THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 270, No. 27, Issue of July 7, pp. 16470–16475, 1995
Printed in U.S.A.

Genomic Organization and Chromosomal Localization of the Gene Encoding Human P-selectin Glycoprotein Ligand*

(Received for publication, March 3, 1995, and in revised form, May 1, 1995)

Geertruida M. Veldman‡§, Kevin M. Bean‡, Dale A. Cumming‡, Roger L. Eddy¶, Sheila N. J. Sait¶, and Thomas B. Shows¶

From the ‡Small Molecule Drug Discovery Group, Genetics Institute, Cambridge, Massachusetts 02140 and the ¶Department of Human Genetics, Roswell Park Cancer Institute, Buffalo, New York 14263

The gene for P-selectin glycoprotein ligand (PSGL-1) has been cloned from a human placenta genomic DNA library. A single intron of approximately 9 kilobases was found in the 5′-untranslated region and the complete coding region resides in exon 2. The genomic clone differs from the cDNA clone isolated from HL-60 cells in that it encodes an extra copy of the decameric repeat located in the extracellular domain of PSGL-1. Further analysis indicated that the PSGL-1 genes of HL-60 and U-937 cells contain 15 repeats, whereas the PSGL-1 genes of polymorphonuclear leukocytes, monocytes, and several other cell lines contain 16 repeats. Transfection experiments did not indicate a functional difference between these two variants of PSGL-1. The two previously observed PSGL-1 mRNA species of 2.5 and 4 kilobases most likely arise from differential utilization of polyadenylation signal sequences. The organization of the PSGL-1 gene closely resembles those of CD43 and human platelet glycoprotein GPIbα, both of which have an intron in the 5′-noncoding region, a long second exon containing the complete coding region, and TATA-less promoters. The gene for human PSGL-1, which has been designated *SELPLG* by the Human Gene Nomenclature Committee, was mapped to chromosome 12q24 using Southern blot analysis of DNA from a set of human-mouse cell hybrids, and fluorescent *in situ* hybridization on metaphase chromosome spreads.

P-selectin glycoprotein ligand (PSGL-1)[1] is the high affinity counter-receptor for P-selectin on myeloid cells and stimulated T lymphocytes[2] (1–5). As such it plays a critical role in the tethering of these cells to activated platelets or endothelia expressing P-selectin (5). A cDNA clone for human PSGL-1 has been isolated from HL-60 cells and encodes a unique, mucin-like glycoprotein (1).

To date little is known about the organization and attributes of the human PSGL-1 gene (*SELPLG*). In a previous study, two PSGL-1 mRNA transcripts (a predominant species of approximately 2.5 kb and a minor 4-kb species) were identified in PMNs and HL-60 cells (1). Furthermore, Southern blot analysis indicated that the PSGL-1 gene is present as a single copy in the human genome, suggesting that the two mRNAs are alternative transcripts derived from the same gene, resulting from either alternative splicing events or differential utilization of promoter and/or polyadenylation signal sequences (1).

The presence of multiple mRNA transcripts for PSGL-1 raises questions about the organization of the PSGL-1 gene, in particular with respect to the number and placement of introns. Although the coding regions of the majority of eukaryotic genes are interrupted by introns, a number of integral membrane proteins lack introns in their coding region. In this regard, the glycoproteins genes can be divided into two classes. The first class represents genes whose coding region is contained on one long exon. Examples of this class are human and mouse CD43 and human GPIbα, which also share several other common features such as a promoter lacking the typical TATA box (6–8). The second class represents genes whose coding region consists of multiple exons, and the promoters of these genes generally (albeit with a few exceptions) resemble the typical eukaryotic promoter containing a TATA consensus sequence. Examples of this class include glycophorin, CD4, and the L-selectin ligands GlyCAM-1 and CD34 (9–12).

Another potentially interesting question concerns the chromosomal localization of the PSGL-1 gene. The genes for P-selectin, E-selectin, and L-selectin as well as the L-selectin ligand CD34 all map to chromosome 1 (13, 14). The clustering of the selectins and their counter-receptor(s) to the same chromosome has obvious evolutionary implications and therefore it is of interest to ascertain whether PSGL-1 also maps to chromosome 1. Here we report the isolation, genomic organization, and chromosomal mapping of the human PSGL-1 gene.

EXPERIMENTAL PROCEDURES

Cells and Cell Lines—The tumor cell lines HL-60, U-937, THP-1, Ramos, CEM, MOLT-4, and SB were obtained from the American Type Culture Collection and maintained in RPMI 1640 containing 10% fetal calf serum (Sigma). Human polymorphonuclear leukocytes (PMNs) were isolated from whole blood by centrifugation over a Mono-Poly Resolving Medium Ficoll-Hypaque density gradient (ICN Biomedicals, Costa Mesa, CA). Peripheral blood mononuclear cells were isolated from heparinized blood of normal donors by centrifugation over a Ficoll/Hypaque density gradient (Sigma), and monocytes were subsequently isolated by adherence to plastic.

Isolation and Sequencing of Genomic Clones Encoding Human PSGL-1—A human placenta genomic library in the λ Fix II vector (Stratagene, La Jolla, CA) was plated at ~20,000 plaques/plate, and nitrocellulose replicas of the plates were hybridized according to standard procedures (15). The PSGL-1 cDNA probe comprising nucleotides (nt) 60–389, was amplified by polymerase chain reaction (PCR) and labeled to a specific activity of 5×10^8 cpm/μg with [^{32}P]dCTP (Amersham Corp.) using the random-primed DNA labeling kit (Boehringer

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) U25955 and U25956.

