Geraci, G., Parkhurst, L. J., & Gibson, Q. H. (1969) *J. Biol. Chem. 244*, 4664.
Heindall, H. C., Lui, A., Paddock, G. V., Studnicka, G. M., & Salser, W. A. (1978) *Cell (Cambridge, Mass.) 15*, 43–54.
Hull, W. E., & Sykes, B. D. (1975) *J. Mol. Biol. 98*, 121–153.
Jardetzky, O., & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, pp 71, 98, Academic Press, New York.
Jarema, M. C., Miller, J. H., & Lu, P. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2707.
Kalk, A., & Berendsen, H. J. C. (1976) *J. Magn. Reson. 24*, 343–366.
Kuznetsov, A. N., Wasserman, A. M., Volkov, A. I., & Korst, N. N. (1971) *Chem. Phys. Lett. 12*, 103–109.
Likhtenshtein, G. I. (1976) *Spin-Labeling Methods in Molecular Biology*, p 86, Wiley, New York.
McConnell, H. M. (1971) *Annu. Rev. Biochem. 40*, 227–236.
McConnell, H. M., Deal, W., & Ogata, R. T. (1969) *Biochemistry 8*, 2580–2585.
Moffat, J. K. (1971) *J. Mol. Biol. 58*, 79–88.
Morrow, J. S., & Gurd, F. R. N. (1975) *CRC Crit. Rev. Biochem. 10*, 221–287.
Neya, S., & Morishima, I. (1980) *Biochemistry 19*, 258–265.
Ohe, M., & Kajita, A. (1980) *Biochemistry 19*, 4443–4450.
Ohnishi, S., Maeda, T., Ito, T., Huang, K., & Tyuma, I. (1968) *Biochemistry 7*, 266.
Perutz, M. F. (1980) *Proc. R. Soc. London, Ser. B 208*, 135–162.
Perutz, M. F., Muirhead, H., Cox, J. M., & Goaman, L. C. G. (1968) *Nature (London) 219*, 131–139.
Raiford, D. S., Fisk, C. L., & Becker, E. D. (1979) *Anal. Chem. 51*, 2050.
Riggs, A. (1981) *Methods Enzymol. 76*, 14–18.
Rummens, F. H. A. (1976) *Prog. Nucl. Magn. Reson. Spectrosc. 9*, 1.
Schroeder, W. A., & Huisman, T. H. J. (1980) *The Chromatography of Hemoglobin*, pp 56–65, Marcel Dekker, New York.
Sharma, V. S., Vedvick, T. S., Magde, D., Luth, R., Friedman, D., Schmidt, M. R., & Ranney, H. M. (1980) *J. Biol. Chem. 255*, 5879–5884.
Sheard, B., Yamane, T., & Shulman, R. G. (1970) *J. Mol. Biol. 53*, 35–48.
Shulman, R. G., Withrich, K., Yamane, T., Patel, D. J., & Blumberg, W. E. (1970) *J. Mol. Biol. 53*, 143–157.
Sykes, B. D., & Weiner, J. H. (1980) in *Magnetic Resonance in Biology* (Cohen, J. S., Ed.) Vol. I, Chapter 4, Wiley, New York.
Tong, J. H., Petitclerc, C., D'Iorio, A., & Benoiten, N. L. (1971) *Can. J. Biochem. 49*, 877–881.
Van Geet, A. L. (1970) *Anal. Chem. 42*, 679.
Waterman, M. R. (1978) *Methods Enzymol. 52*, 457.
Westhead, E. W., & Boyer, P. D. (1961) *Biochim. Biophys. Acta 54*, 145–156.
Wilson, T. H. (1962) *Intestinal Absorption*, W. B. Saunders, Philadelphia.
Winter, M. R. C., Johnson, C. E., Lang, G., & Williams, R. J. P. (1972) *Biochim. Biophys. Acta 263*, 515–545.
Yamane, T., Wuthrich, K., Shulman, R. G., & Ogawa, S. (1970) *J. Mol. Biol. 49*, 197–202.
Yip, Y. K., Waks, M., & Beychok, S. (1972) *J. Biol. Chem. 247*, 7237–7244.

Characterization of the Complementary Deoxyribonucleic Acid and Gene Coding for Human Prothrombin†

Sandra J. Friezner Degen, Ross T. A. MacGillivray,‡ and Earl W. Davie*

ABSTRACT: The DNA sequences of a complementary deoxyribonucleic acid (cDNA) and a portion of the gene coding for human prothrombin have been determined. The cDNA was 2005 base pairs in length and was found to code for part of a leader sequence of 36 amino acids, 579 amino acids present in the mature protein, a stop codon, a noncoding region of 97 base pairs, and a poly(A) tail of 27 base pairs. It is proposed that the leader sequence consists of a signal sequence and a pro sequence for the mature protein that circulates in plasma. The 10 glutamic acid residues that are present in the amino-terminal region of prothrombin and are converted to γ-carboxyglutamic acid in the mature protein are coded by only the GAG codon. The cDNA for prothrombin was also employed as a probe for screening a human fetal liver genomic DNA library. One of the strongly positive phage containing a human DNA insert of 5 kilobases was mapped with restriction endonucleases and sequenced. This DNA contained approximately half of the gene for human prothrombin and included six introns and five exons coding for amino acid residues 144–448. The two largest intervening sequences in the genomic DNA contained two copies each of *Alu*I repetitive DNA.

Prothrombin (M_r 72 000) is a vitamin K dependent protein that participates in the final phase of blood coagulation (Mann & Elion, 1980). It is synthesized in the liver and secreted into the blood where it participates in the coagulation process. When coagulation is initiated, prothrombin is converted to thrombin by minor proteolysis by factor X_a (activated Stuart factor) in the presence of factor V_a (activated proaccelerin), calcium ions, and phospholipid. Thrombin then converts fibrinogen to an insoluble fibrin clot. The amino acid sequences for bovine and human prothrombin have been reported (Magnusson et al., 1975; Butkowski et al., 1977; Thompson et al., 1977; Walz et al., 1977; Seegers, 1979). Each protein contains approximately 8% carbohydrate, including three carbohydrate chains and 10 residues of γ-carboxyglutamic acid

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received January 6, 1983.* This work was supported in part by Grant HL 16919 from the National Institutes of Health.

‡ Present address: Department of Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5.

0006-2960/83/0422-2087$01.50/0

(Stenflo et al., 1974; Nelsestuen et al., 1974; Magnusson et al., 1975). Prothrombin also contains two regions of internal homology called kringle structures (Magnusson et al., 1975). These structures occur in the amino-terminal region of the protein.

In order to gain a clearer understanding of the biosynthesis and regulation of prothrombin, we have recently isolated and characterized a cDNA coding for the bovine molecule (MacGillivray et al., 1980). In this paper, we describe the isolation and characterization of a cDNA coding for the human molecule and a λ phage containing a portion of the gene coding for the human protein.

Materials and Methods

Screening Human Liver cDNA Library. A human liver cDNA library containing approximately 18 000 transformants was screened for human prothrombin by using two different probes. This cDNA library was kindly provided by Drs. S. L. C. Woo and T. Chandra of Baylor College of Medicine and contained cDNAs inserted into the *Pst*I site of pBR322 (Chandra et al., 1983). A portion of the library was screened initially with a nick-translated bovine prothrombin cDNA probe by the method of Grunstein & Hogness (1975). The probe was an *Ava*I/*Bam*HI restriction fragment that contained 1200 base pairs (bp)[1] coding for amino acids 109–500.[2] It was employed at reduced stringency. These conditions included hybridization at 60 °C in 2 × Denhardt's solution, 6 × SSC, 0.5% sodium dodecyl sulfate, and 1 mM EDTA and washing the filters at 60 °C in 6 × SSC with 0.5% sodium dodecyl sulfate. Two positive clones were identified (pHII-1 and pHII-2) and their plasmids purified (Katz et al., 1973, 1977).

