Biochemical and Biophysical Research Communications **284,** 887–899 (2001)
doi:10.1006/bbrc.2001.5052, available online at http://www.idealibrary.com on IDEAL®

Cloning and Molecular Characterization of Two Splice Variants of a New Putative Member of the Siglec-3-like Subgroup of Siglecs

George Foussias,*,† Steve M. Taylor,* George M. Yousef,*,† Michael B. Tropak,‡ Michael H. Ordon,* 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 and ‡Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada*

Received May 7, 2001

The sialic acid binding immunoglobulin-like lectin (Siglec) family is a recently described member of the immunoglobulin superfamily. Within this Siglec family there exists a subgroup of molecules which bear a very high degree of homology with the molecule Siglec-3 (CD33), and has thus been designated the Siglec-3-like subgroup of Siglecs. The members of this subgroup have been localized to chromosome 19q13.4, through both *in situ* hybridization and precise genomic mapping at the nucleotide level. Through the positional cloning approach we have identified and characterized a Siglec-like gene (SLG), a putative novel member of the Siglec-3-like subgroup of Siglecs. We have characterized the complete genomic structure of SLG, as well as two alternative splice variants, and determined its chromosomal localization. The short isoform, SLG-S, consists of seven exons, with six intervening introns, while the longer isoform, SLG-L, consists of eight exons and seven intervening introns. The SLG gene is localized 32.9 kb downstream of Siglec-8 on chromosome 19q13.4. The putative SLG-S and SLG-L proteins, of 477 and 595 amino acid residues, respectively, show extensive homology to many members of the Siglec-3-like subgroup. This high degree of homology is conserved in the extracellular Ig-like domains, as well as in the cytoplasmic tyrosine-based motifs. Interestingly, the SLG-L protein contains two N-terminal V-set Ig-like domains, as opposed to SLG-S and other Siglec-3-like subgroup members which contain only one such domain. Through RT-PCR we have examined the expression profile of both SLG splice variants in a panel of human tissues and have found that SLG-S is highly expressed in spleen, small intestine and adrenal gland, while SLG-L exhibits high levels of expression in spleen, small intestine, and bone marrow. This gene is quite likely the latest novel member of the CD33-like subgroup of Siglecs, and given its high degree of homology, it may also serve a regulatory role in the proliferation and survival of a particular hematopoietic stem cell lineage, as has been found for CD33 and Siglec7.

Key Words: **Siglec gene family; CD33-like subgroup; immunoglobulin superfamily; sialoadhesins; new gene discovery; positional cloning.**

Siglec-like gene: GenBank Accession No. AF277806.

Abbreviations used: Siglec, sialic acid-binding immunoglobulin-like lectin; 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; SHP, SH2 domain-containing protein tyrosine phosphatase; SHIP, SH2 domain-containing inositol phosphatase; AML, acute myelogenous leukemia; SLG, Siglec-like gene; BAC, bacterial artificial chromosome.

[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. Fax: 416-586-8628. E-mail: ediamandis@mtsinai.on.ca.

Sialic acid binding immunoglobulin-like lectins (Siglecs) are a novel family of type I transmembrane proteins belonging to the immunoglobulin superfamily. They mediate protein–carbohydrate interactions through their ability to bind sialic acid moieties found on glycolipids and glycoproteins (1, 2). These receptors are characterized by the presence of an N-terminal V-set Ig-like domain and a variable number of downstream C2-set Ig-like domains, ranging from 1 in Siglec-3 to 16 in Siglec-1. Each of the nine members of the Siglec family characterized in humans to date is expressed by a specific hematopoietic cell lineage, with the exception of Siglec-9 which is found on several cell types. Siglec-1 (Sialoadhesin) is a macrophage-restricted adhesion molecule (3), Siglec-2 (CD22) is found on B-lymphocytes

0006-291X/01 $35.00

and regulates their activation (4), Siglec-3 (CD33) is a myeloid-specific inhibitory receptor (5, 6), and Siglec-4a (Myelin-associated glycoprotein, MAG) is found on myelinating oligodendrocytes and Schwann cells and is involved in myelin formation and maintenance (7, 8). Siglec-5 is expressed on neutrophils (9) and Siglec-6 on B-lymphocytes (10). Siglec-7 (AIRM1/p75) is an inhibitory receptor expressed on natural killer cells (11, 12), while Siglec-8 is restricted to eosinophils (13), and Siglec-9 is found on monocytes and neutrophils (14–16).

Among the Siglecs, a subgroup of proteins exist which share a greater degree of sequence homology to Siglec-3. This subgroup is found in a cluster on chromosome 19q13.4 and includes Siglec-3, -5, -6, -7, -8, -8-L, and -9 (14–17). These Siglec-3-like Siglecs are characterized by the presence of two tyrosine-based motifs in their cytoplasmic tails: (i) an immunoreceptor tyrosine kinase inhibition motif (ITIM), with a consensus sequence (I/L/V)xYxx(L/V) (18, 19); and (ii) a motif similar to that identified in the signaling lymphocyte activation molecule (SLAM), referred to as a SLAM-like motif, with the sequence TxYxx(I/V) (20, 21). ITIM motifs have been found to serve as binding sites for the SH2 (src homology 2) domains of the SH2-domain containing protein tyrosine phosphatases SHP1 and SHP2 (22, 23), as well as the SH2-domain containing inositol phosphatase SHIP1 and SHIP2 (24). The second, SLAM-like, motif was originally identified in SLAM and found to recruit both the SLAM-associated protein (SAP) and the tyrosine phosphatase SHP2, both through their SH2 domains (20, 21). The presence of such cytoplasmic motifs capable of recruiting various phosphatases suggests an inhibitory role for these Siglec-3-related Siglecs in intracellular signaling pathways. Functional studies performed on a few members of this subgroup appear to support this hypothesis. Siglec-7, which was initially identified as a natural killer cell inhibitory receptor, is able to inhibit natural killer cell cytotoxicity upon tyrosine phosphorylation of the ITIM motif and subsequent SHP1 recruitment (11). Further, Siglec-3 has been shown to recruit both SHP1 and SHP2 following tyrosine phosphorylation in its ITIM motif. For both of these receptors it has been found that their engagement with monoclonal antibodies results in the inhibition of both normal and leukemic myeloid cell proliferation (25).

In this paper we describe the identification and molecular characterization of a novel putative member of the Siglec family of genes, designated Siglec-like gene (SLG), and two alternative mRNA splice variants. We have determined the genomic organization of both these splice variants, finding them to be encoded by seven and eight exons, for SLG-Short (SLG-S) and SLG-Long (SLG-L), respectively. We have localized the SLG gene to chromosome 19q13.4, 32.9 kb downstream of Siglec-8. The resultant putative SLG-S and SLG-L proteins, of 477 and 595 amino acid residues, respectively, display a high degree of homology to Siglec-7, -8-L, and -9. Detailed examination of their homology with known members of the Siglec-3-like subgroup of Siglecs shows that they possesses all the structural features characteristic of this subgroup, including the extracellular Ig-like domains, as well as the cytoplasmic tyrosine-based motifs. In the case of the SLG-L variant, it appears to encode two N-terminal V-set Ig-like domains, a feature not found in other Siglecs to date. We examined the tissue expression profile of both SLG splice variants through RT-PCR, finding that SLG-S is highly expressed in spleen, small intestine, and adrenal gland, while SLG-L is highly expressed in spleen, small intestine, and bone marrow.

MATERIALS AND METHODS

Identification of the Siglec-like gene (SLG). Based on the high degree of homology among the Siglec-3-like subgroup of Siglecs, in both the extracellular Ig-like domains as well as the cytoplasmic tyrosine-based motifs, we screened the human expressed sequence tag (EST) database with these sequences, using the BLAST alignment tool (26). Any matching ESTs we identified were obtained from the I.M.A.G.E. consortium through Research Genetics Inc. (Huntsville, AL). These clones were then propagated according to the suppliers instructions, purified, and sequenced from both directions with an automated sequencer, using the flanking T3 and T7 vector primers. Further, we obtained genomic sequence derived from BAC clones which cover the area of chromosome 19q13.4, believed to contain the Siglec-3-like subgroup locus, from the Lawrence Livermore National Laboratory (LLNL) Human Genome Center. We utilized the BLAST alignment tool to determine the exact location of any EST identified above in this genomic sequence. We then proceeded to investigate the genomic sequence surrounding the EST sequence with an exon prediction program, Grail2Exons (27).

