RAPID COMMUNICATION

Organization of the Gene for Human Platelet/Endothelial Cell Adhesion Molecule-1 Shows Alternatively Spliced Isoforms and a Functionally Complex Cytoplasmic Domain

By Nancy E. Kirschbaum, Richard J. Gumin, and Peter J. Newman

Platelet endothelial cell adhesion molecule-1 (PECAM-1) is a cell-cell adhesion molecule that is expressed on circulating platelets, on leukocytes, and at the intercellular junctions of vascular endothelial cells and mediates the interactions of these cells during the process of transendothelial cell migration. The cDNA for PECAM-1 encodes an open reading frame of 738 amino acids (aa) that is organized into a 27-aa signal peptide, a 574-aa extracellular domain composed of 6 Ig homology units, and a relatively long cytoplasmic tail of 118 aa containing multiple sites for posttranslational modification and postreceptor signal transduction. To provide a molecular basis for the precise evaluation of the structure and function of this transmembrane glycoprotein, we have determined the organization of the human PECAM-1 gene. The PECAM-1 gene, which has been localized to human chromosome 17, is a single-copy gene of approximately 65 kb in length and is broken into 16 exons by introns ranging in size from 86 to greater than 12,000 bp in length. Typical of other members of the Ig superfamily, each of the extracellular Ig homology domains is encoded by a separate exon, consistent with PECAM-1 having arisen by gene duplication and exon shuffling of ancestral Ig superfamily genes. However, the cytoplasmic domain was found to be surprisingly complex, being encoded by seven short exons that may represent discrete functional entities. Alternative splicing of the cytoplasmic tail appears to generate multiple PECAM-1 isoforms that may regulate phosphorylation, cytoskeletal association, and affinity modulation of the mature protein. Finally, a processed pseudogene having 76% identity with PECAM-1 cDNA was identified and localized to human chromosome 3. These findings should have important implications for structure/function analysis of PECAM-1 and its role in vascular adhesive interactions.

© 1994 by The American Society of Hematology.

From the Blood Research Institute, The Blood Center of Southeastern Wisconsin, Milwaukee, WI; and The Departments of Cellular Biology and Pharmacology, The Medical College of Wisconsin, Milwaukee, WI.

Submitted July 14, 1994; accepted September 20, 1994.

Supported by Grant No. HL-40926 (to P.J.N.) and Training Grant No. HL-07902 (to N.E.K.) from the National Institutes of Health.

Address reprint requests to Peter J. Newman, PhD, Blood Research Institute, The Blood Center of Southeastern Wisconsin, 1701 W Wisconsin Ave, Milwaukee, WI 53233.

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

© 1994 by The American Society of Hematology.

0006-4971/94/8412-0120$3.00/0

by polymerase chain reaction (PCR)-based screening using PECAM-1–specific primers. All inserts were characterized by restriction endonuclease mapping. Restriction fragments containing exons were identified by Southern blot hybridization of restriction endonuclease digests with 32P-labeled PECAM-1 cDNA or oligonucleotide probes. Genomic inserts were subcloned as fragments into plasmid vectors, pzl8r (Stratagene, La Jolla, CA) or pGEM-7 (Promega, Madison, WI), for further gene mapping and direct sequence analysis. All exons, exon/intron junctions, and the majority of intronic sequence were determined by a combination of manual and automated cycle sequencing according to the method of Sanger et al. 5

Genomic Southern blot hybridization. Human genomic DNA was isolated from peripheral blood leukocytes separated from 50 mL of human blood drawn from a normal, healthy volunteer. Ten micrograms of human genomic DNA was digested with various restriction endonucleases, blotted onto nylon membranes, and hybridized with either 32P-labeled PECAM-1 cDNA or oligonucleotide probes. Genomic inserts were subcloned as fragments into plasmid vectors, pzl8r (Stratagene, La Jolla, CA) or pGEM-7 (Promega, Madison, WI), for further gene mapping and direct sequence analysis.

