Structure of the Human Gene Encoding Granule Membrane Protein-140, a Member of the Selectin Family of Adhesion Receptors for Leukocytes*

Geoffrey I. Johnston†, Greg A. Bliss, Peter J. Newman§, and Rodger P. McEver¶

From the Department of Medicine, St. Francis Medical Research Institute, University of Oklahoma Health Sciences Center and Cardiovascular Biology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104 and §The Blood Center of Southeastern Wisconsin, Milwaukee, Wisconsin 53233

GMP-140, an inducible granule membrane protein of platelets and endothelial cells, is a member of the selectin family of cell surface receptors that mediate interactions of leukocytes with the blood vessel wall. These molecules all contain an N-terminal lectin-like domain, followed by an epidermal growth factor-like domain, a variable number of consensus repeats related to those in complement-binding proteins, a transmembrane domain, and a cytoplasmic tail. Two variant cDNAs for GMP-140 have been identified, one predicting a soluble form of the molecule lacking the transmembrane domain and the other predicting a molecule containing eight instead of nine consensus repeats. Here we describe the organization of the human gene encoding GMP-140, which spans over 50 kilobase pairs and contains 17 exons. Almost all exons encode distinct structural domains, including the lectin-like domain, the epidermal growth factor-like domain, each of the nine consensus repeats, and the transmembrane region. Each of the two deletions found in the variant cDNAs is precisely encoded by an exon, suggesting that these forms of GMP-140 are derived from alternative splicing of mRNA. By using the polymerase chain reaction, transcripts encoding the putative soluble form of GMP-140 can be amplified from both platelet and endothelial cell RNA. The structure of the GMP-140 gene supports the concept that the selectins evolved as a result of exon duplication and rearrangement.

The selectins are structurally related cell surface receptors that mediate interactions of leukocytes with the blood vessel wall (1-4). Three members of this family are currently known. The first is the murine Mel14 antigen (3,4) or human LAM-1 (Leu-8/TQ1 antigen) (5-8). This molecule, which is found on neutrophils, monocytes, and a subset of lymphocytes, of peripheral lymph nodes and may also promote leukocyte adhesion to endothelium at inflammatory foci (9,10). The second selectin is ELAM-1, a human cytokine-inducible endothelial cell receptor for neutrophils (2). The third is GMP-140 (PADGEM protein, CD62), a membrane glycoprotein located in secretory granules of human platelets and endothelium (1). When these cells are activated by agonists such as thrombin, GMP-140 is rapidly redistributed to the cell surface (11-14) where it mediates adhesion of neutrophils and monocytes (15-17).

Each of the selectins contains an N-terminal domain homologous to Ca2+-dependent lectins, followed by an epidermal growth factor (EGF)-like domain, a variable number of repeating units similar to those in complement-binding proteins, a transmembrane segment, and a short cytoplasmic tail (1-4). The extensive sequence identity and shared domain organization of the selectins suggest that they comprise a gene family produced by duplication and rearrangement of ancestral exons. This hypothesis is strengthened by the tight clustering of all three genes on chromosome 1 in both mouse and man (18).

Variant cDNA clones encoding GMP-140 have been isolated, one predicting a molecule containing eight repeating units instead of the typical nine and the other predicting a soluble form in which the region encompassing the transmembrane domain is removed (1). These different cDNAs may account for the observation that GMP-140 precursors of slightly different apparent Mr, are synthesized by HEL cells, a human cell line with features of megakaryocytes (19).

In this study we describe the organization of the human gene encoding GMP-140. The intron-exon boundaries support the concept that GMP-140 is assembled from exons encoding structurally distinct domains. Furthermore, the variant forms of GMP-140 appear to be derived from alternative splicing of mRNA.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Genomic Clones—Three human genomic DNA libraries were screened: a lymphocyte library in the bacteriophage vector EMBL3 from Clontech Laboratories (Palo Alto, CA), a colonic mucosa library in EMBL3 provided by Dr. John Metter (University of Texas Health Science Center at San Antonio), and a lymphocyte library in the bacteriophage Charon 4A vector from the American Type Culture Collection (Rockville, MD). Between 1.4 and 3 million recombinant phages were plated out on LE292 bacteria at a density of 200,000 plaques/230-mm plate of NZCYM agar. The EMBL3 libraries were screened with full-length or partial 32P-labeled GMP-140 cDNA probes by previously described methods (1). To obtain missing regions of the gene, the Charon 4A library was screened with a labeled fragment encoding the EGF-like domain and

* This work was supported by Grants HL 34364 (to R. P. M.) and HL 40926 (to P. J. N.) from the National Institutes of Health. 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.

