Human von Willebrand Factor Gene and Pseudogene: Structural Analysis and Differentiation by Polymerase Chain Reaction†,‡

David J. Mancuso,§ Elodee A. Tuley,§ Lisa A. Westfield,§ Teresa L. Lester-Mancuso,§ Michelle M. Le Beau,∥ James M. Sorace,§,⊥ and J. Evan Sadler*,§

Howard Hughes Medical Institute Laboratories, Department of Medicine, and Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, and Joint Section of Hematology/Oncology, Department of Medicine, University of Chicago Medical Center, Chicago, Illinois 60637

Received August 3, 1990; Revised Manuscript Received September 11, 1990

ABSTRACT: Structural analysis of the von Willebrand factor gene located on chromosome 12 is complicated by the presence of a partial unprocessed pseudogene on chromosome 22q11–13. The structures of the von Willebrand factor pseudogene and corresponding segment of the gene were determined, and methods were developed for the rapid differentiation of von Willebrand factor gene and pseudogene sequences. The pseudogene is 21–29 kilobases in length and corresponds to 12 exons (exons 23–34) of the von Willebrand factor gene. Approximately 21 kilobases of the gene and pseudogene were sequenced, including the 5′ boundary of the pseudogene. The 3′ boundary of the pseudogene lies within an 8-kb region corresponding to intron 34 of the gene. The presence of splice site and nonsense mutations suggests that the pseudogene cannot yield functional transcripts. The pseudogene has diverged ~3.1% in nucleotide sequence from the gene. This suggests a recent evolutionary origin ~19–29 million years ago, near the time of divergence of humans and apes from monkeys. Several repetitive sequences were identified, including 4 Alu, one Line-1, and several short simple sequence repeats. Several of these simple repeats differ in length between the gene and pseudogene and provide useful markers for distinguishing these loci. Sequence differences between the gene and pseudogene were exploited to design oligonucleotide primers for use in the polymerase chain reaction to selectivity amplify sequences corresponding to exons 23–34 from either the von Willebrand factor gene or the pseudogene. This method is useful for the analysis of gene defects in patients with von Willebrand disease, without interference from homologous sequences in the pseudogene.

Human von Willebrand factor (vWF)[1] is a complex multimeric plasma glycoprotein that is required for platelet adhesion to damaged endothelium and for factor VIII survival in the circulation [reviewed in Sadler (1989)]. Deficiency of vWF results in von Willebrand disease (vWD), which is the most common inherited bleeding disorder in man if all degrees of severity are included (Rodeghiero et al., 1987). Many deficient phenotypes have been identified (Ruggeri & Zimmermann, 1987).

The human vWF gene has been cloned, and its structure was reported in detail recently. It spans ~178 kb of DNA and contains 52 exons. A high-resolution restriction map was constructed for ~75% of the locus, and ~20% of the gene sequence has been reported (Mancuso et al., 1989b). In the course of characterizing the vWF gene, restriction fragments from the center of the vWF cDNA were found to hybridize not only to the vWF locus on chromosome 12 but also to sequence on chromosome 22 (Shelton-Inloes et al., 1987; Mancuso et al., 1988). This cross-hybridizing locus proved to be a pseudogene for vWF. The presence of this pseudogene intereferes with analysis by Southern blotting of gene defects in patients with vWD and has caused some difficulty in the identification of restriction fragment length polymorphisms for the gene.

The vWF pseudogene corresponds to a region of the vWF gene that encodes important functional domains of the protein and also has an interesting evolutionary history. The human vWF precursor consists of four types of repeated domains, present in 2–5 copies each in the following order: D1-D2-D′-D3-A1-A2-A3-D4-B1-B2-B3-C1-C2 (Bonthron et al., 1986; Verweij et al., 1986; Shelton-Inloes et al., 1986). The segment of the gene that is duplicated in the vWF pseudogene encodes domains A1–A3. The vWF A repeats appear to be homologous to segments of several proteins: complement factor B (Shelton-Inloes et al., 1986; Sadler et al., 1986; Mole et al., 1984), complement component C2 (Bentley, 1986), cartilage matrix protein (Kiss et al., 1989), all three α-chains of collagen type VI (Bonaldo et al., 1989, 1990; Chu et al., 1989; Koller et al., 1989), the α-subunits of the leukocyte adhesion receptors Mac-1 (Arnaout et al., 1988; Corbi et al., 1988; Pytela, 1988), p150,95 (Corbi et al., 1987), LFA-1 (Larson et al., 1989), and the α-subunit of the platelet collagen receptor (Takada & Hemler, 1989). The present distribution of vWF A domain containing genes may have arisen in part

† This work was supported by the Howard Hughes Medical Institute and (M.M.L.B.) by the University of Chicago Cancer Research Foundation and the Leukemia Society of America.

‡ The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL Data Bank under Accession Numbers M36184 and M36185.

* To whom correspondence should be addressed at the Howard Hughes Medical Institute, Washington University School of Medicine, 660 S. Euclid Avenue, Box 8045, St. Louis, MO 63110.

§ Washington University School of Medicine.

∥ University of Chicago Medical Center.

⊥ Present address: Chief, Blood Bank and Hematology, Laboratory Service (113), Veterans Administration Medical Center, 3900 Loch Raven Blvd., Baltimore, MD 21218.

[1] Abbreviations: bp, base pair(s); kb, kilobase(s); PCR, polymerase chain reaction; vWF, von Willebrand factor; vWD, von Willebrand disease; R, a purine nucleoside, adenine or guanosine; Y, a pyrimidine nucleoside, thymidine or cytidine.

0006-2960/91/0430-253$02.50/0

by repeated gene segment duplication and exon shuffling (Gilbert, 1978). The A domains of vWF contain binding sites for heparin, collagen, and platelet glycoprotein Ib (Mohri et al., 1988; Fujimura et al., 1987; Pareti et al., 1986; Roth et al., 1986). One can speculate that the homologous A domains in other proteins might also recognize one or more similar ligands. In particular, the collagen-binding properties of the platelet collagen-receptor might be dependent upon the presence of the vWF A domain in the α-chain.

We report here ~21 kb of DNA sequence for the vWF pseudogene and for the corresponding DNA sequence of the vWF gene, including the 5′ boundary of the pseudogene. the pseudogene is ~21–29 kb in length and has diverged ~3.1% from the gene. Sequence differences between the gene and pseudogene were exploited to selectively amplify segments of either locus by using the polymerase chain reaction (PCR). This method permits the analysis of gene defects in patients with vWD without interference from the pseudogene.

Experimental Procedures

Materials. Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, and T4 DNA polymerase were purchased from Bethesda Research Laboratories or New England Biolabs. The Klenow fragment of *Escherichia coli* DNA polymerase and calf intestine alkaline phosphatase were purchased from Boehringer Mannheim. Deoxycytidine 5′-[α-^{32}P]triphosphate ([α-^{32}P]dCTP), deoxyadenosine 5′-[α-^{35}S]thiotriphosphate ([^{35}S]dATPαS), and [γ-^{32}P]ATP were purchased from Amersham. Taq polymerase (Amplitaq) was purchased from Perkin-Elmer Cetus. Placental DNA was purchased from Sigma.

Isolation of vWF Pseudogene DNA Clones. The isolation of clones λ6, pWE64, and pWE113 was described previously (Mancuso et al., 1989b). The bacteriophage λ Charon 4A library was derived from fetal liver and was provided by Tom Maniatis (Harvard University, Boston) (Lawn et al., 1978). The pWE15 cosmid library was derived from human placenta (pWE15) and was provided by Glen Evans (Wahl et al., 1987) (Salk Institute, San Diego).

Chromosomal Localization by in Situ Hybridization. Human metaphase cells were prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes. The vWF cDNA fragment of plasmid λ3/SacI-2a, which contains the 5′ 3.8-kb *Sac*I–*Sac*I fragment of λHvWF3 (Shelton-Inloes et al., 1987) cloned in the *Sac*I site of pUC18, was radiolabeled by nick translation of the entire plasmid with all four ^{3}H-labeled deoxynucleoside triphosphates to a specific activity of 1.0×10^{8} dpm/μg. In situ hybridization was performed as described previously (Le Beau et al., 1984). Metaphase cells were hybridized at 2.0 and 4.0 ng of probe/mL of hybridization mixture. Autoradiographs were exposed for ~11 days.

Polymerase Chain Reaction. The polymerase chain reaction was performed (Saiki et al., 1988) with a Perkin-Elmer Cetus DNA thermal cycler. Typically, 100 ng of a cosmid or bacteriophage λ DNA isolate was amplified with 2.5 units of Taq polymerase for 30 cycles of the following: 30 s at 94 °C (denaturing), 30 s at 50–60 °C (annealing), and 1 min at 72 °C (extension), with a final extension time of 5 min at 72 °C. Human peripheral blood leukocyte DNA was amplified similarly with the following modifications. Reactions without enzyme were preheated at 99 °C for 5 min and then at 96 °C for an additional 5 min after addition of Taq polymerase. DNA was amplified for 30 cycles of the following: 1 min at 96 °C (denaturing), 1 min at 53–55 °C (annealing), and 2 min at 72 °C (extension) with a final extension time of 5 min at 72 °C.

FIGURE 1: Chromosomal localization of the human vWF pseudogene. Shown in the figure is the distribution of labeled sites on chromosome 22 in normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes hybridized with a vWF cDNA probe. Of 38 labeled sites observed on chromosome 22, 23 (86.8%) were clustered at 22q11–13, with the largest cluster at 22q11–12.

Computer-Assisted Methods. Computer programs were those of the Genetics Computer Group (Madison, WI) or Compugene (St. Louis, MO). Nucleotide sequences were analyzed and compared to the Genbank Genetic Sequence Data Bank (BBN Laboratories, Inc., Cambridge, MA, release 62.0, December 1989) and the EMBL database (release 19.0, May 1989) with the program FASTA (Lipman & Pearson, 1985).

Other Methods. Restriction mapping and Southern blotting, DNA subcloning and sequencing, and preparation of peripheral blood leukocyte DNA were performed as described previously (Mancuso et al., 1989b; Grunebaum et al., 1984).

Results

Chromosomal Localization of the vWF Pseudogene. A restriction fragment from the middle of the vWF cDNA, corresponding to amino acids 144–544 of the mature vWF subunit (Shelton-Inloes et al., 1986), was hybridized to human metaphase chromosomes. Specific labeling of chromosomes 12 and 22 was observed due to the presence of the vWF gene and pseudogene on these chromosomes, respectively. Of the 38 labeled sites observed on chromosome 22, 36 (94.7%) were clustered at bands q11–13 with the largest number of grains at bands q11–12. This cluster of grains represented 16% (38/238) of all labeled sites ($p \ll 0.005$). The distribution of grains on chromosome 22 thus indicates that the vWF pseudogene is localized to chromosome 22q11–13 (Figure 1). This result agrees with a previous report using similar methods (Patracchini et al., 1989).

Restriction Mapping of Genomic DNA Clones. We previously reported the isolation of clones that contain vWF gene and pseudogene sequences from bacteriophage λ and cosmid human genomic DNA libraries (Mancuso et al., 1989b, 1988). Two of the cosmid isolates (p113 and p64) and one bacteriophage λ isolate (λ6) contain sequence from the vWF pseudogene on the basis of the presence of characteristic restriction sites as well as splice site and nonsense mutations.

The region of the pseudogene cloned is contained in six *Eco*RI restriction fragments spanning ~21 kb (Figure 2). Three *Eco*RI restriction fragments (2.2, 1.3, and 12 kb) cross-hybridized with DNA probes corresponding to vWF exons 23–34. Each of the cloned exon-containing fragments correlates with a hybridizing fragment on Southern blots of genomic DNA from normal individuals and from patients with total deletion of the vWF locus on chromosome 12, when hybridized with vWF cDNA probes under conditions of high stringency (Shelton-Inloes et al., 1987). Thus, rearrangement of the cloned fragments probably has not occurred. All fragments outside of the vWF locus that cross-hybridize strongly with vWF cDNA probes appear to be contained within this pseudogene locus.

Detailed restriction maps for the two loci are compared in Figure 3. The 5′ boundary of the pseudogene is ~500 bp

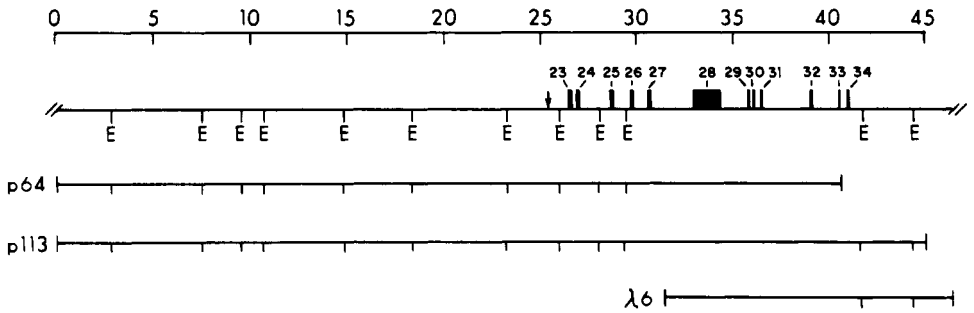


FIGURE 2: *Eco*RI restriction map of cosmid and bacteriophage λ clones for the human vWF pseudogene. The scale at the top is in kb. The regions corresponding to vWF exons are shown by solid bars above the line. *Eco*RI sites are shown below the line. The arrow (↓) indicates the 5′ boundary of the pseudogene. The pWE15 cosmid library (Wahl et al., 1987) yielded two isolates, p64 and p113, which were employed in the construction of this map. The Charon 4A bacteriophage λ library (Lawn et al., 1978) yielded one isolate, λ6.