Cloning and molecular characterization of two SLG splice forms. Based on the alignment of the EST and the exon prediction results, we proceeded to design sets of primers to be used with reverse transcription-coupled polymerase chain reaction (RT-PCR) in order to determine the exact sequence of the SLG mRNA species. This design allowed for the production of overlapping RT-PCR fragments, thus enabling us to determine the entire mRNA sequence. Based on results from RT-PCR with a panel of human tissues (see below), we used bone marrow as the tissue with which to work with. Bone marrow cDNA was prepared as described below. The primer combinations which we used were: (i) SLG-F3: 5′-AGGAAGCCTCTGCCTCAGAG-3′, and SLG-R3: 5′-CCTTCATCCTTCACATGCAC-3′; and (ii) SLG-F2: 5′-ATCACTCGCTCCTCGATGCT-3′, and SLG-R2: 5′-TCTCTCCTTCCTCTGGGGAG-3′. Due to the high degree of homology, even at the nucleotide level, among the Siglec-3-like subgroup of Siglecs, we utilized semi-nested PCR, to ensure we were amplifying the correct mRNA species, using the above forward primers, SLG-F3 and SLG-F2, and the nested reverse primers SLG-R3N (5′-GAGGACTGTGAGGGGCTCAG-3′) and SLG-R2N (5′-GATTCAATCAGGGGTCC-3′), respectively. The PCR conditions were as follows: 2.5 units HotStarTaq polymerase (Qiagen, Valencia, CA), 1× PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 1 μl cDNA, 200 μM dNTPs (deoxynucleoside triphosphates), and 200 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 62°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 electrophoresis on a 2% agarose gel containing ethidium bromide.

The PCR products were extracted from the gel and the purified DNA was directly sequenced using an automated sequencer.

During our examination of this genomic area through exon prediction, an additional putative upstream exon was also identified, which also possessed a signal peptide and Ig-like sequence highly homologous to existing members of the Siglec-3-like subgroup of Siglecs. In order to examine the possibility of an additional upstream exon with a slightly different signal peptide and Ig-like sequence, we designed a forward PCR primer specific for this exon, lying within its 5′ untranslated region. This upstream primer, SLG-FU (5′-TGGCAC-CTCCAACCCGACAC-3′), was used in combination with the reverse primer SLG-R3 for RT-PCR, under identical conditions to those described above. The PCR product was subjected to electrophoresis on a 2% agarose gel containing ethidium bromide. The resultant DNA bands were extracted from the gel, purified, and directly sequenced using an automated sequencer.

Following final characterization of both the SLG mRNA sequences as well as the genomic organization for SLG, the putative protein products were determined. These protein sequences were then aligned with those of the other known members of the Siglec-3-like subgroup of Siglecs using the ClustalX multiple alignment tool (28). Shading of similar and identical residues was accomplished using the BOXSHADE alignment shading program (http://www.ch.embnet.org/software/BOX–form.html). Further, the SLG protein sequences were examined for the presence of transmembrane domains, with both TMpred (29) and through their Kyte–Doolittle hydrophobicity profile (30). The presence of putative signal peptides was examined through the use of the SignalP v1.1 signal peptide prediction program (Nielsen *et al.*, 1997).

Tissue expression of two SLG splice variants. The tissue expression profile for both SLG alternative splice forms was elucidated by performing RT-PCR using total RNA from 25 normal human tissues (Clontech, Palo Alto, CA). The PCR primers used were SLG-F3 and SLG-R3, and SLG-FU and SLG-R3, for the short and longs forms of SLG, respectively. The PCR conditions were the same as those described above, and reverse transcription was performed using SuperScript II, according to the manufacturer's instructions (Gibco BRL, Gaithersburg, MD). The temperature profile for the PCR was identical to that described previously. From the cDNA that was produced we also performed a PCR for actin, as described elsewhere (31), as a control for cDNA quality.

Mapping and chromosomal localization of SLG. As indicated above, the SLG gene was identified in the genomic sequence from a BAC clone covering chromosome 19q13.4. We subjected the sequence encompassing the SLG gene to the Webcutter restriction analysis tool to determine the size of the resultant EcoR1 fragments. We then compared these results to the published EcoR1 map for chromosome 19 (32), which is also available through the LLNL Human Genome Center. In addition, the SLG mRNA sequences were aligned with the final genomic sequence from this BAC clone (GenBank Accession No. AC020914) using the BLASTN nucleotide alignment program, allowing for precise localization of all exons. Further, by knowing the precise location of both the Siglec-9 gene (14) and the Siglec-8 gene (17), we determined the distance between these and the SLG gene.

Phylogenetic analysis of the expanded Siglec family. In order to examine the evolutionary relationship among the members of the Siglec family, including the novel members identified in this work, we proceeded to perform phylogenetic analysis. This was achieved, initially, through use of the ClustalX multiple alignment tool, utilizing the protein sequence for all the human Siglec family members and those of SLG-S and -L. Phylogenetic analysis was then performed with the Phylip software package, through the use of ClustalX and a distance matrix. Parsimony trees were constructed using neighborhood-joining, UPGMA, and protpars parsimony methods. Selection of the appropriate tree was based on which of these was consistent with the findings by other groups (15).

RESULTS

Identification of the Siglec-like Gene (SLG)

Our screening for ESTs homologous to previously published members of the Siglec-3-like subgroup revealed the presence of an EST (GenBank Accession No. AI132995) which showed extensive homology to the tyrosine-based motifs found in the cytoplasmic tail of other members of this subgroup. This EST was then compared to genomic sequence derived from BAC clones covering chromosome 19q13.4. We identified one clone, CTD-3073N11, which contained this EST in the form of three exons. Subsequent exon prediction using the genomic sequence from this clone indicated the presence of an additional four putative exons.

Cloning and Molecular Characterization of Two SLG Splice Forms

Based on the results of exon prediction, followed by verification through RT-PCR and sequencing, we cloned and fully characterized the entire mRNA structure of SLG-Short (SLG-S). Through alignment with the genomic sequence we determined the precise genomic organization of the SLG gene. We found that this mRNA species, similar to our findings for Siglec-8 and -9 (14, 17), is encoded by seven exons, with six intervening introns. The first two predicted exons mentioned above were found to be a single exon, based on our experimentally defined mRNA sequence. Further, the exon prediction did not detect the third exon, which was identified through RT-PCR. The lengths of the exons which encode the SLG-S mRNA are 474, 279, 48, 270, 97, 97, and 471 bp (Fig. 1). All intron/exon splice sites are closely related to the consensus splice sites (-m**GT**AAGT . . . C**AG**m-, where m is any base) (33).

