Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4393–4397, July 1987
Biochemistry

Molecular cloning of the human gene for von Willebrand factor and identification of the transcription initiation site

(human genomic clones/promoter/cross-species homology)

CAROLYN J. COLLINS*†‡, JERALD P. UNDERDAHL*†, RICHARD B. LEVENE*†, CHRISTINA P. RAVERA§, MELINDA J. MORIN*†, MARKAR J. DOMBALAGIAN§, GEORGE RICCA§, DAVID M. LIVINGSTON*†, AND DENNIS C. LYNCH*†

*Dana–Farber Cancer Institute and †Department of Medicine, Harvard Medical School, Boston, MA 02115; and §Meloy Laboratories, Inc., Springfield, VA 22151

Communicated by Gerald N. Wogan, March 13, 1987

ABSTRACT **A series of overlapping cosmid genomic clones have been isolated that contain the entire coding unit of the human gene for von Willebrand factor (vWf), a major component of the hemostatic system. The cloned segments span ≈175 kilobases of human DNA sequence, and hybridization analysis suggests that the vWf coding unit is ≈150 kilobases in length. Within one of these clones, the vWf transcription initiation site has been mapped and a portion of the vWf promoter region has been sequenced, revealing a typical "TATA box," a downstream "CCAAT box," and a perfect downstream repeat of the 8 base pairs containing the transcription start site. Sequencing of a segment of another genomic clone has revealed the vWf translation termination codon. Where tested, comparative restriction analysis of cloned and chromosomal DNA segments strongly suggests that no major alterations occurred during cloning and that there is only one complete copy of the vWf gene in the human haploid genome. Similar analyses of DNA from vWf-producing endothelial cells and nonexpressing leukocytes suggest that vWf gene expression is not accompanied by gross genomic rearrangements. In addition, there is significant homology of C-terminal coding sequences among the vWf genes of several vertebrate species.**

von Willebrand factor (vWf) is a large, multimeric glycoprotein that is of special interest because of its critical role in hemostasis. One of the earliest responses to vascular injury is the specific, vWf-mediated adherence of activated platelets to damaged areas of the subendothelium. Further, vWf serves as a physiologically important plasma "carrier" of factor VIIIC, the antihemophilic factor. Quantitative and qualitative deficiencies in vWf are manifest in von Willebrand disease, the most common inherited human bleeding disorder. This disease is characterized by a prolonged bleeding time and is heterogeneous in its clinical and laboratory manifestations (reviewed in ref. 1).

vWf appears to be synthesized only in endothelial cells and megakaryocytes. Its biosynthetic pathway is complex and involves the initial production of a preproprotein of ≈350 kDa, which then dimerizes and undergoes further posttranslational processing prior to secretion from the cell. Specific proteolytic cleavage of the pro-vWf dimer is temporally associated with the formation of a series of vWf multimers composed of up to ≈50 subunits, each of ≈250 kDa (reviewed in ref. 2).

Human vWf cDNA clones have now been isolated and characterized in several laboratories, including our own (3–8). These studies, as well as direct amino acid sequence analysis (9), have revealed the complete vWf primary structure and indicate that the protein is composed of a signal peptide, a pro-specific portion, and a mature subunit of 22, 741, and 2050 amino acids, respectively. In keeping with the length of the protein, vWf cDNA clones hybridize to an apparently single ≈8.8-kilobase (kb) vWf mRNA present in human endothelial cells. Sequence analysis has also shown that vWf is largely composed of irregularly located repeats of five unrelated "domains" (7, 10).

Until now there has been little information on the structure and organization of the vWf gene, although preliminary evidence suggested that it was likely to be quite large (6). It is clear that an appreciation of its structure will be essential to efforts aimed at elucidating the mechanisms that govern normal vWf biosynthesis. Thus, we report here the cloning and initial characterization of the complete coding unit of the human vWf gene and certain 5′ and 3′ flanking sequences. In addition, we have mapped the vWf transcription initiation site and sequenced a portion of the vWf promoter region.

MATERIALS AND METHODS

Isolation of cDNA Clones. A partial vWf cDNA clone, pDL34, has been described (3). A full-length (≈8.8 kb) vWf cDNA was subsequently assembled from clonal members of a new, human endothelial cell cDNA library. First strand synthesis for this library was primed with synthetic oligonucleotides complementary to specific segments of the vWf mRNA. A description of this full-length cDNA and its full-length SP6 promoter-driven transcript will be presented elsewhere. The sequences of these oligonucleotides were derived from the data of Sadler *et al.* (4).

