Genomics **67,** 171–178 (2000)
doi:10.1006/geno.2000.6208, available online at http://www.idealibrary.com on IDEAL®

Identification and Molecular Characterization of a Novel Member of the Siglec Family (SIGLEC9)

George Foussias, George M. Yousef, and Eleftherios P. Diamandis[1]

Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1L5, Canada

Received January 12, 2000; accepted March 20, 2000

Using the positional cloning approach, we have identified siglec-9 (HGMW-approved symbol SIGLEC9) a novel member of the sialic acid-binding Ig-like lectin (Siglec) family, which belongs to the immunoglobulin superfamily (IgSF). We characterized the genomic structure of this gene and determined its chromosomal localization, its homology to other members of the siglec family, and its tissue expression profile. The siglec-9 gene is composed of seven exons, with six intervening introns. The coding region consists of 1392 nucleotides and produces a 463-amino-acid protein. Furthermore, we have localized this gene to 19q13.4, 43.19 kb more telomeric than KLK14 (a member of the kallikrein gene family) through genomic sequencing data and restriction mapping with *Eco*RI. This novel siglec shows a high degree of homology to many members of the siglec family, including siglec-7 (80%), siglec-8 (72%), siglec-5 (65%), and CD33 (64%). This high degree of homology is also conserved in the extracellular Ig-like domains. Through RT-PCR, we have examined the expression of siglec-9 in a large number of tissues and have found relatively high-level expression in bone marrow, placenta, spleen, and fetal liver. Based on its homology to CD33, we speculate that this gene may also have some utility as a target for immunological antineoplastic therapy.

INTRODUCTION

The immunoglobulin superfamily (IgSF)[2] encompasses a large number of cell surface molecules that play a vital role not only in immunity, but also in controlling the behavior of cells in various tissues, through their ability to mediate cell surface recognition events. These molecules are characterized by the presence of at least one immunoglobulin (Ig) domain, a sandwich of two β-sheets stabilized by a conserved disulfide bond. The core of this domain is composed of β-strands A, B, and E in one sheet and G, F, and C in the other and arises from the ends of the domain sequence (Williams and Barclay, 1988). In between, however, there is a great deal of sequence length variation. Such Ig domains occur in two types, the V set and the C set, and can be distinguished based on patterns of conserved amino acid residues responsible for forming the characteristic β-sheet sandwich. V-set domains consist of about 65–75 amino acid residues between conserved cysteines, whereas C-set domains have about 55–60 residues (reviewed in Williams and Barclay, 1988). The C-set domains can be further divided into C1 and C2 sets and are distinguished by the fact that, although showing signs of a C-set domain, the latter halves of the C2-set domains exhibit sequence patterns more homologous to V-set domains rather than C1-set domains (Williams *et al.*, 1989).

Recently, a novel family of structurally related IgSF molecules were identified, which mediate protein–carbohydrate interactions through specific interactions with sialic acid-containing glycoproteins and glycolipids (Crocker *et al.*, 1996). This family was originally referred to as the sialoadhesins, but has recently been designated the sialic acid-binding Ig-like lectin (Siglec) family (Crocker *et al.*, 1998). These molecules are characterized by the presence of one N-terminal V-set domain and a variable number of downstream C2 set domains, ranging from 16 in sialoadhesin to 1 in CD33 (Crocker *et al.*, 1996). Furthermore, these Ig-like domains possess some unique features. In the V-set domain, the conserved cysteine in β-strand F of classic V-set domains is absent, while a highly conserved cysteine is present in β-strand E in all siglecs identified thus far. This results in the cysteines in β-strands B and E being next to each other in one β-sheet, which likely results in an intrasheet disulfide bond (Crocker *et al.*, 1996; Williams *et al.*, 1989). There is also an

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF135027.

[1] To whom correspondence and reprint requests should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Telephone: (416) 586-8443. Fax: (416) 586-8628. E-mail: ediamandis@mtsinai.on.ca.

[2] Abbreviations used: Siglec, sialic acid-binding immunoglobulin-like lectin; OB-BPL, OB-binding protein-like; KLK, kallikrein; RT-PCR, reverse transcription-polymerase chain reaction; IgSF, immunoglobulin superfamily; EST, expressed sequence tag; ITIM, immunoreceptor tyrosine kinase inhibitory motif; SLAM, signaling lymphocyte activation molecule; SAP, SLAM-associated protein; SH2, src homology 2; AML, acute myelogenous leukemia.

0888-7543/00 $35.00

additional highly conserved cysteine residue in both the V-set and the first C2-set domains of all siglecs. In the V-set domain, it is located at the beginning of β-strand B, while in the C2-set domain it is found between β-strands B and C. These two additional cysteines have been found to form an interdomain disulfide bond, a feature unique to siglecs (Crocker *et al.*, 1996; Pedraza *et al.*, 1990).

