

Cloning and Characterization of a Novel Mouse Siglec, mSiglec-F

DIFFERENTIAL EVOLUTION OF THE MOUSE AND HUMAN (CD33) Siglec-3-RELATED GENE CLUSTERS*[§]

Received for publication, September 6, 2001, and in revised form, September 27, 2001
Published, JBC Papers in Press, September 28, 2001, DOI 10.1074/jbc.M108573200

Takashi Angata[‡], Ravi Hingorani[§], Nissi M. Varki[‡], and Ajit Varki^{‡¶}

From the [‡]Departments of Medicine and Cellular and Molecular Medicine, the Glycobiology Research and Training Center, University of California, San Diego, La Jolla, California 92093-0687 and [§]BD Biosciences Pharmingen, Research Immunocytometry, San Diego, California 92121

A novel mouse Siglec (mSiglec-F) belonging to the subfamily of Siglec-3-related Siglecs has been cloned and characterized. Unlike most human Siglec-3 (hSiglec-3)-related Siglecs with promiscuous linkage specificity, mSiglec-F shows a strong preference for α 2-3-linked sialic acids. It is predominantly expressed in immature cells of the myelomonocytic lineage and in a subset of CD11b (Mac-1)-positive cells in some tissues. As with previously cloned Siglec-3-related mSiglecs, the lack of strong sequence similarity to a singular hSiglec made identification of the human ortholog difficult. We therefore conducted a comprehensive comparison of Siglecs between the human and mouse genomes. The mouse genome contains eight Siglec genes, whereas the human genome contains 11 Siglec genes and a Siglec-like gene. Although a one-to-one orthologous correspondence between human and mouse Siglecs 1, 2, and 4 is confirmed, the Siglec-3-related Siglecs showed marked differences between human and mouse. We found only four Siglec genes and two pseudogenes in the mouse chromosome 7 region syntenic to the Siglec-3-related gene cluster on human chromosome 19, which, in contrast, contains seven Siglec genes, a Siglec-like gene, and thirteen pseudogenes. Although analysis of gene maps and exon structures allows tentative assignments of mouse-human Siglec ortholog pairs, the possibility of unequal genetic recombination makes the assignments inconclusive. We therefore support a temporary lettered nomenclature for additional mouse Siglecs. Current information suggests that mSiglec-F is likely a hSiglec-5 ortholog. The previously reported mSiglec-3/CD33 and mSiglec-E/MIS are likely orthologs of hSiglec-3 and hSiglec-9, respectively. The other Siglec-3-like gene in the cluster (mSiglec-G) is probably a hSiglec-10 ortholog. Another mouse gene (mSiglec-H), without an apparent human ortholog, lies outside of the cluster. Thus, although some duplications of Siglec-3-related genes predated separation of the primate and rodent lineages (about 80–100 million years ago), this gene cluster underwent extensive duplications in the primate lineage thereafter.

Siglecs¹ (sialic acid-binding immunoglobulin superfamily lectins) are a family of cell surface proteins defined by certain shared structural motifs in the first two Ig-like domains and by their ability to recognize sialic acids via the first Ig V-sect domain (1–4). Each Siglec has a distinct expression pattern (3, 4), suggesting that these molecules play unique roles in the cells expressing them. In humans, ten canonical Siglecs (3–5) and one Siglec-like molecule (6, 7) have been reported so far, and our analysis of human genome sequences suggests that there is only one additional Siglec.²

In view of the usefulness of mice for functional gene analysis, especially by gene disruption, it is highly desirable to identify and analyze mouse orthologs of human Siglecs (hSiglecs). Mouse Siglecs (mSiglecs) orthologous to hSiglec-1/Sialoadhesin, hSiglec-2/CD22, and hSiglec-4/myelin-associated glycoprotein (MAG) have been well-characterized, with the latter two including phenotypes with gene disruptions (8–12). Siglec-3-related Siglecs, which include all other members not listed above, are characterized by their sequence similarity in the first two Ig-like domains, and by the presence of characteristic tyrosine-based signaling motifs in their cytosolic domains, an immunoreceptor tyrosine-based inhibitory motif (ITIM) and another putative signaling motif situated nearby (3, 4). Most of these Siglec-3-related human genes are clustered together on a region of ~500 kilobases (kb) of chromosome 19q (cytological band 19q13.3–13.4), suggesting their origin through repeated gene duplications (3, 4). The Siglec-3-related gene cluster in the human genome contains seven functional Siglec genes (Siglecs 3 and 5 through 10), one Siglec-like gene (Siglec-L1/S2V), and at least 13 Siglec-like pseudogenes (6). On the other hand, only two Siglec-3-related Siglecs have been reported so far in the mouse, *i.e.* mSiglec-“3”/CD33 (13) and mSiglec-E/MIS (14, 15). However, neither of these shows a clear one-to-one orthologous correspondence to hSiglecs (14, 16). Although it is likely that these mSiglec-3-related Siglecs, as well as ones yet to be reported, are present in a similar gene cluster in mouse chromosome 7 (17), there has so far been no report of a comprehensive

* This work was supported by United States Public Health Service Grants R37-GM32373 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 on-line version of this article (available at www.jbc.org) contains Tables I and II.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF293371.

[¶] To whom correspondence should be addressed: University of California, San Diego, School of Medicine 0687, La Jolla, CA 92093-0687. Tel.: 858-534-3296; Fax: 858-534-5611; E-mail: avarki@ucsd.edu.

¹ The abbreviations used are: Siglec, sialic acid-binding lectins of immunoglobulin superfamily; hSiglec and mSiglec, human and mouse Siglec, respectively; Ig, immunoglobulin; ITIM, immunoreceptor tyrosine-based inhibitory motif; PAA-Bio, biotinylated polyacrylamide; RACE, rapid amplification of cDNA ends; sialyl Le^x, Neu5Ac α 2-3Gal β 1-4[Fuc α 1-3]GlcNAc; EST, expressed sequence tag; BAC, bacterial artificial chromosome; kb, kilobase(s); UTR, untranslated region; GlcNAc, N-acetylglucosamine; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; DHFR, dihydrofolate reductase; ELISA, enzyme-linked immunosorbent assay; TBS, Tris-buffered saline; NK cells, natural killer cells; HAS1, hyaluronan synthase 1; ETFB, electron transfer flavoprotein β subunit; MAG, myelin-associated glycoprotein.

² T. Angata, S. Kerr, N. M. Varki, P. R. Crocker, and A. Varki, manuscript in preparation.

analysis of Siglec-3-related mouse Siglecs.

Here we report the molecular cloning and characterization of a novel mSiglec, as well as a genomic overview of Siglec-3-related gene clusters on mouse chromosome 7 and human chromosome 19. Through multiple approaches, the human and mouse Siglec orthology issue is addressed. Our analyses provide evidence of complex and differential evolutionary paths that CD33-related Siglecs have taken in the rodent and primate lineages.

EXPERIMENTAL PROCEDURES

Materials—Reagents were purchased from Fisher or Sigma, unless otherwise stated. Access rights to the assembled mouse genomic DNA sequence was purchased from Celera Genomics (Rockville, MD), under an agreement with the University of California System. Corresponding genomic DNA sequences in GenBank™ deposited by publicly funded sequencing projects were also utilized.

