The EMBO Journal vol.5 no.8 pp.1839-1847, 1986

Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protein considerably larger than the mature vWF subunit

Cornelis L.Verweij, Paul J.Diergaarde, Margreet Hart and Hans Pannekoek

Department of Molecular Biology, Central Laboratory of The Netherlands Red Cross Blood Transfusion Service, Amsterdam, The Netherlands

Communicated by P.Borst

Full-length human von Willebrand factor (vWF) cDNA was assembled from partial, overlapping vWF cDNAs. This cDNA construct includes a coding sequence of 8439 nucleotides which encode a single-chain precursor of 2813 amino-acid residues, representing a putative signal peptide, a pro-sequence and mature vWF of 22, 741 and 2050 amino acids, respectively. This represents the longest coding sequence determined to date. *In-vitro* expression of full-length vWF cDNA revealed the synthesis of a polypeptide with a mol. wt corresponding with that of the unglycosylated precursor. The precursor is a highly repetitive protein which consists of two duplicated (B, C), a triplicated (A), a quadruplicated (D) and a partly duplicated domain (D'), in the following order: H-D1-D2-D'-D3-A1-A2-A3-D4-B1-B2-C1-C2-OH. Both the pro-sequence, composed of two D domains (D1, D2), and mature vWF harbor an arg-gly-asp ('R-G-D') sequence which has been implicated in cell-attachment functions. It is argued that the pro-sequence is equivalent to von Willebrand Antigen II (vW AgII).

Key words: von Willebrand factor/cDNA cloning/*in-vitro* expression/domain structure/RGD tripeptide

Introduction

The von Willebrand factor (vWF) is a large, multimeric plasma protein, composed of an apparently single glycoprotein with a mol. wt of about 225 000. These subunits are linked together by disulfide bonds. In plasma, vWF circulates as multimers, ranging from dimers to multimers of more than 50 subunits (Van Mourik and Bolhuis, 1978; Hoyer and Shainoff, 1980; Ruggeri and Zimmerman, 1980). Dimers consist of two subunits joined, probably at their C-termini, by flexible 'rod-shaped' domains and are presumed to be the protomers in multimerization (Perret *et al.*, 1979). The protomers are linked through large, probably N-terminal, globular domains to form multimers (Fowler *et al.*, 1985).

vWF is synthesized by endothelial cells (Jaffe *et al.*, 1973) and megakaryocytes (Nachman *et al.*, 1977). It is believed that this protein is initially produced as a 240 000-260 000-glycosylated precursor (Wagner and Marder, 1983; Lynch *et al.*, 1983) that is subsequently subjected to carbohydrate processing, dimerization, multimerization and to proteolytic cleavage to yield the mature 225 000 subunit (Sporn *et al.*, 1985; Wagner and Marder, 1984). vWF is stored in the Weibel-Palade bodies within the endothelial cells (Wagner *et al.*, 1982). It cannot be excluded that these organelles play a role in the processing of the precursor protein.

vWF participates in critical steps in hemostasis. It is involved in platelet-vessel wall interactions after vascular injury, leading to platelet plug formation (Sakariassen *et al.*, 1979). On the vWF protein, domains have been assigned which show specific interaction with the platelet glycoproteins IB (Jenkins *et al.*, 1976), IIB/IIIA (Fujimoto and Hawiger, 1982), collagens type I and III (Houdijk *et al.*, 1985) and with another, yet unidentified, component (Ph.G. De Groot, M.Ottenhof-Rovers, J.A. Van Mourik and J.J.Sixma, personal communication) in the subendothelium. These assignments are based on studies with monoclonal anti-vWF antibodies which are able to inhibit a particular interaction of vWF. For a full analysis of structure-function relationships of the vWF protein, a full-length vWF cDNA will be indispensable. Introduction of well-defined mutations within this cDNA and expression of the mutated cDNA in a suitable host will allow a detailed localization of functional domains within the vWF protein. Recently, we and others (Lynch *et al.*, 1985; Ginsburg *et al.*, 1985; Verweij *et al.*, 1985; Sadler *et al.*, 1985) have cloned partial vWF cDNA sequences. The presence of a short 3' untranslated region (136 nt) on vWF mRNA, which extends to about 9000 nt, led us to assume that the precursor protein for vWF has a mol. wt considerably larger than the reported 240 000-260 000. A full-length vWF cDNA will enable us to elucidate the enigma of the mol. wt of the precursor, characterize its processing pathway and establish the primary structure.

In this paper, we report on the isolation and the nucleotide sequence of cDNAs, spanning the entire vWF mRNA, and on the assembly of these sequences into a full-length, functional vWF cDNA.

Results

Construction of partial vWF cDNA clones and assembly of full length vWF cDNA

Previously, we and others have reported on the construction of plasmids containing part of a full-length human vWF cDNA (Verweij *et al.*, 1985; Lynch *et al.*, 1985; Ginsburg *et al.*, 1985; Sadler *et al.*, 1985). The most extended vWF cDNA, that we obtained from an oligo(dT)-primed human endothelial cDNA library, comprised about 2280 bp (pvWF2280). Nucleotide sequence analysis revealed that this cDNA insert has been initiated at the poly(A) tail of vWF mRNA. To construct a full-length vWF cDNA, we have isolated additional, overlapping vWF cDNA sequences which are located upstream of pvWF2280. For that purpose, two biochemical selections were employed to enrich for the number of vWF cDNA harboring plasmids. Firstly, oligonucleotide primers, derived from the partial nucleotide sequence (Sadler *et al.*, 1985), were synthesized to direct cDNA synthesis with human endothelial poly(A)$^+$ RNA as substrate. Secondly, cDNA preparations were digested with particular restriction endonucleases, known to dissect vWF cDNA at a limited number of sites (Ginsburg *et al.*, 1985; Sadler *et al.*, 1985). The cloning strategy is outlined in Figure 1A. The plasmids, containing adjacent vWF cDNA sequences, were designated pvWF1330, pvWF1800, pvWF2600, pvWF2084 and pvWF2280. The nucleotide sequence of the 5' end of the cDNA insert of

