Laboratory Press, Cold Spring Harbor, NY.
Rafferty, J. B., Somers, W. S., Saint-Girons, I., & Phillips, S. E. V. (1989) *Nature 341*, 705–710.
Saint-Girons, I., Duchange, N., Cohen, G. N., & Zakin, M. M. (1984) *J. Biol. Chem. 259*, 14282–14285.
Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P., & Susskind, M. (1983) *J. Mol. Biol. 168*, 699–713.
Sorger, P. K., & Nelson, H. C. M. (1989) *Cell 59*, 807–813.
Vershon, A. K., Youderian, P., Susskind, M. M., & Sauer, R. T. (1985) *J. Biol. Chem. 260*, 12124–12129.
Vershon, A. K., Bowie, J. U., Karplus, T. M., & Sauer, R. T. (1986) *Proteins: Struct., Funct., Genet. 1*, 302–311.
Vershon, A. K., Liao, S.-M., McClure, W. R., & Sauer, R. T. (1987a) *J. Mol. Biol. 195*, 323–331.
Vershon, A. K., Liao, S.-M., McClure, W. R., & Sauer, R. T. (1987b) *J. Mol. Biol. 195*, 311–322.
Vershon, A. K., Kelley, R. D., & Sauer, R. T. (1989) *J. Biol. Chem. 264*, 3267–3273.
Zagorski, M. G., Bowie, J. U., Vershon, A. K., Sauer, R. T., & Patel, D. J. (1989) *Biochemistry 28*, 9813–9825.

Nucleotide Sequence of the Gene for the b Subunit of Human Factor XIII[†,‡]

Ralph E. Bottenus, Akitada Ichinose, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received May 22, 1990; Revised Manuscript Received August 10, 1990

ABSTRACT: Factor XIII (M_r 320 000) is a blood coagulation factor that stabilizes and strengthens the fibrin clot. It circulates in blood as a tetramer composed of two a subunits (M_r 75 000 each) and two b subunits (M_r 80 000 each). The b subunit consists of 641 amino acids and includes 10 tandem repeats of 60 amino acids known as GP-I structures, short consensus repeats (SCR), or sushi domains. In the present study, the human gene for the b subunit has been isolated from three different genomic libraries prepared in λ phage. Fifteen independent phage with inserts coding for the entire gene were isolated and characterized by restriction mapping, Southern blotting, and DNA sequencing. The gene was found to be 28 kilobases in length and consisted of 12 exons (I–XII) separated by 11 intervening sequences. The leader sequence was encoded by exon I, while the carbonyl-terminal region of the protein was encoded by exon XII. Exons II–XI each coded for a single sushi domain, suggesting that the gene evolved through exon shuffling and duplication. The 12 exons in the gene ranged in size from 64 to 222 base pairs, while the introns ranged in size from 87 to 9970 nucleotides and made up 92% of the gene. The introns contained four Alu repetitive sequences, one each in introns A, E, I, and J. A fifth Alu repeat was present in the flanking 3′ end of the gene. Two partial *Kpn*I repeats were also found in the introns, including one in intron I and one in intron J. The *Kpn*I repeat in intron J was 89% homologous to a sequence of approximately 2200 nucleotides flanking the gene coding for human β globin and approximately 3800 nucleotides from the L1 insertion present in the gene for human factor VIII. Intron H also contained an "O" family repeat, while two potential regions for Z-DNA were identified within introns G and J. One nucleotide change was found in the coding region of the gene when its sequence was compared to that of the cDNA. This difference, however, did not result in a change in the amino acid sequence of the protein.

Factor XIII is a zymogen that circulates in blood as a tetramer (a_2b_2) and is converted to an active transglutaminase by thrombin in the presence of calcium and fibrin [for reviews, see Lorand et al. (1980) and Folk (1983)]. The two pairs of a and b subunits separate during the activation reaction, giving rise to two activated peptides and factor XIIIa, an enzyme composed of two a′ subunits. Factor XIIIa catalyzes the formation of ε-(γ-glutamyl)lysine bonds between fibrin monomers as well as a number of other plasma proteins. The function of the b subunits is not clear, but it is likely that they protect and stabilize the plasma zymogen prior to activation (Folk & Finlayson, 1977). It may also play a role in the regulation of the contact activation pathway (Halkier & Magnusson, 1988).

Plasma levels of factor XIII have been determined by Yorifuji et al. (1988), who reported that all of the a subunits are in the tetrameric form, which is present in plasma at a concentration of 11 μg/mL. An excess of the b subunits also circulates in plasma in a monomeric form at a concentration of 10 μg/mL. Electron microscopy has shown that the tetrameric form of factor XIII appears as two globular a subunits surrounded by two flexible, rod-shaped b domains (Carrell et al., 1989).

The primary structure of both the a and b subunits has been determined by a combination of cDNA cloning and amino acid sequence analysis in our laboratory (Ichinose et al., 1986a,b) and those of others (Grundmann et al., 1986; Takahashi et al., 1986). The gene for the a subunit is over 160 kb in length (Ichinose & Davie, 1988) and has been localized to chromosome 6 at p24–25 (Board et al., 1988). The b subunit is 641 amino acids in length and is made up of 10 repeating domains (approximately 60 amino acids each) that have been called GP-I domains (Davie et al., 1986), short consensus repeats, or sushi domains (Ichinose et al., 1990). The term GP-I

[†]This work was supported in part by a research grant (HL 16919 to E.W.D.) and a National Research Service Award (HL 07843 to R.E.B.) from the National Institutes of Health.
[‡]The nucleotide sequence in this paper has been submitted to GenBank under Accession Number J05294.
*Author to whom correspondence should be addressed.

0006-2960/90/0429-11195$02.50/0

domain originated from β_2-glycoprotein I, a plasma protein with characteristic disulfide bonds between the first and third and the second and fourth Cys residues of each GP-I domain (Lozier et al., 1984). In the complement proteins these domains have been called short consensus repeats (SCR), while more recently they have been called sushi domains in factor XIII since they resemble the well-known Japanese sushi structures (Ichinose et al., 1990). The gene for the b subunit of factor XIII has been localized to bands q31–q32.1 of chromosome 1 (Webb et al., 1989) that are linked to the regulator of the complement activation (RCA) gene cluster in the same chromosome (Rodriguez-de-Cordoba et al., 1988). The genes of a number of the other proteins containing sushi domains have also been found in a similar region on chromosome 1 at band q32 (Morton & Burns, 1987).

In order to compare the gene for the b subunit of factor XIII with the codings of genes for other proteins in this family, it was important to determine its structure and organization. This knowledge may also provide some insight as to the evolution of these genes and their regulation and also make it possible to compare the normal gene with abnormal factor XIII genes. In the present study, we report the entire DNA sequence of the gene for the b subunit of human factor XIII including the 5′ and 3′ flanking regions.

Experimental Procedures

Restriction endonucleases, T4 DNA ligase, T4 kinase, bacterial alkaline phosphatase, M13mp18 and -mp19, pUC18 and -19, kilobase sequencing kits, and molecular biology grade reagents were purchased from Bethesda Research Laboratories. Restriction endonucleases and the Klenow fragment of *Escherichia coli* DNA polymerase were also purchased from Boehringer Mannheim Biochemicals and Promega. Klenow sequencing reagents, deoxynucleotides, and dideoxynucleotides were purchased from Pharmacia LKB Biotechnology Inc., while T7 DNA polymerase (Sequenase) was obtained from U.S. Biochemical Corp. ^{32}P-Labeled nucleotides were obtained from New England Nuclear, and [α-^{35}S]dATPαS was purchased from Amersham. Cyclone I, the rapid deletion subcloning system for M13 derivatives, was purchased from International Biotechnologies, Inc. Reagents for the polymerase chain reaction were purchased from Perkin-Elmer Cetus.

Three human genomic libraries were screened by using full-length as well as partial cDNAs coding for the b subunit of human factor XIII (Ichinose et al., 1986b). A human leukocyte library in EMBL3, obtained from Clontech Laboratories, a human fetal liver library in Charon 4A (Lawn et al., 1978), and a human fibroblast 5X chromosome library in EMBL3 (Yoshitake et al., 1985) were employed. The cDNA probes were used to screen approximately 2×10^7 phage by the plaque hybridization technique of Benton and Davis (1977) as modified by Woo (1979). Positive clones were plaque purified, and phage DNA was then prepared by the liquid culture lysis method (Maniatis et al., 1982).

Positive clones were analyzed by a combination of restriction mapping and Southern hybridization analysis (Southern, 1975). Nick-translation cDNA probes (Thomashow et al., 1980) were used to identify overlapping clones. Oligonucleotides were used as sequencing primers and as probes for selecting some restriction fragments for subcloning into pUC and M13. Two primers of 30 nucleotides in length were also used for amplifying a section of the DNA near exon X by the polymerase chain reaction (PCR), as described later. These were prepared by Patrick S. H. Chou and Yim Foon Lee or by Jeff Harris, employing a nucleotide synthesizer (Applied Biosystems, Foster City, CA). The PCR was carried out in a thermal cycler from Perkin-Elmer, with denaturation at a temperature of 94 °C, annealing at 45 °C, and extension at 72 °C for 30 cycles.

The phage DNA inserts were generated by *Sal*I and/or *Eco*RI restriction digestion and the appropriate fragments subcloned into pUC18 or pUC19 for amplification. Other restriction sites of the phage DNA insert were chosen later to produce fragments for subcloning directly into M13 to confirm the continuity of the *Eco*RI fragments. The polymerase chain reaction was also used to amplify an intervening region prior to exon X to confirm the continuity between *Eco*RI fragments pEco5.7 and pEco4.9. Restriction fragments of the pUC-amplified clones were inserted into M13mp18 or M13mp19 for sequence analysis. Subclones were generated by sequential deletion for several areas of the gene according to the method of Dale et al. (1985). This procedure employs the 3′ to 5′ exonuclease activity of T4 DNA polymerase to generate deletions within single-stranded DNA insertions in M13 clones. One-third of the gene was sequenced by the procedure described by Sanger et al. (1977), employing the Klenow fragment of DNA polymerase I, while the remainder was sequenced by the modified technique of Tabor and Richardson (1987), employing either the Klenow fragment or the Sequenase version of T7 polymerase.

Areas of the gene that were difficult to sequence by restriction subcloning or sequential deletion were primed by using specific sequencing primers 15 nucleotides in length. Sequencing reactions used [α-^{35}S]dATPαS and were run on 6% polyacrylamide buffer gradient gels (Biggen et al., 1983) or ion gradient gels (Lurquin, 1988) in 60-cm plates. Initially, the exons and the intron–exon boundaries in the gene were identified. Southern transfer and hybridizations with specific cDNA probes were used to locate the exons within the various plasmid subclones. Specific restriction sites within the cDNA were also identified, and the subclones containing the desired section of genomic DNA were digested with the appropriate restriction enzyme (Figures 1 and 2). The resulting fragments were then cloned into M13 and sequenced. All the intron–exon boundaries and their relative locations within the gene were determined in this manner except for exon X, for which oligonucleotide primers were used to sequence from *Eco*RI fragments inserted into M13. After the boundaries on one coding strand were sequenced, new restriction sites were chosen on the basis of the sequence of the intron, and the new fragments were used to subclone into M13 and to sequence the noncoding strand. When there was no convenient restriction site available, oligonucleotide primers were made to sequence the noncoding strand and to confirm the sequence.

DNA sequences were analyzed by using the GENEPRO program (version 4.2, Riverside Scientific Enterprises, Seattle, WA) and then compared to the GenBank DNA sequence library, release 60 (Los Alamos National Laboratory), and the Protein Identification Resource protein library, release 21 (National Biomedical Research Foundation), with an IBM PC/AT compatible computer.

Results

Isolation of the Gene for the b Subunit of Factor XIII. Approximately 2×10^7 recombinant phage were screened from three human genomic libraries by hybridization with a cDNA coding for the b subunit of human factor XIII. Twelve unique recombinant phage were isolated, purified, and characterized by restriction mapping, Southern transfer, and hybridization experiments (Figure 1). The first recombinant phage isolated, λCT9 (Clontech leukocyte library), contained the 5′ end of the gene, while λSY1 from a second library (Yoshitake et al.,

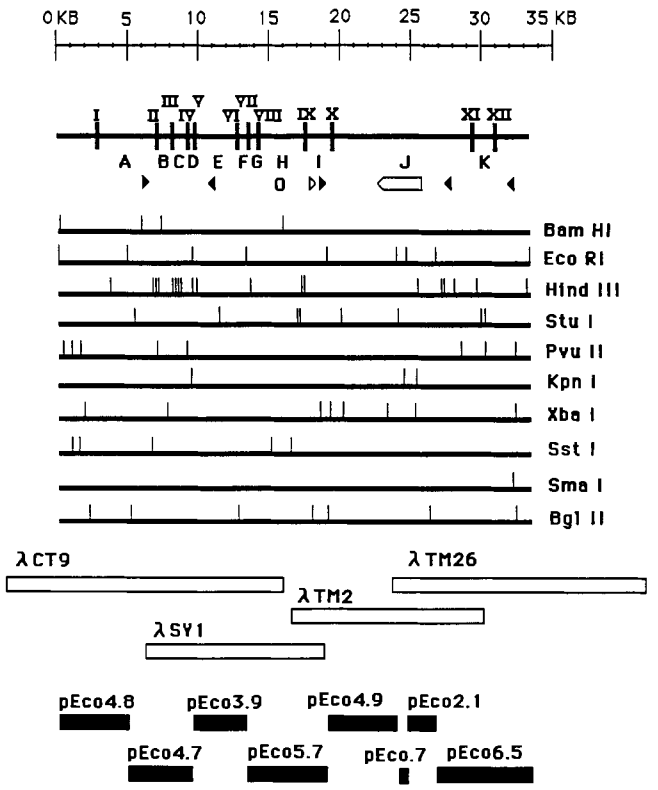

FIGURE 1: Organization and restriction map of the gene for the b subunit of human factor XIII. Exons I–XII are shown by the solid vertical bars. Solid arrowheads below the gene map show the location and direction of Alu sequences, while *Kpn*I repeats are shown by open arrowheads and the "O" family repeat is shown by an O. Cleavage sites for a few restriction enzymes, the λ phage inserts, and the plasmid subclones are shown in the remainder of the figure.

1985) contained exons II–IX. λTM26 and λTM2 from the Charon 4A library (Lawn et al., 1978) contained the 3′ end of the gene and exons IX and X, respectively.

DNA Sequence of the Gene. Genomic DNA inserts from the four phage (λCT9, λSY1, λTM2, and λTM26) were digested with *Eco*RI and the fragments subcloned into plasmids (Figures 1 and 2). The detailed sequencing strategy and the complete nucleotide sequence of the gene are shown in Figures 2 and 3. These figures also include data for the 5′ and 3′ flanking regions of the gene. A total of 33 207 nucleotides were determined in this study (Figure 3). Approximately 89% of the gene and its flanking region were sequenced two or more times. Approximately 54% of the sequence analysis was also performed on both strands, including the coding sequences, their flanking regions, and the 5′ and 3′ flanking regions.

Exon, Intron, and Intron–Exon Boundaries. Table I shows the 11 splice junctions and the splice junction types (Sharp, 1981) identified in the gene coding for the b subunit of factor XIII. The splice junctions all follow the GT-AG rule of Breathnach et al. (1978) and are similar to the consensus sequence published by Mount (1982). The first 10 splice junction types were all type I in that the introns divided the codon between the first and second nucleotides. The last splice junction (intron K) was type II since it divided the codon for Ser-631 between the second and third nucleotides.