The proposed open reading frame for the SLG-S mRNA transcript consists of 1736 bp, which results in a 477 amino acid protein, with a molecular weight of 51.7 kDa, excluding any posttranslational modifications. The translation initiation codon (ATG), located at position 21 (based on the numbering of our GenBank submission) was chosen for two reasons: (i) the sequence surrounding this initiation codon conforms to the Kozak consensus sequence for translation initiation, especially with the most highly conserved purine at position -3 (34); (ii) with this translation initiation codon the resultant protein product shows extensive homology with other members of the Siglec-3-like subgroup of Siglecs (see below), as well as the fact that no other initiation codon was found that produced a long, continuous open reading frame. The first exon contains a 5′ untranslated region of at least 20 bp, while in the seventh (last) exon there is a 3′ untranslated region of 282 bp. Through the presence of a poly dA tail at the 3′ end of the EST, we were able to identify the end of the SLG-S mRNA transcript.

```
TGGCACCTCCAACCCGACACATG CTA CTG CTG CTG CTA CTG CTG CCA CCC CTG CTC TGT GGG AGA GTG GGG
                     M  L  L  L  L  L  L  P  P  L  L  C  G  R  V  G
GCT AAG GAA CAG AAG GAT TAC CTG CTG ACA ATG CAG AAG TCC GTG ACG GTG CAG GAG GGC CTG TGT
A  K  E  Q  K  D  Y  L  L  T  M  Q  K  S  V  T  V  Q  E  G  L  C
GTC TCT GTG CTT TGC TCC TTC TCC TAC CCC CAA AAT GGC TGG ACT GCC TCC GAT CCA GTT CAT GGC
V  S  V  L  C  S  F  S  Y  P  Q  N  G  W  T  A  S  D  P  V  H  G
TAC TGG TTC CGG GCA GGG GAC CAT GTA AGC CGG AAC ATT CCA GTG GCC ACA AAC AAC CCA GCT CGA
Y  W  F  R  A  G  D  H  V  S  R  N  I  P  V  A  T  N  N  P  A  R
GCA GTG CAG GAG GAG ACT CGG GAC CGA TTC CAC CTC CTT GGG GAC CCA CAG AAC AAG GAT TGT ACC
A  V  Q  E  E  T  R  D  R  F  H  L  L  G  D  P  Q  N  K  D  C  T
CTG AGC ATC AGA GAC ACC AGA GAG AGT GAT GCA GGG ACA TAC GTC TTT TGT GTA GAG AGA GGA AAT
L  S  I  R  D  T  R  E  S  D  A  G  T  Y  V  F  C  V  E  R  G  N
ATG AAA TGG AAT TAT AAA TAT GAC CAG CTC TCT GTG AAT GTG ACA Ggtaa....INTRON 1....acag
M  K  W  N  Y  K  Y  D  Q  L  S  V  N  V  T

AGGAAGCCTCTGCCTCAGAGATG CTG CTG CCC CTG CTA TGG GCA AAT GAA GAG AGG GAC AGT GGG GGC TGG
                     M  L  L  P  L  L  W  A  N  E  E  R  D  S  G  G  W
GCT GAC CCT CGT TTC TCC ACA GCG TCC CAG GAC CTA CTG TCA AGA TAC AGG CTG GAG GTG CCA GAG
A  D  P  R  F  S  T  A  S  Q  D  L  L  S  R  Y  R  L  E  V  P  E
TCG GTG ACT GTG CAG GAG GGT CTG TGT GTC TCT GTG CCC TGC AGT GTC CTT TAC CCC CAT TAC AAC
S  V  T  V  Q  E  G  L  C  V  S  V  P  C  S  V  L  Y  P  H  Y  N
TGG ACT GCC TCT AGC CCT GTT TAT GGA TCC TGG TTC AAG GAA GGG GCC GAT ATA CCA TGG GAT ATT
W  T  A  S  S  P  V  Y  G  S  W  F  K  E  G  A  D  I  P  W  D  I
CCA GTG GCC ACA AAC ACC CCA AGT GGA AAA GTG CAA GAG GAT ACC CAC GGT CGA TTC CTC CTC CTT
P  V  A  T  N  T  P  S  G  K  V  Q  E  D  T  H  G  R  F  L  L  L
GGG GAC CCA CAG ACC AAC AAC TGC TCC CTG AGC ATC AGA GAT GCC AGG AAG GGG GAT TCA GGG AAG
G  D  P  Q  T  N  N  C  S  L  S  I  R  D  A  R  K  G  D  S  G  K
TAC TAC TTC CAG GTG GAG AGA GGA AGC AGG AAA TGG AAC TAC ATA TAT GAC AAG CTC TCT GTG CAT
Y  Y  F  Q  V  E  R  G  S  R  K  W  N  Y  I  Y  D  K  L  S  V  H
GTG ACA Ggtaa....INTRON 2.... cagCC CTG ACT CAC ATG CCC ACC TTC TCC ATC CCG GGG ACC CTG
V  T                        A  L  T  H  M  P  T  F  S  I  P  G  T  L
GAG TCT GGC CAC CCC AGG AAC CTG ACC TGC TCT GTG CCC TGG GCC TGT GAA CAG GGG ACG CCC CCC
E  S  G  H  P  R  N  L  T  C  S  V  P  W  A  C  E  Q  G  T  P  P
ACG ATC ACC TGG ATG GGG GCC TCC GTG TCC TCC CTG GAC CCC ACT ATC ACT CGC TCC TCG ATG CTC
T  I  T  W  M  G  A  S  V  S  S  L  D  P  T  I  T  R  S  S  M  L
AGC CTC ATC CCA CAG CCC CAG GAC CAT GGC ACC AGC CTC ACC TGT CAG GTG ACC TTG CCT GGG GCC
S  L  I  P  Q  P  Q  D  H  G  T  S  L  T  C  Q  V  T  L  P  G  A
GGC GTG ACC ATG ACC AGG GCT GTC CGA CTC AAC ATA TCC Tgtga....INTRON 3.... cagAT CCT CCT
G  V  T  M  T  R  A  V  R  L  N  I  S                        Y  P  P
CAG AAC TTG ACC ATG ACT GTC TTC CAA GGA GAT GGC ACA Ggtaa....INTRON 4.... cagCA TCC ACA
Q  N  L  T  M  T  V  F  Q  G  D  G  T                        A  S  T
ACC TTG AGG AAT GGC TCG GCC CTT TCA GTC CTG GAG GGC CAG TCC CTG CAC CTT GTC TGT GCT GTC
T  L  R  N  G  S  A  L  S  V  L  E  G  Q  S  L  H  L  V  C  A  V
GAC AGC AAT CCC CCT GCC AGG CTG AGC TGG ACC TGG GGG AGC CTG ACC CTG AGC CCC TCA CAG TCC
D  S  N  P  P  A  R  L  S  W  T  W  G  S  L  T  L  S  P  S  Q  S
TCG AAC CTT GGG GTG CTG GAG CTG CCT CGA GTG CAT GTG AAG GAT GAA GGG GAA TTC ACC TGC CGA
S  N  L  G  V  L  E  L  P  R  V  H  V  K  D  E  G  E  F  T  C  R
GCT CAG AAC CCT CTA GGC TCC CAG CAC ATT TCC CTG AGC CTC TCC CTG CAA AAC GAG TAC ACA Ggt
A  Q  N  P  L  G  S  Q  H  I  S  L  S  L  S  L  Q  N  E  Y  T
gg....INTRON 5....cagGC AAA ATG AGG CCT ATA TCA GGA GTG ACG CTA GGG GCA TTC GGG GGA GCT
                     G  K  M  R  P  I  S  G  V  T  L  G  A  F  G  G  A
GGA GCC ACA GCC CTG GTC TTC CTG TAC TTC TGC ATC ATC TTC GTT GTgtga. ...INTRON 6....cagA
G  A  T  A  L  V  F  L  Y  F  C  I  I  F  V  V
GTG AGG TCC TGC AGG AAG AAA TCG GCA AGG CCA GCA GTG GGC GTG GGG GAT ACA GGC ATG GAG GAC
V  R  S  C  R  K  K  S  A  R  P  A  V  G  V  G  D  T  G  M  E  D
GCA AAC GCT GTC AGG GGC TCA GCC TCT CAG gtga....INTRON 7....cag GGA CCC CTG ATT GAA TCC
A  N  A  V  R  G  S  A  S  Q                        G  P  L  I  E  S
CCG GCA GAT GAC AGC CCC CCA CAC CAT GCT CCG CCA GCC CTG GCC ACC CCC TCC CCA GAG GAA GGA
P  A  D  D  S  P  P  H  H  A  P  P  A  L  A  T  P  S  P  E  E  G
GAG ATC CAG TAT GCA TCC CTC AGC TTC CAC AAA GCG AGG CCT CAG TAC CCA CAG GAA CAG GAG GCC
E  I  Q  Y  A  S  L  S  F  H  K  A  R  P  Q  Y  P  Q  E  Q  E  A
ATC GGC TAT GAG TAC TCC GAG ATC AAC ATC CCC AAG TGA GAAACTGCAGAGACTCAGGCCTGTTTGAGGGCTCA
I  G  Y  E  Y  S  E  I  N  I  P  K  *
CGACCCCTGCAGCAAAGAAGCCCGAGACTGATTCCTTTAGAATTAACAGCCCTCCATGCTGTGCAACAGGACATCAGAACTTATTCC
TCTTGTCAAACTGAAAATGCGTGCCTGATGACCAAACTCTCCCTTTCTCCATCCAATCGGTCCACACTCCCCGCCCCCGGCCTCTGG
TACCCACCATTCTCTTCTCTACTTCTCTGAGGTCGACTATTTTAGGTTCCAAATATAGTGAGATCGTAGAGTG
```

