cDNA Sequences for Human von Willebrand Factor Reveal Five Types of Repeated Domains and Five Possible Protein Sequence Polymorphisms†

Beverley B. Shelton-Inloes,‡ Koiti Titani,§ and J. Evan Sadler*,‡

Howard Hughes Medical Institute Laboratories, Departments of Medicine and Biochemistry, Washington University School of Medicine, St. Louis, Missouri 63110, and Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received January 14, 1986; Revised Manuscript Received April 11, 1986

ABSTRACT: A human umbilical vein endothelial cell cDNA library in λgt11 was screened with two previously described cDNA inserts for human von Willebrand factor. Among 16 positive isolates, two that hybridized with a probe corresponding to the amino terminus of von Willebrand factor were sequenced. Together, these four cDNA inserts span 6.5 kilobases of the von Willebrand factor mRNA sequence, completely specifying the 2050 amino acids of the subunit of mature, secreted von Willebrand factor and 24 residues of a precursor peptide. Approximately 77% of the sequence is contained in five types of repeated domains. Domain A consists of 193–220 amino acids and is present in three tandem copies between residues 497 and 1111. Domain B contains 25–35 amino acids and is present in three copies between residues 1533 and 1636. Domain C consists of 116–119 amino acids and is duplicated between residues 1637 and 1899. In contrast to the essentially contiguous repetition of domains A–C, the two copies of domains D and E are each separated by 804 and 1383 amino acids, respectively. Domain D1 contains 289 amino acids between residues 79 and 367, while domain D2 consists of 270 amino acids between residues 1171 and 1440. Domain E1 consists of 46 amino acids between residues 25 and 70, and domain E2 consists of 46 amino acids between residues 1453 and 1498. The triplicated A domains are notably poor in Cys content, while the remaining domains are Cys-rich. The A domains appear to be homologous to a 225-residue segment of complement factor B. Otherwise, von Willebrand factor is not closely related to any protein in the National Biomedical Research Foundation Protein Sequence Database, nor is the portion of the von Willebrand factor cDNA sequence that encodes the secreted protein homologous to any sequence in the Genbank Genomic Sequence Data Bank. Thus, four of the five types of repeated domains in von Willebrand factor have no homologues among other known proteins. The tetrapeptide Arg-Gly-Asp-Ser occurs at the carboxy-terminal end of domain C1 and may mediate the binding of von Willebrand factor to the GPIIb/IIIa complex of activated platelets. All of domain A1 lies within a 50-kilodalton tryptic fragment of von Willebrand factor that binds to GPIb of resting platelets [Fujimura, Y., Titani, K., Holland, L. Z., Russell, S. R., Roberts, J. R., Elder, J. H., Ruggeri, Z. M., & Zimmerman, T. S. (1986) *J. Biol. Chem. 261*, 381–385]. The remaining domains (B, D, and E) have not been correlated with specific functions. The sequence of the von Willebrand factor precursor before the amino-terminal *Ser* of plasma von Willebrand factor is His-Arg-Ser-Lys-Arg-*Ser*. The mature subunit is generated by proteolytic cleavage after the Lys-Arg dipeptide. This sequence resembles that of several mammalian, viral, fungal, and yeast protein precursors that are also proteolytically processed after paired basic residues during biosynthesis. Among the four cDNA isolates sequenced by this laboratory, there are eight single-nucleotide discrepancies that may reflect polymorphism in the von Willebrand factor gene sequence. Of these, seven are transitions, and one is a transversion. Five do not affect the translated protein sequence, but four result in single amino acid substitutions. Together with the previously reported disagreement at residue 7 between the translated sequence of λHvWF1 and the protein sequence, there are now five potential protein sequence polymorphisms for von Willebrand factor.

von Willebrand factor is a glycoprotein that is required for the adhesion of platelets to damaged endothelium. The concentration of von Willebrand factor in plasma is approximately 10 μg/mL. It is also found in subendothelial connective tissue and platelet α granules. The form of the protein found in plasma contains a single type of subunit, with M_r 225 000, that forms oligomers ranging in size from dimers with M_r 500 000 to species with M_r over 12 000 000, held together by disulfide bonds [reviewed in Hoyer (1981)]. von Willebrand factor is synthesized by endothelial cells (Jaffe et al., 1973, 1974) and by megakaryocytes (Nachman et al., 1977). The biosynthesis of von Willebrand factor is very complex and requires several posttransitional modifications, including proteolytic processing (Wagner & Marder, 1983; Lynch et al., 1983), glycosylation (Wagner & Marder, 1983, 1984), and sulfation (Browning et al., 1983).

von Willebrand factor is not an enzyme but participates in hemostatis through a variety of binding interactions. It forms a bridge between platelets and subendothelial connective tissue at sites of endothelial damage, and there are specific receptors for von Willebrand factor on each of these surfaces. In the presence of the antibiotic ristocetin, von Willebrand factor binds to glycoprotein Ib of resting platelets, causing platelet aggregation in plasma (Jenkins et al., 1976). The importance of this interaction is suggested by the bleeding diathesis that affects patients with the Bernard–Soulier syndrome, a disorder

† This work was supported in part by Research Grant HL29595 (to K.T.) from the National Institutes of Health. K.T. is a visiting professor of the Fujita-Gakuen Health University, Nagoya, Japan.

* Correspondence should be addressed to this author.

‡ Washington University School of Medicine.

§ University of Washington.

0006-2960/86/0425-3164$01.50/0

characterized by the lack of functional glycoprotein Ib. von Willebrand factor binds to a second site on activated platelets, and binding to this site is competitive with the binding of fibronectin or fibrinogen (Pietu et al., 1984; Haverstick et al., 1985; Plow et al., 1985). The corresponding platelet receptor appears to be the glycoprotein IIb/IIIa complex. The target in subendothelial connective tissue may be collagen, since von Willebrand factor will bind specifically to purified fibrillar collagens but not to gelatin (Santoro, 1981; Santoro & Cowan, 1982; Morton et al., 1983). However, von Willebrand factor also binds to extracellular matrix that is devoid of detectable collagen, so that additional receptors may exist in the subendothelium (Wagner et al., 1984). Finally, factor VIII binds to von Willebrand factor and is stabilized in the circulation by this interaction. In the absence of von Willebrand factor, the survival of factor VIII is dramatically reduced, causing a secondary deficiency of factor VIII (Tuddenham et al., 1982). Consequently, patients with von Willebrand's disease, who lack functional von Willebrand factor, may exhibit bleeding characteristic either of platelet dysfunction or of classical hemophilia A.