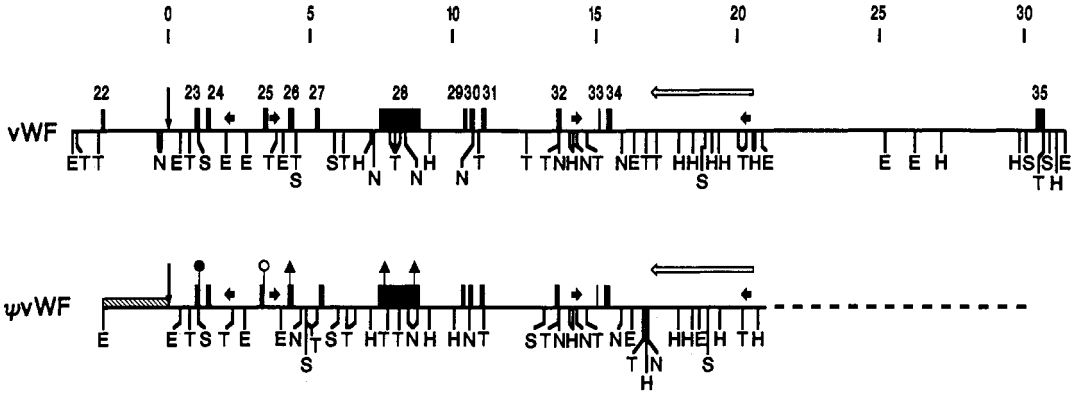


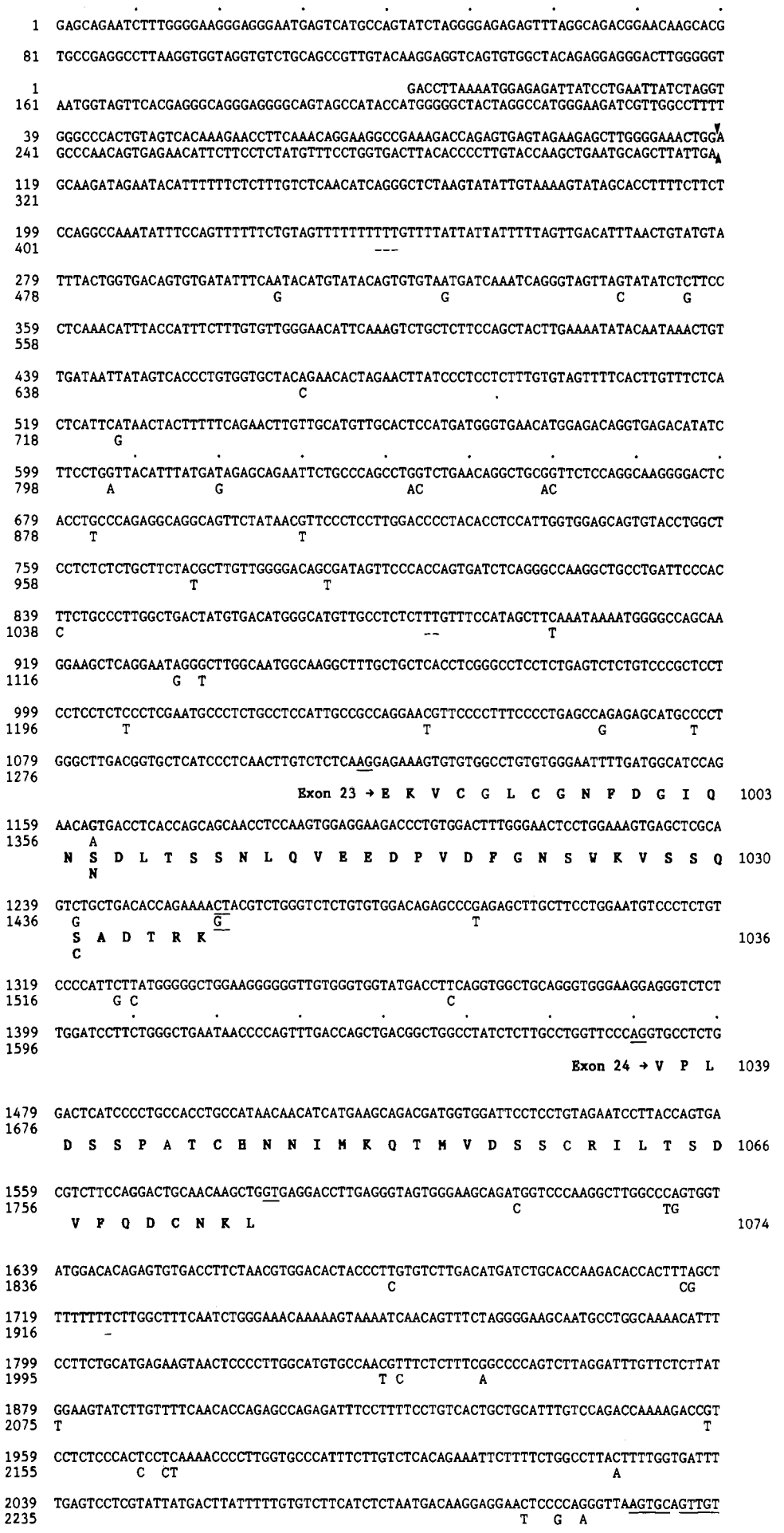
FIGURE 3: Restriction map and comparison of the vWF gene and pseudogene loci. The scale in kb is indicated at the top of the figure. Regions corresponding to vWF exons are indicated by bars above the line in each map and are numbered consecutively from 5′ to 3′. The human vWF gene is represented between exons 22 and 35, with exons represented by boxed areas. The corresponding region in the vWF pseudogene is shown below. The 5′ boundary of the pseudogene is indicated in each map by a vertical arrow (↓). Nonsense mutations (▲) and splice junction mutations (●, GT → CT; ○, GT → AT) are indicated. Filled arrows above each map show the location and direction of Alu repeats. The open arrows show the location of a Line-1 repeat. For the region sequenced (Figure 4), corresponding to coordinates 0–21 kb, the restriction map is complete. For regions 5′ or 3′ of the sequence shown in Figure 4, the *Eco*RI map is complete, but other enzymes may cut at sites that are not shown. The sites shown are as follows: E, *Eco*RI; H, *Hin*dIII; T, *Taq*I; N, *Nco*I; S, *Sac*I.

5′ of the 2.2 kb *Eco*RI fragment corresponding to the 1.5-kb *Eco*RI fragment of the gene that contains exons 23 and 24. the cloned portion of the pseudogene comprises ~21 kb of DNA, ending at a *Hin*dIII site in a region corresponding to intron 34 of the gene. The next exon in the gene, exon 35, has no cross-hybridizing homologue in the pseudogene and is ~8 kb further 3′. This was determined by hybridization of radiolabeled DNA probes containing exon 35 to genomic DNA from a patient (S.G.) with total deletion of the vWF gene (Shelton-Inloes et al., 1987) (data not shown). Thus, if no significant insertions or other structural differences with the gene are present in the remainder of the pseudogene, it is ~21–29 kb in length.

Nucleotide Sequence of the vWF Gene and Pseudogene. The pseudogene and the corresponding segments of the normal gene, not previously reported, were sequenced on both strands. A total of ~21 kb of DNA sequence for both loci are compared in Figure 4. The 5′ boundary of the pseudogene can be identified between nucleotides 117 and 118 by an abrupt change from lack of sequence similarity to nearly complete conservation of sequence with the gene. Counting insertions or deletions as one difference regardless of length, the vWF gene and pseudogene have diverged ~3.14% (Table I). Considering the gene to be the reference sequence, there are 24 insertions (total length 225 bp, range 1–162 bp), 32 deletions (total length 167 bp, range 1–71 bp), and 595 single nucleotide substitutions. Transitions account for ~73% of the substitutions, and ~29% of the substitutions could be explained by CG → TG mutations. There was slightly less variation, ~2.47%, in the sequences corresponding to exons.

There are several mutations in the pseudogene that should prevent it from generating functional transcripts that could encode a vWF-like protein. The generally conserved GT dinucleotide at the donor splice junction of intron 23 contains a GT → CT mutation. Similarly, the donor splice junction of intron 25 contains a GT → AT mutation. The sequence corresponding to exon 26 contains a nonsense mutation, and the sequence corresponding to exon 28 contains two nonsense mutations (Figures 3 and 4). These sequence differences indicate that this locus is a pseudogene rather than a related functional gene.

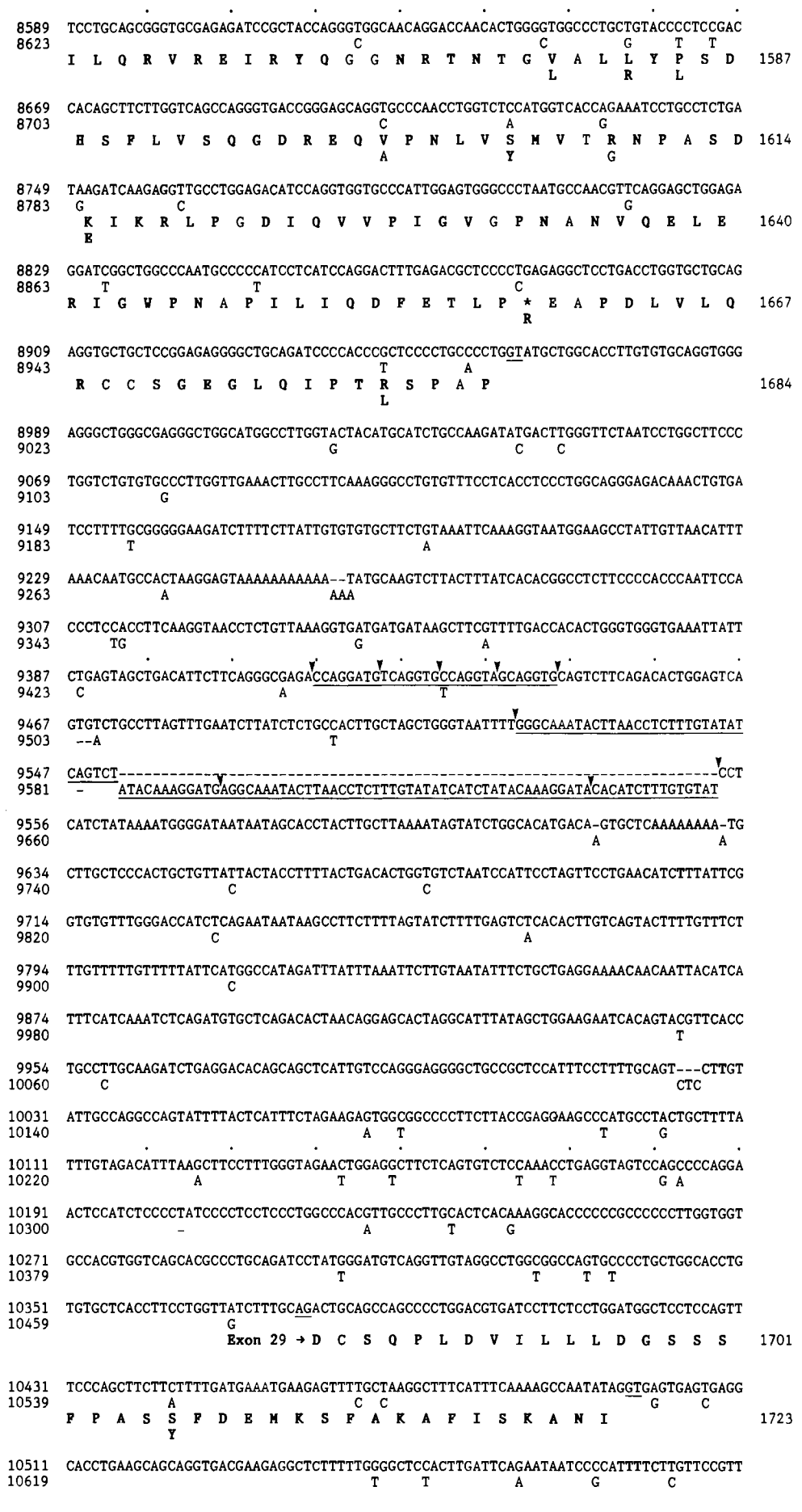
Repetitive Sequence Elements. Four complete or partial Alu repeats (Deininger & Schmid, 1979) were found in both the gene and pseudogene (Figures 3 and 4), each flanked by short direct repeats (Table II). One partial ~3.7-kb Line-1 repeat in inverted orientation was identified near the 3′ end of the pseudogene, between nucleotides ~17104 and 20790, and in the same location in the gene. The boundaries of this Line-1 element are difficult to define because it shows relatively modest conservation of sequence compared to a representative human Line-1 (Skowronski et al., 1988) with which it shares only ~50–60% identical residues. The 3′ end of the Line-1 was more easily identified by comparison with the Line-1 element in the porcine relaxin gene (Haley et al., 1987) with which it shares >64% identical residues between pseudogene nucleotides 17104 and 17800. The inverted Alu repeat repeats between pseudogene nucleotides 20377 and 20724 apparently is inserted into the Line-1 element. Direct repeats