Fig 1. Schematic representation of the organization of the human PECAM-1 gene. Genomic clones encompassing all coding regions of the PECAM-1 gene are positioned relative to the structure of the entire gene and the PECAM-1 protein. Restriction maps for EcoRI, Xba I, and Sac I sites are shown. The position and relative size of each coding exon is illustrated by solid boxes. Noncoding regions are illustrated by hatching, and dashed borders indicate that exon borders have not yet been determined. Arrows point directly from coding exons to corresponding regions of the PECAM-1 protein.

PECAM-1 protein structure

<table>
<thead>
<tr>
<th>SP</th>
<th>Ig1</th>
<th>Ig2</th>
<th>Ig3</th>
<th>Ig4</th>
<th>Ig5</th>
<th>Ig6</th>
<th>TM</th>
<th>Cyto</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Genomic clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

Restriction sites

- EcoRI
- Xba I
- Sac I

Exon/intron organization

PECAM-1 mRNA splicing variants. Total RNA was isolated from human umbilical vein endothelial cells by the method of Chomczynski and Sacchi. cDNA was generated using the antisense primer 5'-TGCTGTGTTCTGTGGGAG-3', corresponding to nucleotides 2322-2324 of the PECAM-1 cDNA, using MMLV reverse transcriptase (Boehringer Mannheim Biochemicals, Indianapolis, IN). PCR amplification of cDNA used a forward primer in exon 11 with the sequence, 5'-TGCTGTGTTCTGTGGGAG-3', corresponding to nucleotides 2085-2102, and an antisense primer spanned the exon 15/16 junction having the sequence, 5'-GGAGCCCTTCCGTTCTAGAGT-3', corresponding to nucleotides 2341-2322, of the PECAM-1 cDNA. PCR products were separated by 2% agarose gel electrophoresis. On occasion, PCR products were excised directly from the gel, subcloned into pGEM-5 (Promega), and sequenced as described above.

Identification of PECAM-1 mRNA splicing variants. Total RNA was isolated from human umbilical vein endothelial cells by the method of Chomczynski and Sacchi. cDNA was generated using the antisense primer 5'-TGCTGTGTTCTGTGGGAG-3', corresponding to nucleotides 2482-2465, of the PECAM-1 cDNA, using MMLV reverse transcriptase (Boehringer Mannheim Biochemicals, Indianapolis, IN). PCR amplification of cDNA used a forward primer in exon 11 with the sequence, 5'-TGCTGTGTTCTGTGGGAG-3', corresponding to nucleotides 2085-2102, and an antisense primer spanning the exon 15/16 junction having the sequence, 5'-GGAGCCCTTCCGTTCTAGAGT-3', corresponding to nucleotides 2341-2322, of the PECAM-1 cDNA. PCR products were separated by 2% agarose gel electrophoresis. On occasion, PCR products were excised directly from the gel, subcloned into pGEM-5 (Promega), and sequenced as described above.

Chromosomal localization analysis of human/hamster somatic cell hybrid clones. A chromosome panel consisting of genomic DNA derived from well-characterized human-hamster hybrids was obtained from BIOS, Inc (New Haven, CT). Seventeen different DNA samples, which together encompass the full complement of human chromosomes, were screened by PCR analysis according to the manufacturer's directions. The PECAM-1-specific primer pair consisted of a forward primer located in exon 10 (cDNA sequence 5'-2038-2056-3') and a reverse primer located in exon 11 encompassing...
Fig 2. Nucleotide sequence of the human PECAM-1 gene. The nucleotide sequence of all 16 exons and flanking intronic sequences are shown. The numbering of the sequence reflects only those basespairs actually shown in the figure. Nucleotides within exons are displayed in uppercase letters, and corresponding amino acids are shown above each codon. For personal use only. www.bloodjournal.org From
Table 1. Exon Organization of the Human PECAM-1 Gene

