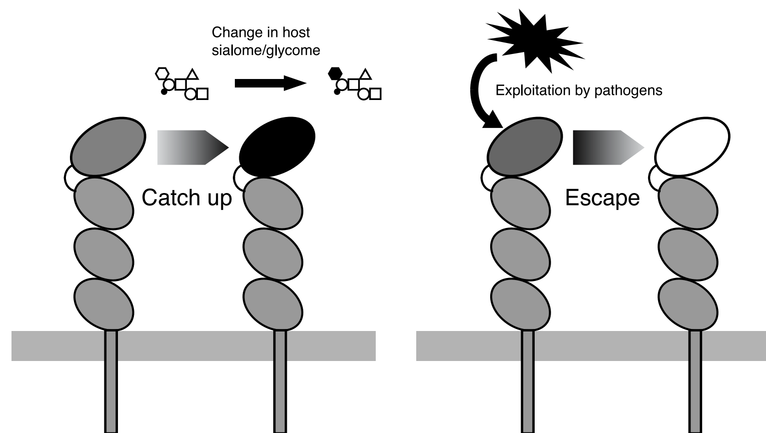

Figure 5. Two possible evolutionary forces behind accelerated evolution of Siglecs. Siglecs may be undergoing accelerated evolution, driven by their need to catch up with the changing landscape of own glycans, which may be changing to avoid exploitation by pathogens (*Left* panel). Alternatively, Siglecs may be rapidly evolving by direct competition against pathogens that utilize Siglecs for their benefit (*Right* panel).

evolution. Biochemical analysis of sialomes in these species is necessary to test this hypothesis.

If "change in sialome" is not enough to explain the observed pattern of evolution, then what else could? An additional hypothesis may be "direct competitions between Siglecs and pathogens". Siglecs may be functioning as "receptors" for some viruses and/or bacteria, which may or may not be related to the Sia-recognition of Siglecs. Some pathogenic bacteria synthesize and modify cell surface with Neu5Ac (but not Neu5Gc, which has not been found so far in bacteria), and can be recognized by some Siglecs [79]. Some viruses are coated by "envelope" glycoproteins, which are derived from host cell membrane and should contain sialylated glycans. In fact, it has been reported that a porcine envelope virus utilizes Sn/Siglec-1 to initiate infection in macrophages, and this interaction is dependent on the presence of sialic acid on virus surface [99, 100]. It is also possible that cross-linking of Siglecs by sialylated bacteria may transmit negative regulatory signaling into the host cells involved in innate immunity, and thus dampening Toll-like receptor-mediated pro-inflammatory response (e.g., cytokine release), an important initial reaction against bacterial infection. Thus, not only sialic acids but also Siglecs themselves may have to evolve to avoid exploitation by pathogens.

These two hypotheses are not mutually excusive, and Siglecs may be evolving both to catch up with the changing host sialome (which in turn may be evolving to avoid utilization by pathogens), and to escape exploitation by pathogens (Figure 5). Our finding that the first Ig-like domain is the fastest-evolving domain of Siglecs, and the mutations is accumulating more rapidly on the face of the domain which interact with sialylated glycans (which is the face opposite to the one protected by an *N*-glycan, attached to the conserved *N*-glycosylation site [22]), is consistent with both hypotheses.

Future directions

Which of the two forces driving the evolution of CD33rSiglecs is dominant, and/or if there are any other driving forces, is a question to be answered by using multiple approaches, such as biochemical analysis of sialome and glycome, comparative genomics of Siglec gene cluster, and analysis of glycan binding preference of Siglecs, all in multiple related species. We still lack efficient means to synthesize glycans containing most of naturally occurring sialic acids, and development of such methodologies would greatly facilitate our understanding of sialic acid and Siglec biology. With regard to comparative genomics, sequences of Siglec gene cluster of some New World monkeys (e.g., marmoset) are awaited. Loss of α1-3-galactosyltransferase in Old World primates (monkeys and apes) [101–103] should have resulted in overall increase in the amount of sialic acids expressed, for the α1-3-galactosyltransferase competes against several sialyltransferases over the common acceptor structure, Galβ1-4GlcNAc. Such a drastic change in glycan landscape may have also left some marks on the CD33rSiglecs. With an increase in the availability of genome sequences of other animals, more comprehensive comparative "sialome – Siglec-ome" studies may become feasible. Analysis of glycome, especially the analysis of glycans containing sialic acids is still considered a great technical challenge, and development of such technology, probably based on mass spectrometry, is awaited. Thus, the quest for understanding Siglec biology requires development of several related fields.