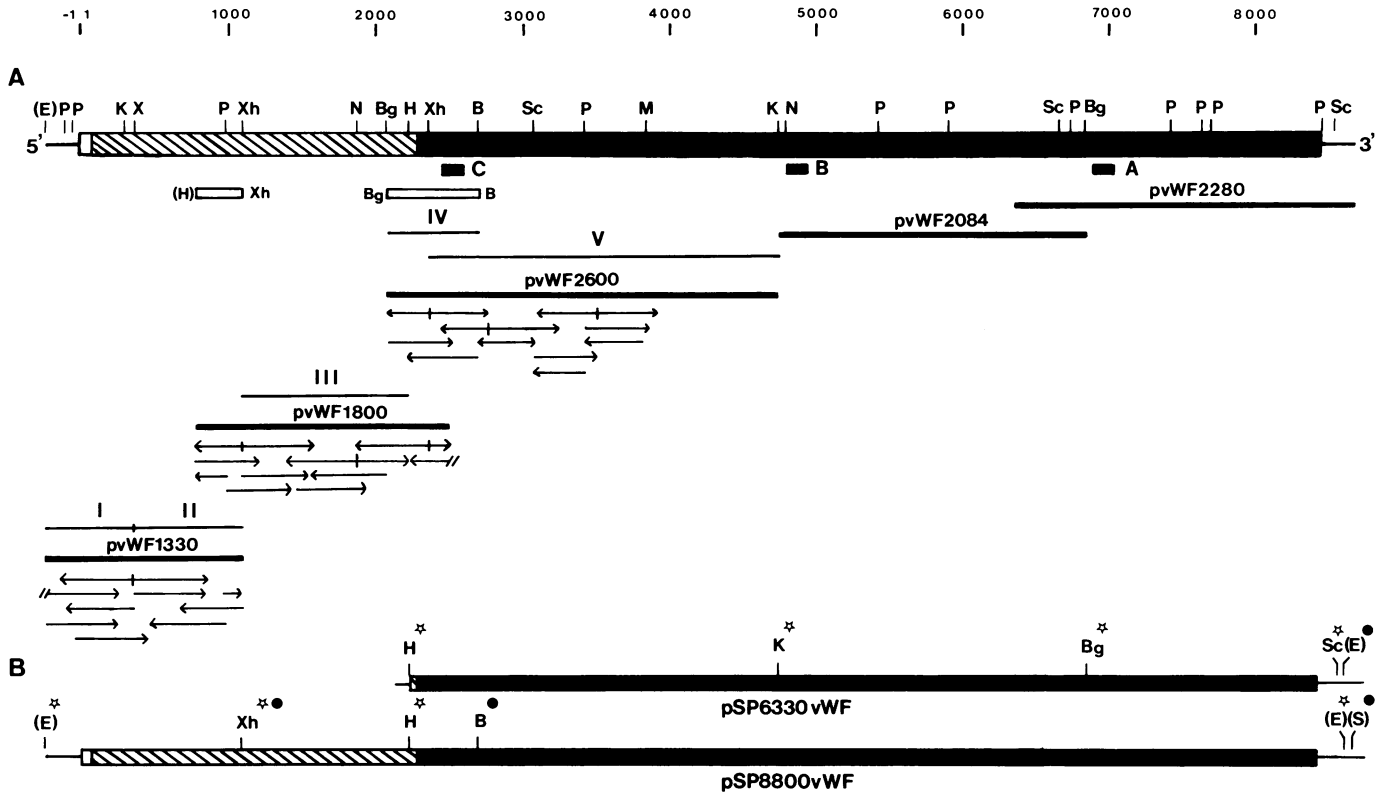


Fig. 1. Strategy for the construction of vWF cDNAs, the assembly of full-length vWF cDNA and the determination of the nucleotide sequence. (**A**) vWF mRNA is indicated by a bar; open area, signal peptide coding region; hatched area, pro-sequence coding region; solid area, mature vWF coding region. The oligonucleotides (20-mers) A (6901–6921), B (4819–4839) and C (2467–2487), which were used for primer-directed cDNA synthesis and/or as probe for hybridizations, are indicated by small bars. The 575-bp *Bgl*II–*Bam*HI and the 350-bp *Hin*dIII–*Xho*I fragments which were used as probes for colony screening are indicated by open bars. Below the schematic representation of vWF mRNA, the five partial, adjacent vWF cDNAs are given which were used for the assembly of full-length vWF cDNA and for nucleotide sequencing. The fragments I, II, III, IV and V, which were used in the S1 nuclease protection experiments, are shown above the vWF cDNA insert from which they were derived. The arrows indicate the nucleotide sequencing strategy. In the case of sequence analysis according to the procedure of Maxam and Gilbert (1977), the position of the radioactive labeling is given by a short vertical line at the end of an arrow. The slashes at the end of arrows mean that the end labeling was at a terminus, specified by vector DNA. Only restriction endonuclease sites which are relevant in this study are given. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; M, *Mst*I; N, *Nar*I; P, *Pvu*II; S, *Sal*I; Sc, *Sac*I; X, *Xba*I; Xh, *Xho*I. (**B**) Assembly of full-length vWF cDNA. Plasmid pSP6330vWF contains a 6331-bp vWF cDNA sequence, extending from the *Hin*dIII site (position 2235) till the *Sac*I site (position 8562), subcloned in vector pSP64. Plasmid pSP8800vWF includes full-length vWF cDNA, extending from the *Eco*RI site (see **Panel A**) till the *Sac*I site (position 8562), subcloned in vector pSP65. Restriction endonuclease sites, delimiting the fragments used for the assembly of full-length vWF cDNA, are indicated with an asterix. The *Eco*RI site at the 5′ end of full-length vWF cDNA originates from the *Eco*RI linker, used for the construction of pvWF1330 DNA. The sites for restriction enzymes, which were employed to linearize plasmid DNAs for *in-vitro* 'run off' transcription by SP6 RNA polymerase, are indicated by a dot. The *Sal*I site in plasmid pSP8800vWF and the *Eco*RI site in plasmid pSP6330vWF are present in the polylinkers of the pSP-type vectors.

pvWF1330 DNA (corresponding with the 5′ part of vWF mRNA) revealed that nonsense codons were present in all three reading frames. From this finding, we conclude that pvWF1330 DNA extends beyond the translation initiation codon.

S1-nuclease protection experiments with human endothelial RNA were performed to prove that the various vWF cDNA inserts are fully complementary to vWF mRNA. The construction of the probes and the conditions used are described in Materials and methods. The results are shown in Figure 2. In all cases, the length of the protected fragments is in accord with the length of the vWF cDNA sequences present in the different probes. From these data, we conclude that the vWF cDNA inserts of, respectively, pvWF1330, pvWF1800 and pvWF2600 DNA are entirely complementary to vWF mRNA. The nucleotide sequence of the remaining cDNA inserts of plasmids pvWF2084 and pvWF2280 were shown to correspond with the published sequence (Sadler *et al.*, 1985). Consequently, the different, adjacent vWF cDNA sequences are genuine copies of vWF mRNA.

The vWF cDNA sequences that we have constructed span a length of about 8900 bp. This length is consistent with the size of vWF mRNA, determined by Northern blot analysis of human endothelial poly(A)$^+$ RNA (Lynch *et al.*, 1985; Ginsburg *et al.*, 1985; Verweij *et al.*, 1985). A detailed description of the assembly of full-length vWF cDNA is given in Materials and methods and Figure 1B. The correct composition of the assembled, full-length vWF cDNA was established by restriction enzyme analysis.

Nucleotide sequence of full-length vWF cDNA

Nucleotide sequence analysis of vWF cDNA fragments was carried out both by the chemical degradation method (Maxam and Gilbert, 1977) and by the dideoxy chain-termination procedure (Sanger *et al.*, 1977), according to the scheme outlined in Figure 1A. In Figure 3, the nucleotide sequence of 4429 residues, extending from the 5′ end of vWF mRNA, and the corresponding predicted amino-acid sequence is presented. The remaining nucleotide sequence of the 3′ part of vWF mRNA has been reported before (Sadler *et al.*, 1985). In general, the overlapping part of our nucleotide sequence of vWF cDNA with that of Sadler *et al.* (1985) reveals no differences. However, the first 12 nucleotides (corresponding with position 2217–2229) at the 5′ terminus of vWF cDNA on phage lambda-HvWF1, con-