The library was also screened by Drs. K. Kurachi, S. Leytus, and S. Yoshitake in our department by using a mixture of synthetic DNAs 14 nucleotides in length. This mixture was purchased from P-L Biochemicals and contained the following sequences:

$$5'\quad C\text{-}C\text{-}\begin{matrix}A\\G\\C\\T\end{matrix}\text{-}G\text{-}C\text{-}\begin{matrix}A\\G\end{matrix}\text{-}C\text{-}A\text{-}\begin{matrix}A\\G\end{matrix}\text{-}A\text{-}A\text{-}C\text{-}A\text{-}T\quad 3'$$

The probe was labeled with T4 kinase and [γ-^{32}P]ATP and screened by a modification of the method of Wallace et al. (1981). Approximately 30 positive clones were found and 14 hybridized with the cDNA probe coding for prothrombin. The remainder were primarily cDNAs coding for other serine proteases, including factor IX. Two of the positive clones for prothrombin (pHII-3 and pHII-4) were further characterized, and their plasmids were also isolated by the method of Katz et al. (1973, 1977).

The four plasmids were further characterized by restriction enzyme mapping. All restriction enzymes (Bethesda Research Laboratories or New England Biolabs) were used according to the manufacturer's instructions.

DNA Sequence Analysis. Restriction fragments were labeled in a number of ways depending on the restriction site and the desired strand to be labeled. The 5′ ends of fragments with overhanging 5′ ends were labeled with [γ-^{32}P]ATP (New England Nuclear; approximately 3000 Ci/mmol) by using polynucleotide T4 kinase (New England Nuclear or Bethesda Research Laboratories) by the exchange reaction of Maxam & Gilbert (1980). The 3′ ends of fragments with 5′-overhanging ends were labeled with the appropriate [α-^{32}P]dNTP (New England Nuclear; 800 Ci/mmol) by using the Klenow fragment of *Escherichia coli* DNA polymerase (P-L Biochemicals; Smith et al., 1979). The 3′ ends of fragments that were blunt or contained 3′-overhanging ends were labeled with cordycepin 5′-[α-^{32}P]triphosphate (New England Nuclear; approximately 6000 Ci/mmol) and calf thymus terminal transferase (New England Nuclear; Tu & Cohen, 1980) according to the manufacturer's instructions. The 5′ ends of fragments that were blunt or contained 3′-overhanging ends were labeled with [γ-^{32}P]ATP (New England Nuclear; approximately 3000 Ci/mmol) by using polynucleotide T4 kinase after prior treatment of the DNA with bacterial alkaline phosphatase (Maxam & Gilbert, 1980). The alkaline phosphatase was purchased from Worthington Biochemical Co. Labeled fragments were subjected to base modification and cleavage (Maxam & Gilbert, 1980) and subjected to electrophoresis on 6% and 20% polyacrylamide gels (Sanger & Coulson, 1978). DNA sequences were stored and analyzed by the computer programs of Staden (1977, 1978). The programs were adapted for use on a local computer facility by Dr. Jon Herriott in our department.

Screening of Human Genomic DNA Library. The *Alu*I/*Hae*III human fetal liver genomic DNA library (Lawn et al., 1978; Maniatis et al., 1978) was kindly provided by Dr. Tom Maniatis of Harvard University. The library containing 2 × 10^6 phage was screened by the in situ plaque hybridization technique of Benton & Davis (1977) as modified by Woo (1979). In these experiments, a nick-translated bovine prothrombin *Ava*I/*Bam*HI probe was hybridized and washed under the same conditions used for screening the human cDNA library. Twelve positive phage were identified, and three were found to code for human prothrombin. One of these phage, named λ10, was fully sequenced. Another of the positive phage coded for amino acid residues 392–510 in prothrombin and was only partially characterized.

Preparation of Phage DNA. Approximately 1 × 10^6 phage were plated on each of 24 L-agar plates and incubated at 37 °C for 9 h. Tris-HCl buffer (10 mM at pH 7.6 containing 20 mM $MgCl_2$) was overlaid on each plate, and the set of plates was stored at 4 °C overnight. The overlay was removed, centrifuged to remove agar, and digested with DNase I and RNase A (both 1 μg/mL) for 1 h at 37 °C. The phage were precipitated by addition of NaCl (to 0.4 M) and poly(ethylene glycol)-6000 (to 9%) followed by incubation on ice for at least 2 h. Phage were pelleted at 6000 rpm for 30 min in a GSA rotor, resuspended, and again digested with DNase I (5 μg/mL) and RNase A (25 μg/mL) for 30 min at 37 °C. After centrifugation, the supernatant was applied to a cesium chloride step gradient (1.3–1.7 specific density) and centrifuged at 30 000 rpm for 1 h at 10 °C in a SW40 rotor. Phage were extensively dialyzed against 0.1 M Tris-HCl (pH 7.6) containing 0.3 M NaCl and 1.25 mM EDTA. Phage DNA was extracted by digestion with proteinase K (100 μg/mL) in the presence of 1% sodium dodecyl sulfate at 37 °C for at least 1 h followed by several phenol/chloroform extractions. Approximately 50 μg of DNA was obtained by this procedure.

Southern Hybridization Analysis. Restriction enzyme digests were subjected to electrophoresis on 0.7% agarose and transferred to nitrocellulose (Schleicher & Schuell, Keene, NH) by the Southern procedure (Southern, 1975) as modified by Smith & Summers (1980). This procedure was developed for bidirectional transfer but can be employed for transfer in one direction by placing a used X-ray film on one side to block the transfer in that direction.

[1] Abbreviations: bp, base pair(s); SSC, 0.15 M sodium chloride and 0.015 M trisodium citrate (pH 7.0); EDTA, ethylenediaminetetraacetic acid; kb, kilobase(s); Tris, tris(hydroxymethyl)aminomethane.

[2] R. T. A. MacGillivray and E. W. Davie, unpublished results.

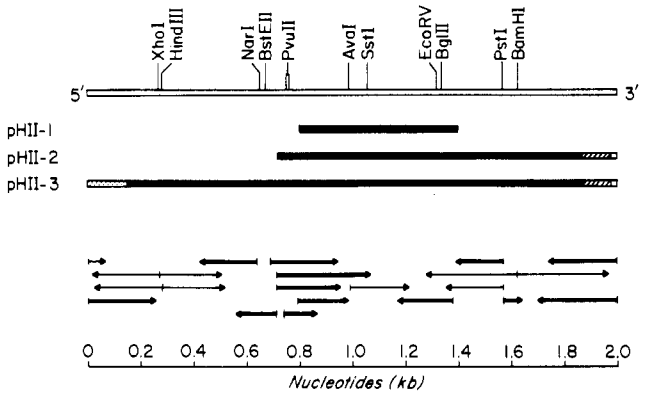

FIGURE 1: Restriction map and sequencing strategy for three human prothrombin cDNAs. The bars below the restriction map represent the three clones (pHII-1, pHII-2, and pHII-3). The region coding for the leader sequence is represented by a dotted bar, the coding region by a solid bar, and the 3'-noncoding region by a slashed bar, followed by an open bar for the poly(A) tail. The extent of sequencing is indicated by the length of each arrow, and the site of labeling is represented by a vertical line. Fragments labeled at their 3' and 5' ends are indicated by heavy and light arrows, respectively. The scale shown at the bottom represents nucleotides in kilobases.

Subcloning. Preparation of phosphatase-treated linear pBR322 and subcloning conditions are as described by Goodman & MacDonald (1979). The vector to target DNA ratio was determined by the method of Dugaiczyk et al. (1975).

E. coli RR1 cells were subjected to $CaCl_2$ shock (Bolivar et al., 1977) and transformed with the subcloning mixture. Colonies were plated on the appropriate antibiotic containing L-agar plates and grown at 37 °C overnight. Colonies containing the desired subcloned plasmids were identified by colony hybridization (Grunstein & Hogness, 1975).

Results

Human Prothrombin cDNAs. A human liver cDNA library of 18 000 transformants was screened for cDNAs coding for prothrombin employing two different probes. The first probe was a cDNA of 1200 bp coding for residues 109–500 in the bovine molecule (*AvaI*/*BamHI* probe). The second probe was a synthetic nucleotide 14 bases in length containing 16 different DNA sequences complementary to the amino acid sequence of Met-Phe-Cys-Ala-Gly (amino acid residues 505–509). Four of the plasmids for prothrombin that were identified with these two probes were further characterized by restriction mapping and DNA sequencing.