Cloning of a Novel mSiglec (mSiglec-F) cDNA—An EST clone (GenBank™ accession number AA672534), which shows sequence similarity to hSiglec-5, was identified and kindly provided by Dr. Paul Crocker (Dundee University, UK). The clone contained the 3'-end of the cDNA (~1.6 kb), including the coding sequence for the third (partial) and fourth Ig-like domains, a transmembrane domain and cytosolic tail, and a 3'-untranslated region (UTR). This putative mSiglec was tentatively given the letter designation mSiglec-F,³ because it is the sixth mSiglec reported so far and because of the difficulty in unequivocally identifying the human ortholog of this molecule. In an independent attempt to clone novel mSiglecs in our laboratory, a few cDNA fragments corresponding to the first Ig-like domain of mSiglecs were cloned, one of which was later revealed to be identical to mSiglec-F.⁴

Mouse bone marrow total RNA (C57BL/6 strain) was prepared with an RNeasy Midi kit (Qiagen) and used for the cDNA cloning. The 5'-end cDNA sequence of mSiglec-F was elucidated by 5'-rapid amplification of cDNA ends (RACE), as follows: the first-strand cDNA was synthesized with Superscript II (Invitrogen) using mSL5-SP1 (5'-CTGAACCTAT-TGTCCTCC-3') as primer, and oligo(dA) was added to the 3'-end of the first-strand cDNA using terminal deoxynucleotidyl transferase (Invitrogen) and dATP. Gene-specific cDNA fragments were amplified using the Expand Long Template PCR system (Roche Molecular Biochemicals), with first-round PCR using dT-NXB (5'-GCTAGCTCGAG-GATCCT₁₈-3') and mSL5-SP2 (5'-TGTCCTCCAAGCTTAC-3'), followed by nested PCR using NXB (5'-TTTGCTAGCTCGAGGATCCTT-TT-3') and mSL5-SP3 (5'-GACTTAACCCGTGAAGGAGCC-3') as primers. DNA fragments obtained were cloned into pCRII-TOPO TA cloning vector (Invitrogen), and sequences were analyzed. Primers were designed based on the revealed sequence, and a full-length cDNA was cloned by reverse transcription-PCR, as follows: random-primed first-strand cDNA was synthesized using SuperScript II with random DNA hexamer as primer, and the cDNA was amplified by PCR with primers mSL5 5'-UTR (5'-TAGAGCCTTTCACCTGTGTGATGCGG-3') and mSL5 3'-UTR1 (5'-CTTCATTCCTTCCAAAACCTCCCTCC-3'), followed by a nested PCR with mSL5 5'-Expr (5'-CCCTCTAGAGCCACCATTGCGGT-GGGCATGGCTGCTGCC-3'; *Xba*I site is underlined) and mSL5 3'-UTR2 (5'-CGCGGATCCGGACTCCAGCCAGCACTGGG-3'; *Bam*HI site is underlined). The resulting DNA fragments were digested with *Xba*I and *Bam*HI, cloned into *Xba*I-*Bam*HI sites of pBluescript II (Stratagene), and sequences were analyzed.

Preparation of a Full-length mSiglec-F Expression Construct and Introduction of an R114A Point Mutation—The cDNA fragment of mSiglec-F (*Xba*I-*Bam*HI) was subcloned into *Xba*I-*Bam*HI sites of pcDNA3.1(-) for expression of the full-length protein in mammalian cells. A point mutation (R114A) was introduced into mSiglec-F/pBlue-script II using a QuikChange site-directed mutagenesis kit (Stratagene), using primers mSL5 R114A Fwd (5'-GGGAGCTACTTCT-TCGCCTTGGATGGTTCTGTG-3'; the codon for A114 is underlined) and mSL5 R114A Rev (complementary to mSL5 R114A Fwd). The DNA

sequence was analyzed, and an appropriate fragment (*Xba*I-*Bam*HI) was subcloned to *Xba*I-*Bam*HI sites of pcDNA3.1(-).

Erythrocyte Rosetting Assay—This assay was performed as described previously (16, 18). In short, COS-7 cells were transfected with wild-type or R114A point-mutant mSiglec-F/pcDNA3.1(-) using LipofectAMINE Plus reagent (Invitrogen). The transfected cells were split at 1:2 into new plates 24 h after transfection and analyzed for human erythrocyte rosetting 48 h after the transfection, with or without *Arthrobacter ureafaciens* sialidase (Sigma Chemical Co.) pretreatment of COS-7 cells or erythrocytes.

mSiglec-F-Fc Protein Preparation—The mSiglec-F cDNA fragment encoding the first two Ig-like domains was amplified by PCR using *Pwo* polymerase (Roche Molecular Biochemicals) with primers mSL5 5'-Expr and mSL5 3'-Chi2D (5'-ATCTCCAGGACACCCTGATGGTC-3'), digested with *Xba*I and cloned to *Xba*I-*Eco*RV sites of EK-Fc/pEDDc vector (16, 18, 19). Upon mammalian cell transfection, the resulting construct (mSiglec-F-EK-Fc/pEDDc) expressed a recombinant soluble mSiglec-F-Fc protein (a fusion protein of the first two Ig-like domains of mSiglec-F and human IgG Fc fragment, with an enterokinase cleavage site/FLAG tag (DYKDDDDK) between them). The construct also contains mouse dihydrofolate reductase (DHFR) cDNA as a selection marker. A CHO cell line stably expressing mSiglec-F-Fc protein was established by transfection of CHO-DHFR⁻ cells with mSiglec-F-EK-Fc/pEDDc, followed by selection in minimal essential medium- α supplemented with 10% dialyzed fetal calf serum and 20–200 nM methotrexate. Clones expressing the recombinant protein were established by manual colony isolation, and the best clone was expanded. Recombinant fusion protein secreted into culture medium was purified by adsorption to Protein A-Sepharose (Amersham Pharmacia Biotech).

Sialic Acid Binding Specificity of mSiglec-F—Sialic acid linkage specificity was analyzed by a standard ELISA assay as previously described (16, 18, 19). Briefly, protein A (5 μ g/ml in 50 mM sodium bicarbonate buffer, pH 9.5) was immobilized to the wells of 96-well plates (Nunc, 269620) overnight at 4 °C. The wells were washed three times with ELISA buffer (20 mM HEPES, 125 mM NaCl, 1 mM EDTA, 1% bovine serum albumin, 0.01% Na₂S₂O₃, pH 7.45), blocked with the same buffer at room temperature for 1 h, followed by successive incubation with the following (each followed by three washes with the ELISA buffer): mSiglec-F-Fc (5 μ g/ml), 2 h; biotinylated polyacrylamide probes (PAA-Bio) multiply substituted with sialylated oligosaccharides (Glycotek, Rockville, MD; 10 μ g/ml), 2 h; alkaline phosphatase-conjugated streptavidin (Invitrogen, 1:1000 dilution), 1 h. After the final wash, phosphatase substrate (10 mM *p*-nitrophenyl phosphate in 100 mM sodium carbonate and 1 mM MgCl₂) was added to each well, and absorbance at 405 nm was measured using a microplate reader after incubation at room temperature up to 3 h.

Preparation of Rat Monoclonal Antibodies against mSiglec-F—Lou rats (Harlan, Indianapolis, IN) were immunized intraperitoneally with mSiglec-F-Fc. Briefly, rats were immunized three times with 50 μ g of immunogen in Titermax Gold (Cytrx, Norcross, GA) followed by a final boost with 50 μ g of immunogen in phosphate-buffered saline. After fusion of the spleen cells with the Yb/20 myeloma cells, the hybridomas were screened by ELISA on antigen-coated plates. The initial screen on mSiglec-F-coated plates was followed by a second screen on human IgG-coated plates. Clones positive only for mSiglec-F were tested by flow-cytometry on primary cells. Positive clones were subcloned twice and isotyped using the monoclonal antibody-based rat Ig Isotyping kit (BD Biosciences Pharmingen, San Diego, CA). All clones gave a similar staining pattern on primary cells. The selection of clone E50–2440 (rIgG2a) was based on its staining intensity on primary cells.