† Present address: Pharmaceutical Proteins, Ltd., School of Biochemistry, University of Birmingham, Birmingham B15 2TT, United Kingdom.

‡ Recipient of Research Career Development Award HL 01733 from the National Institutes of Health. To whom correspondence should be addressed: Dept. of Medicine, University of Oklahoma Health Sciences Ctr., 825 N. E. 13th St., Oklahoma City, OK 73104.

¶ The abbreviations used are: GMP-140, granule membrane protein-140; EGF, epidermal growth factor; bp, base pair(s).
the first four consensus repeats that was amplified from the GMP-140 cDNA using the polymerase chain reaction. Positive clones were rescreened and plaque purified. The EMBL3 genomic DNA inserts were excised from the phage DNA by digestion first with Sall, then with either EcoRI, HindIII, or XbaI. The Charon 4A inserts were removed with EcoRI and, in some cases, PstI. Overlapping inserts spanning the entire gene were subcloned into the plasmids pBluescript II KS+ (IBI Biotechnologies, Inc., New Haven, CT), BlueScript (Stratagene, San Diego, CA), or PUC18 (BRL, Gaithersburg, MD). Overlaps between inserts were confirmed by sequencing.

Intron-exon boundaries were sequenced (1) using oligonucleotide primers located at predicted borders of exons, for example, at each end of the lectin-like domain, the EGF-like domain, and each of the nine consensus repeats. Additional primers were synthesized as required. All coding regions were sequenced on both strands. The locations and sizes of introns were determined by a combination of sequencing, partial restriction mapping, agarose gel electrophoresis of genomic fragments amplified by the polymerase chain reaction, and Southern blots probed with specific oligonucleotides.

Amplification of GMP-140 mRNA by the Polymerase Chain Reaction—Previous analysis of several endothelial cell cDNA clones predicted variant forms of GMP-140 with deletions of either the seventh consensus repeat or the transmembrane domain (1). To determine whether such variations were present in RNA transcripts, the relevant regions of platelet or endothelial cell RNA were amplified using the polymerase chain reaction as described previously (20).

Briefly, the first strand of cDNA was synthesized from cellular RNA with reverse transcriptase and then amplified with Taq1 polymerase. For the region encompassing the seventh consensus repeat, primers in the coding strand were synthesized from sequence in the sixth repeat and primers in the noncoding strand were synthesized from sequence in the eighth repeat. For the region surrounding the transmembrane domain, the coding primers contained a sequence from the ninth repeat, and the noncoding primers were from the second exon encoding the cytoplasmic domain.

To increase the ratio of specifically to nonspecifically amplified DNA fragments, the products were subjected to repeat amplification with one or two internally nested primers. Prior to each additional amplification, the products of the previous amplification were extracted with phenol/chloroform, ethanol-p precipitated, and resuspended in water. The sequences of the primers were designated in a 5' to 3' direction according to the nucleotide sequence of the GMP-140 cDNA (1). For the region surrounding the transmembrane domain, the nucleotide sequences of the primers were as follows: first amplification, coding 2148–2164, noncoding 2598–2519; second amplification, coding 2171–2189, noncoding 2528–2512. For the area encompassing the seventh repeat, the primers as follows: first amplification, coding 1538–1557, noncoding 1996–2016; second amplification, coding 1568–1588, noncoding 2022–2116; third amplification, coding 2171–2189, noncoding 2528–2512. Portions of the final amplification products were excised by agarose gel electrophoresis and restriction digests. The smaller cDNA product with the putative deleted transmembrane domain was excised from the gel and directly sequenced (21). The products of amplification between the sixth and eighth repeats were electrophoresed, transferred to a nylon membrane, and probed consecutively with 32P-labeled oligonucleotides specific for repeats 6 and 7. The labeled oligonucleotide from repeat 6 was designed from nonoverlapping sequence internal to that of the repeat 6 primer used in the polymerase chain reaction.

