THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 29, Issue of July 21, pp. 22127–22135, 2000

Printed in U.S.A.

Cloning, Characterization, and Phylogenetic Analysis of Siglec-9, a New Member of the CD33-related Group of Siglecs

EVIDENCE FOR CO-EVOLUTION WITH SIALIC ACID SYNTHESIS PATHWAYS*

Received for publication, April 2, 2000, and in revised form, May 3, 2000
Published, JBC Papers in Press, May 5, 2000, DOI 10.1074/jbc.M002775200

Takashi Angata‡ and Ajit Varki§

From the Glycobiology Research and Training Center, Department of Medicine and Cancer Center, University of California, San Diego, La Jolla, California 92093

The Siglecs are a subfamily of I-type lectins (immunoglobulin superfamily proteins that bind sugars) that specifically recognize sialic acids. We report the cloning and characterization of human Siglec-9. The cDNA encodes a type 1 transmembrane protein with three extracellular immunoglobulin-like domains and a cytosolic tail containing two tyrosines, one within a typical immunoreceptor tyrosine-based inhibitory motif (ITIM). The N-terminal V-set Ig domain has most amino acid residues typical of Siglecs. Siglec-9 is expressed on granulocytes and monocytes. Expression of the full-length cDNA in COS cells induces sialic-acid dependent erythrocyte binding. A recombinant soluble form of the extracellular domain binds to α2–3 and α2–6-linked sialic acids. Typical of Siglecs, the carboxyl group and side chain of sialic acid are essential for recognition, and mutation of a critical arginine residue in domain 1 abrogates binding. The underlying glycan structure also affects binding, with Galβ1–4Glc[NAc] being preferred. Siglec-9 shows closest homology to Siglec-7 and both belong to a Siglec-3/CD33-related subset of Siglecs (with Siglecs-5, -6, and -8). The Siglec-9 gene is on chromosome 19q13.3–13.4, in a cluster with all Siglec-3/CD33-related Siglec genes, suggesting their origin by gene duplications. A homology search of the *Drosophila melanogaster* and *Caenorhabditis elegans* genomes suggests that Siglec expression may be limited to animals of deuterostome lineage, coincident with the appearance of the genes of the sialic acid biosynthetic pathway.

Sialic acids are a family of α-keto acids with 9-carbon backbones that are expressed abundantly in animals of the deuterostome lineage (1–3). They are found mostly at distal positions of oligosaccharide chains of glycoproteins and glycolipids and are thus exposed to the extracellular environment, allowing them to be recognized during the initial contact of cells with various pathogenic agents such as viruses, bacteria, protozoa, and toxins (4, 5). The marked structural complexity of sialic acids can thus be interpreted as a result of evolutionary arms race between the hosts and the pathogens (6). Recent studies have revealed another prominent role of sialic acids, namely their functions in generating ligands for endogenous lectins (4). Siglecs (sialic acid-binding Ig superfamily lectins) are a family of such sialic acid-recognizing lectins that have been recently defined (7). These proteins are all single-pass type 1 transmembrane polypeptides, with an N-terminal Ig V-set domain, followed by variable numbers of Ig C2-set domains, a transmembrane domain, and a cytoplasmic tail. The first V-set Ig-like domain is the most important in carbohydrate recognition, and the second Ig-like domain may also contribute to the binding (8–11). Eight members of the family have been described so far in humans, and each shows highly cell type-specific expression: Siglec-1/sialoadhesin (expressed on macrophages) (12); Siglec-2/CD22 (on B lymphocytes) (13); Siglec-3/CD33 (on myeloid precursors and monocytes) (14); Siglec-4a/myelin-associated glycoprotein (on oligodendroglia and Schwann cells) (15); Siglec-5 (on neutrophils and monocytes) (16); Siglec-6/OBBP-1 (on B lymphocytes and placental trophoblasts) (17); Siglec-7/AIRM1 (on natural killer cells and monocytes) (18–20); and Siglec-8 (on eosinophils) (21).

Many of the Siglecs have potential tyrosine phosphorylation sites in the context of an immunoreceptor tyrosine-based inhibitory motif in their cytoplasmic tails, suggesting their involvement in intracellular signaling pathways. In fact, Siglecs-3 and -7 have been shown to be capable of transmitting negative regulatory signals upon cross-linking by specific antibodies (18, 22, 23). In case of CD22/Siglec-2, a negative regulatory role was further proven by the studies using genetically engineered mice (24–27). On the other hand, the precise functional importance of sialic acid-binding property of Siglecs is not well understood, although the phenotypic similarity between Siglec-2 null mice and ST6Gal-I (Galβ1–4GlcNAc α2–6 sialyltransferase) null mice (28) suggests that sialic acid binding does affect the signaling activity of this Siglec.

Expression of sialic acids is well documented in animals of the deuterostome lineage (primarily in echinoderms and vertebrates), but their expression in another major group of animals, the protostomes (including nematodes, arthropods, and mollusks), is inconspicuous (1, 2). From evolutionary point of view, it is also an open question whether there are any Siglec homologs in the protostome lineage. However, given occasional reports of sialic acids in insects (29–31), it is possible that the expression of sialic acids and sialic acid binding lectins is simply under more strict spatio-temporal regulation that has diminished possibilities for cDNA cloning of the relevant genes.

Here we report the molecular cloning and characterization of a new member of Siglec family in humans, Siglec-9, and show that it is closely related to a Siglec-3/CD33-related subgroup that have arisen by gene duplications. We describe the expression pattern and glycan binding specificity of the molecule. The

* This work was supported by United States Public Health Service Grants R01-GM323373 and P01-HL57345. 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 nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF227924.

‡ Partially supported by the Naito Foundation (Tokyo, Japan).

§ To whom correspondence should be addressed: Glycobiology Research and Training Center, La Jolla, CA 92093-0687. Tel.: 858-534-3296; Fax: 858-534-5611; E-mail: avarki@ucsd.edu.

possible co-evolution of sialic acids and Siglecs is also explored and discussed, taking advantage of the recent near-completion of the genomic DNA sequencing of a fruit fly (*Drosophila melanogaster*) (32) and a nematode (*Caenorhabditis elegans*) (33).

EXPERIMENTAL PROCEDURES

Materials—Biotinylated polyacrylamide (PAA-Bio)[1] polyvalently substituted with sialylated oligosaccharides were purchased from Glycotech. Culture media were obtained from Life Technologies, Inc. Unless otherwise specified, all other reagents were purchased from Sigma.

Cloning and Mutation of a Full-length cDNA Encoding Siglec-9—Total RNA was prepared from human peripheral blood mononuclear cells using RNeasy Midi Kit (Qiagen). First strand cDNA was synthesized by reverse transcription (Superscript II, Life Technologies, Inc.) of 1 μg of peripheral blood mononuclear cell total RNA using random DNA hexamers as primers. The reaction product was subjected to PCR using Expand Long Template PCR System (Roche Molecular Biochemicals) using primers SLY-5′-untranslated region (5′-CTCGGATCCCTGGCACCTCTAACCC-3′) and SLY-3′-untranslated region (5′-CCTCTAGAATCAGCCTCTGACTTCTCCC-3′). The product was further amplified by nested PCR using SLY-5′-Chi (5′-CCTGTCGACGCCACCATGCTGCTGCTGCTGCTGCC-3′) and SLY-3′-untranslated region as primers. The PCR product was digested with *Sal*I and *Xba*I, ligated to *Sal*I-*Xba*I sites of pBluescript II KS (−) (Stratagene), and sequenced. A point mutation converting an arginine residue (Arg^{120}) to lysine was introduced using QuickChange site-directed mutagenesis kit (Stratagene), following the manufacturer's protocol. Individual clones were isolated, and plasmids were purified and sequenced.

Northern Blot Analysis—A 267-base pair 3′ end cDNA fragment of Siglec-9 was amplified by PCR and labeled with the Strip-EZ DNA kit (Ambion) and [α-^{32}P]dATP (NEN Life Science Products). A human 12-lane Multiple Tissue Northern blot (CLONTECH) was probed with the labeled probe. Hybridization signals were visualized using a Storm 860 PhosphorImager (Molecular Dynamics).