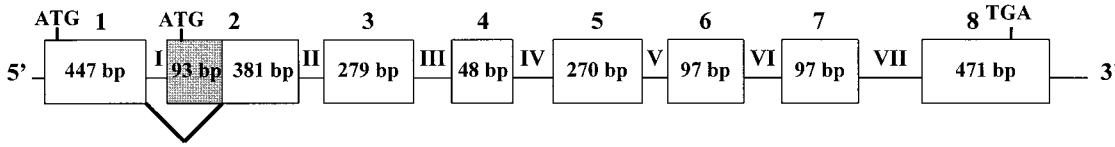


FIG. 2. Genomic organization of the SLG gene. Shown schematically above is the organization and splicing pattern for the two SLG mRNA transcripts we identified. The SLG-L transcript begins at exon 1, joining exon 2 as indicated by the solid line. The SLG-S mRNA transcript begins at the beginning of exon 2, with its corresponding ATG and continues through the remaining 6 exons, as does the SLG-L transcript. The shaded region denotes the nucleotide sequence that is contained in the SLG-S mRNA transcript, but absent in the SLG-L transcript. The sizes of the exons are indicated, with introns being denoted by roman numerals. The location of the stop codon (TGA), which is the same for both mRNA species, is also shown.

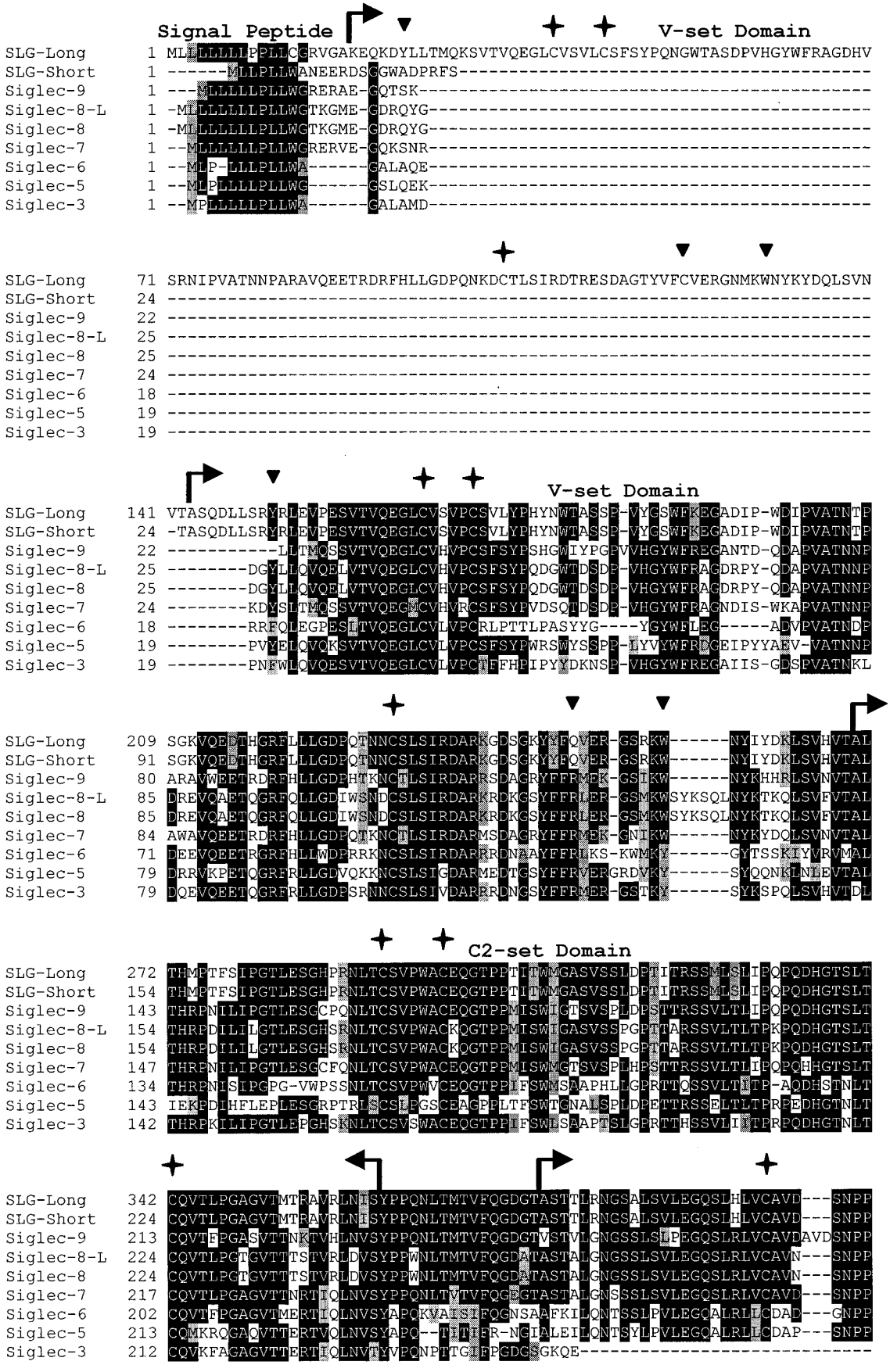
Through our examination of the additional putative upstream exon encoding an alternate signal peptide, we discovered that this produces an alternative mRNA species which differs from the SLG-S mRNA at its 5′ end (Fig. 1). This novel SLG transcript, designated SLG-Long (SLG-L) is comprised of 8 exons, with 7 intervening introns. Through nucleotide alignment of this mRNA species with the genomic sequence from the BAC clone, we were able to determine the exon boundaries for this alternative splice form as well. The first exon present in the SLG-L mRNA is at least 447 bp in length, including at least 20 bp of 5′ untranslated sequence, with a second exon of 381 bp. These 381 bp of the second SLG-L exon are identical to those of the SLG-S mRNA transcript. However, in the case of SLG-S, these are part of its first exon, which extends at least 93 bp, including at least 20 bp of 5′-untranslated sequence, upstream of the splice site for SLG-L (Fig. 2). The remainder of the SLG-L mRNA sequence is identical to that of SLG-S. As in the case of SLG-S, the SLG-L mRNA splice sites are all closely related to the consensus splice sites, as described above.

The putative protein product derived from the SLG-L mRNA transcript is composed of 2070 bp, producing a protein product of 595 amino acid residues, with a predicted molecular weight of 65.0 kDa, excluding post-translational modifications. The translation initiation codon (position 21 in the SLG-L GenBank submission) was chosen because it is the only reading from which produces a continuous open reading frame, maintaining this frame after splicing and assuming identity with the SLG-S mRNA sequence. Further, this open reading frame exhibits extensive homology to known members of the Siglec-3-related subgroup of Siglecs.

Examination of both SLG-S and SLG-L protein sequences revealed that they are highly homologous to other members of the Siglec-3-like subgroup of Siglecs (Fig. 3). Like other members of this subgroup, SLG-S and SLG-L also appear to possess similar N-terminal signal sequences. These hydrophobic N-terminal sequences are evident in the hydrophobicity plots for SLG-S and SLG-L, with the hydrophobic N-terminal sequence of SLG-L being longer than that of SLG-S (Fig. 4). This is also reflected in the signal peptide prediction results, which detects a strong signal putative signal peptide sequence only in the SLG-L protein sequence. These are followed by an N-terminal V-set Ig-like domain, two in the case of SLG-L, and two C2-set Ig-like domains, similar to other Siglec-3-like Siglecs. The single transmembrane domain, predicted by TMpred and evident in the Kyte-Doolittle hydrophobicity plots (Fig. 4), is in keeping with observations for other members of this subgroup. Furthermore, we noticed that both SLG splice forms also contain the two characteristic tyrosine-based motifs, ITIM and SLAM-like, noted in members of the Siglec-3-like subgroup of Siglecs. Further, as is evident in Fig. 3, there is also conservation of all the key cysteine residues that are responsible for the characteristic folding of the extracellular Ig-like domains in all Siglecs (1, 35). With regards to the residues believed to be responsible for sialic acid binding there is conservation of both aromatic residues, however in SLG-S there is a glutamine, and in SLG-L there is a cysteine, in place of the otherwise conserved arginine (36).