Structure–function relationships of von Willebrand factor have been partly elucidated by protein chemistry, immunology, and molecular biology methods. Staphylococcal V8 protease separates native von Willebrand factor into two major fragments that are not connected by disulfide bonds. The amino-terminal fragment is a homodimer of 170-kilodalton peptides that retains the ability to bind to platelet glycoprotein Ib (Girma et al., 1986) and to collagen (Fressinaud et al., 1985). A 50-kilodalton tryptic fragment that also binds to platelet glycoprotein Ib has been placed within the linear sequence of this fragment (Fujimura et al., 1986). The carboxy-terminal V8 protease fragment is a homodimer of 100-kilodalton peptides that binds to activated platelets at the glycoprotein IIb/IIIa complex (Girma et al., 1984). Monoclonal antibodies to von Willebrand factor that selectively inhibit these binding activities have been shown to recognize still smaller fragments of von Willebrand factor produced by digestion with other proteases (Sixma et al., 1984). We have previously reported that a tetrapeptide segment, Arg-Gly-Asp-ser, occurs in the carboxy-terminal V8 protease fragment. The same sequence is required for the cell attachment and platelet binding activities of fibronectin (Pierschbacher & Ruoslahti, 1984a,b), and small peptides containing this sequence inhibit the binding of both fibronectin and von Willebrand factor to activated platelets (Haverstick et al., 1985; Plow et al., 1985). Thus, the domain that binds to the platelet glycoprotein IIb/IIIa complex probably contains this tetrapeptide.

In this paper, we describe the isolation of cDNA[1] inserts for von Willebrand factor that complete the nucleotide sequence corresponding to the species found in plasma. In addition, analysis of the translated amino acid sequence reveals the presence of two previously unrecognized duplicated domains within the sequence, and one additional copy of domain B (Sadler et al., 1985). There are a total of five unrelated repeated domains in the sequence, two of which are triplicated and three of which are duplicated. Together, these repetitive sequences comprise over three-fourths of the entire subunit. The A domains may be homologous to a segment of complement factor B. The remaining domains do not have any homologue within other known protein sequences. In addition, the sequence of the von Willebrand factor precursor surrounding the amino terminus of the mature subunit is similar to that of several mammalian, viral, and yeast protein precursors, suggesting that the posttranslational proteolytic processing of these diverse proteins may proceed by similar mechanisms. Eight potential point mutations have been identified by comparing the nucleotide sequences of several cDNA isolates, of which four alter the predicted amino acid sequence.

Materials and Methods

The human umbilical vein endothelial cell cDNA library in λgt11, cDNA isolates λHvWF1 and λHvWF3, and methods for DNA preparation, subcloning, and sequencing have been described (Sadler et al., 1985). The reported nucleotide sequences were determined at least once on both strands. Deoxyadenosine 5′-[α-^{35}S]thiotriphosphate ([^{35}S]dATPαS) and cytidine [α-^{32}P]triphosphate ([α-^{32}P]CTP) were purchased from Amersham. Deoxycytidine [α-^{32}P]triphosphate ([α-^{32}P]dCTP) was purchased from New England Nuclear.

cDNA Library Screening. The 404 base pair cDNA insert of λHvWF1, corresponding to amino acids −24 to 110 of von Willebrand factor, was ligated into the *Eco*RI site of pGEM-1 (Riboprobe Gemini System, Promega-Biotec, Madison, WI), and the recombinant plasmid was propagated in *E. coli* HB101. Plasmid DNA was linearized with *Pvu*II (or *Bam*HI) and employed as a template for the SP6 (or T7) RNA polymerase to prepare RNA labeled with [α-^{32}P]CTP to a specific activity of 2×10^8 cpm/μg according to the instructions provided by the supplier.

A 2752 base pair *Fsp*I–*Sac*I fragment, corresponding to amino acids 544–1461 of von Willebrand factor, was prepared by digestion of a *Sac*I–*Sqc*I subclone of λHvWF3 in pUC18 (Sadler et al., 1985). The fragment was purified by polyacrylamide gel electrophoresis and electroelution and then labeled with [α-^{32}P]dCTP by nick translation (Maniatis et al., 1975) to a specific activity of $(1–2) \times 10^8$ cpm/μg.

Approximately 1 500 000 recombinant phage from the λgt11 cDNA library were plated on *E. coli* Y1088 at a density of 50 000 per 150-mm plate of LB agar and screened in duplicate by hybridization according to the method of Benton and Davis (1977) with both probes described above. Positive isolates were plaque-purified, and DNA was prepared for subcloning and sequencing.

Northern Blotting. Poly(A)+ RNA was prepared from cultured human umbilical vein endothelial cells (Sadler et al., 1985), and 5 μg was employed for Northern blotting according to Thomas (1983). Size standards consisting of bovine liver 28S and 18S RNA and *Hin*dIII fragments of phage λ DNA were electrophoresed in adjacent lanes and stained with ethidium bromide. The blot was prehybridized in 7.5 M sodium citrate, 0.75 M sodium chloride, pH 7.0 (5× SSC), 50% (v/v) formamide, 5 mM sodium phosphate, 0.1% (w/v) $NaDodSO_4$, 1 mM EDTA, 2.5× Denhart's solution, and 200 μg/mL denatured salmon sperm DNA, at 55 °C. The blot was hybridized in the same solution containing 500 000 cpm/mL of the λHvWF1 probe (antisense) at 55 °C for 18 h, washed 3 times for 20 min at 65 °C in 0.1× SSC and 0.1% (w/v) $NaDodSO_4$, and exposed to Kodak XAR-5 film for 20 h.

Computer Analysis of Sequences. The von Willebrand factor protein sequence was compared to all entries in the NBRF Protein Sequence Database (Georgetown University,

[1] Abbreviations: $NaDodSO_4$, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid; SSC, standard saline citrate (15 mM citrate, 0.15 M NaCl, pH 7.0); cDNA, complementary deoxyribonucleic acid; Denhardt's solution, 0.02% (w/v) bovine serum albumin, 0.02% (w/v) poly(vinylpyrrolidone), and 0.02% (w/v) Ficoll; NBRF, National Biomedical Research Foundation.