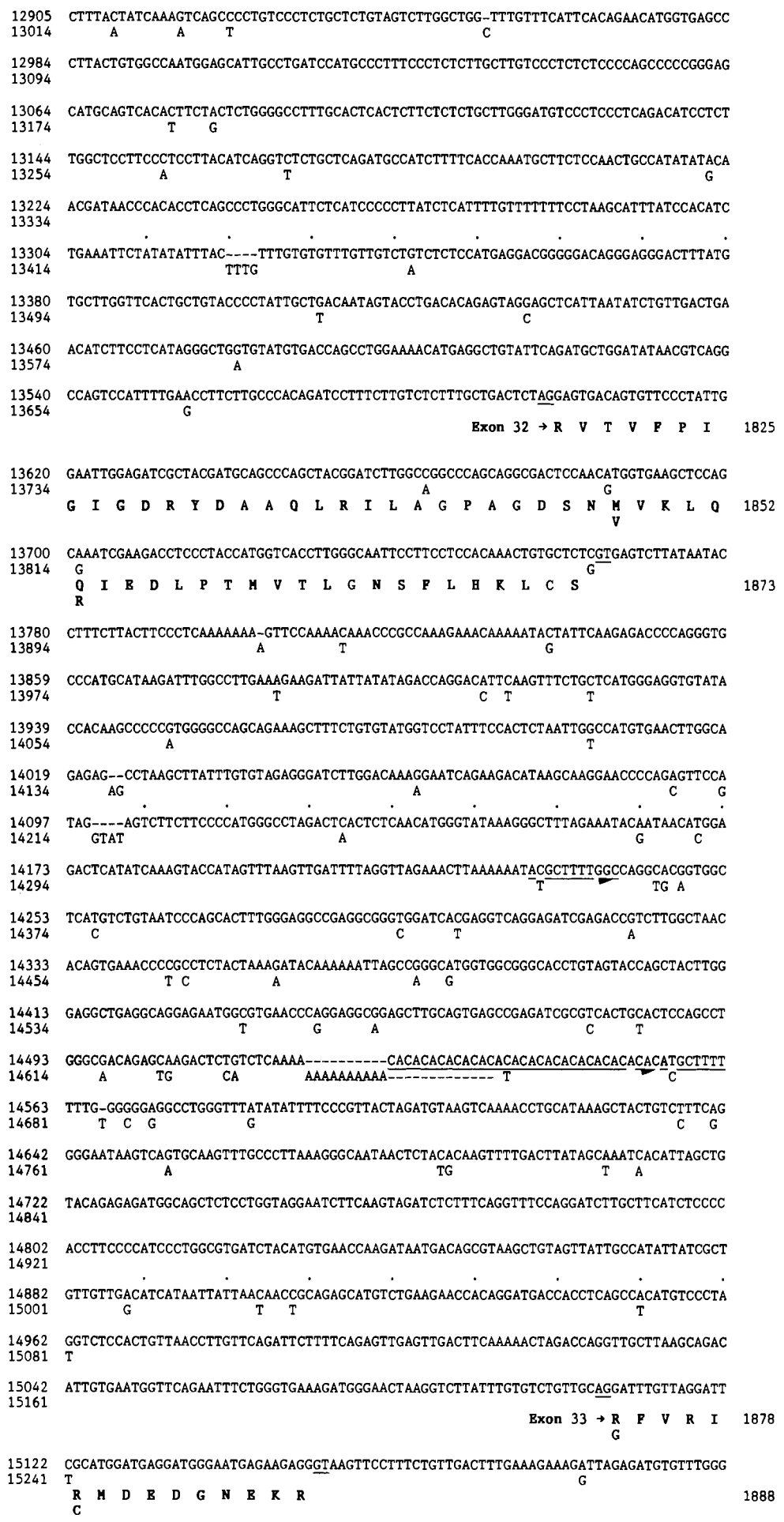
```
2119  CTCTCTTCTTTTTTTTT------GAGACAGAGTCTCGCTCTGTCACTCAGGCTGGAGTGCAATGGCATGATCTTGGCTTA
2315        -          TTTTTT                    C
                .          .          .          .          .          .          .
2193  GTGCAGCCTCTGCTTCCCGGGTTCAAGCGATTCTCCTGCCTCAACCTCCCGAGTAGCTGGGATTACAGGTACCCGCCACC
2394              C  T  C                                   C          C  T

2273  ACACCCAGCTAATTTTTGTATTTTTAGTAGAGACGAGGTTTCACCACGTTGGTCAGGCTGGCCTCGAACTCCTGACCTCA
2474    G                          T A        T              T        -

2353  GGTGATCTACCCACCTGAGCCTCCCAAAGTGCTGGGATTACAGGTGTAAGCCACCATGCCCGACCATAGTGCGGTTGTCT
2553              CG                          G                    A

2433  CTTATCTGCTGCATGTATGTGTGTACACACATGTGGACATGGGCTTGTACATGGACATATGTGACTACAGAAGTCTATGT
2633                                                                          G

2513  ACACACACACAGCCATGCACACATACATGTTGCTGTGTAGCACCACCAATGTGGGAGAGACTGGTTGAAAACATGGATCT
2713          --                  C        A

2593  CAATTCTCTTTCTATCTATGCAGTGATTTTCAGTCTCTGAGGACTCCAAAGGATACTTACATTCCC-GGTTTGGTGGAAA
2791                              G                                  T

2672  TCCTGGGCATCTGAGTTGGAAGAGTGAGGACAGGGGAGGAGCTGGGGACATTGGGACTGTTGGAACGTCTTGGAAACAAT
2871                                        T          A

2752  GACCCACTCAGTGTCTGAATTCATTTTCTGTCATAACTGCCCTGAAAAAGTCCAGTGTGTTTAGAGGCGTGTTTTGGG-A
2951                                                                          G

2831  TGAGGAAGGGTGGGAACTGGTTGAACTGGATTTGGAATCAGAGTCTAGGCCCTATTGTTCTGCATACCTGCCCCATAGCA
3031            T                                        C

2911  CTGCAGTGAGGCAGCTGCAGGGCTTGATGTATTTCCCCATTCTGCTTGCTGAATTGAGGCAAAGAAAGACAGTGACCAGC
3111                        C          CT
                .          .          .          .          .          .          .
2991  ACATATGTGTGTTTGTGTTTTTGTAAAAACACCCACATGCTCATGAGGCCAAGAGTGGGTTGTGAGGACAGATGGGTGGC
3191                              G                    T

3071  TGAGCAGGGAGGCAGGCAGAGGGACAGAGGGAATGTTCTTCTGGAATATCCTCAGGCTCATTGTGTTCTGCAGAAGGCCA
3271              T                                  A

3151  GCAGCACTGCATTATTCAACTCTTCTTGCTGGAATGCAGATTAGAAACTAAGAATCTTGCCTTCCCACTCATTCCCTCTT
3351

3231  TGAGACCATTGAGCTACATTTCTCCTTCTACCTGGACCCCCCTTATCCTTAAATTGACCATCAGAACATTTGCACCCAGA
3431                  G                          -

3311  CTAAGAGCCAGAGTTTCTGACACCTGGCCATAGGCCTGGGCCACCTGAGGCTACCTTTGCAGGTGGACCCCGAGCCATAC
3510                  C                                    G                        T
                                                                  Exon 25 → V  D  P  E  P  Y   1080

3391  CTGGATGTCTGCATTTACGACACCTGCTCCTGTGAGTCCATTGGGGACTGCGCCTGCTTCTGCGACACCATTGCTGCCTA
3590
      L  D  V  C  I  Y  D  T  C  S  C  E  S  I  G  D  C  A  C  F  C  D  T  I  A  A  Y  1107

3471  TGCCCACGTGTGTGCCCAGCATGGCAAGGTGGTGACCTGGAGGACGGCCACATTGTGCCATGAGTCCTGACACCCTCATG
3670                                                                  G    A    G
      A  H  V  C  A  Q  H  G  K  V  V  T  W  R  T  A  T  L  C                        1126

3551  TTCTCAGATGCCCTCCCTTCTTCCCATGTGTCTATGCTTGAAGACCTTGTGAGTGCAGGGGGATATCTTCATGGGCGAGA
3750

3631  GAAGACAG--TGCTAGTATGGGACTGGATGGCCAAGGGCTAACGAGGGTTAAGACGTTGGCTGAGTAAGAAGTTTATATT
3830          AG  T     -                          G            A       T

3709  ATGGACTTTGAGGCGGGACATGGATTCAAGGCATGAACATGTGGATACTTTTCTTCTGGAGAGATTCTGGCAGAGGAGAA
3909   C  G---     T                                G                          G
                .          .          .          .          .          .          .
3789  GAGGGAATACTGATGAAAAGAAGGAAGTTGATTTATGTCTTTAATTAGGCGTGATTATATTTGCCAATATGAGTGATCAA
3986                              C

3869  CTCATACATTCATGTCTATAGAATCTTCCTTCTTTGGACATAAGCAACTTAATGCTTTTATGTAGAAAACTGGGCCAGGC
4066                        G          G          T                    G

3949  ACAGTGACTCACACCTGTGATACCTGTGTTTTGGGAGGCTGAGGCCAGAGGATTCCTTGAGCCCAGGAGTTCAAGACCAG
4146              T      A  C

4029  CCTGGGCAACATAGCAAGACCCCATCTCTACAAAAATTAAAAAAAGAAATGAATTCAGAACCAATAGATTCTGGTTTAGG
4226                        G

4109  TGCTTCAACAATCCAGAAGTCTCTAATATTGGTGACGCCCATAGTCCCTTAGTTCCCCAACATTATCTCCAGATGCTGCA
4306                                                  C                          GC

4189  GGCCATCACCACATGGGTCTGCAGTCCTGGAGGCTTTGCCTTTTGTGGCCACAGCCTTGTCTCCTGTCTACACAGCCCAG
4386                                          G
                                                                        Exon 26 → P  Q   1128

4269  AGCTGCGAGGAGAGGAATCTCCGGGAGAATGGGTATGAGTGTGAGTGGCAATAGAACAGCTGTGCACCTGCCTGTCAAGT
4466                              C                    GC  T
      S  C  E  E  R  N  L  R  E  N  G  Y  E  C  E  W  Q  *  N  S  C  A  P  A  C  Q  V  1155
                                                            Y
```

