

- Takasaki, S., Murray, G. J., Furbish, F. S., Brady, R. O., Barranger, J. A., & Kobata, A. (1984) *J. Biol. Chem.* 259, 10112-10117.
- Timonen, T., Patarroyo, M., & Gahmberg, C. G. (1988) *J. Immunol.* 191, 1041-1046.
- Timonen, T., Gahmberg, C. G., & Patarroyo, M. (1990) *Int. J. Cancer* (in press).
- Todd, R. F., III, Arnaout, M. A., Rosin, R. E., Crowley, C. A., Peters, W. A., & Babior, B. M. (1984) *J. Clin. Invest.* 74, 1280-1290.
- Yamashita, K., Mizuochi, T., & Kobata, A. (1982) *Methods Enzymol.* 83, 105-126.
- Yamashita, K., Ueda, I., & Kobata, A. (1983) *J. Biol. Chem.* 258, 14144-14147.
- Yamashita, K., Kochibe, N., Ohkura, T., Ueda, I., & Kobata, A. (1985) *J. Biol. Chem.* 260, 4688-4693.

Intron-Exon Organization of the Human Gene Coding for the Lipoprotein-Associated Coagulation Inhibitor: The Factor Xa Dependent Inhibitor of the Extrinsic Pathway of Coagulation[†]

C. Paul E. van der Logt,* Pieter H. Reitsma, and Rogier M. Bertina

Haemostasis and Trombosis Research Unit, University Hospital, Leiden, The Netherlands

Received July 25, 1990; Revised Manuscript Received October 29, 1990

ABSTRACT: Blood coagulation can be initiated when factor VII(a) binds to its cofactor tissue factor. This factor VIIa/tissue factor complex proteolytically activates factors IX and X, which eventually leads to the formation of a fibrin clot. Plasma contains a lipoprotein-associated coagulation inhibitor (LACI) which inhibits factor Xa directly and, in a Xa-dependent manner, also inhibits the factor VIIa/tissue factor complex. Here we report the cloning of the human LACI gene and the elucidation of its intron-exon organization. The LACI gene, which spans about 70 kb, consists of nine exons separated by eight introns. As has been found for other Kunitz-type protease inhibitors, the domain structure of human LACI is reflected in the intron-exon organization of the gene. The 5' terminus of the LACI mRNA has been determined by primer extension and S1 nuclease mapping. The putative promoter was examined and found to contain two consensus sequences for AP-1 binding and one for NF-1 binding, but no TATA consensus promoter element.

Blood coagulation is a host defense system that is involved in maintaining the integrity of the vascular circulatory system after blood vessel injury. The coagulation system consists of several plasma glycoproteins, including factor VII, factor IX, and factor X, which are zymogens of serine proteases. They are converted from an inactive form to an active enzyme by limited proteolysis. Coagulation is initiated when factor VII(a) binds to the transmembrane glycoprotein tissue factor [reviewed in Furie and Furie (1988) and Bach (1988)]. This factor VIIa/tissue factor complex proteolytically activates factors IX and X, triggering a cascade of events which eventually leads to the formation of insoluble fibrin.

Early studies regarding the regulation of the tissue factor initiated coagulation showed that incubation of tissue factor (in crude extracts) with serum inhibited its procoagulant activity (Schneider, 1947; Thomas, 1947; Lanchantin & Ware, 1953). Hjort (1957) confirmed and extended these observations and concluded that serum contains a component that inactivates the factor VIIa/tissue factor complex. Recent studies (Sanders et al., 1985; Hubbard & Jennings, 1987; Broze & Miletich, 1987a) have shown that this inhibitor, that is variously called the tissue factor inhibitor (Broze & Miletich, 1987b), the extrinsic pathway inhibitor (EPI) (Rao & Ra-

poort, 1987), or the lipoprotein-associated coagulation inhibitor (LACI) (Broze et al., 1988), binds to factor Xa and inhibits the formation of factors IXa and Xa by the factor VIIa/tissue factor complex in a factor Xa dependent manner. The inhibition of the factor VIIa/tissue factor complex is thought to involve the formation of an LACI/factor Xa complex which binds noncovalently to the factor VIIa/tissue factor complex, producing a quaternary factor VIIa/tissue factor/factor Xa/LACI complex (Broze et al., 1988).

The complete cDNA of LACI has recently been cloned (Girard et al., 1989a; Wun et al., 1988). The predicted amino acid sequence reveals that LACI contains several discernible domains, including a negatively charged NH₂ terminus and a positively charged COOH terminus. The center portion of the 32-kilodalton protein consists of three tandemly arranged homologous domains which have the typical cysteine backbone of the Kunitz-type inhibitor domain, a structure very common in basic protease inhibitors (Wun et al., 1988). Mutation experiments indicated that the first Kunitz domain binds to the active site of the factor VIIa/tissue factor complex (Girard et al., 1989b, 1990) and that the second Kunitz domain binds to the active site of factor Xa. The function of the third Kunitz domain is not known (Girard et al., 1989b).

The tissue distribution of LACI expression has not been studied in detail, but LACI transcripts have been identified in liver-derived cell lines (Wun et al., 1988) and platelets (Novotny et al., 1988). Furthermore, LACI activity has been demonstrated in conditioned media from endothelial cell cultures (Warn-Cramer et al., 1989) and in the media from

[†] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05312.

* Correspondence should be addressed to this author at the Haemostasis and Thrombosis Research Unit, University Hospital, Building 1, C2-R, P.O. Box 9600, 2300 RC Leiden, The Netherlands.

Table I: Nucleotide Sequence and Position in the LACI cDNA of Oligonucleotides Used for the Synthesis of LACI cDNA Probes and the Analysis of LACI Genomic Clones

oligo	nucleotide sequence	positions
LACI 1	GCGAGGTAAGAATTTGACTA	51-70
LACI 2	TCGCTCTTTCCTGCTAGTA	67-86*
LACI 3	GCAATCTGATCTTACTAGCA	79-98
LACI 4	GTTCTGGTGAGGGC	320-334
LACI 5	AGAATCAGCATTAAAGAGGGCAGGG-GCAAGATTAA	437-471
LACI 6	TTAATGCTGATTCTGAGGAA	458-477*
LACI 7	TTACATGGGCCATCATCCGC	553-572
LACI 8	AATGCAAACAGGATTATAAA	703-722*
LACI 9	CGTTCACACTGTTTTGTCTG	817-836
LACI 10	GGAACCTTGGTTGATTGCGG	979-998
LACI 11	CTGTACTTAAATGGGCGGCA	1105-1124
LACI 12	GATATCTTTGGATGAAACC	1189-1208

*Position of the oligonucleotides is according to the LACI cDNA (Girard et al., 1989a). An asterisk indicates that the oligonucleotide is antisense.

stimulated monocytic cells (Rana et al., 1988).