Currently, the siglec family consists of sialoadhesin (Siglec-1), CD22 (Siglec-2), CD33 (Siglec-3), myelin-associated glycoprotein (MAG) (Siglec-4a), Schwann cell myelin protein (SMP) (Siglec-4b), OB-binding protein 2 (Siglec-5), OB-binding protein 1 (Siglec-6), p75/AIRM1 (Siglec-7), and Siglec-8 (Cornish *et al.*, 1998; Crocker *et al.*, 1998; Falco *et al.*, 1999; Floyd *et al.*, 2000; Nicoll *et al.*, 1999; Patel *et al.*, 1999). Each member of the Siglec family is expressed by specific cell types and exhibits a distinct function. Sialoadhesin is a macrophage-restricted adhesion molecule (Crocker *et al.*, 1994), CD22 is B lymphocyte-specific and regulates its activation (Stamenkovic and Seed, 1990), CD33 is a myeloid-specific inhibitory receptor (Ulyanova *et al.*, 1999), and MAG functions in the formation and maintenance of axonal myelin structure (Li *et al.*, 1998). Siglec-5 and -6 (OB-BP2 and -BP1, respectively) are expressed in several tissues including placenta and peripheral blood leukocytes and have shown an *in vitro* ability to bind leptin (Cornish *et al.*, 1998; Patel *et al.*, 1999), siglec-7 (p75/AIRM1) is an inhibitory receptor expressed predominantly on human natural killer cells (Falco *et al.*, 1999; Nicoll *et al.*, 1999), and siglec-8 is specifically expressed by eosinophils (Floyd *et al.*, 2000).

In this paper we describe the identification and molecular characterization of the gene encoding siglec-9,[3] a novel member of the siglec family (OB-binding protein-like, GenBank Accession No. AF135027). The putative protein product displays a high degree of homology with the recently identified siglec-7 and siglec-8, as well as with siglec-5 and siglec-6. Further, it possesses all the structural features found in other siglecs, mentioned previously. We have characterized the genomic structure of this novel siglec, finding that it is encoded by seven exons. In addition, we have examined its tissue expression through RT-PCR, which revealed high-level expression in bone marrow, placenta, spleen, and fetal liver, as well as expression in other tissues at lower levels.

MATERIALS AND METHODS

New gene identification. Nucleotide sequencing data for approximately 130 kb on chromosome 19q13.4 were obtained from the Lawrence Livermore National Laboratory (LLNL) Web site (http://www-bio.llnl.gov/genome/genome.html), in the form of one contig (clone BC349142 contig 56). This genomic sequence was subjected to a number of computer algorithms (gene prediction programs) designed to predict the presence of putative new genes. All programs used were previously thoroughly evaluated in our laboratory using a large number of known genes (Yousef *et al.*, 1999a). Based on these results, we selected the most reliable algorithms—GeneBuilder (gene prediction) (http://125.itba.mi.cnr.it/~webgene/genebuilder.html) and GeneBuilder (exon prediction) (http://125.itba.mi.cnr.it/~webgene/genebuilder.html); Grail 2 (http://compbio.ornl.gov); and GENEID-3 (http://apolo.imim.es/geneid.html)—for further use.

Expressed sequence tag (EST) identification. The genomic sequence of the putative new gene was subjected to a homology search against the human EST database using the BLASTN algorithm (http://www.ncbi.nlm.nih.gov/BLAST) (Altschul *et al.*, 1997). Clones showing >95% homology were obtained from the IMAGE Consortium through Research Genetics Inc. (Huntsville, AL). The clone obtained was then propagated according to the supplier's instructions, purified, and sequenced from both directions with an automated sequencer, using the insert-flanking vector primers T3 and T7.

Molecular characterization of siglec-9. The sequence derived from the computer-predicted exons of our putative new gene was also used to search the nonredundant protein sequence database, using the BLASTP algorithm (Altschul *et al.*, 1997). Several proteins showing a high degree of homology were selected, and their nucleotide coding sequences were aligned with our predicted coding sequence using the ClustalX multiple alignment program (Jeanmougin *et al.*, 1998). From this, we selected regions on our putative gene that showed the least amount of homology to the others and designed PCR primers F1 (TCACCGGCTCTCTGTGAATG) and R1 (GTCTTCTGCCCAAGGTTCAG). Using these primers, we performed PCR on bone marrow cDNA, prepared as discussed below and chosen based on the tissue expression results. The PCR conditions were as follows: 2.5 units HotStar *Taq* polymerase (Qiagen, Valencia, CA), 1× PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 1 μl cDNA, 200 μM dNTPs (deoxynucleoside triphosphates), and 250 ng of primers, using the Mastercycler gradient thermocycler (Eppendorf Scientific, Inc., Westbury, NY). The temperature profile was denaturation at 95°C for 15 min followed by 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 1 min for a total of 35 cycles, followed by a final extension at 72°C for 10 min. The PCR product was subjected to electropheresis on a 2% agarose gel and stained with ethidium bromide. Aliquots of the PCR products were subsequently extracted from the gel, and the purified DNA was directly sequenced using an automated sequencer.

To verify the sequence surrounding the proposed start codon, another set of primers were designed, again derived from regions showing a low level of homology with other known genes: F3 (TCCTCTAAGTCTTGAGCCCG) and R3 (CAGACGTTGAGATGGACGGT). PCR was performed using bone marrow cDNA, prepared as described below. The conditions used for the PCR were identical to those discussed previously, with electrophoresis of the PCR product on a 2% agarose gel, gel extraction, and automated sequencing as before.

Following final characterization of the genomic structure of siglec-9, the putative protein product was aligned with the protein sequences of the other siglec family members using the ClustalX multiple sequence alignment tool. Further, phylogenetic analysis was performed using ClustalX in combination with TreeView (Page, 1996).

Our putative protein was assessed for the presence of a possible signal peptide, using SignalP v1.1 (http://www.cbs.dtu.uk/) (Nielsen *et al.*, 1997). Further, for the prediction of transmembrane domains, two independent algorithms were used, TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) and DAS (http://www.biokemi.su.se/~server/). In addition, the hydropathic profile of this novel siglec was determined, using the Kyte–Doolittle method (http://bioinformatics.weizmann.ac.il/hydroph/plot_hydroph.html).

