Molecular Characterization of a Siglec8 Variant Containing Cytoplasmic Tyrosine-Based Motifs, and Mapping of the Siglec8 Gene

George Foussias,*† George M. Yousef,*† and Eleftherios P. Diamandis*†,1
*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; and †Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

Received October 23, 2000

Through efforts to investigate the CD33-like subgroup of sialic acid binding immunoglobulin-like lectins (Siglecs), which are believed to be located on chromosome 19q13.4, we have identified the precise genomic region containing the Siglec8 gene. It is located on chromosome 19q13.4, approximately 330 kb downstream of the Siglec9 gene. Further, we have identified a novel Siglec8 variant, named Siglec8-Long (Siglec8-L), which differs in its last two exons from the previously published mRNA sequence of Siglec8 (GenBank Accession No. AF195092). Both Siglec8 and Siglec8-L are comprised of seven exons, of which the first five are identical, followed by marked differences in exon usage and mRNA splicing. The 499 amino acid protein encoded by the Siglec8-L open reading frame has a molecular weight of 54 kDa. Like the other members of the CD33-like subgroup of Siglecs, except for the previously published Siglec8, Siglec8-L also contains the two tyrosine-based motifs that have been found to recruit both SH2 domain-containing tyrosine and inositol phosphatases.

Key Words: siglec gene family; CD33-like subgroup; immunoglobulin superfamily; sialoadhesins; immunoreceptor tyrosine kinase inhibition motif (ITIM); alternative splicing; gene mapping.

Sialic acid-binding immunoglobulin-like lectins (Siglecs) are a family of recently discovered type 1 transmembrane proteins belonging to the immunoglobulin superfamily (IgSF) (1). In addition to their type 1 transmembrane topology, these proteins are characterized by the presence of one N-terminal V-set Ig-like domain, a variable number of downstream C2-set domains, and the ability to bind sialic acid in glycoproteins and glycolipids (2). So far there have been nine Siglec family members described in humans, each with its own unique expression pattern: Siglec1 (sialoadhesin) expressed on macrophages (3); Siglec2 (CD22) on B lymphocytes (4); Siglec3 (CD33) on myeloid progenitor cells and monocytes (5); Siglec4a (myelin-associated glycoprotein (MAG)) on oligodendrocytes and Schwann cells (6); Siglec5 on neutrophils (7); Siglec6 on B lymphocytes (8); Siglec7 (p75/AIRM1) on natural killer cells (9–11); Siglec8 on eosinophils (12); and Siglec9 on neutrophils, monocytes, and various lymphocytes (13–15).

The members of the Siglec family are highly homologous in their extracellular, Ig-like, domains, particularly for the CD33-like subgroup, which includes Siglec3, 5, 6, 7, 8, and 9 (13, 15). This subgroup of genes has been localized to human chromosome 19q13.3-13.4, and it has been suggested that it may be the result of relatively recent gene duplication and exon shuffling (14, 15). In addition to the highly homologous extracellular domains, all the CD33-like Siglecs except Siglec8 show conservation of two cytoplasmic tyrosine-based motifs (12). The first of these contains the consensus sequence for the immunoreceptor tyrosine kinase inhibitory motif (ITIM), (ILV)xYxx(LV) (x being any amino acid) (16, 17). This motif has been shown to be the binding site for the SH2 (src homology 2) domains of the SH2 domain-containing tyrosine phosphatase SHIP (18) and the SH2 domain-containing inositol phosphatase SHIP2 (19). The second motif displays homology to a

Abbreviations used: Siglec, sialic acid binding immunoglobulin-like lectin; Siglec8-L, Siglec8-Long; RT-PCR, reverse transcription-polymerase chain reaction; IgSF, immunoglobulin superfamily; EST, expressed sequence tag; ITIM, immunoreceptor tyrosine kinase inhibitory motif; RACE, rapid amplification of cDNA ends; SLAM, signaling lymphocyte activation molecule; SAP, SLAM-associated protein; SH2, src homology 2; SHP, SH2 domain-containing protein tyrosine phosphatase; SHIP, SH2 domain-containing inositol phosphatase.