§ To whom correspondence should be addressed. Tel.: 617-876-1170; Fax: 617-498-8993.

[1] The abbreviations used are: PSGL-1, P-selectin glycoprotein ligand-1; PMN, polymorphonuclear leukocyte; kb, kilobase(s); *SELPLG*, gene symbol for PSGL-1; nt, nucleotides; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription PCR; bp, base pair(s).

[2] G. Vachino, X.-J. Chang, G. M. Veldman, R. Kumar, L. A. Fouser, M. C. Berndt, and D. A. Cumming, manuscript in preparation.

Mannheim). Filters were hybridized for 16 h at 65 °C with the labeled probe in 6 × SSC, 10 × Denhardt's solution, 0.1% SDS, and 50 μg/ml denatured, sheared salmon sperm DNA. The filters were washed three times for 30 min in 0.3 × SSC, 0.1% SDS at 65 °C and exposed to film. Duplicate positive plaques were rescreened until purified plaques were obtained, and the λ DNAs were prepared using a Qiagen lambda kit (Qiagen, Chatworth, CA). Ten independent λ clones were obtained and rescreened with PSGL-1 sequence-specific oligonucleotide probes, which were end-labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. Clones λ2 and λ4 were characterized by restriction mapping, and suitable restriction fragments were subcloned into the pBluescript vector (Stratagene). Subclones containing PSGL-1 gene exons and flanking regions were sequenced using the Applied Biosystems *Taq* DyeDeoxy Terminator Kit and ABI 373 autosequencer (16). DNA sequences were analyzed using the Sequencher software package (Gene Codes, Ann Arbor, MI).

RNA Isolation and Northern Blot Analysis—Total RNA was purified using the guanidine isothiocyanate/cesium chloride cushion method (15). Poly(A$^+$) mRNA was purified using the PolyATtract® mRNA isolation system II (Promega, Madison, WI). RNA was fractionated by 1.0% formaldehyde-agarose gel electrophoresis and transferred to nitrocellulose via capillary blotting. The PSGL-1 cDNA or gene fragments to be used as probes were amplified by PCR and labeled to a specific activity of at least 1 × 10^9 cpm/μg with [^{32}P]dATP and [^{32}P]dCTP and hybridized to filters as described above.

Rapid Amplification of cDNA Ends (RACE)—Rapid amplification of cDNA 5′ and 3′ ends was performed as described by Innis *et al.* (17). Briefly, three primers were used for the 3′-RACE experiment. The dT_{17}-adapter, a 35-mer containing an oligo(dT) stretch (17 residues) linked to a unique 18-nt adapter, was used to prime reverse transcription of total RNA isolated from PMNs, using the AMV Reverse Transcriptase System (Life Technologies, Inc.). The 18-nt adapter primer and a PSGL-1-specific primer, a 19-mer corresponding to nt 1564–1582, were subsequently used to amplify the cDNA by PCR. PCR products were analyzed on a 1% agarose gel. For 5′-RACE experiments, HL-60 poly(A$^+$) RNA and PMN total RNA were reverse-transcribed using a PSGL-1-specific 18-mer (antisense to nt 294–311 in exon 2). The first-strand reaction products were extended with dATP and terminal transferase and subsequently amplified by PCR using a mixture of the dT_{17}-adapter and the 18-nt adapter as the 5′ primers and a PSGL-1-specific 21-mer (antisense to nt 243–263) as the 3′ primer. The PCR products were analyzed on a 1.5% agarose gel and identified by Southern hybridization with a PSGL-1 oligonucleotide probe corresponding to nt 1–14 in exon 1.

Primer Extension Mapping—Primer extension was done according to described methods (15). In brief, a 30-mer 5′-ACCACCGTGCTCAGCAGAGCATGGGACAGC-3′ (antisense to nt 27–56 in exon 1) was end-labeled to a specific activity of 8 × 10^6 cpm/pmol as described above. Approximately 5 ng of labeled primer was hybridized to 3 μg of poly(A$^+$) RNA (HL-60) or 3 μg total RNA (PMNs) at 48 °C overnight. The extension reaction was done using the AMV Reverse Transcriptase System (Life Technologies, Inc.) with 20 units of AMV reverse transcriptase (Life Sciences, St. Petersburg, FL) for 2 h at 48 °C. The extension products were sized by electrophoresis on a denaturing 12% polyacrylamide gel using Polyoma/*Dde*I fragments as size markers.

PCR Analysis of the Repeat Region—The number of repeats in PSGL-1 from different cell types was determined by PCR using primers that flank the repeat region. The 5′ primer corresponds to nt 381–400, and the 3′ primer is antisense to nt 880–903 of the PSGL-1 gene, resulting in PCR products of 522 bp and 492 bp, representing 16 and 15 copies of the repeat, respectively. mRNA templates were analyzed by reverse transcription PCR (RT-PCR), using the reverse transcriptase reaction described above to generate a cDNA template for PCR amplification. The PCR reactions contained ~1 ng of cDNA or genomic DNA template and were performed according to standard procedures (17) using *Taq* DNA polymerase (Promega). The PCR products were sized by electrophoresis on a 2.5% NuSieve (FMC Bioproducts, Rockland, ME), 0.5% agarose gel.

Chromosomal Mapping of the PSGL-1 Gene (SELPLG)—Mapping of the human PSGL-1 gene was achieved by probing Southern blots of DNAs from 29 mouse-human somatic cell hybrids that were made from 17 unrelated human cell lines and 4 mouse cell lines (18–20). The hybrids were characterized by karyotypic analysis and by mapped enzyme markers (18, 20). Chromosome translations with no intact chromosome present were not tabulated for the percent discordance. Cell hybrid DNA was digested with *Xba*I and subjected to Southern analysis using the ^{32}P-labeled PSGL-1 cDNA probe comprising nt 60–389 (21). Control DNA from human and mouse gave a single human *Xba*I hybridizing fragment of 8 kb, with no cross-hybridization with mouse DNA.