The size and location of the 11 introns in the gene are shown in Table II. Intron J was the largest (9952 nucleotides) and intron F the smallest (87 bp). The introns in the gene for the b subunit of factor XIII make up 92% of the total DNA. The size of the exons ranged from 64 to 221 bp and averaged 192 bp, excluding exon I. The full size of this exon is not known. The 10 tandem repeats that have been called sushi structures

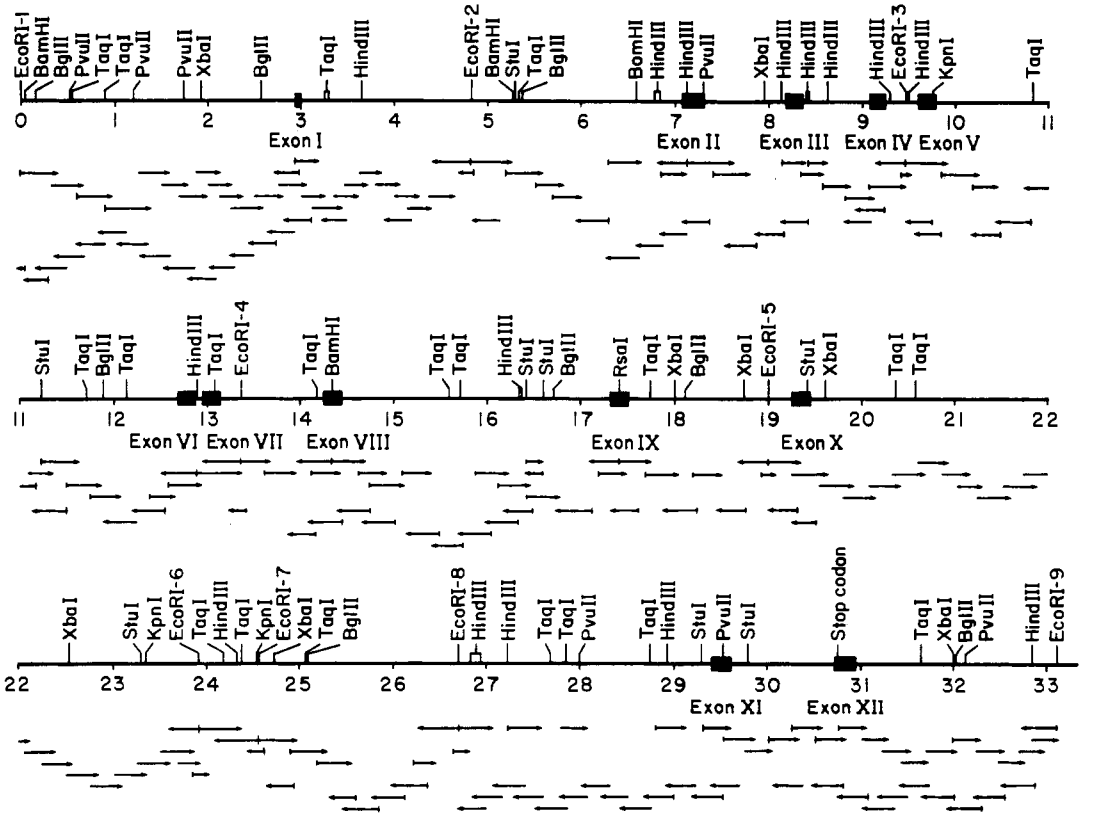

FIGURE 2: Sequencing strategy and detailed restriction map for the b subunit of human factor XIII. The nucleotides are numbered in kilobases from 0 to 33 and the locations of the 12 exons are shown by solid bars. The *Eco*RI sites are numbered 1–9. The length and direction of the sequencing by the dideoxy method are shown by the arrows. Arrows beginning with a vertical bar indicate that the DNA was sequenced by employing a specific oligomer primer. Arrows lacking a vertical bar indicate that the sequencing came from sequential deletion subclones as described in the method of Dale et al. (1985).

Table I: Intron–Exon Splice Junction Sequences in the Gene for the b Subunit of Human Factor XIII

EXON	5′ INTRON 3′	EXON	TYPE[a]
	+2	+2	
	Ala Glu Glu|	|Glu Lys Pro	
I	CCA GAA G |GTAAAA..Intron A..CATTTCAG| AG AAA CCC		I
	69	69	
	Cys Phe Lys|	|Lys Lys Cys	
II	TGC TTC A |GTTAAG..Intron B..TCCCATAG| AA AAA TGC		I
	131	131	
	Glu His Glu|	|Glu Thr Cys	
III	GAA CAT G |GTATAA..Intron C..AATTTTAG| AA ACG TGT		I
	190	190	
	Cys Thr Lys|	|Lys Leu Lys	
IV	TGT ACC A |GTAGGT..Intron D..ATTTGCAG| AA TTA AAG		I
	249	249	
	Cys Glu Gly|	|Gly Arg Arg	
V	TGC GAA G |GTAATT..Intron E..TCCATAAG| GA AGA AGA		I
	309	309	
	Cys Ile Glu|	|Glu Gly Gln	
VI	TGC ATT G |GTTAGT..Intron F..AAGGACAG| AA GGA CAG		I
	371	371	
	Cys Val Glu|	|Glu Asn Asn	
VII	TGT GTT G |GTAGT...Intron G..TTTCTTAG| AA AAT AAT		I
	432	432	
	Cys Leu Glu|	|Glu Pro Cys	
VIII	TGC TTG G |GTAAGA..Intron H..TTTTACAG| AA CCA TGT		I
	499	499	
	Arg Lys Glu|	|Glu Ser Lys	
IX	AGA AAA G |GTAAAA..Intron I..TAGAAAAG| AA TCT AAA		I
	560	560	
	Cys Leu Glu|	|Glu Pro Cys	
X	TGT TTA G |GTATGT..Intron J..GTTCCCAG| AG CCA TGC		I
	631	631	
	Arg Gln Ser|	|Ser Thr Leu	
XI	AGA CAA AG |GTAAGA..Intron K..CCCTCAAG| C ACT CTG		II
	C | A TT T |		
CONSENSUS[b]	AG|GT AGT..... N AG|G		
SEQUENCE	A | G CC C |		

[a] After Sharp (1981). [b] Taken from Mount (1982).

Table II: Location and Size of the Introns in the Gene for the b Subunit of Human Factor XIII

exon	nucleotide position	size (bp)	intron	nucleotide position	size (bp)
I	+1–64	(64)[a]	A	65–4162	4098
II	4163–4363	201	B	4364–5250	887
III	5251–5436	186	C	5437–6143	707
IV	6144–6320	177	D	6321–6676	356
V	6677–6853	177	E	6854–9754	2901
VI	9755–9934	180	F	9935–10021	87
VII	10022–10207	186	G	10208–11321	1114
VIII	11322–11504	183	H	11505–14378	2874
IX	14379–14579	201	I	14580–16334	1755
X	16355–16517	183	J	16518–26469	9952
XI	26470–26683	214	K	26684–27793	1110
XII	27794–28014	221			

[a] The length of the 5′ noncoding region for the cDNA coding for the b subunit is not known; therefore, the length of exon I is shown in parentheses starting from the initiator methionine.

are coded by exons II–XI, and their size is relatively constant with a mean of 189 bp. These exons are only slightly larger than the average length of 150 bp found in other eukaryotic genes (Blake, 1983).

The 5′ and 3′ Flanking Regions. The sequence of the cDNA coding for the b subunit contained a partial leader peptide of 19 amino acids but lacked an initiator Met and a transcription start site (Ichinose et al., 1986b). Accordingly, a human liver cDNA library was screened with the cDNA probe in an attempt to find a full-length transcript. These experiments, however, were unsuccessful. A Met is present, however, in the genomic sequence just prior to the first amino acid codon of the cDNA, and this Met is located at position −20 in the leader sequence. This potential translation initiation site was designated as the 5′ end of the gene. The next in-frame codon for Met occurred upstream an additional 57 nucleotides from the Met at position −20, but this triplet was located prior to several stop codons. This supports the conclusion that the leader sequence for the b subunit of factor XIII is 20 amino acids in length. If the nucleotide prior to the polyadenylation site is arbitrarily designated as the 3′ end of the gene, the length of the gene for the b subunit of factor XIII is 28 014 nucleotides. An additional 2130 nucleotides of the 5′ flanking region and 2262 nucleotides of the 3′ flanking sequence were also determined, as shown in Figure 2.

Further analysis of the 5′ end of the gene revealed several potential transcriptional regulatory elements (Breathnach & Chambon, 1981). Thirty "TATA"-like sequences exist in the 727-bp region upstream from nucleotide 1, the four closest being TAAAA (−29), TAGAA (−38), TATTA (−94), and TATAA (−101). There is also a reverse "CCAAT" box (−320) that has been shown to function in either orientation (Graves et al., 1986). There were no "GC" clusters, however, in the 5′ region.

The protein levels for the a and b subunits of factor XIII show some coordinate control in patients with factor XIIIa deficiencies (Ikematsu, 1981; Rodeghiero et al., 1981). A comparison of the 5′ end of the gene coding for the a subunit of factor XIII (Ichinose & Davie, 1988) with the 5′ end of the gene coding for the b subunit shows several short sequence identities that may be possible common trans-acting regulatory regions. These sequences include AAATTAAA (0713, −714), GAAAATTT (−306, −695), GTGTTATTTT (−175, −642), TAAAAT (−30, −710), and ACATGGTA (−205, −930). The positions of the nucleotides are shown in parentheses for each of the sequences for the b and a subunits, respectively. It is not known, however, which of these sequences, if any, function as regulatory elements.

The 3′ end of the gene contains an AATAAA polyadenylation signal starting at position 27990 and a T(A) polyadenylation site at 28014. There were no other sequences found that commonly surround transcription termination sites.

Other Features of the DNA Sequence. When the DNA sequence of the exons and the sequence previously reported for the cDNA (Ichinose et al., 1986b) were compared, only 1 out of 2171 nucleotides was found to be different. This occurred within exon XI at nucleotide 26537 where a T was replaced by a C. The change identified in λTM26 did not change the Asn, however, at position 582 in the polypeptide chain. This must be due to a polymorphism since another cDNA of 1.6 kb had the C residue at this position (A. Ichinose, unpublished data).

The base composition of the gene for the b subunit of factor XIII was enriched in A and T residues. The composition was 30.1% A, 35.8% T, 17.5% G, and 16.6% C. The dinucleotide frequency was very similar to those summarized by Nussinov (1981) for eukaryotic genes, with a significant deficiency of CG and TA pairs when compared to predictions from random distribution. The frequency of dinucleotides was the following: AA, 9.9%; AC, 4.4%; AG, 6.1%; AT, 9.5%; CA, 6.3%; CC, 3.5%; CG, 0.5%; CT, 6.8%; GA, 5.2%; GC, 3.2%; GG, 3.9%; GT, 5.6%; TA, 8.4%; TC, 6.1%; TG, 7.4%; TT, 13.3%.