In order to obtain information on how the domain structure of LACI is reflected in the structural organization of the gene, and to gain insight in the regulation of LACI expression in different cell types, we have isolated the complete human LACI gene, established its intron-exon organization, and mapped the transcriptional start site.

MATERIALS AND METHODS

LACI cDNA Probe Preparation. LACI cDNA probes were prepared by using the DNA polymerase chain reaction technique (Saiki et al., 1988). DNA isolated from a human placenta cDNA library in λ gt11 (Clontech Laboratories, Palo Alto, CA), containing approximately 10^6 independent recombinant phages, was used as template for the polymerase chain reaction. Two sets of specific oligonucleotides (set I, LACI 6-LACI 12; set II, LACI 2-LACI 5) (Table I) were used to generate respectively LACI cDNA fragments LcDI (751 bp) and LcDII (405 bp). The amplifications were performed in 33 cycles, each cycle consisting of a denaturing step at 94 °C for 1 min, an annealing step at 58 °C for 2 min, and a primer extension reaction at 71 °C for 3 min. The reaction mixture contained 500 ng of DNA isolated by the plate lysate method (Maniatis et al., 1982), 400 ng of each primer, 150 mM of each deoxynucleotide, 100 ng of BSA/mL, 67 mM Tris-HCl, pH 8.8, 6.7 mM MgCl₂, 10 mM β -mercaptoethanol, 6.7 μ M EDTA, 16.6 mM (NH₄)₂SO₄, 10% dimethyl sulfoxide, and 2.5 units of *Taq* DNA polymerase (Cetus Corp., Boston, MA) in a final volume of 100 μ L. The unreacted primers were removed by Qiagen-tip-5 (Diagen, Düsseldorf, FRG) elution. The complete amplification reaction was treated with 5 units of mung bean nuclease (Promega, Madison, WI) for 30 min at 30 °C. The amplified DNA was purified by ultra low gelling temperature agarose (Sigma, St. Louis, MO) gel electrophoresis. The excised blunt-ended DNA fragments were cloned into the *Sma*I site of plasmid pUC13. The cloned LACI cDNA fragments were analyzed by DNA sequencing.

Construction and Screening of Genomic Libraries in Phage λ EMBL. Two human genomic libraries were constructed from DNA isolated from peripheral blood leukocytes. High molecular weight DNA was partially digested with the enzyme *Sau*3A and ligated in the arms of the phage λ EMBL3 (Promega) essentially as described (Maniatis et al., 1982). After in vitro packaging, the recombinant phages were plated on either *Escherichia coli* NM539 or *E. coli* KW251. The libraries contained respectively 7.5×10^5 and 2×10^6 independent recombinant phages. Screening of the unamplified

libraries was performed by plaque in situ hybridization with LACI cDNA fragments LcDI and LcDIIa. The DNA fragments were radiolabeled by random priming using [α -³²P]-dCTP (New England Nuclear, Boston, MA). The filters were washed 2 times with $2 \times$ SSC/1% SDS for 30 min at 65 °C. Positive clones were plaque-purified, and DNA was isolated by the plate lysate method (Maniatis et al., 1982).

Characterization of LACI Genomic Clones. Positive clones were characterized by Southern blot analysis. *Eco*RI, *Hind*III, *Bgl*II, and *Pst*I digests of the recombinant phages were fractionated by agarose gel electrophoresis. The DNA fragments were transferred onto a Gene Screen Plus membrane (New England Nuclear). Restriction fragments containing exonic sequences were visualized with exon-specific oligonucleotides (Table I) which had been labeled at the 5' end with T₄ polynucleotide kinase (Boehringer Mannheim, Mannheim, FRG) using [γ -³²P]dATP (Amersham International, Amersham, U.K.). Hybridization was performed at 42 °C with $6 \times$ SSC, 0.5% SDS, $5 \times$ Denhardt's solution, and 100 μ g/mL denatured salmon sperm DNA. The filters were washed 2 times for 15 min at 42 °C with $6 \times$ SSC/0.5% SDS. The nucleotide sequence of the oligonucleotides was derived from the nucleotide sequence of the reported LACI cDNA (Girard et al., 1989a). Oligonucleotides were synthesized on a Cyclone DNA synthesizer (Millipore, Bedford, MA).

DNA Sequencing. *Eco*RI, *Bgl*II, or *Pst*I fragments of the LACI genomic clones that hybridized to the exon-specific oligonucleotides were subcloned in plasmid pUC13 vectors. Except for a major portion of the 3'-untranslated region, the DNA sequence for all exons and intron-exon junctions was determined for both strands by the dideoxy chain termination reaction (Sanger et al., 1977) using [α -³⁵S]dATP (Amersham International). The sequencing reactions were primed with exon-specific oligonucleotides (Table I). Additional oligonucleotides, the sequences of which were based on the obtained nucleotide sequences, were synthesized and used to prime the reaction in the opposite direction.

Southern Blot Analysis of Genomic DNA. High molecular weight DNA, isolated from human peripheral blood leukocytes, was digested with various restriction enzymes. The DNA fragments were separated on 0.8% agarose gels in TAE buffer with 0.5 μ g/mL ethidium bromide. Subsequent to electrophoresis, the DNA samples were blotted onto Gene Screen Plus membranes. Prehybridization and hybridization were performed at 65 °C in 1 M NaCl, 1% SDS, 50 mM Tris-HCl, pH 7.5, 10% dextran sulfate, and 100 μ g/mL salmon sperm DNA. The filters were probed with LACI cDNA fragments LcDI, LcDIIa, and LcDIIb. Labeling of the fragments was performed by random priming using [α -³²P]dCTP. The filters were washed 2 times for 30 min in $0.2 \times$ SSC/0.1% SDS at 65 °C.

