A Second Uniquely Human Mutation Affecting Sialic Acid Biology*[$]

Received for publication, June 26, 2001, and in revised form, August 1, 2001
Published, JBC Papers in Press, August 23, 2001, DOI 10.1074/jbc.M105926200

Takashi Angata, Nissi M. Varki, and Ajit Varki‡

From the Glycobiology Research and Training Center, Departments of Medicine and Cellular and Molecular Medicine, University of California at San Diego, La Jolla, California 92093-0687

Siglecs are immunoglobulin superfamily member lec-
tins that selectively recognize different types and link-
ages of sialic acids, which are major components of cell
surface and secreted glycoconjugates. We report here a
human Siglec-like molecule (Siglec-L1) that lacks a con-
served arginine residue known to be essential for opti-
mal sialic acid recognition by previously known Siglecs.
Loss of the arginine from an ancestral molecule was
caused by a single nucleotide substitution that occurred
after the common ancestor of humans with the great
apes but before the origin of modern humans. The chim-
panzee Siglec-L1 ortholog remains fully functional and
preferentially recognizes N-glycolylneuraminic acid,
which is a common sialic acid in great apes and other
mammals. Reintroducing the ancestral arginine into the
human molecule regenerates the same properties. Thus,
the single base pair mutation that replaced the arginine
on human Siglec-L1 is likely to be evolutionarily related
to the previously reported loss of N-glycolylneuraminic
acid expression in the human lineage. Siglec-L1 and its
chimpanzee Siglec ortholog also have a different expres-
sion pattern from previously reported Siglecs because
they are found on the luminal edge of epithelial cell
surfaces. Notably, the human genome contains several
Siglec-like pseudogenes that have independent muta-
tions that would have replaced the arginine residue re-
quired for optimal sialic acid recognition. Thus, addi-
tional changes in the biology of sialic acids may have
taken place during human evolution.

The great apes are the closest evolutionary relatives of hu-
mans, with the chimpanzee/bonobo clade likely having shared
a last common ancestor with humans about 6–7 million years
ago (1–5). Human genomic DNA sequences differ on average by
only 1–2% from those of these great apes (3–5). Thus, the altered expression of relatively few gene products may underlie
some of the obvious morphological and functional differences
between the species. We and others recently discovered an
inactivating mutation in the human gene encoding CMP-
N-acety neuraminic acid hydroxylase (6, 7), an enzyme that is
functional in the great apes (6). This explains the human-
specific loss of a major sialic acid, N-glycolylneuraminic acid
(Neu5Gc)[$] (8).

The sialic acids are a family of 9-carbon sugars that are
abundantly expressed on cell surfaces and secreted glycoconju-
gates of animals of the deuterostome lineage (9–11). Located
mostly at the outer end of glycan chains on glycoproteins and
glycolipids, sialic acids mediate a variety of recognition events
involving pathogenic microbes and toxins, as well as endoge-
nous animal lectins (12, 13). Siglecs are the largest family of
such endogenous sialic acid-recognizing lectins defined to date
(14, 15). 10 reported human Siglecs are type I membrane
proteins, consisting of an amino-terminal Ig V-set domain,
variable numbers of Ig C2-set domains, a single-pass trans-
membrane domain, and a cytoplasmic tail typically containing
tyrosine-based signaling motifs. Sialic acid recognition is me-
diated by the first Ig V-set domain (15–19), and certain amino
acid residues invariant to this domain are known to be involved
in interactions with the sialic acid ligand (20). In particular, all
Siglec V-set domains have a conserved arginine residue that
forms a salt bridge with the carboxylate group of sialic acids.
Experimental mutation of this residue markedly diminishes
binding in all Siglecs studied to date (18, 19, 21–23).

Here we report a human molecule that has many features of
Siglecs but lacks robust sialic acid recognition because of a
specific mutation of the “essential” arginine residue, which
remains conserved in its great ape orthologs. We also consider
potential connections to the previously described human-spe-
cific mutation involving sialic acid biology and search the hu-
man genome for Siglec-like pseudogenes, several of which turn
out to have independent mutations replacing the “essential
arginine residue. After this work was completed, another group
independently reported the cloning of the same molecule,
which they called S2V, a “putative” Siglec (24). Our study
shows that the “essential” arginine residue, which was mutated
to cysteine specifically in the human lineage, is required for
easily detectable sialic acid-dependent recognition typical of
previously reported Siglecs. By comparison, we demonstrate
robust sialic acid-dependent recognition by the fully functional
chimpanzee ortholog and by the human molecule when the arginine is restored. Thus, we do not consider the native hu-
man molecule a bona fide Siglec; instead, we call it Siglec-like
molecule 1 (Siglec-L1).

EXPERIMENTAL PROCEDURES

DNA Cloning and Preparation of Expression Constructs—A human
expressed sequence tag clone (GenBank™ accession number A132995)
encoding transmembrane and cytosolic domains of Siglec-L1 was iden-

* This work was supported by United States Public Health Service
Grants R01-GM323373 and P01-HL57345 and by the G. Harold and
Leila Y. Mathers Charitable Foundation. The costs of publication of this
article were defrayed in part by the payment of page charges. This
article must therefore be hereby marked “advertisement” in accordance
with 18 U.S.C. Section 1734 solely to indicate this fact.

[$] The on-line version of this article (available at http://www.jbc.org)
contains supplementary information.