1 To whom correspondence should be addressed at Mount Sinai Hospital, Department of Pathology and Laboratory Medicine, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. Fax: 416-586-8628. E-mail: ediamandis@mtsinai.on.ca.
tyrosine-based motif, TxYxx(IV), identified in the signaling lymphocyte activation molecule (SLAM) and was found responsible for its association with the SLAM-associated protein (SAP), which in turn blocks the binding of SHP-2 to phosphorylated SLAM (21, 22).

As mentioned previously, of the CD33-like subgroup of Siglecs, only Siglec8 does not possess the characteristic two tyrosine-based motifs (12). In the present study, we describe the identification and characterization of a novel variant of Siglec8, termed Siglec8-Long (Siglec8-L), which possesses these two tyrosine-based motifs, consistent with findings for the other members of the CD33-like subgroup of Siglecs. Further, we have determined the genomic organization of the Siglec8 gene, finding that it is comprised of seven exons, and is localized approximately 330 kb downstream of Siglec9 on chromosome 19q13.4. We have further identified the exact nucleotide at which the Siglec8 and Siglec8-L mRNAs deviate, producing proteins of 431 and 499 amino acids, respectively. The difference arises in exon six, where the Siglec8-L mRNA has a splice site after 97 nucleotides followed by a unique exon seven, while the Siglec8 mRNA continues right through exon 6 and then into a further downstream exon seven.

MATERIALS AND METHODS

Identification of the genomic area containing Siglec8. Genomic DNA sequences derived from BAC clones covering chromosome 19q13.4 were identified and obtained from the Lawrence Livermore National Laboratory (LLNL) Human Genome Center. These sequences were compared to the mRNA sequence for Siglec8 (GenBank Accession No. AF195092), which has been reported to be linked to 19q13.4 through the aid of the WebCutter restriction analysis program (25). Based on the results of an R1 map for chromosome 19q13.4 (25), we proceeded to map the exons of the Siglec8 gene. This was achieved through the use of the BLASTN nucleotide alignment tool, which enabled us to localize the mRNA sequences to specific regions of genomic DNA.

Following the initial characterization of Siglec8-L, the primary structure of the protein encoded by the Siglec8-L mRNA was determined. This was compared to the published protein sequence for Siglec8, as well as to all the members of the CD33-like subgroup of Siglecs using the CLUSTALX multiple alignment tool (26).

RESULTS

Identification of the Genomic Area Containing Siglec8

The CD33-like subgroup of Siglecs has been mapped, through various means, to chromosome 19q13.4. The Siglec9 gene has recently been characterized and localized to this area (13), immediately following the end of the kallikrein gene family. Further, the mRNA species deviates, producing proteins of 431 and 499 amino acids, respectively. The difference arises in exon six, where the Siglec8-L mRNA has a splice site after 97 nucleotides followed by a unique exon seven, while the Siglec8 mRNA continues right through exon 6 and then into a further downstream exon seven.
Based on the sequence information from the clone on which Siglec8 is localized, we used the WebCutter restriction analysis tool to determine the size of the EcoR1 fragments produced. By comparing these results with the available EcoR1 restriction map for chromosome 19q13.4, we determined that the Siglec8 gene is located in the more centromeric region of 19q13.4, and is approximately 330 kb downstream from the Siglec9 gene.

Molecular Characterization of Siglec8-L

Using RT-PCR and primers derived from sequence shared by both Siglec8 and Siglec8-L, as well as from the 3' termini of Siglec8 and the putative Siglec8-L, we were able to confirm the existence of both putative exons mentioned above. Through automated sequencing and subsequent alignment of the mRNA sequence with the genomic sequence for our BAC clone, we were able to determine the exact genomic organization for these last two exons. However, using the PCR primers specific for the published sequence of Siglec8 we obtained a very faint band, after agarose gel electrophoresis, which we were unable to sequence. In contrast with the genomic sequence for our BAC clone, we were able to determine the exact genomic organization for these last two exons. However, using the PCR primers specific for the published sequence of Siglec8 we obtained a very faint band, after agarose gel electrophoresis, which we were unable to sequence. In contrast with the genomic sequence for our BAC clone, we were able to determine the exact genomic organization for these last two exons.