Screening of Human Genomic Libraries. A bacteriophage λ Charon 4A library, constructed from random partially digested *Alu* I and *Hae* III restriction fragments of human embryonic DNA, was kindly provided by T. Maniatis and screened by standard methods (11). A human genomic library, constructed from partially digested *Sau*3AI fragments (average insert size, ≈40 kb) cloned into the cosmid vector pCos2-EMBL (12), was kindly provided by H. Lehrach. Colony screening was performed by established techniques (13).

Primer Extension and S1 Nuclease Protection Analysis. Total cell RNA from cultured human umbilical vein endothelial cells was isolated by the guanidium isothiocyanate method of Chirgwin *et al.* (14). Poly(A)-RNA was purified by chromatography of total cell RNA on a column of oligo(dT)-cellulose. For primer extension, 0.5 pmol of a 5′ end-labeled oligonucleotide primer, complementary to bases −126 to −87 of vWf cDNA (7), was incubated with 1 μg of endothelial cell poly(A)-RNA in H_2O for 10 min at 68°C. The solution was

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.

Abbreviations: vWf, von Willebrand factor; nt, nucleotide(s).

‡To whom reprint requests should be addressed.

made 0.2 M KCl, the incubation was continued at 68°C for another 30 min, and the reaction mixture was cooled to 42°C. Reverse transcription was carried out under standard conditions (11) with 5 units of avian myeloblastosis virus enzyme (Life Sciences, St. Petersburg, FL). For S1 nuclease protection analysis, a modification of the method of Weaver and Weissman was employed (15). A 209-base-pair (bp) fragment [nucleotides (nt) 1–209 in Fig. 4] of pCos5, extending from the *Hin*dIII site to the downstream *Hpa* II site ending at base 209, was dephosphorylated and 5′ end-labeled in the presence of T4 polynucleotide kinase. One hundred nanograms of this probe and 1 μg of endothelial cell poly(A)-RNA were dissolved in 15 μl of 80% formamide/0.4 M NaCl/1 mM EDTA/40 mM Pipes, pH 6.5. The solution was heated to 80°C for 5 min and then incubated at 50°C for 16 hr. After ethanol precipitation and drying, the nucleic acids were digested at 37°C for 90 min in 20 μl of buffer containing 16 units of S1 nuclease (Sigma), as described (11).

RESULTS

Isolation and Verification of Human Genomic vWf Clones. Human vWf genomic clones were isolated by using restriction fragments and oligonucleotides derived from the sequence of vWf cDNA as probes. Initially, a bacteriophage λ Charon 4A library containing randomly generated, partial human *Alu* I and *Hae* III restriction fragments (12–20 kb) was screened with radiolabeled fragments from one of the cloned vWf cDNA plasmids [pDL34 (3)]. The first cDNA probe used was a restriction fragment that encodes the C-terminal 83 amino acids of vWf and includes the TGA termination codon (M–N in Fig. 1). A positive clone, LvW1, was identified and found to contain an ≈14-kb human DNA insert. Restriction hybridization analysis indicated that it contained sequences homologous to only ≈500–600 bp of vWf cDNA (data not shown). LvW1 contains sequences encoding the 3′ portion of vWf and extends ≈5 kb into the 3′ flanking region of the vWf genome (Fig. 1). Three additional overlapping phage clones, LvW2–4, were isolated using a series of restriction fragment probes generated from the 3′ region of the vWf cDNA. Restriction and hybridization analyses suggested that, together, LvW1–4 spanned ≈40 kb of continuous vWf genomic sequence and contained ≈3 kb of vWf cDNA sequence (Fig. 1.)