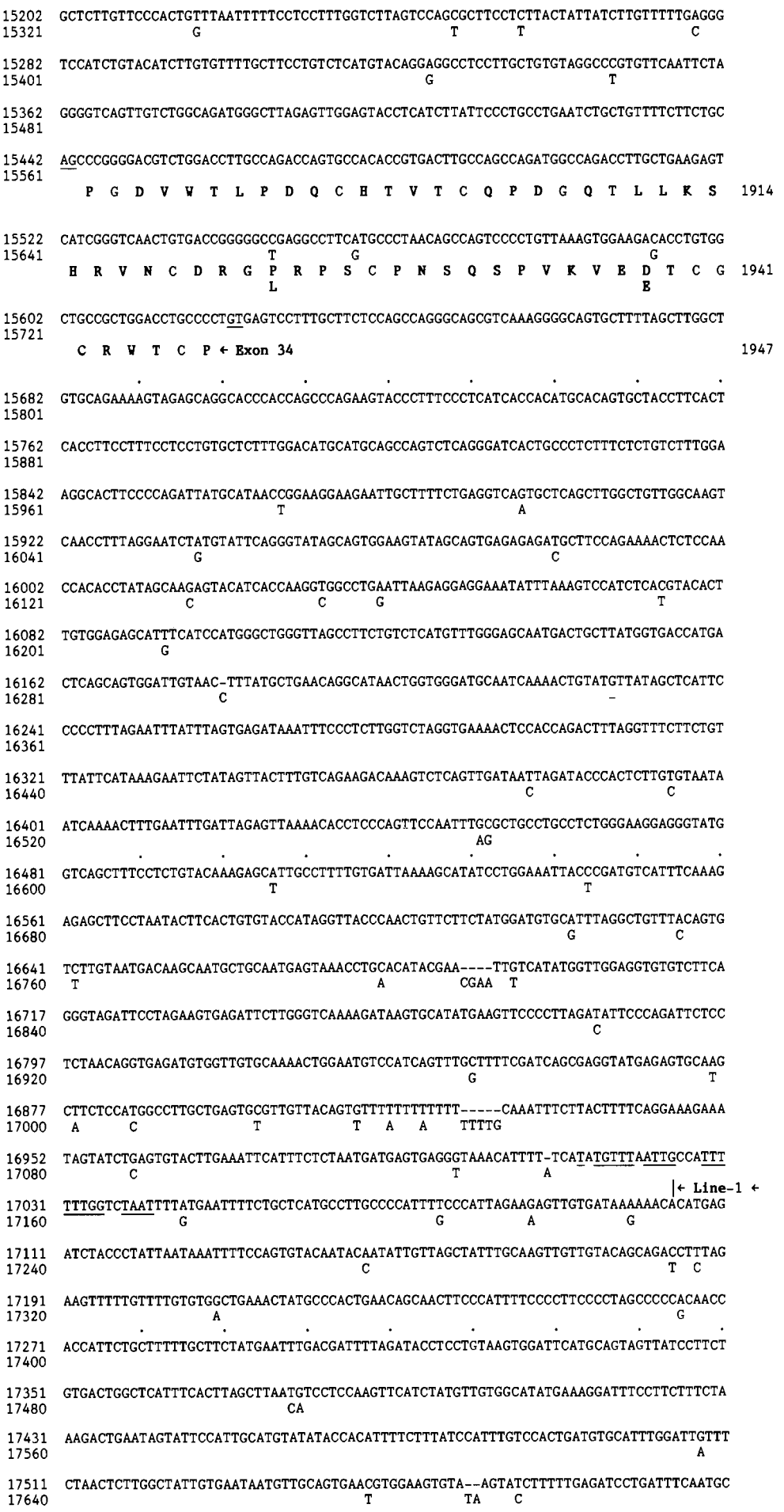
```
4349  TACGTGCCAGCACCCTGAGCCGCTGGCCTGCCCTGTGCAGTGTGTGGAGGTCTGCCATGCCCACTGCCCTCCAGGTGAGG
4546  C    T           A                       G
        T  C  Q  H  P  E  P  L  A  C  P  V  Q  C  V  E  V  C  H  A  H  C  P  P          1179
                                                G

4429  CCTCTATCCCTGGGGGTCGGGCTGGTGGGATGGGATAGGGATGGATGGAAAGGTGCTTCTAGGTCTTGCTTCATCTCAGC
4626            A

4509  CTCCACCTGCCACATCCCATCTCTGACCTGCAAGCCTGCCACCCCATCTCTGACCTGCAAGCCTGCCACCCCATCTCTGA
4706        G  T  ----------------------------------------------------------------------
          .     .     .     .     .     .     .     .     .     .     .     .     .
4589  CCTGCAAGCCTGCCGCCCCATCTCTGACCTGCAAGCCTGCCGCCCCATCTCTGACCTGCAAGCCTGCCGCCCCATCTCTG
      ----------------------------------------------------------------------------------

4669  ACCTGCAAGCCTGCTGCCCCATCTCTGACCTGCAAGCCTGCCGCAGGTTCCATGGGTTCTTTCATCAGAGTCAGGCCAGT
4740  ----------------------------------G  T     G                          A

4749  TGTGATTTTTCTCAAGTCGAGCTCCTCCAAAATGCTTTTCTGTGCCTATTTGTGGCATTCTCATCTAAAGCAGCCCCTGC
4784  C                                                 A  G     C

4829  CGATAGAACTTTCTGCAGTGGGGGAATGTTGTATTGAATGTAACTGTGACGAGTGTGTCTGAGGAACTGAATTTTTGAAT
4864                                                                          G

4909  TTTATTGAATTTTACTTAATTTAACTTAAATAGCCACATGCAGTTTATGACTATCCTATGGTACAGCACAGCTCGAGAGT
4944                          G           C     G  G                       T

4989  TACTGAAGGTGTTATTCAGAAGAGCAGAAAGAGCCCTGGAGATAAGATTTCATTTGTCCTGAGGCTTGGGGAGGTGAGGT
5024                                C

5069  AGGATGAAGGAATCCCCGCTCCCAGTTTTGCAGAGGGATCAATCAAGGCACAAGGAGGAGAGATGCTCCTTGAGTGGCAG
5104    G                                                 C                    ATG

5149  GGTGACCCCTAGGGGTGCAGGCAGGAGGAGTTGGCTTCTAGGGCAGGAGGAGGAGTTGGCTCCTCCCTTTTAGTTAAAAA
5184          G  A

5229  TGAGACTTCCTTGTGGGAAAGGGGAGCGTTTTGGTTCCTAATGAGAGCTTTCTTTTGCAGGGAAAATCCTGGATGAGCTT
5264     G     C
                                                Exon 27 → G  K  I  L  D  E  L   1186

5309  TTGCAGACCTGCGTTGACCCTGAAGACTGCCCGGTGTGTGAGGTGGCTGGCTGGCGTTTTGCCTCAAGAAAGAAAGTCAC
5344                        T  A                 C                 G
      L  Q  T  C  V  D  P  E  D  C  P  V  C  E  V  A  G  W  R  F  A  S  R  K  K  V  T  1213
                                                   R              G
          .     .     .     .     .     .     .     .     .     .     .     .     .
5389  CTTGAATCCCAGTGACCCTGAGCACTGCCAGATTTGGTAAAACAGATGCCTGGATTGTTTGAAGTGACGAATCTTATTGC
5424                                                T     G              T
        L  N  P  S  D  P  E  H  C  Q  I  C                                          1225

5469  TTCTCCAGGGGCCAGCTCAGAGACTAGGTTTTGGGTAAAAAGTCCTAGCCACATGAGTTGAGAGACTGAGTTTATTTTAT
5504                  T

5549  TTGTTTATTTATTTAATTCATTATTTAGTGTAGTTTATTGTTTTGTTTGTTCCTAGGGTTTGCTTCTTTTTCTTTAGGGA
5584                        A

5629  TGGACTAGACTGCTTCCTCCACATATATTCCTCTACAATGTTTTTGCTCTTAATAGGTTTTTAAGTCCTAACAGGTTTTT
5664                                    C  G

5709  GCTTTTCTCACTGAAGGTGGGGGTATGGTCACTTATTGTCCTTGTAGGTCATGACTAGTCTGAGCACAGGTGTCCATATA
5744

5789  CTGGAGGTGGCGACTCAGGAGAGGAAAGTGAGATGGAACGCGGGGCCCTCAGGGTAACATGAACGTCGTGGGGGCCATAG
5824            G                       C A         -

5869  TCAAGGTGGCAGGGTGTGGCTGTCATAATGGGGAGGGAAGGTGCTCGTTGCTCTCACCCTGGGTTCTCACTGGACTTGCC
5903                                                          C

5949  TCTGGGAAACTGGCCTTTGGGGAGGTGAGAAGGAGTCCTCAGGGTGAGCTCTGAGCCCTCTGAAGTCCTGCTCATAAGTA
5983

6029  GGGAACATTTATTAAAATCCACATCTATTCAACACCACTCCCTGACACTGATCAGGCCTTCTGGGAGCTTAAACTTGAGA
6063     G                  G                          A    A

6109  ATGTAAATTCGCTGGGGAGGGGCTATGGTCACAGCTGCCAGAGTCAGATCCTGCCACCACCATCTGCACAAACAAGAACG
6143  C                                    T                 G          T           T
          .     .     .     .     .     .     .     .     .     .     .     .     .
6189  AACACTGCCTGGGGTGAGTTTTCTAGCAAACACAGGAAACTAGAAAGTATGTCACTCACGCACCGTTAACCAAGGATTGA
6223                                                            T    T

6269  AGCAATAAATCTCATGGATCATGGGAAGGAGGAATTAACTTCCTAGTAAGACTCTACCTGCGCACACTCTTTATGCTGAA
6303               TG       C

6349  ACAGGGACTGTAGGGAGCAAAGGATTGGGTGATTCTGATGCTGGCTTGGGGTTCTGTGCCAAAAGGGCATGAGAGCTTCG
6383     A                   C

6429  AGCCCTCTCCCCACCTCCTTTGCTGGCACACGGGCAGGGGCAGCTGCCACGGTGCACGCTGAAGCTGCAAGCAGTGTGTG
6463

6509  GACAGAAGTCCCACAGCAGGCTGGGGGCATCACCTGGTATCAAGCCACAGGGGGACACTTCCCAGGCTCAGGGCACTGCC
6543                          T                   G                       G

6589  GCCTTGTCAGCTTTACTGGGAGTGCTGAAGGCCAGGTGCCTCCAAGGCTGAGAGTCCAAAGAGGAGATGGTCCTGGCCCA
6623
```

```
6669  TATTGGTCTCCACCACATGGCCACTTGATTTCACCTGTGGGCTGGCAGGCCGTGCCACTTTCCGGGACACTGGCAAAAAG
6703                  C T

6749  GTGATGTTACCAGCAGAGCCTCGAAATTTGGCATAGTAGTGTAGGAATTTGTGTGTGTGGGGCCACTGCTCTACTACAGA
6783                        G           G                    G

6829  CTTGTGGTGTGTGCTTGGGGCTGGGCACAGAGCCTGGCCAATACCTGGCACATGAGAGGTGCTCACCCTATGTCTGCTGG
6863                                      G

6909  AGGAATGAGTGAATGTCTTTTCTTAGGCCTAACACCTTGGGGACCCCTGTCTGCTACTCCCAAATAATTTGTTCTGCTAG
6943                              G     G   A
                .         .         .         .         .         .         .         .
6989  ATAAAGCTGGGGTTTTGAATCTTTATTCATCTTTCCACAATGTTTTTGCTCCAATTTTTTCTCTCTTCGTTAGGGCCCAA
7023                                    T

7069  CGTATTAAACATGTTTTAAAACTCATGCAATTTACATTCTGTTAGAATATGTACTGTAAAGCATTTATCTTCAACAGAAC
7103   A   C

7149  ATGTACAAGCACAGGTTAGCGTCTACATTTTAGTAACATTTGTTCTTGGAGACACTTGTAAGAAGGCTTAGATTATAATT
7183          A                                                             G

7229  AAGTATTCATGTTGCTAAAGATGTAGGTTGAAAAGAAATCAGCTTTGAGTGGCTGTGTGGATGGCCTAGAACAACGAGTA
7263                                                  T

7309  TATTTTTTCATATCCTAGGATGGGCTAAAGCTTTGGTTTCTAGAATCAAAGGAAGTCGGCTATGTGTGTGTTTTGAAGTG
7343                                                                              GT

7389  GGGGGGCATGCTATTTGGGGACAGATGTTAAACAGTGACATCTCACTTGGGTGTGGAATGGTCCGTGGGATCTCAAGTTC
7423                                    A               A             A

7469  AGGTGGAACAGAGGAGATTCTGTGGGAATATGGAAGTAGTCATACACTGTAGGGCTCAGAAGTGTCCATAGGTTCTTTCT
7503                                    CA TG                           C       C

7549  GAACCATTTTAATTTCTTCGCTCTTTTCTGCAGCCACTGTGATGGTGTCACCCTCACCTGTGAAGCCTGCCAGGAGCCGG
7583                                                T     A
                                Exon 28 → H  C  D  G  V  T  L  T  C  E  A  C  Q  E  P   1240
                                                V     N

7629  GAGGCCTGGTGGTACCTCCCACAGATGCCCCGGTGAGCCCCACCACTCTGTATGTGGAGGACATCTCAGAACCGCTGTTG
7663            G                                                   G     C
      G  G  L  V  V  P  P  T  D  A  P  V  S  P  T  T  L  Y  V  E  D  I  S  E  P  L  L   1267
                                                                              P

7709  CACGATTTCTACTGCAGCAGGCTACTGGACCTGATCTTCCTGCTGGATGGCTCCTCCAGGCTGTCCGAGGCTGAGTTTGA
7743                              G
      H  D  F  Y  C  S  R  L  L  D  L  I  F  L  L  D  G  S  S  R  L  S  E  A  E  F  E   1294
                                    V
                .         .         .         .         .         .         .         .
7789  AGTGCTGAAGGCCTTTGTGGTGGACATGATGGAGCGGCTGCGCATCTCCTAGAAGTGGGTCCGCGTGGCTGTGGTGGAGT