Expression of Full-length Siglec-9 on COS-7 Cells and Erythrocyte Rosetting—The *Xho*I-*Xba*I fragment of Siglec-9/pBluescript II KS (−) containing full-length coding sequence of Siglec-9 (wild type and R120K mutant) was subcloned into the *Xho*I-*Xba*I sites of pcDNA3.1 (+) (Invitrogen), and sequenced. Constructs were transfected using LipofectAMINE reagent (Life Technologies, Inc.) into COS-7 cells. Erythrocyte rosetting was performed as described previously (20), 48 h after transfection, with or without *Arthrobacter ureafaciens* sialidase (Calbiochem) pretreatment of COS-7 cells or erythrocytes.

Production of Recombinant Chimeric Siglec-9-Fc—A DNA fragment of Siglec-9 encoding the three Ig-like domains was amplified by PCR using SLY-5′-Chi and SLY-3′-Chi-2 (5′-ATCTGATGTGGCTTTGCTCTGCAGGG-3′) as primers and Siglec-9/pBluescript II KS (-) (wild type and R120K) as template. The fragment was cloned into the expression vector EK-Fc-pEDdC (prepared in this laboratory by Hui-Ling Han), giving rise to a fusion protein of Siglec-9 extracellular domains and a human IgG Fc tail with a FLAG epitope tag/enterokinase cleavage site (DYKDDDDK) in between. The constructs were transfected using LipofectAMINE into CHO-TAg cells, culture supernatants collected, and the chimeric proteins were purified on protein A-Sepharose (Amersham Pharmacia Biotech).

Binding Specificity of Siglec-9—Binding to sialylated oligosaccharides on PAA-Bio arrays was performed as described (17, 20). Briefly, microtiter plate (Nunc, catalog number 269620) wells were coated with protein A (0.5 μg/well) in 50 mM sodium carbonate-bicarbonate buffer, pH 9.5, at 4 °C overnight. Wells were washed three times with ELISA buffer (20 mM HEPES, 1% bovine serum albumin, 125 mM NaCl, 1 mM EDTA, pH 7.45), blocked with ELISA buffer (room temperature for 1 h), and sequentially incubated at room temperature with the following (each incubation followed by three washes with ELISA buffer): Siglec-9-Fc (0.5 μg/well, human γ globulin as negative control), 2 h; probes (1 μg/well), 2 h; streptavidin-conjugated alkaline phosphatase (1:1000 diluted from stock solution, Life Technologies, Inc.), 1 h. After final wash, *p*-Nitrophenyl Phosphate Liquid Substrate (Sigma) was added and incubated at room temperature for 30 min, and absorbance was measured at 405 nm.

Modification of Sialic Acids in Probes—The PAA-Bio probes were subjected to mild periodate treatment, as described previously (20). Typically, 10 μg of the probe was incubated in 100 μl of 2 mM $NaIO_4$/phosphate-buffered saline for 30 min on ice in the dark. After the incubation, 100 μl of 20 mM $NaBH_4$ in phosphate-buffered saline was added to the mixture and further incubated at room temperature for 1 h in the dark. The mixture was diluted with 800 μl of ELISA buffer and directly used in the assay. The PAA-Bio probes were also treated with iodoethane, as described previously (34). The PAA-Bio probes carrying Neu5Acα2–3Galβ1–4Glc, Neu5Acα2–6Galβ1–4Glc, and Neu5Acα2–3Galβ1–4[Fucα1–3]GlcNAc (10 μg) were lyophilized in Reacti-Vials (Pierce) and dissolved in 35 μl of dry Me_2SO. To each vial 7 μl of iodoethane (Aldrich) was added and incubated at room temperature for 30 min. Remaining iodoethane was removed by evaporation under reduced pressure. To the residue was added 460 μl of 20 mM HEPES buffer (pH 7.45) containing 125 mM NaCl, and then 2.5 μl of 1 M $NaBH_4$ was added and incubated at room temperature for 1 h to reduce ethylester at the C-1 position of sialic acids to an alcohol. At the end of the reaction 500 μl of 20 mM HEPES buffer containing 125 mM NaCl, 2 mM EDTA, and 2% bovine serum albumin was added, and the mixture was directly used in the binding assay as described above.

Generation of a Monospecific Chicken Polyclonal Antibody (IgY) against the Extracellular Domain of Siglec-9—Siglec-9-Fc chimeric protein (1 mg) was treated with enterokinase (a generous gift from Dr. J. Evan Sadler) and incubated with protein A-Sepharose overnight at 4 °C, to remove the IgG-Fc portion and any intact protein. The supernatant containing only the Siglec-9 extracellular domain (450 μg) was concentrated and used as an antigen to raise polyclonal antibodies in chickens. Two Rhode Island Red hens were each immunized with 75 μg of the antigen in Freund's complete adjuvant, followed by injections of 50 μg of antigen in Freund's incomplete adjuvant 22 and 44 days later. Eggs laid after the second booster injection were collected, egg yolks were separated, and immunoglobulin fraction was purified using EGGstract IgY Purification System (Promega). IgY is the major serum immunoglobulin in chicken that also accumulates in egg yolk. To subtract any IgY subfraction that might cross-react with Siglec-7, the purified IgY (200 μg) was incubated with Siglec-7-Fc (100 μg) immobilized on protein A-Sepharose, and supernatant was used as monospecific polyclonal antibody against Siglec-9.

Flow Cytometry Analysis of Peripheral Blood Leukocytes—Human peripheral blood was collected from healthy adult donors, and granulocytes and lymphocytes/monocytes were separated using Mono-Poly Resolving Medium (ICN). Each cell population (1×10^6 cells/sample) was incubated with polyclonal IgY against Siglec-9 (prepared as described above), followed by fluorescein isothiocyanate-labeled $F(ab')_2$ fragment of donkey anti-chicken IgY (Jackson ImmunoResearch). Two-color staining was performed to distinguish lymphocyte populations, using CyChrome-conjugated anti-CD4, phycoerythrin-conjugated anti-CD8, CyChrome-conjugated anti-CD56 (from Pharmingen), and TRICOLOR-conjugated anti-CD19 (Caltag) as secondary staining, following standard protocols (35). Cells were analyzed using FACScan (Becton Dickinson), and data were processed using CellQuest software (Becton Dickinson).

Chromosomal Localization of Siglec-9—From the physical map of human chromosome 19 (Human Genome Center, Laurence Livermore National Laboratory), BAC clones contiguously covering the region of chromosome 19 containing the Siglec-3/CD33 gene were identified and obtained through Research Genetics. The BAC DNA clones were prepared by a modified alkaline lysis method recommended by BACPAC Resources, Rosewell Park Cancer Institute, and subjected to PCR using SP1 (5′-GCCTTCTCCTTGGAAGACAG-3′) and SP-5′ (5′-AAACTCGGGACCGATTCCAC-3′) as primers. PCR products were purified and directly sequenced.

Phylogenetic Analysis of Siglecs—DNA sequences of human Siglecs 1–9 encoding the first two Ig-like domains (750 nucleotides; the human Siglec-1 sequence was kindly provided by Dr. P. Crocker, Dundee, UK) were aligned using ClustalW at the European Bioinformatics Institute web site and then subjected to phylogenetic analysis using PAUP 4.0 (Sinauer Associates). The phylogenetic tree was constructed using the neighbor joining method (36). The distance matrix was based on Tamura-Nei genetic distance (37).

Homology Search for Siglecs in Fruit Fly and Nematode Genomic Sequences—The sequences of the first 150 amino acids of Siglecs 1–9 (encoding the first Ig V-set domain in each case) were used as templates in homology search of the nr, htgs, and *Drosophila* genome divisions of GenBank™ data base at the National Center for Biotechnology Information web site using tblastn program (38). These divisions include all cDNA and genomic DNA sequences made public. Genomic DNA sequences that showed significant homology (expectation value, <1) to

[1] The abbreviations used are: PAA-Bio, biotinylated polyacrylamide; SLAM, signaling lymphocyte activation molecule; SAP, SLAM-associated protein; NCAM, neural cell adhesion molecule; BAC, bacterial artificial chromosome; EST, expressed sequence tag; PCR, polymerase chain reaction; ELISA, enzyme-linked immunosorbent assay.