Mapping and chromosomal localization of siglec-9. As mentioned previously, the contig on which the siglec-9 gene was identified was obtained from the LLNL. *Eco*RI restriction maps were obtained from the LLNL and were also generated using the Webcutter restriction analysis tool (http://www.firstmarker.com/cutter/cutter2.html), for both this contig and the adjacent, more centromeric contigs, contain-

[3] The HGMW-approved symbol for the gene described in this paper is SIGLEC9.

TABLE 1

Genomic Organization of Siglec-9

Exon no.	Coding region[a] From (bp)	Coding region[a] To (bp)	No. of basepairs	EST match[b]	Intron phase	Exon predicted[c]
1	1083	1591	509	−	I	B, C
2	1793	2071	279	−	I	A, B, C, D
3	2277	2324	48	−	I	A, B, D
4	3226	3492	267	−	I	A, B, C
5	4145	4235	91	−	0	A, B, C, D
6	4610	4706	97	+	0	—
7	6087	6503	417	+	—	A, B, C

[a] The coding region shown includes the 5′ untranslated region in exon 1 and the 3′ untranslated region in exon 7. Numbers refer to GenBank Accession No. AF135027.

[b] EST; GenBank Accession No. AA936059.

[c] The exon prediction programs are as follows: (A) GeneBuilder (gene prediction); (B) GeneBuilder (exon prediction); (C) Grail 2; (D) GENEID-3.

ing the recently identified kallikrein gene family (Diamandis *et al.*, 2000; Yousef *et al.*, 1999a). Overlapping restriction fragments were identified and used to order the contigs and determine the distance between KLK14, the most telomeric member of the kallikrein gene family, and this novel siglec.

Tissue expression. Total RNA from 28 normal human tissues was obtained (Clontech, Palo Alto, CA), and reverse transcription was performed using SuperScript II, according to the manufacturer's instructions (Gibco BRL, Gaithersburg, MD). PCR was then performed using primers F2 (CGTGGGAGATACGGGCATAG) and R2 (AAAAGGGAGGGCACAGTGTG), using the same PCR conditions described previously. PCR for actin was also performed as described elsewhere (Yousef *et al.*, 1999b), as a control for cDNA quality.

RESULTS

Identification of Siglec-9 on 19q13.4

Computer analysis of the approximately 130-kb contig predicted a putative new gene consisting of six exons. Five of these were predicted by at least three programs, with only one exon being predicted by two of the four programs (Table 1). A homology search for the putative new gene against the human EST database revealed the presence of one unique EST (GenBank Accession No. AA936059), which showed 98% identity to the sixth predicted exon.

We proceeded to sequence the entire insert of this EST, followed by alignment of this nucleotide sequence with the genomic sequence of our putative gene, using the "BLAST 2 sequences" program. This revealed the presence of an additional area, between predicted exons 5 and 6, with 98% identity to the EST. This suggested that there was an additional exon in this area that was not detected by the prediction algorithms used.

Characterization of the Genomic Structure of the Siglec-9 Gene and Its Protein Product

Based on the existence of an EST almost identical to part of our putative gene, we postulated that this, in fact, is a novel gene. With the aid of unique primers, designed as discussed under Materials and Methods, we were able to perform RT-PCR on bone marrow cDNA and isolate two additional products, both encompassing multiple predicted exons. Upon sequencing of these PCR products, we confirmed the presence of all six predicted exons, as well as the newly identified exon, found from the EST sequence. With both cDNA and genomic sequence at hand, we determined the genomic organization of this new gene (Fig. 1). We found that the gene encoding this novel siglec encompasses a genomic area of 5421 bp. It is composed of seven exons, with six intervening introns. The lengths of the exons are 509, 279, 48, 267, 91, 97, and 417 bp. All the intron/exon splice sites and their flanking sequences are closely related to the consensus splice sites (-m**GT**AAGT. . .C**AG**m-, where m is any base) (Iida, 1990).

The proposed protein coding region of the siglec-9 gene consists of 1392 nucleotides, producing a 463-amino-acid protein, with a predicted molecular mass of 50.1 kDa, excluding any posttranslational modifications. The translation initiation codon (ATG) at position 1171 of the first exon (according to the numbering of our GenBank Accession No. AF135027) was chosen because: (1) the flanking region surrounding that codon closely matches the Kozak consensus sequence for translational initiation, particularly at position −3 (a purine), which appears to be the most highly conserved (Kozak, 1991), and (2) using this initiation codon, the proposed protein contains an N-terminal signal sequence that shows a high degree of homology to other similar proteins (see below). The 3′ terminus of the siglec-9 gene was verified by the presence of a poly(dA) tail present in the EST sequence. Further, the coding sequence of this gene possesses a 5′ untranslated region of at least 88 nucleotides, as well as a 3′ untranslated region of 228 nucleotides.