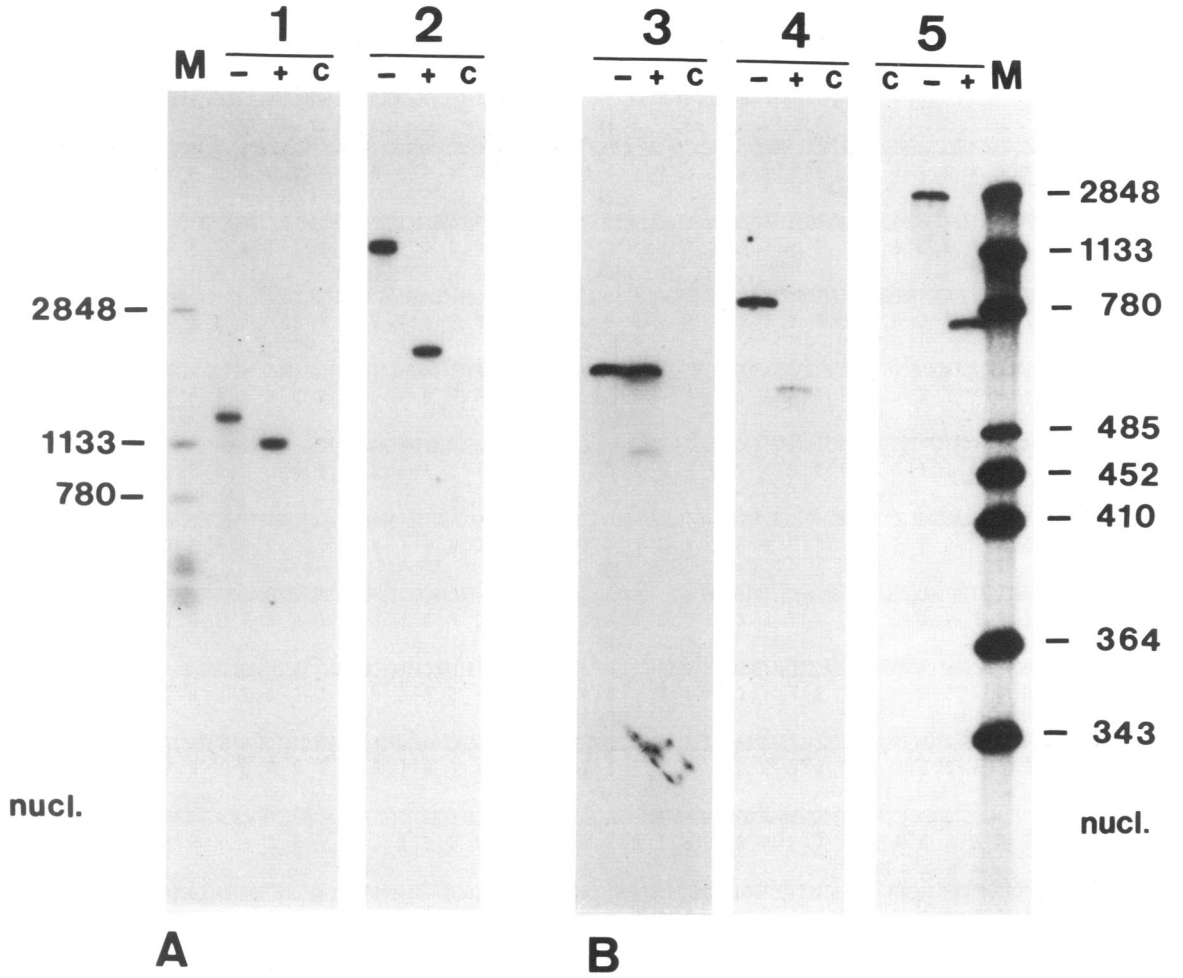


Fig. 2. S1 nuclease protection analysis. Endothelial poly(A)$^+$ RNA was hybridized with ^{32}P-labeled probes, containing vWF cDNA sequences. The construction of the different probes and the conditions used are described in Materials and methods. The vWF cDNA segments, present in the probes, are shown in Figure 1A. **Panel A** shows the results after electrophoresis of the samples in a 1% alkaline agarose gel. **Panel B** gives the results after electrophoresis in a 6% polyacrylamide−8 M urea gel. **Lanes 1**, hybridization with probe III, containing the 1144-bp vWF cDNA fragment III. **Lanes 2**, hybridization with probe V, containing the 2396-bp vWF cDNA fragment V. **Lanes 3**, hybridization with probe I which is equivalent to the 587-bp vWF cDNA fragment I. **Lanes 4**, hybridization with probe IV, containing the 576-bp vWF cDNA fragment IV. **Lanes 5**, hybridization with probe II, containing the 734-bp vWF cDNA fragment II. Symbols: −, incubation of hybridized components in the absence of S1 nuclease; +, incubation of the hybridized components in the presence of S1 nuclease; c, incubation of the samples with S1 nuclease after hybridization in the absence of endothelial poly(A)$^+$ RNA; M, single-stranded DNA-length markers.

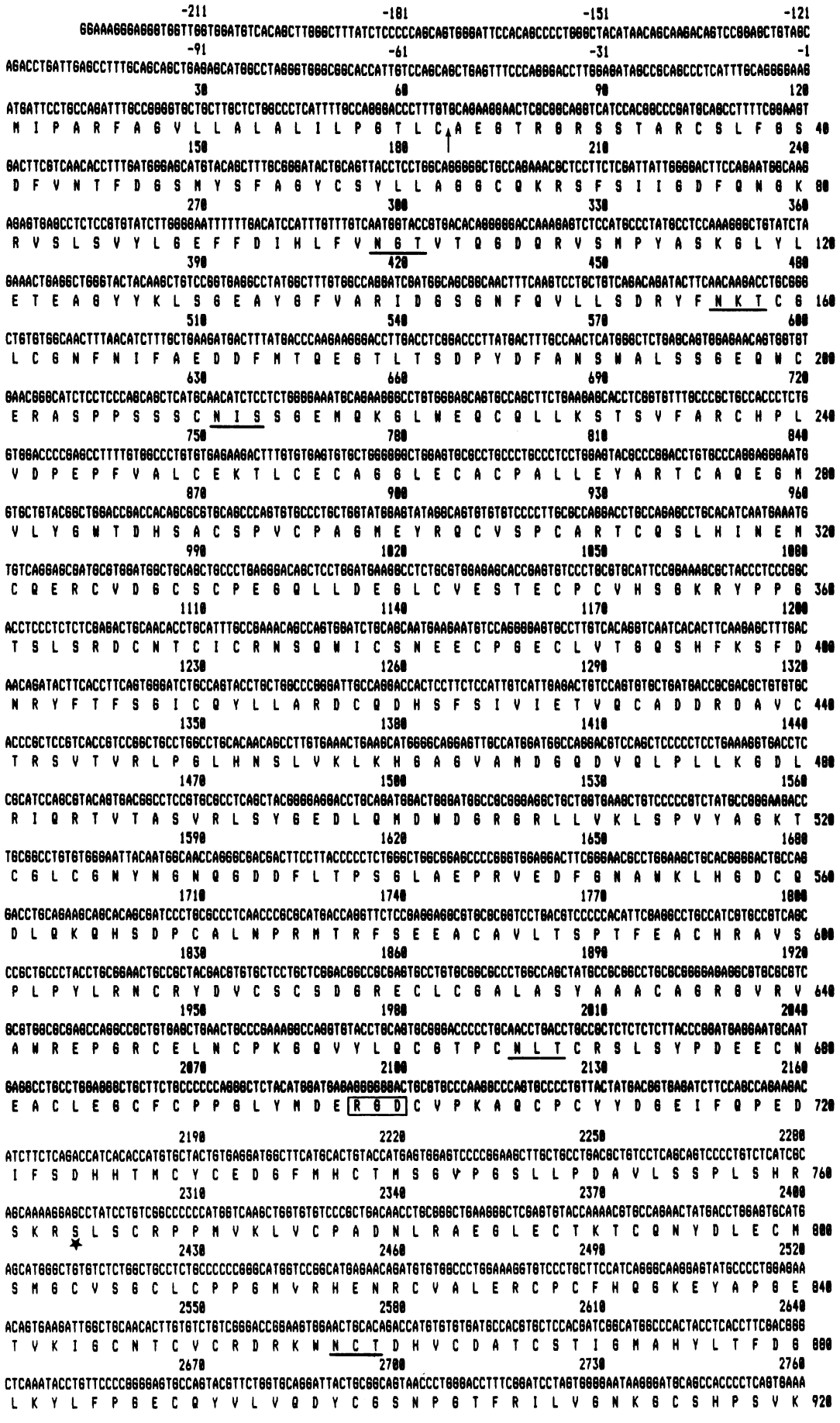
structed by those authors, are completely divergent from our sequence which has been established on three independent, overlapping cDNA inserts. Another discrepancy was observed at position 2309. Our analysis reveals a C residue, whereas Sadler *et al.*, (1985) report an A residue, resulting in, respectively, a proline and a histidine at that particular position. The proline residue has been established independently by automated amino-acid sequence analysis of the mature vWF (Hessel *et al.*, 1984). This difference might be due to a polymorphism in the vWF gene.