The analysis of the phage clones implied that the human vWf gene was substantially larger than the ≈8.8-kb mRNA and could prove to be >100 kb. We, therefore, decided to utilize a cosmid library containing human DNA fragments (average size, ≈40 kb) cloned into the vector pCos2EMBL. A plating of approximately eight genome equivalents of this library was screened, in parallel, with vWf cDNA restriction fragments and large synthetic oligonucleotides corresponding to defined segments of the vWf coding unit. A set of six overlapping cosmids (pCos1A–5) spanning ≈175 kb of human chromosomal sequence was identified and found to contain the entire vWf mRNA coding region, which appears to be ≈150 kb (Fig. 1). The portion of the mRNA coding region represented in each cosmid was determined by hybridization to the above-noted panel of oligonucleotides and restriction fragments (Fig. 1). It is clear from their relative sizes that each cosmid insert contains substantial amounts of intron sequences and, thus, it appears that these intervening sequences are distributed throughout the vWf coding unit. Various fragments from individual genomic clones were also checked for hybridization to vWf mRNA. As shown in the examples in Fig. 1 *Inset*, all tested genomic fragments contained sequences homologous to the ≈8.8-kb vWf mRNA. Overlaps between segments of adjacent cosmids were confirmed by demonstrating identical restriction sites at the same location in their homologous segments. Together, pCos1A–5 were found to contain ≥175 kb of contiguous human genomic sequences and to span the entire vWf coding unit.

For each cosmid clone, we found that representative restriction fragments detected by hybridization to vWf cDNA probes could subsequently be identified in comparative

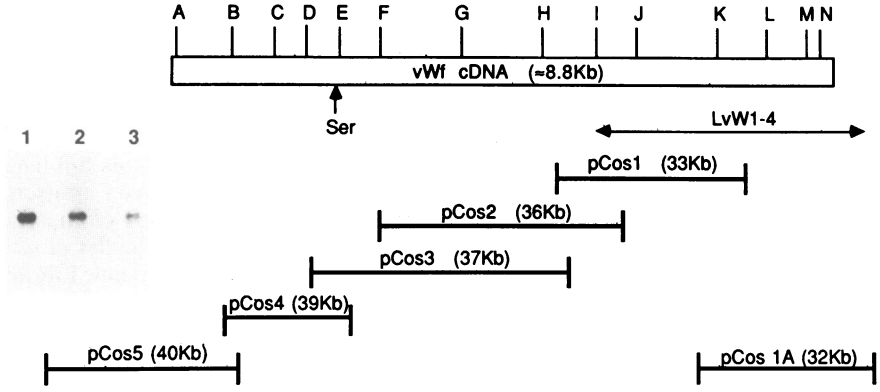

FIG. 1. Schematic representation of human vWf genomic clones. The cDNA sequence corresponds to the intact vWf coding unit. The position of the N terminus (serine) of the mature plasma protein is indicated. Letters above the cDNA sequence indicate locations of either the restriction sites (A–D, J–N) or of the oligonucleotide sequences (E–I) used as probes in the isolation of the clones. The oligonucleotide probes were originally synthesized by using the sequence information of Sadler *et al.* (4). E, a 31-mer whose sequence is complementary to nt 2328–2359 (8) in vWf cDNA; F, G, H, and I, oligonucleotides whose sequences are complementary to cDNA nt 2586–2618, 3866–3899, 4858–4892, and 5670–5715, respectively. Restriction sites in the cDNA are as follows: A, *Eco*RI; B, *Xba* I; C, L, and N, *Pvu* II; D, *Nar* I; J, K, M, and N, *Pst* I. At N, the *Pst* I and *Pvu* II sites overlap. All of these oligonucleotides and restriction fragments were used in the analytic hybridization analysis of each isolated clone. The termination codon (TGA) of the vWf cistron is located between restriction sites M and N. LvW1–4 are the phage genomic clones isolated from the Charon 4A library described in the text. Solid lines indicate the inclusive areas of the vWf cDNA to which the respective cosmid clones display homology in hybridization experiments. Each cosmid contains substantial amounts of intron sequences. We do not, as yet, know the precise 5′ and 3′ boundaries of each clone. The size of each of the six cosmid clones is indicated as is the approximate overlap between the clones. Also shown is the approximate overlap between the phage clones LvW1–4 and cosmid clones pCos1 and pCos1A. (*Inset*) RNA transfer blot. Total RNA (8.5 μg per lane) from human endothelial cells was denatured, electrophoresed through a 0.85% agarose/formaldehyde gel, and transferred to nitrocellulose, as described (3, 16). Individual lanes from the blot were hybridized with the following nick-translated probes: lane 1, a 263-bp *Pst* I cDNA fragment (fragment M–N); lane 2, an 870-bp *Pst* I genomic fragment from LvW1, which was used in the sequence analysis shown in Fig. 2; lane 3, a 3.1-kb *Eco*RI genomic fragment subcloned from LvW3, which contains sequences homologous to cDNA sequences in fragment J–K.

digests of chromosomal DNA. Although not conclusive, this comparative analysis of large segments of each clone and the corresponding cellular DNA strongly suggests that no major rearrangements of the vWf gene have occurred during cloning. In addition, probes from multiple regions of the vWf cDNA identified the same restriction fragments in vWf expressing endothelial cells and in nonexpressing leukocytes, suggesting that no major sequence rearrangement is associated with the expression of the vWf gene.