Future research of Siglecs should also include analysis of Siglec-mediated effects on host-pathogen interactions. To facilitate this, and to understand "regular" or house-keeping functions of Siglecs, search for biologically relevant endogenous ligands, analysis of signal transduction pathway(s) elicited by the engagement of Siglecs and their ligands, and

analysis of the mechanism which couple extra- and intracellular events, are necessary.

With regard to the issue of the loss of Neu5Gc in humans, establishment of chimpanzee (and other ape/monkey) monocytic cell lines would be of great value, enabling us via simulation to analyze what functional consequences have resulted by the loss (e.g., by shutting down the Neu5Gc synthesis with RNA interference against CMP-Neu5Ac hydroxylase mRNA). In vitro evolution of chimpanzee Siglecs preferring Neu5Gc in an environment that selects Neu5Ac-preferring molecules may give us clues to recognize pattern of sequence changes required for the adjustment. Such an experiment may eventually lead to creation of "designer" Siglecs that recognize defined structure of sialylated glycans, which may be of use for diagnosis of some diseases that accompany change of sialylation pattern of serum and/or cell-surface glycoconjugates.

There is little doubt that the research of Siglecs will lead to advancement of our understanding of sialic acid biology and glycobiology in general, and also to technological developments that will benefit human health.

Acknowledgments

The author's research is supported by "Support of young researchers with a term" from Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Varki, A., *Biological roles of oligosaccharides: All of the theories are correct*, Glycobiology, 3 (1993) 97–130.
2. Gagneux, P. and Varki, A., *Evolutionary considerations in relating oligosaccharide diversity to biological function*, Glycobiology, 9 (1999) 747–755.
3. Freeze, H.H., *Update and perspectives on congenital disorders of glycosylation*, Glycobiology, 11 (2001) 129R–143R.
4. Jaeken, J. and Carchon, H., *Congenital disorders of glycosylation: A booming chapter of pediatrics*, Curr. Opin. Pediatr., 16 (2004) 434–439.
5. Aebi, M. and Hennet, T., *Congenital disorders of glycosylation: Genetic model systems lead the way*, Trends Cell Biol., 11 (2001) 136–141.
6. Angata, T. and Varki, A., *Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective*, Chem. Rev., 102 (2002) 439–470.
7. Kelm, S. and Schauer, R., *Sialic acids in molecular and cellular interactions*, Int. Rev. Cytol., 175 (1997) 137–240.
8. Suzuki, Y., *Sialobiology of influenza: Molecular mechanism of host range variation of influenza viruses*, Biol. Pharm. Bull., 28 (2005) 399–408.
9. Holmgren, J., Lonnroth, I., Mansson, J. and Svennerholm, L., *Interaction of cholera toxin and membrane GM1 ganglioside of small intestine*, Proc. Natl. Acad. Sci. U. S. A., 72 (1975) 2520–2524.
10. Mahdavi, J., Sonden, B., Hurtig, M., Olfat, F.O., Forsberg, L., Roche, N., Angstrom, J., Larsson, T., Teneberg, S., Karlsson, K.A., Altraja, S., Wadstrom, T., Kersulyte, D., Berg, D.E., Dubois, A., Petersson, C., Magnusson, K.E., Norberg, T., Lindh, F., Lundskog, B.B., Arnqvist, A., Hammarstrom, L. and Boren, T., *Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation*, Science, 297 (2002) 573–578.
11. Orlandi, P.A., Klotz, F.W. and Haynes, J.D., *A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognizes the terminal Neu5Ac(alpha 2-3)Gal- sequences of glycophorin A*, J. Cell Biol., 116 (1992) 901–909.