Examination of the hydrophobicity profile of the putative siglec-9 protein revealed two regions with long stretches of hydrophobic residues. The first of these occurs at the N-terminus, suggesting the presence of a signal peptide, and is consistent with findings from a

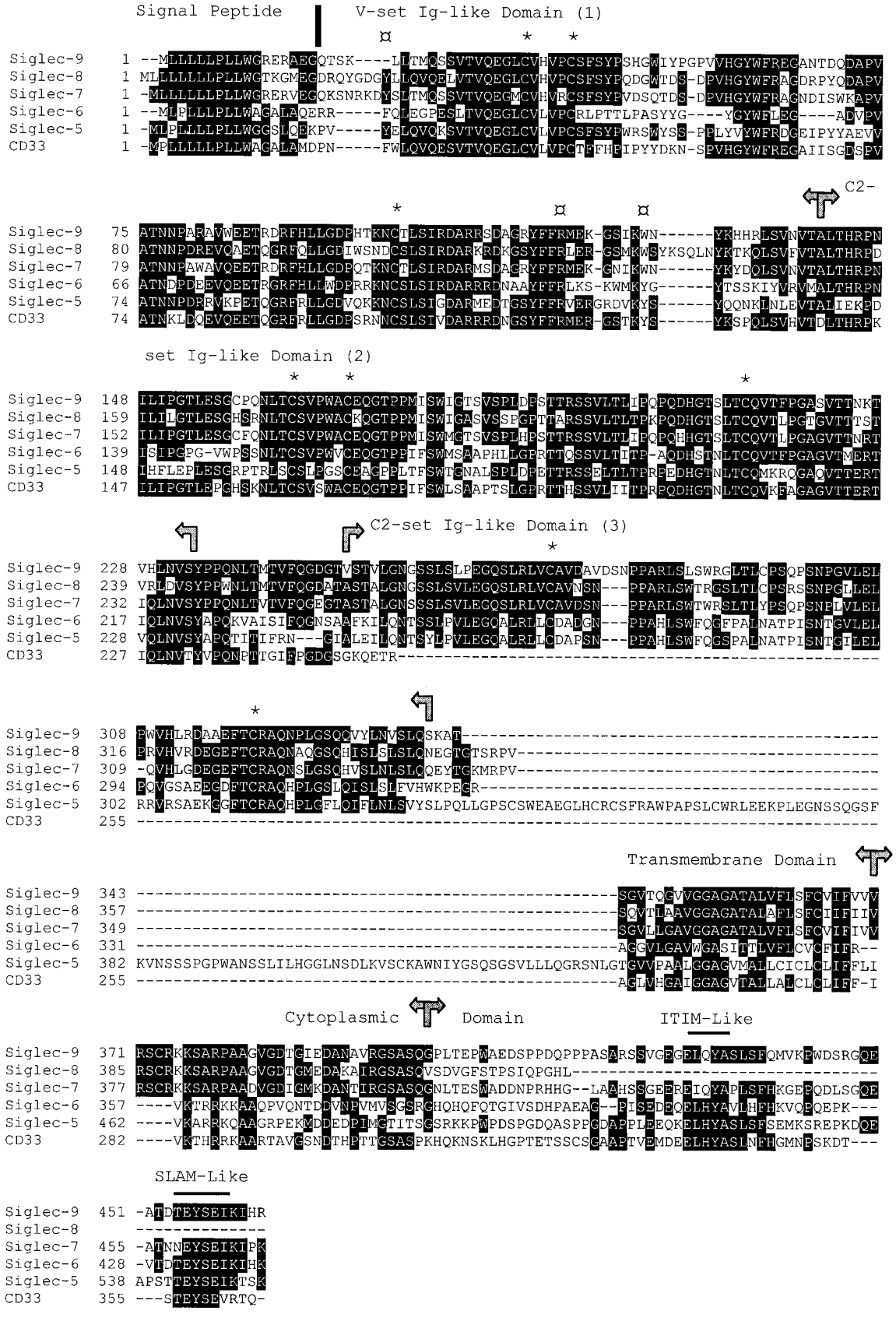

FIG. 1. Siglec family multiple alignment. Siglec-9 was aligned with siglec-5 to -8 and CD33, using ClustalX (Jeanmougin *et al.*, 1998). The signal peptide was determined through computer prediction, and the Ig domain boundaries were assigned based on exon boundaries (shown with bent arrows). The transmembrane domain was also predicted, while taking into consideration exon boundaries as well. The ITIM-like and SLAM-like motifs are indicated, as are the conserved cysteines (*) that form the disulfide bonds of the Ig-like domains in siglecs and the conserved arginine and aromatic residues (¤) that are responsible for sialic acid binding and specificity.

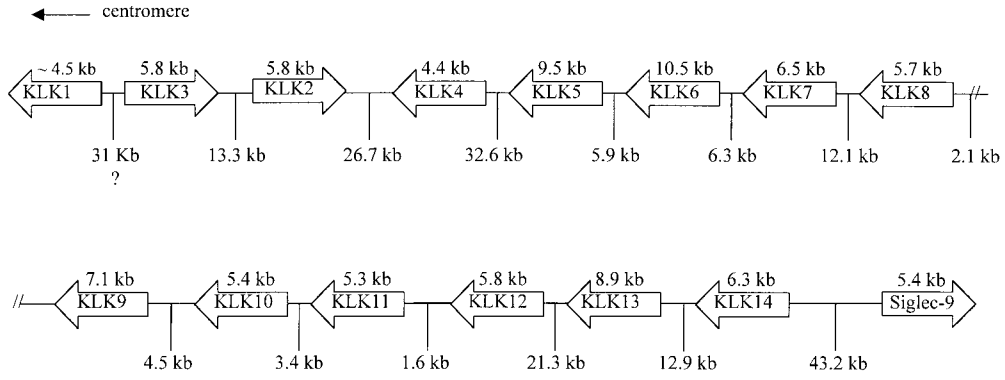

FIG. 2. Localization of the siglec-9 gene. The physical map of the 314-kb genomic area around chromosome 19q13.3–q13.4 where the kallikrein gene family resides. Gene lengths are presented above each arrow, and distances between genes are shown below. Arrows denote the direction of transcription. The siglec-9 gene resides 43.2 kb telomeric to the KLK-L6 gene. For more details on this genomic region, see Diamandis *et al.* (2000). KLK, kallikrein.