Hybridization of several small, individual cDNA restriction fragments or oligonucleotides to digests of total human DNA yielded, in each case, only a single, homologous band of cellular DNA. These findings strongly suggest that there is a single or a small number of identical copies of the vWf gene in the genome. In this regard, we and others have localized the human vWf gene to chromosome 12 (refs. 5 and 6; unpublished data).

To confirm that the initial genomic clone, LvW1, contains authentic vWf cDNA sequences and includes the C terminus of the molecule, an 870-bp *Pst* I fragment of this clone that hybridized to the cDNA segment encoding the 3′ terminus of vWf (M–N, Fig. 1) was subcloned and sequenced (Fig. 2). Comparison of the genomic sequence with that of the cDNA fragment confirms that the cloned, genomic fragment encodes the C-terminal 83 amino acids of vWf and includes the vWf termination codon (TGA). It also reveals the existence of an ≈600-bp intron that interrupts the coding sequence after the triplet encoding glutamine at position 2751 (8). The sequences at each boundary of the intron correspond closely to consensus splice donor and acceptor sequences. Furthermore, size analyses suggest that LvW1 and the corresponding cosmid, pCos1A (Fig. 1), extend ≈5 kb downstream of the vWf termination codon.

Analysis of the 5′ End of the Human vWf Genome. To identify the vWf transcription start site, we constructed a 5′ end-labeled, anti-message sense 40-base oligonucleotide primer complementary to a sequence located 87–126 bases upstream from the initiation codon for prepro-vWf in the cDNA molecule (also see Fig. 4). This primer was hybridized to human endothelial cell poly(A)-RNA and extended with reverse transcriptase. The products of the reaction were sized by denaturing polyacrylamide gel electrophoresis (Fig. 3*A*), and the major species was found to migrate as a 159/160-bp doublet (Fig. 3*A*, lane 3). Such a doublet could be caused by incomplete reverse transcription of the capped nucleotide and/or alternative transcription starts on adjacent bases. From the size of extended primers, we concluded that

```
CTG CAG TAT GTC AAG GTG GGA AGC TGT AAG TCT GAA GTA GAG
Leu Gln Tyr Val Lys Val Gly Ser Cys Lys Ser Glu Val Glu

GTG GAT ATC CAC TAC TGC CAG gtaagggctctgcttcaataagggct..
Val Asp Ile His Tyr Cys Gln

...≈550 bp.....atcttcctctgctttctgcagGGC AAA TGT GCC AGC
                                    Gly Lys Cys Ala Ser

AAA GCC ATG TAC TCC ATT GAC ATC AAC GAT GTG CAG GAC CAG
Lys Ala Met Tyr Ser Ile Asp Ile Asn Asp Val Gln Asp Gln

TGC TCC TGC TGC TCT CCG ACA CGG ACG GAG CCC ATG CAG GTG
Cys Ser Cys Cys Ser Pro Thr Arg Thr Glu Pro Met Gln Val

GCC CTG CAC TGC ACC AAT GGC TCT GTT GTG TAC CAT GAG GTT
Ala Leu His Cys Thr Asn Gly Ser Val Val Tyr His Glu Val

CTC AAT GCC ATG GAG TGC AAA TGC TCC CCC AGG AAG TGC AGC
Leu Asn Ala Met Glu Cys Lys Cys Ser Pro Arg Lys Cys Ser

AAG TGA GGCTGCTGCAG
Lys ---
```

FIG. 2. Partial nucleotide sequence analysis of a genomic fragment containing the 3′ end of vWf. The 870-bp *Pst* I fragment of LvW1 that hybridized to a 263-bp *Pst* I cDNA fragment (fragment M–N, Fig. 1) was subcloned and partially sequenced (17). Exon sequences are in uppercase; those at the beginning and end of the ≈600-bp intron are in lowercase. The predicted amino acid sequence is given, and the termination codon (TGA) is indicated by dashes. The exon sequences match precisely that for the corresponding portion of the vWf cDNA clone, pDL34 (3). The exon sequence corresponds to nt 4328–4591 of Sadler *et al.* (4).