7823                                                C                   C
        V  L  K  A  F  V  V  D  M  M  E  R  L  R  I  S  *  K  W  V  R  V  A  V  V  E   1320
                                                      Q

7869  ACCACGACGGCTCCCACGCCTACATCGGGCTCAAGGACCGGAAGCGACCGTCAGAGCTGCGGCGCGTTGCCAGCCAGGTG
7903                                                                        A
      Y  H  D  G  S  H  A  Y  I  G  L  K  D  R  K  R  P  S  E  L  R  R  V  A  S  Q  V   1347
                                                                        I

7949  AAGTATGCGGGCAGCCAGGTGGCCTCCACCAGCGAGGCCTTGAAATACACACTGTTCCAAATCATCAGCAAGATCGACCG
7983                                    T                       T
      K  Y  A  G  S  Q  V  A  S  T  S  E  A  L  K  Y  T  L  F  Q  I  I  S  K  I  D  R   1374
                                          V                       F

8029  CCCTGAAGCCTTCCGCATCGCCCTGCTCCTGATGGCCAGCCAGGAGCCCCAACGGATGTCCCGGAACTTTGTCCGCTACG
8063            C
        P  E  A  F  R  I  A  L  L  L  M  A  S  Q  E  P  Q  R  M  S  R  N  F  V  R  Y   1400
                  S

8109  TCCAGGGCCTGAAGAAGAAGAAGGTCATCGTGATCCCGGTGGGCATTGGGCCCCATGCCAACCTCAAGCAGATCCGCCTC
8143                                    T
      V  Q  G  L  K  K  K  K  V  I  V  I  P  V  G  I  G  P  H  A  N  L  K  Q  I  R  L   1427

8189  ATTGAGAAGCAGGCCCCTGAGAACAAGGCCTTCGTGCTGAGCGGTGTGGATGAGCTGGAGCAGCAAAGGGAAGAGATCGT
8223    C                                       A                       C
      I  E  K  Q  A  P  E  N  K  A  F  V  L  S  G  V  D  E  L  E  Q  Q  R  E  E  I  V   1454
                                                S                       D

8269  TAGCTACCTCTGTGACCTTGCCCCTGACGCCCCTCCTCCTACTCTGCCCCCCCACATGGCACAAGTCACTGTGGGCCCGG
8303                                    A
        S  Y  L  C  D  L  A  P  D  A  P  P  P  T  L  P  P  H  M  A  Q  V  T  V  G  P   1480
                                  E

8349  GGCTCTTGGGGCTTTCGACCCTGGGGCCCAAGAGGAACTCCATGGTTCTGGATGTGGCATTTGTCCTGGAAGGATCGGAC
8383            G                                                   G  C
      G  L  L  G  L  S  T  L  G  P  K  R  N  S  M  V  L  D  V  A  F  V  L  E  G  S  D   1507
                  V

8429  AAAATTGGTGAAGCCAACTTCAACAGGAGCAAGGAGTTCATGGAGGAGGTGATTCAGCGGATGGATGTGGGCCAGGACAG
8463                  G
      K  I  G  E  A  N  F  N  R  S  K  E  F  M  E  E  V  I  Q  R  M  D  V  G  Q  D  S   1534
                        D

8509  CATCCACGTCACGGTGCTGCAGTACTCCTACATGGTGACCGTGGAGTACCCCTTCAGCGAGGCACAGTCCAAAGGGGACA
8543
        I  H  V  T  V  L  Q  Y  S  Y  M  V  T  V  E  Y  P  F  S  E  A  Q  S  K  G  D   1560
```



```
10591  TAGGGCCTCATCTCACTCAGGGGTCAGTGCTCCAGTATGGAAGCATCACCACCATTGACGTGCCATGGAACGTGGCCCCG
10699         G          T        G                                    T
         G  P  H  L  T  Q  G  S  V  L  Q  Y  G  S  I  T  T  I  D  V  P  W  N  V  A  P   1749
               R           V                                                   V

10671  GAGAAAGCCCATTTGCTGAGCCTTGTGGACGTCATGCAGCGGGAGGGAGGCCCCAGCCAAATCGGTAACGTTGGTGCCAC
10779
       E  K  A  H  L  L  S  L  V  D  V  M  Q  R  E  G  G  P  S  Q  I ← Exon 30          1770

10751  AGGCTGGATGCAGAAGCTGCATTCTGGTTCTTATTTTTGGCATAAGTGACTGTGTGACCTTGGCCAGTCACTTTGCTCCT
10859                                                                  C

10831  TGGCCTTAGTTTCTTCCCCTGGAAAGTGAGGGGCTAGATGCTCTTCCACGTCTCTCCAGATCTCAACTGGGTGTTCCTTG
10939            T
               .        .        .        .        .        .        .        .
10911  GAGTCTCTGAATCATTCAGCTTTTAAGTGACTTAAGGATCCACCGT-GAGACAGGGTGTCCAGCCGCAGTCAGTACTGAC
11019     T                                          T              G

10990  TTGGTGTGATCTGTTCTCCATCCTCAGGGGATGCCTTGGGCTTTGCTGTGCGACACTTGACTTCAGAAATGCATGGTGCC
11099     C                                                  T
                       Exon 31 → G  D  A  L  G  F  A  V  R  H  L  T  S  E  M  H  G  A   1788
                                                             Y

11070  AGGCTGGGAGCCTCGAAGGCGGTGGTCATCCTGGTCACGGATGTCTCTGTGGATTCAGTGGATGCAGCAGCTGATGCCGC
11179     C        A                    C
       R  L  G  A  S  K  A  V  V  I  L  V  T  D  V  S  V  D  S  V  D  A  A  A  D  A  A  1815
          P

11150  CAGGTCCAACAGTAAGAATCTGGTGTACAGTCCTCAATTCAGGAGAGCGAGACTTCAGAAATGATAGAGGATTAAAAGTG
11259
         R  S  N                                                                        1818

11230  GCCTGGGGTTACTTTTGGATGTTGTGGCTGATATTAAATATGCTCAGTACCAACCTCTGTCCTGGTTTCAAACAGCAGGA
11339   G                                          A

11310  AATTATCCAATCACGATGTGATTAGGAAACCAAACAATAATAAGTGTGTGTTGATCCTATATCTTTATAACTAATTATAA
11419            C                        A

11390  CCCTTTAATTTCTCCGCATTTTCCCTTTGTGTCCAGGAGTGATCATACCTCTATGCCCTTTCTTTCTAAAGTAGGATTCG
11499  -T              A

11470  TGGTCTCTTACTAAAATGTCATGTCTGAGGCCAGAGTGATTCATTGTGGGGCCTTGAGAACCAATGGGGACTTTCATGTA
11578                                              A                    G

11550  TTTAGAGTCCTTTCCTGGTACCAGGTAATGAGCTTGGCAGGGTTGATGGATGAGAGACAAAGGATCATGCTGTAGGAAAG
11658                             G

11630  GGAAGGATCCAGATGTCCAGCCTTGTGCTCCATAAGAACATGTGCTTATCATGTTCTTAATATCAACAGGATGTGAGCTG
11738
               .        .        .        .        .        .        .        .
11710  GGACAGGTGGTAGAGTCTGAGGGACCACCTCCCAGCTGAAGGAGGTCTGGGTGAGTGGAGTGACATTGAGTCCCCTCGTG
11818                    -

11790  GCCACCTGGAGAACTGTTCATCTTTCAGAACCCTGCCTAGGAGTGCTGTCAGAACCCTGTCTTTTCTCTGCTGTCTTCCT
11897

11870  GATGTTCTTCCTGAGGAAGTTATTTGCTCCGTTCTCTATGTCACATATGAGCTACTAAAAATGTGCAAATC-----GTAT
11977                                                    C                GTATT

11945  TGTATTGTATTTGTAACATATCTATTCCTTGCAAAAAACTGGGAGTTCTTTGAAGGCAAGGGAAGATCTTGCTTATTTTT
12057

12025  GTATCATCTGTGTATGATGCACAAAATAGGTGACCATTTAATGTTTGTTGAGTGAATGAATGTGACATGTTTATGAAATC
12137                                        C

12105  TATAATTTTAGGGTAAAGGACACTAAAAAGGAAAGGATATATATTTGGATGATATAACGAACTTACAAGGGCTGAAAAGC
12217     G                        G                    A                    A

12185  TGCTAGTAGAAAATCCATTCCAACCTGAAGCACATTGCAAATTCCTAGAACTTTCTGGAGAGCTGAATTGAGGAATTTCA
12297       A          G

12265  GGATGGAGATTTTTTCTCCTTTAGTATCAAGAATCCAAAGAAAGATTTTCACTGGGAAAGGAAGCAGTTGTGGTCTTGAG
12377                       C                                        A                -

12345  GAAGCAGATATGATCCAGAGCACGTCATTTAACCATCTAAAGCCTCGGTTTCTCATGTGTAAATGAGCATGCACATTCTC
12456  --                     A           G  C

12425  TGGGCTCAGCCCTGTGATATTTTCTCTTCTCTTCTTACATATACCCTGTTGGTGGTCTAGTTCAGTGTCATGGGTCTGAA
12534                                              CC                                 A
               .        .        .        .        .        .        .        .
12505  TATATTTATACATCTCTGACTTTGAAATTCGTATCTCCAGCCTCACCTCAACTCAGCTTGCATGCTCTAACAGGCAAAAC
12614    C           T                                          T            T

12585  TAAACTCCTGATTTTCACATCCGTCTTTCTTATCTTAGAAAATGGCGATTCCAACCTTCCAGTTGTTCAGTTCAAAAAGC
12694                       T                    A

12665  CCAGCATCCTTGATTCTTCTTCCTTTTTCCTGATACCTCCATGCAAGCCACGAACAAACTATAGCTCCATCTTCAAAATA
12774  T                          C

12745  CTTTGAGAGTTGCACCTCTCCACACCATGTCCACCAGGGTCACCTCAATCTGAGTCCCCAGCATCTCTCACCTGGACTAT
12854     C      CA      C                T                C         T

12825  TGAAATAGCTGTCTAACTGGAATCCCAGCTTCCACCTTTGTCCCCTGCAGCTCATTCTCAGCACAGCAGCTGCAGTAGCC
12934                       C                              A                        G
```