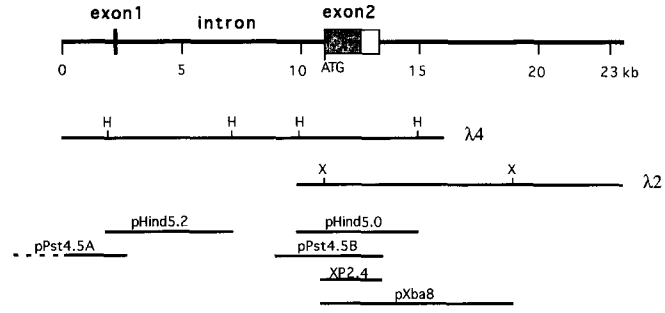

FIG. 1. **Organization of the human PSGL-1 gene.** Exon 1 and 2 are represented by *boxes*, the *shaded area* represents the coding sequence starting at the ATG codon in exon 2, and the *open areas* represent the 5′- and 3′-untranslated regions. The intron and flanking sequences are shown by *thick horizontal lines*. Two overlapping genomic clones λ4 and λ2 are shown with restriction maps for *Hin*dIII (*H*) and *Xba*I (*X*), respectively. The subclones pHind5.2, pHind5.0, pPst4.5A (containing sequences derived from the λ4 clone and the λ phage arm), pPst4.5B, and pXba8 were generated. The 2.4-kb fragment XP2.4, which contains exon 2, is indicated.

In situ fluorescent hybridization on metaphase chromosome spreads was done according to the method of Trask (22). The PSGL-1 gene fragment XP2.4 was labeled with digoxygenin-11-dUTP (Boehringer Mannheim) by random priming (23). Digital images were obtained using a Nikon epifluorescence microscope coupled to a cooled CCD camera (Photometrics). The images were merged and enhanced using the IP Lab Spectrum image analysis software supported on an Apple Macintosh IIci.

RESULTS

Isolation and Genomic Structure of the Human PSGL-1 Gene—The PSGL-1 gene was isolated by hybridization of a human placenta genomic library using a ^{32}P-labeled cDNA probe comprising nucleotides 60–389 of PSGL-1 (1). Ten independent genomic clones were isolated after screening approximately 10^6 phages. Hybridization of these 10 clones with a ^{32}P-labeled 14-mer derived from the 5′ end of the PSGL-1 cDNA revealed that only two clones were positive, suggesting that a large intron interrupted the PSGL-1 gene separating the sequences defined by the two probes. Two overlapping clones, denoted λ2 and λ4, were found to contain inserts of approximately 13 and 16 kb, respectively, and were selected for further fine mapping of the gene. The λ2 clone, which hybridized only to the PSGL-1 probe used in the primary screen, and the λ4 clone, which hybridized with both probes, were digested with *Hin*dIII, *Pst*I, or *Xba*I and subcloned into the pBluescriptII vector. By this procedure several subclones, designated pPst4.5A, pPst4.5B, pHind5.0, pHind5.2, and pXba8, were generated (Fig. 1). Restriction enzyme mapping and hybridization of the generated fragments with oligonucleotides derived from the 5′ terminus (nt 1–14) and 3′ terminus (nt 1636–1649) of the PSGL-1 cDNA clone further established the genomic map for human PSGL-1 as presented in Fig. 1. Indeed, the PSGL-1 gene contains an approximately 9-kb intron inserted in the 5′-noncoding region between nt 54 and nt 55, based on the numbering system of the original PSGL-1 cDNA clone (1). The coding region which starts at nt 60 is fully encoded by exon 2 and does not contain any introns (Fig. 2).

Correct splicing of the intron and linking of exons 1 and 2, as observed in the PSGL-1 cDNA clone derived from an HL-60 library, also occurs in RNA from different sources. RT-PCR on freshly isolated human PMN, monocyte, and HL-60 RNA with oligonucleotides spanning the intron yielded the expected PCR product in all three cases (data not shown).

Preliminary Analysis of the Transcription Initiation Region of the Human PSGL-1 Gene—The putative transcription initi-