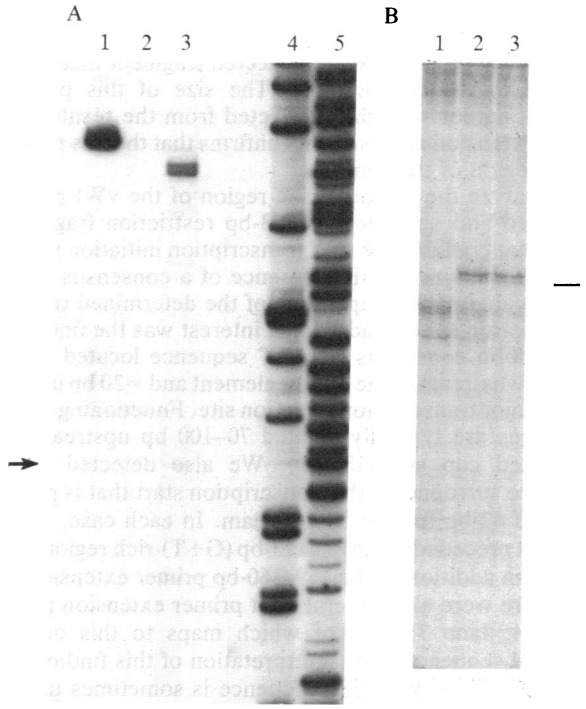

FIG. 3. Determination of the vWf transcription start site. (*A*) Primer extension experiment using a synthetic oligonucleotide that hybridized to bases −126 to −87 of the vWf mRNA (7). Reaction products were run on an 8% polyacrylamide sequencing-type gel. A 10-day exposure is shown. Lane 1, 163-base end-labeled marker. Lane 2, product of a reaction containing 1 μg of yeast RNA and ≈1 ng of synthetic vWf RNA made with SP6 RNA polymerase using subcloned vWf cDNA as template. The predicted ≈100-base product was seen at the bottom of the gel (not shown). Lane 3, product of a reaction containing 1 μg of endothelial poly(A)-RNA. Lanes 4 and 5, sequencing reactions run for size markers. Pilot experiments showed no specific products when endothelial poly(A)-depleted RNA was used as a template. The upper band of the doublet in lane 3 is 160 bases, indicating that vWf mRNA extends 246 bases upstream of the translation start. The possible significance of a collection of minor bands, the largest of which is indicated by the arrow at the left, is discussed in the text. (*B*) S1 nuclease protection experiment using an end-labeled 209-bp fragment of pCos5, extending from an upstream *Hin*dIII site (base 1, Fig. 4) to the *Hpa* II site ending at base 209. Reaction products were run on an 8% polyacrylamide sequencing-type gel. A 12-day exposure is shown. Lane 1, product of a reaction containing 1 μg of yeast RNA and ≈1 ng of SP6-vWf synthetic RNA. The predicted ≈46-base protected fragment was seen at the bottom of the gel (not shown). Lanes 2 and 3, independent duplicates with 1 μg of endothelial poly(A)-RNA. In each lane, a small amount of undigested probe was seen (not shown). Pilot experiments showed no specific products when rabbit liver RNA was used. The position of a 114-base marker is indicated by the bar at the right. A sequencing reaction was also run on the same gel for size markers (not shown). The protected fragment at 116 bases indicates vWf mRNA extends to 246 bases upstream of the translation start, in agreement with the result of primer extension analysis. Additional bands present in each lane are presumably the result of some degree of incomplete probe digestion that is not dependent upon the presence of endothelial mRNA.

the vWf transcription initiation site was located 246/245 nt upstream of the initiator methionine codon (nt 95/94 in the 5′ flanking genomic sequence shown in Fig. 4). S1 endonuclease protection analysis of the 5′ segment of the vWf mRNA was performed to verify the location of the transcription start site defined by primer extension and to eliminate the possibility of an intron in this area of the gene. A 209-bp *Hin*dIII–*Hpa* II 5′ end-labeled restriction fragment of pCos5 (nt 1–209, Fig. 4), which is homologous to sequences at the 5′ end of the vWf cDNA, was hybridized to endothelial cell poly(A)-RNA. Heteroduplexes were then exposed to S1 endonuclease, and the protected fragment(s) were sized by denaturing gel electrophoresis. The autoradiogram shown in Fig. 3*B* reveals the presence of an ≈116-bp protected fragment that was not detected in the control lane. The size of this protected fragment agrees with that predicted from the results of the primer extension analysis and confirms that there is no intron in this region of the genome.