Flow-cytometric Analysis and Sorting Using the mSiglec-F-specific Antibodies—Bone marrow, spleen, blood, and lymph node cells obtained from Balb/c, C57BL/6, and C3H mice (Jackson Laboratories, Bar Harbor, ME) were labeled with either E50–2440 or control rat IgG2a (R35–95). Cells were counter-stained with CD3-FITC (145–2C11), CD11c-FITC (HL3), CD14-FITC (rmC5–3), CD45R/B220-FITC (RA3–6B2), CD54-FITC (3E2), CD80-FITC (1G10), CD95-FITC (Jo2), CD117-FITC (2B8), Ly6A/E-FITC (D7), or Ly6G-FITC (Gr-1) along with CD11b-APC (M1/70). All antibodies used were from BD Biosciences Pharmingen. Briefly, 1×10^6 cells were suspended in staining buffer (phosphate-buffered saline containing 3% fetal calf serum) and mouse Fc block (2.4G2) added to block the Fc receptors. Staining was carried out in three steps. Initially the cells were incubated with either purified E50–2440 or isotype control followed by the second incubation with biotin-labeled anti-rat IgG2a (RG7/1.30). Finally, cells were incubated with fluorescein isothiocyanate-conjugated antibodies (given above), streptavidin-PE, and CD11b-APC. All incubation steps were for 20 min at 4 °C followed by two washes with staining buffer. Cells were acquired

³ The cDNA sequence of mSiglec-F was deposited under GenBank™ accession number AF293371. Much of the data on mouse genomic sequences discussed here was generated through use of the Celera Discovery System and Celera's associated databases. As per the agreement between Celera and the University of California, the relevant sequences have been transferred to the publicly available section of the Celera Genomics databases (available at www.celera.com/publication-library) and assigned accession number CSN011.

⁴ E. C. M. Brinkman-van der Linden, unpublished results.

on FACSCalibur (BD Biosciences, San Jose, CA), and the profiles were analyzed using the software CellQuest (BD Biosciences).

For sorting experiments cells were stained as above with E50–2440 and CD11b-APC. After staining the cells were suspended at 10×10^6 cells/ml and sorted to obtain CD11b^{lo}E50–2440^{hi}, CD11b^{hi}E50–2440^{lo}, CD11b^{lo}E50–2440^{lo}, and CD11b^{neg}E50–2440^{neg} populations on the FACS Vantage fluorescence-assisted cell sorter (BD Biosciences).

Immunohistochemical Analysis Using the mSiglec-F-specific Antibodies—Immunohistology was performed on frozen sections of a number of mouse tissues using purified rat antibody E50–2440 to mSiglec-F. Following a 10-min fixation in acetone, the sections were washed in Tris-buffered saline (TBS), and nonspecific binding sites were blocked with 1% bovine serum albumin in TBS. Endogenous biotin in tissues was blocked by sequential incubations in 0.1% avidin, followed by washes and then 0.01% biotin, followed by washes. Anti-mSiglec-F antibody, at 1 μ g/ml, was incubated on the tissue sections for at least 30 min in a humidified chamber. Irrelevant primary antibodies were used as controls. Unbound antibody was washed off, and the sections were sequentially incubated with biotinylated anti rat antibody (BD Biosciences Pharmingen, San Diego, CA) for 30 min, followed by washes, and then with alkaline phosphatase-conjugated streptavidin (Jackson ImmunoResearch), followed by washes. Bound antibody was then detected using the Vector Blue substrate, nuclei were counter-stained using nuclear fast red (Vector Laboratories, Burlingame, CA), and the sections were then mounted for viewing under a bright field microscopy. Digitized images were captured using a Sony DKC 5000 charge-coupled device camera, IMAGE (National Institutes of Health), and Photoshop (Adobe).

For double immunostains, fluorescent labels were used with sections of mouse thymus and spleen. Serial sections were first singly immunostained using either anti-Siglec-F or biotinylated anti-CD11b antibodies, and separate sections were used for the double labeling. Specific binding was detected using CY3-conjugated donkey anti-rat or CY2-labeled streptavidin (Jackson ImmunoResearch). Digitized photographs were taken using epifluorescence with a Zeiss microscope and red or green barrier filters. Overlays were performed digitally using Photoshop.

Genomic Sequence Analysis—We searched for Siglec genes and pseudogenes in mouse genomic DNA sequence in the Celera mouse genome data base using the TBLASTN program (20) with known human and mouse Siglec sequences (the N-terminal 150 amino acids, approximately covering the first Ig-like domain) as templates. Identification of hSiglec pseudogenes was previously reported (6). Sequence matches with expectation (*E*) value < 1 were considered significant. Only DNA segments encoding a remnant (fossil) of full V-set domain exon were considered as pseudogenes. Thus, a very short (~100 nucleotides) DNA segment immediately upstream (~0.5 kb) of mSiglec-3/CD33 gene showing significant similarity to mSiglec-3 was not counted as a pseudogene. A DNA segment showing very high (>95%) sequence identity to mSiglec-G but lacking the V-set domain was also excluded.

Genomic structures of human and mouse Siglec genes were deduced from sequence alignment of cDNA with genomic DNA, with adjustments to conform to standard sequence context of splicing sites.

Phylogenetic Analysis—Nucleotide sequences for most of the Siglec cDNAs from mouse (mSiglecs 1 through 4 and E) and human (hSiglecs 1 through 11 and L1) were obtained from GenBank™. The sequence of mSiglec-H was determined by sequencing an EST clone (IMAGE clone number 1430130) containing a full-length cDNA of this gene.⁵ The cDNA sequence of mSiglec-G was deduced from a corresponding genomic DNA region based on sequence similarity to hSiglec-10. Nucleotide sequences encoding the signal peptide and the first and second Ig-like domains of Siglecs were translated and aligned using ClustalW (available at www2.ebi.ac.uk/clustalw) and used for phylogenetic analysis using PAUP 4.0 (Sinauer Associates). A phylogenetic tree was reconstructed using the neighbor-joining method (21) from a distance matrix based on mean amino acid difference. Phylogenetic trees of essentially identical topology were obtained by the maximum parsimony method (22, 23) as well as from a distance matrix based on nucleotide sequences (data not shown). The sequence of hSiglec-L1 was omitted from the analysis because of its unusual structure (two V-set Ig-like domains) (6, 7). A phylogenetic tree of only the Siglec-3-related Siglecs and Siglec-4, based on the amino acid sequences of transmembrane + cytosolic domains, was also reconstructed by the same method. The sequence of mSiglec-H was omitted from this analysis due to its

very short cytosolic domain, which greatly restricts the amount of information usable for phylogenetic analysis if included.

RESULTS

Cloning of a Novel mSiglec cDNA—A partial mouse EST clone showing high sequence similarity to hSiglec-5 was identified, and full-length cDNA of this molecule was cloned through 5'-RACE and reverse transcription-PCR from mouse bone marrow. The cDNA encodes a Siglec-like molecule (569 amino acids) with four Ig-like domains, a single-pass transmembrane domain and cytosolic tail (Fig. 1A). The amino acid sequence of the translated product contains the defining structural features of Siglecs, *i.e.* three amino acids whose side chains are known to contact sialic acid in Siglec-1/Sialoadhesin (a conserved arginine residue and two aromatic amino acids, one near the N-terminal and the other near the conserved arginine residue) (24), as well as three conserved cysteine residues in each of the first and second Ig-like domains (Fig. 1A). Judging from the sequence similarity of the first two Ig-like domains and cytosolic tail, this new molecule belongs to a group of Siglecs related to Siglec-3, which typically contains two putative tyrosine-based signaling motifs in both human and mouse (3, 4). However, despite high sequence similarity to hSiglec-5 in the fourth Ig-like domain and shared Ig-like domain numbers, the similarity is not exclusive to hSiglec-5 in the other regions of the molecule (Fig. 1B). Because this new mouse molecule is the sixth mSiglec to be reported and its human ortholog could not be unequivocally established at this point, it was tentatively given the letter designation mSiglec-F. The previously reported mSiglec-“3”/CD33 (13) and mSiglec-E/MIS (14, 15) also lack clear orthologous correspondence to hSiglecs. As discussed below, this issue is addressed here by multiple approaches, including phylogenetic analysis and whole-genome survey of human and mouse Siglecs, as well as analyses of gene structure and chromosomal localization.