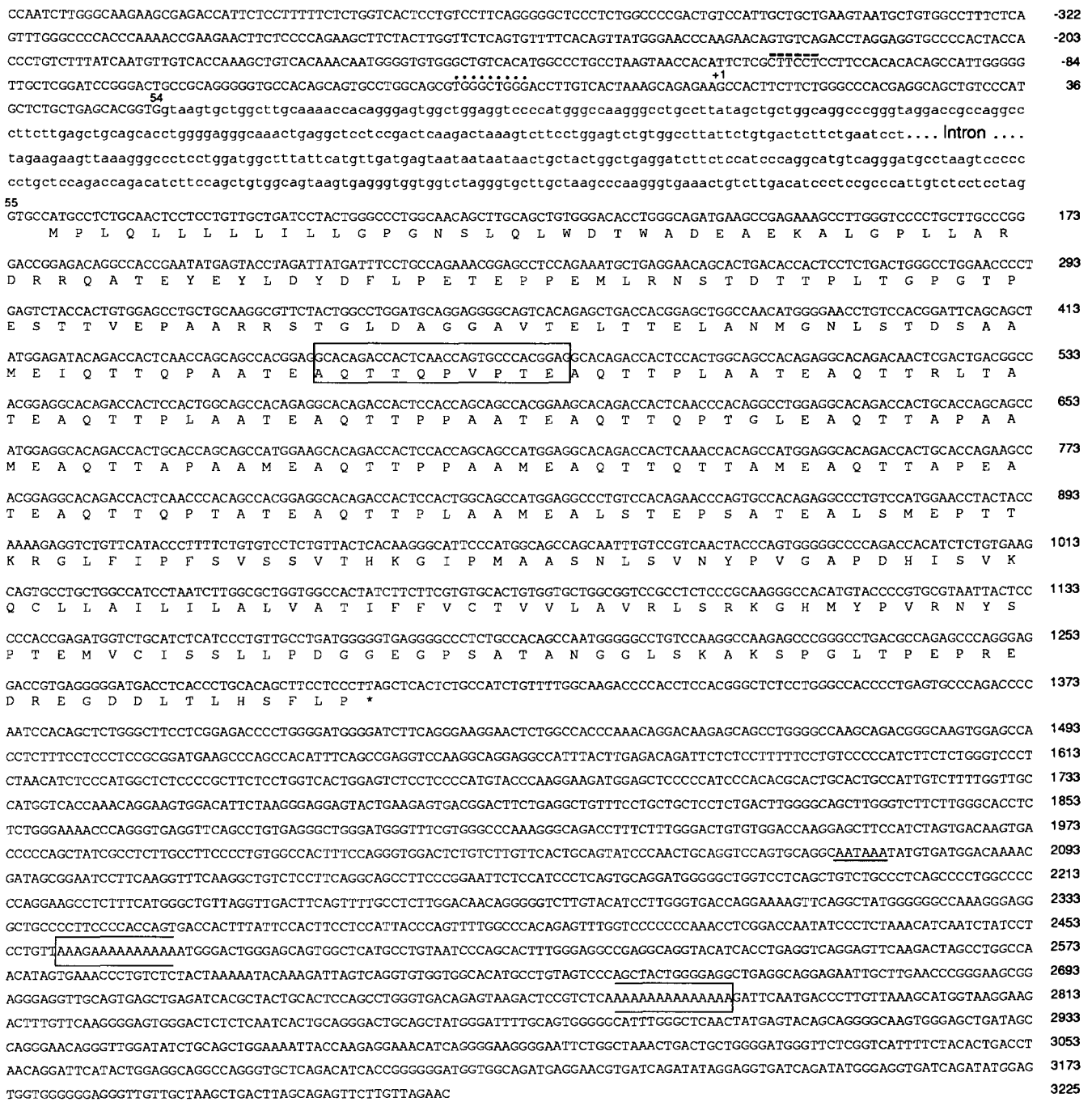

FIG. 2. **Nucleotide sequence of the human PSGL-1 gene.** Position +1 corresponds to the first nucleotide of the PSGL-1 cDNA (Sako). The sequences shown include 440 nt of 5′-flanking sequence, exon 1 (nt 1–54), part of the 9-kb intron (*lowercase letters*), exon 2 (nt 55–2075, including the complete coding region), and additional 3′-flanking sequences. The amino acid sequence is shown in the one-letter code. The potential binding sites for Ets (*dashed line*) and Sp1 (····) are indicated. The extra repeat encoding 10 amino acids is boxed. The major polyadenylation signal is underlined and the *Alu* sequence is indicated between brackets.

ation region of the PSGL-1 gene was defined by 5′-RACE and by primer extension on RNA isolated from human PMNs and HL-60 cells. 5′-RACE was performed with a primer complementary to nt 243–263 in exon 2 as described under "Experimental Procedures." Hybridization of the PCR products with an oligonucleotide probe derived from exon 1 identified a broad band of 280–340 bp from PMN as well as HL-60 RNA templates (data not shown). Considering that the requisite dT_{17} adapter is 35 nt long, this tentatively places the transcription initiation region between −42 and +18. Primer extension analysis of human PMN and HL-60 RNA was performed using a 30-mer complementary to nt 27–56 in exon 1. Several primer extension products were observed including two predominant bands between 55 and 70 nt in length, suggesting initiation from multiple sites in the region from −14 to +1 (Fig. 3). The region immediately upstream from these transcriptional start sites contains no TATA or CAAT box, but is GC-rich and contains putative binding sites for the regulatory elements Sp1 and Ets (Fig. 2).

Analysis of the 3′ Termini of Human PSGL-1 mRNAs—Two PSGL-1 transcripts, a predominant mRNA of approximately 2.5 kb and a minor 4-kb mRNA, have previously been identified in PMNs and HL-60 cells (1). The 3′-noncoding region of the PSGL-1 gene contains a polyadenylation signal at position

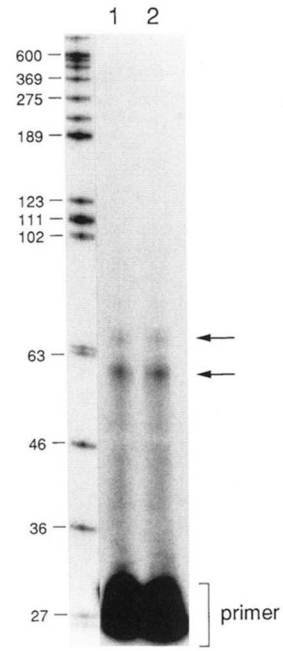

FIG. 3. **Primer extension analysis of the transcription start site(s) of the PSGL-1 gene.** The 5′-end-labeled primer complementary to nt 27–56 in exon 1 was hybridized to PMN total RNA (*lane 1*) and HL-60 poly(A+) RNA (*lane 2*) and extended with reverse transcriptase as described under "Experimental Procedures." The extended products were analyzed on a 12% polyacrylamide sequencing gel with Polyoma/*Dde*I fragments as size markers. Major bands are indicated by *arrows*.

