Nakamura, K. D., & Weber, M. J. (1982) *Mol. Cell. Biol. 2*, 147-153.
Nigg, E. A., Cooper, J. A., & Hunter, T. (1983) *J. Cell Biol. 96*, 1601-1609.
Pepinsky, R. B., Tizard, R., Mattaliano, R. J., Sinclair, L. K., Miller, G. T., Browning, J. L., Chow, E. P., Burne, C., Huang, K.-S., Pratt, D., Wachter, L., Hession, C., Frey, A. Z., & Wallner, B. P. (1988) *J. Biol. Chem. 263*, 10799-10811.
Radke, K., & Martin, G. S. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5212-5216.
Radke, K., Gilmore, T., & Martin, G. S. (1980) *Cell 21*, 821-828.
Radke, K., Carter, V. C., Moss, P., Dehazya, P., Schliwa, M., & Martin, G. S. (1983) *J. Cell Biol. 97*, 1601-1611.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Saris, C. J. M., Tack, B. F., Kristensen, T., Glenney, J. R., Jr., & Hunter, T. (1986) *Cell 46*, 201-212.
Saris, C. J. M., Kristensen, T., D'Eustachio, P., Hicks, L. J., Noonan, D. J., Hunter, T., & Tack, B. F. (1987) *J. Biol. Chem. 262*, 10663-10671.
Silver, J., & Buckler, C. E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1423-1427.
Simmen, R. C. M., Tanaka, T., Ts'ui, K. F., Patkey, J. A., Scott, M. J., Lai, E. C., & Means, A. R. (1985) *J. Biol. Chem. 260*, 907-912.
Smith, M., Leung, D. W., Gillam, S., & Astell, C. R. (1979) *Cell 16*, 753-761.
Staden, R. (1980) *Nucleic Acids Res. 8*, 3673-3694.
Staden, R. (1982) *Nucleic Acids Res. 10*, 4731-4751.
Taylor, B. A. (1978) in *Origins of Inbred Mice* (Morse, H. C., III, Ed.) pp 423-428, Academic Press, New York.
Taylor, B. A. (1981) in *Genetic Variants and Strains of the Laboratory Mouse* (Green, M. C., Ed.) pp 397-407, Gustav Fischer, Stuttgart, FRG.
Zokas, L., & Glenney, J. R., Jr. (1987) *J. Cell Biol. 105*, 2111-2121.

Organization of the Gene for Platelet Glycoprotein IIb†

Randy Heidenreich,‡ Robin Eisman,§ Saul Surrey,§,∥ Kathleen Delgrosso,§ Joel S. Bennett,⊥ Elias Schwartz,§,∥ and Mortimer Poncz*,§

Divisions of Hematology and Metabolic Diseases, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, Division of Hematology and Oncology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania 19104, and Departments of Pediatrics, Human Genetics, and Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received July 12, 1989; Revised Manuscript Received September 21, 1989

ABSTRACT: The glycoprotein (GP) IIb/IIIa heterodimer functions as a receptor for fibrinogen, von Willebrand factor, and fibronectin on activated platelets; it is dysfunctional in the bleeding diathesis Glanzmann's thrombasthenia. This receptor is a member of the integrin family, which includes homologous membrane receptors involved in a number of different cell-cell and cell-matrix adhesive interactions. Knowledge of the sequence and organization of the GPIIb and GPIIIa genes will help in understanding evolutionary relationships and functional homologies of this family of adhesion protein receptors and will facilitate analysis of molecular defects responsible for thrombasthenia. Using the GPIIb cDNA as a probe, we have isolated overlapping genomic clones encompassing the entire coding region, the 5′- and 3′-untranslated sequences, and the immediate flanking regions for the GPIIb gene. The gene spans approximately 17.2 kilobases (kb); all but approximately 2.6 kb of intronic DNA sequence has been determined. The GPIIb gene contains 30 exons whose demarcations do not correlate with previously suggested functional domains. Two intron/exon borders have the rare GC splice donor sequence instead of the consensus GT sequence. There are at least seven complete and three partial *Alu*I sequence repeats within the intron sequences. RNase protection, S1 nuclease analysis, and primer extension studies using human erythroleukemia (HEL) cell RNA and platelet RNA map a major transcription start site 32 base pairs (bp) 5′ to the beginning of the coding region; however, there are no canonical consensus TATA or CAAT boxes in the region immediately 5′ to the proposed cap site. The immediate 5′-flanking sequence of rodent GPIIb demonstrates complete identity near the proposed cap site with its human counterpart, but again, no TATA or CAAT boxes are apparent.

The GPIIb/IIIa complex, a calcium-dependent heterodimer present in platelet membranes, is the binding site for fibrinogen (Jennings & Phillips, 1982; Bennett et al., 1982, 1983), fibronectin, and von Willebrand factor (Ruggeri et al., 1982; Ginsberg et al., 1983) on activated platelets. Glycosylated GPIIb has a M_r of 136 000 on SDS-polyacrylamide gels and contains two disulfide-linked subunits: a heavier M_r 125 000 subunit and a lighter M_r 23 000 subunit (Jennings & Phillips, 1982). GPIIIa is a single-chain disulfide bond rich glycoprotein with an unreduced apparent M_r of 95 000 that increases to 110 000 following disulfide reduction (Jennings & Phillips, 1982). Defects or deficiencies of the GPIIb/IIIa heterodimer underlie the autosomal recessive bleeding disorder Glanzmann's thrombasthenia. GPIIb/IIIa is a member of the integrin family of membrane adhesion receptors that mediate cell-cell and cell-matrix interactions. This family includes

† Supported in part by grants from the National Institutes of Health (HL37419, AM16691, HL40387, HL28157, HD07107), the March of Dimes, and the Council for Tobacco Research-U.S.A., Inc.
* Author to whom reprint requests and correspondence should be addressed at The Children's Hospital of Philadelphia.
‡ Division of Metabolic Diseases, The Children's Hospital of Philadelphia.
§ Division of Hematology, The Children's Hospital of Philadelphia, and Department of Pediatrics, University of Pennsylvania.
∥ Department of Human Genetics, University of Pennsylvania.
⊥ Division of Hematology and Oncology, Hospital of the University of Pennsylvania, and Department of Medicine, University of Pennsylvania.

0006-2960/90/0429-1232$02.50/0

the fibronectin and vitronectin receptors found on a wide variety of cells, and the LFA-1, p150,95, and Mac-1 complexes found on leukocytes and lymphocytes (Hynes, 1987). As in the GPIIb/IIIa complex, other integrins are composed of two different chains, an α (GPIIb-like) chain and a β (GPIIIa-like) chain. The cDNA sequences for a number of the integrin α and β chains, including GPIIb and GPIIIa, are now known (Tamkun et al., 1986; Argraves et al., 1986; Suzuki et al., 1987; Poncz et al., 1987a; Zimrin et al., 1988; Fitzgerald et al., 1987a,b; Bray et al., 1987; Kishimoto et al., 1987b; Corbi et al., 1987; Bogaert et al., 1987; Rosa et al., 1988; Aranout et al., 1988). Analysis of the deduced amino acid sequences indicates that all α and β chains appear to have an N-terminal signal peptide and a single transmembrane domain near the C-terminus. GPIIb, like a number of the other α chains, is cleaved after translation into two disulfide-bonded subunits. As in other α chains, the heavier GPIIb subunit contains four potential calcium-binding domains, while the lighter subunit contains a hydrophobic transmembrane domain and a short hydrophilic cytoplasmic domain. GPIIIa contains four cysteine-rich repeats and a hydrophilic cytoplasmic domain. While the α and β chains have no protein homology, there is approximately 40% amino acid identity among the various α chains as well as among the β chains of the integrin receptors (Hynes, 1987; Phillips et al., 1988).

The integrins can be divided into at least three families, on the basis of the specific β chain that is present (Hynes, 1987; Phillips et al., 1988). GPIIIa is the β subunit present in the platelet GPIIb/IIIa receptor and in the widely distributed vitronectin receptor (Ginsberg et al., 1987; Zimrin et al., 1988; Fitzgerald et al., 1987b). The white cell LFA-1, Mac-1, and p150,95 receptors contain a common β subunit (Hynes, 1987; Phillips et al., 1988). In leukocyte adhesion disease, all three receptors are defective, and in several well-studied examples, the disorder appears to be secondary to an absence of expression of the β subunit (Kishimoto et al., 1987a).

Knowledge of the organization of the genes for α and β integrin genes is essential for complete understanding of both their evolutionary relationships and the regulation of their tissue-specific expression. To this end, we have used a full-length GPIIb cDNA (Poncz et al., 1987a) to isolate and characterize overlapping genomic clones containing the entire gene and its 5'- and 3'-flanking regions.

Materials and Methods

Isolation and Characterization of Genomic GPIIb. The human genomic library used in these studies was prepared from leukocyte DNA of a patient with thalassemia (Semenza et al., 1984). Clones positive for GPIIb were isolated from this amplified library by screening with a 1.9-kb GPIIb cDNA fragment representing the 3' end of the GPIIb coding region and with a 460-bp cDNA fragment [generated by *Bal*31 digestion of the full-length cDNA (Poncz et al., 1982b)] containing the 5' end of the 3.3-kb GPIIb cDNA. Probes were radiolabeled with [α-^{32}P]dCTP by use of DNAse-generated calf thymus DNA random primers (Maniatis et al., 1982).

Purified clones were subjected to single and double restriction enzyme digestion. The digests were size fractionated on agarose gels, stained with ethidium bromide, and photographed. Gels were then blotted (Southern, 1975) onto Genescreen-Plus (New England Nuclear, Boston, MA) according to the manufacturer's recommendations and probed with radiolabeled GPIIb cDNA to localize exon sequences in the genomic inserts.

Rodent GPIIb cDNA (Poncz et al., 1989) was isolated from a rat megakaryocyte cDNA library in λgt11 (kindly provided by Robert Rosenberg, MIT). A *Hin*dIII 250-bp 5' fragment from the rodent GPIIb cDNA was used to isolate 5'-GPIIb genomic clones from a partial *Sau*3A rodent genomic library in EMBL3A (kindly provided by Richard Hynes, MIT).

Subcloning of Restriction Fragments, Generation of Deletion Mutants, and Sequence Determination. Restriction fragments from genomic clones were electroeluted from agarose gels and subcloned into linearized M13mp18 and/or 19 phage (Messing, 1983). Clones in both orientations were identified with S1 nuclease digestion combinations of single-stranded DNA from M13 clones (Poncz et al., 1982b). The size of the *Bal*31 deletion mutants generated by digestion of double-stranded DNA was evaluated by S1 nuclease digestion of heteroduplexes with single-stranded phage containing the full-length complementary insert (Poncz et al., 1982b). The DNA sequence of overlapping deletion mutants was determined by the dideoxy chain termination technique (Sanger et al., 1977) using either the Klenow fragment of *Escherichia coli* DNA polymerase or modified T_7 DNA polymerase (Sequenase) (Tabor & Richardson, 1987) with either a universal M13 primer or specifically designed oligonucleotide primers.