A partial restriction map of three of the plasmids (pHII-1, pHII-2, and pHII-3) is shown in Figure 1. pHII-3 has a DNA insert of approximately 2000 bp and was the longest cDNA clone obtained. Its entire DNA sequence was determined by the method of Maxam & Gilbert (1980) by using the strategy shown in Figure 1. The complete sequence for this plasmid is shown in Figure 2 along with the predicted amino acid sequence. The DNA for this plasmid coded for part of a leader sequence of 36 amino acids as well as the entire coding region of the mature protein present in plasma. This plasmid was 2005 bp in length, including 16 G's and 15 C's on the 5' and 3' ends, respectively.

pHII-1 was found to code for amino acids 224–427 in human prothrombin and was 610 bp in length. It did not contain the 3' end of the cDNA. This was probably due to excessive S1 nuclease digestion during the cloning procedure or failure of the reverse transcriptase to synthesize a second full-length strand. pHII-2 was found to be 1288 bp in length and contained DNA coding for amino acids 200–579, a 3'-noncoding region, and a 14-bp poly(A) tail. The size of the insert in pHII-4 was estimated at 950 bp by restriction mapping.

Table I: Differences in the Amino Acid Sequence of Human Prothrombin As Determined by cDNA and Amino Acid Sequencing Techniques

amino acid residue[a]	cDNA	Walz[b]	Butkowski[c]	Thompson[d]
76	His	Asn		
78	Asn	Ser		
121	Asn	Ile		
133	Val	Ala		
140	Ile	Thr		
151	Ala	Met		
152	Met	Val		
266	absent	Glu		
267	absent	Glu		
279 (281)	Thr	Pro	Thr	
283 (285)	Pro	Thr	Pro	
292 (294)	Asp	Asp	Asn	Asp
306 (308)	Asp	Asp	Asn	Asp
310 (312)	Arg	Lys	Arg	Arg
326 (328)	Asp	Asp	Asn	Asp
355 (357)	Asp	Asp	Asn	Asp
360 (362)	Thr	Thr	Thr	Ser
371 (373)	Asp	Asn	Asn	
442 (444)	Asp	Asp	Asn	Asp
451 (453)	Gln	Gly	Gly	Gln
461 (463)	Trp	Trp	Tyr	Trp
466 (468)	Glu	Ser	Ser	
468 (470)	Trp	Val	Val	
471 (473)	Asn	Asp	Asp	
486 (488)	Pro	Ala	Ala	
487 (489)	Ile	Leu	Leu	
489 (491)	Glu	Gln	Gln	

[a] The numbers shown in parentheses refer to the residue number as determined by amino acid sequencing techniques. [b] From Walz et al. (1977) and Seegers (1979). [c] From Butkowski et al. (1977). [d] From Thompson et al. (1977).

The amino acid sequence predicted by the DNA sequence of pHII-3 corresponds to a mature protein of 579 amino acids. When this sequence was compared to that reported by amino acid sequencing techniques (Butkowski et al., 1977; Thompson et al., 1977; Walz et al., 1977; Seegers, 1979), several differences were noted (Table I). The most notable difference was the absence of a Glu-Glu dipeptide reported by amino acid sequence analysis (Walz et al., 1977). Because of this difference, the numbering system for human prothrombin following residue 265 differs from that reported previously by amino acid sequencing techniques.

The amino acid composition of mature human prothrombin was determined to be as follows: Asp_{35}, Asn_{26}, Thr_{35}, Ser_{35}, Glu_{41}, Gln_{21}, Gla_{10}, Pro_{31}, Gly_{47}, Ala_{36}, Val_{33}, Met_8, Ile_{22}, Leu_{42}, Tyr_{21}, Phe_{20}, Lys_{29}, His_{10}, Arg_{39}, Trp_{14}, and $^1/_2$-Cys_{24}. The molecular weight for the protein was calculated to be 65 740 without carbohydrate and 71 612 with the addition of 8.2% carbohydrate (DiScipio et al., 1977).

The amino-terminal sequence of the mature human prothrombin is Ala-Asn-Thr-Phe-Leu and is indicated in Figure 2 as starting at +1. Preceding this sequence is a portion of a leader sequence of 36 amino acids (−1 to −36). This leader sequence is a partial signal sequence since it does not contain a potential methionine start site. This leader sequence includes a hydrophobic stretch of amino acids (residues −31 to −26) which is characteristic of signal sequences (Blobel et al., 1979). The leader sequence ends with Arg just prior to the amino-terminal Ala that is present in the mature protein. Since an Arg–Ala bond is not a typical cleavage site for signal peptidase (Blobel et al., 1979), it appears likely that the newly synthesized prothrombin contains a pre–pro leader sequence

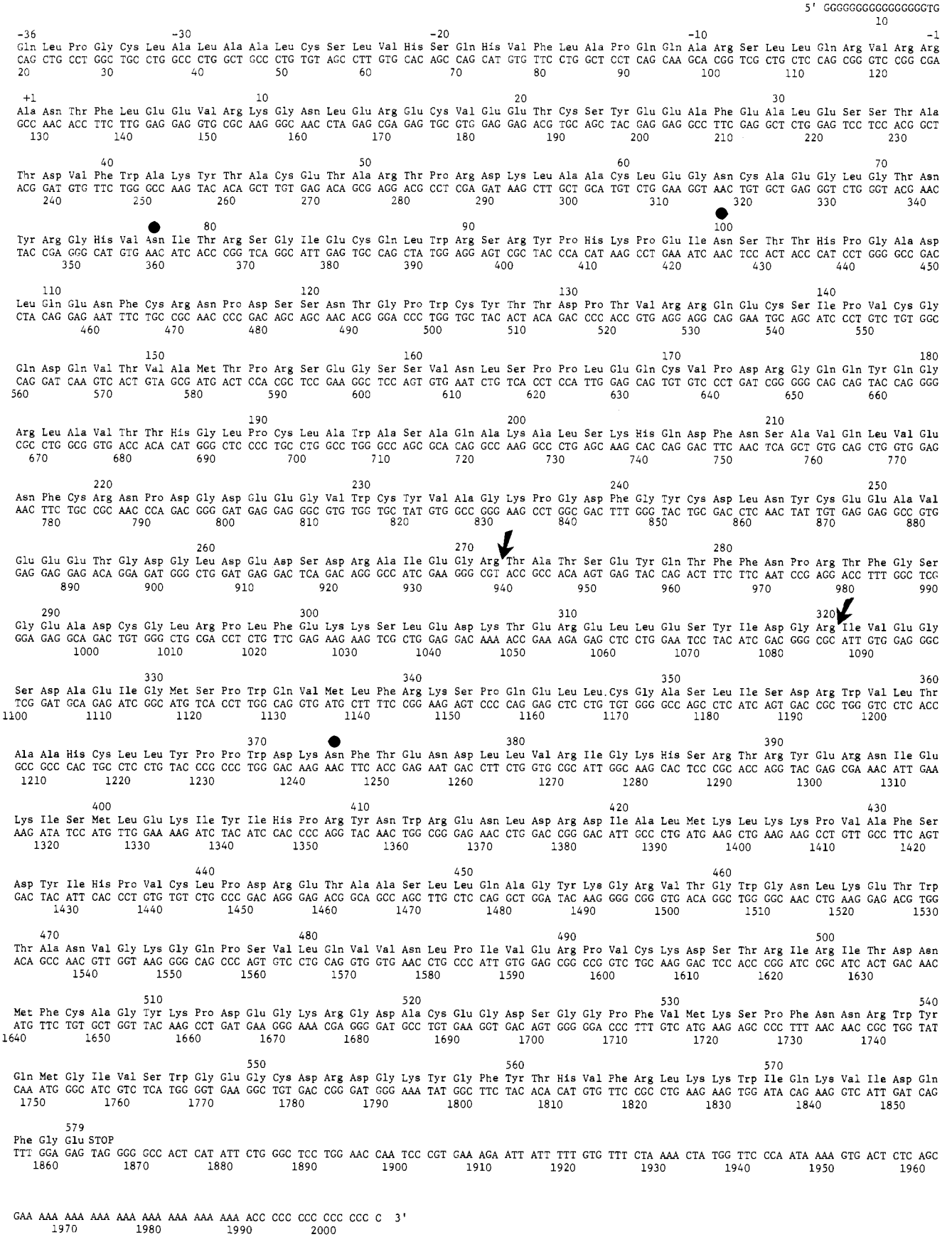

FIGURE 2: Complete nucleotide sequence of the insert in pHII-3 coding for human prothrombin. DNA sequence of the coding strand and the corresponding predicted amino acid sequence are also shown for human prothrombin (pHII-3). Residues −36 to −1 represent a part of the leader sequence for human prothrombin, and residues 1–579 represent in the sequence of the mature protein present in plasma. The apparent carbohydrate binding sites are shown with (♦), and the cleavage sites for factor X_a are shown by the heavy arrows.