The nucleotide sequence(s) reported in this paper has been submitted
to the GenBank™/EBI Data Bank with accession number(s) AF282256
(human), AF293372 (chimpanzee), AY029752 (bonobo), AY029754
(gorilla), and AY029756 (orangutan).

‡ To whom correspondence should be addressed: University of Cali-
fornia at San Diego School of Medicine, 9500 Gilman Dr., MC 0867, La
Jolla, CA 92093-0687. Tel.: 858-534-3296; Fax: 858-534-5611; E-mail:
avarki@ucsd.edu.

1 The abbreviations used are: Neu5Gc, N-glycolylneuraminic acid;
Neu5Ac, N-acetylneuraminic acid; PCR, polymerase chain reaction;
Siglec-L1, Siglec-like molecule 1; UTR, untranslated region; LacNac,
N-acetyllactosamine.
Inactivation of a Siglec by a Uniquely Human Mutation

40823

tified by BLAST search (25) of the GenBank™ data base. A cDNA for the Siglec-L1 full-length coding sequence was obtained by 5'-rapid amplification of cDNA ends from human fetal liver Marathon-Ready cDNA library (CLONTECH) using primers 5'-R5P1 (5'-GGGAGGTTTGCTTGATCAGGCAAG-3') and AP1 (included in the kit) followed by nested PCR using 3'-R5P2 (5'CGATGGCTCTTGGCAGACTGTCATG-3') and AP2 (included in the kit). The 3'-untranslated region (UTR) was obtained likewise, using 3'-R5P1 (5'-GGTGGGCTTGAGATACGGGCTAGG-3') and human IgG Fc tail primers 5'-R5P2 (5'-CCAGCTCGGG-3') and AP1 for the first PCR, followed by a nested PCR using 3'-R5P2 (5'-CCAGATGCACTCTCCATGCTCC-3') and AP2. The DNA fragments were cloned into pCRII-TOPO (Invitrogen), and the sequences were analyzed. A 32P-labeled DNA fragment was used as a probe to screen clones containing full-length cDNA. A cDNA for Siglec-L1/pcDNA3.1 using the QuickChange Site-Directed Mutagenesis kit (Stratagene) was digested with XhoI and HindIII and ligated into appropriate sites of pcDNA3.1 for functional analysis. The PCR product with 5' Exr and 3' UTR (containing full coding sequence) was digested with XhoI and HindIII and ligated into appropriate sites of pcDNA3.1 for functional analysis. Mammalian cell transfection with this construct gives a fusion protein of Siglec-L1 (first three Ig-like domains), which was verified by Western blot analysis (data not shown).

The Genomic Structures of Human and Chimpanzee Tissues

Chimpanzee tissues were kindly provided by Yerkes Primate Center. Paraffin-embedded formalin-fixed tissue sections mounted on glass slides were deparaffinized and incubated overnight at 4°C with 1:200 dilutions of rabbit polyclonal anti-human IgG Fc or preimmune serum (100 µg/mL) immobilized on protein A-Sepharose (Amersham Pharmacia Biotech). The post-immune serum processed as described above was specific for Siglec-L1, as judged by an enzyme-linked immunosorbent assay using Siglec-L1-Fc, Siglec-7-Fc, and L-selectin-Fc as test samples (data not shown).

Phylogenetic Analysis of V-set Domains of Siglec Genes and Pseudogenes

Amino acid sequences of known Siglec V-set domains were used to find similar sequences (E values < 1) in the human genome data base in GenBank™ using the TBLASTN program. Relevant DNA sequences were selected for functional analysis based on the presence of exons and donor/acceptor sites. The full-length DNA sequence for Siglec-L1 was reconstructed by comparing Siglec-8 cDNA (the closest paralog) with human Siglec-7 pseudogene outside the cluster was numbered P1. The only human pseudogene outside the cluster was numbered P14. This numbering system is as recommended by the Human Gene Nomenclature Committee. One DNA segment that we could not unambiguously assign as a gene or pseudogene was named X. The precise order of some pseudogenes was determined with the help of the publicly accessible portion of the Celera human genome data base (30).

DNA sequences of the exons and exon fossils of V-set domains were aligned by Clustal W (www.ebi.ac.uk/clustalw/) and minimally adjusted using six anchor points (i.e. codons for two aromatic amino acid residues and the arginine residue involved in sialic acid recognition, and three cysteines conserved in known Siglecs (20)). Coding sequences for the signal peptide were then removed because some pseudogenes lack a well-defined signal sequence, and the remaining sequences were analyzed for phylogenetic relationship using PAUP 4.0 (Sinauer Associates). Phylogenetic trees were constructed using the neighbor-joining method (31) with a distance matrix based on absolute distance. Pseudogene P12 was excluded from the analysis because its coding sequence was disrupted by an insertion of an Alu element.