Siglecs were first analyzed for putative protein-coding sequences. Putative protein-coding sequences were either found in annotated GenBank™ entries or using Berkeley Fly Data Base of Berkeley *Drosophila* Genome Project (for fruit fly genomic DNA sequences). These putative protein-coding sequences, as well as the cDNA sequences that showed significant homology (expectation value, <1) to Siglecs, were translated to amino acid sequences and analyzed for domain structures using the SMART program at European Molecular Biology Laboratories at the Heidelberg web site (39, 40). If a candidate sequence showed similar overall domain structure to Siglecs (Ig domains followed by a transmembrane domain), its human homologs were searched using the tblastn program on the GenBank™ data base at the National Center for Biotechnology Information. As a control, the sequence of the first 150 amino acids of human neural cell adhesion molecule (NCAM) was used as template in similar procedure to find homologs in the nematode and fruit fly.

Homology Search of Enzymes Involved in Sugar Nucleotide Synthesis—The amino acid sequences of the following human enzymes were used to identify putative orthologs in mouse, fruit fly, and nematode: UDP-Gal 4-epimerase (GenBank™ accession number NM_000403), GDP-Man pyrophosphorylase (NM_013334), UDP-HexNAc pyrophosphorylase (NM_003115), GDP-Man 4,6-dehydratase (NM_001500), GDP-4-keto-6-deoxy-Man epimerase-reductase (NM_003313), UDP-GlcNAc 2-epimerase/ManNAc kinase (NM_005476), GlcNAc 2-epimerase (NM_002910), and CMP-Neu5Ac synthetase. Because human CMP-Neu5Ac synthetase cDNA is yet to be reported, its sequence was deduced from genomic DNA (AC007671) and EST (AW402305) sequences, using mouse CMP-Neu5Ac synthetase cDNA (AJ006215) as a template. Orthologs of these enzymes were searched for in the GenBank™ data base (nr, htgs, and EST data base, and *Drosophila* genome divisions) using tblastn program, as described above. If a sequence with high homology was found only in genomic DNA, the putative protein-coding sequence was sought in the annotated GenBank™ entry or using the Berkeley Fly Data Base of the Berkeley *Drosophila* Genome Project. In the case of the mouse, sequences with high homology were often found only among EST sequences. In such cases, we deduced the longest contiguous coding sequences possible from both 5′ and 3′ ends and left the gap in between as is. Such cases include: UDP-Gal 4-epimerase (N-terminal coding sequence from AA880071; C-terminal coding sequence from AI327487), GDP-Man pyrophosphorylase (entire coding sequence deduced from AI322967, AA016515, AI118793, and AA572327), UDP-HexNAc pyrophosphorylase (N-terminal from AA474378 and AA798356; C-terminal from AA666936, AA437860, and AA896061), and GDP-Man 4,6-dehydratase (N-terminal from AI852418; C-terminal from AI747296, AI322318, AA403686, and AA538309). Putative orthologs were aligned using ClustalW and analyzed by PAUP 4.0 as described.

RESULTS

Cloning of a Full-length cDNA Encoding the Putative Siglec-9—During a 5′-rapid amplification of cDNA ends experiment that eventually resulted in the cloning of Siglec-7 (20), we noted two independent but highly homologous DNA fragments, one of which turned out to be Siglec-7 and the other an as yet unknown gene product (tentatively named Siglec-Y). When the sequence of the Siglec-Y was compared with data base sequences at GenBank™, we found one entry named "OB-binding protein-like protein gene" (GenBank™ accession number AF135027) that showed 100% nucleotide identity in two segments. This submission was based only on computer-based exon-intron predictions using raw genomic sequences from the human genome project and was accompanied by an incorrect prediction for the protein coding sequence (at the time of comparison, May 1999; the entry was later modified several times).[2] During an independent attempt to find new Siglec candidates from the dbEST data base using 3′ coding sequences (encoding the transmembrane domain and the cytoplasmic tail) of Siglecs, we found a Siglec-like EST sequence (AA936059) that also showed 100% nucleotide identity with the above sequence. We therefore designed primers to clone the Siglec-Y full-length cDNA by reverse transcription-PCR, using RNA prepared from peripheral blood mononuclear cells as template.

The cDNA we finally cloned contained an open reading frame (1392 nucleotides) that encodes a signal peptide, three Ig-like domains (one V-set domain followed by two C2-set domains), a putative transmembrane domain and cytosolic tail (Fig. 1*A*). The first Ig-like domain contains many features conserved among all Siglecs. The most prominent examples are a typically placed arginine (Arg^{120}) and an aromatic amino acid (Trp^{128}), which are known to be involved in sialic acid recognition of Siglec-1/sialoadhesin from x-ray crystallographic analysis of a Siglec-ligand co-crystal (41) and in Siglecs-1, -2, -3, -4, and -7 by mutagenesis experiments (10, 11, 20, 22, 42). One exception is the lack of an aromatic amino acid in the proximity (typically the second amino acid residue) of the putative N terminus of the mature protein. Notably, the first, second and third Ig-like domains of Siglec-9 contained three, four, and three cysteines, respectively, suggesting that there are either interdomain disulfide bonds, as suggested for other Siglecs, or interpolypeptide disulfide bonds as in immunoglobulins. The putative cytoplasmic tail contains two tyrosine residues, of which the first (Tyr^{433}) is found in an immunoreceptor tyrosine-based inhibitory motif ((S/I/L/V)*XYXX*(L/V)). The second tyrosine residue is in a motif ($TEY^{456}SEI$) similar to the putative binding site (TIY*XX*(V/I)) on signaling lymphocyte activating molecule (SLAM) for SLAM-associated protein (SAP) (43). Nucleotide identity with other human Siglecs (in the first 750 nucleotides) are: Siglec-1/sialoadhesin, 49.9%; Siglec-2/CD22, 49.9%; Siglec-3/CD33, 73.3%; Siglec-4a/myelin-associated glycoprotein, 49.2%; Siglec-5, 71.5%; Siglec-6/OB-BP 1, 73.1%; Siglec-7/AIRM1, 90.5%; and Siglec-8, 81.4%. Alignment of closely related Siglec subset (Siglecs-3, -5, -6, -7, -8, and -9; Fig. 1*B*) reveals a high degree of conservation in the extracellular domains among these molecules. The cytoplasmic tail is less highly conserved, except for the amino acids surrounding the two tyrosines mentioned above, suggesting that these motifs are under functional positive selection. The new molecule was renamed as Siglec-9, based on the high homology with known Siglecs, conservation of defining features, and sialic acid-dependent binding properties shown below.

Expression Pattern of Siglec-9—Northern blot analysis indicates the following pattern for Siglec-9 mRNA expression in human tissues: high expression in the liver and less prominent expression in spleen, placenta, and skeletal muscle (Fig. 2*A*). The major mRNA species is estimated to be about 1.8 kilobases long. Although expression of Siglec-9 mRNA in total peripheral blood leukocyte mRNA appears to be relatively low, most known Siglecs to date are expressed on only a subset of blood leukocytes and sometimes on mature cells that expressed their message only during their development in the bone marrow. We therefore examined Siglec-9 expression in peripheral blood leukocytes using a monospecific chicken polyclonal antibody generated against the extracellular domain of Siglec-9. Flow cytometry analysis revealed its expression on the great majority of granulocytes and monocytes (Fig. 2*B*). Natural killer cells (CD56-positive) appeared to be very weakly positive (data not shown). At present we cannot rule out that this could be due to any cross-reactivity of anti-Siglec-9 antibody to Siglec-7, which is known to be highly expressed on natural killer cells (18, 19). Nevertheless, antibody binding to granulocytes and monocytes must be due to expression of Siglec-9 and not due to any cross-reactivity with Siglec-7, because expression of the latter is highest in natural killer cells, less prominent in monocytes, and minimal in granulocytes (19), a staining pattern clearly

[2] The incorrect prediction of intron splicing and putative cDNA and protein sequences were corrected in a later release of their submission in December, 1999. The submitters later published a paper on the adjacent kallikrein gene cluster on chromosome19q13.3–4 (63), in which they referred to the Siglec-like gene as the "unknown gene."