Exon 32 → R V T V F P I 1825
G I G D R Y D A A Q L R I L A G P A G D S N M V K L Q 1852
Q I E D L P T M V T L G N S F L H K L C S 1873
R
Exon 33 → R F V R I 1878
R M D E D G N E K R 1888
C


15202 GCTCTTGTTCCCACTGTTTAATTTTTCCTCCTTTGGTCTTAGTCCAGCGCTTCCTCTTACTATTATCTTGTTTTTGAGGG
15321 G T T C
15282 TCCATCTGTACATCTTGTGTTTTGCTTCCTGTCTCATGTACAGGAGGCCTCCTTGCTGTGTAGGCCCGTGTTCAATTCTA
15401 G T
15362 GGGGTCAGTTGTCTGGCAGATGGGCTTAGAGTTGGAGTACCTCATCTTATTCCCTGCCTGAATCTGCTGTTTTCTTCTGC
15481
15442 AGCCCGGGGACGTCTGGACCTTGCCAGACCAGTGCCACACCGTGACTTGCCAGCCAGATGGCCAGACCTTGCTGAAGAGT
15561
P G D V W T L P D Q C H T V T C Q P D G Q T L L K S 1914
15522 CATCGGGTCAACTGTGACCGGGGGCCGAGGCCTTCATGCCCTAACAGCCAGTCCCCTGTTAAAGTGGAAGACACCTGTGG
15641 T G G
H R V N C D R G P R P S C P N S Q S P V K V E D T C G 1941
L E
15602 CTGCCGCTGGACCTGCCCCTGTGAGTCCTTTGCTTCTCCAGCCAGGGCAGCGTCAAAGGGGCAGTGCTTTTAGCTTGGCT
15721
C R W T C P ← Exon 34 1947
15682 GTGCAGAAAAGTAGAGCAGGCACCCACCAGCCCAGAAGTACCCTTTCCCTCATCACCACATGCACAGTGCTACCTTCACT
15801
15762 CACCTTCCTTTCCTCCTGTGCTCTTTGGACATGCATGCAGCCAGTCTCAGGGATCACTGCCCTCTTTCTCTGTCTTTGGA
15881
15842 AGGCACTTCCCCAGATTATGCATAACCGGAAGGAAGAATTGCTTTTCTGAGGTCAGTGCTCAGCTTGGCTGTTGGCAAGT
15961 T A
15922 CAACCTTTAGGAATCTATGTATTCAGGGTATAGCAGTGGAAGTATAGCAGTGAGAGAGATGCTTCCAGAAAACTCTCCAA
16041 G C
16002 CCACACCTATAGCAAGAGTACATCACCAAGGTGGCCTGAATTAAGAGGAGGAAATATTTAAAGTCCATCTCACGTACACT
16121 C C G T
16082 TGTGGAGAGCATTTCATCCATGGGCTGGGTTAGCCTTCTGTCTCATGTTTGGGAGCAATGACTGCTTATGGTGACCATGA
16201 G
16162 CTCAGCAGTGGATTGTAAC-TTTATGCTGAACAGGCATAACTGGTGGGATGCAATCAAAACTGTATGTTATAGCTCATTC
16281 C -
16241 CCCCTTTAGAATTTATTTAGTGAGATAAATTTCCCTCTTGGTCTAGGTGAAAACTCCACCAGACTTTAGGTTTCTTCTGT
16361
16321 TTATTCATAAAGAATTCTATAGTTACTTTGTCAGAAGACAAAGTCTCAGTTGATAATTAGATACCCACTCTTGTGTAATA
16440 C C
16401 ATCAAAACTTTGAATTTGATTAGAGTTAAAACACCTCCCAGTTCCAATTTGCGCTGCCTGCCTCTGGGAAGGAGGGTATG
16520 AG
16481 GTCAGCTTTCCTCTGTACAAAGAGCATTGCCTTTTGTGATTAAAAGCATATCCTGGAAATTACCCGATGTCATTTCAAAG
16600 T T
16561 AGAGCTTCCTAATACTTCACTGTGTACCATAGGTTACCCAACTGTTCTTCTATGGATGTGCATTTAGGCTGTTTACAGTG
16680 G C
16641 TCTTGTAATGACAAGCAATGCTGCAATGAGTAAACCTGCACATACGAA----TTGTCATATGGTTGGAGGTGTGTCTTCA
16760 T A CGAA T
16717 GGGTAGATTCCTAGAAGTGAGATTCTTGGGTCAAAAGATAAGTGCATATGAAGTTCCCCTTAGATATTCCCAGATTCTCC
16840 C
16797 TCTAACAGGTGAGATGTGGTTGTGCAAAACTGGAATGTCCATCAGTTTGCTTTTCGATCAGCGAGGTATGAGAGTGCAAG
16920 G T
16877 CTTCTCCATGGCCTTGCTGAGTGCGTTGTTACAGTGTTTTTTTTTTTT-----CAAATTTCTTACTTTTCAGGAAAGAAA
17000 A C T T A A TTTTG
16952 TAGTATCTGAGTGTACTTGAAATTCATTTCTCTAATGATGAGTGAGGGTAAACATTTT-TCATATGTTTAATTGCCATTT
17080 C T A
|← Line-1 ←
17031 TTTGGTCTAATTTTATGAATTTTCTGCTCATGCCTTGCCCCATTTTCCCATTAGAAGAGTTGTGATAAAAAACACATGAG
17160 G G A G
17111 ATCTACCCTATTAATAAATTTTCCAGTGTACAATACAATATTGTTAGCTATTTGCAAGTTGTTGTACAGCAGACCTTTAG
17240 C T C
17191 AAGTTTTTGTTTTGTGTGGCTGAAACTATGCCCACTGAACAGCAACTTCCCATTTTCCCCTTCCCCTAGCCCCCACAACC
17320 A G
17271 ACCATTCTGCTTTTTGCTTCTATGAATTTGACGATTTTAGATACCTCCTGTAAGTGGATTCATGCAGTAGTTATCCTTCT
17400
17351 GTGACTGGCTCATTTCACTTAGCTTAATGTCCTCCAAGTTCATCTATGTTGTGGCATATGAAAGGATTTCCTTCTTTCTA
17480 CA
17431 AAGACTGAATAGTATTCCATTGCATGTATATACCACATTTTCTTTATCCATTTGTCCACTGATGTGCATTTGGATTGTTT
17560 A
17511 CTAACTCTTGGCTATTGTGAATAATGTTGCAGTGAACGTGGAAGTGTA--AGTATCTTTTTGAGATCCTGATTTCAATGC
17640 T TA C

```
17589  TTTTGGATGAATACCCAGAAGTGGGATTGCGGGGTCATATGCTAATACCAGTTTTAATTTTCTGAGGAACCTTCTTACCG
17720                              A T                T           T               T

17669  TTTTCTGTAGTGGCTACACCACTTTACATTCCTACCATCAATGCACAAGGGTTTCAGTTTCTTCACATTCTTGTTAACAC
17800            C  G                                          A

17749  TTGGTATTTCCTG----GTGTGTGTGTGTGTGTGTGTGTGTGTTTGATAATGGCCATTCTAGCACGTATGAGGTAATA
17880  G            GTGT                                                          G

17825  CCTCATTATGGTTTTGATTTGCATTTCCCTGATGATCAGTAATGTTGGATATTTTTTCATATAATTGACTTTTTAGATGT
17960                                                                 G     G

17905  CTCTTTGGAGGAATGTCTTTTCAAGTCCTTTGCTCATTTTTAAATTGGGTTATTTGCATTTTTTTTTTTTTTGCTATTGA
18040

17985  GTTGTGTGAGTTCCTAATATATTTTGGATATTAACCTTTTATCAAATGTACGGTTTGCAAATATTTTCTTCCATTCTGTA
18120       A                          C
                .         .         .         .         .         .         .         .
18065  GGCTGCCTTTTTATTCTGTTGATTGTTTCTTTGGCTGCACAGAAGCTTTTTGGTTTGATGTAGTCCCACTTGTCTATTTT
18200

18145  TGCTTTTGTTGCCTTGTTTTTGGTAACGATATCCAAGAAATCGTCGCCAAGACCAATGTTGCAAAGTTCTTGTGTTTTCT
18280                        C  T

18225  TTTAGGAGTTTTATAATTTCAGGTTTTACGTATCAGTGTTTAAGTTTGATTTTGGGTTAACTTTGTGTATGATATAAGAT
18360                                                                           G

18305  AAGGGTCCAACTTCATTCTTTTTCATGTGGACAGGT-CATTGTTAGTGTATAAAAACAAATGGTTTTCTATGTTGATTTT
18440           T                          T              TG             G

18384  GTATTCTGCAACTGAATTTGTTTCAACTTTGCTGAATTTGCTAACAAATTTTTGGTGAAATCTTTAGAGTTTTCTACACA
18520                         T     C                                           TG

18464  TAAGATTATGTCATCTGCAAATTGAGATAATTTTACTTCTTTTTTGATTTGGGTGATGTTTATATCTTCTTGATTCATTG
18600    T   C               G                                  C

18544  CTCTAGCTAAGACTTCTGGTACTATGTTGAATAGAAGTGGTGAGAGTGGGCATCCTTGCTGTCTTCCTGATTAGAGGAAA
18680                  C                                        C

18624  AGCTTTTAGTTTTGCACTATTCAG-GTGATGTTAGCTGTGGTATTTTCATTAATGGCCTTAATTTTGTTGAGGTAATTTC
18760                       G T                                  T                 T

18703  CTTCTATTCTTAGTTTGTTGAATGTTTTAATCATGAAATGGTGTTTGAATTTTGACAAATGCTTTATCTGCATCTATTGA
18840                                                    A                 T    G

18783  CATGACAGTATGATTTTTATCCTTTGTCCTGTTAATGCTGTGTATCACATTTACTGATTTTTAGATGCTGAGCCATTCTC
18920       T                    T          A        C    T          C T
                .         .         .         .         .         .         .         .
18863  CCATCTGAGGGATAAATCCCACTC----AGAGTGTATGATCCTTTGATTGTGCTGTTGAATTCAGTTTGCTATCATTTTG
19000   T                      CGTC        T                   T             G

18939  TTGAGAATTTTTGCATCTATGTTCATCAGGAATATTGACTTGTAGTTTTCTTT-AACGTAAAGTCTTTGGCTTTGGTATC
19080                    G                          T   T  T           A

19018  AGGGTAATTCTGGCTTC--------AGGTTTCCCACTTCAATGTTTTGGAAGAGTTTGAGAAAGACTGGCATTAATTCTTC
19160                  ATAAAGT               G             G         T

19091  TTTAAATCTTTGGTAGAATTATCCAGTGAAAGCATCTGAGCTCTTCTTTGTTGGGAGATTATGATTATTAATTTGATCTC
19240                                            A   A   A             C      C

19171  CTTACTAATTGTGGGGCTGTTCAGATTTTCTGTTTATTCATCATTCAGTCTTGGTAGGTTGTATGTTTCAGGGATTTATT
19320           GA                                    G

19251  CATTTCTTTAGGTTATCCAATTTGTTAGTGTGTAGTTGTTCATAGTAGTCTCTTACTAGCCTTTTTATTTCTGTGCTATC
19400                                                                 A              C

19331  AATTGTAATATAACCTCTTCTATTTCTGATTTTATTTATTTGAGTCTT---TTTTTGTCTTGGTTAATCTAGCTAAAGAT
19480           G   G                                    CTC                          C

19408  TTGTTAATTTTGTTGACCTTTAAAAAACCAATTCTAAGTTTTATTGATATTTTGTATTCTATATTTCATTTATTTTTCCT
19560

19488  CAAATCTTA---ATTTCCTTCCTTCTGCTAACTTAGGGCTTAGTTTATTCATTTTCTAGTTCCTTATGTTGTAAAATTAG
19640          TATT                          C                              -   CA

19565  GTT-----GTTTGAGATCTTTCTTACTGTAGACATTTACTGCTGTAAACTTTCCTTTTATTGCTTTTGCTGTTTCCCATAA
19719     ATTT                                  A
                .         .         .         .         .         .         .         .
19641  GCTTTGCTATGTTGTGCTTTAGTTTTTGTTTTTCTCAAAATATTTTCTGATTTTTCTTTTGATTTCTTCTTTGACCCATT
19799                                                                           T

19721  GGTTGTTCAAGAGTGTTTTTTAATTTCCACATATTTGAGAACTTTCTGAAATTATTTCTATTAGCAATGTCTCATTCCGT
19879                                                                  C                TTA

19801  T--GTGGTCAGAAAAGATACTTGGTATGATTTTAGTCTCCTTAAATTTGTTAAGATTTGTTTTGTGGGCTAGCATGTGAA
19959   CT  T                                C                              C

19879  CTATTCTGGAGACTGTTCTATGTGTGTTTGAGGAGAATGTATAATTTGCTGTTGTTGGGTAGAATATTCTGTATTTGTTT
20039                                     A                                 G

19959  GATAGGTCCACTTGGGCCATAGTGTTGTTTGAGTTATCTGTTTCCTTATTGATCTTCTGTCTGGATTTTTTATCCATTAT
20119                                     C                        A

20039  TGAAAGTGAGATATTACAATTTCTCACTATTATGTTGTTCTATATTTATCATCACTTCTGTCAACATTTGCTTCTTATAT
20199                  A   C                                                  T
```

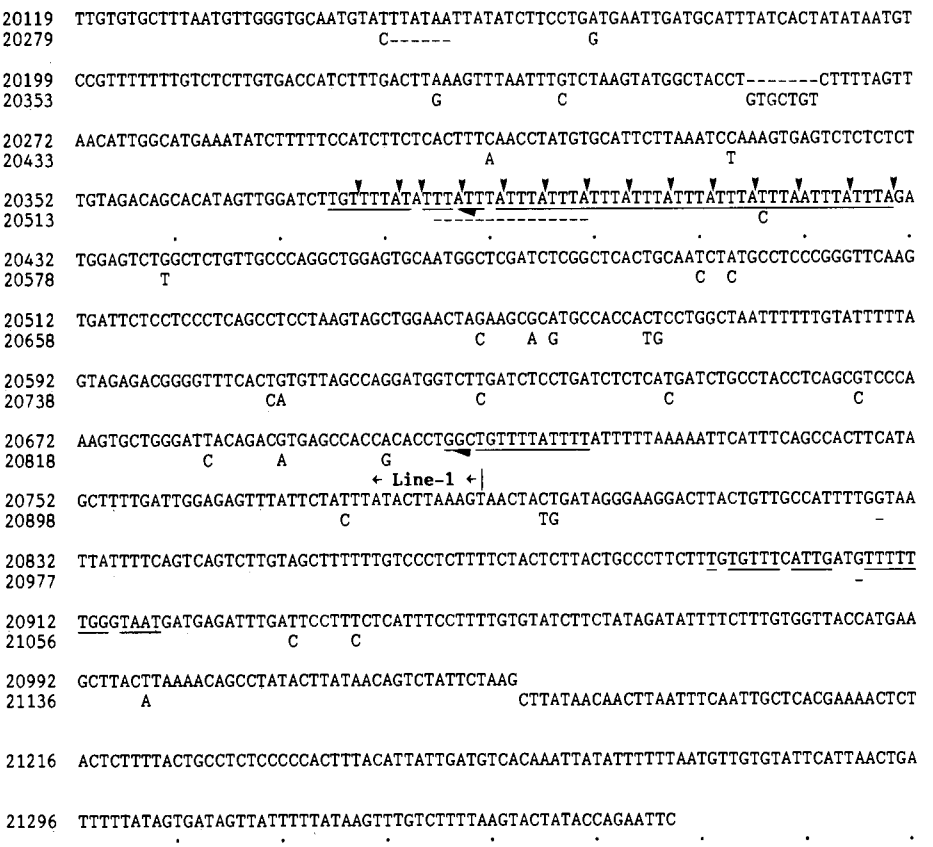


FIGURE 4: Nucleotide sequence of the vWF pseudogene and gene. Dots above and within the sequence are placed at 10-bp intervals. The pseudogene sequence is shown in the first line of each comparison. Where different, nucleotides corresponding to the vWF gene sequence are indicated in the second line, below the pseudogene sequence. Regions of the pseudogene corresponding to vWF exons are numbered 23–34. Translated amino acid sequence for the pseudogene is shown in single letter code in the third line. Where different, the translated amino acid sequence of the vWF gene is shown in the fourth line. The 5′ boundary of the pseudogene is indicated by double arrowheads, pointing up and down. Alu repeats are flanked by arrows that indicate their extent and direction, and flanking direct repeats are underlined. The approximate boundaries of a Line-1 repeat near the 3′ end of the sequence are indicated (Line-1), and candidate flanking direct repeats are underlined. Simple sequence repeats described in Table III are underlined, and the extent of monomer units is indicated by single arrowheads, pointing down. Insertions and deletions necessary to maintain alignment of the two sequences are indicated by dashes (---).

Table I: Divergence between the vWF Gene and Pseudogene

		substitutions				gaps		
sequences compared	BP	TV	TS	TS/TV	CG → TG	Ins	Del	% divergence
exons	2875	21	50	2.38	31	0	0	2.47
introns	17816	142	382	2.69	141	24	32	3.25
total	20691	163	432	2.65	172	24	32	3.14

[a] BP is the length of the sequences compared in nucleotides minus the length of any sequence opposite a gap; TV, transversion; TS, transition; TS/TV, the ratio of transitions to transversions; CG → TG, transitions that could be accounted for by deamination of methylcytosine; Ins, insertions relative to the gene sequence; Del, deletions relative to the gene sequence. Substitutions are the sum of TV + TS. Gaps are the sum of Ins + Del. Percent divergence is equal to [(100)(substitutions + gaps)/(BP + gaps)], in which gaps are counted as single events regardless of their length.

Table II: Alu and Line-1 Repeats in the Human vWF Gene and Pseudogene

Location/ Nucleotides	Repeat/ Orientation	Flanking direct repeats Length	Source	Sequence
		bp		
Intron 24 2123–2417	Alu/ inverted	15	ψ/g5′	AGTGCAGTTGTCTCT
			ψ/g3′	G
Intron 25 3941–4072	Alu/ direct	9	ψ/g5′	AGAAAACTG
		7	ψ/g3′	--
Intron 32 14237–14554	Alu/ direct	8	ψ5′	ACGCTTTT
			g5′	T
			ψ3′	T
			g3′	C
Intron 34 20388–20709	Alu/ inverted	12	ψ/g5′	TGTTTTATATTT
		11	ψ/g3′	TGTTTTAT-TTT
Intron 34 17104–20790	Line-1/ inverted	29	ψ/g5′	TATGTTTAATTGCCATTTTTTGGTCTAAT
		28	ψ3′	G C ATG -G
		27	g3′	G C AT- -G

flanking this Line-1 element were not identified with certainty; candidate 29 bp repeats are underlined in Figure 4 and listed

Table III: Simple Sequence Repeats in the Human vWF Gene and Pseudogene

		no. of repeats	
consensus	location	gene	pseudogene
$(CA)_n$	2513–2539	~12	~13
	14522–14555	17	10
CCTGCCGCCCCATCTCTGA-CCTGCAAG	4514–4713	~1.5	~7.5
$(TTTA)_n$	5539–5624	15	15
	20379–20429	12	12
$(GT)_n$	6799–6806	3	4
	7372–7379	4	4
	11353–11360	4	4
	17762–17789	14	16
$(YCAGGTC)_n$	9416–9444	4	4
RGGCAAATACTTAACCTCT-TTGTATATCATCTATA-CAAAGGATR	9520–9552	~2.5	<1
$(GTATT)_n$	11941–11970	6	5

in Table II. Several short simple sequence repetitive elements were identified, and some of these show differences in length between the gene and pseudogene (Table III).