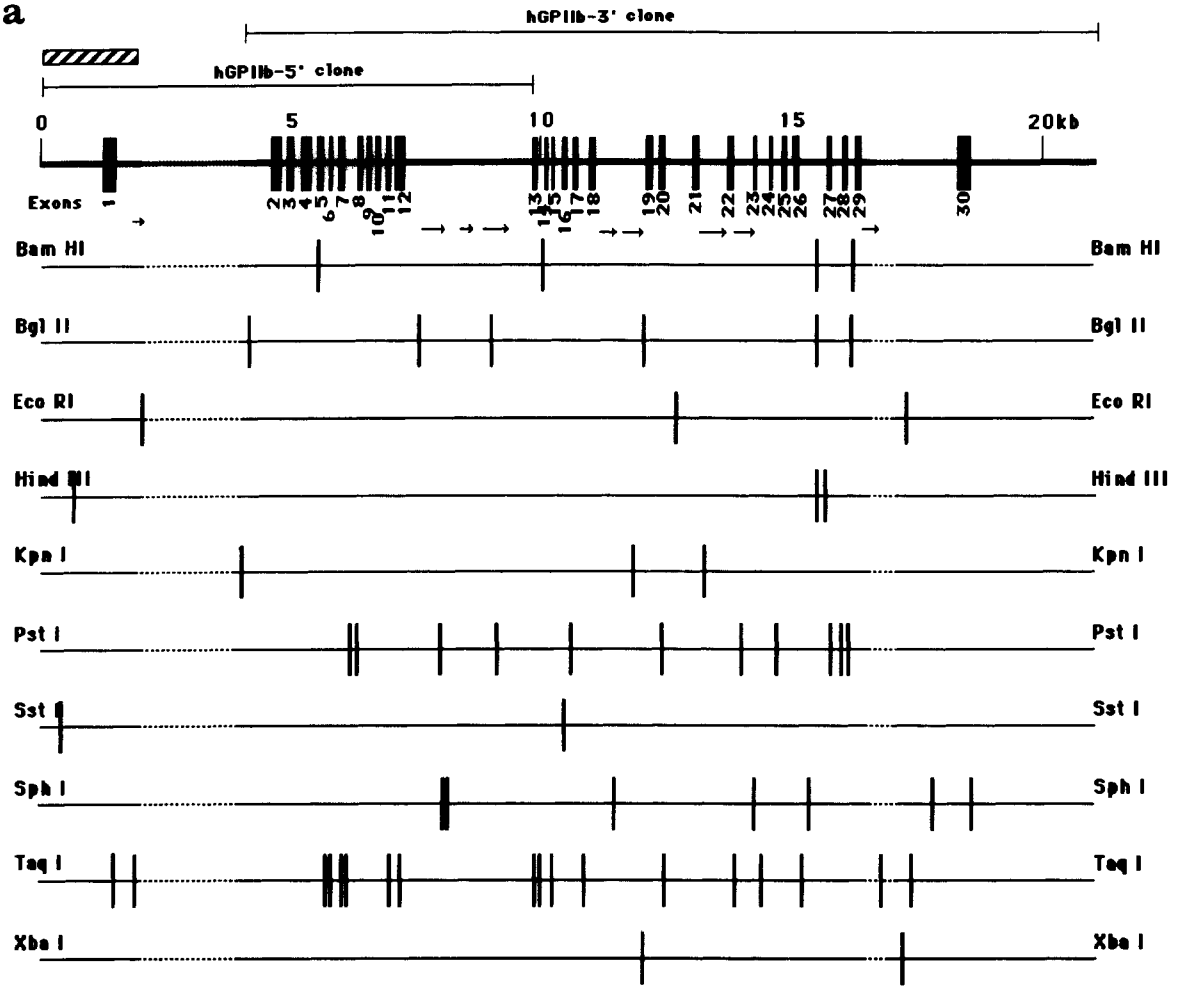
Southern Blot Analysis of Genomic DNA. Southern blot analysis of genomic DNA from a panel of normal controls was done as recommended by the manufacturer (GeneScreen Plus, Du Pont, DE) and probed with either radiolabeled 1.5-kb 5', 1.0-kb middle, or 0.8-kb 3' *Sac*I fragments of GPIIb as well as with the full-length cDNA. Information derived from these blots and from DNA sequence analysis was used in the construction of the restriction map shown in Figure 1a.

RNA Preparation. Total cellular RNA from HEL cells induced with 1.25% (v/v) DMSO (Tabilio et al., 1984), HL60 cells (Collins et al., 1977), and K562 cells induced with 10^{-8} M phorbol 12-myristate-13-acetate (TPA) (Lozzio & Lozzio 1975) were isolated either by the guanidinium thiocyanate (GTC)/cesium chloride banding technique (Ullrich et al., 1977; Poncz et al., 1987b) or with GTC/acid phenol (Chomczynski et al., 1987). Total platelet RNA was prepared from fresh platelets obtained from a normal individual by GTC/acid phenol.

Primer Extension. Primer extension of GPIIb mRNA was performed with a 30-mer oligonucleotide (double underlined in Figure 1) and AMV reverse transcriptase (Geliebter et al, 1986). Briefly, the oligonucleotide was end labeled with T4 polynucleotide kinase to 1×10^9 dpm/μg, annealed to 10 μg of total HEL cell RNA, and extended with AMV reverse transcriptase in the presence of 100 μg/mL actinomycin D and triphosphates. The products of the primer extension reaction were fractionated on a 6% (w/v) denaturing polyacrylamide sequencing gel.

RNase Protection. Protection of uniformly labeled RNA transcribed from genomic DNA was performed as previously described (Zinn et al., 1983). An *Eco*RI/*Sal*I 1.9-kb fragment (which contains 1253 bp of 5'-flanking sequence, exon 1, and 403 bp of the first intron and is shown as the cross-hatched area in Figure 1a) was cloned into pGEM 3Z. This construct was then truncated by the removal of eight bases of GPIIb exon 1 and all of the first intron by double digestion with *Eco*RI and *Nco*I. The DNA fragment was then treated with S1 nuclease to produce blunt ends and religated with T_4 ligase. To perform RNase protection experiments, 0.5 μg of each construct was first linearized with *Hin*dIII, and then transcribed in vitro with [α-^{32}P]dATP, the other three triphosphates, and T_7 RNA polymerase. The ^{32}P-radiolabeled

