Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6394–6398, October 1985
Biochemistry

Cloning and characterization of two cDNAs coding for human von Willebrand factor

(cDNA cloning/DNA sequence analysis)

J. EVAN SADLER*†‡, BEVERLY B. SHELTON-INLOES*†‡, JAMES M. SORACE*§, JOHN M. HARLAN¶, KOITI TITANI||, AND EARL W. DAVIE||

*Howard Hughes Medical Institute Laboratory, Departments of †Medicine, ‡Biochemistry, and §Laboratory Medicine, Washington University School of Medicine, St. Louis, MO 63110; and Departments of ||Biochemistry and ¶Medicine, University of Washington, Seattle, WA 98195

Contributed by Earl W. Davie, May 29, 1985

ABSTRACT **A cDNA library was prepared in λgt11 bacteriophage from poly(A)$^+$ RNA isolated from primary cultures of endothelial cells from human umbilical vein. Approximately 2.5 million independent recombinants were screened and 2 of those were found to synthesize a fusion protein with β-galactosidase that reacted with rabbit antibody against human von Willebrand factor. Comparison of the amino acid sequence translated from the cDNA insert of the two clones with the amino acid sequence determined by Edman degradation of the protein established that both phage isolates code for von Willebrand factor. The first clone (λHvWF1) contained an insert of 404 nucleotides that corresponded to amino acid residues 1–110 in the mature protein circulating in blood, in addition to a portion (24 amino acids) of a prepro leader sequence. The second cDNA clone (λHvWF3) contained an insert of 4.9 kilobases that coded for the carboxyl-terminal 1525 amino acids of von Willibrand factor, a stop codon of TGA, 134 nucleotides of 3′ noncoding sequence, and a poly(A) tail of 150 nucleotides. The two clones together code for >80% of the molecule circulating in blood. The same carboxyl-terminal lysine residue was identified in the mature protein as well as in the cDNA, indicating that all of the proteolytic processing that occurs during the biosynthesis and assembly of von Willebrand factor is associated with the amino-terminal portion of the precursor protein. The amino acid sequence of von Willebrand factor indicates the presence of two different internal gene duplications and one triplication. These repetitive amino acid sequences account for about one-half of the amino acids present in the mature protein. The tetrapeptide sequence of Arg-Gly-Asp-Ser, which mediates the cell attachment and platelet binding activity of fibronectin, was also identified in the carboxyl-terminal portion of von Willebrand factor.**

von Willebrand factor is a multimeric plasma glycoprotein that consists of subunits (M_r, 260,000) that are held together by disulfide bonds. It circulates in blood as multimers that range in size from dimers of ≈500,000 to multimers of >10,000,000. von Willebrand factor is synthesized by endothelial cells throughout the body (1, 2) and also by megakaryocytes (3). It participates in the initial reactions of hemostasis by forming a bridge between platelets and the damaged vascular subendothelium, and this leads to platelet plug formation. Specific receptors for von Willebrand factor have been identified on the platelet membrane as well as the subendothelium. With platelets, its binding is associated with platelet glycoprotein IB (4–6) and, under some circumstances *in vitro*, to platelet glycoprotein IIB/IIIA (7, 8). The principal receptor for von Willebrand factor is probably collagen in the subendothelial connective tissue (9–11). von Willebrand factor also forms a complex with factor VIII (antihemophilic factor), and this interaction is necessary for the survival of the coagulant protein *in vivo* (12). The function of von Willebrand factor in platelet plug formation appears to be dependent on the assembly of the protein subunits into large multimers. Accordingly, the decreased biological activity of the plasma protein may be due to abnormalities of polymerization or decreased levels of the protein in blood. Both types of defect have been described in von Willebrand's disease.