RESULTS

Isolation and Characterization of the Gene for GMP-140—Three human genomic DNA libraries in bacteriophage vectors were screened with cDNAs encoding GMP-140. Seven clones spanning the GMP-140 gene were isolated and overlapping genomic DNA inserts were subcloned into plasmids for analysis as described under "Experimental Procedures." Fig. 1 depicts the organization of the GMP-140 gene and the location of genomic inserts from each of the seven phage clones. Table I indicates the size and location of the exons and introns as well as the nucleotide sequences surrounding the splice donor and acceptor sites. The gene for GMP-140 spans at least 50 kilobase pairs and is composed of 17 exons and 16 introns. The nucleotide sequences at the 5' donor and 3' acceptor sites of all introns conform to the GT...AG rule (22).

The GMP-140 cDNA sequence predicts a protein composed of a number of discrete modular elements (1). Analysis of the genomic organization indicates that individual exons tend to encode each of these domains, as illustrated schematically in Fig. 2. Exon 1 ends with the published 38 bp of 5'-untranslated cDNA sequence followed by the ATG codon for the methionine residue that initiates translation (1). A large intron of 11.3 kilobase pairs separates this exon from exon 2, which encodes the first 30 amino acids of the signal peptide. Exon 3 encodes the remaining 10 amino acids of the signal peptide joined to the sequence encoding the lectin domain. Following exon 3 is another large intron spanning at least 13.1 kilobase pairs, whose entire length has not been determined. Next are a series of exons that encode the EGF domain and each of the nine consensus repeats. Exon 14 encodes 40 amino acids surrounding the 24-residue hydrophobic transmembrane domain, including the first seven charged amino acids of the cytoplasmic segment. Two additional exons encode the remaining 28 amino acids of the cytoplasmic domain. Finally, exon 17 begins after the stop codon and contains all of the 3'-untranslated sequence. All but two of the intron-exon splice junctions are split between the first and second nucleotides of codons (phase 1). The exceptions are exon 1, which ends after the three bases of the ATG codon for the initiating methionine (phase 0), and exon 15, which encodes the first part of the cytoplasmic tail and ends after the first two bases of a codon (phase 2).

The sequence of the coding regions of the GMP-140 gene matched that determined previously from endothelial cell cDNAs (1). Five of eight polymorphisms previously noted in the cDNAs were also observed in the genomic sequence, as well as a new polymorphism in which C replaces A at base 2304. This nucleotide change results in a codon encoding a proline instead of a threonine in the ninth repeat of the molecule. All nine polymorphisms thus far identified are in sequence encoding the consensus repeats.

Variant Forms of GMP-140 Generated by Alternative Splicing—We previously identified two in-frame deletions in certain independent GMP-140 cDNA clones that predicted alternative forms of the molecule (1). One was a 186-bp deletion encoding 82 amino acids in the seventh consensus repeat, predicting a soluble form of GMP-140. Genomic analysis indicates that exon 11 encodes the second portion of the cytoplasmic tail. Products of the 357 and 237 bp were excised from the phage DNA by digestion first with SalI, then with XbaI (not shown). The size difference of 120 bp between the two amplified fragments was consistent with the predicted deletion of the 120-bp transmembrane exon in the smaller product. To confirm this, the fragment was gel purified and sequenced. As shown in Fig. 3B, the sequence of this fragment matched the predicted splice site created by joining exon 13 encoding the...
FIG. 1. Organization of the human gene encoding GMP-140. Exons are indicated by vertical bars, with filled bars representing coding regions and open bars indicating 5'- and 3'-untranslated sequence. The domains encoded by each exon are identified as follows: 5'UN, 5'-untranslated region; SI, first portion of signal peptide; SZ/LEC, second portion of signal peptide fused to lectin-like domain; EGF, epidermal growth factor-like domain; CR1...CR9, consensus repeats 1 through 9; TM, transmembrane domain; Cl, first portion of cytoplasmic domain; C2, second portion of cytoplasmic domain; 3'UN, 3'-untranslated region. Introns are represented by the thin line and are drawn to scale; the diagonal lines in the intron following exon 3 represent a region of indeterminate length that has not been characterized. A partial restriction map with four different endonucleases is depicted; cleavage sites for these enzymes were not examined in all regions of the gene. At the bottom are depicted the positions of seven overlapping genomic clones from which the genomic structure was determined. EMBL clones 1, 4, and 5 were from a lymphocyte library, EMBL3 clones 2 and 4 were from a colonic mucosa library, and Charon 4A clones 1 and 2 were from a lymphocyte library. The diagonal lines at the ends of genomic clones indicate regions that were not studied.