a

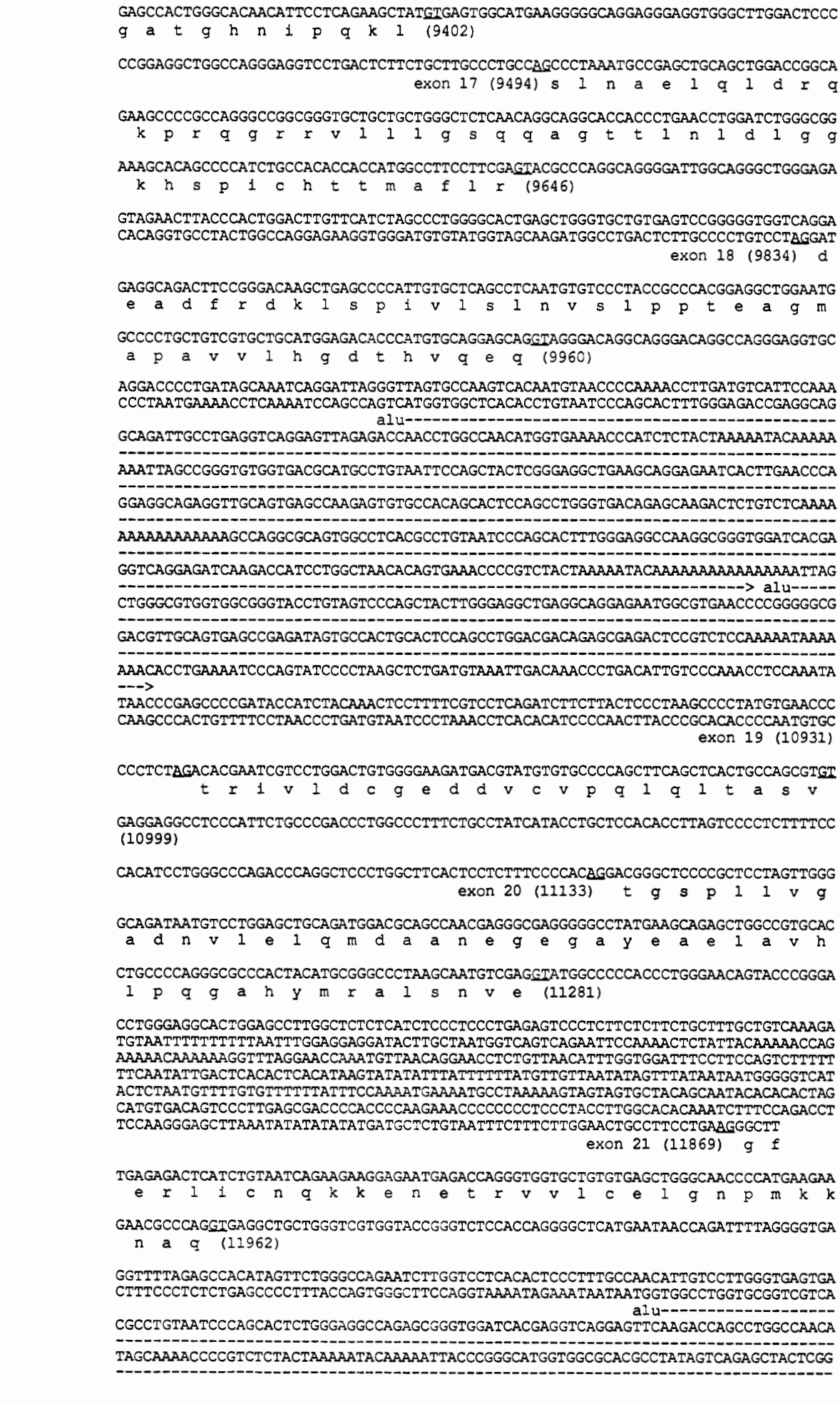


b

```
-1253
     GTCAACGGATCAGAAAATAGAAATCAAAAGGAAAATGTGGCTATGGTTACCCCTAGCGGACCTCTTAAATCTT
CCTGAGAACCTGCTTTTTTGGGAAGGCATGAGTGCCAGTAAGACTTGGCACTCCTCCTCTTCCGCTTACCGAGAGAAA
ATGACTTTGCCTTTCTGCTCAAAACTCATCCCTTCACTTTGTCACCCTATGTTTGCATCTTCCATCCTTAGTGTGTGT
TTCCATCCATCCAGTCTTTCAGCAATACACGTACTACACATTGGACTCTTGGGTAGTCTCTAGGGCTGTAGCAAGGAG
CCTTGCTCCCAAGGGACTCATTTACACAATCCTGTGAACGGACCAAGAGTAAACAGTGTGCTCAATGCTGTGCCTACG
TGTGTTAGCCCACGCGGCCAGCCTGAGGAGTCAGGGAAGGCTCCCCTAGGCAAAGCCCCCAACCAGAATCAAGTCTTA
ATGGTTAAAGAGCTCCATCACCCAAAAAGGATTGAGGGCCTACCTTCAACTGAACAGCTAATGCATAATCTCAGAAAC
TGTGAGTCAAAATTCCCTGGAATAACTCCACTTTATCCCCAATCTCCTTGCCACCTAGACCAAGGTCCATTCACCACC
CTGTCCCCAGCACTGACTGCACTGCTGTGGCCACACTAAAGCTTGGCTCAAGACGGAGGAGGAGTGAGGAAGCTGCTG
CACCAATATGGCTGGTTGAGGCCGCCCAAGGTCCTAGAAGGAGGAAGTGGGTAAATGCCATATCCAAAAAGATACAGA
AGCCTCAGGTTTTATCGGGGGCAGCAGCTTCCTTCTCCTTCCCCGACCTGTGGCCAAGTCACAAAGCACCACAGCTGT
ACAGCCAGATGGGGGAAGGGAGGAGATTAGAACTGTAGGCTAGAGTAGACAAGTATGGACCAGTTCACAATCACGCTA
TCCCAAGCAGAAAGTGATGGTGGCTTGGACTAGCACGGTGGTAGTAGAGATGGGGTAAAGATTCAAGAGACATCATTG
ATAGGCAGAACCAATAGGACATGGTAATAAACTATTCTCAGGAAAGGGGAGGAGTCATGGCTTTCAGCCATGAGCATC
CACCCTCTGGGTGGCCTCACCCACTTCCTGGCAATTCTAGCCACCATGAGTCCAGGGGCTATAGCCCTTTGCTCTGCC
CGTTGCTCAGCAAGTTACTTGGGGTTCCAGTTTGATAAGAAAAGACTTCCTGTGGAGGAATCTGAAGGGAAGGAGGAG
GAGCTGGCCC -1

+1 (cap site)
  ATTCCTGCCTGGGAGGTTGTGGAAGAAGGAAGATGGCCAGAGCTTTGTGTCCACTGCAAGCCCTCTGGCTTCTGGA
                    exon 1 (33)    m  a  r  a  l  c  p  l  q  a  l  w  l  l  e

GTGGGTGCTGCTGCTCTTGGGACCTTGTGCTGCCCCTCCAGCCTGGGCCTTGAACCTGGACCCAGTGCAGCTCACCTT
  w  v  l  l  l  l  g  p  c  a  a  p  p  a  w  a  l  n  l  d  p  v  q  l  t  f

CTATGCAGGCCCCAATGGCAGCCAGTTTGGATTTTCACTGGACTTCCACAAGGACAGCCATGGGAGGTGAGCCGTAAG
  y  a  g  p  n  g  s  q  f  g  f  s  l  d  f  h  k  d  s  h  g  r (220)

GGAAGTTGGGGTATTGGGAGAGAGCAGGACCCCTCCCCATCACTGCTTCTGGGGGCTTCGAGTTTCCCATTTGCGATA
GCAGTTGAGCAAGGTGACTTGTGGGGCCTATTCAGGTTGATTTCTTGTCAAGAATGTTGGGGTCCAGGGGACTGGCTC
AGGTGAAGGTATAAGGGCAGGGCACATGTGGGCTGATGGGCACTGAAAACTACAGCAAGAACAAAGGGAAGACAAGAG
TTGATGCTTTATTTTTTCCCCAAGGGTCAGTTGTATGAACCACTCCACCCTCAACACCTTGAAATGCAGAGAGGAGGC
CGGGCGCGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCGAGAATTCG
      alu----------------------------------------------------------------------
AGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTACTAAAAATACAAAAAAAAAAAAAAAAAAAAGAGGCCAGGCA
----------------------------------------------------------------------> alu----
CAGTGGCTCACACCTGCAATCCCAGCACTTTGGGAGGCAGAGGTGGGCAGATCATGAGGTCAGGAGTTCAAGACCAGC
----------------------------------------------------------------------------
```

```
CTGGCCAATATGGTGAAACCCTGTCTCTATTAAAAATACAAAAATTAGCTGAGCATGGTGGCACACTCCTGTAGTCCC
AGCTACTCGGGAGGCTGAGGTAGGAGAATCACTTGAACCCGGGAGGTGGAGGTTGCAGTGAGCTGAGACTATGCCACT
--------------------------------------------------------------------------------
GCACTCCCAGCCCTGGGGTTGACAGAGTGACACTCCGTCTCAAAAAAAAAA
-------------------------------------------------->
<----------------------approximately 2000 bp of intron ----------------------->
CTGCAGGTCAACGGATCTGCTAGGGTCCTCCTATCAGCACACACACTCCAGCCCCACTTTAGAGGTACCCGCTACCTT
CCCTCATTAAAACCAGCTCTCAAGAGGGGATCTGGTAACAGTCTAGGCAGGCATTCCAGGGAGCATGTGAACCGCTGG
TTCTTGTTGCGGGTGGAGGATGGAGGTGTTGTACAGAGTTTAGGTCTTTTTCAGCAAAGATCTCCAAACCCCGGGTGT
TCAAAATCAAACCAAAGGGGATTATAGTCCCAGCTCTACTCACAACTCACTGGTTACTTTAGCCACGAGATTGCCCTC
GCTGAGAGTCGGTTTCACTGTCCATAAGATGAAGAAGTACATCACGGTGGTCTGTGAGGTGTCATTGAGGAAAGATGG
TCCAGTGCCCCCATGCCACATGGCCTTCGGGCAGTGCTCCCAGCGCCGGCGCCAGGGCCTGGGATACGCTGGAATCTG
CGCGGCGCTCACCCAGCTTTCCTATGCAGAGTGGCCATCGTGGTGGGCGCCCCGCGGACCCTGGGCCCCAGCCAGGAG
                  exon 2 (3660) v a i v v g a p r t l g p s q e

GAGACGGGCGGCGTGTTCCTGTGCCCCTGGAGGGCCGAGGGCGGCCAGTGCCCCTCGCTGCTCTTTGACCTCCGTGAG
 e t g g v f l c p w r a e g g q c p s l l f d l (3782)

TCCCAGGCAAGGAGAGCAAGGTTGGGGTCAGAGGGACGTGGACTGCCCGGGCTTCAGCGCCCCACCCCTTCTTGTGCC
                                                          exon 3 (3873)

TTCCAGGTGATGAGACCCGAAATGTAGGCTCCCAAACTTTACAAACCTTCAAGGCCCGCCAAGGACTGGGGGCGTCGG
     r d e t r n v g s q t l q t f k a r q g l g a s

TCGTCAGCTGGAGCGACGTCATTGTGGTGGGCCCCGCGGTACAGGGCACAGGGAACAATCGGGGGCAGGGACACTGGG
v v s w s d v i v (3970)

GCCAGGAGGAGCCCAAGTCTCGCGCCCCGTCCCCATCTGTGGCCCTTTCTCAGGCCTGCGCCCCCTGGCAGCACTGGA
                                        exon 4 (4007) a c a p w q h w

ACGTCCTAGAAAAGACTGAGGAGGCTGAGAAGACGCCCGTAGGTAGCTGCTTTTTGGCTCAGCCAGAGAGCGGCCGCC
n v l e k t e e a e k t p v g s c f l a q p e s g r

GCGCCGAGTACTCCCCCTGTCGCGGGAACACCCTGAGCCGCATTTACGTGGAAAATGATTTTAGTAAGCGCCAGCTAC
r a e y s p c r g n t l s r i y v e n d f . (4171)

GACCTGGCCCCGCCCACTCGCGACGGCTTGGCCCCGCCCCCCATCGGATCCCGCCCCCAGCGCCGCAGCCCTTGCTTT
GGATCTGGCCTCGCCCCAGGGCCCCGCCGACTCAAGGCCCCGCCCCTGTCCCCCAGCCCTCCTCCGGGCTCGCGCGCG
CCTCCCTTCACCCCTGGGCTGACCCCTCCTCCTTGTCTCCTCAGGCTGGGACAAGCGTTACTGTGAAGCGGGCTTCAG
                                exon 5 (4421) s w d k r y c e a g f s

CTCCGTGGTCACTCAGGCGAGTAGGGAGCAAAAGCGCAGTGGGGGCGGCTCCCAAACAGGGCCCCCTCTCACCCTCAG
  s v v t q  (4471)

GACTTCCCTTCCAGGCCGGAGAGCTGGTGCTTGGGGCTCCTGGCGGCTATTATTTCTTAGGTACGTGCCCATCCGTAC
 exon 6 (4545) a g e l v l g a p g g y y f l (4591)

ACCTCCCTCCCTTCTCGCGGCCGAAGGAGACCGCTTTGGGCTTCACACCCGCTGTCCCTCCCGCCCTAGGTCTCCTGG
                                                  exon 7 (4589) g l l

CCCAGGCTCCAGTTGCGGATATTTTCTCGAGTTACCGCCCAGGCATCCTTTTGTGGCACGTGTCCTCCCAGAGCCTCT
a q a p v a d i f s s y r p g i l l w h v s s q s l

CCTTTGACTCCAGCAACCCAGAGTACTTCGACGGCTACTGGGGTAACACCGCCATTCCAGACTTCCAGCACCCCGAGG
s f d s s n p e y f d g y w (4719)

GTCACCGCCCACCGCAGACGGTCAGGTCCTGCCCCTGTGGGAGCCTCCATGGCCACCCCTGCCGGCCAACCCACCGCC
TAAGCCGCTCCCGCCCTCCGCTCCTGCGCTTCCCCGCAGACCGCCCACCTCCCATGCGCCCACCGCTCCCTTCCACTG
CGGACTCGTAGCGCAGCCTGGGGCAGGGCTTGGCCCCTCGAAGGCCTCCGTTTTTCCATCTGCACAATGCAGGGCTGG
GGCTGAGTGGCCTTAATCTCCTCCTTCTTTGCCCTCCGTCCCCTCTGTGCTTCCTCCCCTGGAAAAGACTAATTTGCG
                                                          exon 8

CCCTTGTCCTCAGGGTACTCGGTGGCCGTGGGCGAGTTCGACGGGGATCTCAACACTACAGGCAAGAAATCCACTTAG
     (5080) g y s v a v g e f d g d l n t t (5128)

GGCGGGAGTTGGGTAGCCCAGCCCGGGGAGGAGCGCCTTCCTGAAATCTCCCCTATGTAGCTGGGTGCAGAACGGGGA
GCGGGAAGTGGGTAGGTTCTAAGGCTCTCATTCCCTGAGCCTGGCTCTCCCTATCGCCAGAATATGTCGTCGTGCCCC
                                        exon 9 (5283) e y v v g a

CCACTTGGAGCTGGACCCTGGGAGCGGTAAGTGCCCCCACCACTGGGCCTCCCGAAGCCCCTTATCCCAGTTCTCAGG
p t w s w t l g a (5327)

CTGACAACTCCTGAGCGCCCCCCACCCCCGCCCCGCCTCCACCAAACCACCCTTTCTCACCTGGAGTGGGAGGTTGCT
TTGGGTACAAGAATGATGCTCTCGCCTGCGCTGTCCGTGCAGGTGGAAATTTTGGATTCCTACTACCAGAGGCTGCAT
                          exon 10 (5499)  v e i l d s y y q r l h

CGGCTGCGCGGAGAGCAGGTGGGGGCCAGGTCCCAGTGGGCGTGGCTGGGTGGAGGGGGAACTGAGACTTCAGAATAT
 r l r g e q (5553)

TTCATGGGAGGTGAGGGCCCATTTCTTAAAGAGGATGCTTGTCCAGCGGCGTGAATGATGGTGCTCCTCATCTTGCAG
                                                    exon 11 (5691)
```

```
ATGGCGTCGTATTTTGGGCATTCAGTGGCTGTCACTGACGTCAACGGGGATGGGTGAGGAGGGACATGCCCCCACCCC
 m  a  s  y  f  g  h  s  v  a  v  t  d  v  n  g  d  g  (5744)

TACCCAGTTGGGTCCCAAATTACCAGAGCTGCCCCTCTGTCTCCCTTTCCTAGCCCTAGTCTCACGTATCCACTGGAG
GAACAGGAGAGCAAGGGTCGAGGAGATTTGGCCCTAGCCCCAATATACCCCTGGTCCAGTCCCATGTAACCACTCATC
                                                                        exon 12

TGGCCCACAGGAGGCATGATCTGCTGGTGGGCGCTCCACTGTATATGGAGAGCCGGGCAGACCGAAAACTGGCCGAAG
    (5935)  r  h  d  l  l  v  g  a  p  l  y  m  e  s  r  a  d  r  k  l  a  e

TGGGGCGTGTGTATTTGTTCCTGCAGCCGCGAGGCCCCCACGCGCTGGGTGCCCCCAGCCTCCTGCTGACTGGCACAC
v  g  r  v  y  l  f  l  q  p  r  g  p  h  a  l  g  a  p  s  l  l  l  t  g  t

AGCTCTATGGGCGATTCGGCTCTGCCATCGCACCCCTGGGCGACCTCGACCGGGATGGCTACAATGGTGAGGGAAGAG
q  l  y  g  r  f  g  s  a  i  a  p  l  g  d  l  d  r  d  g  y  n  (6145)

AGGAGCCCTACTTGCTGCAGAGGGGTTAACAGCCACTCAAAAAGCATGGAGTTGGCCTGAGGGCAGCCAGAACCAGGA
TGGGTTTTAAGCATATAAGTATGTGGCTTAGACACATGGGGTGCTGAGTGGAGAGCAGATGGGAGAGTTGAAGACTAA
TTAGGAAGTGTTTGCCTTAATCCAAGCAAGAGACAATGACCACCTGGATGTGGATTTTGGCAGTGGAGTTAGAGATGG
GAGTGACTTCACAGATATTTAGGACTCGGATTATTAGGACTTGGTGGGAGACTGGATGTGGGGCCAGGGGAGAGGTTG
GAGTTGGGTGCCTGTGATGGCCTCCACTGCCTGGAACTCAGGCCGTGCAGCAGGTGCTGGGGAGAGGCGGGAGATCAG
CAGTTCAGCTCTGGACCTGTTGAGCTTGAAGGGCTTGGGTGCTTTAGGCGGAAATATCCAAAGAACAGTTGGGAGTGG
CTCTCCCCGCTTCCACAAGAGAGATCTGAATGGGAGACAGGGGTTTGGGGAAAGTGGATGAGGTCCCGGGACCTGTGA
AATAAGAGGCCCAGGATAGAGCCCTAGGGAGCAAAAGCATTTAGGTGACTCCTACAGGAGGTAAGTCTGAGAAGGAGA
CAGAGGAGTGTCCAGAGAGGGAGGAGGGAACCCAGGGGGTCTGATGGCCCGGGACTCAAGGAAGAGCATGCGTTAAAG
AGCATGCACAGGAGGAAGTGGGCGCTGCAGCTCCTGCTGCTGCTGCAAGATACAATTAGGTGGGGCTGGAGAAATATT
CATGGGCTTTAGCAAGAAGAGGGTGCCAGGCATGGTGGCTCATACCTGTAATCCCAGCTACTTGGGAAATTGAAGCAG
                                                alu----------
GAGAATCTCTTGAACCCGGGAAGTGGAGGTTGCACTGAGCTGAGCTTGCGCCACTACTGCACTCCAGCCTGGGTGACA
-------------------------------------------------------------------------------
GAGCAAGACTCCATCTCAACAAAATAAAAAAAAAAATAGAGAAAGAAAGGAAGAAAGAAAAAAGAAGGGGAGGTTATT
------------------------------------------------------------>
GGTGACAGTGACATAAATTGATTCAGGCCAAGATAGGGTCAGAAGCCAGAATGCAATGGGGTAAGGTATGAATGGAGA
TGAAAAATTGGATGCAGCTAATGTAGACAGCTCTTTCAACAGGTTTGTGGTAAAAAGGAATTTGAGGAATAGAAAGGA
AAAAAAAAAACATGTTTGACTATAAGAGGAAAAAGAGAAAAGGTGATCACAGAAAAGAGATGAGGGTCAAGGGAAGAT
TATTTCAATGTGGAAGAACATGTAGTAGGTTGAAAATGATGTTGTGGGGAAATGGGGGGATGAGCCAGCAGAGAGTCC
CTGTGATGCCTCAGGGGGTGGGAGGGTGACTGGCCCAGTGTCAGGGTGAAGGAAGGAAACCTCTTCCAGGGTCAAATG
GGGAAAGGGAAAAAGAAAGTTGGTGTGGGATTATAGCATAACAGTGGGCTGCCTCTCTTCCTGAAGTAAGAGATTACG
TCACCTGCTGAAGGAAGTGTGGGGGGTCTGGGAGTTTGATGGAATGGAGAAGGCTAGAAATAGATGCTAGATGGCCAG
GCACGGTGGCTCACACCTGGAATCCCAGCACTTTGGGAGGCCGAGGCAGGAGGATCACTGGAGCCTAGGAGTTTGACA
alu-------------------------------------------------------------------------
CCAGCCTGGCCAACATAGGGAGATCTCGTCTCCATAAAAATTTTTAAAAATTAGCTGGGCATGGTGGCTATAGTCTCA
---------------------------------------------------->
ACTGCTTGGGAAGCTGAGGTGGGAGGATTGCTTTAGTCCAGAAGGTTGAGGCTGCAGTAAGCCATGGTTGCACCACTG
CACTTCAGCCTGAATGACAAGTGCAAGACTGTCTTAAAATAAAAAATTTAAAGGGCTTGGGCACGGTGGCTCACACCT
                                                          alu-------------
GTAATCCAGCACTTTGGGAGCCCAAGGTGGGCAGATCACTTGAGGTCAGGAGTTCGAGATCAGCCTGGCCAATGTGGT
-------------------------------------------------------------------------------
GAAACCCCGTCTCTACTGAAAATACAAAAATTAGCCGGGCATGGTGGTAGGCGCCTGTAATCCCAGCTACTGAAGAGG
-------------------------------------------------------------------------------
CTGAGGCACAAGAATCACTTTAACGGGGGAGGCAGAGGTTGCAGTGAGCCGAGATCGCACCACTGCACTCCAGCCAGG
-------------------------------------------------------------------------------
ACAACAGAGCGAGACTCCATCTCAAAAAAAAAAAAATTTAGAAAAGGGAATAATGATGCTTAATTTTCAGGATATATT
------------------------------------------>
TTCCTCAATAGACAGTGAGAGTTGTCACTGTTTTTATAACAATCCTACTTGGCAGGTCCCTCTCCCACCTGATTGTTA
ACTCCTGGAGGGTAGGGCAGTGCCTCCTTCACCCACACTTTGCACCCCTTTCCTAGTCTCCTGGGATGTTCCCAGAGA
AGCTCAGGAAAGTTTTACAGTCATCTAGGGAGGCTGAATAACAATCAGCCACTTCCTTTCTGTTACTCCTTCCAGACA
                                                          exon 13 (8577) d

TTGCAGTGGCTGCCCCCTACGGGGGTCCCAGTGGCCGGGGCCAAGTGCTGGTGTTCCTGGGTCAGAGTGAGGGGCTGA
i  a  v  a  a  p  y  g  g  p  s  g  r  g  q  v  l  v  f  l  g  q  s  e  g  l

GGTCACGTCCCTCCCAGGTCCTGGACAGCCCCTTCCCCACAGGCTCTGCCTTTGGCTTCTCCCTTCGAGGTGCCGTAG
r  s  r  p  s  q  v  l  d  s  p  f  p  t  g  s  a  f  g  f  s  l  r  g  a  v

ACATCGATGACAACGGATACCCAGGTGCCCTGGACTGCCTCCAGCTAGAAATGCCCAAGAAAGGCCCTTGGACATTCG
d  i  d  d  n  g  y  p  (8760)

CTGGAAGTGCCAAGAGACACGGCCAGGGCTCATGCCTGGCCTGGTGTCCCACTATGGACTGCCAGAGGGGCTGGGTGA
AACCTCCAGTGGGGGAGGTGGTGTGGGGAACCCCTGGGAAGATGAGATGAGGATCCCCATACCCTAATCGCCAATTCT
GACCCATTCCTCGATGTCTATAGACCTGATCGTGGGAGCTTACGGGGCCAACCAGGTGGCTGTGTACAGGTGAGCACT
       exon 14 (8993) d  l  i  v  g  a  y  g  a  n  q  v  a  v  y  r  (9039)

GGCTCCAGGGGCGGGATGGGGAAGGTCCTGTGCCATCAAGAGGAGGCCAGGCCAGGAGGAGCCACAATGGCAAGCCTA
CCCCATCACCCTATCCCATCAGAGCTCAGCCAGTGGTGAAGGCCTCTGTCCAGCTACTGGTGCAAGATTCACTGAATC
     exon 15 (9158)   a  q  p  v  v  k  a  s  v  q  l  l  v  q  d  s  l  n

CTGCTGTGAAGAGCTGTGTCCTACCTCAGACCAAGACACCCGTGAGCTGGTGAGGAGGCAGAGGGCATGGGCCTTAAA
p  a  v  k  s  c  v  l  p  q  t  k  t  p  v  s  c  (9262)

GGATCTGGGACCTCAGAAAGGCTCCAACCCCTGAGCCCCACTTACGTCTTTGCAGCTTCAACATCCAGATGTGTGTTG
                                          exon 16 (9346)   f  n  i  q  m  c  v
```