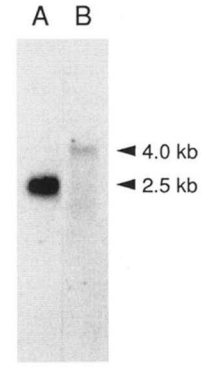

FIG. 4. **Northern blot analysis of PSGL-1 mRNAs.** Northern blots containing 3 μg of HL-60 poly(A^+) RNA/lane were hybridized under high stringency conditions with a ^{32}P-labeled PSGL-1 cDNA probe comprising nt 60–389 (*A*) or a PSGL-1 genomic DNA probe comprising nt 2081–2460 (*B*). The major 2.5-kb and minor 4-kb PSGL-1 mRNAs are indicated.

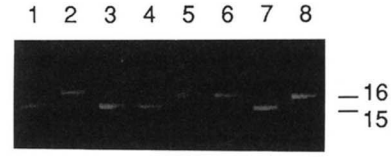

FIG. 5. **PCR analysis of the repeat region in PSGL-1.** The repeat region of PSGL-1 was amplified by RT-PCR of mRNA templates or PCR of cloned and genomic DNA templates as described under "Experimental Procedures." The templates analyzed include: cDNA clone pPL85 (*lane 1*), genomic clone λ4 (*lane 2*), HL-60 RNA (*lane 3*), HL-60 genomic DNA (*lane 4*), human PMN RNA (*lane 5*), human monocyte RNA (*lane 6*), U-937 RNA (*lane 7*), and THP-1 RNA (*lane 8*). The PCR products representing 16 or 15 copies of the repeat sequence were separated by agarose gel electrophoresis.

2075 as indicated in Fig. 2. Mapping of the PSGL-1 mRNA species from PMNs by 3′-RACE with an oligonucleotide corresponding to nt 1564–1582 identified a predominant PCR product of approximately 520 bp, indicating that the major PSGL-1 mRNA species utilizes this polyadenylation signal sequence (data not shown).

Additional mapping of the minor 4-kb mRNA species was performed by Northern blot analysis of HL-60 poly(A^+) RNA with two different PSGL-1 probes. Fig. 4 shows that the PSGL-1 cDNA probe, comprising nt 60–389, hybridized with the 2.5-kb mRNA and also (albeit very weakly in this exposure) with the 4-kb mRNA (*lane A*). The genomic DNA probe, comprising nt 2080–2459 located 3′ of the polyadenylation signal, only hybridized to the 4-kb mRNA (*lane B*), suggesting that the 4-kb mRNA utilizes a downstream, yet unidentified, polyadenylation signal sequence. The region of the PSGL-1 gene that is unique for the 4-kb transcript contains an *Alu* sequence (24) as shown in Fig. 2.

An Extra Repeat Unit is Encoded in the PSGL-1 Gene of Many Cell Types—The PSGL-1 cDNA clone (pPL85) derived from an HL-60 cDNA library encodes 15 consecutive repeats of a 10-amino acid consensus sequence (1). Our genomic clone derived from human placenta, however, contains an extra sequence of 30 nucleotides (*boxed* in Fig. 2) located in the second repeat unit and thus encodes 16 such repeats. RT-PCR analysis with primers spanning the repeat region allowed for rapid analysis of the number of repeats in PSGL-1 mRNA from different cell types. Similarly, the number of repeats in PSGL-1 genomic DNA from different sources was determined by PCR analysis. Fig. 5 shows that freshly isolated HL-60 RNA contains 15 repeats (*lane 3*) similar to the original PSGL-1 cDNA clone pPL85 (*lane 1*), eliminating the possibility that the cDNA clone had lost one repeat due to a cloning artifact. Identical results were obtained for RNA isolated from the monocytic cell line U-937 (*lane 7*). In contrast, 16 repeats were found in the genomic clone λ4 (*lane 2*) and in mRNA isolated from human PMNs (*lane 5*) and monocytes (*lane 6*) and several human cell lines, such as THP-1 (*lane 8*) and Ramos, CEM, MOLT-4, and SB (not shown). To investigate whether splicing was responsible for this phenomenon, genomic DNA isolated from HL-60 cells was analyzed by PCR. Fig. 5 (*lane 4*) shows that the HL-60 PSGL-1 gene encodes just 15 repeats. Similarly, the PSGL-1 gene of U-937 cells only encodes 15 repeats (data not shown). Sequence analysis of the repeat region in the PSGL-1 gene from U-937 cells confirmed that the U-937 and HL-60 PSGL-1 genes are identical and are therefore lacking the same repeat unit.

There does not appear to be a direct functional role for the extra decameric repeat. The extra repeat was introduced into the PSGL-1 cDNA expression vector pPL85, generating pPL85-R16. COS cells cotransfected with vectors encoding the α(1,3/1,4)-fucosyltransferase gene, and pPL85 or pPL85-R16 bound equally well to P-selectin *in vitro* (data not shown).

Chromosomal Localization of the Human PSGL-1 Gene (SELPLG)—Human chromosomal localization of the PSGL-1 gene was performed in a two-step procedure employing analysis of a human-mouse somatic cell hybrid panel and chromosomal fluorescent *in situ* hybridization. Southern analysis of *Xba*I-digested DNA from 29 human-mouse hybrids with the human PSGL-1 cDNA probe comprising nt 60–389 detected a single 8-kb fragment in human DNA. This human PSGL-1 probe did not cross-hybridize with mouse DNA. The results in Table I show that only chromosome 12 gave 0% discordance, indicating that *SELPLG* localizes to chromosome 12.

