Gene for Human Factor X: A Blood Coagulation Factor Whose Gene Organization Is Essentially Identical with That of Factor IX and Protein C†

Steven P. Leytus, Donald C. Foster, Kotoku Kurachi, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received May 6, 1986; Revised Manuscript Received June 30, 1986

ABSTRACT: Factor X is one of six vitamin K dependent proteins known to be involved in blood coagulation, the others being factor VII, factor IX, prothrombin, protein S, and protein C. In the present studies, recombinant bacteriophage containing overlapping DNA inserts coding for the gene for human factor X have been isolated and characterized. These DNA inserts code for almost the entire gene for factor X, extending from the prepro leader peptide through the 3′ noncoding region of the transcription product. The organization of the gene for factor X was established by DNA sequencing to identify the location of the introns and exons in the gene. Seven introns and eight exons were identified and their intron/exon boundaries established. The seven introns interrupt the coding sequence at essentially identical locations in the amino acid sequence as the introns in the genes for human factor IX and protein C. In addition, the introns in the gene for factor X divide the coding sequence into discrete exons that code for potential structural and functional domains of the protein. This information provides strong evidence to support the suggestion that the vitamin K dependent proteins present in plasma have evolved from a single, common gene and that this ancestral gene arose through a process that involved the assembly of small protein coding units of DNA into a single gene.

Factor X (Stuart factor) is a plasma glycoprotein that participates in the middle phase of the intrinsic and extrinsic pathways of the blood coagulation cascade (Davie et al., 1979). The human protein (M_r 59 000) circulates in blood as a two-chain zymogen to a serine protease. It is composed of a light chain (M_r 16 900) and a heavy chain (M_r 42 100), and these two chains are held together by a disulfide bond (DiScipio et al., 1977a,b). In the intrinsic pathway, factor X is converted to factor X_a by minor proteolysis by factor IX_a in the presence of factor $VIII_a$, calcium, and phospholipid. In the extrinsic pathway, factor X is converted to factor X_a by factor VII_a in the presence of tissue factor. When factor X_a is generated, it in turn converts prothrombin to thrombin, an enzyme that converts fibrinogen to fibrin.

Factor X is one of six plasma glycoproteins involved in blood coagulation that requires vitamin K for biosynthesis (Stenflo & Suttie, 1977). The other five vitamin K dependent proteins include factor VII, factor IX, prothrombin, protein S, and protein C. Factor VII, factor IX, factor X, and prothrombin participate in the generation of fibrin, whereas protein C in the presence of protein S regulates the coagulation process by inactivating factor V_a and factor $VIII_a$ (Kisiel et al., 1977; Vehar & Davie, 1980; Marlar et al., 1982). The first 10–12 glutamic acid residues at the amino-terminal regions of the vitamin K dependent proteins are converted to γ-carboxyglutamic acid residues by a carboxylation reaction that requires vitamin K as a cofactor. The γ-carboxyglutamic acid residues are involved in calcium binding and mediate the interaction of the vitamin K dependent proteins to phospholipid surfaces provided by platelets at the site of vascular injury.

Comparison of the primary structures of the vitamin K dependent serine proteases with those of other serine proteases has given rise to the suggestion that these proteins are composed of a number of structural or functional domains (Patthy, 1985). For instance, factor VII, factor IX, factor X, and protein C each contain in corresponding positions (i) a signal peptide that is involved in secretion, (ii) a propeptide that could be involved in the vitamin K dependent γ-carboxylation reaction (Pan & Price, 1985), (iii) an amino-terminal region, referred to as the Gla domain, that contains the γ-carboxyglutamic acid residues that participate in calcium binding, (iv) two potential domains that share homology with epidermal growth factor and whose function is not known, (v) a connecting region or activation peptide, and (vi) a catalytic or protease domain that includes the active site serine, aspartic acid, and histidine residues. In prothrombin, the potential growth factor domains have been replaced by two kringle structures.