GAGCCACTGGGCACAACATTCCTCAGAAGCTATGTGAGTGGCATGAAGGGGGCAGGAGGGAGGTGGGCTTGGACTCCC
g a t g h n i p q k l (9402)
CCGGAGGCTGGCCAGGGAGGTCCTGACTCTTCTGCTTGCCCTGCCAGCCCTAAATGCCGAGCTGCAGCTGGACCGGCA
exon 17 (9494) s l n a e l q l d r q
GAAGCCCCGCCAGGGCCGGCGGGTGCTGCTGCTGGGCTCTCAACAGGCAGGCACCACCCTGAACCTGGATCTGGGCGG
k p r q g r r v l l l g s q q a g t t l n l d l g g
AAAGCACAGCCCCATCTGCCACACCACCATGGCCTTCCTTCGAGTACGCCCAGGCAGGGGATTGGCAGGGCTGGGAGA
k h s p i c h t t m a f l r (9646)
GTAGAACTTACCCACTGGACTTGTTCATCTAGCCCTGGGGCACTGAGCTGGGTGCTGTGAGTCCGGGGGTGGTCAGGA
CACAGGTGCCTACTGGCCAGGAGAAGGTGGGATGTGTATGGTAGCAAGATGGCCTGACTCTTGCCCCTGTCCTAGGAT
exon 18 (9834) d
GAGGCAGACTTCCGGGACAAGCTGAGCCCCATTGTGCTCAGCCTCAATGTGTCCCTACCGCCCACGGAGGCTGGAATG
e a d f r d k l s p i v l s l n v s l p p t e a g m
GCCCCTGCTGTCGTGCTGCATGGAGACACCCATGTGCAGGAGCAGGTAGGGACAGGCAGGGACAGGCCAGGGAGGTGC
a p a v v l h g d t h v q e q (9960)
AGGACCCCTGATAGCAAATCAGGATTAGGGTTAGTGCCAAGTCACAATGTAACCCCAAAACCTTGATGTCATTCCAAA
CCCTAATGAAAACCTCAAAATCCAGCCAGTCATGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGACCGAGGCAG
alu----------
GCAGATTGCCTGAGGTCAGGAGTTAGAGACCAACCTGGCCAACATGGTGAAAACCCATCTCTACTAAAAATACAAAAA

AAATTAGCCGGGTGTGGTGACGCATGCCTGTAATTCCAGCTACTCGGGAGGCTGAAGCAGGAGAATCACTTGAACCCA

GGAGGCAGAGGTTGCAGTGAGCCAAGAGTGTGCCACAGCACTCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAA

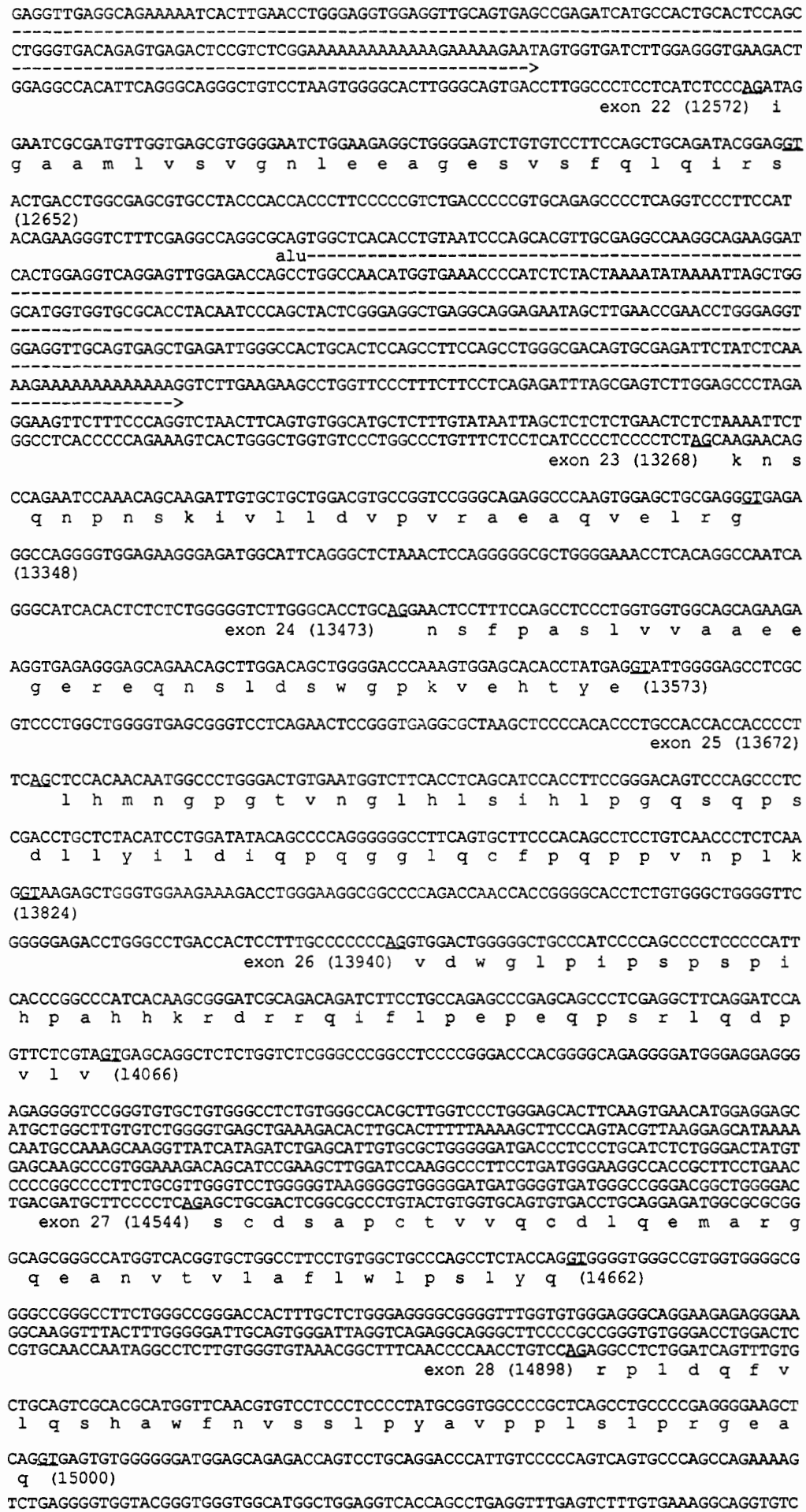
AAAAAAAAAAAAGCCAGGCGCAGTGGCCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGATCACGA

GGTCAGGAGATCAAGACCATCCTGGCTAACACAGTGAAACCCCGTCTACTAAAAATACAAAAAAAAAAAAAAAATTAG
---> alu-----
CTGGGCGTGGTGGCGGGTACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCCGGGGGCG

GACGTTGCAGTGAGCCGAGATAGTGCCACTGCACTCCAGCCTGGACGACAGAGCGAGACTCCGTCTCCAAAAATAAAA

AAACACCTGAAAATCCCAGTATCCCCTAAGCTCTGATGTAAATTGACAAACCCTGACATTGTCCCAAACCTCCAAATA
--->
TAACCCGAGCCCCGATACCATCTACAAACTCCTTTTCGTCCTCAGATCTTCTTACTCCCTAAGCCCCTATGTGAACCC
CAAGCCCACTGTTTTCCTAACCCTGATGTAATCCCTAAACCTCACACATCCCCAACTTACCCGCACACCCCAATGTGC
exon 19 (10931)
CCCTCTAGACACGAATCGTCCTGGACTGTGGGGAAGATGACGTATGTGTGCCCCAGCTTCAGCTCACTGCCAGCGTGT
t r i v l d c g e d d v c v p q l q l t a s v
GAGGAGGCCTCCCATTCTGCCCGACCCTGGCCCTTTCTGCCTATCATACCTGCTCCACACCTTAGTCCCCTCTTTTCC
(10999)
CACATCCTGGGCCCAGACCCAGGCTCCCTGGCTTCACTCCTCTTTCCCCACAGGACGGGCTCCCCGCTCCTAGTTGGG
exon 20 (11133) t g s p l l v g
GCAGATAATGTCCTGGAGCTGCAGATGGACGCAGCCAACGAGGGCGAGGGGGCCTATGAAGCAGAGCTGGCCGTGCAC
a d n v l e l q m d a a n e g e g a y e a e l a v h
CTGCCCCAGGGCGCCCACTACATGCGGGCCCTAAGCAATGTCGAGGTATGGCCCCCACCCTGGGAACAGTACCCGGGA
l p q g a h y m r a l s n v e (11281)
CCTGGGAGGCACTGGAGCCTTGGCTCTCTCATCTCCCTCCCTGAGAGTCCCTCTTCTCTTCTGCTTTGCTGTCAAAGA
TGTAATTTTTTTTTTTAATTTGGAGGAGGATACTTGCTAATGGTCAGTCAGAATTCCAAAACTCTATTACAAAAACCAG
AAAAACAAAAAAGGTTTAGGAACCAAATGTTAACAGGAACCTCTGTTAACATTTGGTGGATTTCCTTCCAGTCTTTTT
TTCAATATTGACTCACACTCACATAAGTATATATTTATTTTTTATGTTGTTAATATAGTTTATAATAATGGGGGTCAT
ACTCTAATGTTTTGTGTTTTTTATTTCCAAAATGAAAATGCCTAAAAAGTAGTAGTGCTACAGCAATACACACACTAG
CATGTGACAGTCCCTTGAGCGACCCCACCCCAAGAAACCCCCCCCTCCCTACCTTGGCACACAAATCTTTCCAGACCT
TCCAAGGGAGCTTAAATATATATATATGATGCTCTGTAATTTCTTTCTTGGAACTGCCTTCCTGAAGGGCTT
exon 21 (11869) g f
TGAGAGACTCATCTGTAATCAGAAGAAGGAGAATGAGACCAGGGTGGTGCTGTGTGAGCTGGGCAACCCCATGAAGAA
e r l i c n q k k e n e t r v v l c e l g n p m k k
GAACGCCCAGGTGAGGCTGCTGGGTCGTGGTACCGGGTCTCCACCAGGGGCTCATGAATAACCAGATTTTAGGGGTGA
n a q (11962)
GGTTTTAGAGCCACATAGTTCTGGGCCAGAATCTTGGTCCTCACACTCCCTTTGCCAACATTGTCCTTGGGTGAGTGA
CTTTCCCTCTCTGAGCCCCTTTACCAGTGGGCTTCCAGGTAAAATAGAAATAATAATGGTGGCCTGGTGCGGTCGTCA
alu----------
CGCCTGTAATCCCAGCACTCTGGGAGGCCAGAGCGGGTGGATCACGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACA

TAGCAAAACCCCGTCTCTACTAAAAATACAAAAATTACCCGGGCATGGTGGCGCACGCCTATAGTCAGAGCTACTCGG


GAGGTTGAGGCAGAAAAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGC
CTGGGTGACAGAGTGAGACTCCGTCTCGGAAAAAAAAAAAAAGAAAAAGAATAGTGGTGATCTTGGAGGGTGAAGACT
GGAGGCCACATTCAGGGCAGGGCTGTCCTAAGTGGGGCACTTGGGCAGTGACCTTGGCCCTCCTCATCTCCCAGATAG
exon 22 (12572) i
GAATCGCGATGTTGGTGAGCGTGGGGAATCTGGAAGAGGCTGGGGAGTCTGTGTCCTTCCAGCTGCAGATACGGAGGT
g a a m l v s v g n l e e a g e s v s f q l q i r s
ACTGACCTGGCGAGCGTGCCTACCCACCACCCTTCCCCCGTCTGACCCCCGTGCAGAGCCCCTCAGGTCCCTTCCAT
(12652)
ACAGAAGGGTCTTTCGAGGCCAGGCGCAGTGGCTCACACCTGTAATCCCAGCACGTTGCGAGGCCAAGGCAGAAGGAT
alu
CACTGGAGGTCAGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAATATAAAATTAGCTGG
GCATGGTGGTGCGCACCTACAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATAGCTTGAACCGAACCTGGGAGGT
GGAGGTTGCAGTGAGCTGAGATTGGGCCACTGCACTCCAGCCTTCCAGCCTGGGCGACAGTGCGAGATTCTATCTCAA
AAGAAAAAAAAAAAAAAGGTCTTGAAGAAGCCTGGTTCCCTTTCTTCCTCAGAGATTTAGCGAGTCTTGGAGCCCTAGA
GGAAGTTCTTTCCCAGGTCTAACTTCAGTGTGGCATGCTCTTTGTATAATTAGCTCTCTCTGAACTCTCTAAAATTCT
GGCCTCACCCCCAGAAAGTCACTGGGCTGGTGTCCCTGGCCCTGTTTCTCCTCATCCCCTCCCCTCTAGCAAGAACAG
exon 23 (13268) k n s
CCAGAATCCAAACAGCAAGATTGTGCTGCTGGACGTGCCGGTCCGGGCAGAGGCCCAAGTGGAGCTGCGAGGGTGAGA
q n p n s k i v l l d v p v r a e a q v e l r g
GGCCAGGGGTGGAGAAGGGAGATGGCATTCAGGGCTCTAAACTCCAGGGGGCGCTGGGGAAACCTCACAGGCCAATCA
(13348)
GGGCATCACACTCTCTCTGGGGGTCTTGGGCACCTGCAGGAACTCCTTTCCAGCCTCCCTGGTGGTGGCAGCAGAAGA
exon 24 (13473) n s f p a s l v v a a e e
AGGTGAGAGGGAGCAGAACAGCTTGGACAGCTGGGGACCCAAAGTGGAGCACACCTATGAGGTATTGGGGAGCCTCGC
g e r e q n s l d s w g p k v e h t y e (13573)
GTCCCTGGCTGGGGTGAGCGGGTCCTCAGAACTCCGGGTGAGGCGCTAAGCTCCCCACACCCTGCCACCACCACCCCT
exon 25 (13672)
TCAGCTCCACAACAATGGCCCTGGGACTGTGAATGGTCTTCACCTCAGCATCCACCTTCCGGGACAGTCCCAGCCCTC
l h m n g p g t v n g l h l s i h l p g q s q p s
CGACCTGCTCTACATCCTGGATATACAGCCCCAGGGGGGCCTTCAGTGCTTCCCACAGCCTCCTGTCAACCCTCTCAA
d l l y i l d i q p q g g l q c f p q p p v n p l k
GGTAAGAGCTGGGTGGAAGAAAGACCTGGGAAGGCGGCCCCAGACCAACCACCGGGGGCACCTCTGTGGGCTGGGGTTC
(13824)
GGGGGAGACCTGGGCCTGACCACTCCTTTGCCCCCCCAGGTGGACTGGGGGCTGCCCATCCCCAGCCCCTCCCCCATT
exon 26 (13940) v d w g l p i p s p s p i
CACCCGGCCCATCACAAGCGGGATCGCAGACAGATCTTCCTGCCAGAGCCCGAGCAGCCCTCGAGGCTTCAGGATCCA
h p a h h k r d r r q i f l p e p e q p s r l q d p
GTTCTCGTAGTGAGCAGGCTCTCTGGTCTCGGGCCCGGCCTCCCCGGGACCCACGGGGCAGAGGGGATGGGAGGAGGG
v l v (14066)
AGAGGGGTCCGGGTGTGCTGTGGGCCTCTGTGGGCCACGCTTGGTCCCTGGGAGCACTTCAAGTGAACATGGAGGAGC
ATGCTGGCTTGTGTCTGGGGTGAGCTGAAAGACACTTGCACTTTTTAAAAGCTTCCCAGTACGTTAAGGAGCATAAAA
CAATGCCAAAGCAAGGTTATCATAGATCTGAGCATTGTGCGCTGGGGGATGACCCTCCCTGCATCTCTGGGACTATGT
GAGCAAGCCCGTGGAAAGACAGCATCCGAAGCTTGGATCCAAGGCCCTTCCTGATGGGAAGGCCACCGCTTCCTGAAC
CCCCGGCCCCTTCTGCGTTGGGTCCTGGGGGTAAGGGGGTGGGGGATGATGGGGTGATGGGCCGGGACGGCTGGGGAC
TGACGATGCTTCCCCTCAGAGCTGCGACTCGGCGCCCTGTACTGTGGTGCAGTGTGACCTGCAGGAGATGGCGCGCGG
exon 27 (14544) s c d s a p c t v v q c d l q e m a r g
GCAGCGGGCCATGGTCACGGTGCTGGCCTTCCTGTGGCTGCCCAGCCTCTACCAGGTGGGGGTGGGCCGTGGTGGGGCG
q e a n v t v l a f l w l p s l y q (14662)
GGGCCGGGCCTTCTGGGCCGGGACCACTTTGCTCTGGGAGGGGCGGGGTTTGGTGTGGGAGGGCAGGAAGAGAGGGAA
GGCAAGGTTTACTTTGGGGGATTGCAGTGGGATTAGGTCAGAGGCAGGGCTTCCCCGCCGGGTGTGGGACCTGGACTC
CGTGCAACCAATAGGCCTCTTGTGGGTGTAAACGGCTTTCAACCCCAACCTGTCCAGAGGCCTCTGGATCAGTTTGTG
exon 28 (14898) r p l d q f v
CTGCAGTCGCACGCATGGTTCAACGTGTCCTCCCTCCCCTATGCGGTGGCCCCGCTCAGCCTGCCCCGAGGGGAAGCT
l q s h a w f n v s s l p y a v p p l s l p r g e a
CAGGTGAGTGTGGGGGGATGGAGCAGAGACCAGTCCTGCAGGACCCATTGTCCCCCAGTCAGTGCCCAGCCAGAAAAG
q (15000)
TCTGAGGGGTGGTACGGGTGGGTGGCATGGCTGGAGGTCACCAGCCTGAGGTTTGAGTCTTTGTGAAAGGCAGGTGTC

```
AAGGTGACTGAGGAGACACGTGGGTTTGCCCCAGGTGTGGACACAGCTGCTCCGGGCCTTGGAGGAGAGGGCCATTCC
                    exon 29 (15187)  v  w  t  q  l  l  r  a  l  e  e  r  a  i  p

AATCTGGTGGGTGCTGGTGGGTGTGCTGGGTGGCCTGCTGCTGCTCACCATCCTGGTCCTGGCCATGTGGAAGGTGAG
  i  w  w  v  l  v  g  v  l  g  g  l  l  l  l  t  i  l  v  l  a  m  w  k

GTGTGAAGGACGGTGGAGTCCCCAGCGGGGCACAGGCTTGGCTCTGCCCTGCCTCACAGGGAGTCAAGGAGAGATGGT
(15304)

GGCCCACCCAAGTGGGTAATCCAGGGACCAGGGGTCTATGTCTCCACTATTAGAATGTCATTCTCGTCCAGGGGGGTG
                                                                          alu----
GCTCACACCTGTAATCCCAGCACTTTGGCAGGCAAAGCGTTTAGATCACCTGAGGTCAAGAGTTCGAGACCAGCCTGG
-------------------------------------------------------------------------------
CCAACATGGTGAAACCCCATCTCTACTAAAAATACCAAATTAGCCGGGCGTGTTGACACATGCCTGTAATCTCAGCTA
-------------------------------------------------------------------------------
CTCGGGAGGCTGAGGCAGTAGAATTGCATGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCACACCACTGCACT
-------------------------------------------------------------------------------
CCAGCTTGGGCAACAGAGCGAGCCTCCATCTCAAAAAAAAAACAAAAAAATAGAATGTCTTTCTCTAGTAGAGCAAAA
------------------------------------------------------>
GGCAAAACAAACACAAAAATGTCATTCTCCTGGGAACCCTTCCAGACACATACCACTGGAAAGGATAGCACCTGAAAT
TCTGAGGCCTTTAGACACCCCTGCCACCAAAAAGATTCAGAGGATATAGAGGGTATAGAGGGTGTAAGTCCTGCCTTC
AGGAATTCCTGGCTGGTCTCAAGGACAAGATGCACTTCTTCCTAGCCCTGCCCTTCCCCTTGAGTGAGGAAGAGGCCA
AGGATTGGTCTAGACCCTATTCCATACCTTCCTATGTGGCCCTGGAGGGTCACTCGCTCCTCTGCACCTGGAGGAGTC
TCAAGCACACTGAAGGGAAGACATGGTGCTTTTAGGGAAAACCACGCACTAGACCCACAATAATCAAATACATATCAT
CATATGCTCGAGTCATGCAGACACAAACTTCAGTATAAGAAAAATTCCAGGCTGGGCGTTGGTGGCTCACACCGGTAA
AATCCCAGCACTTTGGGAGGCCGAGGTGGG
<----------------------approximately 600 bp of intron ---------------------->
TTCTGCGCTGGTCCAGGGAGGTGCTCATATGCTAGCATACTTCCTCACATGTGCTCTGGGGCCAGCAAATCATCTGTA
TACCCTGACCTTGGCCCCCGTGTACCCCCAGGTCGGCTTCTTCAAGCGGAACCGGCCACCCCTGGAAGAAGATGATGA
                exon 30 (16975)  v  g  f  f  k  r  n  r  p  p  l  e  e  d  d  e

AGAGGGGGAGTGATGGTGCAGCCTACACTATTCTAGCAGGAGGGTTGGGCGTGCTACCTGCACCGCCCCTTCTCCAAC
  e  g  e ter (17035)

AAGTTGCCTCCAAGCTTTGGGTTGGAGCTGTTCCATTGGGTCCTCTTGGTGTCGTTTCCCTCCCAACAGAGCTGGGCT
ACCCCCCCTCCTGCTGCCTAATAAAGAGACTGAGCCCTG*ATGCTGAGCATGC
                                        (17219)
```