12. Kawano, T., Koyama, S., Takematsu, H., Kozutsumi, Y., Kawasaki, H., Kawashima, S., Kawasaki, T. and Suzuki, A., *Molecular cloning of cytidine monophospho-N-acetylneuraminic acid hydroxylase. Regulation of species- and tissue-specific expression of N-glycolylneuraminic acid*, J. Biol. Chem., 270 (1995) 16458–16463.
13. Kawano, T., Kozutsumi, Y., Kawasaki, T. and Suzuki, A., *Biosynthesis of N-glycolylneuraminic acid-containing glycoconjugates. Purification and characterization of the key enzyme of the cytidine monophospho-N-acetylneuraminic acid hydroxylation system*, J. Biol. Chem., 269 (1994) 9024–9029.
14. Schwarzkopf, M., Knobeloch, K.P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W. and Horstkorte, R., *Sialylation is essential for early development in mice*, Proc. Natl. Acad. Sci. U. S. A., 99 (2002) 5267–5270.
15. Eisenberg, I., Avidan, N., Potikha, T., Hochner, H., Chen, M., Olender, T., Barash, M., Shemesh, M., Sadeh, M., Grabov-Nardini, G., Shmilevich, I., Friedmann, A., Karpati, G., Bradley, W.G., Baumbach, L., Lancet, D., Asher, E.B., Beckmann, J.S., Argov, Z. and Mitrani-Rosenbaum, S., *The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy*, Nat. Genet., 29 (2001) 83–87.
16. Martinez-Duncker, I., Dupre, T., Piller, V., Piller, F., Candelier, J.J., Trichet, C., Tchernia, G., Oriol, R. and Mollicone, R., *Genetic complementation reveals a novel human congenital disorder of glycosylation of type II, due to inactivation of the Golgi CMP-sialic acid transporter*, Blood, 105 (2005) 2671–2676.
17. Crocker, P.R. and Varki, A., *Siglecs, sialic acids and innate immunity*, Trends Immunol., 22 (2001) 337–342.
18. Crocker, P.R., *Siglecs in innate immunity*, Curr. Opin. Pharmacol., 5 (2005) 431–437.
19. Varki, A. and Angata, T., *Siglecs – the major subfamily of I-type lectins*, Glycobiology, 16 (2006) 1R–27R.
20. Crocker, P.R., Clark, E.A., Filbin, M., Gordon, S., Jones, Y., Kehrl, J.H., Kelm, S., Le Douarin, N., Powell, L., Roder, J., Schnaar, R.L., Sgroi, D.C., Stamenkovic, K., Schauer, R., Schachner, M., Van den Berg, T.K., Van der Merwe, P.A., Watt, S.M. and Varki, A., *Siglecs: A family of sialic-acid binding lectins [letter]*, Glycobiology, 8 (1998) v.
21. Angata, T., Hingorani, R., Varki, N.M. and Varki, A., *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 (2001) 45128–45136.
22. Angata, T., Margulies, E.H., Green, E.D. and Varki, A., *Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms*, Proc. Natl. Acad. Sci. U. S. A., 101 (2004) 13251–13256.
23. Crocker, P.R. and Varki, A., *Siglecs in the immune system*, Immunology, 103 (2001) 137–145.
24. Patel, N., Brinkman-Van der Linden, E.C.M., Altmann, S.W., Gish, K., Balasubramanian, S., Timans, J.C., Peterson, D., Bell, M.P., Bazan, J.F., Varki, A. and Kastelein, R.A., *OB-BP1/Siglec-6 - A leptin- and sialic acid-binding protein of the immunoglobulin superfamily*, J. Biol. Chem., 274 (1999) 22729–22738.
25. May, A.P., Robinson, R.C., Vinson, M., Crocker, P.R. and Jones, E.Y., *Crystal structure of the N-terminal domain of sialoadhesin in complex with 3′ sialyllactose at 1.85 A resolution*, Mol. Cell., 1 (1998) 719–728.
26. Alphey, M.S., Attrill, H., Crocker, P.R. and Van Aalten, D.M., *High resolution crystal structures of siglec-7. Insights into ligand specificity in the Siglec family*, J. Biol. Chem., 278 (2003) 3372–3377.