To analyze the 5′ noncoding region of the vWf gene, we subcloned and sequenced a 243-bp restriction fragment of pCos5 that contains the vWf transcription initiation site (Fig. 4). The results reveal the presence of a consensus "TATA box" located ≈30 bp upstream of the determined transcription initiation site. Of additional interest was the finding of a perfect 8-bp consensus CCAAT sequence located immediately downstream of the TATA element and ≈20 bp upstream of the major transcription initiation site. Functioning CCAAT sequences are typically located 70–100 bp upstream of an associated cap site (18, 19). We also detected an 8-bp sequence surrounding the transcription start that is perfectly repeated 42 bp further downstream. In each case, the 8-bp repeat is preceded by an 8- to 13-bp (G+T)-rich region. In this regard, in addition to the 159/160-bp primer extension product, there were also several faint primer extension products (Fig. 3*A*, lane 3), one of which maps to this octameric repeated sequence. One interpretation of this finding is that the downstream repeated sequence is sometimes used as a transcription initiation site.

Cross-Species Homology Among vWf Coding Sequences. Hybridization analyses were undertaken to study the conservation of 3′ vWf gene sequences. Total DNA from quail and several mammalian species (human, field vole, rat, hamster, monkey, and mouse) was digested with *Eco*RI and then subjected to Southern blot analysis. The hybridization probe was a human cDNA restriction fragment (L–N, Fig. 1) from the 3′ end of vWf that encodes amino acids 2476–2813 (8). When this cDNA probe was hybridized to the vertebrate DNAs and washed under moderately stringent conditions, hybridizing bands were detected in all of the mammalian DNAs (Fig. 5). This suggests that there is significant homology in this 3′ region of the gene among the human and the other mammalian genomes that were tested. By contrast, there was little or no hybridization between the human cDNA probe and quail DNA sequences (Fig. 5, lane F). The size of the DNA fragments that were homologous to human vWf sequences varied, indicating that the structure and/or the location of the target gene varies from species to species. Conceivably, by analogy with other genes, such differences could result, in part, from nonhomology of intron sequences (see, for example, ref. 21).

```
                                                          40
AAGCTTTATC AGCTTGGAGG TACTTCTAAT ACATTTCCTT
                                                          80
TCATTGTTTC CTTTTGGTAA TTAAAAGGAG GCCAATCCCC
                                                         120
TGTTGTGGCA GCTCACAGCT ATTGTGGTGG GAAAGGGAGG
                                                         160
GTGGTTGGTG GATGTCACAG CTTGGGCTTT ATCTCCCCCA
                                                         200
GCAGTGGGAC TCCACAGCCC CTGGGCTACA TAACAGCAAG
                                                         240
ACAGTCCGGA GCTGTAGCAG ACCTGATTGA GCCTTTGCAG

CAG
```

FIG. 4. DNA sequence of the 5′ region of human vWf gene. A 243-bp *Hin*dIII–*Pvu* II fragment of pCos5, which hybridized to the 5′ end of vWf cDNA, was subcloned in mp18 and mp19; both strands were sequenced (17). The TATA box and CCAAT consensus sequence are underscored with a heavy line. The diamond at nt 95 indicates the major vWf transcription initiation site determined by primer extension and S1 nuclease analysis. This site is 246 bp upstream of the initiator methionine codon. The 8-bp repeated sequences around the primary initiation site and the downstream repeat are underscored with a thin line. The 40-mer synthetic oligonucleotide used in the primer extension analysis begins at nt 213 and extends 40 nt downstream. The *Hpa* II site (nt 206–209) was used in preparation of the fragment for S1 nuclease analysis (see legend to Fig. 3).