Common Repetitive Sequences. Repetitive sequences found within the introns made up 21% of the gene for the b subunit of factor XIII. These sequences belong to the Alu, *Kpn*I, and

```
                                                                                                                GAA  -2928
TTCTTTTTCT GGCAGTTCAG GAATTTCTTC TTGGTTTGGA TCCATTGCTG ATGAGCTAGT GTGATTTTTG GGTGGTGTTA AAGAACCTTG TTTTGTCATA  -2828
ATACCAAAAT TGTTTTTCTA GTTCCTTCTC TTTTGGGTAG GCTATGTCAG AGGGGAGATC TGGGGCTCAA GGCTGCTGTT CAGATTCTTT TGTCCCACGG  -2728
GGTGCTCCCT TGATGTAGTA TTCTCCCCCT TTTCCTAGGG ATGTGACTTC CTGAGAGCCA AACTGTAGTG ATTATTATTA CTTTTCTGGA TCTAGCCACC  -2628
AGAAGGGCTA CCAGGCTCTG GGCTGGTACT GGGGGTTGTC TGCATAGAGT CCTGTGATGT AAACTGTCTT CAGTTCTCTC AGTCATGGAT ACCAGCACCT  -2528
GTTCCAGTAG AGGTGGCAGG GGAGTGAAAG GAACTCTGTG AGGGTCCTTA GTTGTAGTTG TTTGATGCAC TAGTTTGGTG CTGGTTGGCC TCCTGCTGAG  -2428
AGGTGAAGCC AGCTGGACTT CCTGGGTCGA ATGGGGACTT GGAGAACTTT TCTGTCTAGC TAAAGGATTG TAAACACACC AATCAGCGCT CTGTGTCTAG  -2328
CTAAAGGTTT GTAAATGCAC CAATCAGCAC TTTGTAAAAA CAGACCAACC AGCACTCTGT AAAATGGGCC AATCAGCGGG ATGTGGGCGG AGACCAAATA  -2228
AGGGAATAAA AGCTGGCCAC CTGAGCCAGC CGCAGCAACC TGGTCTGGTC CCCTTCCAGG CTGTGGAAAC TTTGTTCTTT CGCTCTTTGC AATAAATCTT  -2128
GCTGCTGCTC TCTTTTTGGG TCTGCACTAC CTTTATGAGC TGTAACACTC ACCGCGAAGG TCTGTGGCTT CACTCCTGAA GTCAGCAAGA CCTCGAACCC  -2028
ACCAGGAGGA AGAAACAACT CCGGACGCAC CATCTTTTAG AGCTGTAACA CTCACTGTGA AGGTCTGCGG CTTCACTCCT GAAGTCAGCA AGACCAGGAA  -1928
CCCACCAGAA GGAAGAAACT CTGGACACAT CTGAACATCA GAGGAACAAA CTCCGGACAC ACCATCTTTA AAAACTGTAA TGCTCACCGC AAGCGTCCGC  -1828
AGCTTCATTC TTGAAGTCAG CAAGACCAAG AACCCATTGG AAGGAACCAA TTCCAGACAC ACTGCCATGA GGTAGTGCTT TCATGAGAGC ATCAGCTGTG  -1728
GTATATAGGG GAGGATCAGG TGGTGGGTGG GACACTAGAA CTTCCAAGAG AATATGACCT TTGTCTTCAG CTACCAGGGT GGGTAGGGAA GGACCATCAG  -1628
GTGGGGGCAG GGTTAGGCAT GTCTGAGCTC AGAGTCTTCT TGGGTGGGGC TTGTTGTGGC TCTGTGGGGA ATGGGGGTGT GATTCCCAGG TTAATGGAGT  -1528
TACCTTCCCA GGAGGATTAT GGCTGCCTCT TCTGTGTCAT GCAGGTTGTC AGGGAAGTGG GGGAAAGCTG GCAGTTACAG ACCTCACCCA GCTCTCACAC  -1428
AACCCAACAA GCAAATCTCA CTTCTACTGT GTTCCCCCAA AAAGCACCGA ATTTGTTTCC AGGCAGTAGG GGAGCAGGAT TGAGAACTTG CCCCAGGCTA  -1328
CTAGCCTCCC AGCCGAGAAT GCAAGCCAGA CTTTTGTGCC TCCCTACCTG TTAAGTCTGC ACACCAGATT CACGCCCTTT CCTGAGTTCT GGACAGGAAA  -1228
GTTTGTGTTC GGTTGGAATT GTTACTAAGT TCAGCTGGAG GTTTCCTTCT CCCTGTGGTC TTTTCCCAGT TCCTCCGGCA GCCCCTGTGA GACAAGTCAG  -1128
AAATGGCTTC TCTGGGGACC CACAGAGACG ACGGGGCTTT ACCCTTTGCT TCCTCTACCC CTGTATTTCG CTCAGCTCTC TAAATTGTCT GAGCTCCAGG  -1028
TAAGGTCAAA TCCTTCTCCT GTGATCTAGA CCTTCAGGTT CCCCAAGGGT GTGTGTTCAG GGGTGGACAA TACCCCTTTC ACACTTTTCA CAGTTTGGGC  -928
ATTCACAGTA TGTGGGCTAT TTTCTGGGTC ATGCAGGAGC AATGTACTTT CTTCAGAGGG TCTGTGGATT ATCTTGGCTT TCCTGTTATA TCCTTGCAGC  -828
AGTTCTTGGA GCAAAAGTTC ACAATGCGAG TCTCCACATC CTGCTCTGTC CATCCAAGTG GGAGCTGCAA GTTAGTCCTG CCTCCTATCC ACCATTTTTT  -728
CCTATTTTGT ATTTTAAATT AAAACTCATT AGAATGTAAA CTCCATGAGG GCAATAATTT TAATCCAGTT TTTTCACTGC TAAATACCTC AAAACTGTTA  -628
CATGTGCTGC CATTATCATG GACGTTCTAA ACATTTCTCA TGAATGAATG TTGATATAGC ACAGTGCTGA TCTCACCACC TGCCACAAAG TAAAAACTCT  -528
CTAAATGTCA AATACTACTA TTATAAGTAA TACTATTCTA TCATTGAGAC AACTATGGTT CTGTATCTCT TTAAGGAGTG CAGTCTAGGC AACAACATAT  -428
TTCAACATCA GGTATCTGTT ATCTCATGTT CCTAAGTTCA TCTTTTTACC TGATCAATAT TAGTAAAGAT CTATGGCCTC AAGTGAACTT TACTCTATGA  -328
ATTTACAATT GGGAACTTAG GTTGAAAATT TACCTAGTAT TATTTATTCT GCTGCTATAA TTTATTATTT ATTCTGCTGC TATTGTTGCC CTAATCAGTT  -228
ATCATGCTCT TACTGGACTC TGACATGGTA AGTAATGTCA ATTAATATAG TTGTGTTATT TTAACATTAT TTCAAAATAT TTTAAGTGAA ATATAAGACT  -128
CGTTACAAAA GGACTTAGAC AGAGGTTATA AATTATTAAC TTTTTCCTCA AGAGAAGTAT TGTGTACTTA ATGACAAGTT CCTAGTGCTT AGAATTGTTA   -28
                          -20
                            Met Arg Leu Lys Asn Leu  Thr Phe Ile Ile Ile Leu Ile Ile Ser Gly Glu Leu Tyr Ala
AAATCTTTGT GAAGCACACC ACTGAAG ATG AGG TTG AAA AAC CTG  ACT TTT ATC ATC ATA TTG ATA ATC TCA GGA GAA CTC TAT GCA    60
 +1
Glu
GAA G GTAA AATTAATTCT TAATATTCTT TATTCTCTAA AACATACAAA TGGAGATTCT AATATTTATA CAACTCAAGT TTTAATATAC TCTTCTATTT   158
    ▲
TCATTATCAA TTTTATAACA TATTTTATAA ATATAACTTT TTATAATATT TAGACCTTTT CCTTTGAGCA AAATTAGTAA AGTGTTCTTA CAGTAGCATA   258
AGTCAGAAAA GCAAAACAAA ACAAAAACCC TTGTATCTCT AGGGGGCATA GGAGACCAGG ATCTAAATAC TTCGAAAAGG TTTTCAGGTT AACAATATGT   358
CCATTTCGAA GTCACATGTT AACTTTAATT CTAAATTTAT GCTAGAATCC ATTCCTAATA CTGACAAAAC TTGTACAATT TTAATTCTGT TCATAAGGTG   458
AGAGTTTAAT GTGGGAAATT TGAAAAGGAA ATAGCCATAA ATATTTTTTA CTTTTGAGAC TTTGATTTTT ATCACAATGA ATTATTCATT TGTTGGAGAC   558
ATTAAGTAGT GATTAGTAGG TGAACCATAT AATTGCTTGT ACAAACAGGG ACACTTAGAG AGTAAAAGAG GGGCTCTATT AATAACTGTT TGGACAACAC   658
ATGTAAATCA GGACTTTCCA AACAAACCAA TACCTATGGA CTCCCTATTA GAAAGCTTTA TAGTCAGATT GCCCAATGTG TAAGATTAGT GTGCGAGTGG   758
AAGCTAGGCA GTGTACATGG GTATGTTGTT CAGGAACATT ATCAATATTA TCAGGCCGTA GATGAAGAAT CCAGGGAGGA CACTTAGCAT TAGCAAGAGT   858
GAACATACAA AGAAGGAAGC TACAGCAATA CAGCTTTTCT AAAATAAACA TTTTTGGGAA TTTGTTGAAA TTAAATACAT AAGTTTTAAC TGTGTTTTAG   958
TTGAATAGTT AACTATAAAT TTTCTACCAA AACATCTATA ATTAAAATTA TAAAAGTGAA AGACACAGAA TTAATCTTAT TTTATACTTT TAATTCTATA  1058
ATAGGATGAT TTTTGCATGA ATTTCTTTTG ATTTTATTTT TCTGTCTAAA TAGGGAAATA CCAGCAGATA TAATTGTCTG TTATCTATAC CATAAAGACA  1158
TTGCAATATT CTAGTTAATT TTCTAATTAT TATAAGGTTA TAATGCTATT TTACACAACA AAAAAAAAAA GTAAGCTGCT TGAAGACCAA GACAATGATT  1258
TTATTTTTGC ATGCTTCTCT CAACTCCCAT TGTGCACTTC AGGGACAAGA ATTGAATTAA TTGTGTTTCT CATTTCATGG TATGCAAGTA TTATTGGCCT  1358
ATGAAAGGTC TCACTCTTGC TTTACTTACA CATTTTATCC CCTTATTCCT AATCTTGAAG TCCATTTCCC CTGGGACCTA CTACAATGTT TTAACATGTT  1458
ACTCTAGCTC TCTCAACCAC AAATAGCTCT GGAATGAATA TGTAATATAA CCAACTTATC TTTCTGTAAG AATGTCATCA GTCTGTATTC ACTCACAGAT  1558
AGCATTATTG AGTCTATTCT TGTACTTGAT TTCTGTAATG AGAGATTACT TTTGCAGAAT GGCAGTTTTG AAGTATATCA TCTCTGAGAG AGGATGTAAA  1658
TTGTTGCTCC TCATCATTCT CACCAACACC TCATATTGTC CATTTTTGTA CTGGTTGCTA TATGGTTATT CATAAAGTGA TATCCCATTT TTATTTATTT  1758
ACCATTCTCT GATCAGTAGT AAGTGTGTCC ATCGCTTCAT GTACTTTTTT TAATATGTTC AGACATCTGT GATTTCCTGG TTTATATGTG ATTAGTATAT  1858
TCCTCACATT TCTCATTAAT TTTTTGTCAG GAATTCATAT TCTCCAAGGC ATTCCAGCTA CCTTAACAAA TAATGTGCCT ATCCTGGTGG TGGAGGGAGA  1958
TGTCCCAGGA CAGAGAGTCT TTAAAGCCAC AACCTAGGAT TGTTGTGCAG AGAATACTTT GCTGCCCCTC CCTGACAAGC AACTGTGCCC CTCAGGAAAT  2058
TTCTTGTGTA TTCACCATTG CACTGCTATT TTGGAAGACT CTTTTTATTG TAAATTCCTT TGGTGCTTGG GTGTGAAAAA ATGGTGAGAA TAATTCTCAA  2158
GGAATGGGGA CTATCCAGCC CATGTTTTTC TGTCTGCTCC ATCAGATAGT GGAAGTTTTG TTGCTCATCA TTCCACATCC ATTTCCTTCA AACCTAGGCT  2258
AGCACTACTA GTATTTTATG TCTCATAGTA GCATTAAGGT ATACAGTGAT CAACCAAGGC ATTGAAGCCG TAGTAATCAC ACTGGATCCA GGCCTTCTTC  2358
CATCCATGAC TGCCCATGTT TCTTCCCACC ACAGCTCCTA GGATTGTTCG AGCATTTCTT TCTTTTAAGT GTTTTTATGA CTTTAAGATC TTCAACATAA  2458
AGATATATTT TTAAAATATA TGTTTTATTT GGAGTATAAC ATATCCAGAA AAATGCAAAA ATCACAGGAT ACACTTATAT AGTTTTTACA AAATTTTACA  2558
```

```
CCATCCAGAT CAAGATATAG AACAGTGATA GTACCCCAGA AATATGATCC TTCTAAGGCA CTAAACCCCA AAGGTAACCA TTCTGACTTC TATTATCATG   2658
GATTAGTTTG TTTTCTTTTA AAAAGTTATG CAAATGTATA CATTAATTCA TATGTGATTG GCTTCTTTCA CTCAATATTA TGTTTATGAG ATTCACCATT   2758
TGTTCATTTC TGTTGTTGGA TAATATTCCA TTGTATGCAT AGTCTAAAAT GTGTTTATCT TTTCTATTAT TTATGGGCAT TTAGATTGTT CCCAGTTTGG   2858
GACTATATGA ATACTGCCAC TATAAGCATT TTTTTCCATG TCTTTTGGTA TACAAATATG TGCTTTTTTT ATTGGGTAAT ATGGTTTGTA CATGTTTTAC   2958
CCCAATATAT ATTGCAAAGG TGTTCTCTCA TAGAGTTTGG ACTGATTTAT ACTTTCAGTT GTAGAAATAT GCGAGTTCCA GTTTCTTCAC TTCTTCACAA   3058
TCATTGAGAT AGCTTGTCTT TTTCATTTCA GCAGTTCTAT TGGATATGCA GTAGTATCTC ATGGTGGTTT AAATTTTCAT TTCCCTGATG ACTAGTAATG   3158
CTAAGTACCA GGTCTCTAAC CATTTAAATA TCCTATTTTG CCAAGTTCTT GCCAAGTATT GCTTTGTTTT ACTTGGTTGT CTGTTTTTTC TTTTAATGCT   3258
ATTTATTTCA AAATTTTTTA TGAGTTAGCA TATGTTAAGG AGAAACTTTA TCTATTCCTG AACTGTACTG TCCAATACTA GCCATCAGCC ACATATAGCT   3358
ATTGGAGACA GTACTTAAAA TGTCACTAGT TCAAGATCAG ATATTCTGTA AGTATAAATA TGTAGTGGAT TTTAAAGAGT TAGAAAAAAA AAGAATGTAA   3458
AATATTTTCA ATATATTGGG TTAAATAAAA ACCTATTATT AAAACTAACT TCAATTGTTC CTTAGTATTT TTATGAATGT GACTATTAGA AATTTTTAAA   3558
TTACATAAGT GACTTACAAT ACATATCTAT TAGGAAGTGC AGGCAGGGTG GAGTGGTTCA AGTCTTGTAA TCCTAGTGCT TTGGGAGGCC AAAGCAGAAG   3658
GATCCTTTGA GTCCAGGAGT TTGAGACCAG CGTGAGCAAC AGAGTGAGAC CCCTCATCTC TACAAAAAAT AAAACAAAAA TTAGCCAGAT GTGGTGGCAT   3758
GTGCCTGCAG TCCTAGCTAC AGGAGAGGAG GCTTGCTTAA GCCCAGGAGC TCCTGGTGGC AGTGAGCTAT ATTTCAGCAC TGCACTGAAG CTTGGATAAC   3858
TGAGGAAGAC TCTGTGTCTA GGAAGAAAAA AAAAAGTACT GTTCTACGAA GTTATGCAAA GCTTCCCTGA AAAAGTGGTC TGCTAAATTT TATCATGACT   3958
GTTGGATGTT TAAAGTCATT TGCAGATCAA TTATGATAAA AGGACTCCTT GAGTTGTCAC AAAAGTACCT TAAAAATTTA AGTAAGAAAA ATAAGTACTA   4058
GTTGAAGTGT CTCTAAAAAC TTTTAGTTTG TATTGTCACC TGATTACAAA TTTATGTTTT TAGATTTGTT TTTCCATTTT CAGCTTGACA TTTACTCATT   4158
    +2
    Glu Lys Pro Cys Gly Phe Pro His Val Glu Asn Gly Arg Ile Ala Gln Tyr Tyr Tyr Thr Phe Lys Ser Phe Tyr Phe
TCAG AG AAA CCC TGT GGT TTT CCT CAT GTG GAA AAT GGA AGA ATT GCC CAA TAT TAC TAT ACT TTT AAA AGC TTT TAC TTT   4239

Pro Met Ser Ile Asp Lys Lys Leu Ser Phe Phe Cys Leu Ala Gly Tyr Thr Thr Glu Ser Gly Arg Gln Glu Glu Gln Thr
CCA ATG AGC ATA GAC AAA AAA TTG TCA TTT TTC TGC TTG GCT GGT TAT ACC ACT GAA AGT GGA AGA CAA GAA GAG CAA ACC   4320
                                                68
Thr Cys Thr Thr Glu Gly Trp Ser Pro Glu Pro Arg Cys Phe
ACG TGT ACA ACA GAA GGC TGG TCT CCA GAG CCA AGG TGC TTC A GTAAGTC AGCTGGATAT GTCACTCAAT GTTTCAATAC TCAAAGAAAT   4410

TTGTATATAA AAATAGGGGT CCAATAAAAA TGGAAAGCGG TTAAACCTTT TAGTAGATAA GAACAAAATC ACTAGCAATC TGTCTTTCTA CAAAATGAAG   4510
GCATAATTTG AAATCCAGTA AATAATGCTT TTACATTAAA ATGTTATTTT CTGTATCTGT TTTTAATAAT ACTCTGTTGA AGGGAATAGG GTAATAACTG   4610
AAATTATCAT ATTGCAAGGC AAAGAGAGGT AATTTACTTT AAGTAACAAG AAACTTATTT TGATAATTAA GCATAGGATG CCAGAAAAAT GCTTGAAACT   4710
CTGGCCTCTG GAAGGGCAAG ATTGCTACTC TAATGCTTTA TCATTTGTAA GACTCATCTC TCTTCATGTG CCAACTTCTC TCATGATACC CATTGAAAAA   4810
TGTGTTTTTC AATTTTGAAT TGAAAGTTTG AGTTCAAAGT TGTGAGAAGG TAATCTAATT GACTCATTTC ATTATTTTAT ATCAGGCTAC ACTATCATCA   4910
AACTATCAAT GATCTTCCAC TAGGGCAGGT GACCCAGAGT AAGGGGCTTT GGTTGAGCCA GTTTCGTTTA AAAGAGGATG TTGAATATGG CTATTGTGGT   5010
TGATATATCT AGTAAACCTA TCTAGAAGTA GATGTGGTAG GGGTAGAAAT TACTATAAGA ACTAAAATTA GTAAATACAA AGAGAGTCAG TGTTAAATTG   5110
AATCCATTTT CAAAAAAATT ATGCAGATAT TTCCAAAATG AAATCGCCAA TAATAACATT ATACTTTTGG CTTAATTTTA CAATTTAGTA AAAGACAAGC   5210
                                        69
                                        Lys Lys Cys Thr Lys Pro Asp Leu Ser Asn Gly Tyr Ile Ser Asp Val
TTAGTTTCAT CATTAAGTTA AATAATTTTT TTTCCCATAG AA AAA TGC ACT AAG CCT GAC CTG AGT AAT GGT TAC ATC TCT GAT GTA   5298

Lys Leu Leu Tyr Lys Ile Gln Glu Asn Met His Tyr Gly Cys Ala Ser Gly Tyr Lys Thr Thr Gly Gly Lys Asp Glu Glu
AAG TTA TTG TAT AAA ATT CAA GAG AAC ATG CAT TAT GGT TGC GCT TCA GGG TAC AAA ACC ACT GGA GGG AAG GAT GAA GAA   5379
                                                        130
Val Val Gln Cys Leu Ser Asp Gly Trp Ser Ser Gln Pro Thr Cys Arg Lys Glu His