27. Pedraza, L., Owens, G.C., Green, L.A. and Salzer, J.L., *The myelin-associated glycoproteins: Membrane disposition, evidence of a novel disulfide linkage between immunoglobulin-like domains, and posttranslational palmitylation*, J. Cell Biol., 111 (1990) 2651–2661.
28. Crocker, P.R., Mucklow, S., Bouckson, V., McWilliam, A., Willis, A.C., Gordon, S., Milon, G., Kelm, S. and Bradfield, P., *Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains*, EMBO J., 13 (1994) 4490–4503.
29. Connolly, N.P., Jones, M. and Watt, S.M., *Human Siglec-5: Tissue distribution, novel isoforms and domain specificities for sialic acid-dependent ligand interactions*, Br. J. Haematol., 119 (2002) 221–238.
30. Domeniconi, M., Cao, Z.U., Spencer, T., Sivasankaran, R., Wang, K.C., Nikulina, E., Kimura, N., Cai, H., Deng, K.W., Gao, Y., He, Z.G. and Filbin, M.T., *Myelin-associated glycoprotein interacts with the Nogo66 receptor to inhibit neurite outgrowth*, Neuron, 35 (2002) 283–290.
31. Liu, B.P., Fournier, A., GrandPré, T. and Strittmatter, S.M., *Myelin-associated glycoprotein as a functional ligand for the Nogo-66 receptor*, Science, 297 (2002) 1190–1193.
32. Zhang, M. and Varki, A., *Cell surface sialic acids do not affect primary CD22 interactions with CD45 and sIgM, nor the rate of constitutive CD22 endocytosis*, Glycobiology, 14 (2004) 939–949.
33. Kumamoto, Y., Higashi, N., Denda-Nagai, K., Tsuiji, M., Sato, K., Crocker, P.R. and Irimura, T., *Identification of sialoadhesin as a dominant lymph node counter-receptor for mouse macrophage galactose-type C-type lectin 1*, J. Biol. Chem., 279 (2004) 49274–49280.
34. Martínez-Pomares, L., Crocker, P.R., Da Silva, R., Holmes, N., Colominas, C., Rudd, P., Dwek, R. and Gordon, S., *Cell-specific glycoforms of sialoadhesin and CD45 are counter-receptors for the cysteine-rich domain of the mannose receptor*, J. Biol. Chem., 274 (1999) 30325–30318.
35. Brinkman-Van der Linden, E.C.M., Sjoberg, E.R., Juneja, L.R., Crocker, P.R., Varki, N. and Varki, A., *Loss of N-glycolylneuraminic acid in human evolution – Implications for sialic acid recognition by siglecs*, J. Biol. Chem., 275 (2000) 8633–8640.
36. Blixt, O., Collins, B.E., Van Den Nieuwenhof, I.M., Crocker, P.R. and Paulson, J.C., *Sialoside specificity of the Siglec family assessed using novel multivalent probes: Identification of potent inhibitors of myelin associated glycoprotein*, J. Biol. Chem., 278 (2003) 31007–31019.
37. Van der Merwe, P.A., Crocker, P.R., Vinson, M., Barclay, A.N., Schauer, R. and Kelm, S., *Localization of the putative sialic acid-binding site on the immunoglobulin superfamily cell-surface molecule CD22*, J. Biol. Chem., 271 (1996) 9273–9280.
38. Kelm, S., Brossmer, R., Isecke, R., Gross, H.J., Strenge, K. and Schauer, R., *Functional groups of sialic acids involved in binding to siglecs (sialoadhesins) deduced from interactions with synthetic analogues*, Eur. J. Biochem., 255 (1998) 663–672.
39. Collins, B.E., Ito, H., Sawada, N., Ishida, H., Kiso, M. and Schnaar, R.L., *Enhanced binding of the neural siglecs, myelin-associated glycoprotein and Schwann cell myelin protein, to Chol-1 (alpha-series) gangliosides and novel sulfated Chol-1 analogs*, J. Biol. Chem., 274 (1999) 27899–27893.
40. Collins, B.E., Kiso, M., Hasegawa, A., Tropak, M.B., Roder, J.C., Crocker, P.R. and Schnaar, R.L., *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 (1997) 16889–16895.
41. Collins, B.E., Yang, L.J.S., Mukhopadhyay, G., Filbin, M.T., Kiso, M., Hasegawa, A. and Schnaar, R.L., *Sialic acid specificity of myelin-associated glycoprotein binding*, J. Biol. Chem., 272 (1997) 1248-1255.
42. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M.E., Alvarez, R., Bryan, M.C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D.J., Skehel, J.J., van Die, I., Burton, D.R., Wilson, I.A., Cummings, R., Bovin, N., Wong, C.H. and Paulson, J.C., *Printed covalent glycan array for ligand profiling of diverse glycan binding proteins*, Proc. Natl. Acad. Sci. U. S. A., 101 (2004) 17033–17038.