Sialic Acid Binding Specificity of the Novel mSiglec—COS-7 cells transiently transfected with mSiglec-F full-length expression construct showed clear sialic acid-dependent binding of human erythrocytes (Fig. 2), justifying the denomination of this molecule as a Siglec (1). A point mutation at the conserved arginine residue (R114A) abolished this binding, proving that this arginine residue is essential in sialic acid recognition by this molecule (Fig. 2), as has been shown in all other Siglecs studied to date (6, 16, 18, 25–28). Unlike some previously reported Siglecs, pretreatment of the COS cells was not essential to liberate the cell surface Siglec from masking by *cis*-ligands (Fig. 2). To determine a more detailed ligand binding specificity, a recombinant fusion protein mSiglec-F-Fc was prepared and used in an ELISA assay, with biotinylated polyacrylamide arrays carrying multiple copies of sialylated oligosaccharides as ligands (16, 18, 19). As shown in Fig. 3, mSiglec-F showed a strong preference toward α 2–3-linked sialic acids. This is in contrast to many Siglec-3-related hSiglecs, which show rather promiscuous binding to both α 2–3- and α 2–6-linked sialic acids (3, 29). Substitution at the proximal GlcNAc with fucose residue (*i.e.* sialyl Le^x) strongly interfered with the interaction. The latter finding is similar to that for hSiglecs recognizing α 2–3-linked sialic acids (29), with the known exception of hSiglec-9 (16).

Expression Pattern of the Novel mSiglec—Rat monoclonal antibodies against mSiglec-F were prepared as described under “Experimental Procedures” and used for flow-cytometry and immunohistochemical analyses. Four clones with identical staining pattern were obtained, and the clone (E50–2440) with the highest staining intensity on primary cells was selected for the phenotyping studies. The staining profile in bone marrow shows that mSiglec-F is present on CD11b^{lo} as well as on

⁵ T. Angata and A. Varki, unpublished.

FIG. 2. Erythrocyte rosetting by COS cells expressing mSiglec-F. When COS-7 cells were transfected with wild-type mSiglec-F expression construct (A–C), the cells show distinct erythrocyte rosetting, with (B) or without (A) pretreatment with *A. ureafaciens* sialidase to remove *cis*-ligands on COS cell surface. This binding is abolished by treating erythrocyte with sialidase (C). On the other hand, COS-7 cells transfected with mSiglec-F expression construct with R114A point mutation (D–F) do not show erythrocyte rosetting, with (E) or without (D) sialidase pretreatment of COS-7 cells.

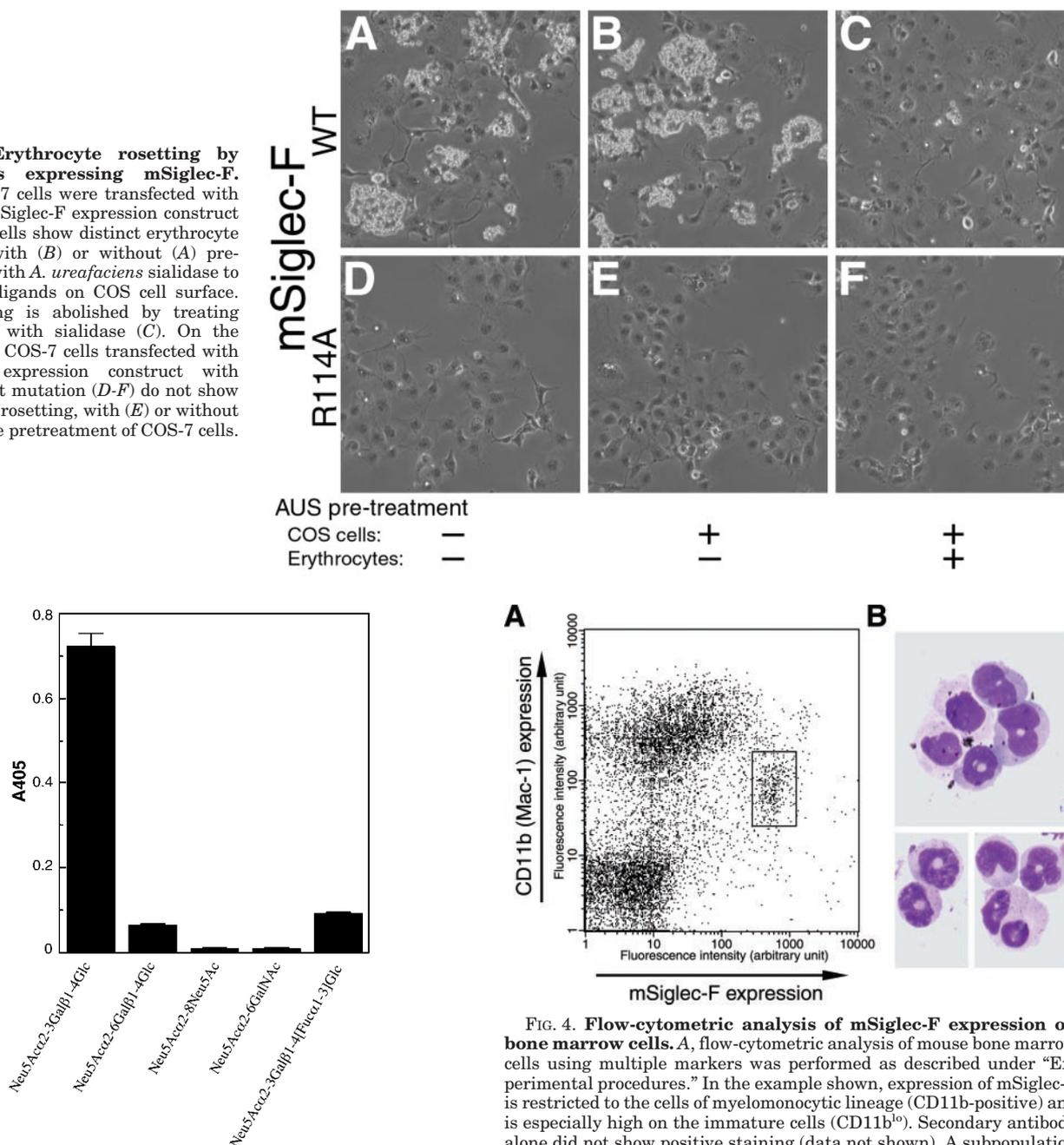


FIG. 3. Sialic acid linkage specificity of mSiglec-F. The mSiglec-F-Fc recombinant fusion protein was immobilized on the wells of a 96-well plate, and binding of PAA-Bio probes (biotinylated polyacrylamide array carrying multiple sialylated oligosaccharides) was analyzed as described under “Experimental Procedures.” Mean read out at 405 nm (*A405*) of triplicate wells for each ligand were plotted. Error bars indicate standard deviation.

medulla (Fig. 5, A–C), most of the cells that expressed mSiglec-F also expressed CD11b (Mac-1); however, there were also rare single positive mSiglec-F cells seen. In the spleen, the expression of mSiglec-F did not completely overlap with that of CD11b (Fig. 5, D–F), whereas it showed better correlation with that of Ly6G (Gr-1) (data not shown). Scattered mSiglec-F cells were also noted in the lung parenchyma. No staining was seen in the heart, liver, kidney, pancreas, lymph nodes, brain, salivary gland, or gastrointestinal tract. Taken together, the data indicate that there is a prominent expression of mSiglec-F on immature bone marrow cells of the myelomonocytic lineage, and a decreased expression in mature neutrophils and mono-