FIG. 1. Genomic organization of the Siglec8 gene. Based on our own experimental results, as well as the previously published sequence for the Siglec8 mRNA (GenBank Accession No. AF195092), we have established that both Siglec8 and our Siglec8-L are composed of seven exons (Arabic numerals). The two mRNA species are identical until exon six, where the Siglec8 mRNA contains the whole of exon six (6a and 6b) and continues to exon 7a, indicated by the broken line, while the Siglec8-L mRNA contains only exons 6b and 7b, shown by the solid line. The location of the stop codon (TGA) is shown for both mRNA species, and differs due to a change in the open reading frame. Splice sites are conserved (-mGT . . .AGm-, where m is any base) in Siglec8-L between exons 6b and 7b, but not between exons 6a and 7a in the Siglec8 sequence reported by Floyd et al. (12).

DISCUSSION

Through efforts to investigate the CD33-like subgroup of Siglec8, we have identified the area of chromosome 19q13.4 which contains the Siglec8 gene, located approximately 330 kb downstream of the recently published Siglec9 gene (13). Examination of this area revealed the existence of an alternative form of Siglec8, named Siglec8-L. The protein product of this mRNA species differs markedly from the previously published Siglec8 (GenBank Accession No. AF195092) at its C-terminus (12). Siglec8-L is a 499 residue protein with a molecular weight of 54.04 kDa, excluding any posttranslational modifications which may be present. The sequence of Siglec8-L is identical to that of Siglec8 until residue 415, after which Siglec8-L contains a sequence homologous to the C-terminal sequences of the CD33-like subgroup of Siglec8, including the two tyrosine-based motifs (Fig. 2).
contains approximately 1.5 kb of untranslated sequence at its 3' end. By comparison, and in keeping with the hypothesis that this subgroup of Siglecs arose through gene duplication relatively recently in vertebrate evolution, none of the other Siglecs that belong to this subgroup have such an extensive 3′ untranslated region, or any untranslated exons at the 3′ end. Furthermore, the intron–exon splice sites for the last two exons of Siglec8 (based on our genomic sequence and the mRNA sequence of Floyd et al.) are not consistent with the characteristic splice donor and acceptor sites, unlike the first five exons and the two exons identified by us (Fig. 1).

The identification of Siglec8 by Floyd et al. (2000) was accomplished through the use of an EST that showed homology to CD33 (12). We have not as yet been able to identify this clone, and believe that the discrepancy between our sequence and theirs may be due to the inclusion of genomic sequence in the EST. Wolfsberg and Landsman have reported that they have indeed found several instances in which ESTs contained intron sequences or unusual splice sites (28). If this is the case with Siglec8, the EST used by Floyd et al. (12) may represent a partially spliced or incorrectly spliced sequence, which results in the inclusion of two pieces of genomic DNA (not present in any mRNA species we were able to identify) as well as the absence of an entire exon. During their identification of Siglec8, the authors report an unsuccessful attempt of Northern blot analysis for Siglec8 using a probe derived from the coding and 3′ untranslated sequences of their mRNA species. Based on the above noted problems with ESTs, the lack of specific bands in the Northern blot may be due to such intronic sequence contamination. This contamination likely resulted in a change in the open reading frame of the Siglec8 mRNA which causes premature termination and loss of the two tyrosine-based motifs. Despite efforts to identify the mRNA species reported by Floyd et al. (12), we have been unsuccessful in obtaining any PCR product that could be sequenced. However, we can still not rule out the possibility that this mRNA species does in fact represent an alternatively spliced form that lacks the two tyrosine motifs and is present in tissues not tested by us.

During their characterization of Siglec8, the authors demonstrated eosinophil–specific expression for this member of the CD33-like subgroup of Siglecs (12). This was accomplished by generating monoclonal antibodies to the extracellular domains of Siglec8. Given that Siglec8 and Siglec8-L are identical in their extracellular domains, the expression of Siglec8-L is also expected to be restricted to eosinophils.