Genomic Structure of Siglec-L1 and Its Close Paralogues—Genomic structures of Siglec-L1, Siglec-7, Siglec-8, Siglec-9, and mouse Siglec-E were deduced by comparing the cDNA (Siglec-7, NM_014385; Siglec-8, NM_014442 and AF287892; Siglec-9, AF227924; mouse Siglec-E, AF317298) and genomic DNA sequences of these genes in the GenBank™ database. The "coding" sequence of pseudogene Siglec-P4 was reconstructed by comparing Siglec-8 cDNA (the closest paralog) with the genomic DNA region containing Siglec-P4. The Siglec-7 genomic DNA buffer saline 8 h after transfection, treated with *Arthrobacter ureafaciens* sialidase (10 million units/106 cells) at 37°C for 1 h, and washed extensively with 1% bovine serum albumin in phosphate-buffered saline. The cells (105) were incubated with 1 µg of PAA-Bio probes carrying Neu5Ac-LacNAc, Neu5Gc-LacNAc, or non-sialylated LacNAc at 4°C overnight. After washing 3 times, the samples were incubated with 1 µg of streptavidin-phycocyanin-erythrin conjugate (Jackson ImmunolResearch) at 4°C for 30 min. The cells were washed again, suspended in 1% bovine serum albumin-phosphate-buffered saline, and subjected to flow cytometry using FACScan (Becton Dickinson). Green fluorescence-positive cells (FL1high, the cells successfully transfected) were gated and analyzed for probe binding.
DNA region in the data base appeared to be misassembled; it was therefore reassembled from some overlapping fragments in draft-stage sequences of BAC clones RP11-423F16 and CTD-3187F8. Repetitive elements were identified by using RepeatMasker (repeatmasker. genome.washington.edu/cgi-bin/RepeatMasker).

RESULTS AND DISCUSSION

A Human Siglec-like Molecule Lacks Arginine Residue(s) Essential for Robust Sialic Acid Recognition—A sequence in the GenBank™ expressed sequence tag data base (GenBank™ accession number AI132995) showed close similarity to the sequences of BAC clones RP11-423F16 and CTD-3187F8. Repetitive elements were identified by using RepeatMasker (repeatmasker. genome.washington.edu/cgi-bin/RepeatMasker).

Great Ape Orthologs Have the Arginine Residue in the First Domain and Can Recognize Sialic Acids—DNA fragments including the first four exons of the Siglec-L1 orthologs of great apes were amplified from genomic DNA of the chimpanzee, bonobo, gorilla, and orangutan, and the exon sequences were analyzed. All great ape orthologs have the “essential” arginine residue in the first Ig-like domain (Fig. 1B), encoded by a CG/T/C) codon. Thus, a single nucleotide substitution (C to T) at the first residue of the Arg codon replaced the ancestral Arg with a Cys residue in humans. On the other hand, all great apes as well as humans lack the “essential” arginine residue in the second Ig-like domain (Fig. 1B). To confirm that the great ape orthologs are indeed fully functional Siglecs, we cloned a full-length cDNA encoding the chimpanzee ortholog (GenBankTM accession number AF293372). When expressed in COS-7 cells by transient transfection, this molecule gave robust sialic acid-dependent erythrocyte rosetting (Fig. 2A).

Restoration of the Arginine Residue in the First V-set Domain of the Human Siglec-L1 Regenerates Robust Sialic Acid Recognition—A reverse mutation (C122R) restoring the ancestral three amino-terminal Ig-like domains of the molecule also failed to show sialic acid binding (data not shown) in a standard enzyme-linked immunosorbent assay (23, 26). Because designation as a Siglec requires sialic acid recognition (14), we have named this protein Siglec-like molecule-1 (Siglec-L1). Given the close sequence similarity of the first V-set domain of Siglec-L1 to the corresponding domains of Siglec-7 and Siglec-9, we hypothesized that the lack of the “essential” arginine residue (universal in humans, see below) represents a derived change from the ancestral state. To explore this possibility, we studied Siglec-L1 orthologs in the great apes.

FIG. 1. Sequences of human Siglec-L1 and its great ape orthologs. A. CDNA and deduced amino acid sequences of human Siglec-L1 (GenBank™ AF282256). Double circles, amino acids occupying the expected positions of the “essential” arginine in typical Siglec V-set domains; circles, aromatic amino acid residues in the two V-set domains typical of Siglec V-set domains; underline with hatched and double lines, signal peptide and transmembrane domain, respectively; arrowheads, exon junctions; underline, potential N-glycosylation sites; boxes with solid and hatched lines, putative ITIM motif and another tyrosine-based motif conserved among Siglecs, respectively. B. Alignment of deduced amino acid sequences of the two amino-terminal V-set domains of human Siglec-L1 and the great ape orthologs. Amino acid residues that differ from human Siglec-L1 are indicated. The expected positions of the “essential” arginine in the first and second V-set domains are indicated with filled and open triangles, respectively.
arginine residue was introduced in the human Siglec-L1 cDNA and studied in the same rosetting assay as described above. As shown in Fig. 2C, the “restored” human molecule also showed robust sialic acid-dependent erythrocyte binding. Thus, the single nucleotide substitution that changed the Arg codon to a Cys codon was primarily responsible for diminishing the sialic acid binding property of the human molecule.

There is an apparent discrepancy with another recent study (24) that claimed sialic acid recognition by the native human molecule. This might be explained by the fact that those authors counted transfected COS cells associated with only a few red cells in an assay that seemed to have a signal:noise ratio of about 4:1. In contrast, we saw robust sialic acid-dependent recognition by the chimpanzee ortholog as well as the arginine-restored human molecule, with large clusters of red cells attached to the transfected COS cells (see Fig. 2). Thus, we are confident that the native human molecule does not have robust sialic acid binding properties typical of other Siglecs. It is possible that the binding reported in the other study represents some residual recognition of other aspects of the sialic acid molecule that can occur even in the absence of the arginine residue.