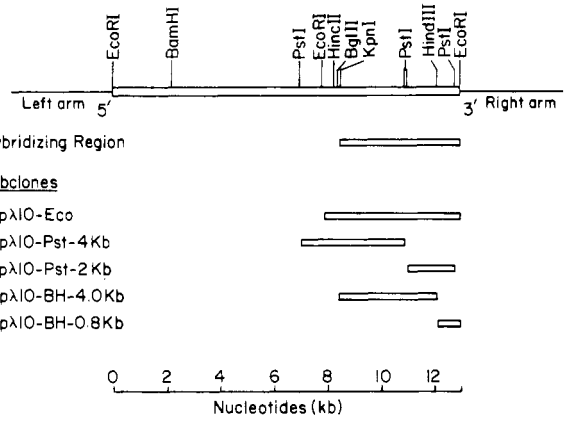

FIGURE 3: Partial restriction map and subcloned fragments for λ10. The orientation of the inserted DNA (13 kilobases) of λ10 with respect to the left and right arms of the phage is shown. The 5′ and 3′ indicate the direction of the coding strand. The region that hybridizes with the bovine prothrombin *Ava*I/*Bam*HI probe is shown along with the fragments subcloned into pBR322. The scale shown at the bottom represents nucleotides in kilobases.

analogous to some of the other plasma proteins, such as factor IX (Kurachi & Davie, 1982) and serum albumin (Patterson & Geller, 1977; Strauss et al., 1978; MacGillivray et al., 1979; Lawn et al., 1981). Accordingly, it is likely that the signal peptidase cleaves at an Ala (−10) or a Ser (−8) resulting in a pro leader sequence of seven or nine amino acids.

The coding sequence for human prothrombin is followed by a stop codon of TAG starting at nucleotide 1864. Other stop codons are present in the 3′-noncoding region, but these are in different reading frames. The noncoding region of 97 bp also contains an AATAAA sequence that appears to be required for polyadenylation (Proudfoot & Brownlee, 1976). This sequence is 14 bp upstream from the poly(A) tail of 27 bp.

Human Prothrombin Gene. The *Alu*I/*Hae*III human fetal liver genomic DNA library (Lawn et al., 1978; Maniatis et al., 1978) was screened with the bovine prothrombin *Ava*I/*Bam*HI probe. These experiments were initiated before a human cDNA probe was available. Phage (2×10^6) were screened at reduced stringency to accommodate some base-pair mismatching due to small differences in the two species. Twelve positive phage were identified that hybridized to varying degrees to the probe. Each phage was plaque purified and the DNA isolated.

The phage that hybridized the strongest was named λ10. It contained a human DNA insert of approximately 5 kilobases. This was much smaller than expected, since most of the human DNA inserts are 15–20 kilobases in size. On digestion of λ10 with *Eco*RI, two fragments of 7.8 and 5.0 kilobases were observed in addition to the right and left arms of the λ phage. By restriction mapping, size, and lack of hybridization, it was found that the 7.8-kilobase fragment was from the native Charon 4A bacteriophage (de Wet et al., 1980).

Table II: Location and Size of the Intervening and Coding Sequences in λ10

intervening sequence	location[a]	size (bp)	coding sequence	region[b]	size (bp)
1	144	1955+	1	144–248	315
2	249	324	2	249–291	129
3	292	84	3	292–334	127
4	334	1157	4	335–390	168
5	390	497	5	391–448	174
6	448	28+			

[a] Location with respect to amino acid residue. [b] Region of amino acid sequence.

Restriction mapping of λ10 was performed by digestion with one or two restriction enzymes at a time. In this manner, the orientation of the human DNA insert was determined with respect to the right arm (10.9 kilobases) and left arm (19.8 kilobases) of the phage, as well as the 7.8-kilobase insert. A restriction map was then obtained for the λ10 human DNA insert (Figure 3).

A number of subclones were constructed in pBR322 to facilitate further mapping and DNA sequencing. A partial restriction map and the sequencing strategy for the 5-kilobase human DNA insert of λ10 are shown in Figure 4. The distribution and frequency of the 6-bp restriction enzyme sites greatly facilitated sequencing by the method of Maxam & Gilbert (1980). As can be seen, all overlaps were obtained for the sequence of the human DNA insert. The actual size of the insert was 4957 bp (Figure 5). Sixty-one percent of the DNA was sequenced on both strands, and 94% was sequenced two or more times.

With the use of the computer programs of Staden (1977, 1978), it was found that the 5-kilobase fragment coded for amino acids 144–448 of human prothrombin, representing 50% of the coding sequence present in the prothrombin gene. Six intervening and five coding sequences were located in this part of the gene (Figure 4). The size of the intervening sequences ranged from 84 bp to greater than 1955 bp, while the coding sequences were around 150 bp except for one of 315 bp (Table

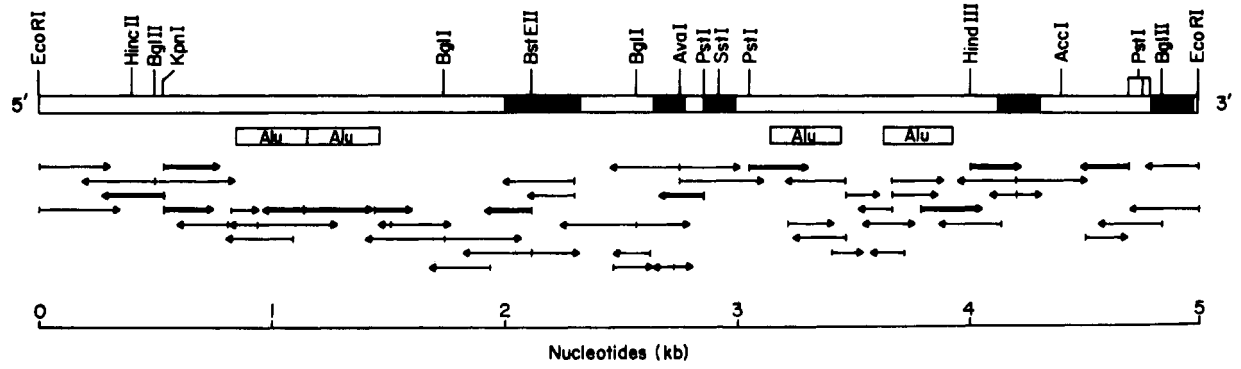

FIGURE 4: Partial restriction map and sequencing strategy for the human DNA present in λ10. The restriction map shows most of the enzymes that recognize six base sequences. The direction of transcription is indicated 5′ to 3′. Dark boxes indicate regions coding for human prothrombin. The boxes with Alu inside indicate the placement of the four Alu repetitive DNA sequences within λ10. The sequencing strategy is shown by the arrows; the vertical lines represent sites of labeling, and the horizontal arrows show the distance sequenced from each site. Fragments labeled at their 3′ and 5′ ends are indicated by heavy and light arrows, respectively. The scale shown at the bottom represents nucleotides in kilobases.

1 GAATTCATGCCACCTTCAGAGCTGGCGTCAGTCATTCAGATCATATCTGTGCCTATTGCTCAGTAAAGTCAGGGAATCAGGGGATCTGAGTGGGGGGATC

101 TGCCAGCCTCCTCCTCCCCCTCCCCACTCTTGACTTCCTTATGGTCTAGGCTGTGGCTCATTCCAAACATGCCTCCTTTCTGATCAAGGCACTCCTCCCT

201 CCGGGAAGCCCTCCCTAGCCATTTCAGTCCACACACCCTGTTCTGAGTATCACAGAGCAAGCCTTGTGCAGTTTGGCCCGCGGGATTCTGTCATTATTAT

301 TTCCTTGGTGTGTTAAGTAGCTATAGCCACCCCTTCCCTGAGGCAGACCACAATAAGCATTTCTTTTTCCCATGAGGGTTGGCAGGTGTGGCTGCACTCG

401 CTAATGCGTCTGTAGGGTCAACTGACGGAGGTTGGCCCTGGCTGGGTGGCTCTGATTCAAATAATGGGTCCAGCTGAGTCTGGCTCCTCGTTGAGGGTTG

501 GGCCTAGATCTGCTCCACGTGCGTTCATGCTGGGGCTGAGGCTGAAAGAAGGTACCTGGGAAAACTCTTCTTATGCTGATGACAGACACAGAAAACAATG

601 AACAGAAAAGCGTCTTCTGTCCTGAAGGCCTGGCTCAGAACAGGCACAGTCAGCCCTGCCCACGTTCCATTGGCCAGAGCAAGTATATGTTCAAGGCCAG

701 GGTCAAGAGGTAAACTACACCTCAGCCTGTAAAATCACAGAGCAAGGGATGTGGATGCAGGCAGGGGTAAAGAATTTGTGCCGATTACCAGTCCACAAAC

801 ATGCGTTAGTGTTTGTTCTCTAGGCAACCCTGTCGGGCCCATTGCTCATTCCTGGGGTTGGTCTTTTTTTTTTTTCTTTCTAAGAAGGAGTCTCACTCCC

901 TTGCCCAGGCTGTTGGAGTGCAGTGGCCCTATCTCAGCTCACTGCAACCTCCGCCTCCTGGGTTCAAGCGATTCCCCTGCTTCAGCCTCCTGAGTAGCTA

1001 GGATTACAGGCGTGTGCCACCACTCCTGGCTAATTTTTTTTTATGTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGATCTCAAACTCCTGACCT

1101 TGTGATCCTCCCGCCTCGGCCTCCCAAACTGCTGAGATTACAGGGGTGAGGCACTGCGCCCAGCCATTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTC

1201 ACTCTCACCCAGGCTGGAGTGCAGTGGCATAATCTTGGCTCACTGCAACCTCCACCTCCTGGGTTCAGGCGATTCTCTGCCTCAGCCTCTCATATAGCTG

1301 GGATTACAGGCACACGCCACCACGCCTTGCTAATTTTGTATTTTTAGTAGAGACGGGGTTTCTTCATGTTGGCCTTGCCTGACTTGAACTCCTTGTTCCG

1401 GTGATCTGCCCAGCTCGGCCTCCCAAAGTTCTGGGATTACAGGTGTAAGCCACTGCGCCTGGCCCCTGGTATTGGTCTTATAGCAAGTTTATCCCAACAA

1501 AAACAGCTACTATTTACTCCCCAACCCCCATACACACGCACACACATTGATGATAAATAAGTTGCAGGCTTGCAGAAATTGGCCCATCCAGGTGAACAGC

1601 CTAGTGATCCGAGCAAGCGTCCTGCTGTGCAGCTATAAAAACATGACTCCTCCAGCAGCTCCAGGCAGCCACTACCAGTTGGTTACAGATGGCCTAGGAG

1701 GCCAAACCTGGTTACTATCTCTGGTTTATTATGTGCCAGACACTTATGCTGTATATTTTGTTTAATCCTCTCAACAAACCTGCAAAAGTGGCATTAGTAA

1801 CCCCTTTAAAGGCAAACGGTCAGAAGCCCAGAGAGGTTAAGTAACCTGAGGTCACACAGGCAGAAAGCAGCAAGACCGGGGTTCACACCCCTGTCTGTTC

144
GlyGlnAspGlnValThrValAlaMetThrProArgSerGluGlySer
1901 CGGTCCATGTGTGGTCTCACTCACTCTGCTGCCTCCTTGCCCCTCACCCACCAGGCCAGGATCAAGTCACTGTAGCGATGACTCCACGCTCCGAAGGCTC

SerValAsnLeuSerProProLeuGluGlnCysValProAspArgGlyGlnGlnTyrGlnGlyArgLeuAlaValThrThrHisGlyLeuProCysLeu
2001 CAGTGTGAATCTGTCACCTCCATTGGAGCAGTGTGTCCCTGATCGGGGGCAGCAGTACCAGGGGCGCCTGGCGGTGACCACACATGGGCTCCCCTGCCTG

AlaTrpAlaSerAlaGlnAlaLysAlaLeuSerLysHisGlnAspPheAsnSerAlaValGlnLeuValGluAsnPheCysArgAsnProAspGlyAsp
2101 GCCTGGGCCAGCGCACAGGCCAAGGCCCTGAGCAAGCACCAGGACTTCAACTCAGCTGTGCAGCTGGTGGAGAACTTCTGCCGCAACCCAGACGGGGATG

248
GluGluGlyValTrpCysTyrValAlaGlyLysProGlyAspPheGlyTyrCysAspLeuAsnTyrCys
2201 AGGAGGGCGTGTGGTGCTATGTGGCCGGGAAGCCTGGCGACTTTGGGTACTGCGACCTCAACTATTGTGGTGAGCTGCCTGGGTAGGGGGCCTGAGTTGC

2301 AGGGACAAATCCTAGTGGGAATAACAACAGCCGCTTCTGCTTATCGAACGCTTACCTCATTGAGTGCGCTCATTACAGCCTTACAGTAACCAGGTGGGGG

2401 GTAAGGTCCTGTGCCCATTTCACAGATAAGTACACTGAGGCCCCAGGAGGTTATTGCCTAGTAGCCCAACTGTGCATGCACGCTTAACCTCTGCACCAAA

249
GluGluAla
2501 TGGCCTCCAAGGCCCGTAGGGGAACTGGGGGGATCTAGGGGATGGGTGAGGAATGGCCCAGCCCAGTCCCGGCCGGTGCCTGGGTCCCAACAGAGGAGGC

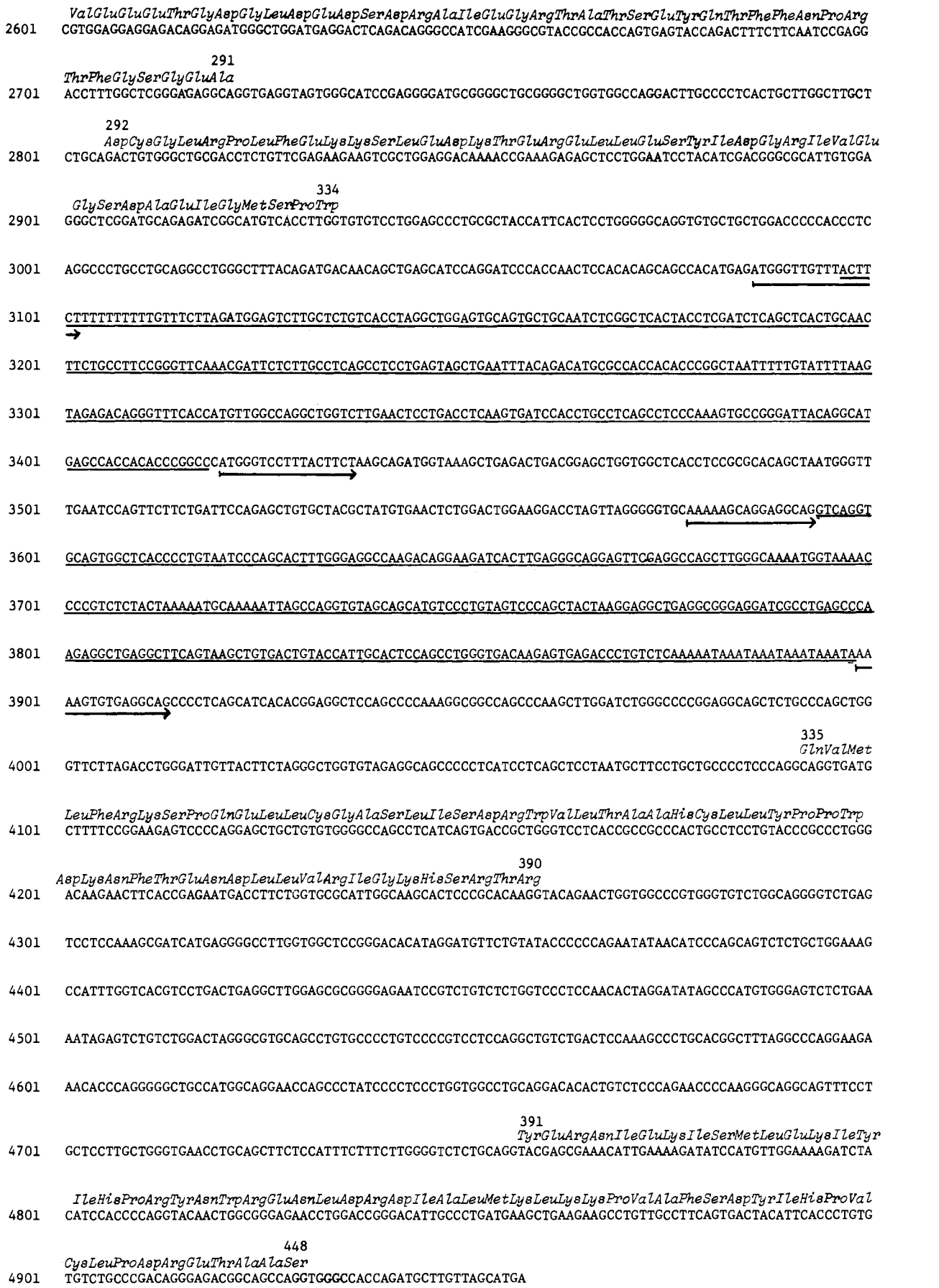

FIGURE 5: Complete nucleotide sequence of the 5-kilobase insert of λ10 coding for human prothrombin. The DNA sequence for the coding strand is shown. Numbers on the left of each row indicate the nucleotide number of the first base in that row from the 5′ end. Coding sequences are indicated with the predicted amino acid sequence above the DNA sequence. All Alu repetitive DNA sequences are underlined, and all flanking repeats are indicated by arrows.