<table>
<thead>
<tr>
<th>Exon</th>
<th>Nucleotide Boundaries (bp)</th>
<th>Size (bp)</th>
<th>Phase</th>
<th>Protein Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>493-760</td>
<td>268</td>
<td>1</td>
<td>5' UT, Signal Peptide</td>
</tr>
<tr>
<td>2</td>
<td>847-873</td>
<td>27</td>
<td>1</td>
<td>Signal Peptide, N-Term.</td>
</tr>
<tr>
<td>3</td>
<td>926-1219</td>
<td>294</td>
<td>1</td>
<td>Ig domain 1</td>
</tr>
<tr>
<td>4</td>
<td>1270-1575</td>
<td>276</td>
<td>1</td>
<td>Ig domain 2</td>
</tr>
<tr>
<td>5</td>
<td>1626-1901</td>
<td>294</td>
<td>1</td>
<td>Ig domain 3</td>
</tr>
<tr>
<td>6</td>
<td>1952-2200</td>
<td>249</td>
<td>1</td>
<td>Ig domain 4</td>
</tr>
<tr>
<td>7</td>
<td>2250-2525</td>
<td>276</td>
<td>1</td>
<td>Ig domain 5</td>
</tr>
<tr>
<td>8</td>
<td>2576-2863</td>
<td>288</td>
<td>1</td>
<td>Ig domain 6</td>
</tr>
<tr>
<td>9</td>
<td>2914-3021</td>
<td>108</td>
<td>1</td>
<td>Transmembrane domain</td>
</tr>
<tr>
<td>10</td>
<td>3072-3099</td>
<td>28</td>
<td>2</td>
<td>Cyto 1</td>
</tr>
<tr>
<td>11</td>
<td>3150-3223</td>
<td>74</td>
<td>1</td>
<td>Cyto 2</td>
</tr>
<tr>
<td>12</td>
<td>3274-3327</td>
<td>54</td>
<td>1</td>
<td>Cyto 3</td>
</tr>
<tr>
<td>13</td>
<td>3378-3440</td>
<td>63</td>
<td>1</td>
<td>Cyto 4</td>
</tr>
<tr>
<td>14</td>
<td>3491-3547</td>
<td>57</td>
<td>1</td>
<td>Cyto 5</td>
</tr>
<tr>
<td>15</td>
<td>3598-3620</td>
<td>23</td>
<td>0</td>
<td>Cyto 6</td>
</tr>
<tr>
<td>16</td>
<td>3671-4571</td>
<td>&gt;900</td>
<td></td>
<td>Cyto 7, 3'UT</td>
</tr>
</tbody>
</table>

nucleotides 5'-2077-2059-3'. This primer pair yields a 1.1-kb product that spans intron X (see Fig 2). PECAM-1 pseudogene-specific primers used were 5'-ATGGTCTGAAGGTGGAACTAAC-3' (forward), corresponding to nucleotides 63-84 of the pseudogene sequence (see Fig 6), and 5'-TTCAAGTTTTTCAGCTGTGAAG-3' (reverse), corresponding to nucleotides 594-573 of the PECAM-1 pseudogene. All amplifications were conducted using Taq polymerase at 2 U/mL, with annealing temperatures between 58°C and 60°C.

RESULTS

Organization of the human PECAM-1 gene. To determine the organization of the gene encoding human PECAM-1, two different λ phage libraries and one PI phagemid library were screened using a combination of PCR amplification (PI phagemid library) and hybridization with PECAM-1-specific probes (phage libraries). As shown in Fig 1, six genomic clones, having inserts averaging approximately 15 kb in size, were obtained and characterized. Restriction mapping of these six clones (Fig 1) suggested that the PECAM-1 gene encompassed 60 to 70 kb. From these genomic clones we determined the nucleotide sequence of 30,127 bp of the PECAM-1 gene, including 561 bp of sequence that is 5' from the previously reported PECAM-1 cDNA.7 The sequence of the PECAM-1 gene has been submitted to the GenBank, and will be available under accession numbers L34631-L34657.