The total length spanned by the adjacent vWF cDNA sequences is 8804 bp, excluding the poly(A) tail of vWF mRNA. The translation initiation site was assigned to the ATG codon indicated (position 1−4), being the first initiator codon downstream of the TAG nonsense codon (position −79 to −82), which is the start of an 'open' translation reading frame of 8439 nt (deduced from our sequence data and those of Sadler *et al.*, 1985). This assignment is supported by the observation that the predicted 22 N-terminal amino-acid residues display the characteristic features of a signal peptide. The cleavage site for a signal peptidase with the highest probability is located between the cysteine and alanine residues at position 22 and 23 (Von Heijne, 1983). The proposed translation initiation codon is preceded by an untranslated region of at least 229 nt.

A continuous vWF cDNA coding sequence of 8439 bp potentially programs the synthesis of a polypeptide of 2813 amino-acid residues, with a calculated mol. wt of 309 000. To our knowledge, this represents the longest coding sequence determined to date. Furthermore, it has been reported that mature vWF protein is a glycoprotein, containing approximately 15% carbohydrate residues (Sodetz *et al.*, 1979). If it is assumed that the carbohydrate moieties also contribute about 15% by weight to the calculated mol. wt of pro-vWF, then the mol. wt of pro-vWF will amount to approximately 350 000.

Peculiarities of the amino-acid sequence of pro-vWF

A comparison of the predicted amino-acid sequence (Figure 3) with the established N-terminal amino-acid sequence of mature vWF protein (Hessel *et al.*, 1984) confirms our earlier assumption (Verweij *et al.*, 1985) and that of others (Lynch *et al.*, 1985; Ginsburg *et al.*, 1985) that the vWF precursor protein is con-



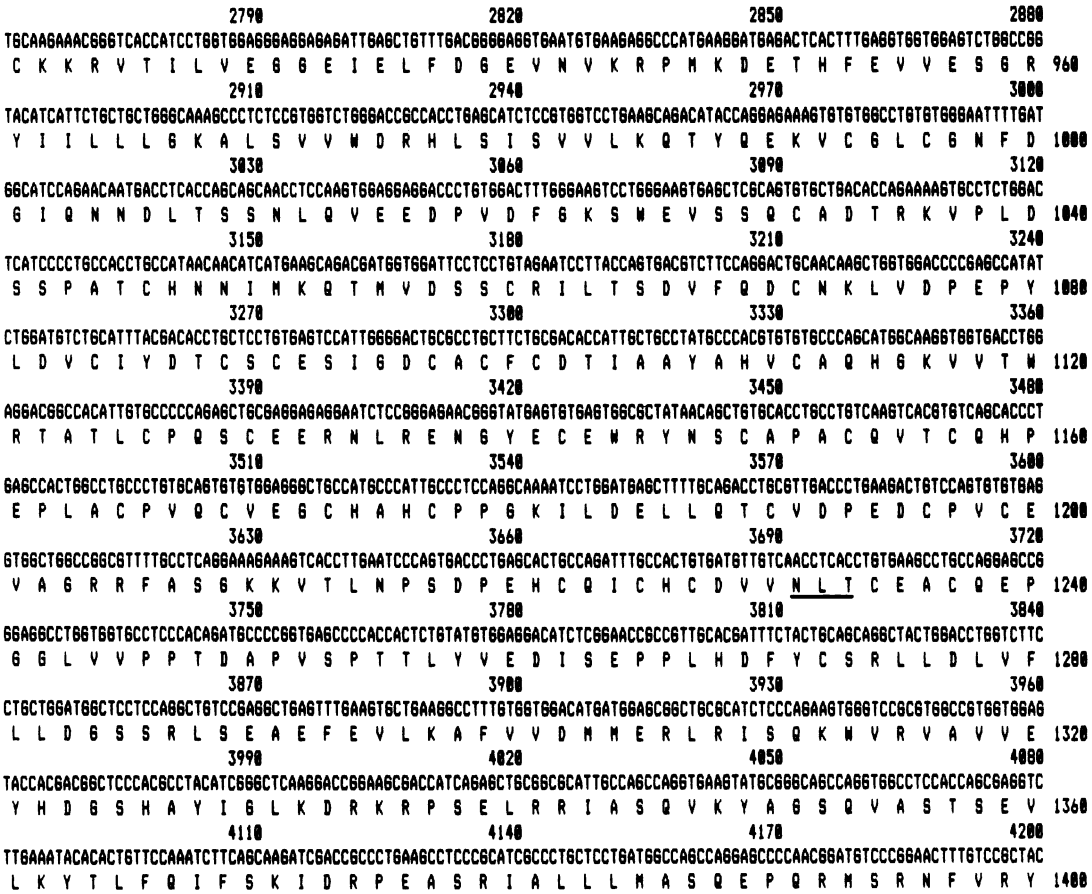


Fig. 3. Nucleotide sequence of 4429 bp of vWF cDNA, derived from the 5′ terminus of vWF mRNA. The numbering starts at the putative ATG translation−initiation codon. The predicted amino-acid sequence is shown beneath the nucleotide sequence and is separately numbered, again starting at the putative methionine translation−initiation codon. Potential N-linked glycosylation sites are underlined. The tripeptide arginine-glycine-aspartic acid is boxed. The putative signal peptide of 22 amino-acid residues is generated by a signal peptidase. The predicted cleavage site (between residues 22 and 23) is shown with an arrow. The pro-sequence with a length of 741 amino acids (residues 23−764) will be cleaved to generate the mature vWF subunit. The N-terminal amino-acid residue of the mature vWF subunit (position 764) is indicated with an asterix.

siderably larger than the reported 240 000−260 000 glycoprotein. Alignment of the predicted amino-acid sequence with the N-terminal sequence of the mature (225 000) vWF protein shows that the nucleotide sequence which codes for mature vWF protein initiates at position 2290. This conclusion implies that the vWF precursor protein is 763 amino-acid residues larger than the mature protein. Obviously, this pro-sequence (calculated mol. wt 81 000) is removed by protein processing to yield the mature vWF glycoprotein with a mol. wt of about 225 000.

A homology matrix comparison of the amino-acid sequence of the vWF precursor protein reveals a quadruplication of a domain (D1, D2, D3, D4) with a length of about 350 amino-acid residues. Part of this domain (D′, about 96 amino acids) appears to be present at the N-terminus of mature vWF. Figure 4A shows the alignment of these repetitive domains. A salient feature of the pro-sequence is that it largely consists of a duplication of the D domain (D1, D2) (see Figure 4B). The repeats are extremely rich in cysteine residues (D1, 9.1%; D2, 8.9%; D3, 9.6%; D4, 8.8%; and D′, 17%) and exhibit a significant conservation of the position of these residues, indicative for a structural similarity of the repeats. Interestingly, the pro-sequence comprises an arginine-glycine-aspartic acid sequence ('RGD' sequence) at position 698−701. It has been shown that a tetrapeptide with the indicated amino-acid sequence can compete with proteins, harboring a similar sequence, which are involved in cell attachment (Pierschbacher and Ruoslahti, 1984). It has been noticed that another RGD sequence is present within the C-terminal part of the mature vWF protein (Sadler *et al.*, 1985).