TABLE I

Intron-exon organization of the gene for human GMP-140

| Number and domain | Location in cDNA | Length | 3' splice acceptor | 5' splice donor | Phase | Length |
|-------------------|------------------|--------|-------------------|----------------|-------|--------|
| 1 5'UN            | 1-41             | 91     | AGAGGA GAGT      | Met            | 0     | 11,300 |
| 2 S1              | 42-132           | 387    | GAGT G TCT G     | gtaacttta      | 1     | 1,600  |
| 3 S2              | 153-319          | 108    | GCC TGC TCT G... | GAGT G TCT G... | 1     | >13,150 |
| 4 EGF /LEC        | 520-627          | 186    | GCC TGC TCT G... | GAGT G TCT G... | 1     | >13,150 |
| 5 CR1             | 628-813          | 133-150| 1 | 1     | 472'  |
| 6 CR2             | 814-989          | 133-150| 1 | 1     | 526'  |
| 7 CR3             | 1000-1185        | 186    | GCC TGC TCT G... | GAGT G TCT G... | 1     | >13,150 |
| 8 CR4             | 1186-1371        | 133-150| 1 | 1     | 1,810  |
| 9 CR5             | 1372-1557        | 133-150| 1 | 1     | 1,900  |
| 10 CR6            | 1558-1743        | 133-150| 1 | 1     | 4,000  |
| 11 CR7            | 1744-1929        | 133-150| 1 | 1     | 6,030  |
| 12 CR8            | 1930-2139        | 133-150| 1 | 1     | 900    |
| 13 CR9            | 2149-2325        | 133-150| 1 | 1     | 1,230  |
| 14 TM             | 2326-2445        | 133-150| 1 | 1     | 956'  |
| 15 C1             | 2446-2476        | 133-150| 1 | 1     | 2,310  |
| 16 C2             | 2477-2532        | 133-150| 1 | 1     | 1,390  |
| 17 3'UN           | 2533-3124        | 133-150| 1 | 1     | 690'  |

*The domain abbreviations are described in the legend to Fig. 1.

a Nucleotide sequence of cDNA numbered according to Ref. 1.

b Length determined by sequencing of entire intron.

d Ninth repeat to exon 15 encoding the first portion of the cytoplasmic tail. This indicates that transcripts encoding a putative soluble form of GMP-140 are produced by megakaryocytes as well as endothelial cells.

Primers from exons encoding the sixth and eighth repeats were used to amplify the region surrounding the seventh repeat (Fig. 3C). A prominent 543-bp fragment was produced from both platelet and endothelial RNA. A minor fragment of 357 bp was also amplified from endothelial RNA. The size difference of 186 bp between the minor and major fragments was consistent with the predicted deletion of the 186-bp exon encoding the seventh repeat in the smaller product. This is supported by the observation that both the larger and smaller fragments hybridized with an oligonucleotide containing a sequence from the sixth repeat (Fig. 3C) but only the larger fragment hybridized with an oligonucleotide from the seventh
alternative splicing generates transcripts that encode soluble forms of two other cell surface receptors, HLA-A2 (30) and the FCγRII receptor (31).