We also proceeded to perform a more detailed examination of the homology between both SLG splice forms and the other members of both the Siglec-3-like subgroup and the other Siglec family members. This was achieved through the use of the BLASTP protein alignment tool (26). As shown in Table 1, SLG-S has over 70% similarity with Siglecs 7–9, in addition to slightly lower homology with other members of this subgroup.

FIG. 1. Nucleotide and protein sequences of both SLG isoforms. The nucleotide sequences of both splice variants, SLG-L and SLG-S, which we verified experimentally, along with their amino acid translation are shown. The 5′ untranslated region of the SLG-L mRNA transcript is indicated by the solid underline, while that of the SLG-S transcript is indicated by the dotted underline. Further, the predicted SLG-S protein is indicated by the broken underline, until the codon where both splice variants assume the same reading frame. The 3′ untranslated region is indicated by the double underline. The intron/exon splice sites are also shown, in lowercase letters. A putative polyadenylation signal is also indicated by the shaded sequence.


Signal Peptide
V-set Domain
V-set Domain
C2-set Domain
SLG-Long
SLG-Short
Siglec-9
Siglec-8-L
Siglec-8
Siglec-7
Siglec-6
Siglec-5
Siglec-3

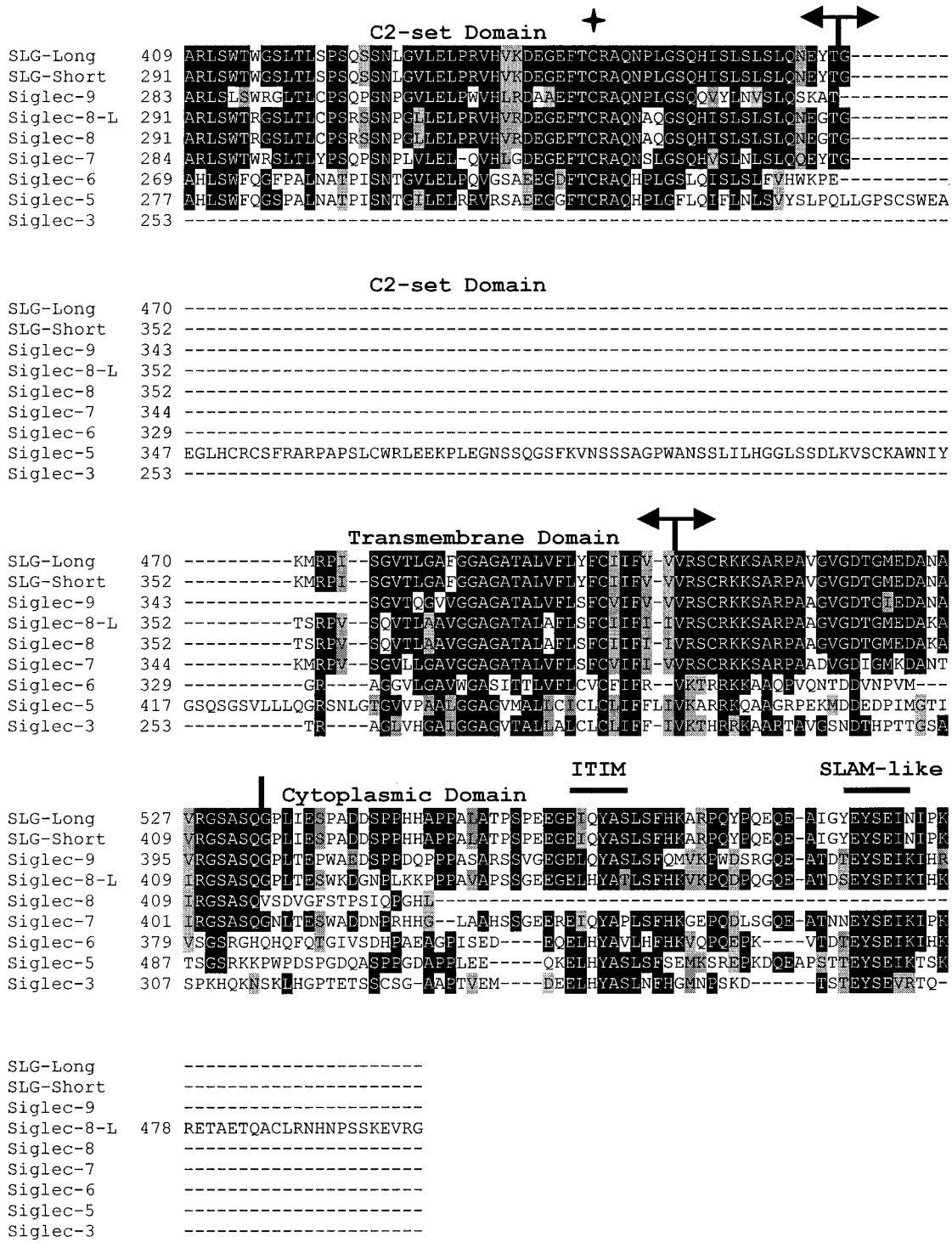


FIG. 3. Protein sequence alignment of both SLG splice variants and the Siglec-3-like subgroup of Siglecs. The sequences of both SLG-S and SLG-L were aligned to those of the entire Siglec-3-like subgroup of Siglecs, using the ClustalX multiple alignment tool (28). The solid vertical lines indicate the positions of the exon boundaries, with the arrows indicating the different domains. The conserved cysteine residues responsible for the intra- and interdomain disulfide bonds are indicated by the star, while the triangles denote the residues believed to be important for sialic acid binding, based on findings for Siglec-1. The signal peptide is indicated and is based on homology with other members of the Siglec-3-like subgroup of Siglecs, as well as signal peptide prediction in the case of SLG-L. The Ig-like domain assignments, as well as those for the transmembrane and cytoplasmic domains, are based on previous reports (17) and the one domain–one exon rule (45). For the predicted transmembrane domain, the TMpred results indicate that it exists between residues 366 and 384 (based on the numbering of the SLG-S transcript). The positions of the two tyrosine-based motifs, ITIM and SLAM-like, are indicated.