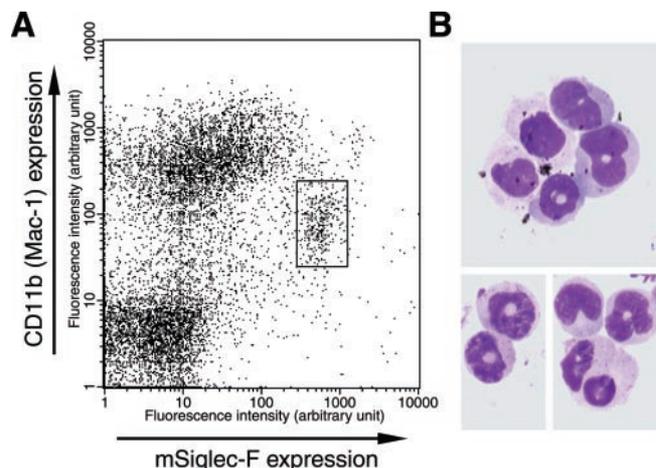
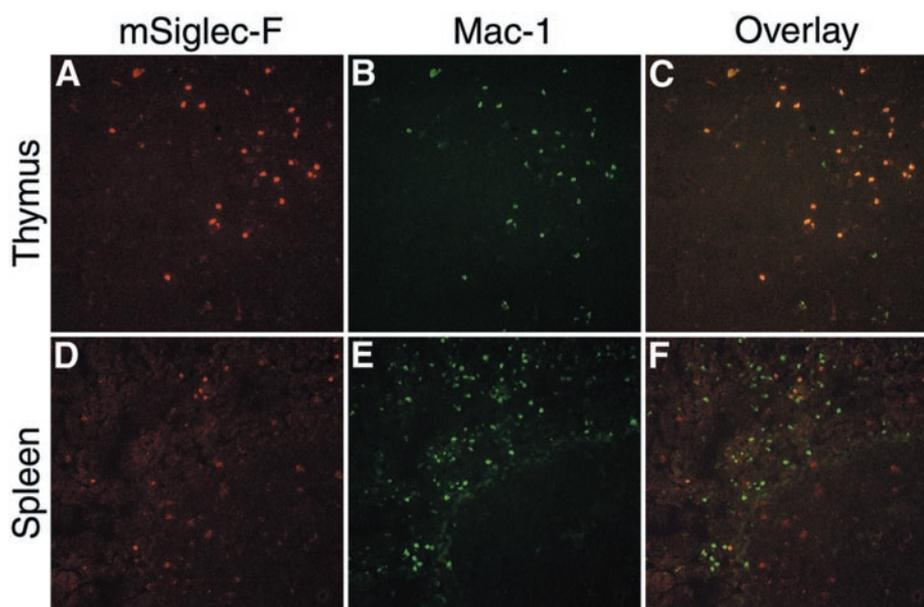


FIG. 4. Flow-cytometric analysis of mSiglec-F expression on bone marrow cells. A, flow-cytometric analysis of mouse bone marrow cells using multiple markers was performed as described under “Experimental Procedures.” In the example shown, expression of mSiglec-F is restricted to the cells of myelomonocytic lineage (CD11b-positive) and is especially high on the immature cells (CD11b^{lo}). Secondary antibody alone did not show positive staining (data not shown). A subpopulation of CD11b^{lo} cells in peripheral blood was also strongly positive for mSiglec-F (data not shown). B, composite picture of a Wright-Giemsa stain of the mSiglec-F^{hi} population (A, boxed), showing examples of immature myelomonocytic precursors at various stages of differentiation.

cytes. The nature of the mSiglec-F-positive cells found in thymus, spleen, and lung requires further investigation. Of particular note, the high level of mSiglec-F expression on immature bone marrow cells of myelomonocytic lineage (see Fig. 4) provides the first known marker that is selective for these cell populations in the mouse.

Comparison of Mouse and Human Siglec-3-related Gene Clusters—To gain an overview of all mSigslecs and to compare the structure of mouse and human Siglec-3-related gene clusters, we analyzed mouse genomic DNA sequences recently assembled to near completion by Celera Genomics. The previously known mSigslecs with clear human orthologs (Siglec-1/Sialoadhesin, Siglec-2/CD22, and Siglec-4/myelin-associated glycoprotein) were confirmed to be single-copy genes. We also found a mouse genomic region (chromosome 7, cytological band

FIG. 5. Immunohistochemical analysis of mSiglec-F expression in mouse tissues. Some examples are shown. Expression of mSiglec-F (A) in thymus is similar to that of CD11b (B), with the latter showing somewhat broader distribution (C; cells expressing both markers are shown in yellow). In spleen, mSiglec-F-positive cells (D) and CD11b-positive cells (E) are only partially overlapping (F).



7B2) syntenic to the hSiglec-3-like gene cluster (chromosome 19, cytological band 19q13.3–13.4) in which a number of mSiglec genes and pseudogenes are clustered (Fig. 6). Although the same genomic region is also being sequenced by the publicly funded mouse genome sequencing project, the relevant scaffolds are not assembled into a contiguous entity as of the time of this writing. In the human genome, the Siglec-3-related gene cluster is flanked by the kallikrein gene cluster on the centromeric side by ~40 kb (the closest one is *KLK-L6/KLK14*), and by a hyaluronic acid synthetase gene (*hyaluronan synthase 1 (HAS1)*) at the telomeric side by ~90 kb. The mSiglec-3-related gene cluster is also flanked by the kallikrein gene cluster on one side. On the other hand, the *HAS1* ortholog was not found near the cluster. According to a recent study (31), there is a discontinuity of mouse-human synteny between *ETFB* (electron transfer flavoprotein β subunit) and *HAS1*. In fact, a region of ~3 Mb containing the mouse Siglec-3-related gene cluster appears to have undergone intrachromosomal translocation, whereas the genomic segment containing mouse *HAS1* is found on mouse chromosome 17 (31). Nevertheless, all but one canonical Siglec-3-like sequence (mSiglec-H) are found in the above cluster in the mouse genome. Thus, we consider the mouse Siglec-3-related gene cluster on chromosome 7 to be unique and the only cluster equivalent to that in the human genome.

Besides mSiglec-F, there is only one other Siglec-like putative gene (mSiglec-G) in the cluster yet to be formally reported. It appears to be an ortholog of hSiglec-10, judging from its high sequence identity to hSiglec-10 and similar chromosomal localization in relation to the proximal genes (*ETFB*, *NKG7*, and *LIM2*). Including this putative gene (which is a functional gene according to Dr. P. Crocker⁶), there are four Siglec genes and two pseudogenes in this region. This is in striking contrast to the syntenic region of the human genome, which contains seven Siglec genes, one Siglec-like gene (*Siglec-L1*), and thirteen pseudogenes (exon fossils of Siglec-like Ig V-set domains). As mentioned earlier, the Siglec-3-related mouse genes do not show strong one-to-one correspondence to hSiglecs, with the only exception of mSiglec-G, the putative hSiglec-10 ortholog. However, comparison of the gene maps of Siglec-3-related human and mouse gene clusters, as well as the gene structures of individual Siglecs (Fig. 7), provide clues to address the orthol-