The functional significance, if any, of the removal of the seventh repeat by alternative splicing is unknown. With one exception, each of the nine exons encoding the repeats in GMP-140 contains 186 bp encoding 62 amino acids. The exception is exon 12 which contains sequence for an additional eight amino acids fused to the rest of the eighth repeat. Perhaps the secondary structure of the precursor mRNA in the region between exons 11 and 12 allows for occasional removal of exon 11. The eight-repeat form of GMP-140 appears to be rare. Only one such cDNA was noted out of six endothelial cDNAs initially examined (1). Furthermore, even with repeated amplifications during the polymerase chain reaction, only a minor band corresponding to this variant was noted in endothelial RNA. No corresponding structure was noted in platelet RNA. However, the amplification was performed with platelet RNA from a single individual, whereas the endothelial RNA was pooled from many donors.

Although precise quantitation has not been performed, transcripts encoding the putative soluble form of GMP-140 may be as common as those encoding the membrane form. The deletion of the exon encoding the transmembrane domain was noted in two out of four independent endothelial cDNAs (1). Furthermore, this transcript was easily amplified from both platelet (Fig. 3B) and endothelial cell RNA. This suggests a potentially important role for a secretable form of GMP-140 that is synthesized by both megakaryocytes and endothelium. A similarly synthesized soluble variant has not been described for the other selectins. Although a soluble form of the homing receptor (Mel 14 antigen, LAM-1) is released from activated leukocytes, it is probably generated by proteolytic cleavage of the membrane-bound molecule (9, 10, 32, 33).

The structure of the human gene encoding LAM-1, another selectin, has been recently reported (34). The intron-exon organization of this gene demonstrates a striking similarity to that found in GMP-140, which suggests that a similar pattern will be found in the gene encoding ELAM-1 and perhaps in genes of other selectins yet to be identified. The genomic structures of LAM-1 and GMP-140 support the role of exon duplication and rearrangement in generating this newly described family of proteins that facilitate cellular interactions during inflammation.

Acknowledgments—We thank Sheryl Christoffersen, Ginger Hampton, Chris Titsworth, and Suzanne Lyman for excellent tech-
Fig. 3. Alternative splicing of GMP-140 transcripts. A, the normal splicing pattern producing full-length transcripts is shown above the gene map. Below the gene map are shown alternative splicing events resulting in removal of either exon 11 encoding the seventh repeat or exon 14 encoding the transmembrane domain of GMP-140 in platelet RNA. Primers from exons 13 and 16 were used to amplify the region surrounding the seventh repeat or exon 14 encoding the transmembrane domain. The amplified products were electrophoresed on an agarose gel and stained with ethidium bromide (left panel) and then transferred to a nylon membrane and probed with a 35P-labeled oligonucleotide (right panel). This probe hybridized only with the larger fragment (not shown), consistent with deletion of exon 11 encoding the seventh repeat from the smaller fragment.