Fluorescent *in situ* hybridization on metaphase chromosome spreads was done with a 2.4-kb PSGL-1 genomic DNA probe (XP2.4 in Fig. 1) and the results are shown in Fig. 6. The double fluorescent signals on both chromosomes were found only at 12q24 in 17 out of 20 metaphase spreads (85%) examined and on no other chromosome. Three out of 20 spreads showed double fluorescent signals on only one chromosome. These results localize *SELPLG* to 12q24.

TABLE I

Segregation of PSGL-1 with human chromosomes in XbaI-digested human-mouse cell hybrid DNA

The table is compiled from Southern analysis of 29 mouse-human hybrids as described under "Experimental Procedures." Scoring was determined by the presence (+) or absence (−) of a human PSGL-1 band on the blots compared with the presence or absence of human chromosomes in each hybrid. A 0% discordance indicates matched segregation of the probe with a human chromosome.

Chromosome	Concordant no. of hybrids		Discordant no. of hybrids		Discordance
	(+/+)	(−/−)	(+/−)	(−/+)	
					%
1	6	10	8	1	36
2	7	7	9	5	50
3	12	7	4	5	32
4	9	9	8	3	38
5	12	6	5	6	38
6	10	7	7	5	41
7	10	3	6	9	54
8	13	7	4	5	31
9	2	11	12	1	50
10	16	4	1	8	31
11	6	3	9	8	65
12	17	12	0	0	0
13	9	8	8	4	41
14	13	4	4	7	39
15	12	7	4	5	32
16	9	10	8	2	34
17	15	4	1	7	30
18	13	8	4	4	28
19	3	9	14	3	59
20	11	5	6	7	45
21	13	4	4	8	41
22	6	7	11	4	54
X	8	6	6	5	44

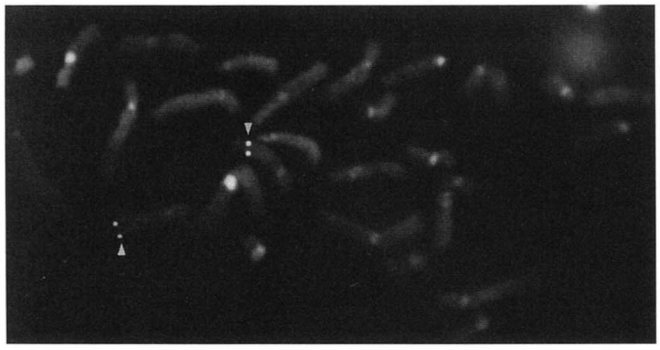

FIG. 6. **Mapping of the human PSGL-1 gene by *in situ* hybridization.** Metaphase chromosome spreads were hybridized *in situ* with the genomic PSGL-1 probe XP2.4. A partial spread is illustrated. Double fluorescent signals are indicated by *arrows*.

DISCUSSION

A single 23-kb human PSGL-1 gene was isolated on two overlapping clones derived from a human placenta genomic library. No clones were isolated from any additional genomic loci, consistent with the existence of a single human PSGL-1 gene and in agreement with previously described Southern hybridization analysis of human genomic DNA (1). The human PSGL-1 gene contains a single intron of approximately 9 kb located 5 bp upstream of the ATG initiation codon, and the complete coding region is contained within exon 2. In contrast, many members of the mucin gene family, such as glycophorin, episialin, or MUC-1, and the L-selectin ligands CD34 and GlyCAM-1, are encoded by multiple exons (9, 11, 12, 25). Although unusual, the lack of introns in the coding region has been reported for other integral membrane proteins, such as the human and mouse sialomucin CD43 and the human blood platelet membrane glycoprotein GPIbα (6–8). These genes, like the PSGL-1 gene, lack introns in the coding region yet also have an intron in the 5′-untranslated region.

Analysis of possible transcription initiation sites also indicated features shared between the PSGL-1 gene and the genes for CD43 and GPIbα. Primer extension and 5′-RACE experiments were employed to identify the transcription initiation sites of the PSGL-1 gene in PMNs and HL-60 cells. These preliminary analyses indicated that transcription may initiate from multiple sites just upstream of exon 1. Our data does not exclude the possibility that transcription of the PSGL-1 gene originates from other promoters as well, since those 5′-RACE PCR products would fail to hybridize with the exon 1 probe. A complete understanding of possible other promoter usage will require sequence analysis and mapping of cloned RACE PCR products. The PSGL-1 promoter region tentatively identified in this study lacks consensus sequences such as TATA and CAAT, which are general components of eucaryotic promoters. Several features of the PSGL-1 promoter, *e.g.* absence of TATA and CAAT boxes, multiple transcription initiation sites, high G + C content (55%), and a potential Sp1 binding site, resemble those of "housekeeping" genes (26). However, several tissue-specific promoters that lack TATA and CAAT boxes, but are nonetheless regulated, have been described recently, including the promoters for human GPIbα and CD43 (8, 27, 28). Other examples of TATA-less promoters include those for the integrins CD11a and CD11b and for CD4 (29–31). Potential binding sites for transcriptional activators, such as Sp1, PU.1, Ets, and AP-2, have been identified in these promoter regions and are thought to direct tissue-specific gene expression (30, 31). The PSGL-1 promoter has potential binding sites for Sp1 and Ets and therefore may very well fall in this category of promoters (32, 33). Indeed PSGL-1 expression appears to be regulated in a cell type-specific manner, since PSGL-1 transcripts have been found in myeloid and lymphocytic cells[2] but not in the hepatoblastoma cell line HepG2 (1) or the human lung fibroblast line MRC-5.[3] This is consistent with observations by others that cell surface PSGL-1 can be detected on leukocytes, but not on fibroblasts and keratinocytes (5, 34).