Small amounts of highly purified von Willebrand factor have been available for nearly 15 years, but the biochemical characterization of the protein has been hampered by its large size and complexity. Recently, several posttranslational modifications have been described during the biosynthesis of von Willebrand factor, including proteolytic processing (13, 14), glycosylation (13, 15, 16), and sulfation (17). Formation of disulfide bonds leads to the generation of the small and large multimers. Some relationships between protein domains, oligomeric structure, and platelet-related functions have been described (18–20). Many variants of von Willebrand's disease have been identified, and these have been classified on the basis of the structural properties of the mutant proteins as well as *in vitro* tests (8, 21, 22).

As a step toward understanding the structure–function relationships, gene organization, biosynthetic regulatory mechanisms, and evolution of von Willebrand factor, cDNA clones coding for von Willebrand factor have been isolated from a human endothelial cell cDNA library in our laboratory (23) as well as by others (24, 25). In this report, we describe the DNA sequence of two cDNA inserts that together code for >80% of the mature protein present in plasma, in addition to 24 amino acids from an amino-terminal leader peptide.

MATERIALS AND METHODS

Restriction endonucleases, nuclease BAL-31, T4 DNA ligase, T4 polynucleotide kinase, *Eco*RI methylase, and T4 DNA polymerase were purchased from Bethesda Research Laboratories or New England Biolabs. Human placental ribonuclease inhibitor was supplied by Bolton Biologicals (Richmond Heights, MO), and reverse transcriptase was purchased from Seikagaku America (St. Petersburg, FL). Oligo(dT)- cellulose, oligo(dT)$_{12-18}$, *Eco*RI linkers, ATP, dideoxynucleotide, and deoxynucleotide triphosphates were supplied by Pharmacia P-L Biochemicals. The Klenow fragment of *Escherichia coli* DNA polymerase was purchased from Bethesda Research Laboratories and Boehringer Mannheim. Calf intestine alkaline phosphatase and nuclease S1 were products of Boehringer Mannheim. Deoxyadenosine 5′-[[α-^{35}S]thio]triphosphate (dATP[α-^{35}S]) was purchased from Amersham. Na^{125}I and [γ-^{32}P]ATP were purchased from New England Nuclear.

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. §1734 solely to indicate this fact.

For cDNA cloning, umbilical vein endothelial cells were cultured by standard methods (26, 27), and RNA was isolated by centrifugation through 5.7 M CsCl after lysis of the cells in 5 M guanidine thiocyanate (28). Poly(A)$^+$ RNA was selected by chromatography on oligo(dT) cellulose. Double-stranded cDNA was synthesized by using reverse transcriptase for first-strand synthesis, and with both reverse transcriptase and Klenow fragment for second-strand synthesis. These reactions as well as methylation, addition of *Eco*RI linkers, ligation into the λgt11 arms, and packaging *in vitro* were performed as described by Schwarzbauer *et al.* (29). From 10 μg of poly(A)$^+$ RNA, 0.42 μg of size-selected double-stranded cDNA was prepared, yielding ≈20 million independent recombinant phage.

Rabbit antiserum and affinity-purified antibodies to human von Willebrand factor were prepared as described by Canfield and Kisiel (30). The purified antibody was labeled with ^{125}I to a specific activity of 4 million cpm/μg (31) and was used to screen the library at a plating density of 50,000 recombinants per 150-mm plate (32). Positive clones were plaque-purified and DNA was prepared by a plate-lysis method (33, 34).

For DNA sequencing, the cDNA inserts from positive isolates were subcloned into plasmid pUC18 (35) and suitable fragments were further subcloned into M13 mp18 or M13 mp19 (35). Sequencing was performed by the dideoxy method (36) using dATP[α-^{35}S] and buffer-gradient gels (37). Controlled digestions with nuclease BAL-31 were used to generate templates providing overlapping sequences (38). When necessary, oligonucleotides (17–20 bases) were synthesized (Applied Biosystems, Foster City, CA; DNA synthesizer) and used as sequencing primers with appropriate templates. All sequences were determined at least once on each strand by the dideoxy method. The first 220 nucleotides in λHvWF1 were also sequenced by the method of Maxam and Gilbert (39). DNA sequences were analyzed with the computer programs of CompuGene (St. Louis, MO) and the Protein Identification Resources (40).