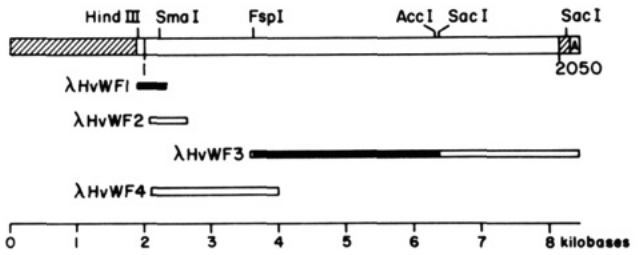


FIGURE 1: Restriction map of four human von Willebrand factor cDNA isolates. Only restriction sites used in subcloning and DNA sequencing are shown. In the summary map at the top of the figure, the hatched region to the left represents the leader peptide and 5′-noncoding sequences of the mRNA that are not represented in the cDNA isolates, and the hatched region at the right represents the 3′-noncoding sequence. The length of the complete von Willebrand factor cDNA (8.5 kilobases) is estimated from the Northern blot of Figure 2. The black regions of λHvWF1 and λHvWF3 indicate the fragments employed as hybridization probes in these studies. The scale at the bottom is in kilobases.

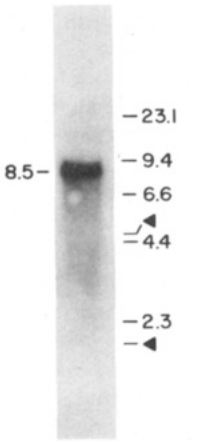


FIGURE 2: Northern blot of RNA from cultured human umbilical vein endothelial cells probed with a human von Willebrand factor cDNA probe. The mobility of 28S and 18S ribosomal RNA (▲) and denatured DNA size markers (kilobases) is indicated to the right of the figure. The interpolated size of the autoradiographic signal is indicated on the left.

Washington, DC, release 6.0, August 28, 1985) and separately to the amino acid sequence of human fibronectin (Kornblihtt et al., 1985), with the programs Search, Relate, and Align (Dayhoff et al., 1983). The nucleotide sequence of the von Willebrand factor cDNA corresponding to amino acid residues −24 through 2050 was compared to all entries in the Genbank genomic sequence data bank (BBN Laboratories Inc., Cambridge, MA, release 38.0, November 11, 1985) with the program Fastn (Wilbur & Lipman, 1983). The amino acid sequence of plasma von Willebrand factor, residues 1–2050, was analyzed for internal segment duplications with the programs Relate and Align (Dayhoff et al., 1983).

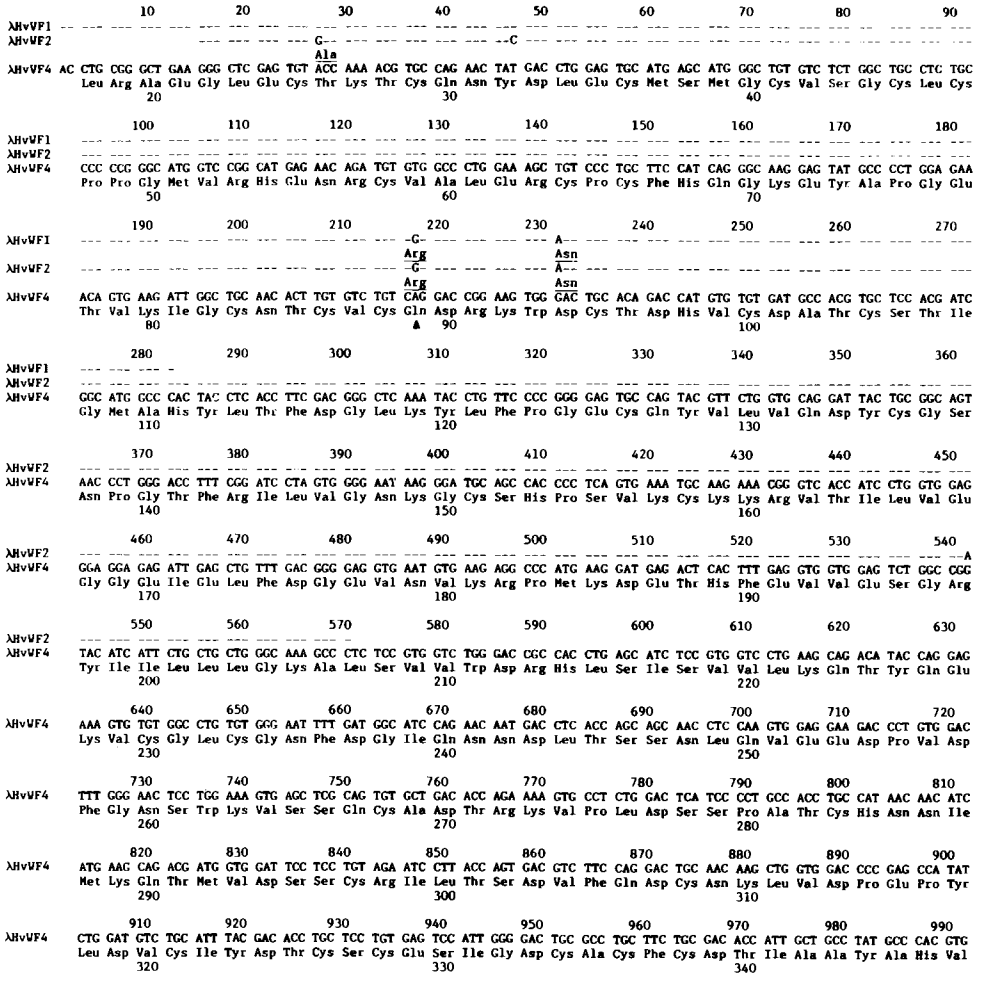
Results and Discussion

cDNA Library Screening and Sequence Comparisons. Among 1 500 000 recombinants screened, 16 hybridized with the λHvWF3 probe only, 1 with the λHvWF1 probe only (λHvWF2), and 1 with both probes (λHvWF4). The cDNA inserts from these two latter isolates were subcloned into M13mp18 for sequencing. A summary restriction map for the von Willebrand factor cDNA indicating the portion contained in these isolates is shown in Figure 1.

The cDNA insert of λHvWF2 consisted of 556 base pairs, encoding amino acids 22–206 of von Willebrand factor. A portion of the sequence overlapped with that determined previously for λHvWF1 (Sadler et al., 1985). The insert of λHvWF4 was 3.2 kilobases in length, and 1.9 kilobases at the 3′ end corresponded to amino acids 18–661 of von Willebrand factor. Therefore, this insert overlapped with both λHvWF1 and λHvWF3, specifying the remaining amino acid sequence between these two clones (Figure 1). However, the sequence of the first 1.3 kilobases of this insert did not match that of λHvWF1, nor did the translated amino acid sequence match that determined by amino acid sequencing of von Willebrand factor. In addition, this segment contains no continuous open-reading frame, and there is no suitable acceptor splice junction sequence prior to the codon for amino acid 18 of von Willebrand factor. Thus, this segment may represent an unrelated DNA sequence that was juxtaposed inadvertently during construction or propogation of the cDNA library. Together, the four cDNA isolates span 6.5 kilobases of von Willebrand factor cDNA, including the poly(A) tail.