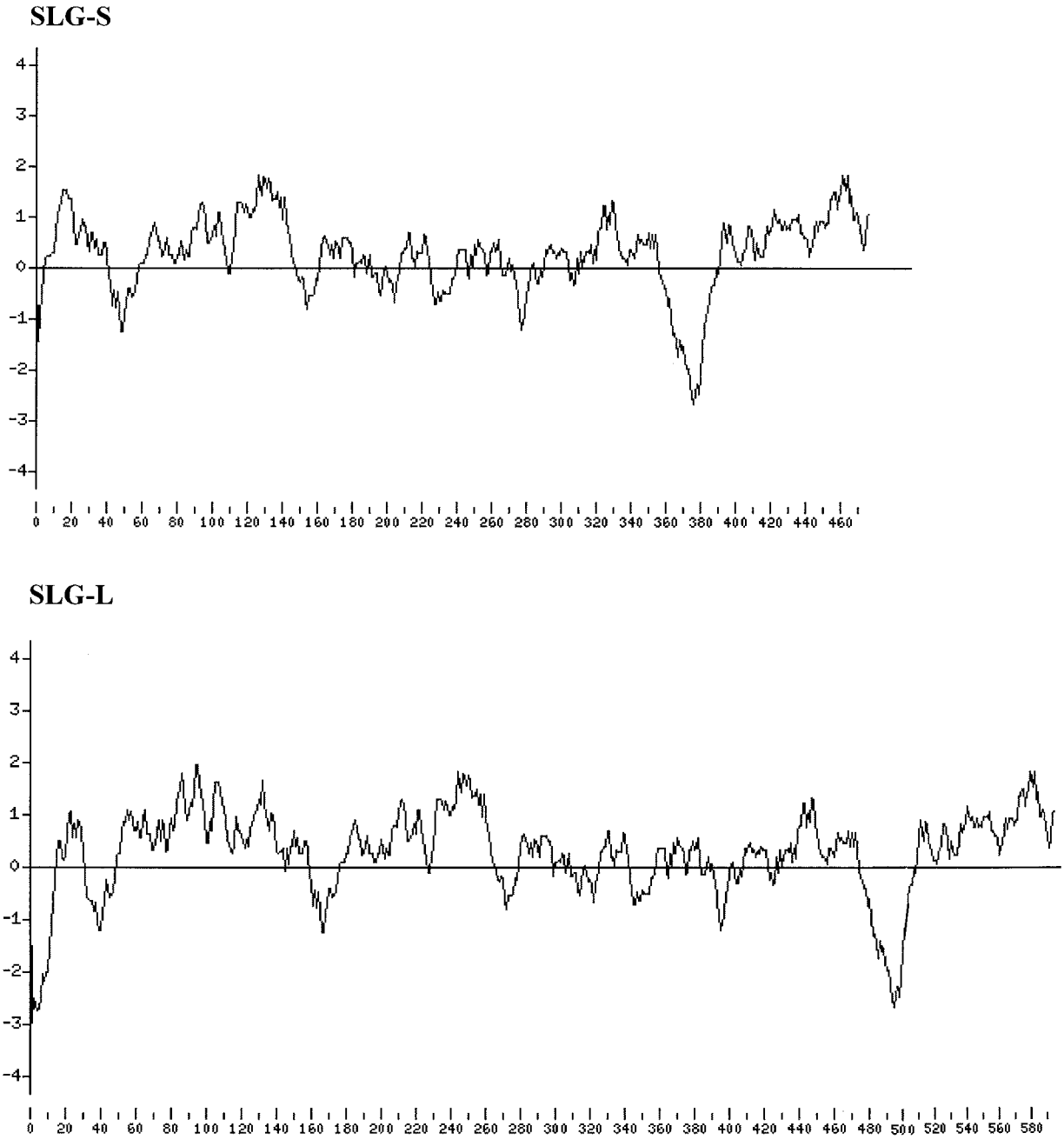


FIG. 4. Hydrophobicity plots of both SLG isoforms. This shows the regions of the putative SLG-S and SLG-L proteins which contain stretches of hydrophobic amino acid residues. As is evident, there is a strong hydrophobic region around residues 360–390 of SLG-S, as well as around residues 475–505 for SLG-L, which contain the predicted transmembrane region. Further, the C-termini of both proteins, especially SLG-L, exhibit a smaller hydrophobic region, which likely correspond to the signal peptide as seen in other members of the Siglec-3-like subgroup of Siglecs.

SLG-L also exhibits in excess of 70% homology with the Siglecs 7–9, with increased homology to Siglec-3, -5, and -6, compared to SLG-S.

Tissue Expression Profile of Two SLG Splice Variants

Through RT-PCR with a total RNA panel of 25 different normal human tissues we examined the tissue expression profile of both SLG splice forms (Fig. 5). SLG-S was found to be relatively highly expressed in spleen, small intestine and adrenal gland. Moderate to mild expression was apparent in thyroid, placenta, brain, stomach, bone marrow, spinal cord, and breast. SLG-L, on the other hand, exhibited relatively high levels of expression in spleen, small intestine, and bone

TABLE 1

SLG Isoform Sequence Homology with the Siglec-3-like Subgroup of Siglecs

Siglec-3-like subgroup member	GenBank Accession No.	Homology to SLG-S[1]		Homology to SLG-L[1]	
		% Identity	% Similarity	% Identity	% Similarity
Siglec-7	NM–014385	68	76	71	79
Siglec-8-L	AF287892	67	75	71	78
Siglec-8	NM–014442	66	74	73	80
Siglec-9	AF135027	64	73	68	77
Siglec-6	NM–001245	48	60	49	61
Siglec-3	M23197	41	50	55	69
Siglec-5	NM–003830	40	50	53	67

[1] Homology was determined using the BLASTP algorithm.

marrow. Further, intermediate levels of expression were observed in thyroid, placenta, thymus, trachea, stomach, lung, adrenal gland, fetal brain, and testis. The PCR products obtained were all of equal size and corresponded to the length of the product obtained during the molecular characterization of the respective SLG mRNA transcript. Further, specificity was ensured through sequencing of RT-PCR products. In the case of SLG-L, where in some tissues two bands are observed, sequencing of both DNA bands revealed that they in fact are identical, and may possibly be due to different secondary structures for the same DNA sequence, resulting in different migration speeds.

Mapping and Chromosomal Localization of SLG

The contig on which we identified the SLG gene was located telomeric to the Siglec-9 gene, which was originally characterized and precisely mapped in our laboratory (14). Furthermore, this same contig contained the Siglec-8 gene, also described by others (13) and by our group (17). Therefore, through EcoR1 mapping, as described above, as well as the known locations of both the Siglec-8 and -9 genes, we were able to determine the location of the SLG gene. We found that the SLG gene is located 32.9 kb more telomeric than the Siglec-8 gene, and approximately 363 kb downstream of the Siglec-9 gene, on chromosome 19q13.4 (Fig. 6).

Phylogenetic Analysis of the Expanded Siglec Family

Phylogenetic analysis of the previously existing members of the Siglec family, as well as the novel Siglecs identified throughout this work (SLG-S and -L) was accomplished through the Phylip software package, using the UPGMA method. As is evident in Fig. 7, the novel Siglecs we have identified are tightly clustered among the other previously known members of

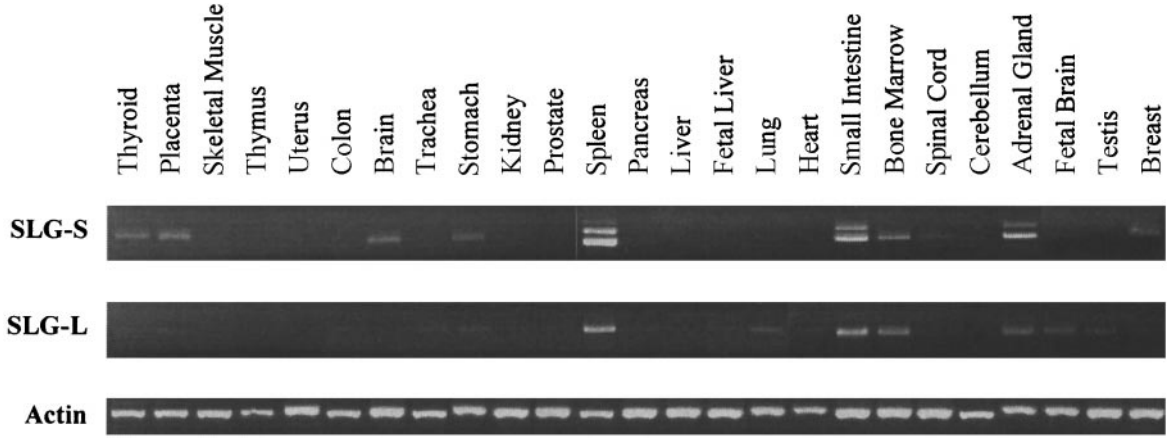


FIG. 5. Tissue expression profile of both SLG splice forms. RT-PCR was performed on 25 tissue total RNAs, for both SLG-S and SLG-L, as well as actin (control gene). SLG-S was found to be highly expressed in spleen, small intestine, and adrenal gland. SLG-L, on the other hand, was found to be highly expressed in spleen, small intestine, and bone marrow. Relatively lower levels of expression were found for both these splice forms in various tissues. The molecular weights of the expected cDNA bands were 991 and 1345 bp, for SLG-S and SLG-L, respectively. The observed bands closely matched these molecular weights. The multiple bands observed for SLG-S in spleen, small intestine, and adrenal gland were found to be of identical sequence, following purification and automated DNA sequencing. Some very faint bands may have been lost during reproduction of this figure.