RNA Isolation and Northern Blot Analysis. Total human liver RNA was isolated by the LiCl-urea method (Auffray & Rougeon, 1980). The RNA was separated by electrophoresis on 1% agarose gels in 2.2 M formaldehyde, 20 mM Mops, 5 mM sodium acetate, pH 7.0, and 1 mM EDTA. The RNA was transferred to Gene Screen Plus membranes. Prehybridization and hybridization were performed at 42 °C in 50% formamide, 4% SDS, $5 \times$ Denhardt's solution, $4 \times$ SSPE, and 100 μ g/mL salmon sperm DNA. The membrane was probed with LACI cDNA fragment LcDI, radiolabeled by random priming. Washing was performed with a final stringency of 0.2% SSC/0.1% SDS at 65 °C.

Primer Extension and S1 Nuclease Mapping. The length of the 5'-untranslated region of the LACI gene was determined

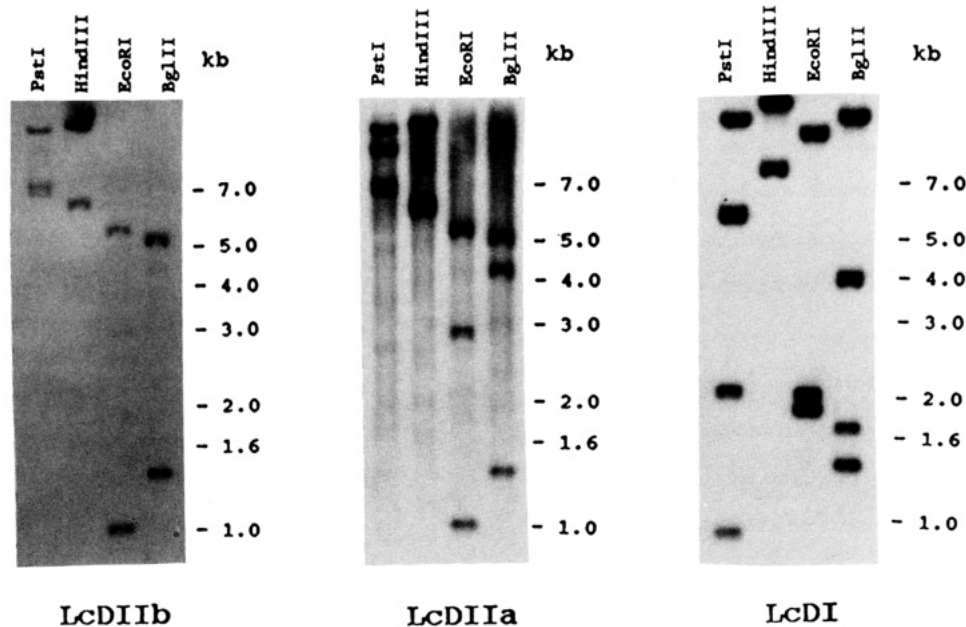


FIGURE 1: Southern blot analysis of the human LACI gene. DNA from human blood leukocytes was digested to completion with the restriction endonucleases *Pst*I, *Hind*III, *Eco*RI, and *Bgl*II and hybridized with LACI cDNA fragments LcDI, LcDIIa, and LcDIIb. The sizes of the markers are shown in kilobase pairs.

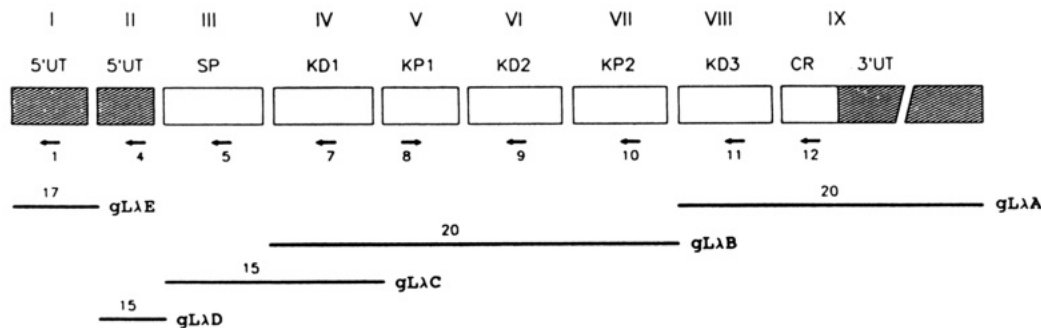


FIGURE 2: Intron-exon organization of the human LACI gene. The positions of the genomic clones with respect to the exons they contain are indicated at the bottom. The length of the clones is given in kilobases. The small arrows indicate the positions of the exon-specific oligonucleotides (numbered 1-9; see Table I). Abbreviations: 5'UT, 5'-untranslated region; SP, signal peptide; KD, Kunitz-type inhibitory domain; KP, Kunitz intervening peptide; CR, carboxy-terminal region; 3'UT, 3'-untranslated region.

by primer extension. Oligonucleotides LACI 1 and LACI 3 (Table I) were labeled at their 5' end with T_4 polynucleotide kinase using $[\gamma\text{-}^{32}\text{P}]\text{dATP}$ to a specific activity of 5×10^7 cpm/ μg ; 10^5 cpm were coprecipitated with 30 μg of total liver RNA. The precipitate was resuspended in 20 μL of buffer containing 10 mM dithiothreitol, 1 mM of each of the deoxynucleotides, 75 mM KCl, 50 mM Tris-HCl, pH 8.3, 3 mM MgCl_2 , 30 units of RNasin (Promega), and 200 units of reverse transcriptase (Bethesda Research Laboratories, Life Technologies, Inc.). Incubation at 47 $^\circ\text{C}$ for 1 h was followed by ethanol precipitation. The recovered extended products were separated on a denaturing sequencing gel. The probe for S₁ nuclease analysis was an LACI-specific genomic *Bst*NI fragment containing sequences from position 292 to position 675 (Figure 3). After dephosphorylation using calf intestinal phosphatase (Promega), the fragment was end-labeled with T_4 polynucleotide kinase using $[\gamma\text{-}^{32}\text{P}]\text{dATP}$. Total liver RNA (42 μg) was coprecipitated with 10^5 cpm of probe DNA. The precipitate was resuspended in 80% formamide, 0.4 M NaCl, 40 mM Pipes, pH 6.4, and 1 mM EDTA and after a denaturing step at 85 $^\circ\text{C}$ for 15 min hybridized at 42 $^\circ\text{C}$ for 16 h. Nuclease S1 digestion was performed during 30 min at 37 $^\circ\text{C}$ using 1000 units/mL S1 nuclease (Boehringer Mannheim), essentially as described (Maniatis et al., 1982). After precipitation, the recovered protected fragments were separated

on a denaturing sequencing gel.