signal sequence prediction program (Nielsen *et al.*, 1997), which predicts a 17-amino-acid residue signal sequence. The second region occurs between residues 349 and 370, suggestive of a transmembrane domain, and is consistent with results from transmembrane region prediction programs. Based on this information, we postulate that the protein product of this novel gene is a type I transmembrane protein, after cleavage of the 17-residue signal sequence.

Mapping and Chromosomal Localization of Siglec-9

The contig in which we identified the gene encoding siglec-9 is located at 19q13.4, telomeric to the kallikrein gene KLK3 (PSA). Previous studies in our laboratory have identified and mapped the kallikrein gene family locus on this region of chromosome 19 (Diamandis *et al.*, 2000; Yousef *et al.*, 1999a). The contig containing the novel siglec gene was found, through *Eco*RI restriction mapping, to be located adjacent to this kallikrein gene family. The siglec-9 gene is located 43.19 kb more telomeric than KLK14, at 19q13.4. A detailed physical map of the area, which contains some known genes and the newly identified siglec gene, is shown in Fig. 2. By computer analysis, we did not predict any other genes between KLK14 and this novel siglec.

Homology with Other Siglec Family Members

Using the predicted protein sequence, we performed a homology search against the GenBank database using the BLASTP program. Siglec-9 showed a high degree of homology to other known members of the Siglec family (Table 2). We further performed a multiple alignment of siglec-9 with the other family members, using the ClustalX alignment program. As is evident in Fig. 1, the N-terminal signal sequence is highly conserved within this family of proteins. Furthermore, our protein contains Ig domains typically found in Siglec family members: an N-terminal V-set domain, followed by multiple C2-set domains (Crocker *et al.*, 1996). This novel siglec contains a total of three Ig domains, a V-set domain and two C2-set domains, based on homology with known Ig domains. As shown in Table 3, the V-set domain and both C2-set domains are highly similar to siglec-7 and -8. Siglec-9 exhibits conservation of the cysteine residues in the V-set domain and first C2-set domain, which form the two characteristic disulfide bridges in other Siglec family members. The V-set domain also possesses a conserved arginine that has been found to be essential for sialic acid binding (van der Merwe *et al.*, 1996), as well as two conserved aromatic residues in β-strands A and G that have been found to make hydrophobic contacts with the *N*-acetyl and glycerol side groups of *N*-acetyl neuraminic acid (May *et al.*, 1998). As is evident from Fig. 1, siglec-9 also possesses this critical arginine, as well as one of the aromatic residues. The domain boundaries were determined based on the one domain:one exon rule (Williams and Barclay, 1988), while taking into consideration the domain assignments of others (Cornish *et al.*, 1998; Crocker *et al.*, 1998; Falco *et al.*, 1999; Nicoll *et al.*, 1999; Patel *et al.*, 1999).

Examination of the transmembrane and intracellular domains of Siglec family members reveals that it is more variable than the extracellular domain. However, there are regions that show a high level of conservation. As shown in Fig. 1, all the Siglecs possess a single

TABLE 2

Overall Homology of Siglec-9 with Other Known Siglecs

Siglec family member[a]	Homology to siglec-9[b]	
	% identity	% similarity
Siglec-7 (p75/AIRM1) (AF170485)	75	80
Siglec-8 (AF195092)	65	72
Siglec-5 (OB-BP2) (U71383)	52	65
CD33 (M23197)	52	64
Siglec-6 (OB-BP1) (U71382)	49	60
Sialoadhesin (Z36293)	27	43
CD22 (X52785)	26	42
Myelin-associated glycoprotein (MAG) (M29273)	25	42

[a] GenBank accession numbers for each of the siglec family members are also shown, in parentheses.

[b] Homology was determined using the BLASTP algorithm.

TABLE 3

Ig-like Domain Homology between Siglec-9 and Other Siglec Family Members[a]

	Homologous protein	Domain	% identity	% similarity
Siglec-9 Ig 1 (V set)	Siglec-7 (p75/AIRM1)	1	75	78
	Siglec-8	1	63	71
	CD33	1	61	71
	Siglec-5 (OB-BP2)	1	54	67
	Siglec-6 (OB-BP1)	1	54	62
	MAG	1	32	48
	Sialoadhesin	1	29	48
	CD22	1	28	44
Siglec-9 Ig 2 (C2 set)	Siglec-7 (p75/AIRM1)	2	89	93
	Siglec-8	2	77	83
	CD33	2	63	75
	Siglec-6 (OB-BP1)	2	58	70
	Siglec-5 (OB-BP2)	2	58	71
	Sialoadhesin	2	30	46
		12	31	44
	MAG	2	25	46
	CD22	2	27	43
Siglec-9 Ig 3 (C2 set)	Siglec-7 (p75/AIRM1)	3	76	79
	Siglec-8	3	70	80
	Siglec-6 (OB-BP1)	3	52	67
	Siglec-5 (OB-BP2)	3	48	62
	Sialoadhesin	13	33	48
		7	31	42
		15	28	40
	MAG	3	27	49