RESULTS AND DISCUSSION

A human endothelial cell cDNA library prepared in λgt11 was screened with affinity-purified, ^{125}I-labeled rabbit antibody to human von Willebrand factor. Among 2.5 million recombinants screened, 2 positive λ phage were identified and plaque purified. The DNA inserts in the two isolates were then subcloned into pUC18 for further characterization. The first isolate (λHvWF1) contained a cDNA insert of 404 nucleotides flanked by *Eco*RI linkers (Fig. 1). The second isolate (λHvWF3) contained an insert of 4.9 kilobases of DNA, but only one linker *Eco*RI site was reconstituted. The DNA in this phage was digested with *Sac* I, which led to the formation of two fragments that together spanned all but 280 nucleotides of the cDNA insert (Fig. 1). The remaining DNA fragment was sequenced from an *Acc* I/*Eco*RI subclone of λHvWF3. The DNA and corresponding amino acid sequences of the DNA inserts in λHvWF1 and λHvWF3 are shown in Fig. 2.

λHvWF1 codes for the amino-terminal end of von Willebrand factor in addition to 24 amino acid residues that constitute a portion of a leader sequence. This conclusion was made possible by the fact that the amino-terminal sequence of the von Willebrand factor subunit has been established as Ser-Leu-Ser-Cys-Arg-Pro-Pro by automated Edman degradation of the intact protein and a cyanogen bromide fragment originating from the amino-terminal end of the protein (ref. 50; M. Chopek, personal communication). The amino acid residue occurring immediately before the amino-terminal serine was arginine. Since signal peptidase does not cleave Arg-Ser bonds, the partial leader sequence of 24 amino acids represents a portion of a prepro peptide (41).

The cDNA sequence in λHvWF3 contained a long open reading frame extending from nucleotide 2 through nucleotide 4576. The last 11 amino acid residues in this reading frame corresponded exactly to the carboxyl-terminal amino acid sequence of von Willebrand factor as determined by automated Edman degradation of the CNBr peptide originating from the carboxyl-terminal end of the protein (M. Chopek, personal communication). Therefore, the DNA insert in λHvWF3 corresponds to the last 1525 amino acids of the protein. The carboxyl-terminal lysine (AAG) was followed by a stop codon (TGA) and 134 nucleotides of 3′ noncoding sequence. These data indicate that all of the proteolytic processing that occurred during the assembly of von Willebrand factor into multimers took place in the amino-terminal region of the precursor polypeptide. A polyadenylylation or processing signal of AATAAA was identified 25 nucleotides prior to the poly(A) tail of 150 nucleotides.

The cDNA inserts in λHvWF1 and λHvWF3 do not overlap. Between them, they code for 1635 amino acids present in the mature von Willebrand factor subunit present in plasma and account for all but three of the known CNBr fragments of the protein (unpublished results). The mature protein present in plasma consists of ≈2000 amino acids (19). Thus, there is a gap of ≈1 kilobase of DNA coding for ≈350 amino acids that is not represented in either λHvWF1 or λHvWF3. Thus far, ≈80% of the sequence of von Willebrand factor has been established by amino acid sequence analysis (unpublished results). The amino acid sequence predicted from the two cDNAs is in excellent agreement with these data except for histidine-7 in λHvWF1, which was clearly a proline by amino acid sequence analysis. This difference may