The length of the von Willebrand factor mRNA was estimated to be 8.5 kilobases by Northern blotting, as shown in Figure 2. This value is consistent with those obtained by others using similar methods (Lynch et al., 1985; Ginsburg et al., 1985; Verweij et al., 1985). Since only 6.4 kilobases are necessary to encode the mature protein plus all 3′-noncoding mRNA sequences, 2 kilobases of mRNA might be available to encode a precursor or leader peptide as large as 700 amino acids, or 77 000 daltons (unglycosylated). On the basis of a comparison of amino-terminal protein sequence with translated von Willebrand factor cDNA sequence, Fay et al. (1985) have suggested that this precursor peptide is identical with von Willebrand antigen II, a 92 000–98 000-dalton (reduced) plasma protein of unknown function that is immunologically distinct from von Willebrand factor, originally described by Montgomery and Zimmerman (1978).

The sequences of isolates λHvWF2 and λHvWF4 are compared with those of λHvWF1 and λHvWF3 in Figure 3. This completes the nucleotide sequence of the cDNA corresponding to plasma von Willebrand factor. Compared to other human cDNA sequences, there are no striking bases in codon usage (Lathe, 1985). In the regions that overlap, there are eight single-nucleotide discrepancies, of which seven are transitions and one (λHvWF4, nucleotide 1526, G to T) is a transversion. Four of the transitions affect the translated amino acid sequence. For these discrepancies, the predicted amino acid residue that agrees with the protein sequence (Titani et al., 1986) is underlined in Figure 3. We previously reported a discrepancy at amino acid residue 7 between λHvWF1 (His) and the protein sequence (Pro) (Sadler et al., 1985). The cDNA library employed for these studies was derived from a pool of 30–60 umbilical veins, and none of the tissue donors was known to harbor an abnormal von Willebrand factor allele. Thus, there are a total of five potential protein sequence polymorphisms implied by these cDNA sequences. One of these, Ala/Thr at residue 26, has been confirmed directly by protein sequencing. In von Willebrand factor prepared from pooled factor VIII concentrate, both Ala and Thr are identified at this position in a ratio of approximately 4:1 (Titani et al., 1986). At present, there is no way to exclude the possibility that the remaining discrepancies arose through errors of transcription during cDNA library preparation or that they represent nonfunctional mRNA sequences. A comparison of the sequence of λHvWF3 with three other independently reported cDNA sequences does not show any additional discrepancies for the region encoding amino acids 1857 through the carboxy-terminus (Lynch et al., 1985; Ginsburg et al., 1985; Verweij et al., 1985). There are some disagreements between the sequence of the 3′-noncoding region



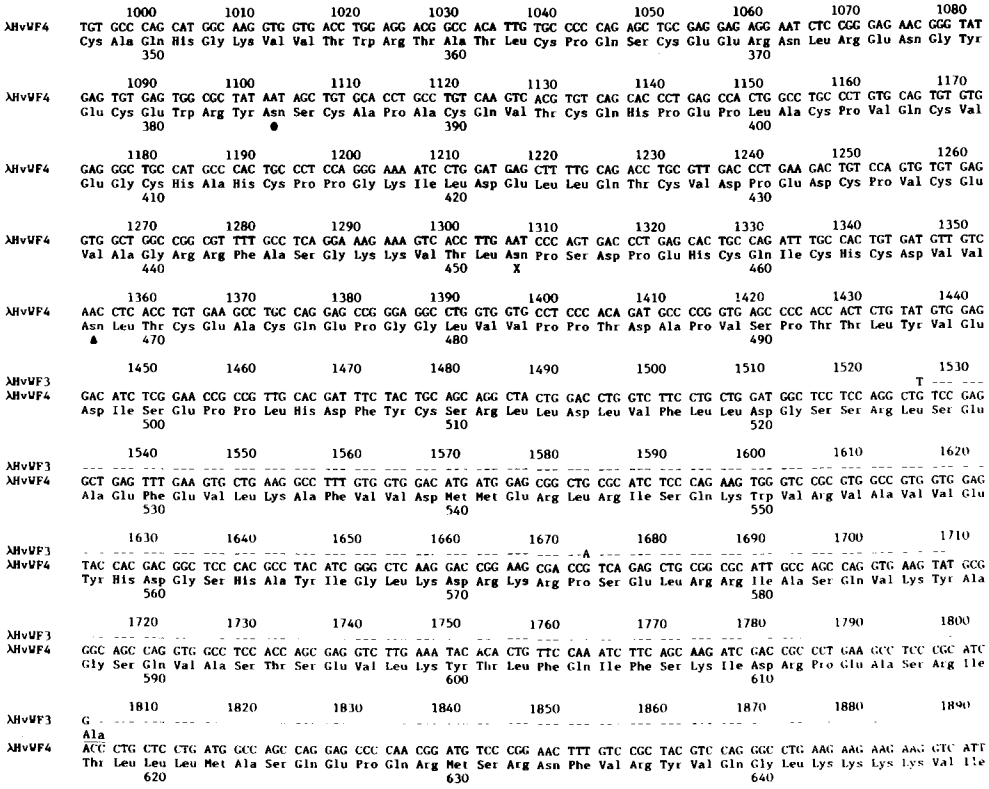


FIGURE 3: Nucleotide and translated amino acid sequence of human von Willebrand factor cDNA isolates. The sequence of λHvWF4 is shown in its entirety, and the nucleotide numbering begins with the first nucleotide of that isolate. The amino acid numbering assumes that the amino-terminal Ser of plasma von Willebrand factor is residue 1. Identity between regions of λHvWF1, λHvWF2, and λHvWF3 that overlap with λHvWF4 are indicated by dashes (---), and discrepancies are shown explicitly. Where the translated amino acid sequences of the cDNA isolates disagree, the amino acid that agrees with the major sequence determined independently for the protein (Titani et al., 1986) is underlined. Potential Asn-X-Thr/Ser glycosylation sites that are utilized in von Willebrand factor are indicated by closed triangles (▲), and one potential site that is not utilized is indicated by an X. The location of a glycosylated Asn in the sequence Asn-Ser-Cys is shown by a closed circle (●).