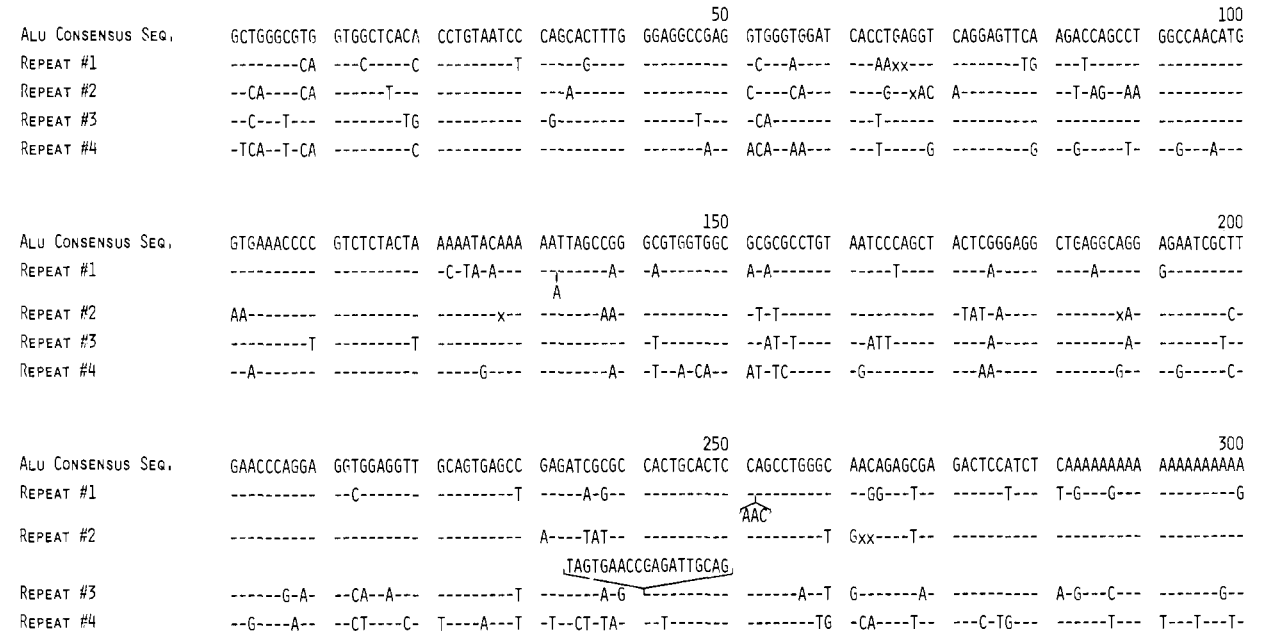

FIGURE 6: Alu repetitive DNA sequences within the human prothrombin gene. The four Alu repetitive DNA sequences of the human prothrombin gene are compared with the Alu consensus sequence of Deininger et al. (1981). Dashes represent identity with the consensus sequence while an x represents a deletion. Three insertion sequences are also shown. All differences are indicated. The position of each repeat with respect to the DNA sequence of λ10 in Figure 5 is as follows: repeat 1 is from nucleotides 861 to 1167, repeat 2 from 1168 to 1464, repeat 3 from 3097 to 3418, and repeat 4 from 3595 to 3898.

II). The first two intervening sequences were located at amino acid residues 144 and 249. These amino acids immediately follow kringle structures 1 and 2 in human prothrombin. Intervening sequences within the human DNA insert of λ10 constitute 87% of the DNA. Splice junctions between intervening and coding sequences follow the GT-AG rule of Breathnach et al. (1978) and agree well with the consensus sequences of Mount (1982). All of the intervening sequences interrupt codons between the first and second nucleotide or the second and third nucleotide and thus are class I or II introns (Sharp, 1981).

The base composition of the gene thus far sequenced was fairly random, including 21.7% A, 23.4% T, 27.0% G, and 28.0% C. The dinucleotide frequencies for the human DNA insert in λ10 corresponded quite well with those found by Nussinov (1981). Most notable was a CG frequency of 0.32, suggesting some constraint on the occurrence of this dinucleotide.

The coding sequences for λ10 are shown in Figure 5 along with the predicted amino acid sequence. The amino acid sequence corresponds exactly with that predicted from the cDNA sequence. Two differences were noted, however, in the DNA sequence, but these did not change the amino acid sequence. The differences in the cDNA sequence were observed at nucleotide 948 where an A was found in the cDNA and a C was found in the gene and at nucleotide 1293 where a C was found in the cDNA and an A was found in the gene. In both cases, the sequence of the three different cDNAs was the same, indicating that an error by the reverse transcriptase was unlikely during the preparation of the cDNA library. This suggests that these differences are probably due to a small amount of polymorphism in the prothrombin gene.

The two largest intervening sequences occur at amino acid residues 144 and 334, and each contains two copies of *Alu*I repetitive DNA. The position of each Alu repeat within the 5-kilobase insert of λ10 is shown in Figure 4. The first two repeats are present in a head-to-tail orientation with no DNA in between. These repeats span nucleotides 861–1464 in the DNA sequence of the 5-kilobase insert. The second Alu sequence starts at position 1168. The other two Alu sequences are located at nucleotides 3097–3418 and 3595–3898. These two Alu sequences are present in an inverted orientation with respect to each other and have 141 bp of DNA in between. The sequence of each repeat with respect to the consensus sequence for Alu repeats of Deininger et al. (1981) is shown in Figure 6. Each repeat is approximately 85% identical with the consensus sequence and 77% identical with each other. There are regions that appear to be conserved when all the Alu repeats are compared, most notably in the regions of 21–50 and 91–120. There are also regions with great variability, as can be seen by nucleotides 2–10, 51–57, and 151–156. It is interesting to note that the region at 151–156 in the consensus sequence has the sequence GCGCGC; in all of the prothrombin Alu repeats, this sequence has changed to ACACGC, GTGTGC, GCATGT, and ATGTCC.

Also typical of Alu repetitive DNA is the presence of short flanking repeats on each end of the Alu sequence. These sequences are not conserved, and of the sequences reported so far, none are the same. The short flanking repeats present on the ends of the Alu repeats in the prothrombin gene are shown in Figure 7. The flanking sequences on the 5′ and 3′ ends of Alu repeats 1 and 2 are 15 bp in length. These sequences are not present between the two Alu repeats. The flanking sequences at the ends of Alu repeats 3 and 4 are 17 and 16 bp in length, respectively. Figure 7 also depicts the relative orientation of each Alu repeat with respect to each other. The sequences shown in Figure 6 for the three repeats with the same orientation are present on the strand complementary to the coding strand for human prothrombin.

Discussion

A human prothrombin cDNA of 2005 bp was isolated from a cDNA library prepared from human liver mRNA. This cDNA was found to code for part of a leader sequence of 36 amino acids followed by the entire coding region of human prothrombin and a 97-bp noncoding region at the 3′ end. A poly(A) tract of 27 bp was found at the extreme 3′ end of the cDNA.