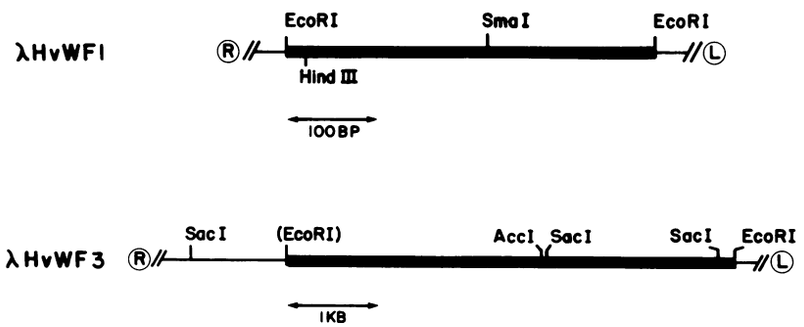

FIG. 1. Partial restriction maps of the cDNA inserts in λHvWF1 and λHvWF3 that code for von Willebrand factor. The cDNA inserts are shown with the 5′ end of the coding strand at the left; thus, the conventional orientation of the left and right arms of the λ phage are reversed. Only those restriction sites used in subcloning and DNA sequencing are shown. BP, base pairs; KB, kilobase.

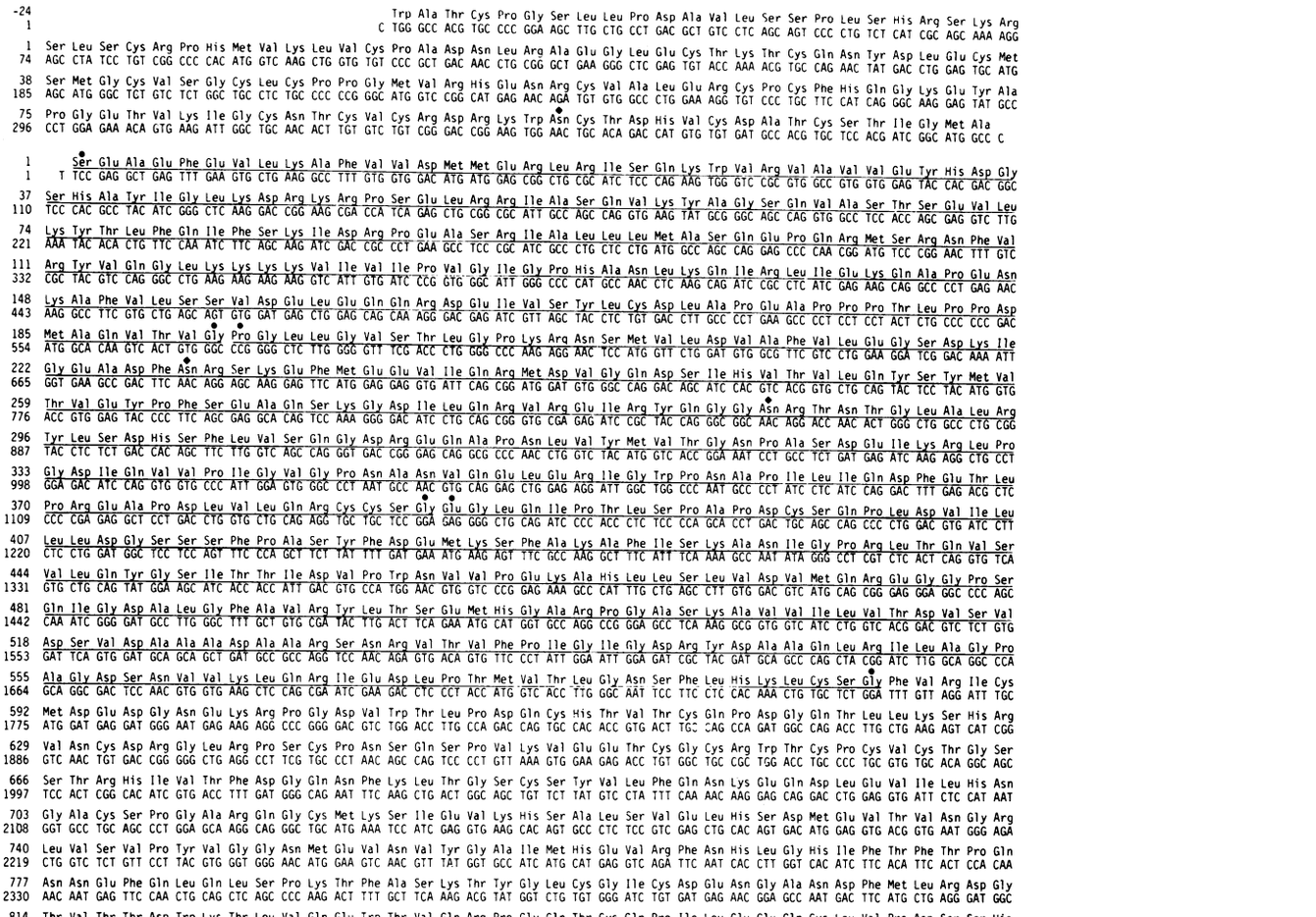

FIG. 2. Nucleotide sequences for the cDNA inserts in λHvWF1 and λHvWF3 and predicted amino acids. Nucleotides are numbered 1–404 in λHvWF1 and 1–4713 in λHvWF3, excluding the poly(A) tail. The serine, which is the amino-terminal residue of von Willebrand factor present in plasma, has been numbered as residue 1 in λHvWF1. The numbering of the amino acids in λHvWF3 (1–1525) is arbitrary and will require revision when the DNA sequence separating these two isolates is established. Asparagine residues that are potentially glycosylated are shown by solid diamonds. Amino acids that are present in repeated segments are underlined. The beginning and ending of the A domains are identified by solid circles, the B domains are shown by solid triangles, and the C domains are shown by solid squares. The tetrapeptide, Arg-Gly-Asp-Ser, is identified by arrowheads; the RNA processing or polyadenylylation signal of AATAAA is enclosed in a box.