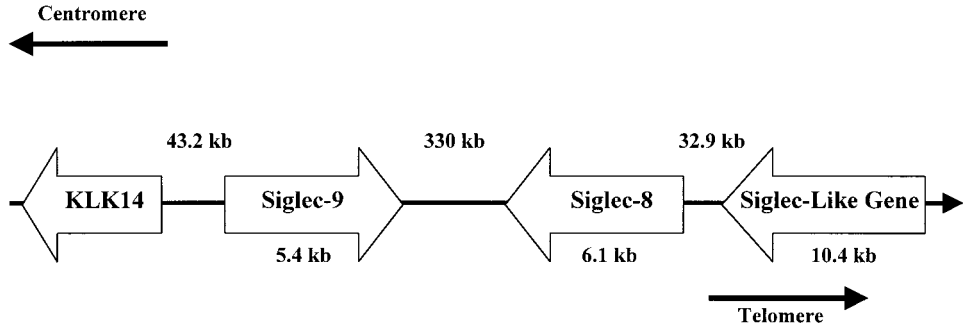


FIG. 6. Physical map of chromosome 19q13.4 and the Siglec-3-like subgroup locus. Shown schematically above is the developing physical map for the Siglec-3-like subgroup of Siglecs. Gene lengths are presented below each arrow, and distances between genes are shown above. Arrows denote the direction of transcription.

the Siglec-3-like subgroup. In particular, the novel Siglecs we have identified are closely related to Siglec-7, as well as the previously reported Siglec-8, while showing a somewhat more distant relationship with Siglec-3, -5, and -6.

DISCUSSION

Through our investigations of chromosome 19q13.4 and attempts to identify novel members of the Siglec-3-like subgroup of Siglecs, we have identified and characterized a novel gene encoding two putative Siglec splice variants, designated Siglec-like gene (SLG)-Short (SLG-S) and -Long (SLG-L). This novel Siglec gene was localized to chromosome 19q13.4, 32.9 kb downstream of the Siglec-8 gene, and approximately 370 kb telomeric to the Siglec-9 gene. The SLG gene we have characterized is comprised of 8 exons, with 7 intervening introns. All intron/exon splice sites are consistent with the consensus sequence for splice donor/acceptor sites (-m**GT**AAGT . . . C**AG**m-, where m is any base) (33). The SLG-S mRNA transcript consists of exons 2 through 8, while the SLG-L transcript contains exons 1 through 8, with part of exon 2 being removed through splicing. The first exon of each of the two isoforms contains a 5′ untranslated region of at least 20 bp, while the last exon possesses a 282 bp 3′ untranslated region. The putative coding sequence of the SLG-S transcript consists of 1736 nucleotides, which encode for a 477 amino acid protein with a predicted molecular mass of 51.7 kDa. The putative translation initiation codon is consistent with the Kozak consensus sequence (34). The SLG-L transcript, on the other hand, has a 2070 bp coding sequence, producing a putative protein product of 595 amino acids with a predicted molecular weight of 65 kDa.

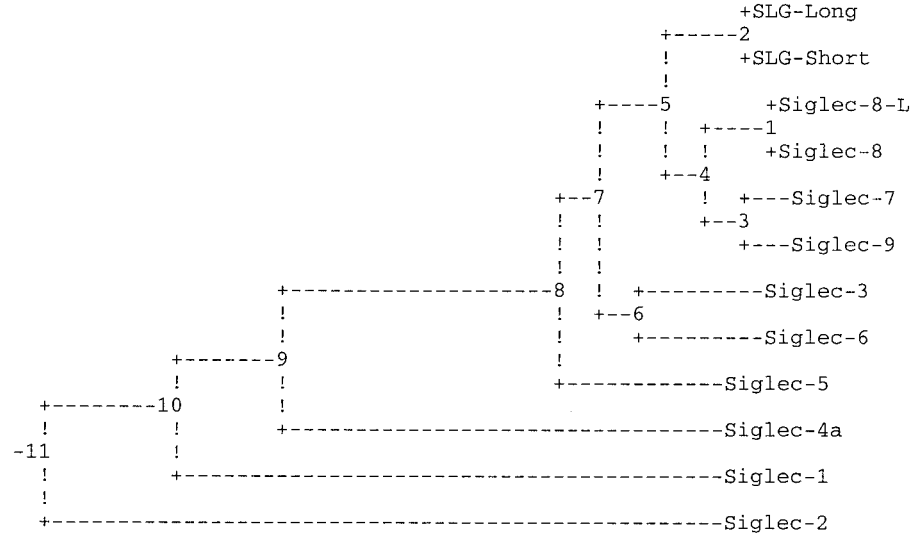


FIG. 7. Phylogenetic analysis of the Siglec family. Phylogenetic analysis was performed for all the known members of the Siglec family along with those we identified during these investigations. As seen by others, the Siglec-3-like subgroup exhibits a much closer relationship among themselves than they do to either of Siglec-4a, Siglec-1, or Siglec-2. Siglec-9, Siglec-8-L, SLG-S, and SLG-L all appear tightly clustered, along with Siglec-7 and Siglec-8, suggesting a much greater evolutionary relationship among these members.

In the case of the SLG-L transcript, unlike that of SLG-S, there is no identifiable Kozak consensus sequence. However, it is the only potential initiation codon which not only produces an extensive open reading frame, but also produces a signal peptide highly homologous to that seen in other Siglec-3-related Siglecs and at the same time maintains the reading frame seen in SLG-S through exon 2 and onwards. Interestingly, although the highly conserved purine at position −3 in the Kozak consensus sequence is absent, this has been noted in a few other vertebrate mRNA species as well (reviewed in 34). These mRNAs that lack the preferred nucleotide in both key positions, −4 and +3, encode potent regulatory proteins such as growth factors and cytokines. This suggests that the absence of strict adherence to the Kozak consensus sequence may provide a mechanism for modulating the yield of proteins which may be harmful if overproduced (34).

Examination of the tissue expression profile of both SLG transcripts revealed that SLG-S was relatively highly expressed in spleen, small intestine, and adrenal gland. On the other hand, the SLG-L transcript exhibited relatively high levels of expression in spleen, small intestine, and bone marrow. From this it is appears quite likely that both transcripts are expressed in tissues, particularly spleen and bone marrow, in which haemopoietic cell lineages are found. This is consistent with findings for other members of the Siglec-3-related subgroup of Siglecs, as well as that of other members of the Siglec family.

Based on examination of the homology between the putative protein products of both the SLG-S and -L transcripts with other known members of the Siglec-3-like subgroup, it is evident that this gene, and its two isoforms, likely represents the newest addition to the expanding Siglec-3-like subgroup of Siglecs. As is evident in Fig. 3, both SLG-S and SLG-L protein products contain many of the structural characteristics possessed by other Siglec-3-like Siglecs discovered thus far. They both contain the distinctive distribution of cysteine residues found in all Siglecs, necessary for the unique folding pattern of their Ig-like domains (1, 35). They both also contains the two highly conserved aromatic residues believed to be involved in sialic acid binding (36). The conserved arginine residue, which is present in all other Siglecs and believed to be essential for sialic acid binding, is replaced by glutamine in the case of the SLG-S protein, and cysteine in the SLG-L protein. Given that both conserved aromatic residues, found to be critical for sialic acid binding by Siglec-1 (37), were not so critical in the case of Siglec-9, it is unclear what effects the absence of this conserved arginine residue will have on sialic acid-dependent binding by these SLG isoforms. Studies are currently in progress to assess the effect these changes have on sialic acid binding.

A unique feature evident in the putative SLG-L protein is the presence of an additional N-terminal V-set Ig-like domain. This is the first report of a putative Siglec in which more than one N-terminal V-set domain has been observed. Interestingly, the current hypothesis regarding the origin of this family in humans is that it developed through the process of gene duplication and exon shuffling through the aid of chromosome 19q13.1-qter-specific minisatellites (15, 38). If this is indeed how this subgroup of Siglecs evolved, then it is possible that during a recombination event in which there was unequal exchange of genetic information in this region of 19q13.4, a Siglec-3-like gene with one N-terminal V-set domain was inserted downstream of the first exon of another Siglec-3-like gene. This would result in generation of a Siglec-3-related gene capable of producing two mRNA transcripts, the first of which would maintain the typical arrangement of Ig-like domains, as in the SLG-S transcript. However, an additional transcript would also be possible, in which transcription was initiated from the pre-existing first exon, yielding a mRNA encoding two N-terminal V-set Ig-like domains followed by two C2-set Ig-like domains, as seen in the SLG-L transcript.