Potential Homology with Other Genes. The vWF gene and pseudogene were compared with the sequences in the GenBank and EMBL databases. A segment of intron 22 (gene nucleotides 370–730) aligned with segments of seven other genes with scores above the mean (~60–70% identity, ≥24 standard deviations) that appeared to be well separated from those of all other comparisons. All but one of these sequences are from noncoding regions of primate genes: the 5′ flanking region of the human immunoglobulin heavy chain λ3 gene (Akahori et al., 1988), intron H of the human aldolase B gene (Tolan & Penhoet, 1986), intron D of the human apolipoprotein B gene (Ludwig et al., 1987), the 3′ flanking region of the human ϵ-globin gene (Collins & Weissman, 1984), the 5′ flanking region and first exon of the human Vκ A22 immunoglobulin pseudogene (Straubinger et al., 1987), the 5′ flanking region of the human major histocompatibility class II Dβ gene (Servenius et al., 1987), and intron 2 of the *Tarsius syrichta* β-globulin gene (Koop et al., 1989). The significance of these sequence similarities is not known, but it is possible that they represent fragments of repeated elements distinct from previously described repeats. No flanking direct repeats were identified, however, for any of these sequences. Aside from these examples, no apparently significant similarity was found between the noncoding regions of the vWF gene or pseudogene and other sequences in these databases.

Discrimination between Gene and Pseudogene Sequences by PCR. Sequence differences between the vWF gene and pseudogene were used to design oligonucleotide primers for the selective PCR amplification (Table IV). Primer combinations and PCR conditions first were tested on cloned cosmid DNA containing gene (c9) or pseudogene (p113) sequences. Specificity for the vWF gene or pseudogene was confirmed either by demonstrating restriction fragment patterns characteristic for each locus or by amplification of fragments of different length due to differences in the number of simple sequence repeats.

The efficacy of this method was confirmed with leukocyte DNA from normal individuals and from a patient (S.G.) with total deletion of the vWF gene (Shelton-Inloes et al., 1987) (data not shown). Specific amplification of regions corresponding to exons 23–34 was achieved. For each PCR product, restriction fragments that are diagnostic of the locus of origin can be identified (Figure 5). Thus, the primers selected (Table IV) permit rapid discrimination between the vWF gene and pseudogene loci.

Table IV: Oligonucleotide Primers for the Selective Amplification of Human vWF Gene or Pseudogene Sequences

Exon Number	Primer Number[a]	Sequence (5′→3′) Gene/Pseudogene[b]	Product (nt)	Annealing Temperature
23-24	286	AAGGAAGCTCAGGAATGGGT	735	55
	285	A G		
	283A	ACTCTGTGTCCATACCACCA		
	284A	TG		
25	290	AGAGTCTAGGCCCTATTGTC	505	55
	289	T		
	288A	ATGAGGGCGTCAGTACTCAC		
	287A	G T		
26	268	ACAGCTGTGCACCTGCCTGT	326/488[c]	55
	164A	CAGTTCCTCAGACACACTCG		
27	239	TTGACCCTGAAGACTGTCCA	326	58
	244	C G		
	241A	AAGCAGTCTAGTCCATCCCT		
	241A			
28	226	TGGGAATATGGAAGTCATTG	936	53
	242	AG CA		
	227A	CCGATCCTTCCAGGACGAAC		
	243A	A T		
29	337	TGTGCTCACCTTCCTGGTTG	216	55
	338	A		
	339A	TTTTGAATCAAGTAGAGCCA		
	340A	C G C		
30	341	AAAATAATCCGCATTTTCTC	305	55
	342	G C T		
	343A	TAGCCCCTCACTTTCCAGGA		
	344A	G		
31	345	ACCGTTAAGACAGGGTGTCG	344	55
	346	CA CG G C		
	347A	ACCAGGACAGAGGTTGGTAT		
	348A	C		
32	349	CCTTTCTTGTCTCTTTGCTA	359	55
	350	G		
	351A	CCATGAACAGAAACTTAAAG		
	352A	G G T		
33-34	333	TGCAGGATTTGTTAGGATTT	464	55
	331	C		
	334	GTTAGGGCACGAAGGCCTCA		
	332A	G		

[a] Numbers followed by "A" are reverse primers. [b] Complete sequences are shown only for vWF gene forward and reverse primers. Differences in the corresponding pseudogene-specific primers are indicated below the vWF gene primer sequences. [c] The length of the PCR product obtained with these primers is 326 bp for the fragment derived from the gene and 488 bp for the fragment derived from the pseudogene.

Discussion

In the course of studies of the human vWF gene on chromosome 12, a cross-hybridizing locus was identified on chromosome 22 (Shelton-Inloes et al., 1987). In the present report, cosmid and bacteriophage λ DNA clones were characterized that contain this locus, which was mapped to chromosome 22q11–13 by in situ hybridization (Figure 1), in agreement with a previous report of localization to 22q11.22–11.23 by similar methods (Patracchini et al., 1989). The most recently published map of chromosome 22 shows the locations of 49 genes including the vWF pseudogene (Kaplan & Emanuel, 1989). The region of chromosome 22 containing the vWF pseudogene has been associated with several acquired and inherited chromosomal alterations. The BCR locus of the Philadelphia chromosome in chronic myelogenous leukemia (de Klein et al., 1986) and acute lymphoblastic leukemia (Hermans et al., 1987) has been mapped to chromosome 22q11 (Groffen et al., 1989). In the same region, somatic deletions were identified in acoustic neuromas (McAlpine et al., 1988) and meningiomas (Kaplan & Emanuel, 1989). Translocations involving chromosome 22 have also been reported in the DiGeorge syndrome (Cannizzaro & Emanuel, 1985) and in cat eye syndrome with trisomy or tetrasomy 22 (Wilson et al., 1984). The vWF pseudogene may be a useful marker for genetic analysis of these diseases because of its proximity to these loci.

The vWF pseudogene is is ~21–29 kb in length and is contained in ≥6 *Eco*RI restriction fragments (Figures 2–4). It includes sequences that correspond to 12 exons (exons 23–34) of the vWF gene. The identification of this locus as a partial unprocessed pseudogene is based upon the preservation of sequences corresponding to vWF introns as well as exons and the presence of mutations that would be expected to prevent the transcription and splicing of a functional mRNA product. These include two donor splice site mutations and three nonsense mutations. The restriction map of the vWF pseudogene is very similar to that of the gene (Figure 3). The differences that exist explain the pattern of bands detected by Southern blotting with vWF cDNA probes and attributed

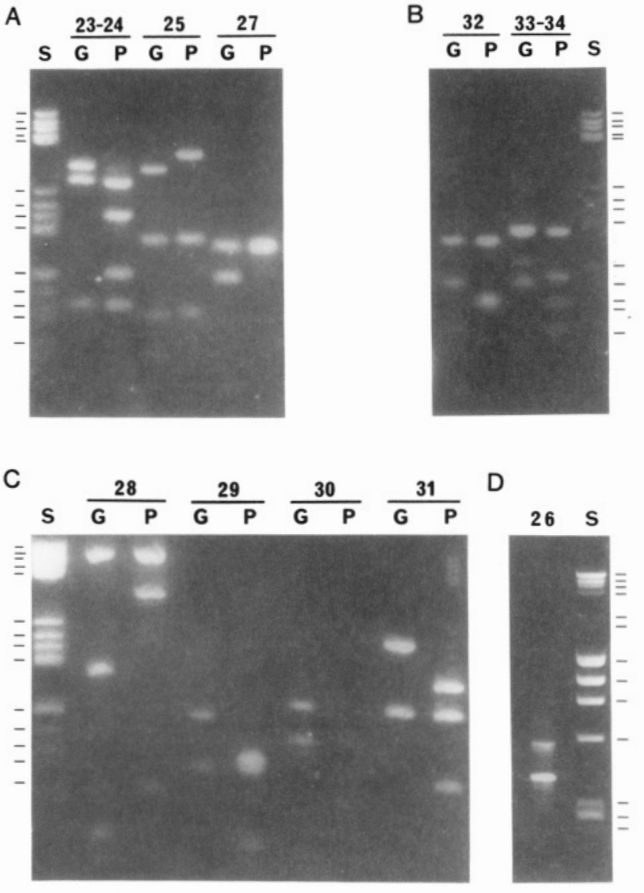


FIGURE 5: Selective PCR amplification of gene and pseudogene sequences. Genomic DNA was amplified with PCR primers specific for the pseudogene or gene as described under Experimental Procedures. The primers employed are listed in Table IV. Regions amplified are identified as "G" for gene or "P" for pseudogene, and the corresponding exon number or numbers is indicated. Panels A–C: The products of the PCR reactions were digested with restriction enzymes that product distinctive patterns for fragments derived from the gene and pseudogene. The enzymes employed were as follows: exons 23–24, *Mae*II; exon 25, *Sau*96I; exon 27, *Hae*III; exon 28, *Taq*I; exon 29, *Mbo*II; exon 30, *Hae*III, exon 31, *Hae*III, exon 32, *Hin*fI; exons 33 and 34, *Sau*96I. Panel D: The amplified product containing exon 26 incorporates a simple sequence repeat that differs in length between the gene and pseudogene. The upper band is derived from the pseudogene. DNA standard lanes are indicated by "S". The position of selected DNA standards is indicated at one side of each panel. The size in base pairs of the DNA standards in descending order is as follows: (panels A–C) 587, 540, 504, 457, 434, 267, 234, 213, 192/184, 124/123, 104, 89, 80, 64; (panel D) 23 130, 9416, 6557, 4361, 2322, 2027, 1353, 1078, 872, 603, 310, 281/271, 234.

previously to the chromosome 22 locus (Shelton-Inloes et al., 1987).

The vWF gene and pseudogene sequences have diverged ~3.14% (Table I). There is slightly less divergence between the sequences corresponding to exons (2.47%) and slightly more divergence between sequences corresponding to introns (3.25%). In the exonlike segments no insertions or deletions were observed, and potential CG → TG transitions account for a relatively large fraction (~44%) of the single-nucleotide substitutions. This apparent bias against insertions and deletions, and in favor of CG → TG mutations, may reflect selective pressure against mutations in the gene that would alter the amino acid sequence of vWF. This is consistent with selective pressure to maintain the sequence of exons, but not of introns, in the active vWF gene.

This degree of divergence suggests a recent origin for the vWF pseudogene. On the basis of an analysis of noncoding sequences flanking the $\psi\eta$-globin locus, the rate of accumulation of mutations at this locus in humans was estimated to be $(1.12–1.68) \times 10^{-9}$ mutations/(site·year) (Miyamoto et al., 1988). Assuming that this rate is approximately the same for the noncoding regions of the vWF gene and pseudogene loci, the vWF pseudogene is calculated to have arisen 19–29 million years ago, which is before the separation of orangutan from human and the African apes (Andrews & Cronin, 1982; Delson, 1985) and near the time of divergence of humans and apes from monkeys (20–30 million years ago) (Li et al., 1987). This result suggests that a vWF pseudogene should be present in the higher primates. To date, however, studies of the vWF gene in other animals have been limited to porcine populations. Bahou et al. (1988) failed to detect bands corresponding to a vWF pseudogene on Southern blots of restriction-digested porcine DNA using a human cDNA probe which hybridizes to both the human vWF gene and pseudogene.

Truncated unprocessed pseudogenes have been identified for a small number of other genes (Flint et al., 1988; Wilde, 1984). The mechanism of formation of these pseudogenes is probably through a translocation event but otherwise is not understood in any detail. Translocations and other DNA rearrangements are frequently associated with repetitive sequences at the boundaries of the rearrangements. For example, Alu repeats were found in the vicinity of the human growth hormone genes which may define the breakpoints of gene duplications (Hirt et al., 1987). Several Alu repeats and one Line-1 repeat were found in the vWF gene and pseudogene (Table II). However, no such repetitive sequences are present at the 5′ boundary of the pseudogene. The 3′ boundary has not yet been cloned and characterized.