RESULTS AND DISCUSSION

LACI cDNA Probes. A 751 bp LACI cDNA fragment (LcDI) was synthesized by using primer set I. Primer set II produced in addition to the expected 405 bp LACI cDNA fragment (LcDIIa) a 284 bp fragment (LcDIIb). Sequence analysis demonstrated that fragment LcDIIb had an internal deletion of 121 base pairs with respect to fragment LcDIIa.

Figure 1 shows that genomic DNA digested with each of the enzymes *Bgl*II, *Eco*RI, *Hind*III, and *Pst*I and probed with fragment LcDIIa gives an additional hybridizing fragment with respect to the hybridization pattern observed using radiolabeled fragment LcDIIb. Since these probes contain none of the recognition sites for the enzymes, this indicates that the LACI cDNA fragment LcDIIb is produced by alternative splicing whereby a 121 bp fragment is excised from the LACI pre-mRNA. Subsequent elucidation of the structure of the LACI gene showed that this 121 bp fragment represents exon II (Figure 3). Finally, probes LcDIIa and LcDIIb detect a 6.4/6.9-kb *Pst*I RFLP (Van der Logt et al., 1990).

Isolation of Genomic Clones Spanning the Human LACI Gene. Three positive recombinants (gLAA, gLAB, and gLAC) were obtained after screening 7.5×10^5 independent recom-

EXON I

30 60
 aaataaactgggctgagtagcaagtagtaagtaggagctcaaaagctgcaacc
 90 120
 tgcataagataatgacattacattgtctcaactaagagagacctccaaagctgagccctcaga
 ap-1 150 180
 ctttaaaaaaataaaatcattgacagtggtgaaacaatgaataaactggaaga
 210 240
 aacaaactgcaaaaaagtttatacaactgtaataaataatgtaattttttaaaccct
 270 300
 aactctagggggaaaaagcattctttcaactgattcaaaaaaatcctggaagta
 330 360
 aaggaaatagctattcaaatgatcgtatctgaaactctggtgtaagctgtttccat
 nbr1 390 420
 ctgttccctccactaaaaaaagaagaagaagaagaagggttagactaaatgga
 nbr2 450 480
 gtcagagtgagcagctaaacaggaagttggctattcccaactgcccagtgatctctg
 ap-1 * * * * *
 aagcCGACTCGAGGCTCCCTCTTGTCTTAACAGACAGCAGCGACTTTAGCGTGATAA
 570 600
 TAGTCAAAATTCCTACCTCGCTCTTCACTGCTAGTAAGATCAGATTGCGTTTCTTTCAGT
 630 660
 TACTCTCAATCGCCAGTTCTGTGACTGCTTCTAAAAGAAGAAGTAGAGAAGATAAATC
 690 720
 CTGTCTCAATACCTGGAAGGAAAAAATAAACCCTCAACTCCGTTTGAAAAAACAT
 750 780
 TCCAAGACTTTCATCAGAGATTTTACTTgtgagtaacctggcaataactgtctagccca
 810
 taagaataaactttaattttctttctctggggg.....

EXON II

20 50
tgatgccttaagaagccttagagaaaagatatctttctctctccatg
 80 110
 gtgcacacagcagcaggaagccctctgtaggacataagggagaaagcagccaccattgcaa
 140 170
 GCCAAGAGAGAGCCCTCACCGAAGCATGGACCAGCAGCACCTTGATCTGGATTTTC
 192
 TAGCCTCCAGAACTgtgagata.....

EXON III

20 50
atatgaggtacacatatgatgtttactagctctgtaaatgagtgcatatta
 80 110
 acactttatttattagatagtgatgggtctgtatttcagagatgatttACACAATGAAG
 MetIleTyrThrMetLys
 -28 -23
 140 170
 AAAGTACATGCACCTTTGGGCTTCTGTAAGCTGCTGCTAACTTGCCTCCCTCCCTCTT
 LysValHisAlaLeuTrpAlaSerValCysLeuLeuLeuLeuLeuAlaProAlaProLeu
 -13 -3
 200 230
 AATGCTGATTCTGAGGAAGATGAAGAACACACAATTTACACAGctaaatattagaagca
 AsnAlaAspSerGluGluAspGluGluHisThrIleIleThr
 -1 1 12
 260 290
 attatcttttaagctagttaaagctctccattttaaacttagggatctaccatagttttc
 302
 cattgtggaggg.....

EXON IV

20 50
caggacattataaaaaagaaataaagaacattattacagtggtgatggag
 80 110
 attacatgttatctcttttcttttctttatagATACGGAGTTGCCACCACTGAAACTT
 AspThrGluLeuProProLeuLysLeu
 13 21
 140 170
 ATGCATTCAATTTGTGCAATCAAGCGGATGATGGCCCATGTAAAGCAATCAATGAAAGA
 MetHisSerPheCysAlaPheLysAlaAspAspGlyProCysLysAlaIleMetLysArg
 31 41
 200 230
 TTTTCTCAATAATTTCACTCGACAGTGGCAAGAAATTTATATATGGGGATGTGAAGGA
 PhePhePheAsnIlePheThrArgGlnCysGluGluPheIleTyrGlyGlyCysGluGly
 51 61
 260 290
 AATCAGAAATCGATTGAAAGCTGGAAGAGTGCAAAAAATGTGTACAAGAGtgatggttt
 AsnGlnAsnArgPheGluSerLeuGluGluCysLysLysMetCysThrArg
 71 78
 320 357
 ctgggaaccttattactcaaaagaccttttaggctattgagcttaattatggatttt...

EXON V

20 50
aaattcaaatcctagttactttataaagtgtgacaatgatgatgata
 80 110
 tattcttttgattacagataATGCAACAGGATTTAAAAGACAACATGCAACAGgtg
 AspAsnAlaAsnArgIleIleLysThrThrLeuGlnGln
 79 91
 140 170
 acatttattgtctgtaaatggaagcatttattgtaagcatttagatattcaattttt
 200 214
 gttgtttgctgtttgtttttctggctgattcagtag.....