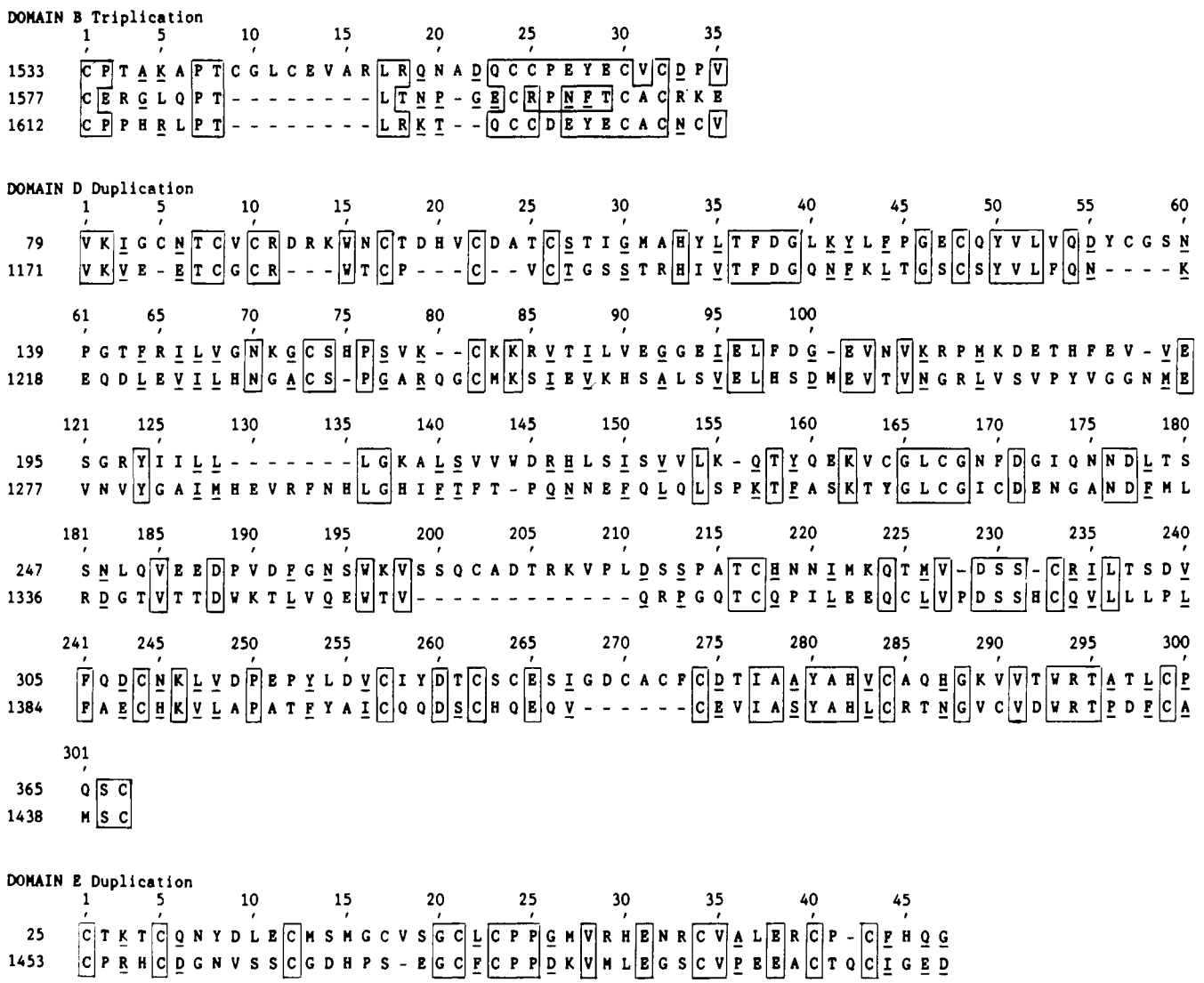


FIGURE 4: Alignment of repeated amino acid sequence domains in human von Willebrand factor. The amino acid numbering begins with the amino-terminal Ser of the plasma von Willebrand factor subunit. The alignments shown were obtained with the Align computer program (Dayhoff et al., 1983). The scoring matrix used was the mutation data matrix (MD+2), with a gap penalty of 6. Amino acids are shown by the single-letter code. Identical residues are shown enclosed in boxes, and conservative substitutions are underlined. In this figure, a conservative substitution is one with a positive value in the mutation data matrix (Dayhoff et al., 1983). Statistical parameters for these alignments are listed in Table I.

of λHvWF3 (Sadler et al., 1985) and that reported by Lynch et al. (1985).

There are two potential *N*-glycosylation sites in the translated amino acid sequence of λHvWF4 that lies in the gap between λHvWF1 and λHvWF3 (Figure 3). Thus, there are a total of 13 Asn-X-Thr/Ser sequences in the von Willebrand factor subunit. The first of these, at Asn-94, is abolished by the A to G transition at nucleotide 231 of λHvWF4. The potential site at Asn-452 is not glycosylated, and Asn-384, in the unusual sequence Asn-Ser-Cys, is glycosylated in plasma von Willebrand factor. The actual sites of Asn-linked and Thr/Ser-linked glycosylation have been determined directly by protein sequencing (Titani et al., 1986).

Internal Homologies in von Willebrand Factor Protein Sequence. Five unrelated sequences are repeated within the von Willebrand factor subunit. Three of these (domains A–C) have been described previously (Sadler et al., 1985). Domain A contains 193–220 amino acids and is present in three tandem copies between residues 497 and 1111 (not shown). The sequence of λHvWF4 (Figure 3) completes the amino-terminal sequence of domain A1. Domain B contains 25–35 amino acids and is present in three copies between residues 1533 and 1636 (Figure 4). Further analysis has revealed a third copy (domain B2) between the previously reported pair. The statistical similarity of domain B2 to domains B1 and B3 is the least significant of all segment comparisons (Table I), but the length of domain B2 is the same as that of domain B3, many residues (particularly Cys) are shared by all three domains, and domain B2 is precisely centered between B1 and B3. Thus, we propose domain B2 as a faint but definite member of the domain B group of peptide segments. Domain C contains 116–119 amino acids and is duplicated between residues 1637 and 1899 (not shown).