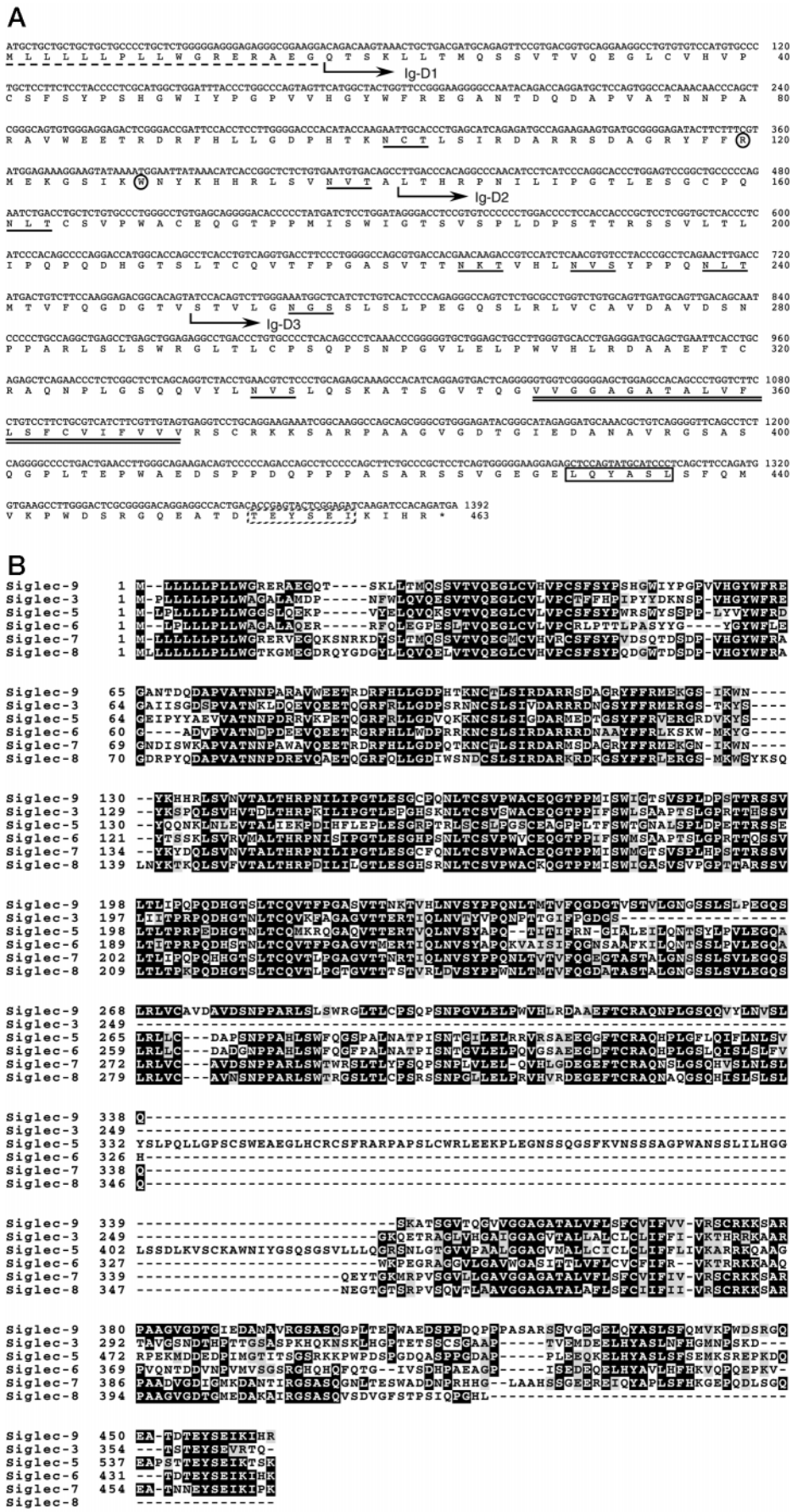

FIG. 1. **Primary sequence of Siglec-9 and comparison with other Siglecs.** *A*, cDNA and derived amino acid sequences of Siglec-9. The *dotted line* and *double line* indicate the putative signal peptide and transmembrane sequence, respectively. Potential *N*-glycosylation sites are *underlined*. The *rectangle* indicates the putative immunoreceptor tyrosine-based inhibitory motif, and the *dotted rectangle* indicates the motif similar to the SAP-docking site. The two amino acid residues essential to sialic acid-binding, Arg^{120} and Trp^{128}, are *circled*. The putative N-terminal ends of the three Ig domains are indicated with *arrows* and labeled for each domain (D1, D2, and D3). *B*, amino acid sequence alignment of Siglec-9 and closely related Siglecs. Amino acid sequences of human Siglec-3, -5, -6, -7, -8, and -9 were aligned using ClustalW multiple sequence alignment program and minimally adjusted manually. GenBank™ accession numbers for Siglec-3, -5, -6, -7, -8, and -9 are NM_001772, NM_003830, NM_001245, NM_014385, NM_014442, and AF227924, respectively.

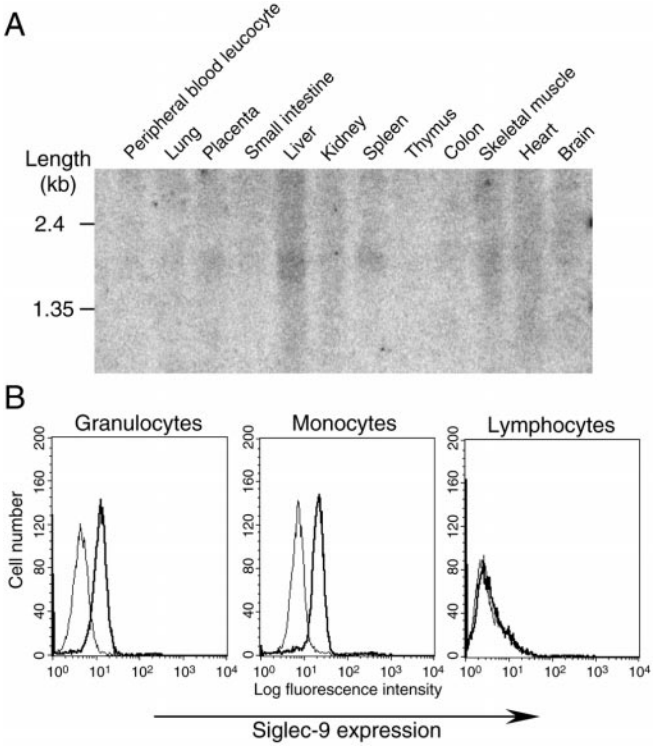

FIG. 2. **Expression of Siglec-9.** *A*, Northern blot analysis. A human 12-lane Multiple Tissue Northern blot (CLONTECH) was probed as described under "Experimental Procedures." Tissue sources of mRNA were indicated above each lane. The region of the blot corresponding to the major ~1.8-kilobase (*kb*) message is shown. *B*, flow cytometry analysis of Siglec-9 expression on peripheral blood leukocytes. Single-color staining signals with anti-Siglec-9 and with negative control antibodies are represented with *thick* and *thin lines*, respectively.

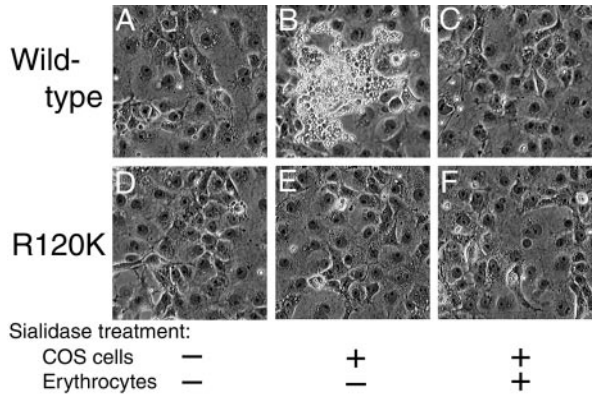

FIG. 3. **Erythrocyte rosetting by Siglec-9-transfected COS cells.** COS cells were transfected with wild type (*A–C*) or R120K mutant (*D–F*) Siglec-9 and treated with (*B*, *C*, *E*, and *F*) or without (*A* and *D*) sialidase, followed by incubation with erythrocytes treated with (*C* and *F*) or without (*A*, *B*, *D*, and *E*) sialidase, as described under "Experimental Procedures."