43. Bochner, B.S., Alvarez, R.A., Mehta, P., Bovin, N.V., Blixt, O., White, J.R. and Schnaar, R.L., *Glycan array screening reveals a candidate ligand for siglec-8*, J. Biol. Chem., 280 (2005) 4307–4312.
44. Tateno, H., Crocker, P.R. and Paulson, J.C., *Mouse Siglec-F and human Siglec-8 are functionally convergent paralogs that are selectively expressed on eosinophils and recognize 6′-sulfo-sialyl Lewis X as a preferred glycan ligand*, Glycobiology, 15 (2005) 1125–1135.
45. Ito, A., Handa, K., Withers, D.A., Satoh, M. and Hakomori, S., *Binding specificity of siglec7 to disialogangliosides of renal cell carcinoma: Possible role of disialogangliosides in tumor progression*, FEBS Lett., 498 (2001) 116–120.
46. Miyazaki, K., Ohmori, K., Izawa, M., Koike, T., Kumamoto, K., Furukawa, K., Ando, T., Kiso, M., Yamaji, T., Hashimoto, Y., Suzuki, A., Yoshida, A., Takeuchi, M. and Kannagi, R., *Loss of disialyl Lewis(a), the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with increased sialyl Lewis(a) expression on human colon cancers*, Cancer Res., 64 (2004) 4498–4505.
47. Ravetch, J.V. and Lanier, L.L., *Immune inhibitory receptors*, Science, 290 (2000) 84–89.
48. Doody, G.M., Justement, L.B., Delibrias, C.C., Matthews, R.J., Lin, J., Thomas, M.L. and Fearon, D.T., *A role in B cell activation for CD22 and the protein tyrosine phosphatase SHP*, Science, 269 (1995) 242–244.
49. Blasioli, J., Paust, S. and Thomas, M.L., *Definition of the sites of interaction between the protein tyrosine phosphatase SHP-1 and CD22*, J. Biol. Chem., 274 (1999) 2303–2307.
50. Ulyanova, T., Blasioli, J., Woodford-Thomas, T.A. and Thomas, M.L., *The sialoadhesin CD33 is a myeloid-specific inhibitory receptor*, Eur. J. Immunol., 29 (1999) 3440–3449.
51. Paul, S.P., Taylor, L.S., Stansbury, E.K. and McVicar, D.W., *Myeloid specific human CD33 is an inhibitory receptor with differential ITIM function in recruiting the phosphatases SHP-1 and SHP-2*, Blood, 96 (2000) 483–490.
52. Falco, M., Biassoni, R., Bottino, C., Vitale, M., Sivori, S., Augugliaro, R., Moretta, L. and Moretta, A., *Identification and molecular cloning of p75/AIRM1, a novel member of the sialoadhesin family that functions as an inhibitory receptor in human natural killer cells*, J. Exp. Med., 190 (1999) 793–801.
53. Whitney, G., Wang, S.L., Chang, H., Cheng, K.Y., Lu, P., Zhou, X.D., Yang, W.P., McKinnon, M. and Longphre, M., *A new siglec family member, siglec-10, is expressed in cells of the immune system and has signaling properties similar to CD33*, Eur. J. Biochem., 268 (2001) 6083–6096.
54. Ulyanova, T., Shah, D.D. and Thomas, M.L., *Molecular cloning of MIS, a myeloid inhibitory siglec, that binds protein-tyrosine phosphatases SHP-1 and SHP-2*, J. Biol. Chem., 276 (2001) 14451–14458.
55. Ikehara, Y., Ikehara, S.K. and Paulson, J.C., *Negative regulation of T cell receptor signaling by Siglec-7 (p70/AIRM) and Siglec-9*, J. Biol. Chem., 279 (2004) 43117–43125.
56. Kitzig, F., Martinez-Barriocanal, A., López-Botet, M. and Sayós, J., *Cloning of two new splice variants of Siglec-10 and mapping of the interaction between Siglec-10 and SHP-1*, Biochem. Biophys. Res. Commun., 296 (2002) 355–362.
57. Bonifacino, J.S. and Traub, L.M., *Signals for sorting of transmembrane proteins to endosomes and lysosomes*, Annu. Rev. Biochem., 72 (2003) 395–447.
58. John, B., Herrin, B.R., Raman, C., Wang, Y.N., Bobbitt, K.R., Brody, B.A. and Justement, L.B., *The B cell coreceptor CD22 associates with AP50, a clathrin-coated pit adapter protein, via tyrosine-dependent interaction*, J. Immunol., 170 (2003) 3534–3543.
59. Lanier, L.L. and Bakker, A.B., *The ITAM-bearing transmembrane adaptor DAP12 in lymphoid and myeloid cell function*, Immunol. Today, 21 (2000) 611–614.