The Arginine Mutation in the First Domain Occurred before the Common Origin of Modern Humans—There is now almost universal agreement that all living humans are very closely related genetically (33, 34). To determine when the arginine-replacing nucleotide change in the first Ig-like domain occurred relative to the common origin of modern humans, we studied human genomic DNA samples from six Asians, eight Europeans/European-Americans, five African-Americans, and five Africans (two Bia and three Mbuti pygmies). All individuals studied had the same point mutation changing the Arg codon to a Cys codon, indicating that this substitution occurred before the common origin of modern humans. We also noted that some human alleles have a frameshift mutation that must have occurred after the common origin of modern humans. This is currently being investigated further.

Both Chimpanzee and Arginine-restored Human Siglecs Preferentially Recognize Neu5Gc, the Sialic Acid That Is Missing in Humans—This finding represents the second human-specific genetic mutation that involves the biology of sialic acids; the first results in the loss of Neu5Gc expression (6). To explore whether the two events are functionally related, we studied the relative preference of the arginine-restored human Siglec-L1 and the chimpanzee ortholog for their ability to recognize Neu5Ac and Neu5Gc. Probes carrying equal amounts of sialic acids in the Neu5Ac or Neu5Gc form were used to study cells transfected with human wild-type Siglec-L1, the human C122R reverse mutant, or the chimpanzee ortholog. As shown in Fig. 3, the human wild-type Siglec-L1-expressing cells showed no binding of either probe in this assay. In contrast, both the human C122R molecule and the chimpanzee ortholog bound both sialic acids. However, both clearly preferred Neu5Gc over Neu5Ac. Because fluorescence intensity is recorded on a log scale, we can also say that any binding of the wild-type molecule to either sialic acid (if present) is markedly diminished.

Tissue Expression Profiles of Human Siglec-L1 and the Chimpanzee Ortholog—A specific polyclonal chicken antibody against Siglec-L1 stained the luminal edge of epithelial cells in organs such as the prostate, ileum, stomach and tonsil, and collecting duct cells in the kidney medulla (see Fig. 4 for examples). A similar expression pattern was noted in chimpanzee tissues. To date, there are only two examples of Siglec expression outside the hematopoietic system: Siglec-4 in the nervous system (35), and Siglec-6 in the placenta (26). This is the first example of a Siglec or Siglec-like molecule that is predominantly expressed in epithelial cells. Not all human samples showed positive staining, which may be explained by the presence of frameshift mutations in some human alleles.

The Human Genome Contains Several Siglec Pseudogenes with Independent Mutations Replacing the Essential Arginine Residue—The characteristic amino acid motifs of the V-set domain of known Siglecs allowed us to search for other Siglec genes and eventually inactivated over the course of evolution. Interestingly, 7 of these 14 pseudogenes have mutations at the codon that would have replaced the “essential” arginine residue. Several of these “arginine-replacing” mutations seemed to have been discrete events (Fig. 5A). The currently available (incomplete) mouse genome sequence shows evidence for four Siglec-3-like Siglec genes and two Siglec-like...
pseudogenes in the syntenic regions of mouse chromosome, and one of the latter has an "arginine-replacing" mutation. Taken together, these data suggest that the Siglec-3-related genes underwent extensive duplications during mammalian evolution, and several of the resulting pseudogenes may have been initially inactivated for sialic acid binding by mutations changing the "essential" arginine residue. A less likely explanation for the high frequency of "arginine-replacing" mutations is that inactivation of functional Siglec genes by some other mutations was followed by CpG dinucleotide mutations to TpG/CpA (Ref. 36 and the references therein) to which the arginine codons (CGN) are susceptible. However, the frequency of these "arginine-replacing" mutations (in ~50% of the pseudogenes) seems to be too high to be explained by the latter mechanism. Additional studies of the corresponding regions of various primate genomes (as well as complete sequence data from the syntenic regions of other genomes such as those of the mouse and rat) are needed to define the series of genetic events that lead to the current human condition.

Genomic Structural Analysis Suggests a Complex Ancestry of Siglec-L1—The genomic structure of human Siglec-L1 was deduced from the human genome sequence and compared with those of its close paralogs. As shown in Fig. 5B, there is a high degree of similarity with the genomic structure of the Siglec-7 gene, suggesting that Siglec-7 and Siglec-L1 are sibling molecules generated by a relatively recent gene duplication. Of particular interest, the second Ig-like domain of Siglec-L1 is very similar to an exon fossil in the Siglec-7 gene (P2 in Fig. 5A; the open box in front of the Siglec-7 C2-set domain in Fig. 5B). The other close human paralogs (Siglec-8, Siglec-9, and pseudogene P4) and the mouse gene orthologous to this group, called “Siglec-E” (37), do not share this exon configuration (Fig. 5B). The tight association of three adjacent exons (V or V2-C2-linker) in all of these genes and the presence of a DNA transposon fossil (type MER20) between exons V1 and V2 of Siglec-L1 (and the corresponding regions of the Siglec-7 gene) suggest that the ancestral V1 exon of Siglec-L1 and Siglec-7 may have been initially inserted in 5′ region of the ancestral gene, possibly in association with the DNA transposon insertion (38). This V-set exon could have then functionally replaced the original one (now V2), allowing the arginine codon in V2 to mutate to glutamine in Siglec-L1 (and allowing this exon to become inactive in Siglec-7).