GTG GTT CAA TGT CTC TCT GAT GGA TGG TCT TCT CAA CCA ACC TGT AGG AAA GAA CAT G GTATAAGAA TCATTTCCTA AAACTGAAGA   5465

TAAGTGCTTA TCAAGCTTAT ATTGAATATA TTGAAGCTTA TATTGATATC TATTATTTAA TAGAATTTAT TTTTCATTAT GTCTACAATG CAGGAAGTTG   5565
ATATTACATT GCTTTTCTCA AGTACTAGGT GATTCCTTAA TCCAAATGTT CTCAATCCCT ACAACCCACC TACGCCAATA GTTCCTACTT TGCATCAATA   5665
AAAGGAATTG GAATGGAATA AAATGAAAGA AGCTTTTTCT TTATATGTCA CTATTTCCTT TATTACAGAA GAATTTAAGA TCATGTACAT TATTTTATAT   5765
TGTGAACCAA TGTCAGATAT ACTTAAAAAT AGTTCAATAA TTTAAATATA TGTATACTTT GTTAGTAAGT TGTTTAAATT GTAAAAATGC ATAGACAGCA   5865
CATACTTTCA CAATTGCTAC TGCACAAGCC TTTAGTTGAA TTTTGGCCTT TGGTCAAGCA CTTGTAGACT AGAATTTTAA AAACAGAAAA CTGGTAGAGT   5965
TAGGTCACAA GGTGTATGAA GAAGAGAATA GAGGCAAAGA AAAATTGGGG CTATTTTTCA ATTTTGCCAA GAGTTGATTT TACCATGTAT TCTTAAGATT   6065
                                                                131
                                                                Glu Thr Cys Leu Ala Pro
TGTAAAACCA TATTTTGAGA TTTTTTATAA TCTAGGTTAT TAAAAAGTTT ATAAAATAAA TAATTTTTAT AATTTTAG AA ACG TGT TTG GCT CCT   6160

Glu Leu Tyr Asn Gly Asn Tyr Ser Thr Thr Gln Lys Thr Phe Lys Val Lys Asp Lys Val Gln Tyr Glu Cys Ala Thr Gly
GAA TTA TAT AAT GGA AAT TAT TCC ACA ACA CAG AAA ACA TTC AAA GTG AAG GAC AAA GTA CAA TAC GAA TGT GCT ACT GGC   6241
                                                                                189
Tyr Tyr Thr Ala Gly Gly Lys Lys Thr Glu Glu Val Glu Cys Leu Thr Tyr Gly Trp Ser Leu Thr Pro Lys Cys Thr
TAC TAC ACA GCT GGA GGA AAG AAG ACA GAG GAG GTA GAA TGT CTC ACA TAC GGA TGG TCT CTC ACA CCA AAA TGT ACC A GTA   6323

GGTTCTCATT TGGAATAAAA TCTACTCTCA CTCTGAAAGC TTTTTCTCAT AGAGAAGTGT ATATAAGTTG TCATGCTCAT TATGTTTCAC CTTTTCATGC   6423
```

```
ATTTTTGCCT TTATTTTTAA TTATGTTGAT AGCAGTCTTT GATAATTAAA GCTCTCAAGT AGAATAAATT CATTTTAGCT CACAGAATTA AATTAGAATA   6523
AGCCTAATGA TATTTAACAA ATATATAAAA TGAATTCATA AGCTTTTTGG CAAATTTAAT TTCTATTCTA GTTATGTCAT GTAAATAGTG CATTTTGTTG   6623
                                                    190
                                                    Lys Leu Lys Cys Ser Ser Leu Arg Leu Ile Glu Asn Gly
TAGCAGAAAA AATAAACATT TGTTTGTAAT TTTATTCATT CTTTTATTTG CAG AA TTA AAG TGC TCT TCT TTA AGA TTA ATT GAA AAT GGT   6714

Tyr Phe His Pro Val Lys Gln Thr Tyr Glu Glu Gly Asp Val Val Gln Phe Phe Cys His Glu Asn Tyr Tyr Leu Ser Gly
TAT TTT CAT CCT GTA AAG CAA ACC TAT GAA GAA GGA GAT GTC GTT CAG TTT TTC TGT CAT GAA AAT TAT TAT CTA AGT GGA     6795
                                                                248
Ser Asp Leu Ile Gln Cys Tyr Asn Phe Gly Trp Tyr Pro Glu Ser Pro Val Cys Glu
TCT GAT TTA ATT CAA TGC TAT AAC TTT GGT TGG TAC CCA GAA TCT CCT GTA TGC GAA G GTAAAT TTTTAAATTT AATACTCGCA      6879

AAATGCCATG TCTTTATCAG CTTTAATCTC TATTAGCTCT GACAAAATTC CTGTGGTCAT ATCTATTTTT TTTCCTGAGT CCTATTTTTA AAATAAATAT   6979
TATCTTTTAA ATCAGTGCTA TTTTTAGCAA TCTTTAATAT ATCCCACCTG AATAAATCTT GTTTTGTCTT AGTTTGTAAA TGACTATATC TTTTAAATTT   7079
ATTATTTTAA AATATCTATT ATACATTTTA AATCCCTTAT TATAACGAAT AACTTTTCTT ATTGGGAATT ATTAGGACTA ATTACCCCAA TCTGGCAGCT   7179
TTGCAAGGAT ATTTTCTCTT AATTTGCCAC ATATTTCCAG ATGTAAAAAT TAATCTCCAA AATAGTGAGG TAGAAAGCAG AGAAAGCAGG AAAAGCAGAA   7279
GAATATAATG CCCAGAAGTG AAAATGGGAA AAGAGGTCAT CTTGTAAAAG CTGCCCAAAA ATGTAATATC TATGGTCACA TTTCTATTAC TGGAAGTGTC   7379
CAAACAGGTA ATGCAATTAC CTGTCAGAAG ACATCCATAG TTATTTAAAC AAACCTACAA TTTAATATGA GGAGGGTCCT ATACTAGCAG AAGATATGAA   7479
AACAGTAATT GCAGATATGA AAAAAATAAT TGCTATAATG TGACAAATGG TACAATCAAC TCAGGTACAA AGTGCAGGAA TAAACATTAG AGATTCTGAT   7579
TGGTATGCTG TGGAATAGCA TTCACAGAAG TGATGGACAA TATTTAAGTT GGGTCTTAAG AGATAAGTAG CAGTTTGAGA ATTGGGAGGA TAGGGTGAGG   7679
CAAAGGGTTT GATAGTCATT TTTTATTTTC TGTAATATGG GATATAAGGT CTATTGACAG GCAGGTGAAA GACATGAAGG TAGAACTTGG AGGAAAGTGG   7779
TGAAGATTTG AAACAGGTGC TAGGGGAAAC AAATGAGGGC TCTGACCATG GGTATGCAGT GGGATTGCCA GAACTATCAT AATTCCCTAA ATGCACTGTA   7879
CAGTAACTTT AAAGATTTTA TGATTTTTCA TTTGTGCTCG ACATCCTTGG TATAGTAGTA GAGAAGTAGA TTTTTTAAGC AAATTTTTTT AAAGTCTATT   7979
ATTCCTTATT TATTTTTTAT TTTTTTGAGA TGGAGTCCTG CTGTGTCGCC TAAGCTAGAG TGCAGTGGCA CGATCTCAGC TCACTGCAAC CTCCACCTCC   8079
CAGGTTCAAG TGATTCTCCT GCCTCAGCCT CCTGAGTAGC TGGGACTACA GGTGTCGGCC ACCATGCCCA GCTAATTTTT TTGTGTTTTT AGTAGAGACA   8179
GGGTTTTGCC ACGTTGGTCA GGCTGGTCTT GAACTCCTGA CCTCAGGTGA TCAGCCTGCC TCAGCCTCTC AAAGTGCTAC TTCAAAAATA AATGGCAGAC   8279
CTACTTAAGG CCTTTTATAC CCTTCTCCTC TTCCAAATTA GCCAGCATGA CAACATGCCT GTTTTTTCCG GTAAACAGTC ATTAGGCATA CAATGGCTAA   8379
TATAATTCAG CCCACAACAA TATAATTCAG TTGTGGGCTG AATTATATCT CTCTTACTCC TCCCAAAATA TAGCCACATT CTAATACTGA GAACCATTAA   8479
ACGTTACCTT ACACGGTACA AGACACTTTG AGGTTAACAA TCTTAAGGTA GGAATATTAT CTGGAACTAT CTATCCAGAC TCTAAATATA ATCACAAATG   8579
CTTTTATAAA AGGGAGGCAG AGGGAGATTG ACTGCAGAAG AGGAAGCAGG AGATGTGACA ATGGAAGCCA CAGGTTGAAA TGTTTTGAGG AAGGGGTTGT   8679
GAGCCAAGGA ATACAGGTGA TTTCAAGAAG CTAGAAAAGG CAAAGTAGCT GATTCTCCTC CAAAGCCTCC AGCAGGAATG TGGCTCTCCT GATATCTCGA   8779
TTTTAGTCCA GTGAAACTGA CTCTGGACTT CTGGCCTCTA GGACTGTAGG AGGATAAATC TGTGCTGTTT TAAGCCACTA AATTGTAATT CATCAAAGCA   8879
GCAATAGGAA AGTAGTGTGG TCAGTATTTT AGAGCAGGAG ACTTTAATTT AGACAGACGT GGGTTCAAGA ACAGATCTGC TACATGCTAT TAATTTATTT   8979
ACTTCATAAG TTTTAGTTTG TCTATAACAT GGGAGAAATA ATAGTTCCTA CTTAATAAAG TTATTTTAGG ACTAAAAGAG GTAATGAATG TAAAATGCCT   9079
GCTGTGTTAG TATTCTATTG GTGCGTAACA ATGTCACCAC AAACTCAGCA GCTTAAAACA ACACAAAGAA ATTATTTCCC ATTGGTGTAG GTTGGGGGTC   9179
TGGGTGCGGC TTAACCGACT CCTTGCATCG ATGGCTATCA AAGCTGCAAT CAAGATGTTA GCTGGACTGC AGTCTCATCT GAGGCTTGAC TGGGCAAGAC   9279
TTCTTCCAAA CTCACTGTAG TTATTGGCAG CATTCAGTGC TTTGCAGACT GCTGAACTGA GAGCCTCTGT TTCTTGTGGC TGTTAACTGA AGGCAGCCCT   9379
CTGCCCCTTG CCTTGCAGCC GTTCTGCTTT GGAAGCTCAT GATGTAGCCT CTTGTTTCTT CAAAACCAGC AAGGAGAGAG AGTCTCTTGG CAAGAATGGC   9479
GTTAAAAACT GATGTTACAT AATTATGTGT GCATAATCAC GTACTAATCC ATCACCTTTG CTGTATTTTG TGGGCTAGAA GCAAGTCACA GGTCCTACTC   9579
ATGGAAAGGG GTATCACACA AGGAAGTTAA CATCAGGAGG TGAGGATCAA AGGAATGCTC CTAGAATCAG TCTGCCATTC TCTCTATGAT GCTTGACACG   9679
                                                                249
                                                    Gly Arg Arg Asn Arg Cys Pro
ATGAATATTT TAATTTAATC TTTATTATTT TTTGTAATTA TATGACTAAA GGTGCTGATT AATTTTTTCC ATAAG GA AGA AGA AAC AGA TGT CCT   9774

Pro Pro Pro Leu Pro Ile Asn Ser Lys Ile Gln Thr His Ser Thr Thr Tyr Arg His Gly Glu Ile Val His Ile Glu Cys
CCT CCA CCT CTG CCC ATA AAC TCC AAA ATT CAA ACA CAT TCA ACA ACT TAT CGT CAT GGA GAA ATA GTT CAT ATA GAA TGT     9855
                                                                                    308
Glu Leu Asn Phe Glu Ile His Gly Ser Ala Glu Ile Arg Cys Glu Asp Gly Lys Trp Thr Glu Pro Pro Lys Cys Ile
GAA CTT AAT TTT GAG ATC CAT GGG TCA GCA GAA ATA CGT TGT GAA GAT GGA AAA TGG ACA GAA CCT CCA AAA TGC ATT G GTT   9937
                                                                                    309
                                                                                    Glu Gly Gln Glu
AGTAACACCT TGAAGAAGCT TTGCTAAAAT GAAATCTGCA TGTGTAGCTA AATGGTCAGT TCTCTTTCAA CTGTACCTTT TCAG AA GGA CAG GAG    10032

Lys Val Ala Cys Glu Glu Pro Pro Phe Ile Glu Asn Gly Ala Ala Asn Leu His Ser Lys Ile Tyr Tyr Asn Gly Asp Lys
AAG GTA GCC TGT GAG GAA CCA CCC TTC ATT GAA AAT GGT GCA GCA AAT TTA CAC TCT AAG ATT TAT TAC AAT GGG GAT AAA    10113

Val Thr Tyr Ala Cys Lys Ser Gly Tyr Leu Leu His Gly Ser Asn Glu Ile Thr Cys Asn Arg Gly Lys Trp Thr Leu Pro
GTG ACA TAT GCA TGT AAA AGC GGC TAC CTT CTC CAT GGA TCG AAT GAG ATA ACT TGT AAT CGT GGA AAA TGG ACA CTT CCT    10194
            370
Pro Glu Cys Val
CCT GAG TGT GTT G GT ATGTATGCTA CATTTACCAT CAGTAGCTAA ACTTATGGTG TAAAATTTTC CATTCTGTTA CTTTAATCTG AATATTCCTA   10289
```

```
GGAAAAGACC ACTGCTTAGG CAAATTAGAA ACACATTGCT ACAGCACCTA CTCTTCTGCT TTGCATGTGT GCATGTGTGT GTGCCTGTAT GTGTAGCTAT  10389
CTGTAGAATT TAAGTCAATG TAAAGAATTC TTAAACAAAA GTTTTTTTTG AACGCTAGTG CTAAGCACGG TTATCTGATC TGTGTCATGC CATTACCTGT  10489
AAGTTCTAAA TGACAACACA TAAATGACGA TGTGCTTCAT TAGATAAGTA AATAAGGGCA TGGTGGATTG AGTGACGGGC CATGAGCATT ATCTACTTCT  10589
TCTCCCCATA TGTCCTGCAA AGTTAGGGTA GTGCATTCTG TTCTTGACCT TGGGTAAGTC CTGCTGATAT TTTGCATCTG GCCATTCTGA ACCTTATGCC  10689
CTTCTGAGAC ATGTCCTGGT CTTTCATTTT ACTGAAAGGC ACCACCAGGC TACACTATCA TTCCTGGAGA CTACTCTTCC TGTCATCTAT CTCATTTCTT  10789
AATCAGATCA GAGCAACATC TCCACACACA TGTTTATCTA GGATATTTAT TAAATAAGAC AAAAGCAATG AAGGATGCTA AAGAAATAGA TAAAGTGAGA  10889
GAAAGATACC AAAAAGAGGA AAAAAATTGA TATTTCCTTC ATGGGTCTAT TGAGACCTGA GATGAAGAGG AATGAAAAAG AAGAAAATAA AAGTGGGGTA  10989
ATGCCAAAGA GCAATTTTAG AATTGACCAT AATCTTTAGG TTGAAGAACA GTTTCTCATT CTATAGGTTT AAGAAATGTG TATATCATAA TTCCTTCACA  11089
ATTGAGAAAA GAGAAAAATA CGTATAGTTG TTTTTAGGTA ACATATTTGC TGTCAACTCT TGCTTAGCAA ATTTTAAAAA TATTTGAGCT GAAAAAATTT  11189
TTCATATTTG TAATGAATTT ATAAAACAAT TGTCCTAAAT TATATAATAC AAATATCGAG ACACTATAGT ATGAGTTACT ATAGTATATT GTTATGAGCT  11289
                                  371
                                  Glu Asn Asn Glu Asn Cys Lys His Pro Pro Val Val Met Asn Gly Ala Val Ala
CATATTAATT TTAAAACAAC CTCTTTTCTT AG AA AAT AAT GAG AAT TGT AAG CAT CCT CCT GTT GTA ATG AAT GGG GCT GTT GCA  11374

Asp Gly Ile Leu Ala Ser Tyr Ala Thr Gly Ser Ser Val Glu Tyr Arg Cys Asn Glu Tyr Tyr Leu Leu Arg Gly Ser Lys
GAC GGG ATA TTG GCA AGC TAT GCA ACA GGA TCC TCA GTG GAA TAT AGA TGC AAT GAA TAT TAC TTA CTG AGG GGA TCA AAA  11455
                                                      431
Ile Ser Arg Cys Glu Gln Gly Lys Trp Ser Ser Pro Pro Val Cys Leu
ATA TCT CGT TGC GAA CAA GGA AAA TGG TCA TCC CCA CCT GTT TGC TTG G GTAAGAA AGAGAACACA TGGAATGTCT ACGTTTGTAC  11541

TTTTATGTGA TTTTCTTACA GTGTTTTATA TCATGAAAAT GATGATTTTG TACAAATCTT TTTTCACAAG TGTTTTCTCT CTCATTTCTT TAAGTAGAAA  11641
GTCCCTAAAA TATTAGTGTA TTTTTTCTAT TGTGGTAAAA TATACATAAC ATAAAATTTA CCACTGTAGC TATTTTTAAG TGTACAATTC AGTGGCATTA  11741
GTATATTAAC ATTTTTGTGA CCAATTAGCT TTTGGGTTGT CTTTCTCCTA GCTGAAATTC CCTCCCTGGT ATACAGCCCT CAATATCTCT GCTTGTATTG  11841
CCCTCAGAAT GTCATTTCAA GAAGAATAAT TAGTCCTATT CTTTCCTTGG TAGAGTATAT TGCTAATATA TTATAGAGTA TATTATATAG AGTATATTGC  11941
TAAATTTTAC CACCGCATCC TCATCTGAAT GGACACGTTT GTTACATCTT GTTGATTATA TAAAAGTGTT TGCTAATGAC TTAGATAATA AATGGTCTTC  12041
TAAATGTTTT ACACACACAA GCACACACTC CACAAAATGC TATGATATGA TGCAGAGGTA TATATCTTAG ATTTTTTTGC TCAAAATATA TTGTAAACAT  12141
GGTTTGTCAT ACTTCTACAA TATATACTTT AATGGCTGCA TAGTATTCCA TAATGCAGAT ATACATATGT AATTTAACCA GTTTATTGAT GGATGTTTAG  12241
TGTTCAACAT TTTTATATTC CAAACAGTAC ACTTTCAAGT AAATCTTTGT ACATGTTAAA CATATTTATT TTTGTTAAAA ATAATTATCA TGAGATCAGG  12341
AATTTTAATT CTTTTTCTAT TAATTGGCAA ATTGTCTATC AGAAAGTTTG TTCTGCTTTG TTAACCCTAC CATAAGTGAA TTAAAGTAGT AATCATCCTA  12441
ATCTTTGTTA ACTTTGAATA TTTTAATCTT GACATTTTTT GTTAAGCAGG CAAGCACAAT TATTTTAATT TGATTTATTT AAGAATTAAT TAATTTGGAA  12541
TTTTTTAACT TTCTAGTAGC CACAGGTTTT TTACATGAAT TAATTATTCA TAGTTTAGCC ATTTTTTCAT ACTTATAAAT CTATATATAA ATTATTTTAT  12641
CCAATTGAGA ATCATGTTTA TTCGAATTAT TTTTAATTTA TTTTTATTTA TAGTGTTTTT AAGTATATAT AAGTGTATGT TTTAAGTCTT GAGCTCTTTC  12741