ogy between mouse and human Siglecs. First, the previously reported mSiglec-“3”/CD33 (13) has a shared domain structure with hSiglec-3 (Fig. 7) and a similar chromosomal localization in relation to other Siglec genes (Fig. 6). Thus, it is the best candidate for an ortholog of hSiglec-3/CD33. By the same criteria, mSiglec-E/MIS (14, 15) is a probable ortholog of hSiglec-9. Although hSiglecs 7 and 9 are localized close together in the human gene cluster and show a high degree of similarity at the mRNA and polypeptide levels, their genomic structures are different: Although hSiglec-7 gene contains a remnant of V-set-domain exon (*Siglec-P2*) between the exons encoding the first and second Ig-like domains of the protein product, hSiglec-9 and mSiglec-E lack this signature motif (Fig. 7). Identification of the mSiglec-F ortholog in humans is not as straightforward, due to the lack of useful genomic landmarks and the presence of two candidate genes, namely hSiglecs 5 and 6; the former has exons for four Ig-like domains, whereas the latter has exons for three Ig-like domains and one exon fossil similar to the fourth Ig-like domain exon in mSiglec-F and hSiglec-5 (Fig. 7). However, in view of the fact that, unlike hSiglec-6, mSiglec-F is not expressed on B lymphoid cells but restricted to cells of myeloid lineage, we consider that hSiglec-5 is more likely to be the ortholog of mSiglec-F. Further support comes from the finding that mSiglec-F is not expressed in the mouse placenta (data not shown), whereas hSiglec-6 is expressed in this fetal organ (19).

Phylogenetic Analysis of Human and Mouse Siglecs—To further understand the phylogenetic relationships of the mSiglecs and hSiglecs, we separately compared the N-terminal (first two Ig-like domains) and C-terminal (the transmembrane domain plus the cytosolic tail) amino acid sequences of the Siglecs. As shown in Fig. 8A, Siglecs 1, 2, and 4 of mouse and human show a clear one-to-one correspondence in the phylogenetic tree based on the amino acid sequences of the first two Ig-like domains. On the other hand, the Siglec-3-related Siglecs do not show such a clear relationship. However, mSiglec-G appears to form a monophyletic clade with hSiglecs 10 and 11, and mSiglec-E with hSiglecs 7, 8, and 9 (Fig. 8A). In these cases, the lack of one-to-one orthologous correspondence can be attributed to specific gene duplication(s) that took place in the lineage leading to humans after the split of the primate and rodent lineages (estimated to be 80–100 million years ago) (32). This interpretation is also supported by the phylogenetic tree reconstructed from the amino acid sequences of transmembrane plus cytosolic domains of Siglecs (Fig. 8B). Here again mSiglec-E

⁶ P. Crocker, personal communication.

FIG. 6. Comparison of the mouse and human *Siglec*-3-related gene clusters. h*Siglec* genes are indicated by number and h*Siglec* pseudogenes are denoted as P1 to P13. Mouse *Siglec* genes are indicated by a letter (see text for discussion). Other gene symbols used are: *KLK-L6*, kallikrein-like 6 (KLK14); *ETFB*, electron transfer flavoprotein β subunit; *NKG7*, natural killer group 7; *LIM2*, lens intrinsic membrane protein 2; *ZNF*, zinc-finger protein.

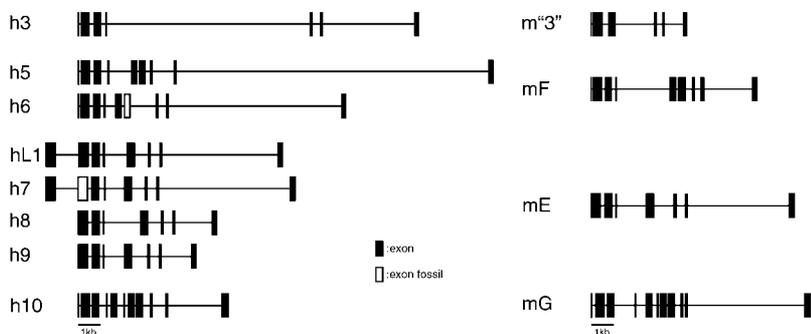
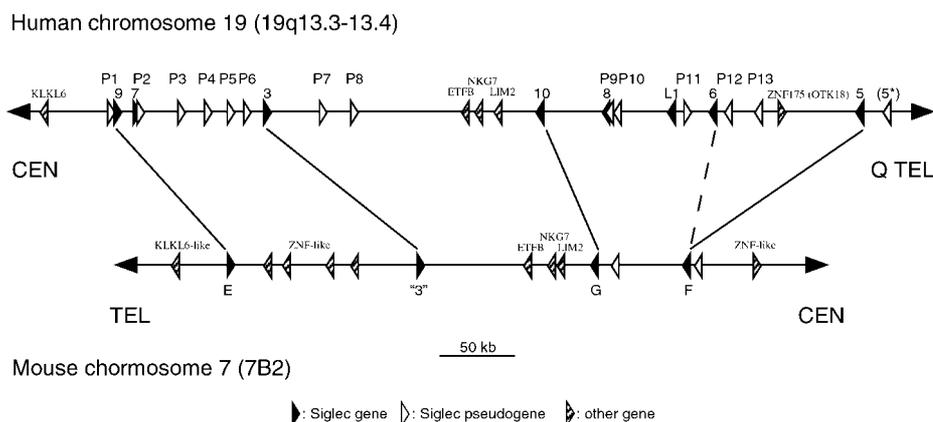


FIG. 7. Gene structures of mouse and human Siglecs in the *Siglec*-3-related gene clusters. Human (left column) and mouse (right column) Siglecs sharing similar domain structures are grouped. Solid boxes and open boxes indicate exons and exon fossils, respectively. The mouse genomic sequences used for these analyses were largely from the Celera Genomics data base, with some input from the publicly funded mouse genome project.

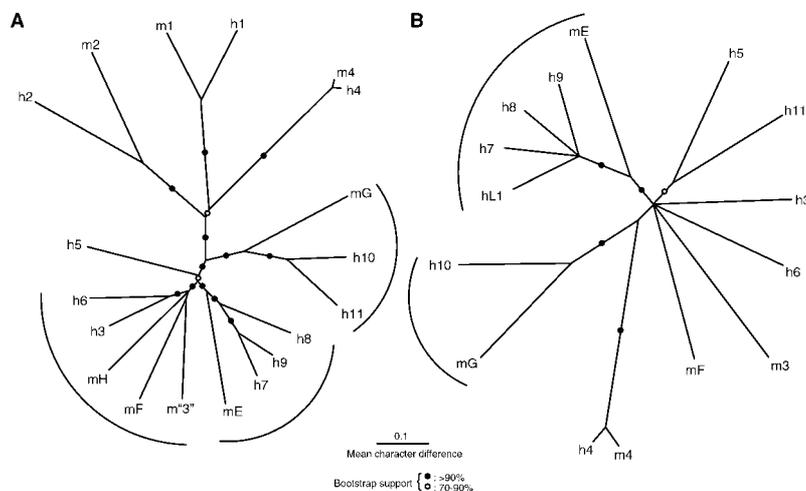


FIG. 8. Phylogenetic trees of mouse and human Siglecs. A, phylogenetic tree derived from N-terminal amino acids (signal peptide + first and second Ig-like domains) of Siglecs. Human *Siglec*-L1 was excluded from the analysis due to its unusual structure (two tandem V-set domains). B, phylogenetic tree derived from C-terminal amino acids (transmembrane + cytosolic domains) of *Siglec*-3-related Siglecs and *Siglec*-4. Mouse *Siglec*-H was excluded from the analysis due to its very short cytosolic tail. Siglecs 1 and 2 were also excluded, due to the lack of C-terminal sequence similarity to the Siglecs being analyzed. An isoform of m*Siglec*-4 lacking exon 12 (42) was used for sequence alignment. Major monophyletic branches containing *Siglec*-3-related mouse and human Siglecs were indicated with quadrants. Internodes supported by high bootstrap values (for 1000 bootstrap replications) were indicated with closed (>90%) and open (70–90%) circles, respectively. The sequences used for these analyses are presented in the supplemental material.