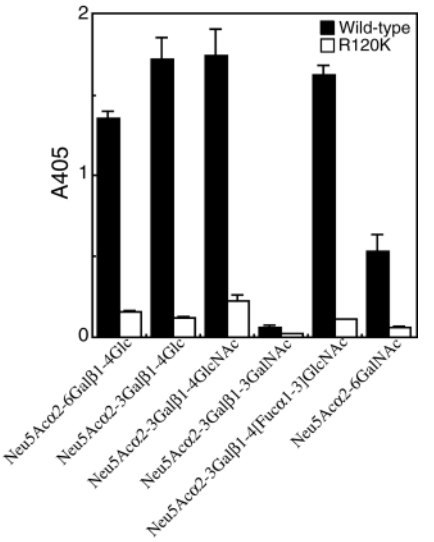

FIG. 4. **Binding of sialylated ligands to the recombinant Siglec-9 extracellular domain.** Siglec-9-Fc chimeric proteins were immobilized via protein A to microtiter plate, and binding of biotinylated polyacrylamide arrays multiply conjugated with sialyloligosaccharides was determined as described under "Experimental Procedures." The data shown are the mean values ± S.D. of triplicates.

different from that shown above.

Erythrocyte Rosetting by Siglec-9 Transfected COS-7 Cells—When COS-7 cells were transfected with the full-length cDNA of wild type Siglec-9, it induced binding of human erythrocytes (Fig. 3). This binding was completely abolished by pretreating erythrocytes with sialidase, indicating that the binding is sialic acid-dependent. On the other hand, the cells transfected with R120K mutant did not bind erythrocytes at all. Notably, sialidase pretreatment of COS cells to eliminate sialylated ligands on the same cell surface (*cis*-ligands) was an absolute prerequisite for the erythrocyte binding.

Sialic Acid Binding Specificity of Siglec-9—A recombinant chimeric protein of the Siglec-9 extracellular domain and human IgG Fc domain was produced and used in the ligand binding assay to further analyze the sialic acid binding specificity of Siglec-9. PAA-Bio probes multiply substituted with sialylated oligosaccharides (20% mol/mol oligosaccharide/acrylamide, 20–25 oligosaccharides/30-kDa polymer) were used in the assay. As shown in Fig. 4, Siglec-9 recognized both α2–3- and α2–6-linked sialic acids (Neu5Acα2–3/6Galβ1–4Glc). Notably, the underlying glycan structures had a profound effect on binding: sialic acids α2–3-linked to Galβ1–4GlcNAc were far better ligands than those α2–3-linked to Galβ1–3GalNAc structure. The weaker binding of Neu5Acα2–6GalNAc, compared with Neu5Acα2–6Galβ1–4Glc, may be also due to the difference in the penultimate sugar structure. The R120K mutant showed very poor binding (<10%) compared with wild type Siglec-9, indicating that this arginine residue is critical in ligand binding, as has been the case with all Siglecs analyzed so far (10, 11, 20, 22, 42). Notably, introduction of fucose on the GlcNAc residue seemed to have no adverse effect on the binding of Neu5Acα2–3Galβ1–4GlcNAc to Siglec-9 (Fig. 4). This is in contrast to previous finding using Siglecs-1, -3, -4, and -5, which all showed marked reduction of binding because of such a fucose residue (34).

Mild periodate treatment of the probes (which specifically truncates the glycerol-like side chains of sialic acids) almost completely abolished binding (Fig. 5*A*), proving that the side chain is recognized by Siglec-9. Also, iodoethane treatment (which modifies the carboxyl group of sialic acid) reduced binding of the probes by ~60% (Fig. 5*B*). The partial loss of binding was concomitant with the extent of sialic acid reduction (~50%, as determined by acid hydrolysis, fluorescence derivatization, and high pressure liquid chromatography analysis; data not shown). This result, along with the loss of binding to the R120K mutant, strongly suggests that the carboxyl group of sialic acid is essential in ligand recognition by Siglec-9.

Chromosomal Localization of Siglec-9—PCR-based screening of five BAC clones contiguously encompassing ~500-kilobase DNA stretch around Siglec-3/CD33 gene (44) revealed that the Siglec-9 gene is localized on band 19q13.3–13.4, within 200 kilobases centromeric to the Siglec-3/CD33 gene (on the adjacent BAC clone). Its closest homolog, Siglec-7, is localized

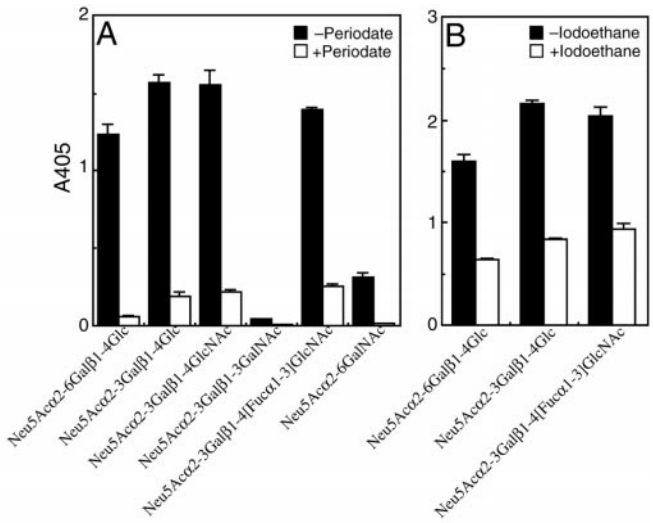

FIG. 5. **Effect of sialic acid modifications on ligand binding to Siglec-9.** PAA-Bio probes were treated with mild periodate to specifically truncate the glycerol-like side chain of sialic acids (*A*) or with iodoethane to modify carboxyl group (*B*), as described under "Experimental Procedures" and used in the same binding assay as in Fig. 4. Data shown are the mean values ± S.D. of triplicates.

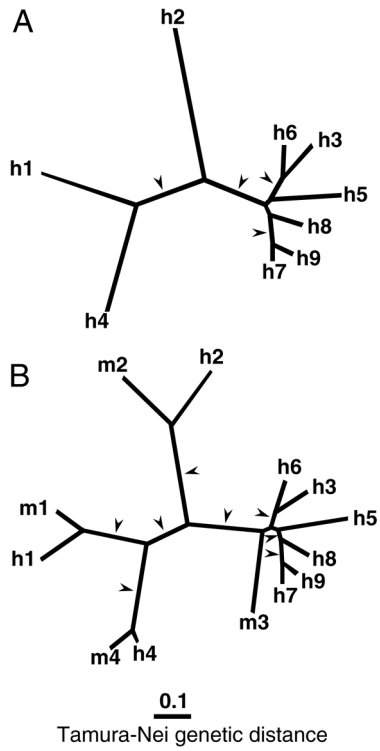

FIG. 6. **Phylogenetic analysis of Siglecs.** The first 750 nucleotides of Siglecs were aligned using ClustalW multiple sequence alignment program and analyzed for phylogenetic relationship using PAUP 4.0. Unrooted phylograms constructed using the neighbor-joining method are shown for human Siglecs (*A*) and for human and mouse Siglecs (*B*). The internodes indicated with *arrowheads* are supported by bootstrap values >90 for 100 resamplings.

on the same BAC clone as Siglec-3 (20).

Phylogenetic Analysis of Siglecs—Phylogenetic analysis of all 9 reported human Siglecs (Fig. 6*A*) revealed a close association of Siglec-9 with a subgroup related to Siglec-3/CD33. As expected from the high sequence identity mentioned above, Siglec-9 is very closely related to Siglec-7. This result, along with published evidence for the presence of Siglecs-3, -5, -6, -7, and -8 genes on human chromosome 19q13.3–13.4 (16, 19–21, 45, 46), suggests that there were multiple gene duplications of Siglec-3-related genes during evolution (the existing data do not allow us to predict which one was the ancestral gene). Interestingly, when the four known mouse Siglec sequences are incorporated into the analysis (Fig. 6*B*), Siglec-1/sialoadhesin, Siglec-2/CD22, and Siglec-4a/myelin-associated glycoprotein all group with their human orthologs, whereas the previously proposed mouse Siglec-3/CD33 does not form exclusive clade with human Siglec-3. Instead, it is loosely associated with Siglec-3-related human Siglecs and stands at about equal distance from human Siglecs-3, -6, and -8.