AACCTTTTGT TAGTAGTTTC CACCTTTCAA GTAATTCTTA GGAGTCGATT CCCATTCAAA TTCAGATACA GATACATATT CACATAGAAT TTTGCTATTG  12841
GTCTCATATT TTTATAAATG TTTTTGATAT TAAGTTTTTG ATGCTTTGGT TGATATTGCA TTGAGTGGAG ATATAGTAAT CTCTAAGGAC AAAGCCTAGA  12941
TTAAAATATT TATGGCAAAG CTGGTAATGA AACAAGCTCA TAAGGAAATA ACAACATTTA AGGAAATACA TCATTCTATA TCTGGAACAA CTGCCTCAAG  13041
AGAGTTAGGT TTGGTCCTAA TGAATCAAAG CAAATACCAA ATGAAGATTA ATCATTGTTA TCAGGGCAAG AGAAAGGGGC TTTGTATATC TTGGTAACTA  13141
TTGAAACTGG AAAGGCTATC AGTACAAATT CACAACAGTA TGATAAAGCG TAGAATTTTC TCTGTAGGAA TTATTATTCC AGACATTAAG AGTCTCTCTA  13241
ACCTCTCCAT ACTCTCTTCC CTGCTCTAAT CACTCCAGTC TTTCTGGTCT CCTTACTTTT CCACAAACAA GCCAAGAATG CTCTCACTTC AGGGCCTTTG  13341
CACTGGTTGT TCCCTTTGCT TGGGACACTC TTTCATAAGA AAATGACATG GTTCACTCAG TCTCCTTCAA GCTTTTGTGC AAAGTCCCCT TCAAGCTTTT  13441
TTACTTCTTA TTTAGGCCTA CCCTAATCAC CTTACTGAAA AGTGCAACTT CTCTTCACCT AGCTCTTTTG TTACCCCCTC CTGCTCCAGT ACTCTTTTTT  13541
TCATCTAATT TATCTTATGC TGTAGTTTAA AGATGGCTAC TGTGTTAGTC TGTTCTCATG CTGCTAATAA AGACATACCT GAGACTGGGT AATTTATAAA  13641
GGAAAGAGAT TTAATGCACT CACAGTTCCA CATGGCTGGG GAGGCCTCAC AATCATGGTA GAAGACAAAG GAGAAGCATG GAACGTCTTA CAAGGTGGCA  13741
AGCAAGAGAG CATGTGCAGG GAAACTCCCC TTTATAAAAC CATCAGATCT CATGAGACTT ATTCACTGTA ATGAGAACAG CATGGGAAAG ACCCACCCCA  13841
GGATTCAATT ACCTCCAACC AGGTTCCTCC CATGACATAT GGGAATTATG GGAGCTACAA TTCAAGATGA GATTTGGGTA GGCACACAGC CAAACCATAT  13941
CAGCCACAAA TTATACAATT CATGTCTTTA TTATATTTAT TATGTATTGT AGGTGTTTCT CCCAGTTAGA AGGTAAGCTC CATGGGAGCA GGAGTCTTGG  14041
TTTTATTCAT TAGCAGGTTT TCTATTATTT TCTACATGGT AGTTAAAGGT TTTTATTTTA TAGAAAGAAA ATTATTATTA ATCTAACTGA TTGATGCCCT  14141
TAAATAGTAT TTTTGATGTT TTAAATTAAG TGGTTTTAGT TTATGACAAT AATGTATAGA TTCATTTGAT AGTAATAAAG TATATATTTT CTTATATATA  14241
TATGTGTGAG AAGCCTTAAT TATGATTTCA ATTATAAAAT ATTGCTTTTG AAACTTGTCA TAATTTTTGT TTTAAAAAAT ATCACTTTTA TATTTGTCAG  14341
                                  432
                                  Glu Pro Cys Thr Val Asn Val Asp Tyr Met Asn Arg Asn Asn Ile Glu Met
ACAAGTAAAG AATTTAAAAG AGCATTTTGT TTTACAG AA CCA TGT ACT GTT AAT GTG GAT TAC ATG AAC AGA AAT AAC ATA GAA ATG  14428

Lys Trp Lys Tyr Glu Gly Lys Val Leu His Gly Asp Leu Ile Asp Phe Val Cys Lys Gln Gly Tyr Asp Leu Ser Pro Leu
AAG TGG AAA TAT GAA GGG AAA GTC TTA CAT GGA GAT TTA ATA GAT TTT GTA TGT AAA CAG GGA TAT GAC TTA TCT CCA TTA  14509
                                                                                  498
Thr Pro Leu Ser Glu Leu Ser Val Gln Cys Asn Arg Gly Glu Val Lys Tyr Pro Leu Cys Thr Arg Lys
ACC CCA TTG TCT GAA TTA TCT GTG CAG TGC AAC AGA GGA GAA GTG AAA TAT CCT TTA TGT ACT AGA AAA G G TAAAATAATA  14590

ATGTTCTCTG CAAGTATCTT TACTTGACCA ATATCTGCTA CCTCGGTTGA AAATTTATTT ATAGTTTGTT GCTATTTTTC TTATTATTAT TCTTTTTAAT  14690
TTTGTTTTAT TTTATTTTAT TTTACTTTGA GTTCTGAGAC ACAAGTGCAG AATGTGCAGG TTTGTTACAT GGGTATACGT GTGCCATGGT GGTTTGCAAT  14790
TCCAGTAATG GAATTGCTGC GTCGAATTTC ATTTCTGGTT CTAGGTCCTT GAGGAATCAC CACACTGTCT TCCACAATGG TTGAACTAAT TTACATTCCC  14890
```

```
ATCAGCAGTG TAAAAGCATT CCTATTTCTC CCCAGCCTTA CCAGCATCTA TTGTTTCTTG GCTTTTTAGT AATTGCCATT CTGACTGGTG TGAAATGGTG  14990
TCTCATTGTG ATTACACCTT ATACAAAAAT TAACTCAAGA TGGATTAAAG ATTTAAATGT AAAAACCCCA AACCATAAAA ACTCTAGAAG AAAACCTAGG  15090
CAATACCATT CAGGACATAG GCATAGGGCA AATACTTAAT GACGAAAATG CCAAAAGCAA TTGAGCTTGA GAGCCAAAAC TGACAAATGA GATCTAATTA  15190
ACCTAAAGAG CTTCTGCACA GCAAAAGAAA CTATCATCAG AGTGAACAGG AAACTTACAG AATGGGAGAA AATTTTTGCA ATCTGCCCAT GTGACAAAGG  15290
TCTAATATCC AGAATCTACA TGGAACTTAA CACAAATTTA CAAGAAAAAA ACGAACAACA CCATCAAAAA GTAGGCAAAG GATATAAACA GTCACTTCTC  15390
AAAAGAAGAC ATTTATGTGG CCAACAAACA TATGAGGAAA AAGCTCAACA TCGCTGATCT TTAGAGAAAA TTTATACCTA TGACCCCAAT TCACCACCAT  15490
CATATTCTAA TCTATAAAGT TGAGAGGAGT GATCCAACTT CCAAACTTCA ATCAGCATAT ATACCTATAT AGGATGGCCA TGCTGACATC TTTTAATTCC  15590
AATAAGTCAA TTTTATGAAG GGAAAAAATC AAGAGGTTTT ATAAATGTTG CCAGACTATC AGAGAATTGA TCACCTTCTT GTGAGCACTC AACAACCAAC  15690
ACTGTAAAGC ATTGCTCTAC TTAGGGCTTT CTGGGGACCC ATTGTCTCAG TACTCACTGG CTATAGGGAC TATTACTTTT AATAAGGAAG AAAAGGAAAA  15790
TAATATATAC AGAAATCATG GCTTCTAGAA AAATCCCATA TACACTTAGG GAGAAATAAA ATCATGATTG TTAATACTCT CATAGTTTTA AATATGCATG  15890
TAAACACTAT AACGTGGGAT TCTAATTATA AAAATACATT GTATGGTATA ATATTATTGC AAAAACATCT TCAATATTTT AAAAATATTT GAGAACTGCC  15990
TAAAAAATGG CTTTATGGCT AGGTGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA GGCCAAGGTG GGCAGATTTC TTGAGCCCAG GAATTCAAGA  16090
CCAACCTGAG CAACATAGTG AGACCCTGTC TCTACAAAAG AAAAATGCAC CTGTAGTCCC AACTAGTTGG GAGGCTAAGA CAGGAGGATC ATTGAGCCTG  16190
GGAGGCAGAG GTTGCAGTGA GTCATGATCA TGCCACCATA CTCCAGCCTG GGTGACAGAG GGAGACTCTG TCTCTGTAGC ACTTATAAAG TACTGGATGT  16290
                                            499
                                            Glu Ser Lys Gly Met Cys Thr Ser Pro Pro Leu Ile Lys His Gly
TCATTATAGC AATTCATTGT ATACTTTAAA ACTTATTTTT GCAG AA TCT AAA GGA ATG TGC ACA TCT CCT CCT CTT ATT AAA CAT GGA    16378

Val Ile Ile Ser Ser Thr Val Asp Thr Tyr Glu Asn Gly Ser Ser Val Glu Tyr Arg Cys Phe Asp His His Phe Leu Glu
GTC ATT ATT AGT TCA ACA GTA GAC ACC TAT GAA AAT GGC TCT TCA GTA GAA TAC AGA TGT TTT GAT CAC CAT TTC CTA GAA    16459
                                                                        559
Gly Ser Arg Glu Ala Tyr Cys Leu Asp Gly Met Trp Thr Thr Pro Pro Leu Cys Leu
GGA TCT AGG GAG GCC TAT TGT TTA GAT GGA ATG TGG ACT ACA CCA CCA TTG TGT TTA G GTATGT ACTACTAAAT ATGCCTCTAA     16543

CAAAGTAAAA CTATATATTT TTAATTTGCT GTCATTTTTG AGTTAACATA ACACTAATAT AATGCTATAT TTGCTACATT TGTGTTATTT TATTCATGTG  16643
CTTATTTTAT CACTGCTTTC TATTCTTTCC TCTAGATCAT AAATACCTAT AGTTCAGTTT GTTGTAATAC ATATTTTTTT AAAGACAAAC CACACATCAG  16743
GTTGGTATGG GCTATAAGTT GAATAAGCCA TAGTGGTGTG AGTAAACTTT AACAATTATC ATTCATTGTG ATTGGTACTA TAATAGAAAA CATTAGGGAA  16843
ACATGTATAA TAGAAAACAA TGCGGAAATA AAGAAATGGA CAACTCAATA TGGAAAGATG AAGCAAGTTA ATGTTTTATT AGCAAGGCAT CCTAGTAAGA  16943
GTAACAACAT TTACTAAGGC CCTCAGAGGC ATGAAATAGC ACATCTGTAA ATAATAAGAA AATATTTTTT ATTTCCAATT CTGACTTTTA TTTACTTCTC  17043
TTATCTTCTT TCACTGGCTA GGACTTCAGT ACAATGTTGA ATGGAAGAGG TAGTGGATAT CTTTATTTGC TTCCCATTCT CATAAGTAAA ACCTTTGATA  17143
ATTCACCTCT ATGTGTTGTT ATAAATATTT TTTTAAAAAT CAAGTTAAAG GTATAAAAAC CTTTCTTTCT ATTTTGCCGA AAGTATTTTT AAAGAATCTT  17243
CAATCTGTGT TTAATTTTTA GCAATATCTG TTGAAATAAT TACATGATTA TTACTTTTGT TTTTTCCCCT GCAAATTTAA TGAATTATAT TCGTTCTTTT  17343
GAATGTTAAA GCATGATTCT ATCCCATAAG AAATTCCACA AGGTTGTGAT AAATTATGCA TATTATATGT TATATTTTTA TTTATCAATA TTTTACTCGA  17443
TCTATTTTCA TGAGAAGACT GGTTGGTTAT TTTCCTTTCT GGTTTGATAT CAAGTCTGTG CTGACCTCAT CCCATGAATG GGAAGTGTTC CGTCTTTTCT  17543
GGTTCTCTGG ATTGATGCTA TTTCTTCCTT TATAATTTTA AGAGTTAACC AGAGAAGCCA TCTGGGACTT GAGTTTTTCC GTGAAAAGAC AAATTGACAG  17643
CTTCGATTTG TTTAACAAAT AATGGTAATT ATTTCTTCTT CTGTCAGTTT TGGTAGGGAG TGGTTTCTTG ATGGTTCATT ACAACTACAT TGTCAAATCT  17743
ACTAACATAA AGTAGTTATG ACATCTTCCT ATCCTCTTTT TGATGTTTTT ATTGATGTAT CTTTTGTTAT TACTTCTATT AATATGTATC TGTCTTGACA  17843
TGCCTTGCTA AGGGTCTATA CATTTTATTA GTGTTTTCAA AAAACCAACA TTTGCATTTG TTGATTCATC AAATTTTGCT TTCTATTTTA TTAATTATGC  17943
TGTTATCTTT ATTATTTCCT TTCTTCTATG TTCTTGGATT TCTTTTGTTA CTCTTTTTCT GGCTTGTTTA GATGAATACA TAGATTACTA ATTTTCAGCA  18043
TTTATATTTT TTCTAATACA TGAATTTAAA AATAGAAATT TCTCTCTGAG CACAACTTTA ACTTCATCCA AACCATTTTT ATATCTTGGA TTTTTACTGT  18143
CATTCAGCTC AAAAATTTTA TAGTATTCAT TTTGATATAT TCTCTGTGCT CTGAATCATT TAGAAGTTTC TTTTTAAAGT TCTGAATATT TAGGGGATTT  18243
TCTAGTTTTC TTTTGGTTAC TGATTTCTAG TTTAATTCAA TTAAGACATA CTTTGAATTA TATCTGCACT TTAAAATTTG TTGATATTTG CTTTATAACA  18343
AGCATGTATT TGCTTTAGGT AAATGTTCTT TTCATTGAAA AGTTAACTTT GTCATTGTTG GGTCACTGTT TTATATATGT GAATTGGATT GAGTTTTGTT  18443
TTTGATGTTT AAGTTAATCC AAAGCTGTAC TGATTTGTGT GTGTTTGTTT TCTATCAATC ACTGAGAGAA GCATGTTAAA GTCTCACTTG GCATTTTAAA  18543
ACATTACTAA TTTTAGTTCT GTCCATTTTT GCTTTTGTAT TTTGATACTA TTTCATTAGG TACACACACA AATTGAGACA CTAAATCTTC GTATCTTGGT  18643
GTATTGGCAT TGTCTAATTT TAAAAGACCT CCTTCATATA TAGTAATGCT TTGCCTTTTT TTTTGCCTTG AGGTCTCCTT TAGCTGCACC AAGGGGCTTT  18743
TGACAAGCAT TCATGTAGTA CACATTTTTC CATCCCTTAA CTTTCAACCT TCTGGGTTTT TACTTTTAGT GTGCCTCTCT TATAAGCAAT ATATACCTTA  18843
TTTTAAAATA TACTTATGCA ATTTATAATT TTTTCCCTTG TAATGGAATA GTCTATTTAC ATATTCATGT AATTACTAAT ATATTTCTGT TTATACCTAC  18943
CAACCCTACC AACTTTCTAT ACTAGGTGTA TGTTACTATA CTCCATGTAT ATCCCATTGG TCATCTCCCT CTTTCTCTCT CTCTCTCCAT ATAGGTGTGT  19043
GTGTATATAT ATATACACAT ATATATATAT AATATTTTAA GTTCTAGGGT ACATATGCAC AATGTGCAGG TTTGTTACAT AGGTATACAT GTGCCATGTT  19143
GGTTTGCTGC ACCCATCAAC TCTTCATTTA CATTAGGTAT TTCTCCTAAA GCTATCCCTC CCCCAGACCC CCACAGACAG GCCCTGGTGT TTGATATTCC  19243
CTGCCCTGTG TCCATGTGTT CTCATTCTTC AACACTCACC TATGAGTGAG AACATGCAGT GTTTGGTTTT CTGTCCTTGT AATAGTTTGC TTACAATGAT  19343
GGCTTCCAGC TTCATCCATG TCCCTGCAAA GGTATGAACT CATCCTTTTT AATGGCTGCA TAGTATTCCA TGGTGTGTAT GTGCCACATT TTCTTAATCC  19443
AGTCTATCAT TGATAGACAT TTGGGTTGGT TCCAAGTTTT GCTATTGTGA ATAGTTCTGC AAAAAACATA TCTGTGCATT TGTCTTTATA GCAGAATGAC  19543
TTATAATCCT TTGGGTATAT ACCCAGTAAT GAGATCACTA GGTCAAATGG TATTTCTAGT TCTAGATCCT TGAGGAGTCG CCACACTGTC TTCCACAATG  19643
ATTGAACTAG TTCATACTCC CACCAACAGT GTAAAAGCGC TCCTATTTCT CCACATCCTT TCCAGCATCT GCTGTTTCCT GACATTTTAA TGATCGCCAT  19743
TGTAACTGGA GTGAGATGGT CTCTCATTGT GGTTCTGATG ACCAGTGATG ATGAGCATTT TTTCATATGT CTGTTGGCTG CATAAATGTC TTCTTTTGAG  19843
AAGTGTCTGT TCGTATCCTT TGCCCACTTT TTGATGGGGT TGTTTTTTTC TCGTAAACTT TTTAAGCTCT TTGTAGATTC TGGATGGTAG CTCTTTGTCA  19943
GATGGATAGA TTGCAAAAAT TTTCTCCCAT TCTGTAGGTT GCCTGTTCAC TCTGATGATC GTTTCTTTTG CTGTGCAGAA GCTCTTTAGT TCAATTAAAT  20043
TCCATTTGTC TATTTTGGCT TTTGTTGCCA TTACTTTTGG TGTTTTAGTC ATGAAGTACT TGCCAATGCC TATGTCCTGA ATGATATTGC CTAGGTTTTC  20143
TTCTAGGGTT TGTATGGTGT TAGGTCTTAC ATTTAAATTT TTAATGCATC TTGAGTTAAT TTTTGTATAA GGTATAAGGA AGTGATCCAG TTTCAGCTTT  20243
```

```
CTACATATGG CTAGCCAGTT TTCCCAACAC CATTTATGAA ATAGGAAATC ATTTCCCATT TCTTGTTTTT GTCAGGTTTG TCAAACATCA GATGGTTGTA  20343
TGTGTGTGCA TTATTTCTGA GGCCTCTGTT CTGTTCCATT GGTCTATATA TCTGTTTTGG TACCAGTATC ATGCTGTTTT GGTTACTGTA GCCTTGTATT  20443
ATAGTTTGAA GTCAGGTAGC CTGATGCCTC CAGCTTTGTT CTTTTTGCTT AGGATTGTCT TGGCAATGCG GGCTCTTTTT TGGTTCCATA TGAACTTTAA  20543
AGTAGTTTTG TCCAATTCTG TGAAGAAAGT CAATGGTAGT TTGATGGGGA TGGCATTGAA TCTATAAATT ACCTTAGGCA GCATGGCCAT TTTCACGATA  20643
TTGATTCTTC CTATCCATGA GCATGGAATG TTCTTCCATT TGTTTGTGTC ATCTTTTATT TCGTTGGGCA GTGGTTAGTA GTTCTTGAAG AGGTCCTTCA  20743
CATCCCTTGT AAGTTGGATT CCTAGGTATT TTATTCTCTT TGAAGTAATT GTGAATGGGG GTTCACTCAT AATTTGGCTC TCTGTTTGTC TATTATTGAT  20843
GTATAGGAAT GCTTGTGATT TTTGCCATTG ATTTTGTATC CTGAGACTTT CCTGAAATTG TTTATCAGCT TAAGGAGATT TTGGGCTGAG ACGATGGGGT  20943
TTTCTAAATA TATAATCATG TTATCTGCAA ACAGGGACAA TTTGAATTCC TCTTTTCCTA ATTTAATACC ATTTATTTCT TTCTCTTGAC AGATTGCCCT  21043
GGCCAGAACT TCCAACACTA TGTTGAATAG GAGTGGTGAG AGAGGGCATC CCTGTCTTGT GCCAGTTTTC AAAGCAAATG CTTCCACTTT TTGCCCATTC  21143
AGTAAGATAT TGGTGGTGGG TTTGTCATGA ACAGCTCATA TTATTTTGAG GTACATTCCC TCAATACCTT GTTTATTGAG AGTTTTTAGC ATGAAGGGCT  21243
GTTGAATTTT GTCGAAGGCT TTTTCTGAAT CTATTGAGAT AATCATGTGG TTTTTGTCAT TGGTCCTGTT TATGTCATGG ATTATGTTTA TTGATTTGTG  21343