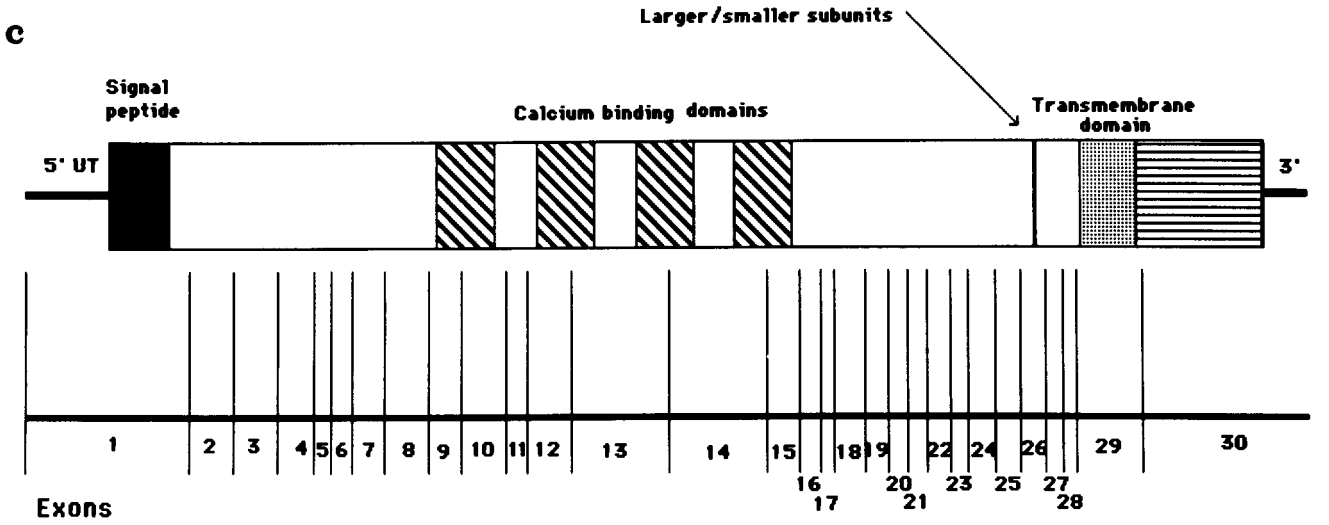


FIGURE 1: Organization of the human GPIIb gene. (a) The area encompassed by the two genomic clones (hGPIIb-5′ and hGPIIb-3′), the distribution of exons (solid boxes) and Alu repeats (arrows), the *Eco*RI/*Sal*I fragment used in defining the transcription start site (cross-hatched area), and the total areas sequenced (solid horizontal line) are indicated on the top three lines. Below is a restriction map derived from sequence analysis and Southern blot studies. Alu repeats are indicated as horizontal arrows under the top exon/intron diagram. The direction of the arrows indicates the orientation of the Alu repeat. (b) The 18.4-kb sequence of the GPIIb gene and its immediate 5′- and 3′-flanking regions is shown, with number +1 starting at the transcriptional start site. Exons are numbered at their first and last coding base, and the encoded amino acid is shown. Splice donor and acceptor sequences of each exon/intron border are underlined. Alu repeats are shown as a dashed-arrow underline. A 13-base region in exon 1 that concurs with the Alu repeat consensus sequence is dot underlined. The 30-mer oligonucleotide used in the primer extension and RNase protection experiments is doubly underlined in exon 1. Potential promoters are boxed in the 5′-UT region, and a pair of near-perfect inverted repeats is underlined by a pair of solid-line arrows. The asterisk after exon 30 refers to the polyadenylation site as previously determined from cDNA sequence analysis (Poncz et al., 1987). (c) Distribution of exons along the cDNA sequence is shown. The term "larger/smaller subunits" refers to the M_r 125 000 and 23 000 subunits into which mature GPIIb is cleaved during its intracellular processing.

antisense RNAs were hybridized with 20 μg of total cellular RNA from DMSO-induced HEL cells and then incubated with RNases A and T_1. RNase-resistant hybrids were denatured, and their sizes were determined on a denaturing urea–polyacrylamide gel.

S1 Nuclease Mapping of the Cap Site. Total HEL cell RNA was hybridized to radiolabeled antisense genomic DNA, essentially as described by Favaloro et al. (1980), except that

the probes were generated by primer extension of the 5′-GPIIb genomic 1.9-kb *Eco*RI/*Sal*I fragment in M13 phage with the universal primer, [α-^{32}P]dCTP, the three other dNTPs, and Sequenase (U. S. Biochemical Co., Cleveland, OH). Probes were fractionated on a Sephadex G-50 column and coprecipitated with 15 μg of total cellular RNA. After annealing overnight at 52 °C, heteroduplexes were digested with 3000 units of S1 nuclease (Bethesda Research Laboratory, Bethesda, MD) at 40 °C for 1 h. The products were resolved on a 6% (w/v) polyacrylamide sequencing gel and bands visualized by autoradiography.

Oligonucleotides and Labeled DNA Size Markers. Oligonucleotides for DNA sequence analysis, primer extension, and S1 nuclease studies were synthesized by the phosphoramidite technique (Beaucage & Caruthers, 1981) on an Applied Biosystems 380B DNA synthesizer. DNA size markers were T_4 polynucleotide kinase generated end-labeled fragments from a *Hae*III digest of ϕX174.

Computer Analysis. Data storage and computer analysis was done on a Macintosh SE microcomputer with DNA Inspector IIe program (TEXTCO, West Lebanon, NH). Alu repeats were found by comparison with the known Alu consensus sequence (Schmid & Jelinek, 1982). Initial alignment of the 5′ rat and human GPIIb flanking sequence was achieved by dot matrix analysis on the same program.

Results

Isolation and Sequence Analysis of Human Genomic GPIIb Clones. The 16.5-kb genomic clone hGPIIb-3′ was isolated after 1×10^6 plaques of the amplified human EMBL3A library were screened (Figure 1a). Restriction endonuclease mapping, Southern blot analysis with full-length GPIIb cDNA, and subsequent DNA sequence analysis demonstrated that hGPIIb-3′ contained the majority of the GPIIb coding region but did not encompass the 5′-UT region and the first 188 bp of the coding sequence. Rescreening of 3×10^6 recombinants using a 460-bp 5′-GPIIb cDNA fragment identified the 14-kb overlapping clone hGPIIb-5′, which contains additional 5′ sequences (Figure 1a).

The continuous DNA sequence of the GPIIb gene is shown in Figure 1b. The distance from the transcript cap site to the polyadenylation signal is 17.2 kb, of which the DNA sequence of 14.3 kb has been determined. Portions of two introns were not fully determined (between exons 1 and 2 and between exons 29 and 30). An additional 1.2 kb 5′ to the transcript cap site was also analyzed (Figure 1b). One difference from the published GPIIb cDNA sequence was detected (Poncz et al., 1987a): a G instead of a C was present at position 10951 of exon 19 in the genomic clone. This change represents an error in the original publication. This results in a change in amino acid 602 of the heavier subunit from a serine to a cysteine, increasing the number of cysteines in mature GPIIb from 17 to 18.

A restriction endonuclease map containing exon locations is shown in Figure 1a. This information is derived from a combination of restriction sites determined from sequence analysis and genomic blotting studies.

Analysis of the GPIIb DNA Sequence. (*A*) *Intron–Exon Organization.* The GPIIb gene consists of 30 exons ranging in size from 45 to 249 bp and introns ranging in size from 77 to approximately 2800 bp. There is no obvious relationship between the position of the exon/intron borders in the gene and proposed functional domains in the protein (Figure 1c). The signal peptide and part of the N-terminus of the mature protein are contained within the first exon; the site of cleavage between the heavier and lighter subunits is located within exon 26; the putative transmembrane domain is present in its entirety within exon 29 along with additional extracellular and cytoplasmic sequence; the 20 amino acid cytoplasmic domain and the 3-UT region are contained in exon 30; and the DNA sequences coding for the four putative calcium-binding domains are divided between exons 9 and 10, 12 and 13, 13 and 14, and 14 and 15.

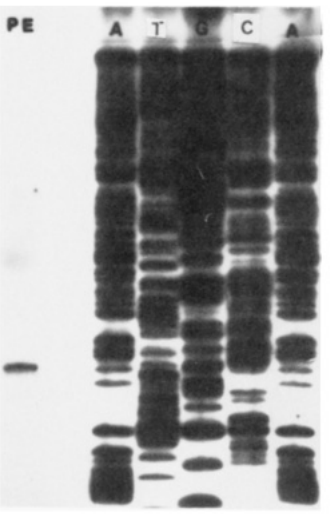


FIGURE 2: Primer extension analysis. In lane PE, a single band indicated by the arrow at 207 bp was seen as the primer extension product by use of total HEL cell RNA and a 30-mer oligonucleotide primer to a region in exon 1 (double underlined in Figure 1b). The DNA sequence ladder used as a size marker was derived from the same oligomer as a primer and the 5′-flanking *Eco*RI/*Sal*I GPIIb gene fragment as template.

(*B*) *Splice Junctions and Introns.* The exon/intron splice junctions conform to the previously described consensus sequence (Mount, 1982) with the exception of the donor sequences between exons 5 and 6 and 8 and 9, which are GC instead of GT. There are at least seven complete and three partial *Alu*I repeat sequences (shown as dashed arrow lines in Figure 1b and as arrows in Figure 1a) (Schmid & Jelinek, 1982). All are in the same 5′ to 3′ orientation and all are found within introns. Of interest, the coding region of the first exon contains a very partial Alu-like sequence, including a perfect 13-base match with the Alu consensus sequence (dot underlined in Figure 1b). This stretch includes the entire conserved CCAGCCTGG motif found in both rodent and primate Alu-like repeat sequences (Schmid & Jelinek, 1982).

(*C*) *Transcription Start Site.* The transcription start site was determined by three separate approaches. Primer extension studies using the 30-mer double underlined in exon 1 in Figure 1b revealed a single band of 207 nucleotides (n), consistent with a 5′-UT region of 32 n beginning with an adenine residue (Figure 2). The position of the transcription start site was aligned with the DNA sequence of the genomic 1.9-kb *Eco*RI/*Sal*I GPIIb fragment by use of the identical 30-mer as a primer for DNA sequence analysis.