Conclusions and Future Prospects—We have described the second human-specific and human-universal genetic mutation that affects the biology of sialic acids, the first of which leads to the loss of expression of the sialic acid Neu5Gc. Given the fact...
that the chimpanzee Siglec-L1 and the “arginine restored” human Siglec-L1 show a preference for binding Neu5Gc over Neu5Ac, it seems particularly likely that the two events are evolutionarily linked. If so, which came first? One possibility is that the “arginine-replacing” mutation in human Siglec-L1 occurred first, thereby making conditions more permissive for the loss of Neu5Gc synthesis. The other possibility is that the loss of Neu5Gc resulted in the effective loss of function of this Siglec, thereby making conditions permissive for it to accumulate mutations. However, in the latter scenario, it is much more likely that a random gene inactivation event would have occurred, rather than this highly specific functional inactivation, mutating the arginine residue required for optimal sialic acid binding.

It is of course possible that human Siglec-L1 has other functions that are unrelated to the ability to recognize sialic acids and that such function(s) resulted in the continued maintenance of the open reading frame. The selective expression of these molecules in epithelial cells also raises some interesting issues. The presence of putative tyrosine-based signaling motifs in the cytosolic tail suggests that the great ape orthologs are (and that the human ancestral Siglec-L1 could have been) involved in the regulation of homeostasis of siaIylated glycoconjugates within the lumen of some organs. Indeed, others have shown that the membrane-proximal tyrosine-based motif of this molecule can interact with the tyrosine phosphatases SHP-1 and SHP-2 (24). The expression of this molecule on surfaces that come into contact with the environment also suggests its possible involvement in microbial infection (e.g., viral particles carrying sialic acids might be accidentally captured). Whereas these possibilities are obviously speculative, they must be taken into consideration when trying to explain the original selection pressure that resulted in the arginine-eliminating mutation in humans.

Many other questions arise from this work. When exactly did this mutation occur during the last ~6 million years since our common ancestor split from the chimpanzee/bonobo clade? Can the unusual expression profile of this Siglec-like molecule on epithelial surfaces explain apparent differences in the incidence of diseases like epithelial cancers between humans and chimpanzees (39)? Why do some human alleles have an additional frameshift mutation? Why are there so many Siglec-like pseudogenes with mutated “essential” arginines? Could these be the result of repeated fine-tuning of endogenous sialic acid recognition? Do these also represent human-specific mutations related to the loss of Neu5Gc in humans, or did some occur earlier during mammalian evolution? We are currently pursuing some of these questions, particularly by comparison of the corresponding regions of the great ape and other primate genomes. Such studies may also determine whether this mutation can explain some of the obvious morphological and functional differences between humans and our closest evolutionary cousins.

Acknowledgments—We thank Pascal Gagneux and Els Brinkman-van der Linden for helpful discussions.


REFERENCES

Supplementary information; Angata et al.

> Siglec-1
---------------------GCCAGGCCTCA------------TGGGGC
GTCTCCAGTCCCGAGGCGTGCAGGGTGTGAAGGGGTCTTGCCTGCTTAT
CCCTGCACTCTTACACGCTCTTCCC---TGCCGAGCT--GGAGTGCCGGACGGC
AT----CACG-GCCATCTGGTACTACGA------CTACTCGGGCCAGCGG
CAGGTGGTGAGCCACTCGGCGGAACCACAAAGCTGGTGCAACCTG
CTGCTGAAGAACCTTGCAGCCCAGGA--GACTCTGTGCTCTCAAACTTCCGC
TTGGAATATGAGGGCTCAAGGCTGCTGAT--ACGCGGAAATGCGAAGATGGAG
---------CACCTTGGTCACAGTAAC----------------------
--------

> Siglec-2
TTCTAGAATCTTGGCTTTCTTCAGTCTCAGTAAA---------TGGGTT
TTGAGCAAGCCTCAAAAAGCCTCTGACCCTGGGAGGAGCCTGCTGGAT
CCCTGCACCTACAGGAGGCCCAGTAGTGTCGGGGGCTGCGTCTGGAT
TCTCCAGTCCCGAGGCGTGCAGGGTGTGAAGGGGTCTTGCCTGCTTAT
CCCTGCACTCTTACACGCTCTTCCC---TGCCGAGCT--GGAGTGCCGGACGGC
AT----CACG-GCCATCTGGTACTACGA------CTACTCGGGCCAGCGG
CAGGTGGTGAGCCACTCGGCGGAACCACAAAGCTGGTGCAACCTG
CTGCTGAAGAACCTTGCAGCCCAGGA--GACTCTGTGCTCTCAAACTTCCGC
TTGGAATATGAGGGCTCAAGGCTGCTGAT--ACGCGGAAATGCGAAGATGGAG
---------CACCTTGGTCACAGTAAC----------------------
--------

> Siglec-3
-----GGGCCCTGGCTATGGATCCAAAT---------------TTCTGG
CTGCAAGTGCAGAAGTCGGTGACGGTGCGAGGAGGGCCTGTGCGTCCTTGT
GCCCTGCACCTACAGGAGGCCCAGTAGTGTCGGGGGCTGCGTCTGGAT
TCTCCAGTCCCGAGGCGTGCAGGGTGTGAAGGGGTCTTGCCTGCTTAT
CCCTGCACTCTTACACGCTCTTCCC---TGCCGAGCT--GGAGTGCCGGACGGC
AT----CACG-GCCATCTGGTACTACGA------CTACTCGGGCCAGCGG
CAGGTGGTGAGCCACTCGGCGGAACCACAAAGCTGGTGCAACCTG
CTGCTGAAGAACCTTGCAGCCCAGGA--GACTCTGTGCTCTCAAACTTCCGC
TTGGAATATGAGGGCTCAAGGCTGCTGAT--ACGCGGAAATGCGAAGATGGAG
---------CACCTTGGTCACAGTAAC----------------------
--------