TATGTTGAAC CAGCATTGCA TCCCAGGGAT GAAGCCATCT TGATCGTGGT GGATAAGCTT TTTGATGTGC TGCTGGATTT GGTTTGCCAG TATTTTATTG  21443
AGGATTTTCA CATCGATGTT CATCAGGAAT ATTGGTCTAA ATTCTGTTTT TTGTTGTTGT TGTGTCTCTG CCAGGCTTTG ATATCAGGAT GATGCTGACT  21543
TCATTAAATG AATTAGGGAG GATTCCCTCT TTTTCTATTG ATTGGAATAG TTTCAGAAGG AATGGTACCA GCTCCTCTTT GTACCTTTGG TAGAATTCGG  21643
CTGTGAATCC ATCTGGTCCT GGGCTTTTTT TGATTTGTAG GCTATTAATT ATTGCCTCAA TTTCAGACCC TGTTATTGGT CTATTCAGAC ATTCAACTTC  21743
TTTCTGGTTT AGTCTTGGAA GGGTGTATGT GTCCAGGAAT TTGTCCATTT CTTCTAGATT TTCTAGTTTA TTTGCATAGA GTGTTTATAG TATTCTCTGA  21843
TGGTAGTTTG TATTTCTGTG GGATTGGTGG TGATATTCCC TTTATTATAT TTTATTGCAT CTATTTGATT CTTCTCTCTT TTCTTCTTTA TTAGTCTTGC  21943
TCTCAGTCTA TGTAGTTTGT TTATCTTTTT TAAAAAAACA GTTCCTGGAT TCATTGATTT TTTTGAAGGG ATTTTTTGTG TCTCTATTTC CTTCAGTTCT  22043
GCTCTGATCT TAGTTATTTC TTGCCTTCTG CTAGCGTTTG AATGTGTTTG CTCTTGCTTC TCTAGTTCTT TCATTGTGAT GTTAGGGTGT CGATTTTAGA  22143
TCTTTCCTGC ATTCTCTTGT GGGCATTTAG TGCTATAAAT TTCCCACTAC ACACTGCTTT AAATGTGTCC CAGAGATTCT GGTACGTAGT GTCTTTGTCC  22243
TTACTGGGTT CAAAGAACAT CTTTATTTCT GCCTTCATTT CGTTATTTAT CTAGTAGTCA TTCAGGAGTA GGCTGTTCAG TTTCCATGTA GTTGTGCAGT  22343
TTTGAGTGAG TTTCTTAATC CTGAGTTCTA ATTTGATTGC ACTGTGGTCT GAGAGACAGT TTGTTGTGAT TTCTGTTCTT TTTCATTTGC CAAAGAGTGT  22443
TTTACTTGCA GTTATGTGGT CAATTTTGGA ATAAGTGCAG TGTGGTGCTG AGAAGAATGT ATATTCTTTT GATTTGGAGT GGAGAGTTAT GCAGATGTCC  22543
ATTAGGTCTG CTTGGTCCAG AGCTGAGTTC AAGTCCTGGA TATTCTTATT TACCTTCTGT CTCGTTGATC TGTCTAGTAT TGACAGTGGG GTGTTAAAGT  22643
CTCCCATTAT TATTGTGTGG GAGTCTAAAT CTCTTTGCAG GTCTCTAAGG ACTTGCTTTA TTAACCTGGG TGCTCCTGAA TTGGGTGCAT ATATATTTAG  22743
AATAGTTAGT TCTTCTTGTT GAATTGATCC CTTTACCATT ATGTAATGGC CTTCTTTGTC TCTTTTGATC TTTGTTGGTT TAAAGTCTGT TTTATCAGAG  22843
ACTAGGATTG TAACCCCTGC TCAGCAATGG TAGACGACCC TCCCCCCATA ACGCTGTCGT GTCACAGGTT GATCTCAGAC TGCTGAGCTA GCAGTGAGCA  22943
AGGCTCCATG GGCGTGGAAA CTGCTGAGCC AGGCAAGGGA GGGTATCTCC TGGTCTGTCG GTGGCTAAGA CCTTGGGAAA AGCACAGTAT TTGGTCAGGA  23043
GTGTACCATT TCTCACTGTG CTGCCCAGGC TGGAGTTCCT TAGAAACAAA TCAGAAACAA ATCTCTCCCT CTCTCTCTTT CTCTCTCTCT CTCTATATAT  23143
ATATAAATAT ATATATATAT GAGTTAGTTC CAAGATGGCC GAATAGGAAC AGCTCCGGTC TGCAGCTCCC AGGGAGATCA GTGCAACACA ATGAGACTCT  23243
CTTTGTACAT TTTGTTCTTC AAGTTAGTAG CATTCCTTGA ACCTTGATGT ACTTGATATC CTTCATGGGT TTTAGAAATG TTTGGCTAAT ATCTTTTAAA  23343
ATGTACATTT CTCACCTCTT CTGTGATGCC ATATTACATA TATTACATCT TTTTGCTATG CCATTGTCTC TTTAAATCTA TATTTTCTCT CATTTTCTTC  23443
TTGTTTCACT GTAGTAATTT TTTGGTAATT TTCTCATTGT ATCATTTCTT TGTTTCTATT TGGTAATTTT CTTCTAGTTT AACAATCCTC TCTTTAGATG  23543
TATCCAATCT GCTTTAAACT CATCTTTTGA GTTCTCAATT TCAGTTATCA TCTTTGCATT TCTATTTTAT ATTTTCTTTT TGTAGATTTG AGTTATCTGA  23643
TGAAAGATCC TAGCTTAACA TCAACTTCTT GAACATATAA TTCATACTTT TTATTAAGGC TTGTCTGCAA ACTCCAATAT CTGGATCTAT TTTTTAGCCC  23743
CAAATTCTTG ATATTTTTGT TAACAGTTGA TTGAATTCCA AATATTATAC GTGGAAAATT TCAGGATGTT TGGATGATGT CAGCTTACTC CAAAGAAGAC  23843
TCATCTCATC CTCTGGCTGA TGATTAGTGT AGAGGCAGTT GGTTTTAATA CAATTAGAAA TAAAGCTTTC TCTATGTTTG TTTTCAGACC TCTGTTTGCC  23943
TTTTCTCACA GACTGTAACT CTCCAGGGCT CTTTACAGAG GGTTTGTTGT GTGTATTAAA GCTTTTTCTC CTTGACTGTC TCTGATATTA CTGGGAGACT  24043
GCCAAATGCC CACTGTAACT TTTTATTTAC TTTCTGCTTG GCTTCTCATC CTATGCCTCT ATACAAGTCA AGGATAAGCA GTACCTTAGG GAAAAGCTGG  24143
TGCAGAGCCT CAAACTCATA TCTCTACACT TCCTATCCTT CTAAATTCTT TATTCTTCAA GTCCTGGCTG CTTTGGTTTA CCGAAATTCT CATTTTTCTA  24243
TCTTCAACTC CATGAGATTT TCAGAAGTTC TGCTGGCTGT ATGCCTAAGC TTACCATTCT CTATCATCTA ACCTCTCAGG TCAGACTTAG AGTCAACAAA  24343
CACCTCAAGA GAAAAAAATG GCACATAGAA TTTTTTTTTT TTTTTAGACA GTCTCCTTTT GTTGCCCAGG CTGGAGTGCA GTGGTGCGAT CCCTGCTCAC  24443
TGCAACCTCC ACCTTCTAGG TTCAAGAGAT TCTTCTGCCT CAGCTTTCCC GAGTAGCTGG GATTACAGGT GCATGCCACC ATGCCCGGCT ATTTTTTTTT  24543
TTTTTTTGTA GAGATGGGGT TTCACCATGT TGGCCAGGCT GGTCTCAAAC TCCTGACCTC AGATGATCCG CCCGCCTTGG CCTCCCAAAG TGCTGGGATT  24643
ACAGGTGTGA GTCACCACGC CCAGCCTCAC ATAGATGTTG AACTCACCTG AATACATTTT TTCCAGAACC TTACTTCCTT GACTAAGTTC CTTGGAGGTT  24743
TTCTGATACC TCGAAATGCT TCCTGTTTTT CAGATTTTCT GGTTCTTAAT AGAAGTGCTG GTTTGATATA ATCTCCTCTG TCATAGCTGA AAGTTAAAGG  24843
TCCATGAATC AACTTTTAAA TTTATTTTAA TTTAAATCTC ATGTTCTGTA GAACAGTGAC ATCGTAAATT ATAGTGCTAT TCGACTTCAT TTTCACAACT  24943
ATTTTGTAAT GAGTGTAATA ATTCTTGTTC ATTTTCATCA CCTGTAGCAC TTCCAGCATA ATGTTTGACA CATAGGCATT TAATAAGTAG TTGTTGAAGA  25043
AAGTAACTGA ACACAGTGCT CTCAGCTGCG GTTTCACTTT CTGTGGTTTC AGTTACCTGC TATAAACTGC AGTCTGAAAA CGTTAAATTG AAAATTTCAG  25143
AAATAAACAA CTTATAAATT TTAAATTATG CACTTAGTAG TATAATAAAA TTTCATGCCA TTCCATCCTT CCTGCCCCAG ATATGCTAAA AAGAAGCTGT  25243
AAAGTGCTTC CTTTAAGTGA AAGGGCGGAA ATTCTTAACT TAAAAAGAAA GGAAAAATAA ATGTATGCTG AGGTTGCTAA GATCCATGGT AACAATGAGT  25343
CTTCTGTCTA TGAAAGTGTG AAGAAGGAAA AAGAAATCCA TGCTAGTTTC GTTGTTGCAC CTCAAACTGC AAAAGTTACA AGCCACAGTG TATAATAAAT  25443
GTTTAGTTGA GGTGGAAGAG GCATTAAATT TGTGGGCAGA AGACATGAAC AGAAACATGT TCCAAAAGAA TGACAACAGG GTTTAGCATT ATCTGCGATT  25543
TCAGGCATGC ACTGGGGCTC TTGAAATGTA TCCCCTGTAG GCAATCAGGG AGGGCATACT GTAAGTAAGT TTGTTAATGT ACTACAGGTA TATGATTGAT  25643
ATGCCATATT AGCAACTCTG GACCTTTTTA CCTGTCTCAG AAGTTATTCC TTAATGCTAG GTCATCCAGG AGACAAATAA GGACATAAGA AATCCCATTT  25743
TCTGAAATAA TTTGATATTT GATATTTTAG AAATATTAAT GTTTTTATTC TGTAAAAAAA ATCGAAGAGG TTTTTTTGTG TATTTTAATA CATACTTTAC  25843
CATTCTCTTT GGTTTACGAT ACATGGGAAA GGACACAACT ATACTTTGTA TGCTTAGAAT TTAAAACGTC TTAATCCAGT CTTGATATTT CTTAACTAGA  25943
AATTTGAAAG AGAAGTAACT GTCAACCATT TTGGATGGCA TGAGGTTATA AAGCTTGGGT CAGTGTTGAG AATCTACATT TCCATAAAAT GTTATTTATA  26043
CAGCAATTAA TAATTTGTCT CATGAGAATT TTATTTGAAC TTTCTGAAGA AGGAAAAGGG AAAAGAGCAT GATATCAATG ACCATTAGTA ACTGTCAATG  26143
CTTCCCCCCA AATACCTTTT TGTTTGTTTG CTATCTTTGA AAAAGCTAAT GACATAAAAC TAAAATACCT TTTAGTTTTT AAAACTGGAA GAAAATACCT  26243
TTTCTCCAAA TCCAGGACCA CAGAGAGAAA TGGCTCACTG TTTTCTGTGA GATTTGAACA TATTTTCTGA CTCAAATATT AATTCTGATT AACTGACTTG  26343
```

```
CTGCTCTTAG GCCTGAGAAT TTTAAGGTCC TCATGAGTAG GCAACACAAC GATTCCTGTT CTTACCAAGA TGTAACAAGA TAGTTTTCTT CTTTTTCTTT  26443
                          560
                          Glu Pro Cys Thr Leu Ser Phe Thr Glu Met Glu Lys Asn Asn Leu Leu Leu Lys Trp Asp Phe
CTTTTTTTTT AAATTGTTGT TCCCAG AG CCA TGC ACA TTA TCT TTT ACT GAA ATG GAA AAG AAT AAT TTA CTT CTG AAA TGG GAT TTT 26531

Asp Asn Arg Pro His Ile Leu His Gly Glu Tyr Ile Glu Phe Ile Cys Arg Gly Asp Thr Tyr Pro Ala Glu Leu Tyr Ile
GAC AAC AGA CCA CAC ATT TTG CAT GGT GAA TAT ATT GAG TTT ATT TGT AGA GGA GAT ACT TAT CCA GCT GAA TTA TAT ATT    26612
                                                                                          631
Thr Gly Ser Ile Leu Arg Met Gln Cys Asp Arg Gly Gln Leu Lys Tyr Pro Arg Cys Ile Pro Arg Gln Ser
ACT GGA TCT ATA CTT AGA ATG CAA TGT GAC AGA GGG CAG TTA AAA TAT CCA AGA TGT ATT CCA AGA CAA AG GTAAGAAGTT     26693

TTTTTTTTTT GGTCAGATTA TTTTTTTTGG TCAGATTGTT ATTCAACATT TCAGAAATGT TATGTTCTAT ACCACTCATA CTTTATGGCA ATTGCAAAGA  26793
AAAAATAATA ATATTTTCTT CATTTTTTTT ACACCAAACA AACAACTGGA AGTGGTAATC AATTCAACAG GCCTTCTGAA TTGTCTGAAG TGAAAGCAAC  26893
GTTTGCCATT TGATATTTGT GACACTTTGG CAAATGGAAA AATAAACCAA TTTATACGTG CAATATTTCA ATGTGCAAAA ATTAACCTAT ATGTTCTGTG  26993
ATAAGGACAA AAATCACACT TTCTCCTACT TCTCAATATA AGTTCAAGCA CTACTAATAA AGGTTGGAAA ATGTGGTATT TATTTTCAGG ACCACATTTT  27093
TGTGTAACTT TTGGAAATTA CATACTAAAG TGTTTCCTCA TATGACAGTT TAAAAAATAT ATCCTCTGCT ACTTACAGGA AATAATTTAA CAATTTTCTT  27193
AGAGAAATGT TAAAAAAGTA TAATAAAATG ATGTAAACCA TTTTGTAGGA TCACAATTTT AACCCTTTTT GAAGACACTG TATATAAAAA ATGGAAATTA  27293
TATTTTTTGT TCAACTAAGA ATATTAAAAA TATAAGTAGT GGCAAAGTGC ATGTGGTCCT GTTTTAAAAA ATGCAAACAA TGAACTGTAT GAACTACATA  27393
AAGTCAGTTT TTTAGATGCT TTGACCACTC TAGTAAAAAA ATATTTAAAA GACACACAAT AGTGGGAAAA ATTTTTATTT TTATTTCTAA TTGTGGTAGC  27493
ATCAGGGTGG AAAAAATTTT ATATGCTATT TAAATGTTTT GGTACAGTAT GAATTTCTCA CTTTTAAGAG AAAAAGGTAA ACACGTAGAA AATATTTAAA  27593
AAATCATGAG TAGTGTAGTA AATAATATTG CATTTTTAAA AGGGAACTGG TGGTATTATT TACTAAGTAA GATCATTTAT TCTTATTAGT ACTTAATAAA  27693
AACAATGACT CATTATATGA AAAATTGAAA GGCTGAGATT GTAATTAACA CCTGACTGCA ATTGATGCTT ATTTCAAAAA TCTCTCTTTT TCCCCTCAAG  27793
  632                                 641
  Thr Leu Ser Tyr Gln Glu Pro Leu Arg Thr Stop
C ACT CTG TCT TAT CAA GAA CCC TTA AGA ACA TAG AAATGAAT GGCAGAAAGA GGAGTCATAT TTCAATACAT CATGAAATTC CTTATAAAAT  27885

ATAATTTTGA GGAAATAAGT TAAAAACTTC GGAGTTTTTC CTTGCTTAAA TATTTGAATC TAAATTATTT GTGCTGAACA TTTCATTATT TATAAATGAA  27985

AACCAATAAA TGTCTAATTT TTAGTTTGTA TGATTTGATG TAGAACTTAA TCTCTTTTGT ATGTCTTCCA TCTGTTACAA ATTAATCTAC ACATCATATT  28085
TTCATTTCCT AATGCCCATT GACTAGACTA TATAGTTGTT TTCAATATAC TGTATGTGCA AAAATTTCAC ATTCACATTC GTCATCATCT TTATTAATAT  28185
TAAGGCTTTT GTTACCCTTA ATTCAAAGTT TATGTCACTA TTAAAGGGCC AAACACACTA GAGATAAGAA ACTATATGCT GTATTAGAGG ATCTGCCTTA  28285
TTAAAGGCAG AAGAGACTGC CCTTCAGACT TTCTAAATGT AAAAAACTTT TAAAATTAAC TGTAATATTT GCTACAACGT TAATAACCAA ATTGTTTATG  28385
AGGTGGTGTA CTACCATATT TGAACATGTG CTCAAATATT GTTAAAGAGA CACAATTAAA GAAAGAATGA CCCTTGGAAT TTTATTTAAT TTTATTTATT  28485
TATTTATTTA TTTATTTATT TATTTATTTA GAGACAGAGT CTTGCTCTGT CGCCCAGCCT AGAGTGCAAT GGCATGATCT TGGCTCACTG CAATTTTTGC  28585
CTCCCGGGTT CAAGCAATTC TCCTTCCTCA GCCTTCCAAG TAGCTGGGAT TACAGGCGTG TACCACCACG CCCGCTACGG TTTTTTTTTT TTTTTGTATT  28685
TTTAGTAGAG ACAAGGTTTC ACCATGTTGA CCAGGCTGGC CTCGAACTCC TGACCTCATG TGATCCACCC GCCTTGGCCT CCCAGGACTC TTGGCATTTT  28785
TACATTTTAC ACTTTCAAGT CTTTTTTTAT ATCAAATAAT TTTTGGAATT ATGCAGACTA CAAAATCTGA AATAAGAGAG TCCTACAAGC ATCAAAAAAT  28885
GAGTGTCAGT ATTATCAAGG TCTGTTAGAG ACATGAAACA GCAACACTGT CACTTTTTCA TTTGAAGTGA GAGTTATACT TCTAATTATT TCAGCTCATT  28985
CATAAGCATA TGTAATAGCA CTAGTACCCA GATACTAGGA CCCCGATTAG GATTGGTTGT TGGTATTAGA ACGAGTTTTC CTTTCTAGAA ACTACTCAGT  29085
TTTTCTAACT CAAGATCTTA GTTTCTCATT ACGATGATAT TCTCAAATGC CCAATGTACA ATTCTCTACT GCTGGGTATG AAATAGCTTC TTTTGGAAAT  29185
GATGCCACAG CTGACCTAAC TCTGTGGATT AAATAAGTAT TAATTCATTC AATGTTAAAC ATGTATATTT CAGAACTGTC CTTGAGTTCA TATTCATTAT  29285
TTTAATTAAT CTCCTGTAGT AGGTTCCCAC AGTGGATCTC ATAAAAATGG AAATTTAATT AAAATTAAAT ATCAACAGAA AAGCATTAAT CTTTTTAAAT  29385
AAATTTCCCA ATCTATCAAT TAAATAACAT TATTGAGCAC CTGCTATATC CCAGGGAACA TGCTGGGTAG TACAGGATTC AAAATGATTA AAATTCACCT  29485
CGTTATTTCA AAGAACTCAC AAAGTAAAAA ACTACTTACT TAGTAAAAGA GATAAACTTG TACTAGTGAT TGCAGACTCA AAATGGGTGA TGATTGTGGC  29585
TTATCTAGTG ATATTTCTAG TGAACTGGAG ATGGCATGCA CTGTTTAACC TGGACAACTA CTTCGCTCAA GGTGAATAAT TGTCATGGGG GAATATTGTA  29685
TCCAGTATTG CCATATGTTC TGATTTGTTA ATTGTGAAAT GCACTGATTT TTATATGTTG CGAGGAGATA AAAAATTATT TAGGCAGTGC AGCTCAAAGT  29785
AAACTCCACA TGCTAAAGAA ATAAAAATTT TGAAAAGTGC AACAGTGGAA GCATGTACGA AGTAAAGTGG TGGCACAAAG TAGGTAGTAT TTAACCTTTT  29885
TTTTTTTTTG TATGAATGTG GTATAGGAAG CTTCTTGGAA AAAAGTTAAA GTAAATCTGA GGTTTACAAA ACTAGAAAAC ACACTTCAAG TAGAAGGAAT  29985
GGCATTTGCA AAGTCAAAGC TCATACTCTA TGCTAGGCAT TACACCGAGC ACTGAAAATA AAACAATGAA TAAGACACAG ACTGTTTCTG CTCTTGTGGA  30085
GCTTGTCTCT CGGAGCGGAA GAAAGATATA AGCAAATCCT TACAAAATAA GGCATGGCAG CAAAGACACC ATCTTTGTCA GTAAACACCA GATGATACTG  30185
CCTAGTACTG GGTGTTTAAA AAGTCTTCAG TGAGGAAGTG ATATATAAGT GACATATAGG CTGAGATTGG ATGGGAGAAA TGATAAGAAT TC         30277
```