forms a monophyletic clade with hSiglecs 7, 8, and 9 (and L1), and m*Siglec*-G groups with h*Siglec*-10. The fact that h*Siglec*-11 does not belong to “*Siglec*-10 clade” in this tree suggests that this gene is a product of a gene duplication and recombination. On the other hand, the evolutionary relationships of other *Siglec*-3-related genes in this analysis is not quite consistent with our interpretation of orthology: Thus, in the tree based on N-terminal amino acid sequences (Fig. 8A), m*Siglec*-3 and h*Siglec*-3 are paraphyletic rather than monophyletic, and the same is true of m*Siglec*-F and h*Siglec*-5. Also, this clade contains Siglecs with 2 and 4 Ig-like domains (h*Siglec*-6 has an exon fossil corresponding to the fourth Ig-like domain, as dis-

cussed above). These anomalies may be due to unequal genetic recombination(s), including gene conversion of some *Siglec* genes, which consist mostly of modular exons encoding one Ig-like domain each, with the potential for “exon shuffling.” The phylogenetic tree of C-terminal amino acid sequences (Fig. 8B) does not provide significant clues to resolve this problem, either. A more precise understanding of the evolutionary paths taken by this gene clade will require more data on related genomic sequences from other mammalian species.

Another clear implication of this phylogenetic analysis is that the ancestral *Siglec*-3-related genes (proto-Siglecs 3, 9, 10, and possibly 5) were already present in the common ancestor of

primates and rodents. In addition, at least one of these members of distinct clades seems to have survived in each lineage through repeated gene duplications and gene losses. This finding suggests that, despite common basic properties (sialic acid recognition and cellular signaling), these molecules are not redundant but rather are playing distinct roles, as has been proposed previously, based on their distinct expression patterns in different cell types (3).

DISCUSSION

When the Siglecs were first recognized as a distinct family of sialic acid binding lectins, there were only four members (1). Of these, Siglec-1/Sialoadhesin, Siglec-2/CD22, and Siglec-4/MAG have clear orthologs in humans and mice, each with a very similar domain structure, cell-type distribution, and sialic acid binding specificity. An apparent murine ortholog of human CD33 (subsequently called Siglec-3 in humans) had already been reported (13). However, the subsequent discovery of an entire subfamily of human Siglec-3-related Siglecs complicated this interpretation (16). The sequence of another Siglec-3-related mSiglec called MIS (15) or mSiglec-E (14) did not serve to resolve the situation. Here we report the molecular cloning and characterization of another novel mSiglec, as well as a comprehensive genomic overview of mouse and human Siglecs, with a particular focus on comparing the mouse and human Siglec-3-related gene clusters. The overall conclusion is that humans have many more Siglec-3-related genes and pseudogenes than mice do. This analysis has also revealed differential and complex evolutionary paths, which seemed to have been more extensively exploited in the primate lineage, as well as the ancient origin of this multigene cluster, predating the split of the primate and rodent lineages (estimated to be 80–100 million years ago).

Based on expression analysis with monoclonal antibodies, phylogenetic analysis, gene structure, and gene maps, we propose that the new mouse Siglec, mSiglec-F, is likely to be a hSiglec-5 ortholog. The general expression pattern of mSiglec-F on hematopoietic cells (on the myelomonocytic cell lineage) is indeed similar to that of hSiglec-5 (33) and dissimilar to that of the other human homolog to mSiglec-F, hSiglec-6, which is found on B cells and placental trophoblasts (19). However, there is no report on the analysis of hSiglec-5 expression in tissues other than peripheral leukocytes. Also, it remains possible that a major chromosomal rearrangement in the past might have eliminated the true ortholog of hSiglec-5 in the lineage leading to mouse, and that another Siglec ancestor might have been recruited into this expression pattern. In this respect, it is of note that recent analysis of mouse chromosome 7 indicates a chromosomal breakage between ETV6 and HAS1 in the lineage leading to the mouse, right adjacent to the Siglec-3-related gene cluster (31). A related report has suggested that this and some other breakages in chromosome 7 may be unique to the mouse (or perhaps to rodents), because the linkage maps of chromosome regions syntenic to human chromosome 19 q-arm in primates, dog, cat, and cattle are more or less similar to that of human (17). To resolve this and other Siglec orthology issues with better precision, we need more information on genomic DNA sequences of Siglec-3-related gene clusters of other mammalian species, including other rodents (*e.g.* rat) and primates (*e.g.* baboon and chimpanzee), as well as other mammals such as bovines and carnivores. Many of these taxa are currently being actively pursued or considered as candidates for genome sequencing and/or major EST projects.

Human chromosome 19 is unusual in terms of its higher gene density compared with other chromosomes (34). Hence the syntenic chromosomal regions in other mammalian organisms

may well become good candidates for pilot study by BAC-based genomic sequencing. To understand the evolution of Siglecs on a deeper time scale, we would have to evaluate other vertebrate branches as well. In this regard, we have noted that some EST clones from African clawed frog *Xenopus laevis* (GenBankTM accession numbers BG017920 and BG731140) and some genomic DNA fragments of puffer fish (available at www.jgi.doe.gov/programs/fugu.htm) show distinct similarities to mammalian Siglecs, including some of the highly conserved amino acid residues.⁷ This strongly suggests that the Siglec family of molecules is present in all vertebrates. On the other hand, we failed to find Siglec orthologs in the *Drosophila* and *Caenorhabditis elegans* genomes (16). Thus, true Siglecs may have emerged only in the deuterostome lineage of animals, where sialic acids are prominently expressed.

The presence of putative intracellular signaling motifs in the cytosolic tail of mSiglec-F suggests that it is involved in intracellular signaling, as has been shown in some other Siglec-3-related Siglecs (7, 14, 15, 28, 35, 36). Although it is beyond the scope of this report, the identification of proteins interacting with the cytosolic tail of mSiglec-F is of interest, especially those that interact with the distal motif, which does not conform to canonical ITIM motif but is highly conserved among all Siglec-3-related Siglecs. We are also pursuing natural ligand(s) for the extracellular domain of mSiglec-F, which may regulate intracellular signaling. Although α 2–3-linked sialic acids are abundantly expressed in mouse tissues, mSiglec-F may prefer a limited number of specific ligands expressing this structure in a preferred manner, or in a particular polypeptide structural context, as in the case of the specific interaction of P-selectin with P-selectin glycoprotein ligand-1 (37). Alternatively, α 2–3-linked sialic acid recognition by mSiglec-F, regardless of other structural contexts, may constitutively send suppressive signals via its cytosolic tail. As in the interaction of killer cell Ig-like receptors and major histocompatibility complex molecules (38), such a suppressive signal may cease in the absence of ligand, *e.g.* upon contact with abnormally under-sialylated cells (*e.g.* virus-infected cells), which may in turn lead to elevation of the cellular activity (16). The apparent absence of mSiglec-F masking by *cis*-ligands conforms to this hypothesis. We are currently pursuing the functional aspects of mSiglec-F, especially the importance of sialic acid recognition by this molecule, via a targeted exon-replacement approach in mice.

Although we consider our genomic sequence-based analysis of human and mouse Siglecs “comprehensive,” there are several caveats. First, we are assuming the human and mouse genomic data bases we have probed are complete. Second, we have noted a genomic DNA segment (~1.2 kb) that shows >99% sequence identity to a 5'-region of the hSiglec-5 gene, containing coding sequences for the signal peptide and Ig-like domains 1 and 2 (indicated as a *triangle* labeled “5*” in Fig. 6). This segment is upstream of the true hSiglec-5 gene, is found in both the public and Celera human genome data bases, and could be confirmed by PCR analysis.⁸ Given the extremely high sequence identity to the Siglec-5 genomic region, this DNA segment is likely to be a relic of a very recent partial gene duplication. However, due to the lack of coding sequences for the downstream components (Ig-like domains, transmembrane domain, or cytosolic tail), we cannot be certain if this genomic DNA segment yields a functional mRNA or translated polypeptide. On the other hand, we cannot confidently assign it as a pseudogene. Third, we have noted a GenBankTM entry by the Human Genome Project derived from an unfinished human

⁷ T. Angata and A. Varki, unpublished.