The two previously identified PSGL-1 mRNA transcripts of 2.5 and 4 kb appear to arise from the single gene by differential utilization of polyadenylation sequences. Analysis of the 3′ end

[3] G. M. Veldman and G. Vachino, unpublished observation.

of PSGL-1 mRNAs indicated that the major mRNA species utilizes the polyadenylation signal at position 2075. This finding (together with the 5′ start mapping data) would predict a mRNA species of approximately 2.1 kb, not including the poly(A) tail. This is in agreement with the major PSGL-1 mRNA on Northern blots which was sized at approximately 2.5 kb (1). The minor 4-kb PSGL-1 mRNA hybridized to sequences located 3′ of the polyadenylation signal at position 2075, suggesting that the size difference of the two mRNA species may primarily arise from utilization of different polyadenylation sequences. A similar situation has been reported for the 1.9- and 4.3-kb mRNA species observed for CD43 (35).

The PSGL-1 gene in HL-60 and U-937 cells encodes 15 repeats of a 10-amino acid consensus sequence (1). This is in contrast to our human placenta genomic PSGL-1 clone, as well as the PSGL-1 in all other cells and cell lines tested, including the recently reported leukocyte PSGL-1 cDNA (5), which contain 16 repeats. However, the number of consensus repeats does not appear to correlate with PSGL-1 function. HL-60 and U-937 cells bind P-selectin as do freshly isolated PMNs, monocytes, and THP-1 cells, all of which contain 16 copies of the repeat. In contrast, the lymphocytic cell lines Ramos, CEM, MOLT-4, and SB, which also express PSGL-1 with 16 repeats, do not bind P-selectin.[2] Moreover, COS cells cotransfected with α(1,3/1,4)-fucosyltransferase and recombinant PSGL-1 with either 15 or 16 copies of the repeat exhibited indistinguishable binding to P-selectin. The PSGL-1 gene in the promyelocytic leukemia cell line HL-60 and the histiocytic lymphoma cell line U-937 are lacking the same copy of the 10-amino acid repeat in both alleles, yet these cell lines are unrelated. It is not clear how this loss of a repeat has occurred in the PSGL-1 genes in both cell lines, although this may have been caused by some form of polymorphic variation. Other examples of polymorphic and allelic variations in the repeat regions of mucin-type glycoproteins, such as the carcinoma-associated mucin episialin and human intestinal mucin, have been reported (36, 37).

Mapping of the PSGL-1 gene to chromosome 12 prompted a search of the genome data base for colocalized disease markers or genes. Although the genes for interferon-γ, protein phosphatase 1, and aldehyde dehydrogenase 2 also map to 12q24, the search did not reveal any readily apparent phenotypes that could be attributed to abnormal PSGL-1 expression (21, 38, 39). The chromosomal loci for several other mucin-type glycoproteins and cell adhesion molecules have been mapped and appear to be clustered. In particular, the genes that encode the selectin family of leukocyte adhesion molecules are clustered on chromosome 1 (13). The gene for P-selectin (*SELP*) maps to chromosome 1q21–24, as do genes for L-selectin (*SELL*), E-selectin (*SELE*), and episialin (13, 40). In addition, CD34, a ligand for L-selectin, also maps to chromosome 1q (14). The PSGL-1 gene, however, does not map to this cluster of selectin and selectin ligand genes, and neither is the PSGL-1 gene linked to the organizationally related genes for CD43 and GPIbα. The gene for human CD43 is located on chromosome 16p11.2, close to the genes encoding the α subunits of the leukocyte integrins LFA-1, Mac-1, and p150,95, and human GPIbα maps to chromosome 17p12-ter (29, 41, 42).

In summary, the organization of the PSGL-1 gene is strikingly similar to those of CD43 and platelet glycoprotein GPIbα. Each of these genes contains a single intron in the 5′-untranslated region and a long exon containing the complete coding region. The promoter regions lack the typical TATA and CAAT elements, and the genes are flanked by *Alu* sequences. Although PSGL-1, CD43, and GPIbα are all sialic acid-carrying *O*-glycosylated proteins, their amino acid sequences are unrelated and their genes do not appear to be linked in the genome.

Acknowledgments—We thank Gloria Vachino for preparing RNA samples, Dave Merberg for computer analysis of the DNA sequence, and Ron Kriz and Ravi Kumar for critical reading of the manuscript.