In-vitro translation of vWF mRNA

A full-length vWF cDNA was assembled to demonstrate its coding capacity for an unglycosylated precursor protein with a mol. wt of about 300 000. Full-length vWF cDNA was inserted into plasmid pSP65. This plasmid contains the *Salmonella typhimurium* bacteriophage SP6 promoter which allows *in-vitro* 'run off' transcription of cloned DNA sequences, specifically directed by SP6 RNA polymerase (Melton *et al.*, 1984). Such mRNA preparations can be efficiently translated, using a reticulocyte lysate.

Initially, plasmid pSP6330vWF was constructed (see Figure 1B), harboring a continuous, 6331-bp vWF cDNA sequence. This plasmid contains the entire coding sequence for mature vWF and, in addition, a sequence coding for 18 amino-acid residues from the C-terminal part of the pro-sequence. Initiation of protein synthesis, directed by RNA transcribed from pSP6330vWF DNA, should occur at the methionine codon eight amino acids downstream of the N-terminus of mature vWF. Translation of the *in-vitro* synthesized vWF mRNA will then yield an unglycosylated polypeptide with a calculated mol. wt of 225 000. Subsequently, pSP8800vWF was constructed harbouring the complete coding

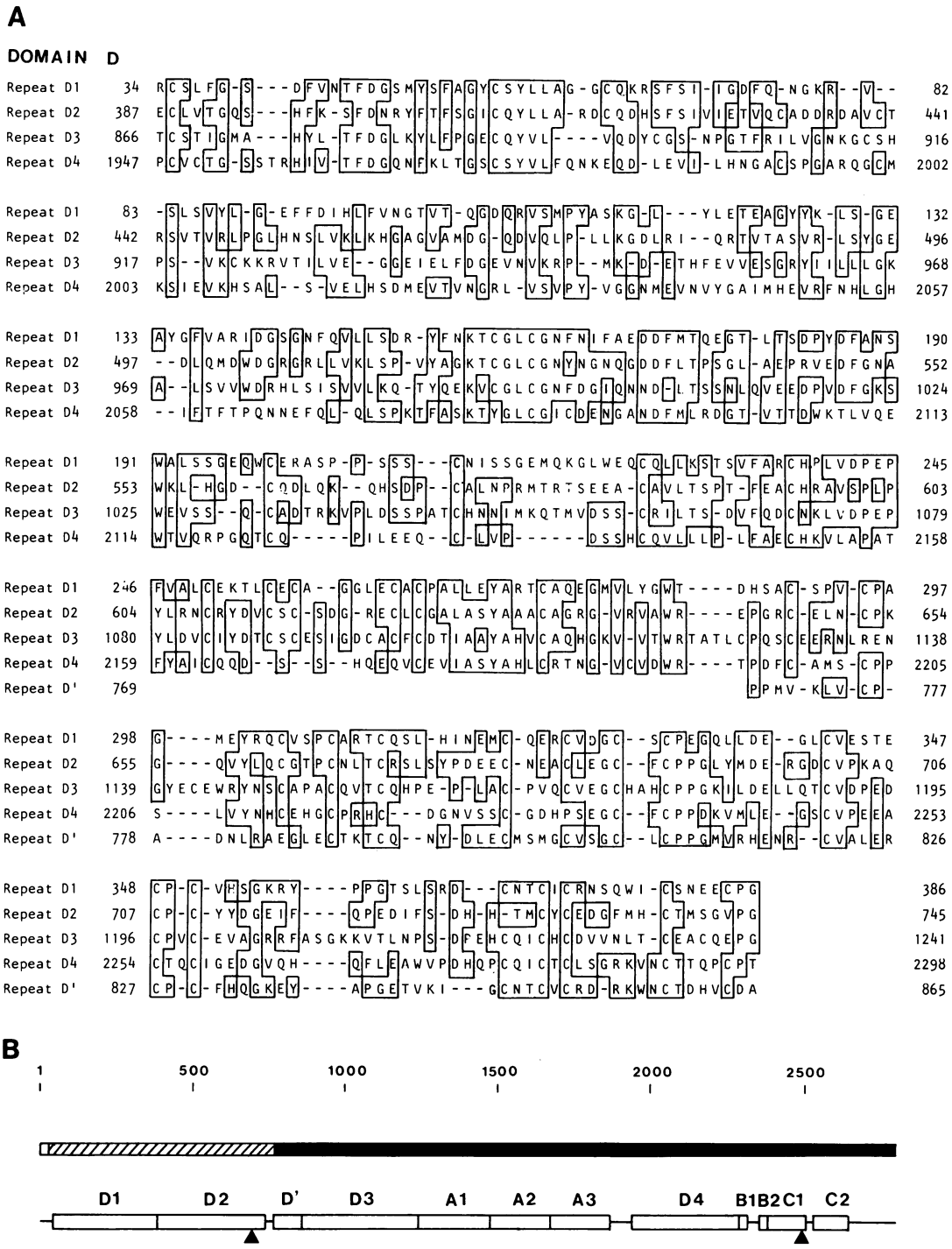


Fig. 4. Internal homology within the precursor for vWF. (**A**) Alignment of the amino-acid sequences of the four repeated domains D1, D2, D3, D4 and the partly duplicated domain D'. The one-letter notation is used and the amino acids are numbered as indicated in Figure 3. Residues which are identical among the four or five repeats are boxed. (**B**) Schematic representation of internal homologous regions within pro-vWF. The upper line in this diagram represents the vWF precursor protein (open area, signal peptide; hatched area, prosequence; dark area, mature vWF). Beneath this line are indicated: the triplicated domain A (A1, A2 and A3) and two duplicated domains B (B1 and B2) and C (C1 and C2), as reported by Sadler *et al.* (1985), the quadruplicated domain D (D1, D2, D3 and D4) and the partly duplicated domain D'. The numerical position of these repeats are listed: A1 (residues 1242–1480), A2 (1480–1673), A3 (1673–1875), B1 (2296–2331), B2 (2375–2400), C1 (2400–2516), C2 (2544–2663), D1 (34–387), D2 (387–746), D3 (866–1242), D4 (1947–2299) and D' (769–866). The position of the RGD tripeptides is shown with a triangle.

sequence for pre-vWF (see Figure 1B). This plasmid will encode a protein with a calculated mol. wt of 309 000. The plasmids pSP6330vWF and pSP8800vWF were linearized with, respectively, *Eco*RI and *Sal*I and trancribed *in vitro*. The results of the *in-vitro* translation of the various vWF mRNAs are given in Figure 5. The polypeptides encoded by pSP6330vWF DNA display a mol. wt of up to about 200 000. The discrepancy of this mol. wt with the calculated mol. wt is probably due to inaccuracy in the mol. wt estimation of large proteins in these gels. The complete coding sequence of pSP8800vWF DNA is translated into a polypeptide with a mol. wt substantially larger than 200 000. To achieve a more accurate mol. wt estimation