60. Blasius, A.L., Cella, M., Takai, T. and Colonna, M., *Siglec-H is an IPC-specific receptor that modulates type I IFN secretion through DAP12*, Blood, 107 (2006) 2474–2476.
61. Jin, L., McLean, P.A., Neel, B.G. and Wortis, H.H., *Sialic acid binding domains of CD22 are required for negative regulation of B cell receptor signaling*, J. Exp. Med., 195 (2002) 1199–1205.
62. Kelm, S., Gerlach, J., Brossmer, R., Danzer, C.P. and Nitschke, L., *The ligand-binding domain of CD22 is needed for inhibition of the B cell receptor signal, as demonstrated by a novel human CD22-specific inhibitor compound*, J. Exp. Med., 195 (2002) 1207–1213.
63. Taylor, V.C., Buckley, C.D., Douglas, M., Cody, A.J., Simmons, D.L. and Freeman, S.D., *The myeloid-specific sialic acid-binding receptor, CD33, associates with the protein-tyrosine phosphatases, SHP-1 and SHP-2*, J. Biol. Chem., 274 (1999) 11505–11512.
64. Avril, T., Floyd, H., Lopez, F., Vivier, E. and Crocker, P.R., *The membrane-proximal immunoreceptor tyrosine-based inhibitory motif is critical for the inhibitory signaling mediated by Siglecs-7 and -9, CD33-related Siglecs expressed on human monocytes and NK cells*, J. Immunol., 173 (2004) 6841–6849.
65. Avril, T., Freeman, S.D., Attrill, H., Clarke, R.G. and Crocker, P.R., *Siglec-5 (CD170) can mediate inhibitory signalling in the absence of immunoreceptor tyrosine-based inhibitory motif phosphorylation*, J. Biol. Chem., 280 (2005) 19843–19851.
66. Grobe, K. and Powell, L.D., *Role of protein kinase C in the phosphorylation of CD33 (Siglec-3) and its effect on lectin activity*, Blood, 99 (2002) 3188–3196.
67. Powell, L.D., Jain, R.K., Matta, K.L., Sabesan, S. and Varki, A., *Characterization of sialyloligosaccharide binding by recombinant soluble and native cell-associated CD22. Evidence for a minimal structural recognition motif and the potential importance of multisite binding*, J. Biol. Chem., 270 (1995) 7523–7532.
68. Han, S., Collins, B.E., Bengston, P. and Paulson, J.C., *Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking*, Nat. Chem. Biol., 1 (2005) 93–97.
69. Sgroi, D., Nocks, A. and Stamenkovic, I., *A single N-linked glycosylation site is implicated in the regulation of ligand recognition by the I-type lectins CD22 and CD33*, J. Biol. Chem., 271 (1996) 18803–18809.
70. Tropak, M.B. and Roder, J.C., *Regulation of myelin-associated glycoprotein binding by sialylated cis-ligands*, J. Neurochem., 68 (1997) 1753–1763.
71. Freeman, S., Birrell, H.C., D'Alessio, K., Erickson-Miller, C., Kikly, K. and Camilleri, P., *A comparative study of the asparagine-linked oligosaccharides on siglec-5, siglec-7 and siglec-8, expressed in a CHO cell line, and their contribution to ligand recognition*, Eur. J. Biochem., 268 (2001) 1228–1237.
72. Bartsch, U., Bandtlow, C.E., Schnell, L., Bartsch, S., Spillmann, A.A., Rubin, B.P., Hillenbrand, R., Montag, D., Schwab, M.E. and Schachner, M., *Lack of evidence that myelin-associated glycoprotein is a major inhibitor of axonal regeneration in the CNS*, Neuron, 15 (1995) 1375–1381.
73. Schachner, M. and Bartsch, U., *Multiple functions of the myelin-associated glycoprotein MAG (siglec-4a) in formation and maintenance of myelin*, Glia, 29 (2000) 154–165.
74. O'Keefe, T.L., Williams, G.T., Davies, S.L. and Neuberger, M.S., *Hyperresponsive B cells in CD22-deficient mice*, Science, 274 (1996) 798–801.