For a few of the other members of the Siglec family, alternative splice forms have been described. For CD22, two isoforms have been identified in humans, with either four or six C2-set Ig-like domains (4, 29, 30). In the mouse there have been three isoforms of MAG identified (31, 32). One of the isoforms lacks an untranslated exon in the 5′ end of the mRNA, which has no effect on the size of the resultant polypeptide. The two other forms differ in the presence of exon 12, which is 45 nucleotides in length and introduces a termination codon when included in the mRNA. In humans, however, there has been no report of any MAG isoforms. Further, two isoforms for CD33 have been reported in the mouse, which differ by an 83 nucleotide in-frame insertion in the cytoplasmic domain (33). Human CD33 has also been reported to exist as two different size transcripts, which is believed to be through the use of alternative polyadenylation signals, with no change in the size of the polypeptide (34). Therefore, although there is evidence of alternative splicing in members of the Siglec family, it appears to occur primarily in nonhuman members, with the exception of the alternative number of Ig-like domains for human CD22. However, even these cases do not compare to the drastic differences seen in the splicing patterns of Siglec8 and Siglec8-L.

The Siglec family of transmembrane proteins, and in particular the CD33-like subgroup, is a very recently described member of the IgSF (1, 14, 15). There has been only limited progress into the investigation of the biological function of these proteins, with the exception of Siglec7, and more recently CD33. Siglec7, identified initially by Falco et al. (9), was found to be tyrosine-phosphorylated in its ITIM motif, resulting in recruitment of SHP1 with a consequent inhibition of natural killer cell cytotoxicity. More recently, it has been reported that engagement of Siglec7 and CD33, through

FIG. 2. Protein sequence alignment for the CD33-like subgroup of Siglecs. The sequence of Siglec8-L was aligned to those of Siglec8, as well as the remaining members of the CD33-like subgroup of Siglecs, using the ClustalX multiple alignment tool (26). The solid vertical lines indicate the positions of the exon boundaries. In all but one case, shown by the broken vertical line, the exon boundaries match those found for Siglec9 (13). The conserved cysteine residues responsible for the intra- and interdomain disulfide bonds (2, 37, 38) are indicated by the star (*), while the triangles (▲) denote the aromatic residues believed to be important for sialic acid binding, based on findings for Siglec1 (sialoadhesin) (39). The signal peptide cleavage site for Siglec8-L, indicated by the solid circle (●), was predicted using the SignalP program (40). The Ig-like domain assignments, as well as those for the transmembrane and cytoplasmic domains, are based on previous reports (12) and the one domain–one exon rule (41). The positions of the two tyrosine-based motifs, ITIM and SLAM-like, are indicated. The GenBank Accession Nos. are as follows: Siglec8-L, AF287982; Siglec8, AF195092; Siglec7, AF170485; Siglec6, NM_001245; Siglec5, NM_003830; Siglec3 (CD33), M23197.
the use of monoclonal antibodies, inhibits the proliferation of both normal and leukemic myeloid cells in vitro (35). These effects are believed to be the result of phosphorylation of the two tyrosine-based motifs present in the cytoplasmic domain of both CD33 and Siglec7. In addition, Taylor et al. (36) have found that CD33 recruits the protein-tyrosine phosphatases SHP1 and SHP2, both in vitro and in vivo, and is the result of tyrosine phosphorylation in the ITIM motif. Mutation of the tyrosine in this ITIM motif of CD33 resulted in increased red blood cell binding by CD33-expressing COS cells. These findings suggest that, in addition to the recruitment of SHP1 and 2 and inhibiting the activating signaling pathways that lead to cell proliferation and survival, this recruitment may also modulate the receptor’s ligand-binding activity (36). It is quite likely, given the high degree of homology within the CD33-like subgroup of Siglecs, that the remainder of this group, including Siglec-8-L, play a similar inhibitory role in their respective cell types.

REFERENCES