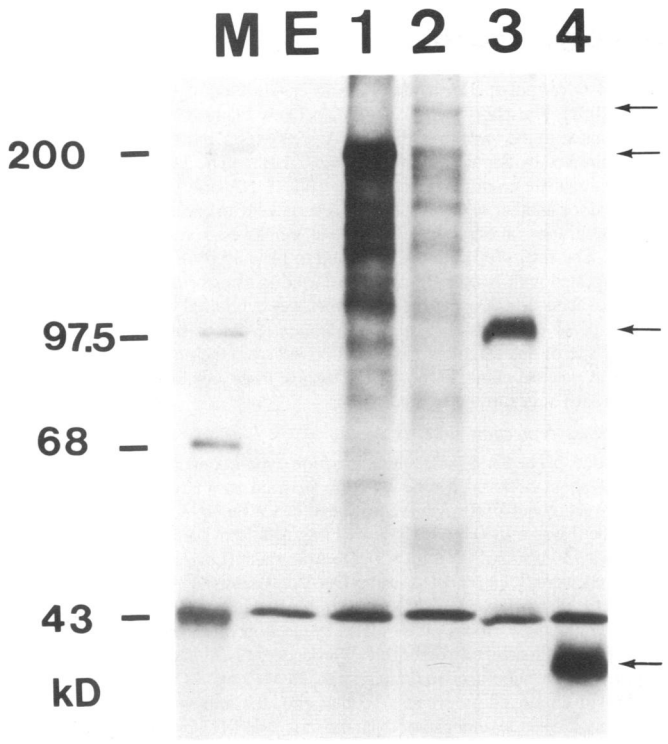


Fig. 5. *In-vitro* translation of vWF mRNA. Capped vWF mRNA was prepared *in vitro*, using 'run off' transcription with SP6 RNA polymerase, as described in Materials and methods. The RNA preparations were added to a reticulocyte lysate translation system, containing [^{32}S]methionine, and polypeptides were synthesized for 90 min. The polypeptides were fractionated on an 8% SDS–polyacrylamide gel and then subjected to fluorography. M, mol. wt marker proteins; E, endogenously synthesized polypeptides (without added RNA). **Lane 1**, polypeptides encoded by vWF mRNA transcribed from pSP6330vWF DNA, digested with *Eco*RI. **Lane 2**, polypeptides encoded by vWF mRNA transcribed from pSP8800vWF DNA, digested with *Sal*I. **Lane 3**, polypeptides encoded by vWF mRNA transcribed from pSP8800vWF DNA, digested with *Bam*HI. **Lane 4**, polypeptides encoded by vWF mRNA transcribed from pSP8800vWF DNA, digested with *Xho*I.

for this extraordinarily long polypeptide, we produced partial, overlapping polypeptides derived from selected portions of full-length vWF cDNA. To that end, pSP8800vWF DNA was digested with *Bam*HI and the transcript (~2855 nt long) was translated. It should be noted that 405 nt at the 3′ end of this transcript constitute the 5′ terminus of the transcript generated from pSP6330vWF cleaved with *Eco*RI. Hence, an enumeration of the mol. wts of the polypeptides, mentioned above, should result in a mol. wt of about 309 000, after subtracting the common protein region. The protein, derived from the *Bam*HI-digested template, has an estimated mol. wt of about 100 000, whereas, as shown before, template pSP6330vWF cleaved with *Eco*RI yields a product of about 200 000. Addition of these mol. wts and subtracting the common region (15 000) results in an estimated mol. wt of 285 000 which is in reasonable agreement with the calculated mol. wt of 309 000. Furthermore, translation of transcripts (~1320 nt), derived from pSP8800vWF DNA digested with *Xho*I, reveals a polypeptide with a mol. wt of 39 000. This result is in agreement with the assignment of the translation initiation site at position 1–4.

From these data, we conclude that we have constructed a full-length vWF cDNA with a coding sequence of 8439 bp which programs the synthesis of a precursor vWF protein consisting of 2813 amino-acid residues.

Discussion

In this paper, we report on the construction of a plasmid containing full-length vWF cDNA. Nucleotide sequence analysis (this paper; Sadler *et al.*, 1985) revealed that the length of the assembled vWF cDNA, excluding cDNA derived from the poly(A) tail, amounts to 8804 bp. This result is in agreement with the length of vWF mRNA (about 9000 nt) as determined by Northern blot analysis of endothelial poly(A)$^+$ RNA. The entire coding sequence for the precursor vWF protein is 8439 bp, corresponding to an unglycosylated polypeptide with a mol. wt of about 309 000. The translation initiation site for this protein could be assigned to the ATG codon at position 1–4 (see Figure 3). This assignment is in accord with the results of *in-vitro* translation experiments with parts of full-length vWF mRNA, generated with the SP6 transcription system. The glycosylated protein, even after removal of a signal peptide, will be considerably larger than 300 000. The mol. wt attributed to the precursor glycoprotein for mature vWF has been reported to be 240 000–260 000 (Wagner and Marder, 1983; Lynch *et al.*, 1983). The discrepancy of the mol. wt that we assign to the precursor protein may be due to inaccuracy of mol. wt estimations by SDS–polyacrylamide gel electrophoresis, inherent to large (glyco)proteins.

The sequence-encoding mature vWF initiates at 2290 bp downstream of the translation initiation codon, as inferred from an alignment of the established N-terminal amino-acid sequence of mature vWF (Hessel *et al.*, 1984) with the predicted amino-acid sequence (Figure 3). This 2289-bp long pre-pro-sequence was shown to be able to encode a polypeptide with a calculated mol. wt of 83 000. Consequently, the pro-vWF protein will be processed by a protease to yield mature vWF. In this respect, it is relevant to note that mature vWF is closely associated with the so-called von Willebrand Antigen II (vW AgII) (Montgomery and Zimmerman, 1978). Several arguments can be advanced which indicate that the pro-sequence of the precursor vWF protein is identical to vW AgII.

(i) The mol. wt of the unglycosylated pro-sequence (81 000) fits with the reported mol. wt of the vW AgII glycoprotein of 98 000 (McCarroll *et al.*, 1985).

(ii) Both vWF and vW AgII are synthesized by cultured endothelial cells and these proteins are simultaneously released *in vivo* upon stimulation with 1-desamino-8-D-arginine vasopressin (DDAVP) (McCarroll *et al.*, 1984a).

(iii) Using immunofluorescence techniques, both proteins are located in the perinuclear region and in the Weibel–Palade bodies (McCarroll *et al.*, 1985).

(iv) Both vWF and vW AgII are present in platelets and released together after platelet activation (Scott and Montgomery, 1981).

(v) The levels of vWF and vW AgII protein are linearly associated in plasma and both proteins are deficient in plasma and platelets of a patient with severe von Willebrand's disease (McCarroll *et al.*, 1984b).

(vi) A complex between vWF and vW AgII can be detected in endothelial cell lysates in the presence of a serine protease inhibitor (PMSF) and not in the absence of the inhibitor (McCarroll *et al.*, 1985).

Studies are in progress to identify the subcellular organelle associated with the proteolytic cleavage between the arginine and the serine residues at positions 763 and 764.

We have compared the predicted amino-acid sequence of pro-vWF with that of other proteins, contained within the NIH Protein Sequence Data Bank, for (partial) homologous amino-acid sequences. This comparison did not reveal any major

similarity of pro-vWF with other proteins. The predicted amino-acid sequence of the pro-sequence displays a remarkable structure. It is composed of a duplicated segment of about 350 amino-acid residues long. These two segments share 37% amino-acid homology. Furthermore, they exhibit a considerable conservation of similarly located cysteine residues, indicating that structural features have been maintained within these direct repeats. Two copies of this repeat within the pro-sequence are also present within mature vWF, whereas part of this repeat is present at the N-terminus of mature vWF. Internal homologous regions have also been reported by Sadler *et al.* (1985), two of which have been duplicated, while one is present in triplicate form. These repeated sequences span a length of about 1070 amino-acid residues within the mature vWF protein (see Figure 4B). The repeated structures that we have found are independent of the ones reported by Sadler *et al.*. From these data, we conclude that about 90% of the precursor vWF protein is constituted of repetitive regions, indicating that the precursor vWF gene has evolved from a series of duplicative events of at least four different regions.