REFERENCES
1. Johnston, G. I., Cook, R. G., and McEver, R. P. (1989) Cell 56, 1033-1044
2. Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A., Jr., and Seed, B. (1989) Science 243, 1160-1165
3. Lasky, L. A., Singer, M. S., Vednock, T. A., Dowhenko, D., Fennie, C., Rodriguez, H., Nguyen, T., Stachel, S., and Rosen, S. D. (1989) Cell 56, 1045-1055
4. Siegelman, M. H., van de Rijn, M., and Weissman, I. L. (1989) Science 243, 1165-1172
5. Tedder, T. F., Isaac, C. M., Ernst, T. J., Demetri, G. D., Adler, D. A., and Diatche, C. M. (1989) J. Exp. Med. 170, 123-133
6. Camerini, D., James, S. P., Stamenkovic, I., and Seed, B. (1989) Nature 342, 78-82
7. Bowen, B. R., Nguyen, T., and Lasky, L. A. (1989) J. Cell Biol. 109, 1-24
8. Siegelman, M. H., and Weissman, I. L. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 5562-5566
9. Kishimoto, T. K., Jutila, M. A., Berg, E. L., and Butcher, E. C. (1989) Science 245, 238-241
10. Jutila, M. A., Rott, L., Berg, E. L., and Butcher, E. C. (1989) J. Immunol. 143, 3318-3324
11. Stenberg, P. E., McEver, R. P., Shuman, M. A., Jacques, Y. V., and Bainton, D. P. (1985) J. Biol. Chem. 101, 880-886
12. Berman, C. L., Yeo, E. E., Wencel-Drake, J. D., Furie, B. C., Ginsberg, M. H., and Furie, B. (1986) J. Clin. Invest. 78, 130-137
13. Hattori, R., Hamilton, K. K., Fugate, R. D., McEver, R. P., and Sims, P. J. (1989) J. Biol. Chem. 264, 7768-7777
14. McEver, R. P., Beckstead, J. H., Moore, K. L., Marshall-Carlson, L., and Bainton, D. F. (1989) J. Clin. Invest. 84, 92-99
15. Larsen, E., Celi, A., Gilbert, G. E., Furie, B. C., Erban, J. K., Boffianti, R., Wagner, D. D., and Furie, B. (1989) Cell 00, 306-312
16. Hamburger, S. A., and McEver, R. P. (1990) Blood 75, 550-554
17. Geng, J.-G., Bevilacqua, M. P., Moore, K. L., McIntyre, T. M., Prescott, S. M., Kim, J. M., Blas, G. A., Zimmerman, G. A., and McEver, R. P. (1990) Nature 343, 78-76
18. Watson, M. L., Kingmore, S. F., Johnston, G. I., Siegelman, M. H., Le Beau, M. M., Lema, R. S., Borrow, N. S., Howard, T. A., Weissman, I. L., McEver, R. P., and Seldin, M. F. (1990) J. Exp. Med. 172, 263-272
19. Johnston, G. I., Kurosky, A., and McEver, R. P. (1989) J. Biol. Chem. 264, 1816-1823
20. Newman, P. J., Gorski, J., White, G. C., II, Gidwitz, S., Cretney, C. J., and Aster, R. H. (1988) J. Clin. Invest. 82, 739-743
21. Newman, P. J., Dubes, R. S., and Aster, R. H. (1989) J. Clin. Invest. 83, 1778-1781
22. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383
23. Patthy, L. (1987) FEBS Lett. 214, 1-7
24. Drickamer, K. (1988) J. Biol. Chem. 263, 9557-9560
25. Leung, J. O., Holland, E. C., and Drickamer, K. (1985) J. Biol. Chem. 260, 12523-12527
26. Suter, U., Bastos, R., and Hofstetter, H. (1987) Nucleic Acids Res. 15, 7295-7308
27. Drickamer, K., and McEver, R. P. (1987) J. Biol. Chem. 262, 2582-2589
28. White, R. T., Damm, D., Miller, J., Spratt, K., Schilling, J., Hawgood, S., Benson, B., and Cordell, B. (1985) Nature 317, 301-303
29. Fujisaku, A., Harley, B. J., Frank, M. C., Gruner, B. A., Frazier, B., and Holers, V. M. (1989) J. Biol. Chem. 264, 2118-2125
30. Krangel, M. S. (1986) EMBO J. 4, 1205-1210
31. Walterhouse, D. O., Cassel, D. L., Schreiber, A. D., Meister, R. P., Schwartz, E., and Rappaport, E. F. (1988) Blood 72, 344a (abstr.)
32. Tedder, T. F., Penta, A. C., Levine, H. B., and Freedman, A. (1990) J. Immunol. 144, 532-540
33. Kishimoto, T. K., Jutila, M. A., and Butcher, E. T. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 2244-2248
34. Ord, D. C., Ernst, T. J., Zhou, L.-J., Ramhaidi, A., Sipetini, O., Griffin, J., and Tedder, T. F. (1990) J. Biol. Chem. 265, 7760-7767
Structure of the human gene encoding granule membrane protein-140, a member of the selectin family of adhesion receptors for leukocytes.

G I Johnston, G A Bliss, P J Newman and R P McEver

J. Biol. Chem. 1990, 265:21381-21385.

Access the most updated version of this article at http://www.jbc.org/content/265/34/21381

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/34/21381.full.html#ref-list-1