Figures 1 and 2 summarize the major features of the PECAM-1 gene. The gene is composed of 16 exons separated by introns ranging from 86 bases to greater than 12,000 nucleotides in length. Exon 1 encodes the 5' untranslated (UT) region of the PECAM-1 cDNA and includes most, but not all, of the signal peptide. Exon 2, which resides in close proximity to exon 1 on the gene, is only 27 bp in length, and contains the remainder of the signal peptide and the first three amino acids of the predicted N-terminus of the mature protein. Thereafter, there is a direct correlation between exon/intron organization and the structure of the extracellular domain of the PECAM-1 protein. Similar to other members of the Ig superfamily, each of the six Ig homology domains of PECAM-1 is encoded by its own exon, numbered 3 through 8. The transmembrane domain and its flanking sequences are also encoded by a separate exon, exon 9. Somewhat unexpectedly, we found that the cytoplasmic tail of PECAM-1 is encoded by seven individual exons, numbered 5' to 3'...
10 through 16, which appear to encode distinct functional domains (see below). Each of these exons are small, ranging in size from only 23 to 74 bp, with the exception of exon 16, which encodes the carboxy-terminus of PECAM-I as well as the 3' UT of the PECAM-I mRNA transcript. Examination of 871 bp 3' to the stop codon for PECAM-I failed to show a consensus primary AATAAA polyadenylation sequence. However, two secondary consensus sequences, GATAAA and AATACA, were noted 330 and 486 bases, respectively, after the stop codon. The size, phase, and relative positions of the exons that comprise the PECAM-I gene are summarized in Table 1.

Primer extension analysis was conducted to localize the 5' end of the PECAM-I mRNA transcript. An antisense oligonucleotide complementary to the 5' region of the PECAM-I mRNA was used to prime reverse transcription of HUVECs, A549, and yeast RNAs. As shown in Fig 3, a specific band unique to HUVEC mRNA corresponding to an "A" at position 493 (see arrow in Fig 2) was obtained. This base is 204 bp upstream from the translation start site. 5' RACE PCR products encompassing the 5' end of PECAM-I mRNA were also generated, and sequence analysis of several of these products showed three additional transcription start sites, all within eight nucleotides of the major primer extension product. Localization of the transcription start site to this region is consistent with the findings of Zehnder et al., who reported a cDNA clone containing a 207-bp 5' untranslated region. Thus, it would appear that PECAM-I mRNA transcription initiates at one of several closely spaced nucleotides, similar to that found for several other vascular cell adhesion molecule genes, including those encoding E-selectin and L-selectin.

Examination of the sequence immediately 5' to the transcription initiation site (Fig 2) showed no consensus TATA or CAAT elements, indicating that the PECAM-I gene is one of a growing number of genes lacking these promoter elements. The 5' flanking sequence does contain several common consensus cis-acting elements, including a consensus GATA motif (AGATA) at position −24, ets sites at positions −28, −48, −192, and an inverted NF-κB site at position −227. A sequence element (CCTGGGA) common to some acute phase reactants is also present at position −240.

Alternative splicing of the PECAM-I gene. Sequences surrounding the splice junctions of each exon are provided in Fig 2. As shown, each intron begins with the consensus splice donor sequence "gt" and ends with the consensus "ag" splice acceptor sequence. Interestingly, all exons with the exceptions of exons 10 and 15 terminate with the first base of the codon, thus having the classification of phase 1 exons (Table 1). In particular, five of seven exons coding for the transmembrane and cytoplasmic domains are of the phase 1 class, leading to the possibility of in-frame alternative splicing events yielding PECAM-I isoforms differing in the sequence of the cytoplasmic tail. To examine whether such isoforms might be generated within the cell, HUVEC mRNA was subjected to RT-PCR amplification of a region encompassing exons 11 through 16. As shown in Fig 4A, two PCR products were visualized by agarose gel electrophoresis: a 260-bp major product, corresponding to a full-length segment containing exons 11 through 16, and a minor 200-bp product that hybridized with a full-length PECAM-I cDNA probe (not shown). When the minor mRNA-derived PCR product was subcloned and sequenced, it was found to be missing exon 14 (Fig 4B). Evidence for the generation of four additional alternatively spliced PECAM-I isoforms differing in their cytoplasmic domains has also been recently obtained (see Discussion).