The cytoplasmic tyrosine-based motifs, ITIM and SLAM-like, found in all other members of the Siglec-3-like subgroup of Siglecs, with the exception of Siglec-8, are also present in our putative members of this subgroup, SLG-S and SLG-L. These motifs have been the focus of investigations in order to elucidate the functional role these Siglecs play within the cell. The primary emphasis has been on the ITIM motif, which has been found to be involved in recruitment of the tyrosine phosphatases SHP1 and 2, and the inositol phosphatases SHIP1 and 2 (22–24). Siglec-7, originally identified as a natural killer cell inhibitory receptor, was found to recruit the tyrosine phosphatase SHP1 following tyrosine phosphorylation of its ITIM motif, leading to the inhibition of natural killer cell cytotoxicity (11). In addition, Siglec-3 has also been found to recruit SHP1 and 2, both *in vitro* and *in vivo,* as a result of phosphorylation of the tyrosine in its ITIM motif (25). Further, mutation of this tyrosine results in increased red blood cell binding by Siglec-3-expressing COS cells. More recently, it has been reported that engagement of Siglec-7 and Siglec-3 with monoclonal antibodies results in the inhibition of proliferation of both normal and leukemic myeloid cells *in vitro* (39). Although Siglec-7 was initially thought to be expressed exclusively in natural killer cells, it has also been found in myeloid cells, at a later stage of differentiation that Siglec-3. The observed inhibitory effects are believed to be the result of phosphorylation of the ITIM motif present in the cytoplasmic domains of both Siglec-3 and Siglec-7. These findings suggest that recruitment of SHP1 and SHP2 by members of the Siglec-3-like subgroup of Siglecs may serve to: (i) to

inhibit the activating signaling pathways that lead to cell proliferation and survival; and (ii) to modulate the receptor's ligand-binding activity (25).

The expression of Siglec-7 on myeloid cells raises the possibility that it too may represent a useful marker for accurate leukemic cell typing, in addition to Siglec-3, which is currently used in the diagnosis of the undifferentiated form of acute myelogenous leukemia (AML) (40–42). Monoclonal anti-Siglec-3 antibodies are already in use in phase I studies for the chemotherapeutic treatment of AML, and have shown selective ablation of malignant hematopoiesis (43, 44). The observed inhibitory effects of both Siglec-3 and Siglec-7 suggest that these molecules may represent useful targets for immunological antineoplastic therapy. By extension, given the extensive homology between both SLG splice variants and members of the Siglec-3-like subgroup of Siglecs, it is quite possible that SLG-S and SLG-L, putative Siglec-3-like Siglec, may also have potential as a therapeutic target for the treatment of hematological malignancies.

REFERENCES

1. 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.
2. 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) Siglecs: A family of sialic-acid binding lectins [letter]. *Glycobiology* **8,** v.
3. 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.
4. Stamenkovic, I., and Seed, B. (1990) The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion. *Nature* **345,** 74–77.
5. Simmons, D., and Seed, B. (1988) Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. *J. Immunol.* **141,** 2797–2800.
6. 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.
7. Kelm, S., Schauer, R., Manuguerra, J. C., Gross, H. J., and Crocker, P. R. (1994) Modifications of cell surface sialic acids modulate cell adhesion mediated by sialoadhesin and CD22. *Glycoconj. J.* **11,** 576–585.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Foussias, G., Yousef, G. M., and Diamandis, E. P. (2000) Identification and molecular characterization of a novel member of the Siglec family (SIGLEC9). *Genomics* **67,** 171–178.
15. Angata, T., and Varki, A. (2000) Cloning, characterization and phylogenetic analysis of Siglec-9, a new member of the CD33-related group of Siglecs. Evidence for co-evolution with sialic acid synthesis pathways. *J. Biol. Chem.* **275,** 22127–22135.
16. Zhang, J. Q., Nicoll, G., Jones, C., and Crocker, P. R. (2000) Siglec-9. A novel sialic acid binding member of the immunoglobulin superfamily expressed broadly on human blood leukocytes. *J. Biol. Chem.* **275,** 22121–22126.
17. Foussias, G., Yousef, G. M., and Diamandis, E. P. (2000) Molecular characterization of a Siglec8 variant containing cytoplasmic tyrosine-based motifs, and mapping of the Siglec8 gene. Submitted.
18. 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.
19. Vivier, E., and Daeron, M. (1997) Immunoreceptor tyrosine-based inhibition motifs. *Immunol. Today* **18,** 286–291.
20. 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., Pavitt, R., Wilkinson, J., Leversha, M., Huckle, E., Shaw-Smith, C. J., Dunham, A., Rhodes, S., Schuster, V., Porta, G., Yin, L., Serafini, P., Sylla, B., Zollo, M., Franco, B., Bentley, D. R., *et al.* (1998) Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene [see comments]. *Nat. Genet.* **20,** 129–135.
21. 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 [see comments]. *Nature* **395,** 462–469.
22. 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.
23. 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 [published erratum appears in *Eur. J. Immunol.* 1998 Mar; **28**(3): 1122]. *Eur. J. Immunol.* **28,** 264–276.
24. Muraille, E., Bruhns, P., Pesesse, X., Daeron, M., and Erneux, C. (2000) The SH2 domain containing inositol 5-phosphatase SHIP2 associates to the immunoreceptor tyrosine-based inhibi-

tion motif of Fc gammaRIIB in B cells under negative signaling. *Immunol. Lett.* **72,** 7–15.

25. Taylor, V. C., Buckley, C. D., Douglas, M., Cody, A. J., Simmons, D. L., and Freeman, S. D. (1999) The myeloid-specific sialic acid-binding receptor, CD33, associates with the protein-tyrosine phosphatases, SHP-1 and SHP-2. *J. Biol. Chem.* **274,** 11505–11512.
26. 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.
27. Murakami, K., and Takagi, T. (1998) Gene recognition by combination of several gene-finding programs. *Bioinformatics* **14,** 665–675.
28. 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.
29. Hofmann, K., and Stoffel, W. (1993) TMbase—A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **347,** 166.
30. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.
31. Yousef, G. M., Obiezu, C. V., Luo, L. Y., Black, M. H., and Diamandis, E. P. (1999) 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.
32. Ashworth, L. K., Batzer, M. A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J. A., Gordon, L. A., Lamerdin, J. E., Lennon, G., *et al.* (1995) An integrated metric physical map of human chromosome 19. *Nat. Genet.* **11,** 422–427.
33. 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.
34. Kozak, M. (1991) An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115,** 887–903.
35. 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.
36. 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.
37. 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 A resolution. *Mol. Cell.* **1,** 719–728.
38. Das, H. K., Jackson, C. L., Miller, D. A., Leff, T., and Breslow, J. L. (1987) The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J. Biol. Chem.* **262,** 4787–4793.
39. Vitale, C., Romagnani, C., Falco, M., Ponte, M., Vitale, M., Moretta, A., Bacigalupo, A., Moretta, L., and Mingari, M. C. (1999) Engagement of p75/AIRM1 or CD33 inhibits the proliferation of normal or leukemic myeloid cells. *Proc. Natl. Acad. Sci. USA* **96,** 15091–15096.
40. Bernstein, I. D., Singer, J. W., Smith, F. O., Andrews, R. G., Flowers, D. A., Petersens, J., Steinmann, L., Najfeld, V., Savage, D., Fruchtman, 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.
41. 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.
42. 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. *Leuk. Res.* **8,** 521–534.
43. 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.
44. 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.
45. Williams, A. F., and Barclay, A. N. (1988) The immunoglobulin superfamily—Domains for cell surface recognition. *Annu. Rev. Immunol.* **6,** 381–405.