EXON VI

20 50
cttcaataacttagccaggtatttataaattttataaatcagtaatt
 80 110
 atttcacaaaagaatattgctttctgacatttttataattctgAAAAAGCCAGATTTC
 GluLysProAspPhe
 92 96
 140 170
 TGCTTTTTGGAAGAAGATCTCGAATAAGCTCGAGCTTATATACCAGGTATTTTTATAAC
 CysPheLeuGluGluAspProGlyIleCysArgGlyTyrIleThrArgTyrPheTyrAsn
 106 116
 200 230
 AATCAGACAAAACAGTGTGAACGTTTCAAGTATGGTGGATGCCCTGGGCAATATGAACAAT
 AsnGlnThrLysGlnCysGluArgPheLysTyrGlyGlyCysLeuGlyAsnMetAsnAsn
 126 136
 260 290
 TTTGAGACACTGGAAGAAATGCAAGAACATTGTGGAAGATGGTCTgtaagtttattcttat
 PheGluThrLeuGluGluCysLysAsnIleCysGluAspGly
 146 150
 320 350
 ttcatatttagttctttaggactctcttttaaaactaaggcatctcagatggcattcta
 362
 tactcaggtttt.....

EXON VII

20 50
cgcttcatgagtcocaaatagagattgtgacaattgtttatctatttta
 80 110
 ttggtgtattttttccagcgaatggtttccagctggataattatggAACCCAGCTCAAT
 ProAsnGlyPheGlnValAspAsnTyrGlyThrGlnLeuAsn
 151 154 164
 140 170
 GCTGTGAAATAACTCCCTGACTCCGCAATCAACCAAGGTTCCAGCCTTTTTGtgaagat
 AlaValAsnAsnSerLeuThrProGlnSerThrLysValProSerLeuPhe
 174 181
 200 230
 cttgtggatttttcttccaggaaactattatcatgctaaactgaagtggatggtgt
 239
 attagaagt.....

EXON VIII

20 50
cctttattcagattactgttttacatcacatgcaacaacattaatctat
 80 110
 aaacaatatacaatacgaaaacctgaaactatcacacatggcttaccatgttttcc
 140 170
 tgattgttttttagaatttccagctccctcactggtctcactccagcagacagaggatg
 GluPheHisGlyProSerTrpCysLeuThrProAlaAspArgGlyLeu
 182 187 197
 200 230
 TGTCGTCCAATGAGAACAGATTTCTACTACAAATTCAGTCAATGGGAAATGCCGCCATTT
 CysArgAlaAsnGluAsnArgPheTyrTyrAsnSerValIleGlyLysCysArgProPhe
 207 217
 260 290
 AAGTACAGTGGATGTGGGGAAATGAAACAATTTACTTCCAAACAAGAAATGTCTGAGG
 LysTyrSerGlyCysGlyGlyAsnGluAsnAsnPheThrSerLysGlnGluCysLeuArg
 227 237
 320 338
 GCATGTAAAAAAGtataagaatactctccatttaattgactagtgtt.....
 AlaCysLysLys
 241

EXON IX

20 50
aataaattgttctacacgaatcacatggacaastaaagcttttaaatca
 80 110
 ataacatcttttggatttaaatataaaagtactcttggagaagatttttagataaaacatt
 140 170
 tcattgtgtatttttaccacacattatttctctccacttatagGTTTCATCCAA
 GlyPheIleGln
 242 245
 200 230
 AGAATATCAAAAGGAGCCCTAATTTAAACCAAAAAGAAAAGAAAGAACAGAGGTGAAA
 ArgIleSerLysGlyGlyLeuIleLysThrLysArgLysArgLysLysGlnArgValLys
 255 265
 260 290
 ATAGCATATGAAGAAATTTTGTGTAATAATGTAATAATTTGTTATAGCAATGTAACATTA
 IleAlaTyrGluGluIlePheValLysAsnMetEnd
 276
 320 350
 ATTCTATAAATATTTTATATGAAATGTTTCACTATGATTTTCTATTTTCTCTAAAAT
 380 410
 GCTTTAAATTAATGTTCAATAAATTTTCTATGCTTATGTACTGTGTTCAACACATT
 440 470
 TGATCAGAGTGTCTTTCTAACTCTGTTAAATGCTTATCTAGGCTGTAAATTTATTA
 500 530
 ACTGGCTACTGGGAAATTACTTATTTTCTGATCTATCTGATTTTCACTTAACTCAAAA
 560 590
 TTATCACTACTACGGCTACATCAAAATCAGTCCCTTTGATTTCCATTGGTGCCATCTGTTT
 620 650
 GAGAAATGATCAATGTAATGATTATCTCCTTTATAGCCTGTAACAGATTAAAGGAATAC
 680 710
 AGCTCTTAAAAAATCAAGAACTTCTGAGTTTACATATAAAAATGGTGACAAAACCTGC