Identification and chromosomal localization of a processed pseudogene for PECAM-1. Southern blot analysis of EcoRI- or HindIII-digested genomic DNA showed a simple pattern of bands that added up to approximately 55 kb when hybridized with a full-length PECAM-1 cDNA probe (Fig 5A). Together with the mapping and sequence analysis of genomic phage clones (see Figs 1 and 2), these data are consistent with the presence of a single PECAM-1 gene within the human genome. However, hybridization of genomic Southern blots of Sac I-, EcoRI-, or Xba I-digested genomic DNA with a probe encompassing only exon 3 (Fig 5B), which lacks restriction sites for these three enzymes, yielded both a major (*) and a minor (+) hybridizing band, suggesting the presence of an additional gene closely related to PECAM-1.
To identify the source of these minor bands, we characterized three human genomic clones previously found to be weakly reactive with PECAM-1 cDNA probes. Restriction analysis (not shown) of these clones using the same PECAM-1 exon 3-specific probe yielded weakly hybridizing Sac I, EcoRI, and Xba I fragments corresponding to the sizes of the minor bands shown in Fig 5B. A 1.8-kb EcoRI fragment common to all three clones was subsequently subcloned and partially sequenced. When aligned with PECAM-1 cDNA (Fig 6), a 666-bp region of this EcoRI fragment was found to share 76% identity with exons 2, 3, and 4 of the PECAM-1 gene. However, unlike the PECAM-1 gene, this closely related sequence contains no introns, splice junctions, or open reading frames, consistent with its identity as a processed pseudogene. Southern blot analysis of Sac I, EcoRI, or Xba I-digested genomic DNA using an oligonucleotide derived from the sequence of this pseudogene fragment yielded bands corresponding in size to the minor bands shown in Fig 5B (not shown), confirming their identity.

The gene for PECAM-1 has recently been localized to the long arm of chromosome 17.26 To determine whether the PECAM-1 pseudogene was physically close to the PECAM-1 gene, we screened a well-characterized human/hamster somatic cell hybrid panel using two different sets of PCR primers specific for the PECAM-1 pseudogene. As shown in Table 2, the PECAM-1 1–specific primers confirmed the presence of the PECAM-1 gene on chromosome 17, whereas PECAM-1 1–specific primers failed to amplify human PECAM-1 pseudogene localized unambiguously to human chromosome 3.

**DISCUSSION**

The purpose of the present investigation was to determine the functional organization of the human PECAM-1 gene. Six clones, encompassing all coding regions, were isolated and characterized, and the nucleotide sequence for nearly 30 kb of the PECAM-1 gene was determined, including all exon-intron boundaries. From these studies, we have found that human PECAM-1 is encoded by a single copy gene that is composed of 16 exons separated by introns ranging in size from 86 to more than 12,000 bp in length. Overall, the PECAM-1 gene encompasses more than 65 kb of DNA (Figs 1, 2, and 5), making it the largest gene thus far reported for an Ig superfamily cell adhesion molecule.

The Ig superfamily genes, like the selectin genes, seem to have been created from a mosaic of exons. Thus, the structure of the PECAM-1 gene also seems to support the concept that proteins uniquely tailored to perform specific tasks may have evolved by duplication, migration, and juxtaposition of separate exons coding for individual structural protein domains. In this regard, it is interesting to note that all Ig homology domains in the PECAM-1 gene (Table 1), as well as in other reported Ig superfamily genes that organize each Ig domain into a separate exon, end with the first nucleotide of the codon (ie, phase 1 exons). In addition, it has been noted that exons coding for lectin-binding, epidermal growth factor homology, and complement repeat structures in selectins and other homologous proteins also exist as phase 1 exons.

Somewhat uniquely, the modular structure of the PECAM-1 gene extends to the cytoplasmic region. Unlike the genes for the homologous proteins, intercellular adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule-1 (VCAM-1), which combine the coding regions for the transmembrane and cytoplasmic domains into a single exon, the cytoplasmic tail of PECAM-1 was found to be encoded by seven separate, small exons separated by long stretches of intron (Fig 7). To date, evidence has been obtained for the presence of at least five variant forms of PECAM-1 that are generated as a result of alternative splicing of the cytoplasmic domain exons: a Δ exon 14 form of
PECAM-1 (Fig 4): murine PECAM-1Δ12,15 and Δ14,15 isoforms expressed in developing cardiac endothelium; a soluble form of PECAM-1 that is generated as a result of splicing out of exon 9 which encodes the transmembrane domain (Fig 2); and a PECAM-1 Δ13 form originally identified in cDNA clones encoding human PECAM-1. These PECAM-1 isoforms and their cell biological properties are summarized in Table 3.