⁸ T. Angata and A. Varki, unpublished.

BAC clone sequence (accession number AC026638, clone RP11-423F16) that contains several putative Siglec genes and pseudogenes. Although it is assigned to chromosome 8, we have found that many fragments (contigs) of this clone show extremely high sequence identity (>99.5%) to another unfinished BAC clone CTD-3187F8 (GenBank™ accession number AC063977), which encompasses a part of the Siglec-3-related gene cluster on chromosome 19. Given the existence of so-called “segmental duplication” (34), or “paralogous genome evolution” (39) in which long genomic DNA segments (up to 200 kb) duplicate in as yet unknown mechanism, it is possible that the chromosome 8 region covered by the clone RP11-423F16 actually represents another Siglec-3-related gene cluster, which was duplicated very recently. However, evidence against this possibility is provided by fluorescence *in situ* hybridization studies of Siglec-3/CD33 (40) and Siglec-7 (41). In both cases hybridization signals were localized exclusively on chromosome 19 (cytological band 19q13.3–13.4). If clone RP11-423F16 was indeed derived from chromosome 8, the DNA sequences nearly identical (>99%) to these genes in the chromosome 8 region represented by this clone should have shown an additional signal in the fluorescence *in situ* hybridization analysis. Also, there is only one unique Siglec-3-related gene cluster in the Celera Genomics human genome data base. Overall, it seems more likely that this is a technical artifact, such as misidentification of a clone or cross-contamination by an adjacent BAC clone, among other possibilities. Finally, there is one more unique Siglec-like putative gene in the human and mouse genomes that has not been studied. This hypothetical molecule was identified by a homology search of genomic and EST data bases, sequencing of a mouse EST clone, and by mutual subtraction of mouse and human genomic DNA sequences.⁹ However, the number and positions of cysteine residues in the first Ig-like domain of this molecule is different from those of typical Siglecs, and one of the cysteines is situated immediately adjacent to the “conserved arginine” residue involved in sialic acid recognition by known Siglecs. Thus, it is questionable if this hypothetical protein will recognize sialic acids. Nevertheless, this gene, which is highly conserved between human and mouse (~85% nucleotide identity), is likely to be of interest in deciphering the relationship of Siglec family to other Ig-like molecules and perhaps the origin of Siglec gene family.

Acknowledgments—We thank Els Brinkman-Van der Linden and Pascal Gagneux for general advice and helpful discussions, and Hui-Ling Han for technical assistance. Some of the mouse genomic data used for this report were generated through the use of the Celera Discovery System, under a Client agreement with the University of California System.

REFERENCES

- 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
- Kelm, S., and Schauer, R. (1997) *Int. Rev. Cytol.* **175**, 137–240
- Crocker, P. R., and Varki, A. (2001) *Trends Immunol.* **22**, 337–342
- Crocker, P. R., and Varki, A. (2001) *Immunology* **103**, 137–145
- Munday, J., Kerr, S., Ni, J., Cornish, A. L., Zhang, J. Q., Nicoll, G., Floyd, H., Mattei, M. G., Moore, P., Liu, D., and Crocker, P. R. (2001) *Biochem. J.* **355**, 489–497
- Angata, T., Varki, N. M., and Varki, A. (2001) *J. Biol. Chem.* **276**, 40282–40287
- Yu, Z., Lai, C.-M., Maoui, M., Banville, D., and Shen, S.-H. (2001) *J. Biol. Chem.* **276**, 23816–23824
- 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
- O’Keefe, T. L., Williams, G. T., Davies, S. L., and Neuberger, M. S. (1996) *Science* **274**, 798–801
- 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
- Nitschke, L., Carsetti, R., Ocker, B., Köhler, G., and Lamers, M. C. (1997) *Curr. Biol.* **7**, 133–143
- Martini, R., and Schachner, M. (1997) *Glia* **19**, 298–310
- Tchilian, E. Z., Beverley, P. C., Young, B. D., and Watt, S. M. (1994) *Blood* **83**, 3188–3198
- Yu, Z. B., Maoui, M., Wu, L. T., Banville, D., and Shen, S. H. (2001) *Biochem. J.* **353**, 483–492
- Ulyanova, T., Shah, D. D., and Thomas, M. L. (2001) *J. Biol. Chem.* **276**, 14451–14458
- Angata, T., and Varki, A. (2000) *J. Biol. Chem.* **275**, 22127–22135
- Dehal, P., Predki, P., Olsen, A. S., Kobayashi, A., Folta, P., Lucas, S., Land, M., Terry, A., Zhou, C. L. E., Rash, S., Zhang, Q., Gordon, L., Kim, J., Elkin, C., Pollard, M. J., Richardson, P., Rokhsar, D., Uberbacher, E., Hawkins, T., Branscomb, E., and Stubbs, L. (2001) *Science* **293**, 104–111
- Angata, T., and Varki, A. (2000) *Glycobiology* **10**, 431–438
- 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
- 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
- Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425
- Li, W.-H. (1997) *Molecular Evolution*, Sinauer Associates, Sunderland, MA
- Page, R. D. M., and Holmes, E. C. (1998) *Molecular Evolution: A Phylogenetic Approach*, Blackwell Science, Oxford, UK
- May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R., and Jones, E. Y. (1998) *Mol. Cell* **1**, 719–728
- 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
- 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
- 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
- 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
- Brinkman-Van der Linden, E. C. M., and Varki, A. (2000) *J. Biol. Chem.* **275**, 8625–8632
- Biermann, H., Pietz, B., Dreier, R., Schmid, K. W., Sorg, C., and Sunderkotter, C. (1999) *J. Leukoc. Biol.* **65**, 217–231
- Kim, J., Gordon, L., Dehal, P., Badri, H., Christensen, M., Groza, M., Ha, C., Hammond, S., Vargas, M., Wehri, E., Wagner, M., Olsen, A., and Stubbs, L. (2001) *Genomics* **74**, 129–141
- Kumar, S., and Hedges, S. B. (1998) *Nature* **392**, 917–920
- 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
- Lander, E. S., et al. (2001) *Nature* **409**, 860–921
- Ulyanova, T., Blasioli, J., Woodford-Thomas, T. A., and Thomas, M. L. (1999) *Eur. J. Immunol.* **29**, 3440–3449
- Vitale, C., Romagnani, C., Falco, M., Ponte, M., Vitale, M., Moretta, A., Bacigalupo, A., Moretta, L., and Mingari, M. C. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 15091–15096
- Wilkins, P. P., Moore, K. L., McEver, R. P., and Cummings, R. D. (1995) *J. Biol. Chem.* **270**, 22677–22680
- Long, E. O. (1999) *Annu. Rev. Immunol.* **17**, 875–904
- Eichler, E. E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N. A., Moyzis, R. K., Baldini, A., Gibbs, R. A., and Nelson, D. L. (1996) *Hum. Mol. Genet.* **5**, 899–912
- Peiper, S. C., Ashmun, R. A., and Look, A. T. (1988) *Blood* **72**, 314–321
- 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
- Fujita, N., Sato, S., Kurihara, T., Kuwano, R., Sakimura, K., Inuzuka, T., Takahashi, Y., and Miyatake, T. (1989) *Biochem. Biophys. Res. Commun.* **165**, 1162–1169

⁹ T. Angata and A. Varki, unpublished.