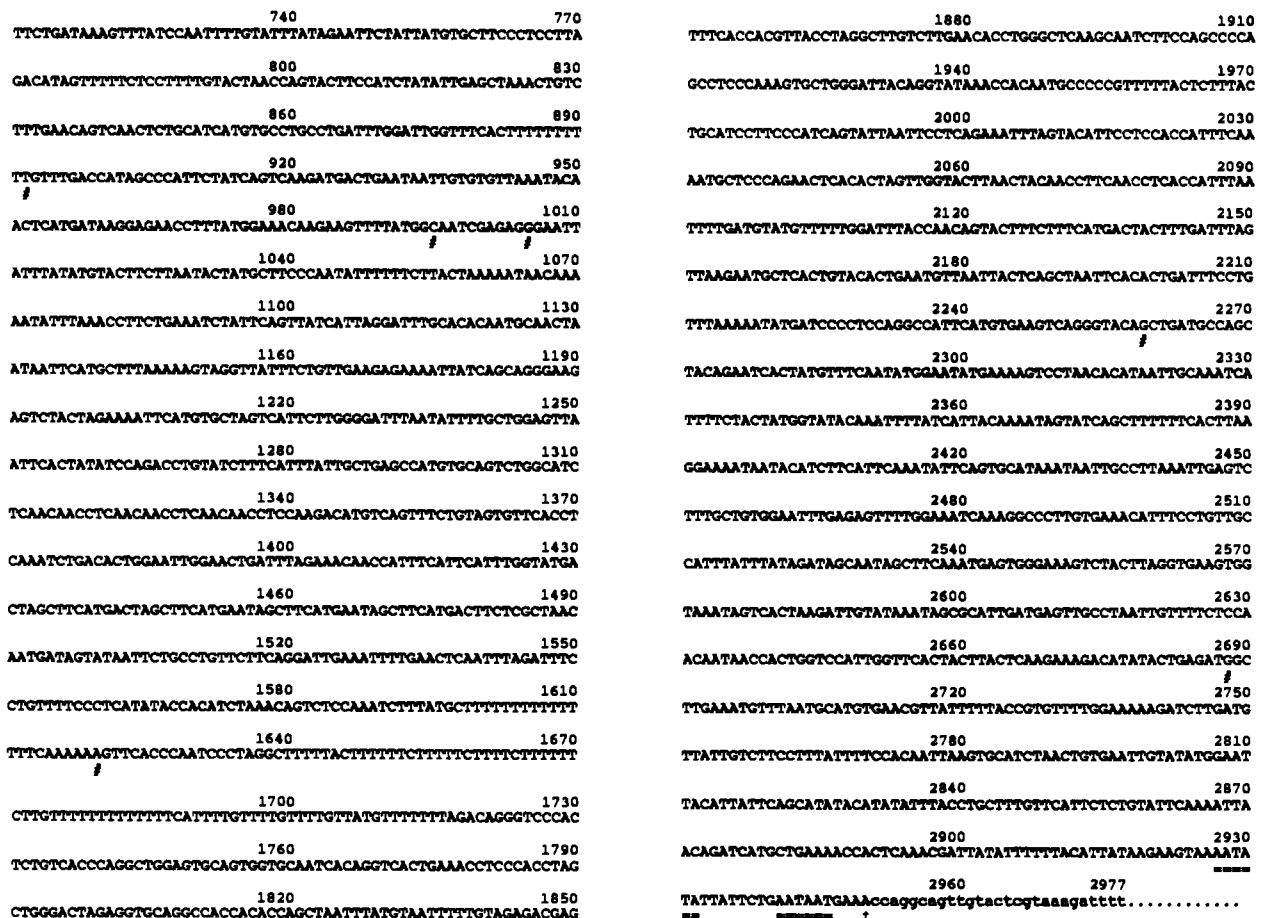


FIGURE 3: Nucleotide sequence of the exons and exon-flanking regions of the human LACI gene. The major transcription initiation sites are indicated by asterisks. The 9 bp direct repeat (nbr1 and nbr2) and the sequences showing homology with consensus sequences for transcription factor binding sites AP-1 and NF-1 are underlined. The exons are numbered, and their nucleotide sequence is presented by capitals. The coding region has been translated by using the three-letter amino acid code. The polyadenylation signals are double-underlined, and the polyadenylation sites are indicated by arrows. Nucleotides in the 3' region that deviate from the published LACI cDNA sequence (Girard et al., 1989a) are indicated by (#).

binant phages with probe LcDI. Screening 2×10^6 independent recombinants with probe LcDIIa resulted in the isolation of two hybridizing clones (gLAD and gLAE). These five positive clones were analyzed by screening with exon-specific oligonucleotides (Table I) that were selected on the basis of a best guess of the intron-exon organization.

The contents of the clones with regard to the exons of the LACI gene are shown in Figure 2. The genomic inserts of the positive clones varied in length from 15 to 20 kb, and together they span approximately 75 kb of the human genome. Because of the size of the gene, no detailed restriction map was made.

Organization of the Human LACI Gene. As shown in Figure 2, the human LACI gene is organized into nine exons separated by eight introns. It is apparently unique, since no cross-hybridizing sequences are detectable in Southern blot analysis of genomic DNA and no related sequences were obtained during nonstringent screening of genomic libraries.

The sequence of the LACI exons and the flanking regions is shown in Figure 3. The intron-exon splice junction sequences agree closely with the consensus sequences as formulated by Shapiro and Senapathy (1987). The eight introns each begin with a GT dinucleotide and end with an AG dinucleotide, sequences thought to be necessary for correct RNA splicing (Shapiro & Senapathy, 1987; Breathnach & Chambon, 1981). All the splice junctions in the human LACI gene are of type I; i.e., the splice occurs after the first nucleotide of a codon (Sharp, 1981).

Overall, our genomic sequence agrees with the cDNA sequence published by Girard et al. (1989a), except for several differences in the 3'-untranslated region. These include an A instead of a G at position 408, a C instead of an A at 995, a G instead of an A at 1005, a G instead of a T at 2688, an extra T at 891, an extra A at 1620, and an extra G at 2259 (Figure 3).

The domains of LACI that can be functionally/structurally identified are encoded by separate exons. The 5'-untranslated region is encoded by two exons. Exon III encodes the NH₂-terminal signal peptide which is removed during processing of the protein. The Kunitz-type inhibitory domains (exons IV, VI, and VIII) and the Kunitz intervening peptide domains (exons V and VII) are all encoded by separate exons. Exon IX encodes the carboxy-terminal domain and all of the extensive 3'-untranslated region.

We have compared the structure of the LACI gene with the genes of several other Kunitz-type protease inhibitors. All share significant structural similarities. In the gene of human LACI, human α -trypsin inhibitor (Vetr et al., 1989), human amyloid precursor (Ponte et al., 1988; Tanzi et al., 1988), bovine pancreatic basic protease inhibitor (Creighton & Charles, 1987), and bovine spleen inhibitor (Creighton & Charles, 1987), the regions coding for the highly conserved cysteine backbone of the Kunitz-type inhibitor domain are organized in one exon. According to the classification scheme of Patthy (1987), all exons of the LACI gene, as well as the exons coding for the inhibitor subunits of all other cloned

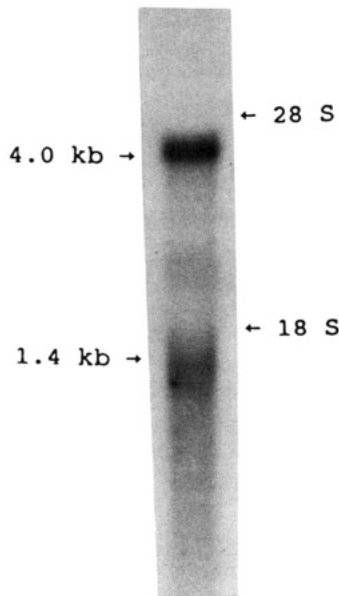


FIGURE 4: Identification of LACI messages by Northern blotting. 15 μ g of total human liver mRNA probed with LACI cDNA fragment LcDI. The relative positions of 28S and 18S are shown on the right.