In line with its mosaic construction, each small cytoplasmic exon of PECAM-1 may represent a discrete functional domain. DeLisser et al have recently shown that deletion of a segment of the cytoplasmic domain encompassing exons 14-16 results in its conversion from a calcium-dependent, heterophilic adhesion molecule to a homophilic, calcium-independent species. Our finding of a naturally occurring PECAM-1 Δ14 transcript would suggest that cells may actually produce PECAM-1 isoforms that differ with regard to the adhesive properties of the extracellular domain. Affinity modulation of this sort is not unprecedented, because the cytoplasmic domains of both GPIIIb/IIa and CD44 have been shown to modulate the interaction of each receptor's extracellular domain with its respective ligand, directly showing the functional significance of the cytoplasmic domains for these receptors. Whether alternative splicing of PECAM-1 mRNA transcripts regulates PECAM-1/ligand interactions in a cell-type specific or developmentally regulated fashion remains an intriguing avenue of future investigation.

It is reasonable to speculate that PECAM-1 isoforms differing in their cytoplasmic tails may also differ in their ability to associate with the cytoskeleton and mediate signal transduction events. PECAM-1 in platelets, endothelial cells, and certain T-cell lines has been shown to be phosphorylated on serine residues in an activation-dependent manner. In this regard, it is interesting to note that PECAM-1 Δ13 lacks 4 of 12 serine residues that represent potential substrates for protein kinases (Table 3). Several cellular events have been correlated with PECAM-1 phosphorylation in different cell types. First, Zehnder et al have implicated phosphorylation of PECAM-1 in phytohemagglutinin-activated T cells as one mechanism by which PECAM-1 expression is downregulated. Second, Hillery and Newman have shown a correlation between PECAM-1 phosphorylation in activated platelets and its association with the cytoskeleton. An interesting
Table 2. Chromosomal Mapping of the Human PECAM-1 Pseudogene Using Human × Hamster Somatic Cell Hybrids

<table>
<thead>
<tr>
<th>Human × Hamster Clone*</th>
<th>Human Chromosome Content</th>
<th>PECAM-1 Gene</th>
<th>Pseudogene</th>
</tr>
</thead>
<tbody>
<tr>
<td>212</td>
<td>5, Y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>324</td>
<td>18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>423</td>
<td>3</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>683</td>
<td>5, 12, 14, 19, 21, 22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>734</td>
<td>5, 9, 18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>750</td>
<td>5, 13, 14, 15, 19</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>756</td>
<td>5, 6, 7, 12, 13, 14, 19, 20, 21, Y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>803</td>
<td>4, 5, 8, 11, 22, X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>811</td>
<td>6, 17, 18</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>852</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>860</td>
<td>3, 5, 6, 10, 19, 21</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>862</td>
<td>5, 9, 19</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>867</td>
<td>1, 5, 13, 14, 18, 19</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>904</td>
<td>5, 6, 12, 16, 21, Y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>937</td>
<td>1, 5, 14, 15, 17, 21</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>967</td>
<td>5, 8, 16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>968</td>
<td>5, 9, 13, X</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Genomic DNA obtained from BIOS, Inc. Two microliters of each sample were amplified using PECAM-1 or pseudogene-specific PCR primers, as described in the Materials and Methods.

† + and – denote the presence or absence of a specific, predicted PCR product as visualized on ethidium bromide-stained agarose gels.

A parallel can be drawn between the expression and function of membrane skeletal protein 4.1 and that of PECAM-1. Protein 4.1 is principally found in erythrocytes but has also been detected in liver, intestine, and lymphocytes. Tissue-specific structural variants have been described resulting from both differential glycosylation as well as from alternative splicing of RNA species. In particular, eight structural variants of a 10-kD domain in the cytoplasmic tail of protein 4.1 are generated in vivo by tissue-specific alternative splicing of three exons encoding 19, 14, or 21 aa. The 21 aa cassette, normally present in the erythrocyte form, but absent in other tissues, is necessary for the ability of protein 4.1 to interact with cytoskeletal proteins, spectrin and actin, thereby enabling protein 4.1 to function properly to maintain erythrocyte membrane strength. By comparison, one may envision that certain regions of the cytoplasmic tail, encoded by separate exons, are necessary for PECAM-1 to interact with cytoskeletal proteins, and this in turn may influence the ability of PECAM-1 to localize to the intercellular junctions and maintain the integrity of the vessel wall.