75. Otipoby, K.L., Andersson, K.B., Draves, K.E., Klaus, S.J., Farr, A.G., Kerner, J.D., Perlmutter, R.M., Law, C.L. and Clark, E.A., *CD22 regulates thymus-independent responses and the lifespan of B cells*, Nature, 384 (1996) 634–637.
76. Sato, S., Miller, A.S., Inaoki, M., Bock, C.B., Jansen, P.J., Tang, M.L. and Tedder, T.F., *CD22 is both a positive and negative regulator of B lymphocyte antigen receptor signal transduction: Altered signaling in CD22-deficient mice*, Immunity, 5 (1996) 551–562.
77. Nitschke, L., Carsetti, R., Ocker, B., Kohler, G. and Lamers, M.C., *CD22 is a negative regulator of B-cell receptor signalling*, Curr. Biol., 7 (1997) 133–143.
78. Brinkman-Van der Linden, E.C., Angata, T., Reynolds, S.A., Powell, L.D., Hedrick, S.M. and Varki, A., *CD33/Siglec-3 binding specificity, expression pattern, and consequences of gene deletion in mice*, Mol. Cell. Biol., 23 (2003) 4199–4206.
79. Jones, C., Virji, M. and Crocker, P.R., *Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake*, Mol. Microbiol., 49 (2003) 1213–1225.
80. Vitale, C., Romagnani, C., Puccetti, A., Olive, D., Costello, R., Chiossone, L., Pitto, A., Bacigalupo, A., Moretta, L. and Mingari, M.C., *Surface expression and function of p75/AIRM-1 or CD33 in acute myeloid leukemias: Engagement of CD33 induces apoptosis of leukemic cells*, Proc. Natl. Acad. Sci. U. S. A., 98 (2001) 5764–5769.
81. Nutku, E., Aizawa, H., Hudson, S.A. and Bochner, B.S., *Ligation of Siglec-8: A selective mechanism for induction of human eosinophil apoptosis*, Blood, 101 (2003) 5014–5020.
82. von Gunten, S., Yousefi, S., Seitz, M., Jakob, S.M., Schaffner, T., Seger, R., Takala, J., Villiger, P.M. and Simon, H.U., *Siglec-9 transduces apoptotic and nonapoptotic death signals into neutrophils depending on the proinflammatory cytokine environment*, Blood, 106 (2005) 1423–1431.
83. Vitale, C., Romagnani, C., Falco, M., Ponte, M., Vitale, M., Moretta, A., Bacigalupo, A., Moretta, L. and Mingari, M.C., *Engagement of p75/AIRM1 or CD33 inhibits the proliferation of normal or leukemic myeloid cells*, Proc. Natl. Acad. Sci. U. S. A., 96 (1999) 15091–15096.
84. Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M., Harafuji, N., Hastings, K.E., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I.A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-Bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D.N., Haga, S., Hayashi, H., Hino, K., Imai, K.S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.I., Makabe, K.W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M.M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P.D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N. and Rokhsar, D.S., *The draft genome of Ciona intestinalis: Insights into chordate and vertebrate origins*, Science, 298 (2002) 2157–2167.
85. Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R., Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L., *A sea urchin genome project: Sequence scan, virtual map, and additional resources*, Proc. Natl. Acad. Sci. U. S. A., 97 (2000) 9514–9518.
86. Poustka, A.J., Groth, D., Hennig, S., Thamm, S., Cameron, A., Beck, A., Reinhardt, R., Herwig, R., Panopoulou, G. and Lehrach, H., *Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters*, Genome. Res., 13 (2003) 2736–2746.
87. Lehmann, F., Gathje, H., Kelm, S. and Dietz, F., *Evolution of sialic acid-binding proteins: Molecular cloning and expression of fish siglec-4*, Glycobiology, 14 (2004) 959–968.
88. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J. III, Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr,