In contrast to the virtually contiguous replication of domains A–C, two additional types of repeated domain can be identified within the sequence of λHvWF4 or λHvWF3 that are separated from their homologues by unrelated repeated domains. Domain D1 contains 289 amino acids between residues 79 and 367, while domain D2 contains 270 amino acids between residues 1171 and 1440 (Figure 4). These domains are separated from each other by a total of 804 residues that contain the triplicated A domains. Domain E1 consists of 46 amino acids between residues 25 and 70, while domain E2 consists of 46 amino acids between residues 1453 and 1498 (Figure

Table I: Alignment Scores for Comparisons between Repeated Domains of Human von Willebrand Factor[a]

first domain (residues)	second domain (residues)	alignment score	
		SD units	probability
A1 (497–716)	A2 (717–909)	10.08	$\leq 10^{-23}$
A1 (497–716)	A3 (910–1111)	8.68	$\leq 10^{-16}$
A2 (717–909)	A3 (910–1111)	10.59	$\leq 10^{-23}$
B1 (1533–1567)	B2 (1577–1602)	3.08	1.04×10^{-3}
B1 (1533–1567)	B3 (1612–1636)	5.26	7.21×10^{-8}
B2 (1577–1602)	B3 (1612–1636)	2.95	1.59×10^{-3}
C1 (1637–1752)	C2 (1781–1899)	10.04	$\leq 10^{-23}$
Cx fragment (63–90)	C1 fragment (1661–1692)	5.07	1.99×10^{-7}
Cx fragment (63–90)	C2 fragment (1812–1844)	6.38	1.00×10^{-10}
D1 (79–367)	D2 (1171–1440)	15.24	$\leq 10^{-30}$
E1 (25–70)	E2 (1453–1498)	6.00	1.00×10^{-9}

[a] Scores are given in standard deviation units and also as estimated probabilities that the proposed alignment could have occurred by chance. Details of the parameters employed for the computer program Align are described in the legend to Figure 4. The number of random runs used to correct for biases introduced by amino acid composition was 100 (Dayhoff et al., 1983).

4). They are separated by 1383 residues that include the triplicated A domains and the duplicated D domains. The region corresponding to the A domains contains only six Cys residues, while the flanking B–E domains are extremely Cys rich, containing a total of 163 Cys. Finally, a short segment (domain Cx fragment, residues 63–90) is very similar to small segments of both domain C1 (residues 1661–1692) and domain C2 (residues 1812–1844). Because this segment overlaps with sequences that we have chosen to include in domains D1 and E1 and because it is very short compared to the other C domains, we have not included it in Figure 4. However, it might represent a remnant of another member of the C-domain family. The segment comparison scores in standard deviation units for the alignments of all proposed repeated domains are shown in Table I.

Homology of von Willebrand Factor to Other Proteins. The amino acid sequence of von Willebrand factor has been compared to all entries in the NBRF Protein Sequence Database. From the ALIGN program (Dayhoff et al., 1983), the A domains of von Willebrand factor appear to be homologous to a portion of complement factor B. For each alignment of an A domain with residues 230–454 of human complement factor B (Mole et al., 1984), 20–24% of the amino acids are identical, and the alignment scores are of range from 5.01 to 5.72 standard deviation units (probability 2.7×10^{-7} to 5.4×10^{-9}). This segment of complement factor B contains the Arg^{234}–Lys^{235} bond that is cleaved during activation by factor D, five residues of the resulting factor Ba peptide, and the amino-terminal 220 residues of the factor Bb polypeptide that precede the serine protease domain. The gene structure of complement factor B indicates that the region encoding this polypeptide contains five intron–exon boundaries at residues 228, 274, 320, 364, and 398 (Campbell et al., 1984). Domains A1 and A2 of von Willebrand factor are encoded together by a single exon; however, domain A3 (residues 910–1111) is interrupted by intron–exon boundaries at amino acid residues 922, 961, 1008, 1056, and 1111 (J. M. Sorace and J. E. Sadler, unpublished results). There is a rough correlation between the location of these intron–exon boundaries in both genes. Twenty-five residues of the corresponding segment of complement component C2a have been reported (Parkes et al., 1983). This sequence also appears to align well with the von Willebrand factor A domains, although the sequence for C2a is too short to evaluate in detail. The biological significance of these potential distant relationships is not known.

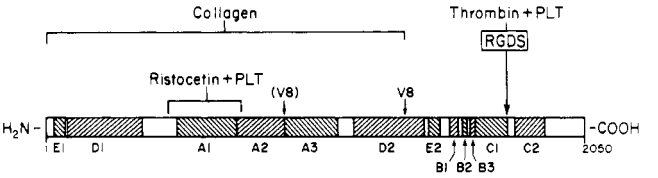


FIGURE 5: Structure–function relationships for human von Willebrand factor. The numbers below the schematic subunit indicate the amino-terminal Ser (1) and the carboxy-terminal Lys (2050). The two principal sites of staphylococcal protease V8 cleavage are indicated by arrows, with the minor site in parentheses. The shaded areas show the extent of internal sequence duplications, or domains, as discussed in the text. The location of a peptide fragment that binds to platelet GPIb is shown by Ristocetin + PLT. The position of the Arg-Gly-Asp-Ser tetrapeptide (RGDS) that may mediate binding to the GPIIb/IIIa complex of activated platelets is shown by Thrombin + PLT. The collagen binding activity of von Willebrand factor is localized to the amino-terminal major V8 protease fragment, as shown.

Aside from these two candidate proteins, the amino acid sequence of von Willebrand factor shows no apparent homology to any protein in the NBRF Protein Sequence Database, and the nucleic acid sequence of the four cDNA isolates is not homologous to any sequence in the Genbank Genomic Sequence Data Bank.

Structure–Function Relationships of von Willebrand Factor. The sequence of the von Willebrand factor precursor surrounding the amino-terminal Ser of the form found in plasma resembles the posttranslational cleavage sites for many other protein precursors. Common features include a Lys-Arg or Arg-Arg dipeptide on the amino-terminal side of the cleavage site. These comparisons suggest that quite distantly related eukaryotic organisms might use similar mechanisms to generate mature secreted protein products. The processing of the von Willebrand factor precursor to its mature form is known to be posttranslational and intracellular and is presumed to occur in the Golgi apparatus (Wagner & Marder, 1983; Lynch et al., 1983). Some proteins, such as proalbumin and many prohormones, are processed intracellularly, too [reviewed in Docherty & Steiner (1982)]. However, despite the similar sequences at the site of protease cleavage, proapolipoprotein A-II is processed extracellularly by a hepatocyte line that processes proalbumin intracellularly (Gordon et al., 1984). Thus, several pathways must exist for the processing of these precursors, even though they share certain structural features. The identity of the required proteases is not known with certainty, although cathepsin B like thiol proteases may participate in both intracellular and extracellular processing reactions (Gordon et al., 1985; Docherty et al., 1982). The *kex2* mutants of yeast are deficient in the protease that is required for the analogous processing of the mating factor α precursor (Julius et al., 1984). So far, no homologue for this enzyme has been found in higher eukaryotes.