The presence of a 'RGD(C)' amino-acid sequence within the pro-sequence may be indicative for a possible function of this protein. It has been shown that an RGD-containing region, on proteins such as fibronectin and vitronectin, carries out a crucial role in the interaction with receptors on a cell surface (Pierschbacher and Ruoslahti, 1984; Pytela *et al.*, 1985). Those interactions are inhibited by RGD-containing peptides. Interaction of mature vWF with activated platelets is also inhibited by RGD-containing peptides, suggesting that this region on vWF is involved in platelet binding (Ginsberg *et al.*, 1985; Haverstick *et al.*, 1985). Based on the presence of an RGD sequence both in mature vWF and the pro-sequence, which may be equivalent to vW AgII, on a striking homology and on a structural conservation between these two proteins, we propose that the pro-sequence might have similar interaction(s) as the mature vWF protein with particular components, such as cell-surface receptors.

Materials and methods

cDNA cloning

Total RNA was purified from cultured endothelial cells, derived from veins of human umbilical cords (Verweij *et al.*, 1985). Primer-directed cDNA was synthesized from poly(A)$^+$ RNA, essentially according to a protocol described (Gubler and Hoffman, 1983; Toole *et al.*, 1985). The cDNA synthesis was arrested by adding EDTA and SDS till a final concentration of, respectively, 20 mM and 0.1%. The cDNA preparations were extracted with phenol–chloroform, then precipitated with ethanol and purified by chromatography on Sephadex G-50. In the case of primer-directed cDNA synthesis with primer A (5′ CACAGGCCACACGTGGGAGC 3′), complementary to nucleotides 6901–6921, the cDNA preparation was digested with *Bgl*II (positions 6836 and 2141) and *Kpn*I (position 4748). Subsequently, the digested cDNA was size-fractionated by chromatography on a Sepharose CL-4B column. Fractions containing cDNA larger than about 600 bp were ligated to plasmid pMBL11, digested with *Bgl*II and *Kpn*I. Plasmid pMBL11 is a derivative of pBR322, containing the promoter and the tryptophan synthetase-A gene of *Escherichia coli*. This plasmid includes unique restriction sites for the enzymes *Kpn*I, *Bgl*II, *Eco*RI and *Xho*I (T.Kos, Medical Biological Lab. TNO, Rijswijk, the Netherlands; personal communication). A cDNA library of about 15 000 independent colonies was established, using strain *E. coli* DH1 as a host, which was screened with two oligonucleotide probes (B and C). Probe B (5′ GAGGCAGGATTTCCGGTGAC 3′), complementary to nucleotides 4819–4839, was employed for the isolation of the plasmid pvWF2084, harboring a 2084-bp *Bgl*II–*Kpn*I vWF cDNA fragment, whereas probe C (5′ CAGGGACACCTTTCCAGGGC 3′), complementary to 2467–2487, was used for the detection of plasmid pvWF2600, harboring an ~2600 bp *Kpn*I–*Bgl*II vWF cDNA fragment. Using probe C for primer-directed synthesis, we divided the resulting cDNA preparation into two parts. One part was C-tailed and annealed to G-tailed plasmid pUC9 as described before (Verweij *et al.*, 1985) and used to transform *E. coli* DH1. Six thousand independent colonies were hybridized with a 'nick-translated' 576-bp *Bgl*II–*Bam*HI vWF cDNA fragment of pvWF2600 DNA. A positive clone, harboring a plasmid with the longest insert (about 1800 bp, designated pvWF1800) was chosen for further study. The other part of the primer C-directed cDNA preparation was treated with *Eco*RI methylase and subsequently with T4-DNA polymerase and dNTPs to ensure blunt-ended termini (Maniatis *et al.*, 1982). Phosphorylated *Eco*RI linkers (New England Biolabs, Beverly, MA) were ligated to the termini of the cDNA preparation and unreacted components were removed by Sephadex G-50 chromatography. The *Xho*I site, located about 350 bp downstream of the 5′ end of the vWF cDNA insert of pvWF1800 DNA, was used for another selection. After digestion with an excess of *Eco*RI and *Xho*I, chromatography on Sepharose CL-4B was employed to remove digested *Eco*RI linkers. The final preparation was ligated to plasmid pMBL11 DNA which had been digested with *Eco*RI plus *Xho*I and used to transform *E. coli* DH1. A collection of about 10 000 independent colonies was hybridized with a 'nick-translated' 350-bp *Xho*I–*Hin*dIII vWF cDNA fragment from plasmid pvWF1800. The *Hin*dIII site of this fragment has been derived from the polylinker of the vector pUC9. A positive clone, harboring the longest insert (about 1330 bp, designated pvWF1330) was further studied.

S1 nuclease protection analysis

We used as probe for S1 nuclease protection experiments an *Xho*I–*Eco*RI fragment of about 5300 bp (probe V) from plasmid pvWF2600 which contains a 2396-bp segment (*Xho*I–*Kpn*I) constituted of vWF cDNA (Fragment V, Figure 1). Probe II was a 4800-bp *Xba*I–*Eco*RI fragment from plasmid pvWF1330 which harbors a 734-bp *Xba*I–*Xho*I vWF cDNA segment (Fragment II, Figure 1). The fragments were 3′ end-labeled, using DNA polymerase I (large fragment) (New England Biolabs, Beverly, MA) to fill in recessed ends (Maniatis *et al.*, 1982). Subsequently, these probes were isolated by electrophoresis on a 0.7% low-melting agarose gel and purified as described (Wieslander, 1979). Three other vWF cDNA fragments were subcloned in double-stranded M13mp18 (Yanisch-Perron *et al.*, 1985) and employed as probes. To that end, the anti-sense DNA strand was uniformly labeled by elongation from the universal M13-primer with DNA polymerase I (large fragment). The subcloned fragments were a 1144-bp *Xho*I–*Hin*dIII fragment of plasmid pvWF1800 (Fragment III, Figure 1), a 585-bp *Xba*I–*Eco*RI fragment of plasmid pvWF1330 (Fragment I, Figure 1) and a 575-bp *Bam*HI–*Bgl*II fragment of plasmid pvWF2600 (Fragment IV, Figure 1). After DNA synthesis, initiated at the M13 primer, double-stranded DNA was digested with both *Hin*dIII and *Pvu*II for fragment III (to yield probe III), with both *Xba*I and *Eco*RI for fragment I (to yield probe I) and with both *Bam*HI and *Pvu*II for fragment IV (to yield probe IV). The rationale for the construction of probes II, III, IV and V is that they contain a segment of vector DNA noncomplementary with endothelial RNA. For example, probes III and IV harbor about 200 bp, derived from M13mp18. These probes were subjected to electrophoresis on a 5% polyacrylamide – 8 M urea gel and the fragments of interest were isolated (Maxam and Gilbert, 1977).

S1 nuclease protection experiments were carried out essentially as described (Berk and Sharp, 1977). One microgram of human endothelial poly(A) RNA was added to 10 000–100 000 c.p.m. of radiolabeled probe, heated for 10 min at 80°C, followed by an incubation overnight at 60°C for probes I, II, III and V and at 57°C for probe IV. Digestion with 200 U of S1 nuclease (Bethesda Research Laboratory, Gaithersburg, MD) per ml was carried out for 20 min at 45°C. Undigested DNA was precipitated with ethanol and the pellets were dissolved in the appropriate loading buffer for electrophoresis on a 1% alkaline agarose gel or on a 6% polyacrylamide sequencing gel (Maniatis *et al.*, 1982). The first procedure was employed for probes I and III, whereas the second one was applied for probes II, IV and V.

Assembly of full-length vWF cDNA

For the construction of plasmid pSP6330vWF, harboring a continuous vWF cDNA segment of about 6331 bp extending from the *Hin*dIII site (position 2235) till the *Sac*I site within the 3′ untranslated region (position 8562), the following vWF cDNA fragments were isolated: the 2517-bp *Hin*dIII–*Kpn*I fragment (position 2236–4753) from pvWF2600 DNA; the 2084-bp *Kpn*I–*Bgl*II fragment (position 4753–6837) from pvWF2084 DNA, and the 1730-bp *Bgl*II–*Sac*I fragment (position 6837–8567) from pvWF2280 DNA. These three vWF cDNA fragments were ligated simultaneously into the vector pSP64 (Melton *et al.*, 1984), digested with both *Hin*dIII and *Sac*I. About half of the resulting transformants contained a plasmid (denoted pSP6330vWF) with the desired vWF cDNA insert of 6331 bp, as verified by restriction-enzyme analysis.