M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.P., Parker, H.G., Pollinger, J.P., Searle, S.M., Sutter, N.B., Thomas, R., Webber, C., Broad Institute Genome Sequencing Platform and Lander, E.S., *Genome sequence, comparative analysis and haplotype structure of the domestic dog*, Nature, 438 (2005) 803–819.

89. Hayakawa, T., Angata, T., Lewis, A.L., Mikkelsen, T.S., Varki, N.M. and Varki, A., *A human-specific gene in microglia*, Science, 309 (2005) 1693.
90. Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T. and Suzuki, A., *The molecular basis for the absence of N-glycolylneuraminic acid in humans*, J. Biol. Chem., 273 (1998) 15866–15871.
91. Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Paabo, S., Satta, Y., Takahata, N. and Varki, A., *Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution*, Proc. Natl. Acad. Sci. U. S. A., 99 (2002) 11736–11741.
92. 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., *A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence*, Proc. Natl. Acad. Sci. U. S. A., 95 (1998) 11751–11756.
93. Muchmore, E.A., Diaz, S. and Varki, A., *A structural difference between the cell surfaces of humans and the great apes*, Am. J. Phys. Anthropol., 107 (1998) 187–198.
94. Angata, T., Varki, N.M. and Varki, A., *A second uniquely human mutation affecting sialic acid biology*, J. Biol. Chem., 276 (2001) 40282–40287.
95. Yu, Z., Lai, C.M., Maoui, M., Banville, D. and Shen, S.H., *Identification and characterization of S2V, a novel putative siglec that contains two V set Ig-like domains and recruits protein-tyrosine phosphatases SHPs*, J. Biol. Chem., 276 (2001) 23816–23824.
96. Sonnenburg, J.L., Altheide, T.K. and Varki, A., *A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor*, Glycobiology, 14 (2004) 339–346.
97. Varki, A., *Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences and implications for hominid evolution*, Yearbook Phys. Anthropol., 44 (2002) 54–69.
98. The Chimpanzee Sequencing and Analysis Consortium, *Initial sequence of the chimpanzee genome and comparison with the human genome*, Nature, 437 (2005) 69–87.
99. Vanderheijden, N., Delputte, P.L., Favoreel, H.W., Vandekerckhove, J., Van Damme, J., Van Woensel, P.A. and Nauwynck, H.J., *Involvement of sialoadhesin in entry of porcine reproductive and respiratory syndrome virus into porcine alveolar macrophages*, J. Virol., 77 (2003) 8207–8215.
100. Delputte, P.L. and Nauwynck, H.J., *Porcine arterivirus infection of alveolar macrophages is mediated by sialic acid on the virus*, J. Virol., 78 (2004) 8094–8101.
101. Galili, U., Clark, M.R., Shohet, S.B., Buehler, J. and Macher, B.A., *Evolutionary relationship between the natural anti-Gal antibody and the Gal alpha1-3Gal epitope in primates*, Proc. Natl. Acad. Sci. U. S. A., 84 (1987) 1369–1373.
102. Galili, U., Shohet, S.B., Kobrin, E., Stults, C.L. and Macher, B.A., *Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells*, J. Biol. Chem., 263 (1988) 17755–17762.
103. Galili, U. and Swanson, K., *Gene sequences suggest inactivation of α-1,3-galactosyltransferase in catarrhines after the divergence of apes from monkeys*, Proc. Natl. Acad. Sci. U. S. A., 88 (1991) 7401–7404.
104. International Chicken Genome Sequencing Consortium, *Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution*, Nature, 432 (2004) 695–716.
105. Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., Gelpke, M.D., Roach, J., Oh, T., Ho, I.Y., Wong, M., Detter, C., Verhoef, F., Predki, P., Tay, A., Lucas, S., Richardson, P., Smith, S.F., Clark, M.S., Edwards, Y.J., Doggett, N., Zharkikh, A., Tavtigian, S.V., Pruss, D., Barnstead, M., Evans, C., Baden, H., Powell, J., Glusman, G., Rowen, L., Hood, L., Tan, Y.H., Elgar, G., Hawkins, T., Venkatesh, B., Rokhsar, D. and Brenner, S., *Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes*, Science, 297 (2002) 1301–1310.
106. Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigo, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J. and Roest Crollius, H., *Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype*, Nature, 431 (2004) 946–957.
107. Angata, T., Kerr, S.C., Greaves, D.R., Varki, N.M., Crocker, P.R. and Varki, A., *Cloning and characterization of human Siglec-11. A recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia*, J. Biol. Chem., 277 (2002) 24466–24474.