Kunitz-type protease inhibitors, are symmetrical and of class I-I. This means that they have introns of the same phase class at both their ends. It has been hypothesized that the use of such nonrandom symmetrical intron phases in a gene is a sign of gene assembly by exon shuffling, a process through which new genes are assembled by using modules of the same phase (Patthy, 1987). Therefore, it is conceivable that LACI is a mosaic protein assembled from several modules including three

Kunitz-type inhibitory domains.

Analysis of the 5' Region of the Human LACI Gene. Northern blot analysis of human liver mRNA shows that LACI cDNA fragment LcDI hybridizes to two mRNA species, a discrete band at 4.0 kb and a broad band centered at 1.4 kb (Figure 4). As demonstrated by Girard et al. (1989a), the occurrence of the 1.4-kb mRNA species is the result of the use of an alternative polyadenylation site.

To determine the length of the LACI mRNA more precisely and to map the transcriptional start site, we analyzed the 5' region of liver LACI mRNA by primer extension and S1 nuclease mapping.

S1 nuclease protection analysis with an end-labeled genomic *Bst*NI fragment of 383 nucleotides yields 3 major protected fragments of 163, 185, and 191 nucleotides, and 4 minor products of 172, 207, 212, and 216 nucleotides (Figure 5A). These S1 nuclease protected products suggest that in the human LACI gene the major transcription initiation sites are located at positions +1, +7, and +29 (Figure 5D).

Primer extension analysis was used to verify the transcriptional start sites mapped by S1 nuclease analysis. As primers for the primer extension reaction, we used oligonucleotides LACI 1 and LACI 3 (Table I) which are complementary to respectively nucleotides 541-560 and 569-588 of the first exon of the LACI gene (Figure 3). Several major and minor extension products were detected (Figure 5B,C).

The lengths of three major extension products correspond to the transcriptional start sites +1, +7, and +29 as mapped by S1 nuclease analysis. The three minor extension products coincide with the minor transcription initiation sites -16, -21, and -25 observed by S1 nuclease mapping. Three additional extension products were found corresponding to positions +30,

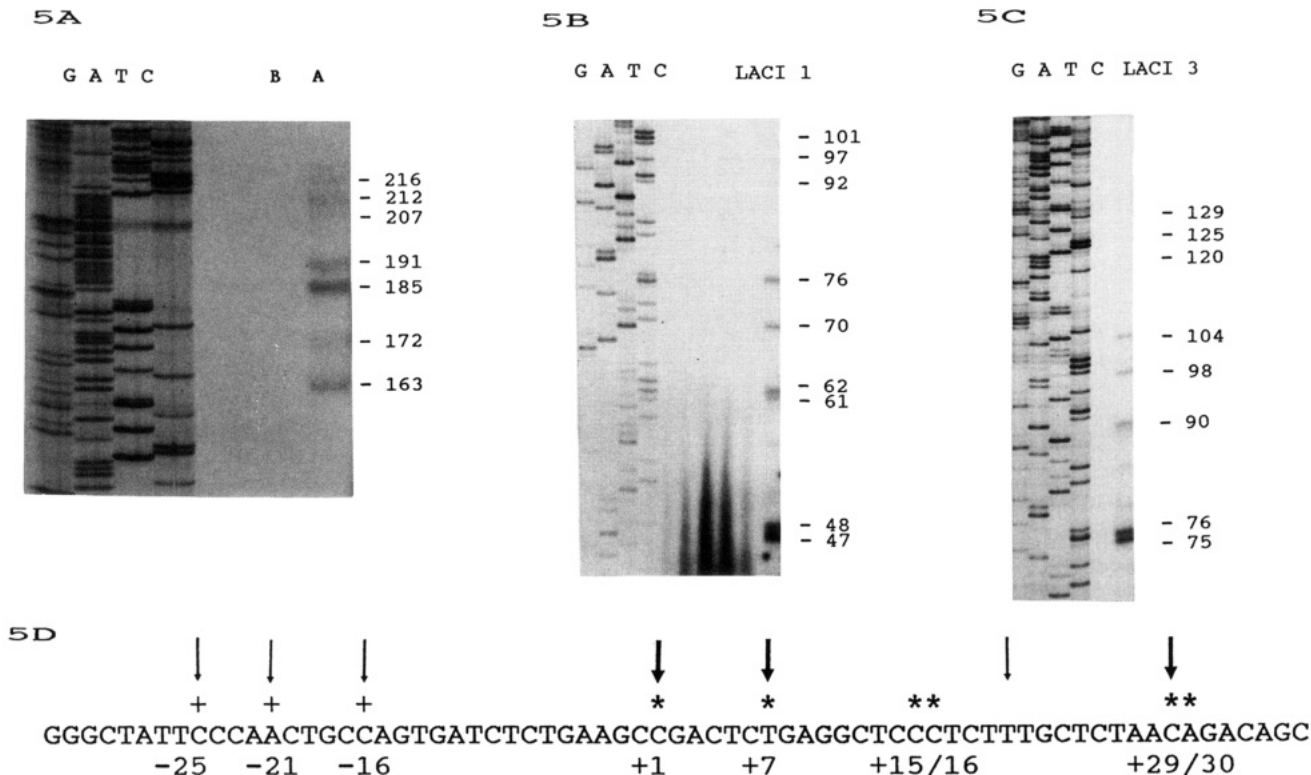


FIGURE 5: Identification of the transcription initiation sites of the human LACI gene. (A) S1 nuclease analysis of human liver LACI mRNA. Lane A, S1 nuclease protection products using 42 μ g of total human liver mRNA. Lane B, control reaction using 42 μ g of peripheral blood lymphocytes mRNA. The length of the protected fragments is indicated in base pairs. Primer extension reactions primed with oligonucleotide LACI 1 (B) and oligonucleotide LACI 2 (C), using 30 μ g of total human liver mRNA as template. Samples were run along with a sequencing reaction as standard. The length of the extension products is indicated in base pairs. (D) Nucleotide sequence of the transcriptional start region. The major transcription initiation sites mapped by primer extension are indicated by asterisks; minor starts are marked by (+). Major start positions of the LACI mRNA determined by S1 nuclease protection are indicated by bold arrows; minor starts are marked by thin arrows.