[a] GenBank accession numbers for the siglecs listed are the same as those shown in Table 2.

transmembrane domain, consisting of approximately 25 residues. In addition, within the cytoplasmic domain, we notice the presence of two highly conserved motifs. The first of these, L(HQ)YA(SV)L, exhibits similarity to an immunoreceptor tyrosine kinase inhibitory motif (ITIM), which has a 6-amino-acid consensus sequence (ILV)xYxx(LV) (Burshtyn *et al.*, 1997; Vivier and Daeron, 1997). The second motif, TEYSE(IV), is homologous to a sequence (TxYxx(IV)) recently found in the signaling lymphocyte activation molecule (SLAM) that is responsible for the binding of the SLAM-associated protein (SAP) (Coffey *et al.*, 1998; Sayos *et al.*, 1998).

Phylogenetic analysis of the entire siglec family was performed using ClustalX and TreeView. This revealed that siglec-9 is very closely related to siglec-7, followed by siglec-8 (Fig. 3). It is evident that this novel gene, which encodes a putative siglec protein, is the newest member of the siglec family. It possesses all the necessary features, including the Ig-like domains, the type I transmembrane topology, and the conserved cytoplasmic motifs, and shows a close phylogenetic relationship to the other siglec family members.

Tissue Expression Profile of Siglec-9

RT-PCR was performed on a panel of tissue-specific total RNA preparations (Fig. 4). We have found high levels of expression of siglec-9 in bone marrow, placenta, spleen, and fetal liver. Lower levels of expression were also evident in fetal brain, stomach, lung, thymus, prostate, brain, mammary, adrenal gland, colon, trachea, cerebellum, testis, small intestine, and spinal cord. Expression of siglec-9 was absent in heart,

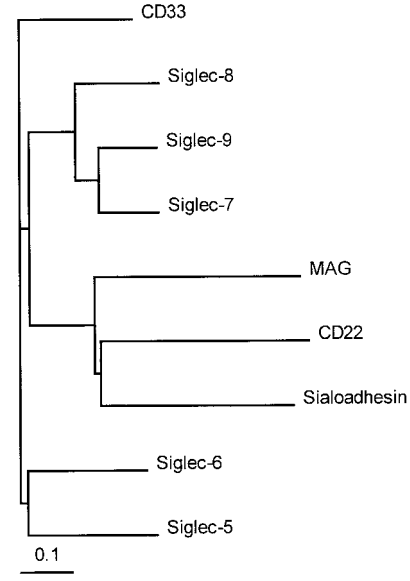

FIG. 3. Phylogenetic analysis of the siglec family. The phylogenetic tree was created using ClustalX (Jeanmougin *et al.*, 1998) and TreeView (Page, 1996). The scale bar shown indicates the number of substitutions per site. As is evident, siglec-7 and siglec-9 are very closely related, and they are both related to siglec-8, in addition to a more distant relation to the other siglecs.

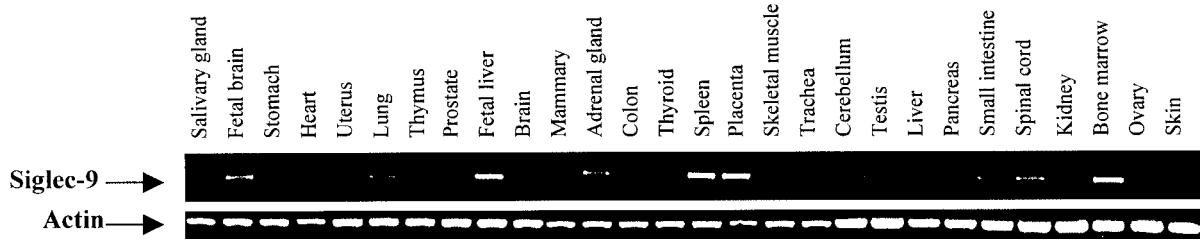

FIG. 4. Tissue expression profile of siglec-9. RT-PCR was performed on 28 tissue total RNAs, for siglec-9 and actin (control gene). Siglec-9 is highly expressed in bone marrow, placenta, spleen, and fetal liver. There is also a lower degree of expression in many of the other tissues, while it is absent in ovary, pancreas, skeletal muscle, and heart.

skeletal muscle, pancreas, and ovary. All PCR products obtained were of equal length and corresponded to the length of the product obtained from an overlapping EST (Accession No. AA936059). Sequencing of the PCR products ensured specificity.

DISCUSSION

Using the positional cloning approach, we have identified siglec-9, a novel gene belonging to the siglec family. This gene is composed of seven exons, with six intervening introns. The coding region of this gene is composed of 1392 nucleotides, producing a 463-amino-acid protein, with a predicted molecular mass of 50.1 kDa. This gene is located at 19q13.4, 43.19 kb telomeric to the newly identified kallikrein KLK14. The high degree of homology between this novel siglec and other siglecs provides strong evidence that this protein also plays a role in sialic acid-dependent protein–glycoprotein or protein–glycolipid interactions. It possesses the unique pattern of conserved cysteine residues in its Ig-like domains, which are found only in members of the siglec family. Furthermore, siglec-9 possesses the conserved arginine residue, which has been found to be essential for sialic acid binding (van der Merwe *et al.*, 1996). Of note, however, is that it possesses only one of the two conserved aromatic residues in the V-set domain, which may be suggestive of a unique sialic acid specificity, differing from that of previously identified siglecs.