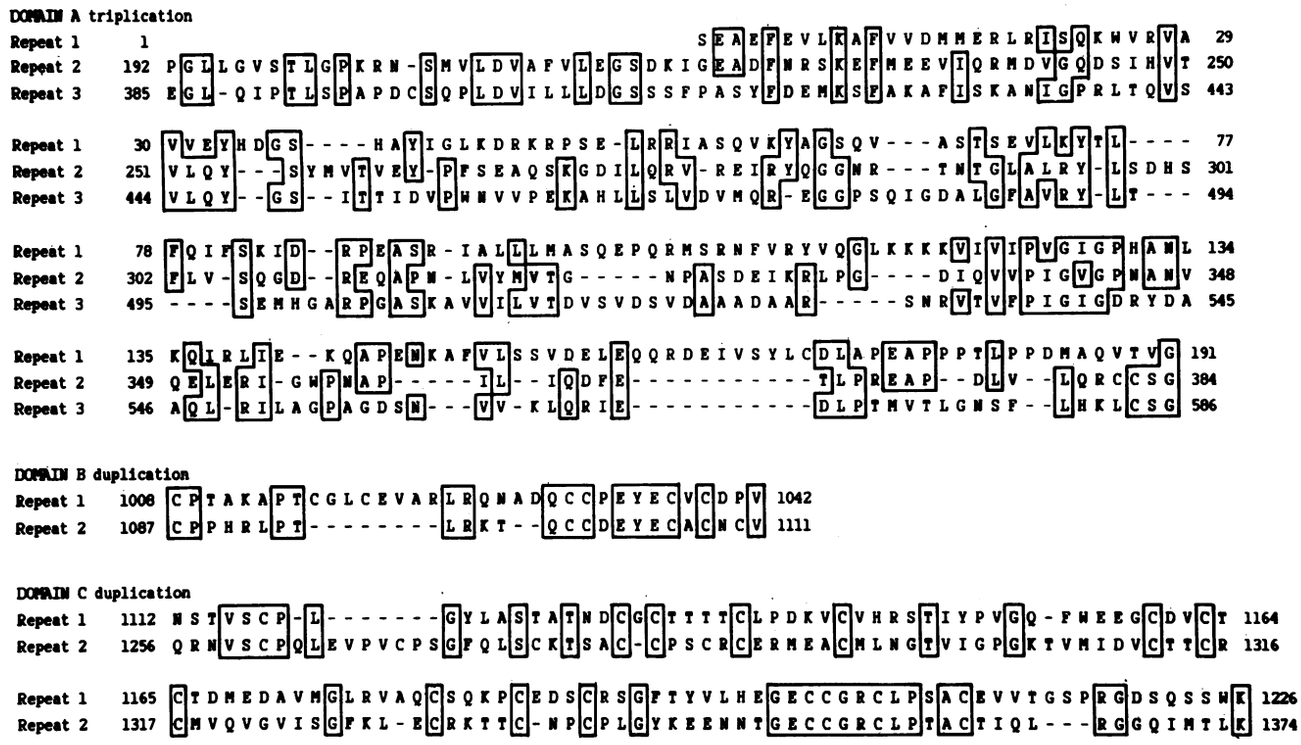

FIG. 3. Internal homologous domains in von Willebrand factor. The amino acids are numbered as shown in Fig. 2 for λHvWF3. The alignments shown were achieved with the ALIGN computer program (40). For each pairwise comparison, the alignment scores in standard deviation units (SD) are followed by the probability (*P*) that this score could occur by chance: Domain A repeat 1 versus 2, 5.96 SD, $P < 10^{-8}$; repeat 1 versus 3, 3.29 SD, $P < 10^{-3}$; repeat 2 versus 3, 10.62 SD, $P < 10^{-23}$. Domain B repeat 1 versus 2, 5.98 SD, $P = 10^{-9}$. Domain C repeat 1 versus 2, 7.76 SD, $P < 10^{-13}$. The scoring matrix used was the mutation data matrix (MD + 2), with a gap penalty of 6 (40). Amino acids are shown by the single-letter code.

be due to polymorphism in the protein or a cloning artifact generated during the preparation of the cDNA library.

von Willebrand factor contains ≈15% carbohydrate distributed among both asparagine-linked and threonine/serine-linked oligosaccharides. Thus far, at least five asparagine-linked structures have been identified (42, 43). The protein sequence predicted from λHvWF1 and λHvWF3 includes 11 potential *N*-glycosylation sites with the sequence of Asn-X-Thr(or Ser). In addition, there are seven potential carbohydrate binding sites with the sequence of Asn-X-Cys. This sequence is glycosylated in protein C (44, 45). The carboxyl-terminal one-third of von Willebrand factor is very rich in cysteine residues, while the methionine residues are distributed throughout the molecule.