We have presented evidence for the existence of a pseudogene having 76% sequence identity with PECAM-1 cDNA (Fig 6). The region of the pseudogene that we characterized encompasses exons 2-4, but does not contain an open reading frame, and completely lacks intronic regions, consistent with its having been generated as a result of packaging by an RNA virus of a processed PECAM-1 mRNA transcript. Although we determined the nucleotide sequence for a 666-bp portion of the PECAM-1 pseudogene, it is likely to extend

Fig 7. Schematic representation of the genes encoding PECAM-1, VCAM-1, and ICAM-1. All three genes encode homologous cell adhesion molecules expressed on endothelial cells. The extracellular domains of each protein are encoded by a series of phase 1 Ig homology domains. In contrast, whereas the genes for VCAM-136 and ICAM-127 combine their transmembrane, cytoplasmic, and 3' untranslated regions into a single exon, the PECAM-1 gene is highly divided, separating these domains into seven short exons separated by long stretches of intron, making it the largest Ig superfamily gene reported to date.
much further, and may explain spurious PCR products obtained using primer pairs having homology with the PECAM-1 pseudogene sequence. Pseudogenes homologous to other genes, such as the von Willebrand factor gene and the human Ig γ heavy chain gene, have been reported, but their significance is unknown. In contrast to the pseudogene for the Ig γ heavy chain gene, whose locus is directly adjacent to other Ig γ genes, the pseudogene homologous to PECAM-1 is situated on a different chromosome from the gene. The human PECAM-1 gene has been localized to chromosome 17, whereas the pseudogene has been localized to chromosome 3 (Table 2).

The tissue distribution of PECAM-1 is unique in that its expression is highly restricted to cells of the vasculature, including platelets, endothelial cells, and most leukocyte subsets. Although we have not yet performed a functional analysis of the PECAM-1 promoter, a number of consensus sequences for cis-acting elements were found immediately upstream from the major transcription initiation site. No consensus TATA or CAAT elements are present within the first 120 upstream base pairs (Fig 1); however, this region does contain the initiator consensus, ATTTC N4-6 GCCA, similar to that found in several other promoters that lack the TATA recognition sequence for RNA polymerase II. A single GATA element, which has been shown to regulate gene expression in cells of the megakaryocytic lineage, is also present at position −24. Finally, three consensus sites, which have been shown in other systems to be recognized by the polyomavirus enhancer A-binding protein, were found at positions −28, −48, and −192, and an inverted NF-κB site was found at position −227. Whether this NF-κB site functions to regulate PECAM-1 in a cytokine-dependent manner is not yet known, but PECAM-1 expression has been shown to be upregulated by the cytokines, transforming growth factor-β and γ-interferon. It will be of future interest to determine how these and other as yet to be identified cis-acting elements within the PECAM-1 promoter function to regulate the expression of PECAM-1 in vascular cells.

Discovery of the human PECAM-1 gene structure has provided insight for identification of functional domains on the PECAM-1 protein, particularly within the cytoplasmic tail. Future investigations based on information derived from the gene structure, focusing on characterization of alternatively spliced PECAM-1 mRNA species and PECAM-1 isoforms will undoubtedly impact on our understanding of the role played by PECAM-1 in mediating the processes of hemostasis, inflammation, and immunity.

ACKNOWLEDGMENT

We are grateful to Drs Steven Albelda, Scott Baldwin, and Clayton Buck for making copies of their manuscripts identifying cytoplasmic domain deletion mutants and murine PECAM-1 isoforms available to us before their publication and for their helpful discussions. We also thank Dr Amy Goldberger for providing unpublished details of her characterization of mRNAs encoding soluble, transmembraneless PECAM-1 isoforms.

REFERENCES

40. Means AL, Farnham PJ: Transcription initiation sites from the dihydrofolate reductase promoter is positioned by HPI binding at the initiation site. Mol Cell Biol 10:653, 1990
Organization of the gene for human platelet/endothelial cell adhesion molecule-1 shows alternatively spliced isoforms and a functionally complex cytoplasmic domain

NE Kirschbaum, RJ Gumina and PJ Newman