We examined the tissue expression profile of siglec-9 and found that it is highly expressed in bone marrow, placenta, spleen, and fetal liver. The high level of expression of this novel siglec in bone marrow, and tissues involved in stem cell differentiation, is consistent with findings from groups investigating the other siglec family members. All currently known siglecs have been found to be expressed in some type of bone marrow stem cell-derived cell, ranging from myeloid progenitor cells for CD33 and eosinophils for siglec-8, to natural killer cells for siglec-7 and B lymphocytes for CD22. We speculate that siglec-9 is predominantly expressed on a distinct subset of immune cells, where it plays an intercellular signaling role. This is supported by the presence of ITIM-like and SLAM-like motifs in the cytoplasmic domain of this novel siglec, with similar domains in other siglecs. ITIM motifs are consensus binding sites for the SH2 (src homology 2) domains of the phosphatases SHP-1 and SHP-2 (Borges *et al.*, 1997; Le Drean *et al.*, 1998). It has been reported that the phosphorylation of the ITIM-like motif in CD22 results in the recruitement of the phosphatase SHP-1, suggesting a possible function of this siglec as a B cell receptor-associated negative coreceptor (Vivier and Daeron, 1997). The second cytoplasmic motif has been identified in SLAM and several SLAM-like proteins, a family of immunoregulatory molecules of the IgSF, and is responsible for the binding of a new SH2-containing molecule, SAP (Coffey *et al.*, 1998; Sayos *et al.*, 1998). The binding of SAP was shown to inhibit the binding of SHP-2 to its respective binding site on these SLAM proteins. The presence of such a motif in siglec-9, and other siglecs, suggests that there may be a similar regulatory mechanism present in the cytoplasmic domains of siglecs, with SAP inhibiting the binding of SHP-1 and SHP-2 to the ITIM-like motif.

The regulation of SHP-1 and SHP-2 binding to ITIM motifs, and thus their activation, very likely affects downstream tyrosine-kinase-dependent pathways by regulating the phosphorylation state of components in these pathways (Burshtyn *et al.*, 1997). Thus, the siglec family of ITIM- and SLAM-bearing receptors probably plays a role in controlling the activation of a number of cell types. By extension, it is possible that these siglecs may be involved in the regulation of tumor growth. CD33 has already been identified as an important marker for the diagnosis of acute myelogenous leukemia (AML), particularly for the undifferentiated form, and serves to distinguish AML from lymphoid leukemias (Bernstein *et al.*, 1992; Dinndorf *et al.*, 1986; Griffin *et al.*, 1984). Recently, Kossman *et al.* (1999) and Sievers *et al.* (1999) have reported the use of anti-CD33 monoclonal antibodies in phase I studies for the treatment of AML and have shown selective ablation of malignant hematopoiesis. It is possible that this newly identified member of the siglec family may also have some utility as a target for immunological antineoplastic therapy.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Bernstein, I. D., Singer, J. W., Smith, F. O., Andrews, R. G., Flowers, D. A., Petersens, J., Steinmann, L., Najfeld, V., Savage, D., Frucht-

man, S., *et al.* (1992). Differences in the frequency of normal and clonal precursors of colony-forming cells in chronic myelogenous leukemia and acute myelogenous leukemia. *Blood* **79:** 1811–1816.

Borges, L., Hsu, M. L., Fanger, N., Kubin, M., and Cosman, D. (1997). A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. *J. Immunol.* **159:** 5192–5196.

Burshtyn, D. N., Yang, W., Yi, T., and Long, E. O. (1997). A novel phosphotyrosine motif with a critical amino acid at position −2 for the SH2 domain-mediated activation of the tyrosine phosphatase SHP-1. *J. Biol. Chem.* **272:** 13066–13072.

Coffey, A. J., Brooksbank, R. A., Brandau, O., Oohashi, T., Howell, G. R., Bye, J. M., Cahn, A. P., Durham, J., Heath, P., Wray, P., *et al.* (1998). Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene. *Nat. Genet.* **20:** 129–135. [See comments]

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). Characterization of siglec-5, a novel glycoprotein expressed on myeloid cells related to CD33. *Blood* **92:** 2123–2132.

Crocker, P. R., Clark, E. A., Filbin, M., Gordon, S., Jones, Y., Kehrl, J. H., Kelm, S., Le Douarin, N., Powell, L., Roder, J., *et al.* (1998). Siglecs: A family of sialic-acid binding lectins. *Glycobiology* **8:** v–vi. [Letter]

Crocker, P. R., Kelm, S., Hartnell, A., Freeman, S., Nath, D., Vinson, M., and Mucklow, S. (1996). Sialoadhesin and related cellular recognition molecules of the immunoglobulin superfamily. *Biochem. Soc. Trans.* **24:** 150–156.

Crocker, P. R., Mucklow, S., Bouckson, V., McWilliam, A., Willis, A. C., Gordon, S., Milon, G., Kelm, S., and Bradfield, P. (1994). Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. *EMBO J.* **13:** 4490–4503.

Diamandis, E. P., Yousef, G. M., Luo, L.-Y., Magklara, A., and Obiezu, C. V. (2000). The new human kallikrein gene family: Implications in carcinogenesis. *Trends Endocrinol. Metab.* **11:** 54–60.

Dinndorf, P. A., Andrews, R. G., Benjamin, D., Ridgway, D., Wolff, L., and Bernstein, I. D. (1986). Expression of normal myeloid-associated antigens by acute leukemia cells. *Blood* **67:** 1048–1053.