Homology Search for Siglecs in Fruit Fly and Nematode Genomic Sequences—The first 150 amino acids of human Siglecs correspond to the unique and defining first V-set Ig domain, contain amino acid residues critical for the sialic acid binding, and are most highly conserved among the Siglecs (10, 11, 22, 41, 42). Amino acid sequences of this region were used as probes to identify possible Siglec orthologs in nematode and fruit fly genomic DNA sequences. In the nematode, genomic sequences under GenBank™ accession number U80022 and AF067220 contained segments with some similarity to Siglecs. However, the predicted protein from the former (AAC25886) contained not only Ig-like domains but also a fibronectin type-III-like domain and showed higher homology to human titin, and the latter DNA segment showed no overlap with any putative gene predicted by the submitters. With regard to the fruit fly, one cDNA clone (L13255 = *lachesin*) showed significant similarity to human Siglecs, but the encoded protein is known to be GPI-anchored (47) and showed higher homology to the limbic system-associated membrane protein (NM_002338) in humans. Thus, we conclude that there are no obvious Siglec homologs in the nematode or fruit fly. As a "positive control" we used another human Ig superfamily molecule, NCAM as a probe. In this case, orthologs were easily found by the same approach, namely SAX-3 (nematode) and fasciculin II (fruit fly). The former is not considered as direct ortholog of human NCAM but belongs to a closely related protein family including NCAM, which share the same overall molecular structure (Ig-like domains + fibronectin type III-like domains + transmembrane domain) and expression in neural tissues (48). Fasciculin in the fruit fly is widely considered a true NCAM ortholog (49).

Homology Search of Enzymes Involved in Sugar Nucleotide Synthesis—As shown in Table I, human genes involved in the synthesis of many major sugar nucleotides have putative orthologs in nematode and fruit fly. The only exceptions we found are apparent lack of the enzymes involved in the synthesis of sialic acids and CMP-sialic acids: UDP-GlcNAc 2-epimerase/ManNAc kinase (50, 51), GlcNAc 2-epimerase/renin-binding protein (52), and CMP-sialic acid synthetase (53). This result suggests that these animals lack expression of sialic acids, at least as synthesized by the conventional pathway in vertebrates. However, it should be noted that some segments of fruit fly genomic DNA showed significant similarity to coding sequences of C-terminal half (putative kinase domain; Ref. 54) of UDP-GlcNAc 2-epimerase/ManNAc kinase (AE003811) and N-terminal half of CMP-sialic acid synthetase (AE003515).

DISCUSSION

The Siglec-9 cDNA encodes a protein with many typical features of previously described Siglecs. The first Ig-like domain contains two amino acid residues critical in sialic acid recognition (41): arginine in β-strand F (Arg^{120} in Siglec-9) and aromatic amino acid in β-strand G (Trp^{128}). Crystallographic study of Siglec-1/sialoadhesin showed that these residues are directly interacting with the carboxyl group and glycerol-like side chain of sialic acid, respectively (41). A notable exception in Siglec-9 was the lack of aromatic amino acid near its puta-

TABLE I
Amino acid sequence identity of enzymes involved in sugar nucleotide synthesis

Human Enzymes (GenBank accession number)	Amino acid identity with human enzymes (GenBank™ accession number)		
	Mouse	Fruit fly	Nematode
		%	
UDP-HexNAc synthesis			
UDP-Gal[NAc] 4-epimerase (NP_000394)	93.4 (NA)[a,b]	60.2 (AAF47398)	57.3 (CAB16861)
UDP-HexNAc pyrophosphorylase (NP_003106)	91.4 (NA)[b]	52.8 (AAD38661)	36.8 (CAA91270)
GDP-Man synthesis			
GDP-Man pyrophosphorylase (NP_037466)	97.8 (NA)[b]	70.2 (AAF52089)	63.8 (AAB97547)
GDP-Fuc synthesis			
GDP-Man 4,6-dehydratase (NP_001491)	92.1 (NA)[b]	67.5 (AAF52189)	64.3 (CAA92455)
GDP-4-keto-6-deoxy-Man epimerase-reductase (NP_003304)	93.5 (AAA39673)[c]	60.2 (AAF46924)	64.4 (AAA50647)
Sialic acid/CMP-sialic acid synthesis			
UDP-GlcNAc 2-epimerase/ManNAc kinase (NP_005467)	98.5 (CAB36908)	ND[d]	ND
GlcNAc 2-epimerase (NP_002901)	86.7 (AAF22157)	ND	ND
CMP-Neu5Ac synthetase (NA)[e]	94.0 (CAA06915)	37.4[f]	ND

[a] NA, not available.
[b] Sequences deduced from EST sequences, as described under "Experimental Procedures."
[c] One nucleotide deletion in the GenBank™ cDNA entry (M30127) was corrected using an EST sequence (AA798845). Thus, the amino acid sequence we used is not identical to the GenBank™ entry.
[d] ND, none detected (expectation value, <0.01).
[e] Human sequence deduced from a genomic sequence (AC007671) and an EST sequence (AW007671), using the mouse sequence as template.
[f] The *Drosophila* genomic DNA segments that show some similarity to this human enzyme do not overlap with any gene coding sequence predicted by the submitters. Therefore, comparison was made between three stretches of amino acids (total, ~200 amino acids) retrieved by tblastn search and the corresponding human enzyme segments only.

tive N terminus (typically the second residue), which in Siglec-1 is in contact with sialic acid 5-acetamido group. The fact that Siglec-9 shows robust binding to ligands, whereas mutation of this residue results in the complete loss of sialic acid recognition in Siglec-1/sialoadhesin (41), clearly shows that the importance of this aromatic side chain in ligand binding is variable among Siglecs. On the other hand, almost complete loss of ligand binding in R120K mutant indicates that the stable salt bridge between the arginine and carboxyl group of sialic acid is indispensable in ligand binding of Siglec-9, as is the case with all other Siglecs examined so far (10, 11, 20, 22, 42).

The cytosolic tail of Siglec-9 contains two tyrosine residues. The amino acid sequence around the first tyrosine (LQY^{433}ASL) is found in the context of an immunoreceptor tyrosine-based inhibitory motif ((S/I/L/V)*XYXX*(L/V)), which is the docking site for the phosphotyrosine phosphatases, SHP-1 and SHP-2 (55, 56). The actual functionality of this motif in Siglec-9 remains to be determined, but this motif is likely to be involved in the signal transduction, judging from the high sequence identity of the motif with those in Siglec-3/CD33 and Siglec-7/AIRM1, which have been shown to interact with SHP-1 (18, 22, 23). Interestingly, the sequence around the second tyrosine (TEY^{456}SEI) does not strictly conform with but is similar to the proposed SAP-docking site (TIY*XX*(V/I)) on SLAM/CDw150 (43). It is known that SAP interacts with this motif in SLAM and 2B4 (Ig superfamily molecules expressed on T/B lymphocytes and natural killer cells, respectively) to prevent docking of SHP-2 (57, 58). A mutation in SAP causes human X-linked lymphoproliferative disease (57). Whether this motif in Siglec-9 actually interacts with SAP and the similar protein EAT-2 (43) remains to be determined.

Despite the relatively low abundance of mRNA in peripheral blood leukocytes as analyzed by Northern blots, flow cytometry revealed that Siglec-9 is actually expressed on granulocytes and monocytes. In view of expression of Siglec-9 on monocytes, the mRNA expression found in the liver and spleen may be due to Kupfer cells and splenic tissue macrophages, respectively, although this remains to be proven by a direct histochemical approach. The wide distribution of Siglec-9 among cells that can be elicited in the innate immune response, along with its potential as negative regulator of signal transduction, raises the intriguing question whether Siglec-9 functions as general negative regulator of "rapidly responding" cells. Experiments to analyze this hypothetical role of Siglec-9, as well as that of another similarly distributed Siglec, Siglec-5 (on neutrophils and monocytes), is now underway in our laboratory.