FIGURE 3: Nucleotide sequence of the gene for the b subunit of human factor XIII. Nucleotide 1 is shown as the first nucleotide of the codon for the initiator methionine. The nucleotide sequence upstream from the initiator methionine is shown with negative numbers. The numbers at the right margin correspond to the last nucleotide on each line, while the amino acid sequence derived from each of the exons is placed above the corresponding DNA sequence. Intron–exon boundaries are shown by arrows below the sequence while the polyadenylation signal AATAAA is underlined. The poly(A) addition site is indicated by a solid circle above the sequence.

O family of repeats (Table III). When the consensus sequences for the Alu repeats, which are roughly 300 bp (Deininger et al., 1981), were compared to the gene for the b subunit, four Alu sequences were found within the gene (introns A, E, I, and J) and one was found flanking the 3′ end of the gene (Figure 1). The average homology to the consensus sequence was 71% and the repeats were present in both orientations. Although Alu sequences have been seen in clusters in several genes (Degen & Davie, 1987), they are present in the b subunit at roughly 6000-bp intervals. This is similar to

Table III: Repetitive Sequences in the Gene for the b Subunit of Human Factor XIII

	nucleotide position	strand	% homology	size of direct repeat	direct repeat	mismatches
Alu						
1	3600–3893	+[a]	65[b]	8	TA$^{C}_{G}$GAAGT	1
2	7985–8312	–	76	8	CAAAT$^{G}_{T}$AG	1
3	16007–16306	+	56	5	CTTTA	0
4	24374–24669	–	85	8	CACATAGA	0
5	28468–28806	–	74	8	TTGGAATT	0
*Kpn*I						
1a	14793–15002	–	45[c]	6	CAATTC	0
1b	15003–15460	+	63			
2	19078–23107	–	89[d]	24	CTCCCTCT$^{T}_{C}$TCTCT$^{C}_{T}$TCT[e]	2
O-LTR						
1	13575–14077	+	62[f]	11	GTAGTT^{T}AAAG	1

[a] (+) insertion from 5′ to 3′ on the mRNA strand; (–) insertion of the sequence in the opposite direction. [b] Compared to the consensus sequence of Deininger et al. (1981). [c] Compared to the pCD-4 sequence of DiGiovanni et al. (1983). [d] Compared to the sequence of Kazazian et al. (1988). [e] A part of the complete repeat that can be found in Figure 3. [f] Compared to the sequence of Sun et al. (1984).

the spacing of about one per 3000–5000 bp in the human genome that contains 300 000–500 000 copies (Schmid & Jelinek, 1982). Direct repeat sequences of 5–17 bp normally flank each of the Alu sequences and are thought to have been duplicated at the time the Alu sequence was inserted in the gene. These flanking sequences were also observed at each end of the Alu sequences in the gene for the b subunit (Table III).

Two regions in the gene for the b subunit contain partial *Kpn*I repeats, and these occur within introns I and J (Table III). *Kpn*I repeats are variable in length, with some being more than 6 kb in length (Adams et al., 1980). They occur about 10 000 times in the human genome. The first partial *Kpn*I repeat in the gene for the b subunit was 666 bp long and was located in intron I. It has been reported that *Kpn*I repeats may contain scrambled arrangements of common sequences (Lerman et al., 1983), and this was observed with the *Kpn*I repeats in the gene for the b subunit. The first *Kpn*I repeat has homologous regions to the *Kpn*I clone pcD-*Kpn*I-4 (DiGiovanni et al., 1983) with a head to tail arrangement of the two homologous regions. It was flanked by a 6-bp direct repeat. Direct repeats also flank *Kpn*I elements as with Alu repeats but are usually shorter and less perfect. The second *Kpn*I repeat, found in intron J, was also made up of scrambled regions homologous to other *Kpn*I members. This repeat was 4029 bp long and was 89% homologous to a *Kpn*I repeat of approximately 2200 nucleotides flanking the 3′ end of the human β globin region on chromosome 11 (70001–73326) (Hattori et al., 1985). It was present, however, in an inverted orientation in the gene for the b subunit. This region in the b subunit was also homologous to other Line-1 (L1) members, including an 89% homology to the human factor VIII gene L1 element insertion of 3785 nucleotides in exon 14 (Kazazian et al., 1988). An additional L1 element of approximately 1500 nucleotides was also present in the 5′ end of the gene starting at nucleotide –2130. The direct repeats are of a $d(CT)_n$, (purine, pyrimidine)$_n$ motif as shown in Figure 3 and include nucleotides 19008–19077 and 23108–23163.

Another type of repetitive element is the "O" family repeat, which is usually 350 bp in length and flanked by direct repeats (Sun et al., 1984). It occurs less frequently than Alu sequences and is estimated at 5000–30 000 copies in the human genome. The 502-bp O repeat for the b subunit is located in intron J and is 62% homologous to the DNA sequence described by Sun et al. (1984).

Regions of alternating purines and pyrimidines have been shown to favor the formation of Z DNA (Wang et al., 1979). Two short regions were found within the gene for the b subunit, the first being 34 bp in length and located within intron G at nucleotides 10351–10384. The second regions occur in the direct repeats before and after the second *Kpn*I repeat in intron J (nucleotides 19032–19077 and 23138–23163) and were 46 bp and 26 bp long, respectively.

Other Internal Homology within the Gene. The remaining sequence of the gene for the b subunit was then analyzed for other dispersed or tandem repeats. A tandem repeat of 25 bp in length was found in the 5′ noncoding region at nucleotides (–290)–(–266) and (–265)–(–241). An inverted tandem repeat of 13 bp was also identified at 2457–2269 and 2271–2283. This sequence has strong homology to sequences within the genes for GM-CSF, as well as to those in yeast and discoideum DNA. Another inverted tandem repeat of 19 bp was located at 8379–8397 and 8409–8427 within intron E, just after the second Alu repeat. There were also a large number of short sequences within the gene for the b subunit that share >85% similarities with other genes. Many of these were found within the introns of other proteins containing sushi domains. The significance of these short homologous regions is not known, but their occurrence seems to be more frequent than would be predicted by random chance.

Comparison of Open Reading Frames. The sequence of the gene for the b subunit of factor XIII was then translated into the six possible reading frames for amino acids and used to search the PIR database. As expected, members of the sushi family of proteins were readily identified by the search. Another prominent feature was the identity of portions of introns I and J with the human retrovirus-related reverse transcriptase sequence (Hattori et al., 1986) and other sequences of the L1 family. The Line-1 or L1 family (Adams et al., 1980) contains the *Kpn*I repeats and codes for large open reading frames that may allow them to function as nonviral retrotransposons (Skowronski et al., 1988). There was an 88% identity of a 250 amino acid portion of intron J in an inverted orientation (nucleotides 23589–22845) and the human reverse transcriptase sequence. The largest open reading frame in the L1 repeat of the b subunit was 570 bp long.

Short portions of other open reading frames coding for 20–30 amino acids in the gene for the b subunit corresponded to other proteins in the protein database, including human platelet glycoprotein IIb (nucleotides 3600–3690 and 16001–16091), and the β-adrenergic receptor of the turkey (18949–19039), and the cell fusion protein of varicella-zoster virus (18919–19009). The significance of these similar sequences is unclear.