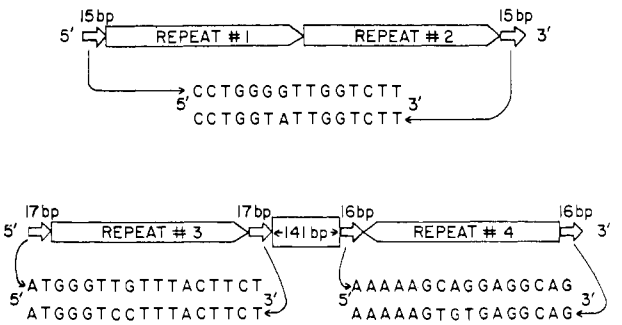

FIGURE 7: Direct flanking repeats of Alu sequences in the human prothrombin gene. The orientation of each Alu repeat is shown with respect to each other. Repeats 1 and 2 are in direct tandem repeat orientation, and repeats 3 and 4 are in an inverted repeat orientation with 141 bp of DNA in between. Repeats 1 and 2 are flanked by direct repeats of 15 bp. Repeat 3 has a flanking repeat of 17 bp, while repeat 4 has a flanking repeat of 16 bp. The orientation of the coding strand is indicated by the 5′ and 3′. The location of each repeat is indicated in Figure 4.

The leader sequence for bovine prothrombin has been partially characterized in a cell-free translation system (MacGillivray et al., 1979). When this partial leader sequence starting with methionine was aligned with the human prothrombin leader sequence, the amino acid sequence was identical in 10 out of the 11 residues that were identified in positions −35 to −22. Since the bovine leader sequence contains an additional eight amino acids upstream from position −35, it appears very likely that the human prothrombin leader sequence is 43 amino acids in length.

The size of the mRNA coding for human prothrombin as determined by Northern analysis (data not shown) was approximately 2100 nucleotides in length. This indicates that the 5′-noncoding region is probably less than 50 nucleotides in length.

The amino acid sequence of the mature human prothrombin molecule is coded by nucleotides 127–1863 (Figure 2), followed by a TAG stop codon. When the amino acid sequence for prothrombin as predicted by the cDNA sequence or the coding region from the genomic DNA was compared with that determined by amino acid sequence analysis (Butkowski et al., 1977; Thompson et al., 1977; Walz et al., 1977; Seegers, 1979), a number of differences were noted (Table I). Many of the discrepancies involve amide–acid assignments. There were, however, other additional differences. A Glu-Glu dipeptide that was found in positions 266 and 267 by amino acid sequence analysis was not present in the sequence translated from the cDNA and genomic DNA. Other differences include a histidine at position 76 instead of asparagine, and an asparagine at position 78 instead of serine. Position 76 was ori-

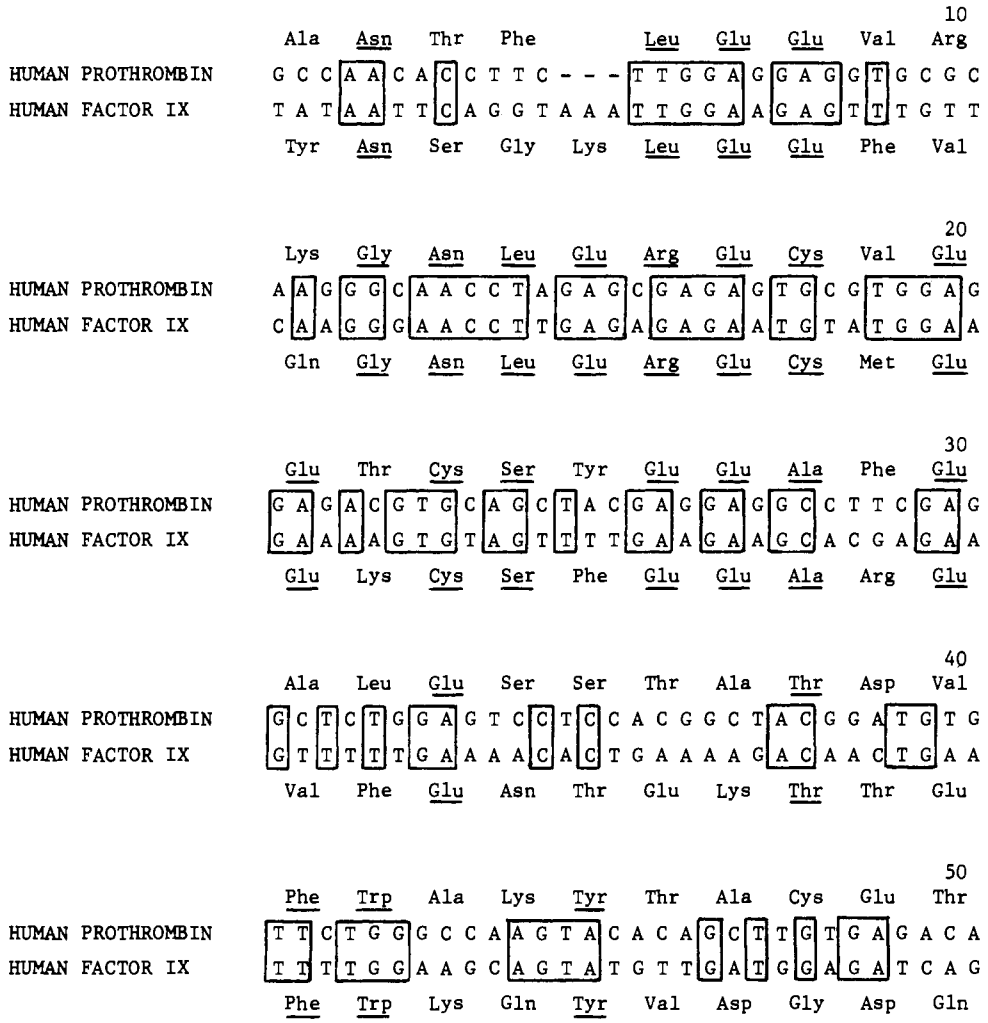

FIGURE 8: Comparison of the DNA sequences coding for the amino-terminal portions of the mature proteins for human prothrombin and human factor IX. The DNA sequences for human prothrombin and human factor IX (Kurachi & Davie, 1982) are compared at the region coding for amino acid residues 1–50 in each mature protein. Regions of identity are indicated in boxes. The amino acid sequences are shown above or below each DNA sequence. Dashes represent gaps inserted for better alignment of the sequences. Amino acids that are identical are underlined. The first 10 and 12 glutamic acid residues in prothrombin and factor IX, respectively, are γ-carboxyglutamic acid residues in the mature proteins.

ginally identified as a carbohydrate attachment site by amino acid sequence analysis. The amino acid sequence predicted from the cDNA indicates that the carbohydrate binding site is at residue 78 instead of 76. Two other potential carbohydrate binding sites include the asparagine residues at positions 100 and 373. These correspond well with the carbohydrate binding sites in bovine prothrombin which are located on asparagine residues 77, 101, and 376 (Magnusson et al., 1975). Other differences were noted at positions 121, 133, 140, 151, 152, 466, 468, 486, and 487. In most of the remaining cases, the translated sequence agrees with at least one of the reported sequences.

The codon usage for human prothrombin was typical of mRNAs coding for mammalian proteins where there is a high frequency of occurrence of G or C in the third position. For human prothrombin, 73% of the codons end in G or C. The base composition for the coding region of the cDNA was 59% G+C. The overall codon usage was fairly random, except that two codons were not used at all, including TTA for Leu and TCT for Ser.

When the cDNAs coding for human and bovine prothrombin were compared, an 85.6% identity was observed for the coding region. The 3′-noncoding region was 70% identical between the two species, although the bovine noncoding region was 119 bp in length while the human molecule was 97 bp in length.

The cDNA and amino acid sequences for human prothrombin and human factor IX (Kurachi & Davie, 1982) showed little homology in their leader sequences. In contrast, the amino-terminal regions showed striking similarities between the two vitamin K dependent clotting proteins (Figure 8). The amino acid sequences for the first 50 residues in the two mature proteins were 48% identical and the cDNA sequences were 49% identical. Following amino acid residue 50, the homology dropped dramatically until residues 320–336 and residues 505–528 where the homology is high for the corresponding amino acids (Davie et al., 1979) as well as their cDNAs. The first 40 amino acids contain all of the γ-carboxyglutamic acids present in the two clotting factors (Stenflo et al., 1974; Nelsestuen et al., 1974; Magnusson et al., 1975; Katayama et al., 1979; Kurachi & Davie, 1982). These amino acid residues result from a posttranslational carboxylation of glutamic acid. In prothrombin, the 10 glutamic acid residues were coded by GAG. In contrast, 10 of the 12 glutamic acid residues in human factor IX were coded by GAA and only 2 by GAG (Kurachi & Davie, 1982).