Some proposed structure–function relationships for mature von Willebrand factor are summarized in Figure 5. Staphylococcal protease V8 cleaves the protein into two major fragments, SPII and SPIII. Fragment SPII is a homodimer of 100-kilodalton peptides containing the carboxy terminus of the subunit, while SPIII is a homodimer of 170-kilodalton amino-terminal peptides. The amino-terminal fragment SPIII retains the ability to bind to both collagen (Fresinaud et al., 1985) and platelet glycoprotein Ib (Girma et al., 1986), and the carboxy-terminal fragment SPII binds to the glycoprotein IIb/IIIa complex of thrombin-activated platelets. The collagen binding activity has not been further localized, but a dimer of 50-kilodalton tryptic fragments of von Willebrand factor has been shown to bind to platelet glycoprotein Ib in the

presence or absence of ristocetin, and this activity is stable to reduction (Fujimura et al., 1986). By protein sequencing, this fragment has been identified as residues 449–729 of the von Willebrand factor subunit. This segment includes all of domain A1, suggesting that at least a part of the binding to glycoprotein Ib is mediated by this small portion of the von Willebrand factor subunit. This fragment was isolated as a dimer from tryptic digests of unreduced von Willebrand factor, and since there are only seven Cys residues in this segment, at least one of them must participate in an intersubunit disulfide bond. Preliminary characterization of the von Willebrand factor gene shows that this functional segment may be encoded by a single exon that species amino acid residues 463–921, including all of domains A1 and A2 (J. M. Sorace and J. E. Sadler, unpublished results).

The tetrapeptide Arg-Gly-Asp-Ser occurs between amino acid residues 1744 and 1747. This sequence also occurs in fibronectin and fibrinogen and appears to mediate the binding of these proteins to the GPIIb/IIIa complex of activated platelets (Pierschbacher & Ruoslahti, 1984a,b). Synthetic peptides containing this sequence inhibit the binding of fibrinogen, fibronectin, and von Willebrand factor to activated platelets (Gartner & Bennett, 1985; Haverstick et al., 1985; Plow et al., 1985). Thus, this small segment of von Willebrand factor may be required for binding to the GPIIb/IIIa complex of activated platelets. It has been placed tentatively at the carboxy-terminal end of domain C1 (Sadler et al., 1985). A factor VIII binding site has not been localized within the linear sequence of von Willebrand factor.

One challenge in the study of proteins with highly repetitive structures, such as fibronectin or von Willebrand factor, is to correlate domains that have been identified by examination of the linear sequence with higher orders of structure such as disulfide bonding patterns, with the organization of the corresponding genomic DNA sequences into introns and exons, and with specific biological functions. Future studies will extend these relationships for von Willebrand factor.

ACKNOWLEDGMENTS

We thank Earl W. Davie for his interest and encouragement, Torben E. Petersen for helpful discussions concerning sequence comparisons, and John R. Bell for his help in synthesizing oligonucleotides.

Registry No. von Willebrand factor, 9001-27-8.

REFERENCES

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–182.

Browning, P. J., Ling, E. H., Zimmerman, T. S., & Lynch, D. C. (1983) *Blood 62*, 281 (Abstr.).

Campbell, R. D., Bentley, D. R., & Morley, B. J. (1984) *Philos. Trans. R. Soc. London, B 306*, 367–378.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524–545.

Docherty, K., & Steiner, D. F. (1982) *Annu. Rev. Physiol. 44*, 625–638.

Docherty, K., Carroll, R. J., & Steiner, D. F. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4613–4617.

Fay, P. J., Kawai, Y., Wagner, D. D., Ginsburg, D., Bonthron, D., Ohlsson-Wilhelm, B. M., Chavin, S. I., Abraham, G. N., Handin, R. I., Orkin, S. H., Montgomery, R. R., & Marder, V. J. (1985) *Blood 66*(Suppl. 1), 333a.

Fujimura, Y., Titani, K., Holland, L. Z., Russell, S. R., Roberts, J. R., Elder, J. H., Ruggeri, Z. M., & Zimmerman, T. S. (1986) *J. Biol. Chem. 261*, 381–385.

Gartner, T. K., & Bennett, J. S. (1985) *J. Biol. Chem. 260*, 11891–11894.

Ginsburg, D., Handin, R. I., Bonthron, D. T., Donlon, T. A., Bruns, G. A. P., Latt, S. A., & Orkin, S. H. (1985) *Science (Washington, D.C.) 228*, 1401–1406.

Girma, J.-P., Pietu, G., Chopek, M. W., Edgington, T. S., & Meyer, D. (1984) *Circulation 70*(Suppl. II), Abstr. 836.

Girma, J.-P., Chopek, M. W., Titani, K., & Davie, E. W. (1986) *Biochemistry* (second paper of four in this issue).

Gordon, J. I., Sims, H. F., Edelstein, C., Scanu, A. M., & Strauss, A. W. (1984) *J. Biol. Chem. 259*, 15556–15563.

Gordon, J. I., Sims, H. F., Edelstein, C., Scanu, A. M., & Strauss, A. W. (1985) *J. Biol. Chem. 260*, 14824–14831.

Haverstick, D. M., Cowan, J. F., Yamada, K. M., & Santoro, S. A. (1985) *Blood 66*, 946–952.

Hoyer, L. W. (1981) *Blood 58*, 1–13.

Jaffe, E. A., & Hoyer, L. W. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1906–1909.

Jaffe, E. A., Hoyer, L. W., & Nachman, R. L. (1973) *J. Clin. Invest. 52*, 2757–2764.

Jenkins, C. S. P., Philips, D. R., Clemetson, K. J., Meyer, D., Larrieu, M.-J., & Luscher, E. F. (1976) *J. Clin. Invest. 57*, 112–124.

Julius, D., Brake, A., Blair, L., Kunisawa, R., & Thorner, J. (1984) *Cell (Cambridge, Mass.) 37*, 1074–1089.

Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K., & Baralle, F. E. (1985) *EMBO J. 4*, 1755–1759.

Lathe, R. (1985) *J. Mol. Biol. 183*, 1–12.

Lynch, D. C., Williams, R., Zimmerman, T. S., Kirby, E. P., & Livingston, D. M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2738–2742.

Lynch, D. C., Zimmerman, T. S., Collins, C. J., Brown, M., Morin, M. J., Ling, E. H., & Livingston, D. M. (1985) *Cell (Cambridge, Mass.) 41*, 49–56.

Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.

Mole, J. E., Anderson, J. K., Davison, E. A., & Woods, D. E. (1984) *J. Biol. Chem. 259*, 3407–3412.

Montgomery, R. R., & Zimmerman, T. S. (1978) *J. Clin. Invest. 61*, 1498–1507.