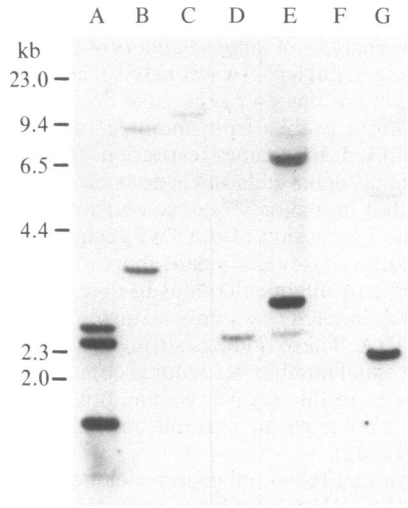

FIG. 5. Southern blot analysis of various vertebrate DNAs. High molecular weight DNA (9 μg per lane) was digested to completion with *Eco*RI, electrophoresed through a 1% agarose gel, and transferred to nitrocellulose. The entire blot was hybridized with a nick-translated 870-bp *Pvu* II fragment of vWf cDNA (fragment L–N, Fig. 1; ref. 11). After hybridization, the blot was washed in 0.15 M NaCl/15 mM sodium citrate/0.1% $NaDodSO_4$ at 68°C for 2 hr. Descriptions of the various cell lines used in this analysis were provided earlier (20). The DNAs displayed were prepared from human endothelial cells (lane A), a normal *Microtus agrestis* (field vole) cell line (lane B), normal rat kidney cell line NRK (lane C), baby hamster kidney cell line BHK-21 (lane D), African green monkey cell line CV-1 (lane E), a chemically transformed quail cell line, Qt6 (lane F), and an immortal but otherwise untransformed mouse cell line, BALB/c 3T3 Cl A31 (lane G). Size markers are indicated on the left.

DISCUSSION

The human vWf gene is clearly a large structure by comparison with most previously studied eukaryotic genes. The overlapping cosmid clones described here contain the complete vWf coding unit and span ≈175 kb of the genome. They include ≈25 kb of upstream sequence and extend ≈5 kb downstream of the vWf termination codon.

Southern blotting and hybridization data reveal that the analyzed portion of each gene clone is composed of fragments present in the corresponding region of chromosomal DNA. In addition, the available clones have yielded a single, unique set of restriction fragments for each region of the genome analyzed. These results suggest that, at least for those portions of the clones examined in detail, no major rearrangements or loss of vWf sequences have occurred

during cloning and that human cells probably contain a single copy of the vWf gene.

The finding of significant homology between human and various mammalian vWf coding units using a cDNA probe derived from the 3′ end of the human gene is consistent with a high degree of conservation of the C-terminal portion of vWf structure among the mammalian species tested. Cysteine-rich sequences within this region of the protein are essential to pro-dimer formation (2). Since this reaction is likely to be critical to the proper processing and secretion of the mature protein, maintenance of certain structural motifs within the C-terminal region during evolution would be expected.

Direct sequence and RNA mapping analyses of the 5′ noncoding region of the vWf gene revealed the site(s) of transcription initiation and two consensus structures known to play a regulatory role in the transcription of other mammalian genes. A TATA box, known to be necessary for accurate and efficient transcription initiation of many genes (22), is located at its customary position, ≈30 bp upstream of the transcription start site. One striking finding was the presence of a perfect, consensus 8-bp CCAAT element located just downstream of the TATA sequence. In promoters from a variety of mammalian genes, the "CCAAT box" appears to be a cis-acting positive regulatory element and is now known to be recognized by sequence-specific transcription factors (23–25). Binding of these factors to the CCAAT sequence appears to modulate transcription activity and may be important in the cell-specific activation of certain genes (26, 27). If the observed vWf CCAAT element is functional and participates in the regulation of transcription from the major observed start site, this would be unusual by comparison with other genes. Downstream CCAAT boxes have been found in the 5′ flanking regions of at least two other human genes, α_1-antitrypsin and prolactin (28, 29). These genes, as well as the vWf gene, encode secreted proteins that are synthesized by a limited repertoire of cells. It will be interesting to determine whether the downstream CCAAT element is involved in some common aspect of the regulation of these genes. Additional sequence analysis of pCos5 will be necessary to determine if there is also a CCAAT element upstream of the primary vWf transcription start site.