> Siglec-4
--------------CCTCCCGAGGGGGTCAC-------------TGGGGT
GTCTCCAGTCCCGAGGCGTGCAGGGTGTGAAGGGGTCTTGCCTGCTTAT
CCCTGCACTCTTACACGCTCTTCCC---TGCCGAGCT--GGAGTGCCGGACGGC
AT----CACG-GCCATCTGGTACTACGA------CTACTCGGGCCAGCGG
CAGGTGGTGAGCCACTCGGCGGAACCACAAAGCTGGTGCAACCTG
CTGCTGAAGAACCTTGCAGCCCAGGA--GACTCTGTGCTCTCAAACTTCCGC
TTGGAATATGAGGGCTCAAGGCTGCTGAT--ACGCGGAAATGCGAAGATGGAG
---------CACCTTGGTCACAGTAAC----------------------
--------

> Siglec-5
-----GGGCCCTGGCTATGGATCCAAAT---------------TTCTGG
CTGCAAGTGCAGAAGTCGGTGACGGTGCGAGGAGGGCCTGTGCGTCCTTGT
GCCCTGCACCTACAGGAGGCCCAGTAGTGTCGGGGGCTGCGTCTGGAT
TCTCCAGTCCCGAGGCGTGCAGGGTGTGAAGGGGTCTTGCCTGCTTAT
CCCTGCACTCTTACACGCTCTTCCC---TGCCGAGCT--GGAGTGCCGGACGGC
AT----CACG-GCCATCTGGTACTACGA------CTACTCGGGCCAGCGG
CAGGTGGTGAGCCACTCGGCGGAACCACAAAGCTGGTGCAACCTG
CTGCTGAAGAACCTTGCAGCCCAGGA--GACTCTGTGCTCTCAAACTTCCGC
TTGGAATATGAGGGCTCAAGGCTGCTGAT--ACGCGGAAATGCGAAGATGGAG
---------CACCTTGGTCACAGTAAC----------------------
--------
---

> Siglec-6
------GGGCCCTGGCTCAGGAGGGAGA------TTCCAG
CTGGAGGAGCCAGATGCACCTGACGAGGCAGGAGGCTGTGCTGCTCTCTGT
ACCTCAGGAGATGCCCCTCACT------ACCCCTCCA------GCCCTGACTAT
GG----TTATG-GCTACTGCTTCC-TGGAAAGGGGCT-------------GAT
GTTCCAGTGGGCCAAACAGCCACAGACAGAAGTGCAAGAGAGAGACCCG
GGGCCATCCTCACCCT--CCTTGGGATCCAGAAGGGAACTCTGCTCTCTG
AGCATAGAGATGGCCAGAAGAG---GACAATGCTGATACCTTCTCTCG
TTGAACTCAAAA------TGGATGAATAACGGTTATATACATC---------
-------------TTCCAAGCTCTCTGTGCTGTGATGG---
---

> Siglec-7
------GGGAGAGGGTGGAAGGACAGAAGAGTAACCGGAAGGATTACTCG
CTGCAAGTGCAGAGTCTCCGACCTGACGAGGGATGGGCTGTCACATGT
GGCTCTGTCTCTCTCTACTACAGCCAGTGGCCTGACGAGGAGAGACCCG
GGTTCACTACTCT------CCTCTGGGACCGAGGAGACCTGCTCTCTG
AGCATAGAGATGGCCAGAAGAG---GATAAGGGGTAGATACCTTCTCTC
CTGAGAGAGAAGGCTCAGAGATGGAATTATAATA-----
-------------TGACCAGCTCTCTGTGACAG---------
---

> Siglec-8
-------CAAAAGGGATGGAGGGAGACAGACAATATGGGGATGGTTACTTG
CTGCAAGTGCAGAGTCTCCGACCTGACGAGGGATGGGCTGTCACATGT
GGCTCTGTCTCTCTCTACTACAGCCAGTGGCCTGACGAGGAGAGACCCG
GGTTCACTACTCT------CCTCTGGGACCGAGGAGACCTGCTCTCTG
AGCATAGAGATGGCCAGAAGAG---GATAAGGGGTAGATACCTTCTCTC
CTGAGAGAGAAGGCTCAGAGATGGAATTATAATA-----
-------------TGACCAGCTCTCTGTGACAG---------
---

> Siglec-9
-------GGGAGAGGGCGGAAGGACAGACAATATGGGGATGGTTACTTG
CTGCAAGTGCAGAGTCTCCGACCTGACGAGGGATGGGCTGTCACATGT
GGCTCTGTCTCTCTCTACTACAGCCAGTGGCCTGACGAGGAGAGACCCG
GGTTCACTACTCT------CCTCTGGGACCGAGGAGACCTGCTCTCTG
AGCATAGAGATGGCCAGAAGAG---GATAAGGGGTAGATACCTTCTCTC
CTGAGAGAGAAGGCTCAGAGATGGAATTATAATA-----
-------------TGACCAGCTCTCTGTGACAG---------
---