Discussion

The gene for the b subunit of human factor XIII is a member of a family of more than 20 genes resulting from

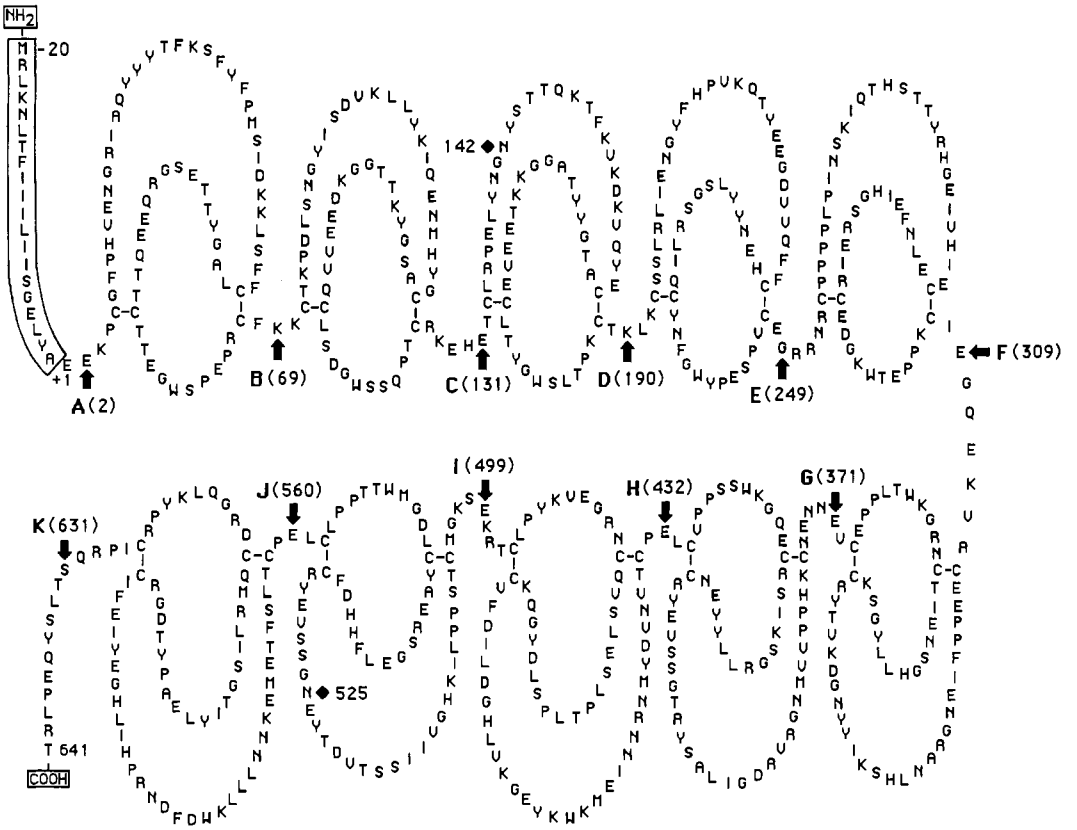

FIGURE 4: Location of the 11 introns (A–K) in the amino acid sequence for the b subunit of human factor XIII. The solid arrows indicate the position of the introns. The leader sequence is shown in the open box and potential N-linked carbohydrate attachment sites are shown by solid diamonds.

duplication of a common ancestral gene (Davie et al., 1986; Kristensen et al., 1987; Ichinose et al., 1990). The division of the b subunit of factor XIII into individual tandem repeats by the intervening sequences is illustrated in Figure 4. The organization of the genes for several other members of this family is also known (Figure 5). A few genes in this family have their sushi domains divided internally by an intron, as in haptoglobin (Bensi et al., 1985), while some have a combination of both complete and divided domains or two domains encoded by a single exon, as in factor H (Vik et al., 1988) and CR2 (Fujisaku et al., 1989). The remaining genes containing sushi domains, however, appear to be encoded by a single exon. These include the b subunit of factor XIII as well as factor B (Campbell et al., 1984) and the IL-2 receptor (Leonard et al., 1985) (Figure 5). It seems likely that the ancestral gene coding for the sushi domain did not contain an internal intron and that the insertion of an intron is a relatively recent event in the few instances where it does occur.

The internal homologies of the sushi domains in the b subunit of factor XIII at the DNA level range from 36% to 43%. Homologies with sushi domains in other proteins are much less, averaging 20–30%. The various proposals for exon shuffling (Gilbert, 1985; Patthy, 1985; Davie et al., 1986) and intron mobility (Perlman & Butow, 1989) are helpful in explaining the duplication and divergence of the proteins containing sushi domains. Typically, these exons are within the boundaries of type I splice junctions that appear to be favorable for exon shuffling. This is also the case for all the exons coding for the 10 sushi domains in the gene for the b subunit of factor XIII. However, there are no homologous areas within the introns that suggest how the sushi domains are duplicated within this gene. The sizes of the introns are variable (Table II) and do not show a common repetitive sequence that they all share.

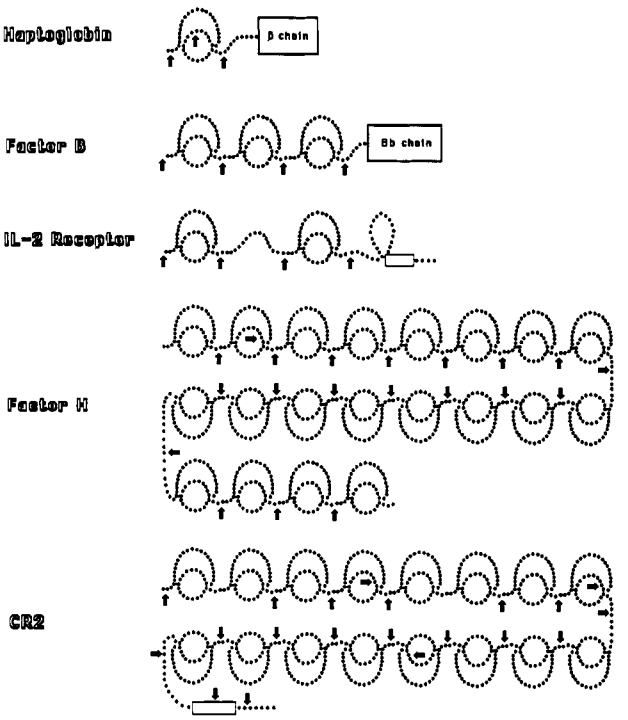

FIGURE 5: Location of the introns in the sushi structures present in haptoglobin, factor B, IL-2 receptor, factor H, and complement receptor type II (CR2). The solid arrows indicate the location of the introns. The transmembrane regions in IL-2 receptor and CR2 are shown by open boxes. The structures are modifications from Campbell et al. (1984), Bensi et al. (1985), Leonard et al. (1985), Vik et al. (1988), and Yorifuji et al. (1988).

There are five Alu repetitive sequences in and around the gene (Figure 1, Table III). Alu sequences are known to undergo homologous recombination (Huang et al., 1989), and

Alu–Alu recombination has resulted in deletions or duplications as seen in portions of a defective LDL receptor gene (Lehrman et al., 1987). The possibility exists that exons II–V were duplicated and resulted in exons VI–IX on the basis of the positions of the Alu sequences. This differs from the previous suggestion of duplication events that was based on homologies at the protein level (Ichinose et al., 1986b). These data suggest that exons II–IV were closely related, that exons V–VII were all similar, and that exons VIII and IX were duplicated to give exons X and XI. By comparing the divergence of the sushi domains to the higher homologies of the Alu repeats, it would appear that insertion of the Alu sequences occurred much more recently than the duplication of the sushi domains and that these two events were not linked.

The second *Kpn*I repeat (Table III) is surrounded by an interesting direct repeat, made up of a repeating domain of $d(CT)_n$ followed by a stretch of alternating purines and pyrimidines. The $d(CT)_n$ repeat is found in the β globin cluster of many species, satellite DNA, and other genes of *Drosophila*, crabs, and sea urchins. The role that this direct repeat plays in the insertional events of the L1 family is still unknown. The high homology of this *Kpn*I repeat with the β globin gene and the L1 insertion in factor VIII (Kazazian et al., 1988) indicates that this repeat was duplicated and inserted recently in these genes.

A few restriction fragment length polymorphisms have been reported for the factor XIIIb subunit locus (Webb et al., 1989). These involved restriction sites of *Eco*RI, *Bgl*II, and *Xba*I. By comparing these results to the present data, it appears likely that these changes are located near the second *Kpn*I repeat within intron J. However, definitive location of the polymorphisms will require more study.

The precise function of the sushi domains is still not known in any of these proteins. In proteins made up exclusively of sushi domains, it is obvious that they contain regions that participate in binding to other proteins, as in the b subunit of factor XIII, C4BP, and factor H. Some of these sushi domains maintain binding functions, while others might have lost their functions during evolution. It remains to be determined which areas of the sushi domains participate in the binding or are involved in other functions.

Acknowledgments

We thank Drs. T. Maniatis and S. Yoshitake for kindly providing genomic libraries. We also thank Drs. Steve Leytus, Joost Meijers, Dominic Chung, Mark Martzen, and Jose Lopez for helpful discussions and Lois Swenson for help in the preparation of the manuscript.

Registry No. DNA (human liver blood coagulation factor XIII b subunit gene), 130095-89-5; blood coagulation factor XIII (human liver b subunit precursor protein moiety reduced), 128986-96-9; blood coagulation factor XIII (human liver b subunit protein moiety reduced), 103067-91-0; blood coagulation factor XIII, 9013-56-3; blood coagulation factor XIII (human b subunit precursor protein moiety), 130095-90-8; blood coagulation factor XIII (human b subunit protein moiety), 130095-91-9.

References

Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M., & Nienhuis, A. W. (1980) *Nucleic Acids Res. 8*, 6113–6128.

Bensi, G., Raugei, G., Klefenz, H., & Cortese, R. (1985) *EMBO J.* 119–126.

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

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Blake, C. (1983) *Nature (London) 306*, 535–537.

Board, P. G., Webb, G. C., McKee, G., & Ichinose, A. (1988) *Cytogenet. Cell Genet. 48*, 25–27.

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

Brethnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853–4857.

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

Carrell, N. A., Erickson, H. P., & McDonagh, J. (1989) *J. Biol. Chem. 264*, 551–556.

Dale, R. M. K., McClure, B. A., & Houchins, J. P. (1985) *Plasmid 13*, 31–40.

Davie, E. W., Ichinose, A., & Leytus, S. P. (1986) *Cold Spring Harbor Symp. Quant. Biol. 51*, 509–514.

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

Deininger, P. L., Jolly, D. J. Rubin, C. M., Friedmann, T., & Schmid, C. W. (1981) *J. Mol. Biol. 151*, 17–33.

DiGiovanni, L., Haynes, S. R., Misra, R., & Jelinek, W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 480, 6533–6537.

Folk, J. E. (1983) *Adv. Enzymol. Relat. Areas Mol. Biol. 54*, 1–56.

Folk, J. E., & Finlayson, J. S. (1977) *Adv. Protein Chem. 31*, 1–33.

Fujisaku, A., Harley, J. B., Frank, M. B., Gruner, B. A., Frazier, B., & Holers, V. M. (1989) *J. Biol. Chem. 264*, 2118–2125.

Gilbert, W. (1985) *Science (Washington, D.C.) 228*, 823–824.

Graves, B. J., Johnson, P. F., & McKnight, S. L. (1986) *Cell (Cambridge, Mass.) 44*, 565–576.

Grundmann, U., Amann, E., Zettlemeissl, G., & Kupper, H. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8024–8028.

Halkier, T., & Magnusson, S. (1988) *Thromb. Res. 51*, 313–324.

Hattori, M., Hidaka, S., & Sakaki, Y. (1985) *Nucleic Acids Res. 13*, 7813–7827.

Hattori, M., Kuhara, S., Takenaka, O., & Sakaki, Y. (1986) *Nature (London) 321*, 625–628.

Huang, L.-S., Ripps, R. E., Korman, S. H., Deckelbaum, R. J., & Breslow, J. L. (1989) *J. Biol. Chem. 264*, 11394–11400.

Ichinose, A., & Davie, E. W. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5829–5833.

Ichinose, A., Hendrickson, L. E., Fujikawa, K., & Davie, E. W. (1986a) *Biochemistry 25*, 6900–6906.

Ichinose, A., McMullen, B. A., Fujikawa, K., & Davie, E. W. (1986b) *Biochemistry 25*, 4633–4638.

Ichinose, A., Bottenus, R. E., & Davie, E. W. (1990) *J. Biol. Chem. 265*, 13411–13414.

Ikematsu, S. (1981) *Acta Haematol. Jpn. 44*, 1499–1505.

Kazazian, H. K., Jr., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., & Antonarakis, S. E. (1988) *Nature (London) 332* 164–166.

Kristensen, T., D'Eustachio, P., Ogata, R. T., Chung, L. P., Reid, K. B. M., & Tack, B. F. (1987) *Fed. Proc. 46*, 2463–2469.

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

Lehrman, M. A., Goldstein, J. L., Russel, D. W., & Brown, M. S. (1987) *Cell (Cambridge, Mass.) 48*, 827–835.

Leonard, W. J., Depper, J. M., Kanehisa, M., Kronke, M., Peffer, N. J., Svetlik, P. B., Sullivan, M., & Greene, W. C. (1985) *Science (Washington, D.C.)* 230, 633–639.
Lerman, M. I., Thayer, R. E., & Singer, M. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3966–3970.
Lorand, L., Losowsky, M. S., & Miloszewski, K. J. M. (1980) *Prog. Hemostasis Thromb. 5*, 245–290.
Lozier, J., Takahashi, N., & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 3640–3644.
Lurquin, P. F. (1988) *BioTechniques 6*, 942–944.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Morton, N. E., & Burns, G. A. (1987) *Cytogenet. Cell Genet. 46*, 102–130.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.
Nussinov, R. (1981) *J. Mol. Biol. 149*, 125–131.
Patthy, L. (1985) *Cell (Cambridge, Mass.) 41*, 657–663.
Perlman, P. S., & Butow, R. A. (1989) *Science (Washington, D.C.) 246*, 1106–1107.
Rodeghiero, F., Morbin, M., & Barbui, T. (1981) *Thromb. Haemostasis 46*, 621–622.
Rodriguez-de-Cordoba, S., Rey-Campos, J., Dykes, D. D., McAlpine, P. J., Wong, P., & Rubinstein, P. (1988) *Immunogenetics 28*, 452–454.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schmid, C. W., & Jelinek, W. R. (1982) *Science (Washington, D.C.) 216*, 1065–1070.
Sharp, P. A. (1981) *Cell (Cambridge, Mass.) 23*, 643–646.
Skowronski, J., Fanning, T. G., & Singer, M. F. (1988) *Mol. Cell. Biol. 8*, 1385–1397.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Sun, L., Paulson, K. E., Schmid, C. W., Kadyk, L., Leinwand, L. (1984) *Nucleic Acids Res. 12*, 2669–2690.
Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767–4771.
Takahashi, N., Takahashi, Y., & Putnam, F. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8019–8023.
Thomashow, M. F., Nutter, R., Montoya, A. L., Gordon, M. P., & Nester, E. W. (1980) *Cell (Cambridge, Mass.) 19*, 729–739.
Vik, D. P., Keeney, J. B., Munoz-Canoves, P., Chaplin, D. D., & Tack, B. F. (1988) *J. Biol. Chem. 263*, 16720–16724.
Wang, A. J. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature (London) 282*, 680–686.
Webb, G. C., Coggan, M., Ichinose, A., & Board, P. G. (1989) *Hum. Genet. 81*, 157–160.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 389–395.
Yorifuji, H., Anderson, K., Lynch, G. W., Van De Water, L., & McDonagh, J. (1988) *Blood 72*, 1645–1650.
Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.

Counterflow of L-Glutamate in Plasma Membrane Vesicles and Reconstituted Preparations from Rat Brain†

Gilia Pines and Baruch I. Kanner*
Department of Biochemistry, Hadassah Medical School, The Hebrew University, P.O. Box 1172, Jerusalem 91010, Israel
Received May 15, 1990; Revised Manuscript Received July 25, 1990

ABSTRACT: Membrane vesicles from rat brain exhibit sodium-dependent uptake of L-[^{3}H]glutamate in the absence of any transmembrane ion gradients. The substrate specificity of the process is identical with (Na^+ + K^+)-coupled L-glutamate accumulation. Although these vesicles are prepared after osmotic shock and are washed repeatedly, they contain about 1.5 nmol/mg of protein endogenous L-glutamate, apparently located inside the vesicles. The affinity of the process ($K_m \approx 1$ μM) is similar to that of (Na^+ + K^+)-dependent accumulation by the L-glutamate transporter. Membrane vesicles have been disrupted by the detergent cholate, and the solubilized proteins have been subsequently reconstituted into liposomes. The reconstituted proteoliposomes also exhibit the above uptake—with the same characteristics—provided they contain entrapped cold L-glutamate. Counterflow is optimal when sodium is present on both sides of the membrane, but partial activity is still observed when sodium is present either on the inside or on the outside. Increasing the L-glutamate concentration above the K_m results in counterflow completely independent of cis sodium. The initial rate of counterflow is 100–200-fold lower than that of net trans potassium dependent flux. The rate of net flux in the presence of trans sodium or lithium is about 10-fold lower than when choline or Tris are used instead. However, the rate of counterflow (no internal potassium present) was not stimulated by replacing internal sodium or lithium by internal choline. Therefore, optimal functioning of the transporter requires internal potassium while internal sodium and lithium are inhibitory. In addition, the membrane vesicles also contain a low-affinity uptake system (K_m about 100 μM) for L-glutamate, which is also dependent on cis sodium and trans potassium. The above data are accommodated in a refined model of the translocation cycle of the (Na^+ + K^+)-coupled L-glutamate transporter.

The reuptake of neurotransmitters from the synaptic cleft by high-affinity transport appears to play an important role in the overall process of synaptic transmission (Iversen, 1975; Kuhar, 1973). The process is catalyzed by sodium-coupled neurotransmitter transport systems [reviewed in Kanner (1983, 1989) and Kanner and Schuldiner (1987)] located in plasma membranes of nerve endings and glial cells. These transport systems have been investigated in detail by using plasma

† This work was supported by Grant NS-16708 from the National Institute of Neurology and Communicative Disorders and Stroke, and by Grant I-02-006.1187 from the German–Israeli Foundation of Scientific Research and Development.

0006-2960/90/0429-11209$02.50/0