As noted earlier, there are two identical repeats of the 8-bp sequence containing the transcription start site, and each repeat is preceded by a (G+T)-rich region 4 nt upstream. In this regard, perfect and imperfect repeats of transcription start sequences and corresponding upstream (G+T)-rich regions have been noted in other genes, but the role, if any, of these sequences in the regulation of transcription is unclear (30, 31). Since appropriate size minor bands were detected in the primer extension analysis, it is conceivable that the region around the downstream site is occasionally used for transcription initiation in endothelial cells. Moreover, the position of the aforementioned CCAAT box ≈70 bp upstream of this downstream site allows one to speculate that there may be two overlapping promoters controlling vWf expression. The primary transcription start at nt 95 (Fig. 4) may be regulated by the TATA box and possibly other, as yet undefined, elements, whereas initiation in the vicinity of the downstream "TATA-less" repeat may be, at least partially, governed by the CCAAT element. Experiments designed to test the functional contribution of each of these possible regulatory elements are necessary.

We thank our colleagues Tom Roberts and Myles Brown, Division of Neoplastic Disease Mechanisms, and Temple Smith and Donald Faulkner, Molecular Biology Computer Research Resource, Dana-Farber Cancer Institute, for helpful discussions during the course of this work. We are grateful to Ann Desai and Esta Bernier for help with preparation of the manuscript. This research was supported by grants from the National Institutes of Health (HL-31311) and Meloy Laboratories, Inc. (Springfield, VA). D.C.L. is an Established Investigator of the American Heart Association with funds contributed by the Massachusetts Division.

1. Zimmerman, T. S., Ruggeri, Z. M. & Fulcher, C. A. (1983) *Prog. Hematol.* **13,** 279–309.
2. Lynch, D. C., Zimmerman, T. S. & Ruggeri, Z. M. (1986) *Br. J. Haematol.* **64,** 15–20.
3. Lynch, D. C., Zimmerman, T. S., Collins, C. J., Brown, M., Morin, M. J., Ling, E. H. & Livingston, D. M. (1985) *Cell* **41,** 49–56.
4. Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., Harlan, J. M., Titani, K. & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6394–6398.
5. Verweij, C. L., de Vries, C. J. M., Distel, B., van Zonneveld, A., van Kessel, A. G., van Mourik, J. A. & Pannekoek, H. (1985) *Nucleic Acids Res.* **13,** 4699–4717.
6. Ginsburg, D., Handin, R. I., Bonthron, D. T., Donlon, T. A., Bruns, G. A. P., Latt, S. A. & Orkin, S. H. (1985) *Science* **228,** 1401–1406.
7. Verweij, C. L., Diergaarde, P. J., Hart, M. & Pannekoek, H. (1986) *EMBO J.* **5,** 1839–1847.
8. Bonthron, D., Orr, E. C., Mitsock, L. M., Ginsburg, D., Handin, R. I. & Orkin, S. H. (1986) *Nucleic Acids Res.* **14,** 7125–7127.
9. 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* **25,** 3171–3184.
10. Shelton-Inloes, B. B., Titani, K. & Sadler, J. E. (1986) *Biochemistry* **25,** 3164–3171.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
12. Poustka, A., Rackwitz, H.-R., Frischauf, A.-M., Hohn, B. & Lehrach, H. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4129–4133.
13. Hanahan, D. & Meselson, M. (1980) *Gene* **10,** 63–67.
14. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
15. Weaver, R. F. & Weissman, C. (1979) *Nucleic Acids Res.* **7,** 1175–1193.
16. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142.
19. McKnight, S. & Kingsbury, R. (1982) *Science* **217,** 316–324.
20. Collins, C. J., Boettiger, D., Green, T. L., Burgess, M. B., Devlin, B. H. & Parsons, J. T. (1980) *J. Virol.* **33,** 760–768.
21. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4853–4857.
22. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
23. Dierks, P., Van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. & Weissmann, C. (1983) *Cell* **32,** 695–706.
24. Jones, K. A., Yamamoto, K. R. & Tjian, R. (1985) *Cell* **42,** 559–572.
25. Myers, R. M., Tilly, K. & Maniatis, T. (1986) *Science* **232,** 613–618.
26. Graves, B. J., Johnson, P. F. & McKnight, S. L. (1986) *Cell* **44,** 565–576.
27. Bienz, M. (1986) *Cell* **46,** 1037–1042.
28. Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Davie, E. W. & Woo, S. L. C. (1982) *Nature (London)* **297,** 655–659.
29. Cooke, N. E. & Baxter, J. D. (1982) *Nature (London)* **297,** 603–606.
30. Seidman, C. E., Block, K. D., Klein, K. A., Smith, J. A. & Seidman, J. G. (1984) *Science* **226,** 1206–1209.
31. Nathans, J. & Hogness, D. S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4851–4855.