Siglec-9 recognizes both α2–3- and α2–6-linked sialic acids. Such promiscuous recognition of sialic acid linkages is seen primarily among Siglecs expressed on monocytes (Siglecs-3, -5, -7, and -9) and granulocytes (Siglecs-5, -8, and -9) (16, 19, 21, 34). Other Siglecs, *i.e.* Siglecs-1, -2, -4, and -6, show much more strict linkage specificity (12, 13, 15, 17, 34, 59, 60). Another unusual aspect of ligand recognition by Siglec-9 is also worthy of note, *i.e.* the preference for the underlying glycan structure. As far as we know Siglec-9 is the only Siglec that can recognize the Neu5Acα2–3Galβ1–4[Fucα1–3]GlcNAc structure (*i.e.* recognition is not disturbed by the fucose residue on the GlcNAc). Promiscuous as its binding is with regard to sialyl linkage, Siglec-9 still distinguishes between Neu5Acα2–3Galβ1–4GlcNAc (typically found in *N*-linked glycans) and Neu5Acα2–3Galβ1–3GalNAc (typically found in *O*-linked glycans and glycosphingolipids). The relevance of this promiscuity and selectivity with regard to Siglec functions is unclear, but it appears to support our hypothesis stated above: that engagement of the Siglecs by polyvalent ligands may be able to elicit negative intracellular signals, thus silencing the cells that are inappropriately activated. Alternatively, Siglecs may be functioning in a manner similar to that of killer cell Ig-like receptors expressed on natural killer cells, which send negative intracellular signals if engaged by major histocompatibility complex class I molecules expressed on target cells. Viral infection or malignant transformation result in down-regulation of major histocompatibility complex class I molecules on cell surface, thus rendering the cell vulnerable to attack by natural killer cells (61). Likewise, if viral infection or malignant transformation results in drastic change in surface sialic acid expression (as occurs with infection by neuraminidase-producing viruses like influenza virus), this may elicit activation of the Siglec-carrying cells in similar manner. Interestingly, the killer cell Ig-like inhibitory receptor genes are clustered on chromosome 19q13.4 (61), near the Siglec-3-related genes.

Six of the nine known Siglecs (Siglec-3, -5, -6, -7, -8, and -9) are closely related to each other both in primary sequence (see "Results") and chromosomal localization (16, 19–21, 45, 46) and are hence grouped as "Siglec-3/CD33-related." The genes for these are all localized on chromosome 19q13.3–13.4. The

presence of this cluster suggests that these genes emerged by repeated gene duplication at some time during vertebrate evolution. We show here that the previously reported mouse Siglec-3/CD33 does not show a clear relationship to the established human counterpart (Fig. 6*B*). One explanation is that the true mouse ortholog has not yet been isolated. Another possibility is that some of the gene duplications in this cluster took place after separation of the ancestors of primates and rodents. Notably, there are other clustered gene families in close vicinity of the Siglec-3 subfamily, such as α1–2 fucosyltransferases (19q13.3) (62), the kallikrein gene cluster (19q13.3–13.4) (63), and the killer cell Ig-like receptor family (19q13.4) (see above), suggesting that there have been frequent gene duplications in this chromosomal region. In this regard, it is interesting that there are chromosome-specific minisatellites in 19q13.3-qter region (64) that could have facilitated duplication of these genes by unequal crossing over of sister chromatids in meiotic recombination (65). Regardless of the similarity in sequences of the Siglec-3/CD33-related group, it is striking that each is expressed in different pattern of cell type specificity and shows distinct patterns of sialic acid recognition. Thus, the gene duplication events may have been selected for by specific new functions in different cell types. In this regard, final definition of the true mouse orthologs may require not only their cloning but also the antibody-based exploration of their expression patterns in different cell types.

Although some Ig superfamily genes are highly conserved from nematode to humans (*e.g.* NCAM and related molecules), there seems to be no distinct Siglec homologs in *C. elegans* or *D. melanogaster*, two of the most extensively studied protostome lineage animals. This result is consistent with the fact that this lineage is also thought to lack constitutive expression of sialic acids. Although not conclusive, the results of our homology search on the newly available comprehensive genomic data also support the notion that protostome lineage animals generally do not synthesize sialic acids. We found that both *Drosophila* and *Caenorhabditis* genomes apparently lack the enzyme genes required for sialic acid biosynthesis, whereas other genes involved in sugar nucleotide biosynthesis are present and well conserved. Thus, the reported capability of insect cells to express sialic acids under certain circumstances (29, 31, 66) should be addressed in other ways, such as the possibility of alternate pathways of biosynthesis and/or uptake of sialic acids or the biosynthetic precursors from the environment. Regardless, our data suggest that the emergence of Siglecs during evolution was dependent on the constitutive expression of sialic acids in deuterostome lineage animals. In this respect, it is particularly interesting to see whether there are any Siglec homologs in echinoderms such as sea urchins and starfishes, which are known to express large amounts of sialic acids. It should be mentioned that when we used full-length amino acid sequences of Siglecs in the homology search of *Drosophila* genome, we did find some gene products that showed significant homology and similar overall molecular structure, such as *irregular optic chiasma C/roughest*, *neuromusculin*, and *faint sausage*. Although the overall homology was significant between these gene products and human Siglec proteins (especially Siglec-2/CD22), it was limited to the Ig domains 2 and later. Because Siglecs are defined by sialic acid binding, molecules lacking homology in the unique first Ig domain were not considered as Siglec homologs. The same consideration applies to *furrowed*, which is considered as fruit fly homolog of selectins (67) but shows low similarity to mammalian selectins in the C-type lectin domain. These phenomena suggest a split of fates between two genes of common ancestry under different biochemical environment characteristic of deuterostome and protostome lineages (presence or absence of sialic acids). The fact that Siglec-2/CD22 shows highest homology with fly proteins also implies that Siglec-2/CD22 may be the closest to the ancestor of mammalian Siglec family; it is unlikely that Siglec-2/CD22 was generated by gene duplication from another Siglec gene and later evolved to acquire similarity to aforementioned fly proteins under different environmental constraints. However, further analysis of the Siglec family in other animals will be needed to address this issue, as well as the issue of the relevant mouse orthologs of Siglec-9 and the other Siglec-3/CD33-related group.

Acknowledgments—We thank Drs. Hiromu Takematsu, Els Brinkman-Van der Linden, and Pascal Gagneux for general advice and helpful discussions.

Note Added in Proof—The results reported here are in good agreement with another paper on Siglec-9 published by Zhang *et al.* (Zhang, J. Q., Nicoll, G., Jones, C., and Crocker, P. R. (2000) *J. Biol. Chem.* **275,** 22121–22126) in this issue of the Journal.