> Siglec-10
-------GGGCCAGCTATGGATGGGAGA------TTCCAG
CTGGAGGAGCCAGATGCACCTGACGAGGCAGGAGGCTGTGCTGCTCTCTGT
ACCTCAGGAGATGCCCCTCACT------ACCCCTCCA------GCCCTGACTAT
GG----TTATG-GCTACTGCTTCC-TGGAAAGGGGCT-------------GAT
GTTCCAGTGGGCCAAACAGCCACAGACAGAAGTGCAAGAGAGAGACCCG
GGGCCATCCTCACCCT--CCTTGGGATCCAGAAGGGAACTCTGCTCTCTG
AGCATAGAGATGGCCAGAAGAG---GACAATGCTGATACCTTCTCTCG
TTGAACTCAAAA------TGGATGAATAACGGTTATATACATC---------
-------------TTCCAAGCTCTCTGTGCTGTGATGG---
---
>Siglec-11
------GGTCCCTGAACAGGTACAGT-----------------TACAGT
CTTCAGTGTCAGAGGCCAGTGTTGCGGAGGGCTTGTCGTCTGC
GTCTGTGGCGCTTACCTACCCGAAGGATGGCAGTCTGCTCTG
GC---TTATG-GCTACTGGTTTCA-AAGGACGAGGGATGCGCCAGAGCTGCTCTATG
GCTCTGTCGGAGCAATACCCGAGATGGAGGTGGAAATGCACCCG
GGACCGAGATTCCACT-CACTGGGATACCGGCAAGGAGCTGCTCTCTG
GTGATCAGAGAGCAGCCGAGGAG--GATGGAGCATTACTTCTTCCG
GTGGAGAGAGGAAAGCGGCTGTGAGACATAGTTTCCTGAG----------
--------CAATGCGTTCTTTTTACAAAGTACACG---------

>Siglec-L1(domain 1)
------GAGTGGGGGCTAAGGAACAGAAGGAT------------TACCTG
CTGCAATGCAGAGTGCAGTTCTGACGGTGCAGGAGGGCCTGTGTGTCTCTGT
GCTTTGGCTCTTCTCTTCTACCCCCAAATGCTG-GACTGCTCCGATCTAC
GT---TCAATG-GCTACTGGTTC-GGAGGAGGACATGTAAGCGGAAC
ATTCAGTGTCAGAGCTCCACAAACCACCAAGATGGCAAGGAAAGAGGGCCAC
GGACCGAGATTCCACT-CCTTGCGGGACCCAGAAACAGGATTTGTACCTCG
AGCTAGAGACAGACGACAGAGAG--GATGCAGGGATACACTTCAGTGGAGG
---------AATATGAATGGAATTATAATA-------------------------
----------TGACCAGCTCTCTGTGGAATGTGACAG--------

>Siglec-L1(domain 2)
------CGTCCCAGGACCTACTGTCAAGA---------------TACAGG
CTGGAGGTGCCAGAAGCTGGTGACGGATGAGGGTCGGATGCTGCTCTG
GCCCTGCAGTGTCTCTCTACCCCCAAATGCTG-GACTGCTCCGATCTAC
GT---TCAATG-GCTACTGGTTAC-AGGATGAAGCCACTGCTCTCT
GCTAGAGAGGAA---AGGAGGAAATGGAATTATAATA-------------------------
----------TGACCAGCTCTCTGTGGAATGTGACAG--------

>Siglec-P1
------GGGCTTTGGCTTAGGATCCAGA-----------------GATGGC
CTTAAAGTGTCGAGAGCTGGTGACGGATGAGGGATGCGAGCATGT
GCTCTGCCCTGTCTCTTCTACCCCAAGTATTACCT-GGCGAGATCCGCGCTC
GC---TCAATG-GTTACTTACCGAAC---AGGATGAAGCCACTGCTCTCT
GCTAGAGAGGAA---AGGAGGAAATGGAATTATAATA-------------------------
----------TGACCAGCTCTCTGTGGAATGTGACAG--------

>Siglec-P2
------CGTCCCAGGCCCCACTGTCAAGA-----------------TACAGG
CTGAGTGTCGAGAGCTGGTGACGGATGAGGGATGCGAGCATGT
GCTCCATCTCTCTTTACTACCCCAAGTATTACCT-GGCGAGATCCGCGCTC
GC---TCAATG-GTTACTTACCGAAC---AGGATGAAGCCACTGCTCTCT
GCTAGAGAGGAA---AGGAGGAAATGGAATTATAATA-------------------------
----------TGACCAGCTCTCTGTGGAATGTGACAG--------

>Siglec-P3
ATGGAAGTCAAGAAG---ATACCAGTGGGAAGGGGCTGTGCACTACAAT
CCCTGTTGCTATTTTGAGG---ATTTTCCA-GAGAACCCCCAAGCAA
-----TTCCATG--ATCATGACTCTGCTCAACAAAAACATCAGCT-----CC
CT-----GATGACTACAAATACAAACCAAATGGTCTCCACTTGGGATACACCAA
GGACAATTTTACT-GACTGGAATTTGATGGAAGAAGACTGTACCCCTA
CTACCCAGTAGATATCCAAAGG-----AACAGCATAACATATTATTTTACT
GCAGATCTAGA---GAACAAAAAGTGCCTTCTGGGGAGAA------
----------TATCACAATTTCCTGGTCAG---------------------