For the construction of plasmid pSP8800vWF, harboring full-length vWF cDNA, the following fragments were isolated: the 6333-bp *Hin*dIII–*Eco*RI insert of plasmid pSP6330vWF (the *Eco*RI site is derived from the polylinker present on the vector); the 1144-bp *Xho*I–*Hin*dIII fragment (position 1092–2236) from pvWF1800, and the 1327-bp *Eco*RI–*Xho*I fragment (position −236 to 1092) from pvWF1330. A five-fold molar excess of each of these three fragments was again ligated simultaneously with vector pSP65 DNA, cleaved with *Eco*RI and treated with calf intestine alkaline phosphatase (Boehringer, Mannheim, FRG). About 30% of the resulting colonies harbored a plasmid with the desired. full-

length vWF cDNA insert of 8794 bp in the correct orientation, as verified by restriction-enzyme analysis and nucleotide-sequence analysis.

In-vitro transcription and translation

In-vitro transcription of linear SP6-based DNA templates wtih SP6 RNA polymerase (New England Nuclear, Dreieich, FRG) was performed in the presence of 0.1 mM UTP, CTP and ATP, 0.05 mM GTP and 2 mM of $m^7G(5')ppp(5')G$ (Pharmacia, Uppsala, Sweden) to provide mRNA preparations with a capped terminus (Melton *et al.*, 1984). *In-vitro* translation of such 5′ capped mRNAs was done in a rabbit reticulocyte lysate system (New England Nuclear, Dreieich, FRG), according to the manufacturer's specifications. Analysis of the *in-vitro* translation products was performed by electrophoresis on an 8% SDS–polyacrylamide gel as described (Laemmli, 1970).

Acknowledgements

We thank Drs J.A.van Mourik and M.N.Hamers for critical reading of the manuscript. This study was supported by The Netherlands Organization for the Advancement of Pure Research (ZWO) (grant No. –13/90–91).

References

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721–732.
Fowler,W.E., Fretto,L.J., Hamilton,K.K., Erickson,H.P. and McKee,P.A. (1985) *J. Clin. Invest.*, **76**, 1491–1500.
Fujimoto,T. and Hawiger,J. (1982) *Nature*, **297**, 154–156.
Ginsberg,M., Pierschbacher,M.D., Ruoslahti,E., Marguerie,G. and Plow,E. (1985) *J. Biol. Chem.*, **260**, 3931–3936.
Ginsburg,D., Handin,R.I., Bonthron,D.T., Donlon,T.A., Bruns,G.A.P., Latt,S.A. and Orkin,S.H. (1985) *Science*, **228**, 1401–1406.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Haverstick,D.M., Cowan,J.F., Yamada,K.M. and Santoro,S.A. (1985) *Blood*, **66**, 946–952.
Hessel,B., Jornvall,H., Thorell,L., Soderman,S., Larsson,U., Egberg,N., Blomback,B. and Holmgren,A. (1984) *Thromb. Res.*, **35**, 637–651.
Houdijk,W.P.M., Sakariassen,K.S., Nievelstein,P.F. and Sixma,J.J. (1985) *J. Clin. Invest.*, **75**, 531–540.
Hoyer,L.W. and Shainoff,J.R. (1980) *Blood*, **55**, 1056–1059.
Jaffe,E.A., Hoyer,L.W. and Nachman,R.L. (1973) *J. Clin. Invest.*, **52**, 2757–2764.
Jenkins,C.S.P., Phillips,D.R., Clemetson,K.J., Meyer,D., Larrieu,M.-J. and Luscher,E.F. (1976) *J. Clin. Invest.*, **57**, 112–124.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Lynch,D.C., Zimmerman,T.S., Kirby,E.P. and Livingston,D.M. (1983) *J. Biol. Chem.*, **258**, 12757–12760.
Lynch,D.C., Zimmerman,T.S., Collins,C.J., Brown,M., Morin,M.J., Ling,E.H. and Livingston,D.M. (1985) *Cell*, **41**, 49–56.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–564.
McCarroll,D.R., Ruggeri,Z.M. and Montgomery,R.R. (1984a) *Blood*, **63**, 532–535.
McCarroll,D.R., Ruggeri,Z.M. and Montgomery,R.R. (1984b) *J. Lab. Clin. Med.*, **103**, 704–711.
McCarroll,D.R., Levin,E.G. and Montgomery,R.R. (1985) *J. Clin. Invest.*, **75**, 1089–1095.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Montgomery,R.R. and Zimmerman,T.S. (1978) *J. Clin. Invest.*, **61**, 1498–1507.
Nachman,R.L., Levine,R. and Jaffe,E.A. (1977) *J. Clin. Invest.*, **60**, 914–921.
Perret,B.A., Furlan,M. and Beck,E.A. (1979) *Biochim. Biophys. Acta*, **578**, 164–174.
Pierschbacher,M.D. and Ruoslahti,E. (1984) *Nature*, **309**, 30–33.
Pytela,R., Pierschbacher,M.D. and Ruoslahti,E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5766–5770.
Ruggeri,Z.M. and Zimmerman,T.S. (1980) *J. Clin. Invest.*, **65**, 1318–1325.
Sadler,J.E., Shelton-Inloes,B.B., Sorace,J.M., Harlan,J.M., Titani,K. and Davie,E.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6394–6398.
Sakariassen,K.S., Bolhuis,P.A. and Sixma,J.J. (1979) *Nature*, **279**, 636–638.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scott,J.P. and Montgomery,R.R. (1981) *Blood*, **58**, 1075–1080.
Sodetz,J.M., Paulson,J.C. and McKee,P.A. (1979) *J. Biol. Chem.*, **254**, 10754–10760.
Sporn,L.A., Chavin,S.I., Marder,V.J. and Wagner,D.D. (1985) *J. Clin. Invest.*, **76**, 1102–1106.
Toole,J.J., Knopf,J.L., Wozney,J.M., Sultzman,L.A., Buecker,J.L., Pittman,D.D., Kaufman,R.J., Brown,E., Shoemaker,C., Orr,E.C., Amphlett, G.W.., Foster,W.B., Coe,M.-L., Knutson,G.J., Fass,D.N. and Hewick,R.M. (1984) *Nature*, **312**, 342–347.
Van Mourik,J.A. and Bolhuis,P.A. (1978) *Thromb. Res.*, **13**, 15–24.
Verweij,C.L., De Vries,C.J.M., Distel,B., Van Zonneveld,A.-J., Geurts van Kessel,A., Van Mourik,J.A. and Pannekoek,H. (1985) *Nucleic Acids Res.*, **13**, 4699–4717.
Von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17–21.
Wagner,D.D. and Marder,V.J. (1983) *J. Biol. Chem.*, **258**, 2065–2067.
Wagner,D.D. and Marder,V.J. (1984) *J. Cell. Biol.*, **99**, 2123–2130.
Wagner,D.D., Olmsted,J.B. and Marder,V.J. (1982) *J. Cell. Biol.*, **95**, 355–360.
Wieslander,L. (1979) *Anal. Biochem.*, **98**, 305–309.
Yanisch-Peron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on 7 April 1986; revised on 4 June 1986

Note added in proof

During submission of this manuscript we learned that our proposal concerning the identity of the pro-sequence and von Willebrand Antigen II (vW AgII), was substantiated by data from other investigators (Fay *et al.*, 1986, *Science*, **232**, 995–998).