RNase protection experiments using *Hin*dIII-linearized pGEM 3Z containing the 1.9-kb *Eco*RI/*Sal*I fragment (shown as the cross-hatched area in Figure 1a) as the transcriptional template annealed to HEL cell RNA demonstrated a major doublet of approximately 219–223 n (Figure 3). Similar studies with total platelet RNA demonstrated the same size RNase-protected bands (data not shown). Additional experiments using a pGEM 3Z construct from which the *Nco*I/*Eco*RI fragment (containing eight bases at the 3′ end

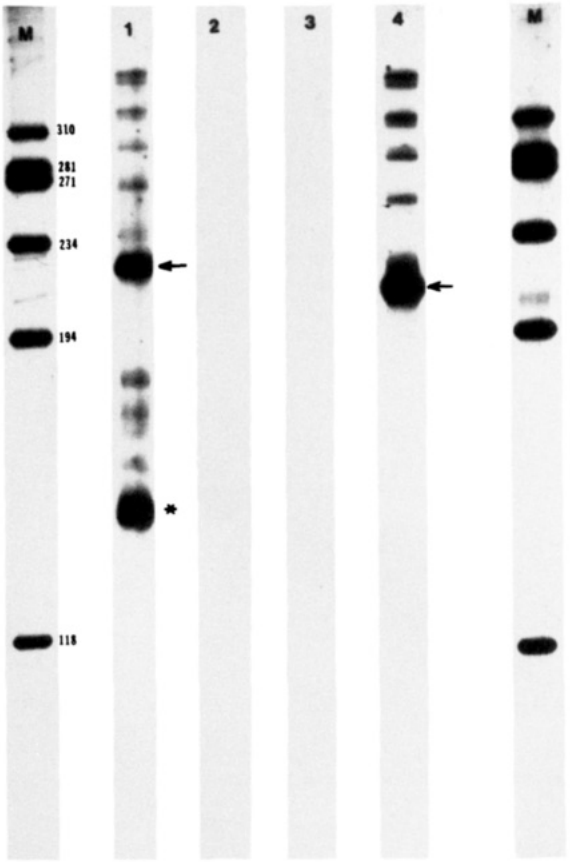


FIGURE 3: RNase protection experiment. RNase protection experiments using total RNA from HEL cells (lane 1), K562 cells (lane 2), and HL60 cells (lane 3) and the full-length *Eco*RI/*Sal*I 1.9-kb insert (cross-hatched area in Figure 1a) in pGEM 3Z that had been linearized with *Hin*dIII are shown. Lane 4 is identical with lane 1 except that the *Nco*I-truncated vector was used. The marker lanes (M) are T_4 polynucleotide kinase end-labeled ϕx174 *Hae*III fragments. The anticipated major bands are indicated by the arrows in lanes 1 and 4. The band indicated by the asterisk (*) is not reproducible.

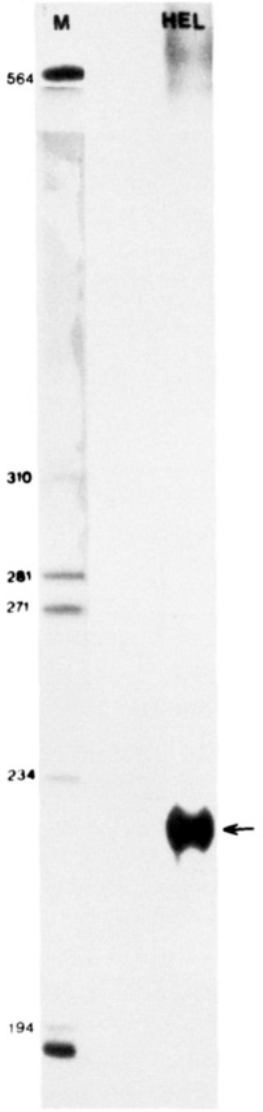


FIGURE 4: S1 nuclease protection. Autoradiograph of S1 nuclease protection using antisense DNA synthesized from the *Eco*RI/*Sal*I 1.9-kb insert in single-stranded M13 and total HEL cell RNA as described under Materials and Methods. The anticipated band is indicated by the arrow. The marker lane (M) represents end-labeled ϕx174 *Hae*III fragments.

of exon 1 and all of intron 1 within the 1.9-kb fragment) had been removed produced a doublet of approximately 208–213 n. These results are consistent with a 5′-UT region of approximately 28–35 n and agree with the primer extension study. The doublet present in these experiments likely represents incomplete digestion of the RNA heteroduplex by RNases.

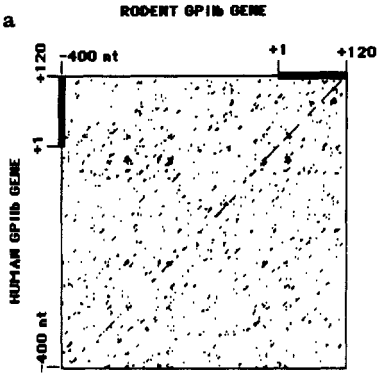
S1 nuclease experiments were performed with single-stranded M13 phage containing the same 1.9-kb *Eco*RI/*Sal*I fragment as a template for synthesizing [α-^{32}P]dATP-labeled antisense strand DNA. A major band of approximately 222–227 n was seen consistent with a 5′-UT region of 34–39 bases (Figure 4).

(*D*) *5′-Flanking Region.* Approximately 1.2 kb of the 5′-flanking sequence was contained within the 5′ genomic GPIIb clone. Inspection of the sequences immediately upstream from the proposed cap site reveals that there are no TATA or CAAT box sequences characteristic of RNA polymerase II transcribed genes (Breathnach & Chambon, 1981). The sequence ATAAGAAAA occurs at −54 and is interrupted by a single G. Such an interrupted TATA consensus sequence can be functional but at a reduced level (Poncz et al., 1982a; Surrey et al., 1985). There is a TATA sequence at −220, and sequences homologous to the CAAT element appear at −137, −234, −248, and −333. A potential SP1 binding site is also present beginning at −534 (Briggs et al., 1986). These potential promoter sequences are boxed in Figure 1b. A near-perfect (15/16 bases) inverted repeat is also present at positions −567 to −552 and −460 to −445, and these are underlined with a solid arrow in Figure 1b. The biological activity of these sequences needs to be explored further in functional assays.

(*E*) *Comparison with Rodent 5′-Flanking Region.* A dot matrix comparison of the 5′-flanking regions of rat and human GPIIb genes is shown in Figure 5a. Analysis demonstrates that there is homology between the two genes that extends for at least 340 n into the 5′-flanking region (Figure 5b). The first 90 n of the coding regions of the two GPIIb genes have approximately 72% homology, while the 5′-UT regions have 78% homology including identity around the proposed transcription start site from −2 to +8. There is about 65% homology in the 5′-flanking region to base −340. Of interest is that neither of the two possible TATA boxes mentioned above at bases −54 and −220 is present in the rat sequence, suggesting that neither is functional.

Discussion

Two genomic clones encompassing the entire GPIIb gene were isolated with the GPIIb cDNA as a probe. The gene for



b

```
                :ggg::a:acca:g:acca:tacaaccgtgg:cc:tggt::agta:g:cag:gc
                GCCAGATGGGGGAAGGGAGGAGATTAGAACTGTAGGCTAGAGTAGACAAGTATG

tgtg:a::::tgt:t::a::t::::gg::g:::::a:::::::::ca:-:::tt:::a:::ac::ag:::
GACCAGTTCACAATCACGCTATCCCAAGCAGAAAGTGATGGTGGCTTGGACTAGCACGGTGGTAGTAGAG

ga:aaaggggat:::tg:gcctg:ggc:::g:aa:t:::a:t:c:::::t:::ag:g-g:gct::tct**
ATGGGGTAAAGATTCAAGAGACATCATTGATAGGCAGAACCAATAGGACATGGTAATAAACTATTCTCAG

ag:gaa::ga::agca:::a:::c:g:t:::c::::-::tat:c:c:::acagac:ct::ct:::::a:t
GAAAGGGGAGGAGT-CATGGCTTTCAGCCATGAGCATCCACCCTCTGGGTGGCCTCACCCACTTCCTGGC

:g:c::::agc:gtgtg::::::a::::::tc::::c:::ca::::t:c::::a:::c:tgag:::----
AATTCTAGCCACCATGAGTCCAGGGGCTATAGCCCTTTGCTCTGCCCGTTGCTCAGCAAGTTACTTGGGG

-----agc ::::::ccg:a::::::g::::t::gga::tg:::::::::-:::-:::::t:a::::::
TTCCAGTTTGATAAGAAAAGACTTCCTGTGGAGGAATCTGAAGGGAAGGAGGAGGAGCTGGCCCATTC

::::t:a:::ca::tg::g:::::::::::::::::::::::ca:: g:gtg a:ta::::g:
CTGCCTGGGAGGTTGTGGAAGAAGGAAGATGGCCAGAGCTTTGTGTCCACTGCAAGCCCTCT
```

FIGURE 5: Comparison of human and rodent GPIIb 5′ sequence. (a) A dot matrix comparison of the 5′ region of rat and human GPIIb genes is shown. Exon 1 for both sequences is shown as thickened lines on the ordinate and abscissa. Numbering is from the transcription start site. (b) Direct sequence comparison is shown with rodent differences from the human sequence shown above the human sequence. The ATG triplet coding for the first amino acid is underlined. Potential TATA and CAT box sequences are underlined. Identity between the two sequences is indicated by a ":" in the rodent sequence. A "-" refers to a missing base in that sequence.

GPIIb spans approximately 17 kb on the long arm of chromosome 17 (Bray et al., 1987; Sosnoski et al., 1988) and contains 30 exons whose sizes range from 45 to 249 bp, a pattern typical of other eucaryotic genes (Traut, 1988). The sum total of the GPIIb exons from this genomic analysis and the previous cDNA work (Poncz et al., 1987) is 3334 bp, consistent with the 3.3-kb GPIIb band seen on Northern blots of HEL cell RNA (Poncz et al., 1987). One nucleotide difference from the published cDNA sequence was noted: a G is present at nucleotide 1899 instead of a C (Poncz et al., 1987a). Review of the cDNA sequence confirmed the nucleotide as G. The amino acid at codon 602 of the heavy chain is, therefore, a cysteine instead of serine at codon 602, indicating that mature GPIIb has 18, instead of 17, cysteine residues. This is consistent with peptide sequence analysis (Phillips, 1988). The 18 cysteine residues occur at homologous positions in the vitronectin and fibronectin receptor α subunits, suggesting similar overall protein folding.

An analysis of the 5′ and 3′ ends of the GPIIb gene has recently been published (Prandini et al., 1988). Our sequence differs from their genomic 5′-flanking sequence in that an extra T between −214 and −213 is present in their sequence. Whether this difference represents a polymorphism or sequencing error is not known.

Loftus et al. (1987) reported the isolation of a cDNA containing a region of GPIIb involved in adhesive function. Comparison of their sequence with ours shows identity of amino acids 38–218 in their sequence with amino acids 784–964 of the full-length GPIIb coding sequence. The 5′ and 3′ sequences bordering these amino acids in their sequence are not present within the GPIIb genomic sequence. The sequence 5′ to amino acids 38–218 of their sequence is a partial Alu repeat. Amino acid 38 corresponds to the start of exon 24. However, the sequence of the intron between exons 23 and 24 does not contain an *Alu*I repeat and has no similarity to this 5′ sequence. This Alu sequence most likely represents unrelated DNA fused to GPIIb cDNA. Amino acid 218 of their sequence terminates in the middle of exon 28. A comparison of the sequence 3′ to their partial GPIIb cDNA sequence with the GPIIb genomic sequence does not reveal any areas of identity or similarity. This 3′ sequence is also likely to represent unrelated DNA fused onto the GPIIb cDNA rather than unprocessed intronic sequence.

Gilbert (1978) and Blake (1978) proposed that exons may have originally represented separate functional or structural units of proteins and that they may have been used as cassettes to construct novel genes. A number of genes, such as that for the LDL receptor, appear to have been constructed from cassetted functional exons (Sudhof et al., 1985). For other genes, the division into exonic subunits that correlate with functional domains is not as clear. It has been suggested that while exons may originally have had separate functions, their borders have become obscured with time, because of either shifts from the original splice junctions or exon fusion or fission (Traut, 1988). Although many of the functional domains of GPIIb have yet to be determined, those identified have no relationship to exon structure. One possible exception is the distribution of the calcium-binding domains within GPIIb. Three of the four calcium-binding domains in the rat calmodulin gene have exon borders near their center similar to those

of GPIIb, suggesting that these domains may have been derived from a common functional unit (Nojima & Sokabe, 1986). Sequence analysis of genes for other integrin α subunits may provide further insights into the relationship between exons and biological function.