Several simple sequence repeats were found in the vWF gene and pseudogene, and many of them show differences in length when the two loci are compared (Table III). These provide useful markers for identifying gene or pseudogene sequences in PCR products or in cloned DNA isolated by other methods. In addition, they are candidate sites for highly polymorphic minisatellite (Jefferys et al., 1985; Nakamura et al., 1987) or variable number of tandem repeat (VNTR) markers (Nakamura et al., 1987) that could be useful for genetic studies of vWD and for linkage analysis near the vWF gene or pseudogene loci. Such a polymorphic VNTR marker has been identified within intron 40 of the vWF gene, and it has been used successfully for the prenatal diagnosis of severe vWD in a family for which the tested restriction fragment length polymorphisms were uninformative (Peake et al., 1990). Preliminary studies of the 27-mer repeat region in intron 26 were performed by PCR with flanking oligonucleotides. No evidence for length polymorphism was seen for either the gene or the pseudogene among six unrelated persons.[2]

The vWF pseudogene complicates the analysis by Southern blotting of genetic lesions in patients with vWD. The pseudogene is only ~12–15% as large as the vWF gene, but it contains exonlike sequences that are ~97.5% identical with ~34% of the exon sequence of the gene. These exons happen to encode many of the known functional domains of the vWF protein, including the binding sites for heparin (Pareti et al., 1986), collagen (Roth et al., 1986), and platelet glycoprotein Ib (Mohri et al., 1987; Fujimura et al., 1987). This region is therefore of particular interest for studies of the genetic basis of vWD, particularly those variants in which a functionally defective protein is made. The high degree of sequence conservation between the gene and pseudogene make it very difficult to distinguish these loci by any hybridization method.

[2] D. J. Mancuso and J. E. Sadler, unpublished results.

Rapid analysis of the vWF gene sequence in the region shared with the pseudogene must rely on methods capable of distinguishing very closely related sequences, such as PCR methods. In the present study, selective amplification of vWF gene and pseudogene sequences was achieved for all of the exons 23–34 and corresponding exonlike sequences (Figure 5 and Table IV). Mismatches at the 3′ ends of primers and, in most cases, one to three additional internal mismatches were used. Mismatches at the 3′ ends of primers are particularly useful in conferring specificity, although not all such mismatches appear to be equally effective in preventing amplification of related sequences (Kwok et al., 1990; Newton et al., 1989; Sommer et al., 1989; Wu et al., 1989). Annealing temperatures were increased, if necessary, to improve selectivity. A single 3′ mismatch and one internal mismatch in one primer was sufficient for selective amplification of exon 27. In this case, selective amplification was achieved by increasing the annealing temperature. A single 3′ mismatch of both primers was sufficient for selective amplification of exons 33 and 34.

These methods will be generally useful for the characterization of vWF gene sequences in patients with vWD. Differences in the restriction maps of homologous vWF gene and pseudogene fragments have been exploited to confirm the localization within the gene of several candidate vWD mutations (Ginsburg et al., 1989; Bernardi et al., 1990). Comparison with the vWF pseudogene sequence suggests that two mutations at the 5′ end of exon 28, proposed to cause vWD type IIB, appear instead to be due to inadvertent amplification and sequencing of the pseudogene (Vink et al., 1989). To date, we have used this approach in two patients to characterize the end point of vWF gene deletions that lie within the region represented in the pseudogene (Mancuso et al., 1989a). In addition, candidate point mutations in several patients with vWD type IIB have been identified in exon 28, without interference from the related sequence in the pseudogene (Randi et al., 1991).

Registry No. vWF, 109319-16-6.

REFERENCES

Akahori, Y., Handa, H., Imai, K., Abe, M., Kameyama, K., Hibiya, M., Yasui, H., Okamura, K., Naito, M., Matsuoka, H., & Kurosawa, Y. (1988) *Nucleic Acids Res. 16*, 9497–9411.

Andrews, P., & Cronin, J. (1982) *Nature (London) 297*, 541–546.

Arnaout, M. A., Gupta, S. K., Pierce, M. W., & Tenen, D. G. (1988) *J. Cell Biol. 106*, 2153–3158.

Bahou, W. F., Bowie, E. J. W., Fass, D. N., & Ginsburg, D. (1988) *Blood 72*, 308–313.

Bentley, D. R. (1986) *Biochem. J. 239*, 339–345.

Bernardi, F., Marchetti, B., Guerra, S., Casonato, A., Gemmati, D., Patracchini, P., Ballerini, G., & Conconi, F. (1990) *Blood 75*, 677–683.

Bonaldo, P., Russo, V., Bucciotti, F., Bressan, G. M., & Colombatti, A. (1989) *J. Biol. Chem. 264*, 5575–5580.

Bonaldo, P., Russo, V., Bucciotti, F., Doliana, R., & Colombatti, A. (1990) *Biochemistry 29*, 1245–1254.

Bonthron, D. T., Handin, R. I., Kaufman, R. J., Wasley, L. C., Orr, E. C., Mitsock, L. M., Ewenstein, B., Loscalzo, J., Ginsburg, D., & Orkin, S. H. (1986) *Nature 324*, 270–273.

Cannizzaro, L. A., & Emanuel, B. S. (1985) *Cytogenet. Cell Genet. 39*, 179–183.

Chu, M.-L., Pan, T., Conway, D., Kuo, H.-J., Glanville, R. W., Timpl, R., Mann, K., & Deutzmann, R. (1989) *EMBO J. 8*, 1939–1946.

Collins, F. S., & Weissman, S. M. (1984) *Prog. Nucleic Acid Res. Mol. Biol. 31*, 315–462.

Corbi, A. L., Miller, L. J., O'Connor, K., Larson, R. S., & Springer, T. A. (1987) *EMBO J. 6*, 4023–4028.

Corbi, A. L., Kishimoto, T. K., Miller, L. J., & Springer, T. A. (1988) *J. Biol. Chem. 263*, 12403–12411.

Deininger, P. L., & Schmid, C. W. (1979) *J. Mol. Biol. 127*, 437–460.

de Klein, A., van Agthoven, T., Groffen, C., Heisterkamp, N., Groffen, J., & Grosveld, G. (1986) *Nucleic Acids Res. 14*, 7071–7081.

Delson, E. (1985) *Nature 313*, 532–533.

Flint, J., Taylor, A. M., & Clegg, J. B. (1988) *J. Mol. Biol. 199*, 427–437.

Fujmura, Y., Titani, K., Holland, L. Z., Roberts, J. R., Kostel, P., Ruggeri, Z. M., & Zimmerman, T. S. (1987) *J. Biol. Chem. 262*, 1734–1739.

Gilbert, W. (1978) *Nature 271*, 501.

Ginsburg, D., Konkle, B. A., Gill, J. C., Montgomery, R. R., Bockenstedt, P. A., Johnson, T. A., & Yang, A. Y. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3723–3727.

Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R., & Grosveld, G. (1989) *Cell 36*, 93–99.

Grunebaum, L., Cazenave, J.-P., Camerino, G., Kloepfer, C., Mandel, J.-L., Tolstoshev, P., Jaye, M., De la Salle, H., & Lecocq, J.-P. (1984) *J. Clin. Invest. 73*, 1491–1495.

Haley, J., Crawford, R., Hudson, P., Scanlon, D., Tregear, G., Shine, J., & Nial, H. (1987) *J. Biol. Chem. 262*, 11940–11946.

Hermans, A., Heisterkamp, N., von Lindern, M., van Baal, S., Meijer, D., ver der Plas, D., Wiedemann, L. M., Groffen, J., Bootsma, D., & Grosveld, G. (1987) *Cell 51*, 33–40.

Hirt, H., Kimelman, J., Birnbaum, M. J., Chen, E. Y., Seeburg, P. H., Eberhardt, N. L., & Barta, A. (1987) *DNA 6*, 59–70.

Jefferys, A. J., Wilson, V., & Thein, S. L. (1985) *Nature 314*, 67–73.

Kaplan, J.-C., & Emanuel, B. S. (1989) *Cytogenet. Cell Genet. 51*, 372–383.

Kiss, I., Deak, F., Holloway, R. G., Jr., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., & Goetinck, P. F. (1989) *J. Biol. Chem. 264*, 8126–8134.

Koller, E., Winterhalter, K. H., & Trueb, B. (1989) *EMBO J. 8*, 1073–1077.

Koop, B. F., Siemieniak, D., Slightom, J. L., Goodman, M., Dunbar, J., Wright, P. C., & Simons E. L. (1989) *J. Biol. Chem. 264*, 68–79.

Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L. Levenson, C., & Sninsky, J. J. (1990) *Nucleic Acids Res. 18*, 999–1005.

Larson, R. S., Corbi, A. L., Berman, L., & Springer, T. (1989) *J. Cell Biol. 108*, 703–712.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) *Cell* 15, 1157–1174.

LeBeau, M. M., Westbrook, C. A., Diaz, M. O., & Rowley, J. D. (1984) *Nature 312*, 70–71.

Li, W.-H., Tanimura, M., & Sharp, P. M. (1987) *J. Mol. Evol. 25*, 330–342.

Lipman, D. J., & Pearson, W. R. (1985) *Science 227*, 1435–1441.

Ludwig, E. H., Blackhart, B. D., Pierotti, V. R., Caiati, L., Fortier, C., Knott, T., Scott, J., Mahley, R. W., Levy-Wilson, B., & McCarthy, B. J. (1987) *DNA 6*, 363–372.

Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G., & Sadler, J. E. (1988) *Circulation 78* (*Suppl. 2*), 506.

Mancuso, D. J., Lester-Mancuso, T. L., Tuley, E. A., Westfield, L. A., & Sadler, J. E. (1989a) *Blood 74* (*Suppl. 1*), 488a.

Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G., & Sadler, J. E. (1989b) *J. Biol. Chem. 264*, 19514–19527.

McAlpine, P. J., Boucheix, C., Pakstis, A. J., Stranc, L. C., Berent, T. F., & Shows, T. B. (1988) *Cytogenet. Cell Genet. 49*, 4–38.

Miyamoto, M. M., Koop, B. F., Slightom, J. L., Goodman, M., & Tennant, M. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7627–7631.

Mohri, H., Fujmura, Y., Shima, M., Yoshioka, A., & Houghten, R. A. (1988) *J. Biol. Chem. 263*, 17901–17904.

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

Nakamura, Y., Lepport, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., & White, R. (1987) *Science 235*, 1616–1622.

Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., & Markham, A. F. (1989) *Nucleic Acids Res. 17*, 2503–2516.

Pareti, F. I., Fujimura, Y., Dent, J. A., Holland, L. Z., Zimmerman, T. S., & Ruggeri, Z. M. (1986) *J. Biol. Chem. 261*, 15310–15315.

Patracchini, P., Calzolari, E., Aiello, V., Palazzi, P., Banin, P., Marchetti, G., & Bernardi, F. (1989) *Hum. Genet. 83*, 264–266.

Peake, I. R., Liddell, M. B., Moodie, P., Standen, G., Mancuso, D. J., Tuley, E. A., Westfield, L. A., Sorace, J. M., Sadler, J. E., Verweij, C. L., & Bloom, A. L. (1990) *Blood 75*, 654–661.

Pytela, R. (1988) *EMBO J. 7*, 1371–1378.

Randi, A. M., Rabinowitz, I., Mancuso, D. J., Mannucci, P. M., & Sadler, J. E. (1991) *J. Clin. Invest.* (in press).

Rodeghiero, F., Castaman, G., & Dini, E. (1987) *Blood 69*, 454–459.

Roth, G. J., Titani, K., Hoyer, L. W., & Hickey, M. J. (1986) *Biochemistry 25*, 8357–8361.

Ruggeri, Z. M., & Zimmerman, T. S. (1987) *Blood 70*, 895–904.

Sadler, J. E. (1989) in *The Metabolic Basis of Inherited Disease* (Scriver, Beaudet, Sly, & Valle, Eds.) 6th ed., pp 2171–2188, McGraw-Hill Book Co., New York.

Sadler, J. E., Shelton-Inloes, B. B., Sorace, J. M., & Titani, K. (1986) *Cold Spring Harbor Symp. Quant. Biol. 51*, 515–523.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487–491.

Servenius, B., Rask, L., & Peterson, P. A. (1987) *J. Biol. Chem. 262*, 8759–8766.

Shelton-Inloes, B. B., Titani, K., & Sadler, J. E. (1986) *Biochemistry 25*, 3164–3171.

Shelton-Inloes, B. B., Chehab, F. F., Mannucci, P. M., Federici, A. B., & Sadler, J. E. (1987) *J. Clin. Invest. 79*, 1459–1465.

Skowronski, J., Fanning, T. G., & Singer, M. F. (1988) *Mol. Cell. Biol. 8*, 1385–1397.

Sommer, S. S., Cassady, J. D., & Bottema, C. D. K. (1989) *Mayo Clin. Proc. 64*, 1361–1372.

Straubinger, B., Osterholzer, E., & Zachau, H. G. (1987) *Nucleic Acids Res. 15*, 9567–9575.

Takada, Y., & Hemler, M. E. (1989) *J. Cell Biol. 109*, 397–407.

Tolan, D. R., & Penhoet, E. E. (1986) *Mol. Biol. Med. 3*, 245–264.

Verweij, C. L., Diergaarde, P. J., Hart, M., & Pannekoek, H. (1986) *EMBO J. 5*, 1839–1847.

Vink, T., de Groot, P. G., Federici, A. B., Mannucci, P. M., & Sixma, J. J. (1989) *Thromb. Haemostas. 62*, 472.

Wahl, G. M., Lewis, K. A., Ruiz, J. C., Rothenberg, B., Zhao, J., & Evans, G. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2160–2164.

Wilde, C. D. (1984) *CRC Crit. Rev. Biochem. 19*, 323–352.

Wilson, G. N., Baker, D. L., Schau, J., & Parker, J. (1984) *J. Med. Genet. 21*, 60–63.

Wu, D. Y., Ugozzoli, L., Pol, B. K., & Wallace, R. B. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2759–2560.