> Siglec-P9
------GGGATCTGCTCAGGATATGAGA---------------TCCCGG
CAGGAAAGCCCCAGATATGCTGGAAAGGAGGTTTCTGGTACACTGC
AC--TGTCCTATTTCCACCCCTTGGTACCT--GTCTGACTCCAGCCTC
GT-----CCAAG-GCTACTGTCC--CAGAAAGGGGACATTTCCATCAGGATC
CTCACGGTACCCACCT--GCTT--CAGACCCCAGGCCAGACCTGCTTTCTG
GAGCAGAGACAGACACAGGAGACAGCTTCTTTAGG
GAGAAAAGGGAGATCTATGTGAA-TGTAATTACAGAAA-------------------
----------TG-----CTCTATGTGCATGTAAGG---------------------

> Siglec-P10
------GAGCCCTGGCTCAGGATGAGA----------TTCTAG
CTGGAAGTCAGGAGTCTGTGACTGTCAAGAGTGTCGTGACCCCTCTAT
GCCTTGCTGCTTCTCTATCA--CAACATCTA------CAAGAATAACTCT
GT-----TCAGT-ACTACTGTTGGCTCCGGAGAGGGAGGAAATTCTGGGAC
ACTTCCAGCCCCACACACAACACCAATATGAAAAATGAGAGGAGGAGACCTC
GGCCCTTCTGGCTC-TTCCAGGAAATCCCCAGAGAAACACTGCTCCCATG
AGATCATAGATGGCAGAGGAGGTGAACAACTGCTATTTCTG
GTGAGAGAGAGAGTACCAAATACAGGTATATAC-----------------
--------TCCCCAGTCTCTTCTGGTGTGACTG---------------------

> Siglec-P11
------GAGGACCAGAAGGCAAATGCTGGCAAGGGGC--TGGGGA
CTGGGGAGGAAGGAAAGGGAGTCTCCTACCCCTTGGTGTG--CCATGT
GCCCTGCTCCTCTGTGACTACCCCAAGCAGACGGT--GACTGACTCTGACCCA
CT-----TCACA-GCTATTGGTAG-GGGAAATGGGCCCTTCCAGCCTCAGAT
GTCCTAGGTCAGCGAGAAACACACAGTAGAAAAATGAGAGGAGGAGACCCA
GGGCACTGTCATACT-TTCCAGGGACCAGATCAATGACTGCTCCCTG
AG-----------------------GGCTCATACTTCTTTCAC
ATGGAGGGAGGA-AATATGAAATGGAATACAALAT-------------------
----------TAACGAGTGTCCATGCTGTGATGC---------------------

> Siglec-P13
------AGGACC---------------------ACGCAGGGG---TATGAG
CTGCAGATTCACAGCTGTTGTGATGTCAGGAGGGCTGCTATGTGCCTGGT
GCCCTGCTCCTCTGTGACTACCCCAAGCAGACAGT--GACTCAGCAGCCCA
CC-----CTACG-TCTCTCTGTGCC-CCGAAGGGGACAATGAACACTGGGAT
CCTCCATGCGGCCACACACAACCAACAGCTAGAAGCTGAAAGAGAAGACT
GGGATATTCCAATTTCTTGGGACCCCAAGACTGACATCTGGCTCCCTG
AGCATGAGACCAACAGGAGAGAA--GATGGGGAGTGTATTTATTTCTGG
TTGAGAGGGCTCTGCCTGGAATACTCAGCTTTCAAGCA-------------------
----------AAAGACACTGTCTTCTGGGACGACTGATAG----------------------

> Siglec-P14
------GTCTCAGACCTCTGTGGATGGGAAGA--------TTCTGG
CTGCAAGTGGCAAGAGTCTGATAACACAGGAGGACTGTGTGTCTTCTTG
GATGTGCTCTTCTCCTACCTACTGAGGAGCTG-GA-CGAGTCTACCTCA
GC---CTTG-GCTACTGGTTCA-AAGAAGGGACCAACATAACATGAGT
GCTCTAGTGGAACAAACAGCTCAAACAAAGTAGTGCGAGATAAGCACCCTA
GGGCCGATTCCAGCT-CATTGGGGATCCCCACTACCAGAAACTGCTCCTTG
GTGATCGAGAGATGTGCAAGATGGAG--GATACGGCAGTGTACTTCTTTCCG
GTAAGAGAGGGAGGCTTTGTGAGATAACATTTATGAGA----------
------------ATACATTCTTTCTGGAACACTGACG------------
> Siglec-X
--------------------------------------GCTGGCCAGCGCAGCGCAGGGTCC
ATGCAGGTGCCACCCGGAGGTAGGCGCCGGAGGCGAGGCGGCCAGTGCT
GCCCGCACCCTTCACGGCAACCCGCACCGGCACCTACGACGGCAGCGGACGG
CCATCTGGCGCGCGGAGGCCGCCCTATGCGGGCCGCCAGGTGGTGTCGCTGC
GCTGGCCGCCGGCCAGCGAGCGTCTCGCCAGACCGGCCCAACGCGCTACCTCGCTG
CGGCGCTTGCCCGCT-GCTGGGCAACCAGCGGCCCGACCGGACCTCTCGCTG
CGCGTGCA---CGCTTCCCTGGCTGACGACCCGCGCTACTTCTGCCCAGT
GTCGAAGTTCGCC---GGCAGCTCCATGA---CGCTACAGAGGCCG----
----------CCACGGCGTCCCGGCTGACGACGACG-------
----------