GPIIb exons 5 and 8 have GC instead of GT splice donor sequences. This rare splice donor sequence has been described in other genes, such as those of α-crystallin (King & Piatigorsky, 1987), adenine phosphoribosyltransferase (Broderick et al., 1987), and avian adult α-globin (Dodgson & Engel, 1983). Studies performed in vitro indicate that genes containing GC at splice donor junctions can be properly spliced (Padgett et al., 1986). Since the thalassemic patient from which the genomic library was constructed had no history of unusual bleeding and the gene frequency for Glanzmann's thrombasthenia is low, it is likely that these GC splice junctions are the normal sequence.

GPIIb contains at least seven complete and three partial *Alu*I repeats, representing approximately 20% of the total intron sequence analyzed. *Alu*I repeats are dispersed throughout the genome at a frequency of 1 per 10^4 bp (approximately 5–10% of the genome) (Schmid & Jelinek, 1982). However, a number of genes have a higher than expected number of *Alu*I repeats, including those for prothrombin (Degen & Davie, 1987), β-tubulin (Lee et al., 1984), thymidine kinase (Flemington et al., 1987), and adenosine deaminase (Wiginton et al., 1986). The reasons for increased frequency of these repetitive DNA sequences is not known, but they may potentiate gene deletions due to unequal homologous recombination events.

Unlike GPIIIa, which is part of the vitronectin receptor as well as of the platelet fibrinogen receptor (Ginsberg et al., 1987; Zimrin et al., 1988; Fitzgerald et al., 1987b) and which is found on many different differentiated cell lines, GPIIb expression is limited in vivo to the megakaryocyte lineage. As part of our studies to better understand the regulation of GPIIb expression, we have analyzed the 5′-flanking region of the GPIIb gene to determine the transcriptional start site by three separate approaches: primer extension, RNase protection, and S1 nuclease experiments. These studies are consistent with a 5′-UT region 32 n in length as also concluded by Prandini et al. (1988). The 5′-UT region is within the 20–100-n range for similar regions found in most vertebrate mRNA species (Kozak, 1987), and the sequence flanking the start site is consistent with those for other genes.

The absence of canonical TATA and CAAT boxes usually found in the immediate 5′-flanking sequence of RNA polymerase II transcribed genes is unusual but has been noted previously for several other nonhousekeeping genes (Bienz-Tadmor et al., 1985; Mignotte et al., 1989). Recently, a short sequence immediately surrounding the cap site of the murine deoxynucleotidyl transferase gene, which also lacks a TATA box in the expected location, was demonstrated to initiate accurate transcription, though at a low level (Smale & Baltimore, 1989). The sequence around both the rat and human GPIIb transcriptional start sites conforms not only to the known cap site consensus (Corden et al., 1980; Bucher & Trifonov, 1986) but also to this weak promoter element, supporting the thesis that the GPIIb gene may be able to function without a TATA box. A comparison of the human gene with the rodent GPIIb gene reveals that several of the potential upstream TATA boxes in the human sequence are not conserved in the rat, suggesting that these may not be functional. Further studies using functional expression assays are needed to confirm the role of this region in the promotion of expression of the GPIIb gene.

ACKNOWLEDGMENTS

We are grateful to Dr. Ann Zimrin for providing the HEL cell total RNA used in the primer extension experiments.

Registry No. DNA (human glycoprotein IIb gene coding region), 124316-27-4; glycoprotein IIb (human precursor protein moiety reduced), 124316-26-3; RNA (human clone hGPIIb-5′/hGPIIb-3′ glycoprotein IIb-specifying messenger), 124316-25-2.

REFERENCES

Aranout, M. A., Gupta, S. K., Pierce, M. W., & Tenen, D. G. (1988) *J. Cell Biol. 106*, 2153–2158.

Argraves, W. S., Pytela, R., Susuki, S., Millan, J. L., Pierschbacher, M. D., & Ruoslahti, E. (1986) *J. Biol. Chem. 261*, 12922–12924.

Beaucage, S. L., & Caruthers, M. H. (1981) *Tetrahedron Lett. 22*, 1859–1862.

Bennett, J. S., Vilaire, G., & Cines, D. B. (1982) *J. Biol. Chem. 257*, 8049–8054.

Bennett, J. S., Hoxie, J. A., Leitman, S. F., Vilaire, G., & Cines, D. B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2417–2421.

Bienz-Tadmor, B., Zakut-Houri, R., Libresco, S., Givol, D., & Oren, M. (1985) *EMBO J. 4*, 3209–3213.

Blake, C. C. F. (1978) *Nature 273*, 267.

Bogaert, T., Brown, N., & Wilcox, M. (1987) *Cell 51*, 929–940.

Bray, P. F., Rosa, J.-P., Johnston, G. I., Shiu, D. T., Cook, R. G., Lau, C., Kan, Y. W., McEver, R. P., & Shuman, M. A. (1987) *J. Clin. Invest. 80*, 1812–1817.

Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.

Briggs, M. R., Kadonaga, J. T., Bell, S. P., & Tijan, R. (1986) *Science 234*, 47–52.

Broderick, T. P., Schaff, D. A., Bertino, A. M., Dush, M. K., Tischfield, J. A., & Stambrook, P. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3349.

Bucher, P., & Trifonov, E. N. (1986) *Nucleic Acids Res. 14*, 10009–10025.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.

Collins, S., Gallo, R., & Gallagher, R. (1977) *Nature 270*, 347.

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

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., & Chambon, P. (1980) *Science 209*, 1405–1414.

Degen, S. J. F., & Davie, E. W. (1987) *Biochemistry 26*, 6165–6177.

Doggson, J. B., & Engel, J. D. (1983) *J. Biol. Chem. 258*, 4623.

Fitzgerald, L. A., Poncz, M., Steiner, B., Rall, S. C., Jr., Bennett, J. S., & Phillips, D. R. (1987a) *Biochemistry 26*, 8158–8165.

Fitzgerald, L. A., Steiner, B., Rall, S. C., Jr., Lo, S.-S., & Phillips, D. R. (1987b) *J. Biol. Chem. 262*, 3936–3939.

Flemington, E., Bradshaw, H., Traina-Dorge, V., Slagel, V., & Deninger, P. (1987) *Gene 52*, 267.

Frischauf, A. M., Lehrach, H., Poustka, A., & Murray, N. (1983) *J. Mol. Biol. 170*, 827–842.

Geliebter, J., Zeff, R., Melvold, R., & Natheson, S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3371–3375.

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

Ginsberg, M. H., Forsuth, J., Lightsey, A., Chediak, J., & Plow, E. F. (1983) *J. Clin. Invest. 71*, 619–624.

Ginsberg, M. H., Loftus, J., Ryckwaert, J., Pierschbacher, M., Pytela, R., Ruoslahti, E., & Plow, E. F. (1987) *J. Biol. Chem. 262*, 5437–5440.

Gitay, J. C., Locksma, O. C., Breederveld, C., & vanMourik, J. A. (1987) *Blood 69*, 809.

Hynes, R. O. (1987) *Cell 48*, 549–545.

Jennings, L. K., & Phillips, D. R. (1982) *J. Biol. Chem. 257*, 10458–10466.

King, C. R., & Piatigorsky, J. (1983) *Cell 32*, 707–712.

Kishimoto, T. K., Hollander, N., Roberts, T. M., Anderson, D. C., & Springer, T. A. (1987a) *Cell 50*, 193–202.

Kishimoto, T. K., O'Connor, K., Lee, A., Roberts, T. M., & Springer, T. A. (1987b) *Cell 48*, 681–690.

Kozak, M. (1987) *Nucleic Acids Res. 15*, 8125.

Lee, M. G. S., Loomis, C., & Covan, N. J. (1984) *Nucleic Acids Res. 12*, 5823–5836.

Loftus, J. C., Plow, E. F., Frelinger, A. L., D'Souza, S. E., Dixon, D., Lacy, J., Sorge, J., & Ginsberg, M. H. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7114–7118.

Lozzio, C., & Lozzio, B. (1975) *Blood 45*, 321–334.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marguerie, G. A., Plow, E. F., & Edgington, T. S. (1979) *J. Biol. Chem. 254*, 5357–5363.

Messing, J. (1983) *Methods Enzymol. 101*, 20–79.

Mignotte, V., Well, L., deBoer, E., Grosveld, F., & Romeo, P.-H. (1989) *Nucleic Acids Res. 17*, 37–54.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Nojima, H., & Sokabe, H. (1987) *J. Mol. Biol. 193*, 439–445.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119–1150.

Phillips, D. R., & Agin, P. P. (1977) *J. Clin. Invest. 60*, 535–545.

Phillips, D. R., Charo, I. F., Parise, L. V., & Fitzgerald, L. A. (1988) *Blood 71*, 831–843.

Poncz, M., & Newman, P. (1989) *Blood* (in press).

Poncz, M., Ballantine, M., Solowiejczyk, D., Barak, I., Schwartz, E., & Surrey, S. (1982a) *J. Biol. Chem. 257*, 5994–5996.

Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., & Surrey, S. (1982b) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4298–4302.

Poncz, M., Eisman, R., Heidenreich, R., Silver, S., Vilaire, G., Surrey, S., Schwartz, E., & Bennett, J. S. (1987a) *J. Biol. Chem. 262*, 8476–8482.

Poncz, M., Surrey, S., LaRocco, P., Weiss, M. J., Rappaport, E. F., Conway, T. M., & Schwartz, E. (1987b) *Blood 69*, 219–233.

Prandini, M. H., Denaier, E., Frachet, P., Uzan, G., & Marguerie, G. (1988) *Biochem. Biophys. Res. Commun. 156*, 595–601.

Rosa, J.-P., Bray, P. F., Gayet, O., Johnston, G. I., Cook, R. G., Jackson, K. W., Shuman, M. A., & McEver, R. P. (1988) *Blood 72*, 593–600.

Ruggeri, Z. M., Bader, T., & DeMarco, L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6038–6041.

Sanger, F., Nicklen, S., & Coulson, A. A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5468.

Schalenghe, F., Turco, E., Edstrom, J. E., Pirrotta, V., & Melli, M. (1981) *Chromosoma 82*, 205–216.

Schmid, C. W., & Jelinek, W. R. (1982) *Science 216*, 1065–1070.

Semenza, G. L., Delgrosso, K., Poncz, M., Malladi, P., Schwartz, E., & Surrey, S. (1984) *Cell 39*, 123–128.

Smale, S., & Baltimore, D. (1989) *Cell 57*, 103–113.

Sosnoski, D. M., Emanuel, B. S., Hawkins, A. L., van Tuinen, P., Ledbetter, D. H., Nussbaum, R. L., Kaos, F.-T., Schwartz, E., Phillips, D., Bennett, J. S., Fitzgerald, L. A., & Poncz, M. (1988) *J. Clin. Invest. 81*, 1993–1998.

Southern, E. (1975) *J. Mol. Biol. 98*, 503.

Sudhof, T. C., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1985) *Science 228*, 815–822.

Surrey, S., Delgrosso, K., Malladi, P., & Schwartz, E. (1985) *J. Biol. Chem. 260*, 6507–6510.

Suzuki, S., Argraves, W. S., Arai, H., Languino, L. R., Pierschbacher, M. D., & Ruoslahti, E. (1987) *J. Biol. Chem. 262*, 14080–14085.

Tabilio, A., Rosa, J.-P., Testa, U., Kieffer, N., Nurden, A., DelCanizo, M., Breton-Gorius, J., & Vainchenker, W. (1984) *EMBO J. 3*, 453–459.

Tabor, S., & Richardson, C. C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767–4771.

Tamkun, J. W., DeSimone, D. W., Fonda, D., Patel, R. S., Buck, C., Horowitz, A. F., & Hynes, R. O. (1986) *Cell 46*, 271–282.

Traut, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2944–2948.

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W., & Goodman, H. (1977) *Science 196*, 1313–1319.

Wiginton, D. A., Kaplan, D. J., States, J. C., Akeson, A. L., Perme, C. M., Bilyk, I. J., Vaughn, A. J., Lattier, D. L., & Hutton, J. J. (1986) *Biochemistry 25*, 8234.

Zimrin, A. B., Eisman, R., Vilaire, G., Schwartz, E., Bennett, J. S., & Poncz, M. (1988) *J. Clin. Invest. 81*, 1470–1475.

Zinn, K., DiMaio, D., & Maniatis, T. (1983) *Cell 34*, 865–879.