Three separate polypeptide segments of von Willebrand factor show evidence of internal duplication (Fig. 3). Amino acid residues 1–586 in λHvWF3 contain a head-to-tail triplication of ≈200 amino acids with an identity ranging from 29% to 43%. The amino acid sequence between residues 1008 and 1042 is duplicated in residues 1087–1111. Also, the amino acid sequence between residues 1112 and 1226 is duplicated in residues 1256–1374. For each comparison, the alignment score in standard deviation units is listed in the figure legend. Among the three groups of duplicated amino acid sequences, there were small regions that show remarkable sequence identity and other regions that have diverged considerably. Altogether, ≈50% of the amino acid sequence shown in Fig. 2 is part of repeated segments. The function of these regions is unknown, but they do indicate that von Willebrand factor has a complex evolutionary history involving gene duplication of the three different sequences or domains.

Nucleotides 3656–3667 in λHvWF3 code for Arg-Gly-Asp-Ser. This sequence has been identified within a region of fibronectin that mediates the cell attachment activity of the protein (46, 47). In addition, both von Willebrand factor and fibronectin bind competitively to thrombin-activated platelets, and these interactions are inhibited by small peptides containing the sequence Arg-Gly-Asp-Ser. This suggests that this region of von Willebrand factor may participate in platelet binding (48, 49). von Willebrand factor, however, is not homologous to fibronectin. In fact, a comparison of the von Willebrand factor sequence with those in the National Biomedical Research Foundation Protein Sequence Data Base (Georgetown University, Washington, DC; ref. 40) did not reveal any major similarity to fibronectin or any other protein. If this region of von Willebrand factor is required for a physiologically significant interaction with platelets, the similarity to the corresponding functional region of fibronectin may represent convergent evolution.