REFERENCES

1. Sako, D., Chang, X.-J., Barone, K. M., Vachino, G., White, H. M., Shaw, G., Veldman, G. M., Bean, K., Ahern, T. J., Furie, B., Cumming, D., and Larsen, G. R. (1993) *Cell* **75,** 1179–1186
2. Moore, K. L., Stults, N. L., Diaz, S., Smith, D. F., Cummings, R. D., Varki, A., and McEver, R. P. (1992) *J. Cell Biol.* **118,** 445–456
3. Norgard, K. E., Moore, K. L., Diaz, S., Stults, N. L., Ushiyama, S., McEver, R. P., Cummings, R. D., and Varki, A. (1993) *J. Biol. Chem.* **268,** 12764–12774
4. Moore, K. L., Eaton, S. F., Lyons, D. E., Lichenstein, H. S., Cummings, R. D., and McEver, R. P. (1994) *J. Biol. Chem.* **269,** 23318–23327
5. Moore, K. L., Patel, K. D., Bruehl, R. E., Fugang, L., Johnson, D. A., Lichenstein, H. S., Cummings, R. D., Bainton, D. F., and McEver, R. P. (1995) *J. Cell Biol.* **128,** 661–671
6. Shelley, C. E., Remold-O'Donnell, E., Rosen, F. S., and Whitehead, A. (1990) *Biochem. J.* **270,** 569–576
7. Cyster, J., Somoza, C., Killeen, N., and Williams, A. (1990) *Eur. J. Immunol.* **20,** 875–881
8. Wenger, R. H., Kieffer, N., Wicki, A. N., and Clemetson, K. J. (1988) *Biochem. Biophys. Res. Commun.* **156,** 389–395
9. Kudo, S., and Fukuda, M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 4619–4623
10. Siu, G., Wurster, A. L., Lipsick, J. S., and Hedrick, S. M. (1992) *Mol. Cell. Biol.* **12,** 1592–1604
11. Dowbenko, D., Andalibi, A., Young, P. E., Lusos, A. J., and Lasky, L. A. (1993) *J. Biol. Chem.* **268,** 4525–4529
12. Brown, J., Greaves, M., and Molgaard, H. (1991) *Int. Immunol.* **3,** 175–184
13. Watson, M. L., Kingsmore, S. F., Johnston, G. I., Siegelman, M. H., Le Beau, M. M., Lemons, R. S., Bora, N. S., Howard, T. A., Weissman, I. L., McEver, R. P., and Seldin, M. F. (1990) *J. Exp. Med.* **172,** 263–272
14. Molgaard, H., Spurr, N., and Greaves, M. (1989) *Leukemia* **3,** 773–778
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
16. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986) *Nature* **321,** 674–679
17. Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J. (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA
18. Shows, T. B., Sakaguchi, A. Y., and Naylor, S. L. (1982) in *Advances in Human Genetics* (Harris, H., and Hirschhorn, K., eds) Vol. 12, pp. 341–352, Plenum Press, New York
19. Shows, T., Eddy, R., Haley, L., Byers, M., Henry, M., Fujita, T., Matsui, H., and Taniguchi, T. (1984) *Somatic Cell Mol. Gen.* **10,** 315–318
20. Shows, T. B. (1983) in *Isozymes: Current Topics in Biological and Medical Research* (Rattazzi, M. C., Scandalios, J. G., and Whitt, G. S., eds) Vol. 10, pp. 323–339, Alan R. Liss, New York
21. Naylor, S. L., Sakaguchi, A. Y., Shows, T. B., Law, M. L., Goeddel, D. V., and Gray, P. W. (1983) *J. Exp. Med.* **157,** 1020–1027
22. Trask, B. (1992) in *Methods in Cell Biology: A Laboratory Guide*, Vol. 35, pp. 1–35, Academic Press, New York
23. Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
24. Batzer, M. A., and Deininger, P. L. (1991) *Genomics* **9,** 481–487
25. Lancaster, C., Peat, N., Duhig, T., Wilson, D., Taylor-Papadimitriou, J., and Gendler, S. (1990) *Biochem. Biophys. Res. Commun.* **173,** 1019–1029
26. Dynan, W. S. (1986) *Trends Genet.* **2,** 196–197
27. Wenger, R. H., Wicki, A. N., Kieffer, N., Adolph, S., Hameister, H., and Clemetson, K. J. (1989) *Gene (Amst.)* **85,** 517–524
28. Kudo, S., and Fukuda, M. (1991) *J. Biol. Chem.* **266,** 8483–8489
29. Cornwell, R. D., Gollahon, K. A., and Hickstein, D. D. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4221–4225
30. Hickstein, D. D., Baker, D. M., Gollahon, K. A., and Back, A. L. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 2105–2109
31. Salmon, P., Giovane, A., Wasylyk, B., and Klatzmann, D. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 7739–7743
32. Faisst, S., and Meyer, S. (1992) *Nucleic Acids Res.* **20,** 3–26
33. Wasylyk, B., Hahn, S. L., and Giovane, A. (1993) *Eur. J. Biochem.* **211,** 7–18
34. Alon, R., Rossiter, H., Wang, X., Springer, T., and Kupper, T. S. (1994) *J. Cell Biol.* **127,** 1485–1495
35. Shelley, C. S., Remold-O'Donnell, E., Davis, A. E., III, Bruns, G. A. P., Rosen, F. S., Carroll, M. C., and Whitehead, A. S. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2819–2823
36. Ligtenberg, M. J. L., Vos, H. L., Gennissen, A. M. C., and Hilkens, J. (1990) *J. Biol. Chem.* **265,** 5573–5578
37. Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamport, D. T. A., and Kim, Y. S. (1989) *J. Biol. Chem.* **264,** 6480–6487
38. Barker, H. M., Craig, S. P., Spurr, N. K., and Cohen, P. T. W. (1993) *Biochim. Biophys. Acta* **1178,** 228–233
39. Hsu, L. C., Yoshida, A., and Mohandas, T. (1986) *Am. J. Hum. Genet.* **38,** 641–648
40. Swallow, D., Gendler, S., Griffiths, B., Kearney, A., Povey, S., Sheer, D., Palmer, R., and Taylor-Papadimitriou, J. (1987) *Ann. Hum. Genet.* **51,** 289–294
41. Pallant, A., Eskenazi, A., Mattei, M.-G., Fournier, R. E. K., Carlsson, S. R., Fukuda, M., and Frelinger, J. G. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1328–1332
42. Corbi, A. L., Larson, R. S., Kishimoto, T. K., Springer, T. A., and Morton, C. C. (1988) *J. Exp. Med.* **167,** 1597–1607