REFERENCES

1. Schauer, R. (1982) *Sialic Acids: Chemistry, Metabolism and Function, Cell Biology Monographs*, Volume 10, Springer-Verlag, New York
2. Varki, A. (1992) *Glycobiology* **2,** 25–40
3. Kelm, S., and Schauer, R. (1997) *Int. Rev. Cytol.* **175,** 137–240
4. Varki, A. (1997) *FASEB J.* **11,** 248–255
5. Karlsson, K. A. (1998) *Mol. Microbiol.* **29,** 1–11
6. Gagneux, P., and Varki, A. (1999) *Glycobiology* **9,** 747–755
7. 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. (1998) *Glycobiology* **8**, v
8. Law, C. L., Aruffo, A., Chandran, K. A., Doty, R. T., and Clark, E. A. (1995) *J. Immunol.* **155,** 3368–3376
9. Nath, D., Van der Merwe, P. A., Kelm, S., Bradfield, P., and Crocker, P. R. (1995) *J. Biol. Chem.* **270,** 26184–26191
10. Vinson, M., Van der Merwe, P. A., Kelm, S., May, A., Jones, E. Y., and Crocker, P. R. (1996) *J. Biol. Chem.* **271,** 9267–9272
11. Van der Merwe, P. A., Crocker, P. R., Vinson, M., Barclay, A. N., Schauer, R., and Kelm, S. (1996) *J. Biol. Chem.* **271,** 9273–9280
12. Crocker, P. R., Mucklow, S., Bouckson, V., McWilliam, A., Willis, A. C., Gordon, S., Milon, G., Kelm, S., and Bradfield, P. (1994) *EMBO J.* **13,** 4490–4503
13. Sgroi, D., Varki, A., Braesch-Andersen, S., and Stamenkovic, I. (1993) *J. Biol. Chem.* **268,** 7011–7018
14. Freeman, S. D., Kelm, S., Barber, E. K., and Crocker, P. R. (1995) *Blood* **85,** 2005–2012
15. Kelm, S., Pelz, A., Schauer, R., Filbin, M. T., Tang, S., De Bellard, M. -E., Schnaar, R. L., Mahoney, J. A., Hartnell, A., Bradfield, P., and Crocker, P. R. (1994) *Curr. Biol.* **4,** 965–972
16. Cornish, A. L., Freeman, S., Forbes, G., Ni, J., Zhang, M., Cepeda, M., Gentz, R., Augustus, M., Carter, K. C., and Crocker, P. R. (1998) *Blood* **92,** 2123–2132
17. 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. (1999) *J. Biol. Chem.* **274,** 22729–22738
18. Falco, M., Biassoni, R., Bottino, C., Vitale, M., Sivori, S., Augugliaro, R., Moretta, L., and Moretta, A. (1999) *J. Exp. Med.* **190,** 793–801
19. Nicoll, G., Ni, J., Liu, D., Klenerman, P., Munday, J., Dubock, S., Mattei, M. G., and Crocker, P. R. (1999) *J. Biol. Chem.* **274,** 34089–34095
20. Angata, T., and Varki, A. (2000) *Glycobiology* **10,** 431–438
21. Floyd, H., Ni, J., Cornish, A. L., Zeng, Z. Z., Liu, D., Carter, K. C., Steel, J., and Crocker, P. R. (2000) *J. Biol. Chem.* **275,** 861–866
22. Taylor, V. C., Buckley, C. D., Douglas, M., Cody, A. J., Simmons, D. L., and Freeman, S. D. (1999) *J. Biol. Chem.* **274,** 11505–11512
23. Ulyanova, T., Blasioli, J., Woodford-Thomas, T. A., and Thomas, M. L. (1999) *Eur. J. Immunol.* **29,** 3440–3449
24. O'Keefe, T. L., Williams, G. T., Davies, S. L., and Neuberger, M. S. (1996) *Science* **274,** 798–801
25. 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. (1996) *Nature* **384,** 634–637
26. Sato, S., Miller, A. S., Inaoki, M., Bock, C. B., Jansen, P. J., Tang, M. L. K., and Tedder, T. F. (1996) *Immunity* **5,** 551–562
27. Nitschke, L., Carsetti, R., Ocker, B., Köhler, G., and Lamers, M. C. (1997) *Curr. Biol.* **7,** 133–143
28. Hennet, T., Chui, D., Paulson, J. C., and Marth, J. D. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 4504–4509
29. Roth, J., Kempf, A., Reuter, G., Schauer, R., and Gehring, W. J. (1992) *Science* **256,** 673–675
30. Davis, T. R., and Wood, H. A. (1995) *In Vitro Cell. Dev. Biol. Anim.* **31,** 659–663
31. Schauer, R., Krisch, B., Lapina, E. B., Shaw, L., Gerardy-Schahn, R., and Malykh, Y. N. (1999) *Glycoconj. J.* **16,** S32
32. Adams, M. D. *et al.* (2000) *Science* **287,** 2185–2195
33. The *C. elegans* Sequencing Consortium (1998) *Science* **282,** 2012–2018
34. Brinkman-Van der Linden, E. C. M., and Varki, A. (2000) *J. Biol. Chem.* **275,** 8625–8632
35. Coligan, E. C., Kruisbeek, A. M., Margulies, D. H., Shevach, E. M., and

Strober, W. (eds) *Current Protocols in Immunology,* John Wiley & Sons, Inc., New York
36. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425
37. Tamura, K., and Nei, M. (1993) *Mol. Biol. Evol.* **10,** 512–526
38. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
39. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5857–5864
40. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000) *Nucleic Acids Res.* **28,** 231–234
41. May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R., and Jones, E. Y. (1998) *Mol. Cell* **1,** 719–728
42. Tang, S., Shen, Y. J., DeBellard, M. E., Mukhopadhyay, G., Salzer, J. L., Crocker, P. R., and Filbin, M. T. (1997) *J. Cell Biol.* **138,** 1355–1366
43. Poy, F., Yaffe, M. B., Sayos, J., Saxena, K., Morra, M., Sumegi, J,. Cantley, L. C., Terhorst, C., and Eck, M. J. (1999) *Mol. Cell* **4,** 555–561
44. Gordon, L. A., Bergmann, A., Christensen, M., Danganan, L., Lee, D. A., Ashworth, L. K., Nelson, D. O., Olsen, A. S., Mohrenweiser, H. W., and Carrano, A. V. (1995) *Genomics* **30,** 187–194
45. Peiper, S. C., Ashmun, R. A., and Look, A. T. (1988) *Blood* **72,** 314–321
46. Takei, Y., Sasaki, S., Fujiwara, T., Takahashi, E., Muto, T., and Nakamura, Y. (1997) *Cytogenet. Cell Genet.* **78,** 295–300
47. Karlstrom, R. O., Wilder, L. P., and Bastiani, M. J. (1993) *Development* **118,** 509–522
48. Zallen, J. A., Yi, B. A., Bargmann, C. I. (1998) *Cell* **92,** 217–227
49. Mendoza, H. L., and Faye, I. (1999) *Dev. Comp. Immunol.* **23,** 359–374
50. Hinderlich, S., Stäsche, R., Zeitler, R., and Reutter, W. (1997) *J. Biol. Chem.* **272,** 24313–24318
51. Stäsche, R., Hinderlich, S., Weise, C., Effertz, K., Lucka, L., Moormann, P., and Reutter, W. (1997) *J. Biol. Chem.* **272,** 24319–24324
52. Maru, I., Ohta, Y., Murata, K., and Tsukada, Y. (1996) *J. Biol. Chem.* **271,** 16294–16299
53. Münster, A. K., Eckhardt, M., Potvin, B., Mühlenhoff, M., Stanley, P., and Gerardy-Schahn, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9140–9145
54. Effertz, K., Hinderlich, S., and Reutter, W. (1999) *J. Biol. Chem.* **274,** 28771–28778
55. Vely, F., and Vivier, E. (1997) *J. Immunol.* **159,** 2075–2077
56. Newman, P. J. (1999) *J. Clin. Invest.* **103,** 5–9
57. Sayos, J., Wu, C., Morra, M., Wang, N., Zhang, X., Allen, D., van Schaik, S., Notarangelo, L., Geha, R., Roncarolo, M. G., Oettgen, H., de Vries, J. E., Aversa, G., and Terhorst, C. (1998) *Nature* **395,** 462–469
58. Tangye, S. G., Lazetic, S., Woollatt, E., Sutherland, G. R., Lanier, L. L., and Phillips, J. H. (1999) *J. Immunol.* **162,** 6981–6985
59. Powell, L. D., Jain, R. K., Matta, K. L., Sabesan, S., and Varki, A. (1995) *J. Biol. Chem.* **270,** 7523–7532
60. Collins, B. E., Kiso, M., Hasegawa, A., Tropak, M. B., Roder, J. C., Crocker, P. R., and Schnaar, R. L. (1997) *J. Biol. Chem.* **272,** 16889–16895
61. Long, E. O. (1999) *Annu. Rev. Immunol.* **17,** 875–904
62. Reguigne-Arnould, I., Couillin, P., Mollicone, R., Faure, S., Fletcher, A., Kelly, R. J., Lowe, J. B., and Oriol, R. (1995) *Cytogenet. Cell Genet.* **71,** 158–162
63. Yousef, G. M., Luo, L. Y., and Diamandis, E. P. (1999) *Anticancer. Res.* **19,** 2843–2852
64. Das, H. K., Jackson, C. L., Miller, D. A., Leff, T., and Breslow, J. L. (1987) *J. Biol. Chem.* **262,** 4787–4793
65. Fitch, D. H., Bailey, W. J., Tagle, D. A., Goodman, M., Sieu, L., and Slightom, J. L. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7396–7400
66. Castellino, F. J., Davidson, D. J., Rosen, E., and McLinden, J. (1993) *Methods Enzymol.* **223,** 168–185
67. Leshko-Lindsay, L., and Corces, V. G. (1997) *Development* **124,** 169–180