We thank Drs. Kenneth Walsh, Kazuo Fujikawa, Santosh Kumar, and Lowell Ericsson for helpful discussions dealing with the protein sequence of human von Willebrand factor; Mark Behlke for performing Maxam–Gilbert sequencing reactions on selected fragments; John Bell for synthesis of oligonucleotides; Penny Thompson for culturing the endothelial cells; and Dr. Michael Chopek for advice and for kindly providing purified von Willebrand factor. Also, the assistance of Drs. Mark Murray and Fred Hagen in constructing the cDNA library in λgt11 is gratefully acknowledged. This work was supported in part by National Institutes of Health Research Grants HL 16919, HL 18645, and HL 29595. J.E.S. was an Associate of the Howard Hughes Medical Institute, University of Washington, Seattle (1982–1984), and is currently an Associate Investigator of the Howard Hughes Medical Institute, Washington University, St. Louis.

1. Jaffe, E. A., Hoyer, L. W. & Nachman, R. L. (1973) *J. Clin. Invest.* **52**, 2757–2764.
2. Jaffe, E. A. & Hoyer, L. W. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1906–1909.
3. Nachman, R. L., Levine, R. & Jaffe, E. A. (1977) *J. Clin. Invest.* **60**, 914–921.

4. Okumura, T. & Jamieson, G. A. (1976) *Thromb. Res.* **8,** 701–706.
5. Nachman, R. L., Jaffe, E. A. & Weksler, B. B. (1977) *J. Clin. Invest.* **59,** 143–148.
6. Jenkins, C. S. P., Phillips, D. R., Clemetson, K. J., Meyer, D., Larrieu, M.-J. & Luscher, E. F. (1976) *J. Clin. Invest.* **57,** 112–124.
7. Fujimoto, T., Ohara, S. & Hawiger, J. (1982) *J. Clin. Invest.* **69,** 1212–1222.
8. Ruggeri, Z. M., Nilsson, I. M., Lombardi, R., Holmberg, L. & Zimmerman, T. S. (1982) *J. Clin. Invest.* **70,** 1124–1127.
9. Santoro, S. A. (1981) *Thromb. Res.* **21,** 689–693.
10. Santoro, S. A. & Cowan, J. F. (1982) *Collagen Relat. Res.* **2,** 31–34.
11. Morton, L. F., Griffin, B., Pepper, D. S. & Barnes, M. J. (1983) *Thromb. Res.* **32,** 545–556.
12. Tuddenham, E. G. D., Lane, R. S., Rotblat, F., Johnson, A. J., Snape, T. J., Middleton, S. & Kernoff, P. B. A. (1982) *Br. J. Haematol.* **52,** 259–267.
13. Wagner, D. D. & Marder, V. J. (1983) *J. Biol. Chem.* **258,** 2065–2067.
14. Lynch, D. C., Williams, R., Zimmerman, T. S., Kirby, E. P. & Livingston, D. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2738–2742.
15. Ling, E. H., Browning, P. J., Zimmerman, T. S. & Lynch, D. C. (1984) *Circulation* **70,** Suppl. 2, abstr. 833.
16. Wagner, D. D. & Marder, V. J. (1984) *J. Cell Biol.* **99,** 2123–2130.
17. Browning, P. J., Ling, E. H., Zimmerman, T. S. & Lynch, D. C. (1983) *Blood* **62,** 281 (abstr.).
18. Sixma, J. J., Sakariassen, K. S., Stel, H. V., Houdijk, W. P. M., In der Maur, D. W., Hamer, R. J., de Groot, P. G. & van Mourik, J. A. (1984) *J. Clin. Invest.* **74,** 736–744.
19. Titani, K., Ericsson, L. H., Kumar, S., Dörsam, H., Chopek, M. W. & Fujikawa, K. (1984) *Circulation* **70,** Suppl. 2, abstr. 837.
20. Girma, J. P., Pietu, G., Chopek, M. W., Edgington, T. S. & Meyer, D. (1984) *Circulation* **70,** Suppl. 2, abstr. 836.