Four copies of Alu repetitive DNA were found within two of the intervening sequences of the human prothrombin gene. Alu repetitive DNA sequences are the predominant class of repeated DNA in the human genome. It has been estimated that 300000–500000 copies of Alu repetitive DNA sequences are present per haploid genome, representing 3–5% of the total mass of the human DNA (Schmid & Deininger, 1975; Rinehart et al., 1981; Schmid & Jelinek, 1982). The Alu repetitive DNA sequences are primarily interspersed among single copy genes at intervals of approximately 5000 bp (Schmid & Deininger, 1975; Rinehart et al., 1981). Alu repetitive DNA has been found in the flanking regions of the human insulin gene (Bell et al., 1980) and in the human β-globin gene cluster (Baralle et al., 1980). Recently, there have been reports of intervening sequences containing Alu repetitive DNA. Chicken α2(I) collagen gene (Wozney et al., 1981), chicken conalbumin (Cochet et al., 1979), and the rat prolactin gene (Weber & Gorski, 1982) have been shown by Southern hybridization analysis to contain repetitive DNA sequences within the intervening sequences. Rat growth hormone (Barta et al., 1981; Page et al., 1981) and human corticotropin–β-lipotropin precursor gene (Tsukada et al., 1982) have been shown to contain Alu repetitive DNA sequences by DNA sequence analysis.

Alu repetitive DNA in a direct tandem repeat orientation was found in one of the intervening sequences in the human prothrombin gene, and this has not been previously reported. Alu repeats are usually present individually in genomic DNA or in inverted repeat orientation. The latter situation occurs with the other two Alu repeats present in the human prothrombin gene. Whether additional Alu repeats are present in the human prothrombin gene remains to be determined. Experiments are now in progress to determine the sequence of the remainder of the prothrombin gene in order to identify the promoter region and additional intervening sequences that may also contain Alu sequences.

Acknowledgments

We thank Drs. Dominic W. Chung, Mark W. Rixon, George Long, Savio L. C. Woo, and T. Chandra for their generous help and fruitful discussions and Drs. Kotoku Kurachi, Steven P. Leytus, and Shinji Yoshitake for their help in screening the cDNA library with the synthetic probe. Thanks are also due to Judy Morishima for excellent technical assistance.

Registry No. Prothrombin, 9001-26-7; prothrombin (human liver clone pHII-3-encoded, reduced), 84959-53-5.

References

Baralle, F. E., Shoulders, C. C., Goodbourn, S., Jeffreys, A., & Proudfoot, N. J. (1980) *Nucleic Acids Res. 8*, 4393–4404.

Barta, A., Richards, R. I., Baxter, J. D., & Shine, J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4867–4871.

Bell, G. I., Pictet, R., & Rutter, W. J. (1980) *Nucleic Acids Res. 8*, 4091–4109.

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

Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H., & Lingappa, R. (1979) *Soc. Exp. Biol. Symp. 33*, 9–36.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., & Falkow, S. (1977) *Gene 2*, 95–113.

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

Butkowski, R. J., Elion, J., Downing, M. R., & Mann, K. G. (1977) *J. Biol. Chem. 252*, 4942–4957.

Chandra, T., Stackhouse, R., Kidd, V. J., & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., & Chambon, P. (1979) *Nature (London) 282*, 567–574.

Davie, E. W., Fujikawa, K., Kurachi, K., & Kisiel, W. (1979) *Adv. Enzymol. Relat. Areas Mol. Biol. 48*, 277–318.

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

de Wet, J. R., Daniels, D. L., Schroeder, J. L., Williams, B. G., Denniston-Thompson, K., Moore, D. D., & Blattner, F. R. (1980) *J. Virol. 33*, 401–410.

DiScipio, R. G., Hermodson, M. A., Yates, S. G., & Davie, E. W. (1977) *Biochemistry 16*, 698–706.

Dugaiczyk, A., Boyer, H. W., & Goodman, H. M. (1975) *J. Mol. Biol. 96*, 171–184.

Goodman, H. M., & MacDonald, R. J. (1979) *Methods Enzymol. 68*, 75–90.
Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.
Katayama, K., Ericsson, L. H., Enfield, D. L., Walsh, K. A., Neurath, H., Davie, E. W., & Titani, K. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4990–4994.
Katz, L., Kingsbury, D. T., & Helinski, D. R. (1973) *J. Bacteriol. 114*, 577–591.
Katz, L., Williams, P. H., Sato, S., Leavitt, R. W., & Helinski, D. R. (1977) *Biochemistry 16*, 1677–1683.
Kurachi, K., & Davie, E. W. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6461–6464.
Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., & Maniatis, T. (1978) *Cell (Cambridge, Mass.) 15*, 1157–1174.
Lawn, R. M., Adelman, J., Bock, S. C., Franke, A. E., Houck, C. M., Najarian, R. C., Seeburg, P. H., & Wion, K. L. (1981) *Nucleic Acids Res. 9*, 6103–6114.
MacGillivray, R. T. A., Chung, D. W., & Davie, E. W. (1979) *Eur. J. Biochem. 98*, 477–485.
MacGillivray, R. T. A., Degen, S. J. F., Chandra, T., Woo, S. L. C., & Davie, E. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5153–5157.
Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) pp 123–149, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–701.
Mann, K. G., & Elion, J. (1980) in *CRC Handbook Series in Clinical Laboratory Science, Section I: Hematology* (Schmidt, R. M., Ed.) Vol. 3, pp 15–31, CRC Press, Boca Raton, FL.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.
Nelsestuen, G. L., Zytkovicz, T. H., & Howard, J. B. (1974) *J. Biol. Chem. 249*, 6347–6350.
Nussinov, R. (1981) *J. Mol. Biol. 149*, 125–131.
Page, G. S., Smith, S., & Goodman, H. M. (1981) *Nucleic Acids Res. 9*, 2087–2104.
Patterson, J. E., & Geller, D. M. (1977) *Biochem. Biophys. Res. Commun. 74*, 1220–1226.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Rinehart, F. P., Ritch, T. G., Deininger, P. L., & Schmid, C. W. (1981) *Biochemistry 20*, 3003–3010.
Sanger, F., & Coulson, A. R. (1978) *FEBS Lett. 87*, 107–110.
Schmid, C. W., & Deininger, P. L. (1975) *Cell (Cambridge, Mass.) 6*, 345–358.
Schmid, C. W., & Jelinek, W. R. (1982) *Science (Washington, D.C.) 216*, 1065–1070.
Seegers, W. H. (1979) *Prog. Chem. Fibrinolysis Thrombolysis 4*, 241–254.
Sharp, P. A. (1981) *Cell (Cambridge, Mass.) 23*, 643–646.
Smith, G. E., & Summers, M. D. (1980) *Anal. Biochem. 109*, 123–129.
Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., & Hall, B. D. (1979) *Cell (Cambridge, Mass.) 16*, 753–761.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Staden, R. (1977) *Nucleic Acids Res. 4*, 4037–4051.
Staden, R. (1978) *Nucleic Acids Res. 5*, 1013–1015.
Stenflo, J., Fernlund, P., Egan, W., & Roepstorff, P. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2730–2733.
Strauss, A. W., Bennett, C. A., Donohue, A. M., Rodkey, J. A., Boime, I., & Alberts, A. W. (1978) *J. Biol. Chem. 253*, 6270–6274.
Thompson, A. R., Enfield, D. L., Ericsson, L. H., Legaz, M. E., & Fenton, J. W., II (1977) *Arch. Biochem. Biophys. 178*, 356–367.
Tsukada, T., Watanabe, Y., Nakai, Y., Imura, H., Nakanishi, S., & Numa, S. (1982) *Nucleic Acids Res. 10*, 1471–1479.
Tu, C.-P. D., & Cohen, S. N. (1980) *Gene 10*, 177–183.
Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879–894.
Walz, D. A., Hewett-Emmett, D., & Seegers, W. H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 1969–1972.
Weber, J. L., & Gorski, J. (1982) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 41*, 1448.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 389–395.
Wozney, J., Hanahan, D., Tate, V., Boedtker, H., & Doty, P. (1981) *Nature (London) 294*, 129–135.