Falco, M., Biassoni, R., Bottino, C., Vitale, M., Sivori, S., Augugliaro, R., Moretta, L., and Moretta, A. (1999). Identification and molecular cloning of p75/AIRM1, a novel member of the sialoadhesin family that functions as an inhibitory receptor in human natural killer cells. *J. Exp. Med.* **190:** 793–802.

Floyd, H., Ni, J., Cornish, A. L., Zeng, Z., Liu, D., Carter, K. C., Steel, J., and Crocker, P. R. (2000). Siglec-8: A novel eosinophil-specific member of the immunoglobulin superfamily. *J. Biol. Chem.* **275:** 861–866.

Griffin, J. D., Linch, D., Sabbath, K., Larcom, P., and Schlossman, S. F. (1984). A monoclonal antibody reactive with normal and leukemic human myeloid progenitor cells. *Leukemia Res.* **8:** 521–534.

Iida, Y. (1990). Quantification analysis of 5′-splice signal sequences in mRNA precursors: Mutations in 5′-splice signal sequence of human beta-globin gene and beta-thalassemia. *J. Theor. Biol.* **145:** 523–533.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23:** 403–405.

Kossman, S. E., Scheinberg, D. A., Jurcic, J. G., Jimenez, J., and Caron, P. C. (1999). A phase I trial of humanized monoclonal antibody HuM195 (anti-CD33) with low-dose interleukin 2 in acute myelogenous leukemia. *Clin. Cancer Res.* **5:** 2748–2755.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell. Biol.* **115:** 887–903.

Le Drean, E., Vely, F., Olcese, L., Cambiaggi, A., Guia, S., Krystal, G., Gervois, N., Moretta, A., Jotereau, F., and Vivier, E. (1998). Inhibition of antigen-induced T cell response and antibody-induced NK cell cytotoxicity by NKG2A: Association of NKG2A with SHP-1 and SHP-2 protein-tyrosine phosphatases. *Eur. J. Immunol.* **28:** 264–276. [Published erratum appears in *Eur. J. Immunol.* 1998, Mar; **28**(3): 1122].

Li, C., Trapp, B., Ludwin, S., Peterson, A., and Roder, J. (1998). Myelin associated glycoprotein modulates glia–axon contact *in vivo*. *J. Neurosci. Res.* **51:** 210–217.

May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R., and Jones, E. Y. (1998). Crystal structure of the N-terminal domain of sialoadhesin in complex with 3′ sialyllactose at 1.85 Å resolution. *Mol. Cell* **1:** 719–728.

Nicoll, G., Ni, J., Liu, D., Klenerman, P., Munday, J., Dubock, S., Mattei, M. G., and Crocker, P. R. (1999). Identification and characterization of a novel siglec, siglec-7, expressed by human natural killer cells and monocytes. *J. Biol. Chem.* **274:** 34089–34095.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.* **8:** 581–599.

Page, R. D. (1996). TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12:** 357–358.

Patel, N., Brinkman-Van der Linden, E. C., 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). OB-BP1/Siglec-6. a leptin- and sialic acid-binding protein of the immunoglobulin superfamily. *J. Biol. Chem.* **274:** 22729–22738.

Pedraza, L., Owens, G. C., Green, L. A., and Salzer, J. L. (1990). The myelin-associated glycoproteins: Membrane disposition, evidence of a novel disulfide linkage between immunoglobulin-like domains, and posttranslational palmitylation. *J. Cell Biol.* **111:** 2651–2661.

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). The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395:** 462–469. [See comments]

Sievers, E. L., Appelbaum, F. R., Spielberger, R. T., Forman, S. J., Flowers, D., Smith, F. O., Shannon-Dorcy, K., Berger, M. S., and Bernstein, I. D. (1999). Selective ablation of acute myeloid leukemia using antibody-targeted chemotherapy: A phase I study of an anti-CD33 calicheamicin immunoconjugate. *Blood* **93:** 3678–3684.

Stamenkovic, I., and Seed, B. (1990). The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion. *Nature* **345:** 74–77.

Ulyanova, T., Blasioli, J., Woodford-Thomas, T. A., and Thomas, M. L. (1999). The sialoadhesin CD33 is a myeloid-specific inhibitory receptor. *Eur. J. Immunol.* **29:** 3440–3449.

van der Merwe, P. A., Crocker, P. R., Vinson, M., Barclay, A. N., Schauer, R., and Kelm, S. (1996). Localization of the putative sialic acid-binding site on the immunoglobulin superfamily cell-surface molecule CD22. *J. Biol. Chem.* **271:** 9273–9280.

Vivier, E., and Daeron, M. (1997). Immunoreceptor tyrosine-based inhibition motifs. *Immunol. Today* **18:** 286–291.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily—Domains for cell surface recognition. *Annu. Rev. Immunol.* **6:** 381–405.

Williams, A. F., Davis, S. J., He, Q., and Barclay, A. N. (1989). Structural diversity in domains of the immunoglobulin superfamily. *Cold Spring Harbor Symp. Quant. Biol.* **54:** 637–647.

Yousef, G. F., Luo, L., and Diamandis, E. P. (1999a). Identification of novel human kallikrein-like genes on chromosome 19q13.3–q13.4. *Anticancer Res.* **79:** 2843–2852.

Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H., and Diamandis, E. P. (1999b). Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated. *Cancer Res.* **59:** 4252–4256.