+15, and +16. The band at position +30 might be caused by methylation of the first residues of the mRNA which interferes with the reverse transcriptase reaction (Calzone et al., 1987). It remains unclear what causes the bands at positions +15 and +16, which do not correspond to the results of the S1 nuclease reaction, but they may represent artifacts of the extension reaction.

We therefore conclude that in the human LACI gene three major transcription initiation sites are present at positions +1, +7, and +29.

Potential Cis-Acting Regulatory Elements. The region upstream of the mapped transcriptional start sites has a high A/T content but does not contain a TATA-box consensus promoter element (Breathnach & Chambon, 1981). Since a single point of transcription initiation is thought to be determined by the TATA box (Breathnach & Chambon 1981), the absence of a TATA box consensus sequence in the putative promoter of the LACI gene may account for the finding of multiple transcriptional start sites.

Further inspection of the 5'-upstream region of the LACI gene reveals the presence of a nine base pair direct repeat. Comparison of 498 bp of DNA sequence upstream of the transcriptional start sites with consensus sequences for transcription factor binding sites provided matches for one NF-1 (Gronostajski, 1987) sequence and two AP-1 (Lee et al., 1987) sequences (see Figure 3). AP-1 binding sites have been demonstrated to act as phorbol ester responsive elements in the 5'-flanking region of several genes (Lee et al., 1987; Angel et al., 1987). The predicted AP-1 binding sites in the putative LACI promoter region may therefore mediate the induction of LACI activity that is observed in the conditioned media of phorbol ester stimulated monocytic cells (Rana et al., 1988).

Registry No. DNA (human lipoprotein-associated coagulation inhibitor gene), 116637-65-1; coagulation inhibitor (human lipoprotein-associated precursor reduced), 116638-33-6; coagulation inhibitor (human lipoprotein-associated reduced), 116638-34-7; RNA (human lipoprotein-associated coagulation inhibitor messenger), 131321-98-7.

REFERENCES

- Angel, P., Baumann, I., Stein, B., Delius, H., Rahmsdorf, H. J., & Herrlich, P. (1987) *Mol. Cell. Biol.* 7, 2256-2266.
- Auffray, C., & Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
- Bach, R. (1988) *CRC Crit. Rev. Biochem.* 23, 339-368.
- Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem.* 50, 349-383.
- Broze, G. J., Jr., & Miletich, J. P. (1987a) *Blood* 69, 150-155.
- Broze, G. J., Jr., & Miletich, J. P. (1987b) *Proc. Natl. Acad. Sci. U.S.A.* 84, 1886-1890.
- Broze, G. J., Jr., Warren, L. A., Novotny, W. F., Higuchi, D. A., Girard, T. J., & Miletich, J. P. (1988) *Blood* 71, 335-343.
- Calzone, F. J., Britten, R. J., & Davidson, E. H. (1987) *Methods Enzymol.* 152, 611-632.
- Creighton, T. E., & Charles, I. G. (1987) *J. Mol. Biol.* 194, 11-22.
- Furie, B., & Furie, B. C. (1988) *Cell* 53, 505-518.
- Girard, T. J., Warren, L. A., Novotny, W. F., Bejeck, B. E., Miletich, J. P., & Broze, G. J., Jr. (1989a) *Thromb. Res.* 55, 37-50.
- Girard, T. J., Warren, L. A., Novotny, W. F., Likert, K. M., Brown, S. G., Miletich, J. P., & Broze, G. J., Jr. (1989b) *Nature* 338, 518-520.
- Girard, T. J., MacPhail, L. A., Likert, K. M., Novotny, W. F., Miletich, J. P., & Broze, G. J., Jr. (1990) *Science* 248, 1421-1424.
- Gronostajski, R. M. (1987) *Nucleic Acids. Res.* 14, 9117-9131.
- Hjort, P. F. (1957) *Scand. J. Clin. Lab. Invest.* 9, (Suppl. 27), 76-97.
- Hubbard, A. R., & Jennings, C. A. (1987) *Thromb. Res.* 42, 489-498.
- Lanchantin, G. F., & Ware, A. G. (1953) *J. Clin. Invest.* 32, 381-389.
- Lee, W., Mitchell, P., & Tjian, R. (1987) *Cell* 49, 741-752.
- Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Novotny, W. F., Girard, T. J., Miletich, J. P., & Broze, G. J., Jr. (1988) *Blood* 72, 2020-2025.
- Patthy, L. (1987) *FEBS Lett.* 214, 1-7.
- Ponte, P., Gonzalez-DeWhitt, P., Schilling, J., Miller, J., Hsu, D., Greenberg, B., Davis, K., Wallace, W., Lieberburg, I., Fuller, F., & Cordell, B. (1988) *Nature* 331, 525-527.
- Rana, S. V., Reimers, H. J., Pathikonda, M. S., & Bajaj, S. P. (1988a) *Blood* 71, 259-262.
- Rao, L. V. M., & Rapoport, S. I. (1987) *Blood* 69, 645-651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science* 239, 487-491.
- Sanders, N. L., Bajaj, S. P., Zivelin, A., & Rapoport, S. I. (1985) *Blood* 66, 204-212.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
- Schneider, C. L. (1947) *Am. J. Physiol.* 149, 123-129.
- Shapiro, M. B., & Senapathy, P. (1987) *Nucleic Acids Res.* 15, 7155-7174.
- Sharp, P. A. (1981) *Cell* 23, 643-646.
- Tanzi, R. E., McClatchey, A. I., Lamperti, E. D., Villa-Komaroff, L., Gusella, J. F., & Neve, R. L. (1988) *Nature* 331, 528-530.
- Thomas, L. (1947) *Bull. Johns Hopkins Hosp.* 81, 26-42.
- Van der Logt, C. P. E., Reitsma, P. H., & Bertina, R. M. (1990) *Nucleic Acids Res.* 18, 5920.
- Vetr, H., Kögler, M., & Gebhard, W. (1989) *FEBS Lett.* 245, 137-140.
- Warn-Cramer, B. J., Almus, F. E., & Rapoport, S. I. (1989) *Thromb. Hemostasis* 61, 101-105.
- Wun, T.-C., Kretzmer, K. K., Girard, T. J., Miletich, J. P., & Broze, G. J., Jr. (1988) *J. Biol. Chem.* 263, 6001-6004.