21. Hoyer, L. W. (1982) in *The Hemophilias*, ed. Bloom, A. L. (Churchill-Livingstone, Edinburgh, Scotland), pp. 106–121.
22. Kinoshita, S., Harrison, J., Lazerson, J. & Abildgaard, C. F. (1984) *Blood* **63,** 1369–1371.
23. Sadler, J. E., Titani, K., Harlan, J. M. & Davie, E. W. (1985) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **44,** 1069 (abstr.).
24. Lynch, D. C., Zimmerman, T. S., Collins, C. J., Morin, M. J., Ling, E. H. & Livingston, D. M. (1985) *Clin. Res.* **33,** 548 (abstr.).
25. Ginsburg, D., Handin, R. I., Bonthron, D. T. & Orkin, S. H. (1985) *Clin. Res.* **33,** 546 (abstr.).
26. Striker, G. E., Harlan, J. M. & Schwartz, S. M. (1980) *Methods Cell Biol.* **21,** 135–151.
27. Thornton, S. C., Mueller, S. N. & Levine, E. M. (1983) *Science* **221,** 623–625.
28. Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
29. Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R. & Hynes, R. O. (1983) *Cell* **35,** 421–431.
30. Canfield, W. M. & Kisiel, W. (1982) *J. Clin. Invest.* **70,** 1260–1272.
31. Salacinski, P. R. P., McLean, C., Sykes, J. E. C., Clement-Jones, V. V. & Lowry, P. J. (1981) *Anal. Biochem.* **117,** 136–146.
32. Young, R. A. & Davis, R. W. (1983) *Science* **222,** 778–782.
33. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 75–85.
34. Degen, S. J., MacGillivray, R. T. A. & Davie, E. W. (1983) *Biochemistry* **22,** 2087–2097.
35. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26,** 101–106.
36. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
37. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3963–3965.
38. Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4298–4302.
39. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
40. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91,** 524–545.
41. Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H. & Lingappa, R. (1979) in *Secretory Mechanisms*, Society for Experimental Biology Symposium, eds. Hopkins, C. R. & Duncan, D. J. (Cambridge University Press, London), Vol. 33, pp. 9–36.
42. Samor, B., Mazurier, C., Goudemand, M., Debeire, P., Fournet, B. & Montreuil, J. (1982) *Thromb. Res.* **25,** 81–89.
43. Debeire, P., Montreuil, J., Samor, B., Mazurier, C., Goudemand, M., Van Halbeek, H. & Vliegenthart, J. F. G. (1983) *FEBS Lett.* **151,** 22–26.
44. Stenflo, J. & Fernlund, P. (1982) *J. Biol. Chem.* **257,** 12180–12190.
45. Foster, D. & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4766–4770.
46. Pierschbacher, M. D. & Ruoslahti, E. (1984) *Nature (London)* **309,** 30–33.
47. Pierschbacher, M. D. & Ruoslahti, E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5985–5988.
48. Ginsburg, M., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. & Plow, E. (1985) *J. Biol. Chem.* **260,** 3931–3936.
49. Haverstick, D. M., Cowan, J. F., Yamada, K. M. & Santoro, S. A. (1985) *Blood*, in press.
50. Hessel, B., Jörnvall, H., Thorell, L., Söderman, S., Larsson, U., Egberg, N., Blombäck, B. & Holmgren, A. (1984) *Thromb. Res.* **35,** 637–651.