Morton, L. F., Griffin, B., Pepper, D. S., & Barnes, M. J. (1983) *Thromb. Res. 32*, 545–556.

Nachman, R. L., Levine, R., & Jaffe, E. A. (1977) *J. Clin. Invest. 60*, 914–921.

Parkes, C., Gagnon, J., & Kerr, M. A. (1983) *Biochem. J. 213*, 201–209.

Pierschbacher, M. D., & Ruoslahti, E. (1984a) *Nature (London) 309*, 30–33.

Pierschbacher, M. D., & Ruoslahti, E. (1984b) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5985–5988.

Pietu, G., Cherel, G., Marguerie, G., & Meyer, D. (1984) *Nature (London) 308*, 648–649.

Plow, E. F., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. A., & Ginsburg, M. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8057–8061.

Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K., & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6394–6398.

Santoro, S. A. (1981) *Thromb. Res. 21*, 689–693.

Santoro, S. A., & Cowan, J. F. (1982) *Collagen Relat. Res. 2*, 31–43.

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.

Thomas, P. S. (1983) *Methods Enzymol. 100*, 255–266.

Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., & Fujikawa, K. (1986) *Biochemistry* (fourth paper of four in this issue).
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.
Verweij, C. L., de Vries, C. J. M., Distel, B., van Zonneveld, A.-J., van Kessel, A. G., van Mourik, J. A., & Pannekoek, H. (1985) *Nucleic Acids Res. 13*, 4699-4717.
Wagner, D. D., & Marder, V. J. (1983) *J. Biol. Chem. 258*, 2065-2067.
Wagner, D. D., & Marder, V. J. (1984) *J. Cell Biol. 99*, 2123-2130.
Wagner, D. D., Urban-Pickering, M., & Marder, V. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 471-475.
Wilbur, W. J., & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 726-730.

Amino Acid Sequence of Human von Willebrand Factor†

Koiti Titani,*,‡ Santosh Kumar,‡ Koji Takio,‡ Lowell H. Ericsson,‡ Roger D. Wade,‡ Katsuro Ashida,‡ Kenneth A. Walsh,‡ Michael W. Chopek,‡,§ J. Evan Sadler,∥ and Kazuo Fujikawa‡

Department of Biochemistry, University of Washington, Seattle, Washington 98195, and Departments of Medicine and Biochemistry, Washington University School of Medicine, St. Louis, Missouri 63110

Received January 14, 1986; Revised Manuscript Received February 25, 1986

ABSTRACT: The complete amino acid sequence of human von Willebrand factor (vWF) is presented. Most of the sequence was determined by analysis of the S-carboxymethylated protein. Some overlaps not provided by the protein sequence analysis were obtained from the sequence predicted by the nucleotide sequence of a cDNA clone [Sadler, J. E., Shelton-Inloes, B. B., Sorace, J., Harlan, M., Titani, K., & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6391-6398]. The protein is composed of 2050 amino acid residues containing 12 Asn-linked and 10 Thr/Ser-linked oligosaccharide chains. One of the carbohydrate chains is linked to an Asn residue in the sequence Asn-Ser-Cys rather than the usual Asn-X-Ser/Thr sequence. The sequence of von Willebrand factor includes several regions bearing evidence of internal gene duplication of ancestral sequences. The protein also contains the tetrapeptide sequence Arg-Gly-Asp-Ser (at residues 1744-1747), which may be a cell attachment site, as in fibronectin. The amino- and carboxyl-terminal regions of the molecule contain clusters of half-cystinyl residues. The sequence is unique except for some homology to human complement factor B.

Human von Willebrand factor (vWF)[1] is a plasma glycoprotein (Legaz et al., 1973; Shapiro et al., 1973; Olson et al., 1977) that is involved in platelet adhesion to the subendothelium, leading to platelet plug formation during vascular injury (Jorgensen & Borchgrevin, 1964; Havig & Stormoken, 1974). The prolonged bleeding time of individuals having low levels of vWF or modified vWF is due to poor platelet plug formation (Ruggeri et al., 1982; Hoyer, 1982; Kinoshita et al., 1984).

vWF is synthesized in endothelial cells (Jaffe et al., 1973; Jaffe & Hoyer, 1974) and megakaryocytes (Nachman et al., 1977) in a large precursor form and secreting into plasma after several processing events, including glycosylation, sulfation, disulfide formation, and proteolytic cleavages (Wagner & Marder, 1983, 1984; Browning et al., 1983; Lynch et al., 1983; Ling et al., 1984). It circulates in plasma as a series of high molecular weight multimers ranging in size from 1×10^6 to 12×10^6 daltons (Counts et al., 1978; Perret et al., 1979; Ruggeri & Zimmerman, 1980; Hoyer & Shainoff, 1980; Meyer et al., 1980). Electron micrographs suggest that extended protomers of 100-120 nm in length assemble into the multimeric structures that circulate in plasma (Slayter et al., 1985; Fowler et al., 1985).

We have established a large-scale purification procedure for human vWF from a commercial factor VIII concentrate and presented preliminary evidence that it is composed of identical subunits of approximate M_r 270000. This conclusion was based on the observation of a single amino acid sequence at the amino terminus as well as at the carboxyl terminus. This was confirmed by the agreement between the number of unique cyanogen bromide fragments and that predicted from the methionine content of the protein (Chopek et al., 1986). We have also shown that limited proteolysis of native vWF by *Staphylococcus aureus* V8 protease produces two major fragments that can be separated without cleaving the disulfide bonds. One of these (fragment III) is a 170K-dalton segment from the amino terminus that retains binding activity to ristocetin-treated platelets (Girma et al., 1986). The other (fragment II) is a 100K-dalton fragment from the carboxyl-terminal end of the protein. It binds platelets activated by either ADP or thrombin (Girma et al., 1984). A third (minor) fragment (fragment I) is a 50K-dalton subdigestion product of fragment III.

† This work was supported by Research Grants HL29595 (to K. Titani) and HL16919 (to E.W.D.) from the National Institutes of Health. K. Titani is a visiting professor of the Fujita-Gakuen Health University, Nagoya, Japan. K. Takio is a senior associate of the Howard Hughes Medical Institute Laboratory in Seattle. J.E.S. is an Associate Investigator of the Howard Hughes Medical Institute in Washington University, St. Louis, MO.

‡ University of Washington.
§ Present address: St. Paul Regional Red Cross, St. Paul, MN 55107, or Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455.
∥ Washington University School of Medicine.

[1] Abbreviations: vWF, von Willebrand factor; HPLC, high-performance liquid chromatography; CM, carboxymethyl; RP, reversed phase; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/86/0425-3171$01.50/0