Recently, it has been shown that the similarity between factor IX and protein C extends beyond their protein sequence and is even evident in the organization of their genes (Anson et al., 1984; Yoshitake et al., 1985; Foster et al., 1985). In both genes, there are seven introns and eight exons that occur in essentially the same positions throughout the amino acid sequence and the prepro leader sequence of the two proteins. Furthermore, the introns in the two genes are located primarily between exon segments coding for the various functional domains in the proteins. Since factor X is also a vitamin K dependent plasma serine protease, it was of interest to determine its gene structure and to correlate its pattern of intron/exon boundaries with that of factor IX and protein C. Observed similarities or differences in gene organization would have significant implications with regard to hypothesizing how this particular class of serine proteases arose from the primordial serine protease gene and at what point in the evolutionary history of vitamin K dependent serine proteases they diverged from other serine proteases.

Experimental Procedures

Human genomic libraries in λ Charon 4A (Lawn et al., 1978) and λ EMBL3 (Yoshitake et al., 1985) were screened

† This work was supported in part by research grants (HL 16919 to E.W.D. and HL 31511 to K.K.) and a postdoctoral fellowship (GM 09118 to S.P.L.) from the National Institutes of Health. D.C.F. was supported by an NIH training grant (GM 07270).

0006-2960/86/0425-5098$01.50/0

for genomic clones of human factor X by plaque hybridization (Benton & Davis, 1977; Woo, 1979) with partial (Leytus et al., 1984) and full-length (unpublished data) cDNAs for factor X as hybridization probes. Genomic DNA inserts were released from purified recombinant phage DNA by restriction enzyme digestion with *Eco*RI and were then subcloned into the *Eco*RI site of a pUC plasmid vector (Vieira & Messing, 1982). Plasmid DNA was prepared by a modification (Micard et al., 1985) of the alkaline extraction procedure (Birnboim & Doly, 1979). Genomic DNA inserts were mapped by single- and double-restriction enzyme digestion followed by agarose gel electrophoresis, Southern blotting, and hybridization to radiolabeled cDNA probes. Selected fragments from restriction enzyme digests of recombinant plasmids were subcloned into M13 bacteriophage vectors (Messing, 1983). Genomic subclones in M13 vectors that hybridized to cDNA probes for factor X were then isolated and sequenced by the dideoxy chain termination method (Sanger et al., 1977), as described in the Amersham cloning and sequencing manual. Sequencing reactions were carried out with [α-^{35}S]dATP and the reaction products electrophoresed in 6% polyacrylamide buffer–gradient gels (Biggin et al., 1983). Nonrandom DNA sequencing by sequential *Bal*31 deletion (Poncz et al., 1982) was also used, employing modifications previously described (Yoshitake et al., 1985).

Results and Discussion

A partial cDNA (Leytus et al., 1984) and a full-length cDNA (unpublished data) coding for human factor X were used as probes to screen approximately 3 000 000 recombinant phage from λ Charon 4A (Lawn et al., 1978) and λ EMBL3 (Yoshitake et al., 1985) human genomic libraries. The full-length clone for human factor X was 1502 nucleotides in length and included 25 nucleotides of 5′ untranslated sequence and 120 nucleotides coding for a prepro leader sequence of 40 amino acids. This sequence, in conjunction with that previously published (Leytus et al., 1984; Fung et al., 1985), establishes the cDNA sequence of human factor X (Figure 1). Seven positive genomic clones were then isolated and characterized by restriction endonuclease mapping and Southern blotting. Following subcloning of genomic DNA fragments into M13 bacteriophage vectors, the subclones that hybridized to the cDNA probes coding for factor X were isolated and sequenced. From these results, the intron/exon structure was established throughout the coding region of the gene for human factor X.

Seven introns (A–G) and eight exons (I–VIII) were identified within the coding region and the 3′ noncoding region of the gene for human factor X (Table I). The DNA sequence of seven exons beginning with amino acid −17 of the prepro leader peptide was established, and these seven exons were found to be distributed over approximately 22 kb of genomic DNA. The overlapping clones that were isolated do not encompass the entire gene for factor X, however, since they do not include the exon(s) coding for the 5′ noncoding region of the mRNA and amino acids −40 through −18 of the prepro leader peptide. By analogy with the gene for factor IX (Yoshitake et al., 1985), an exon coding for residues −40 to −18 in the prepro leader sequence would correspond to exon I, while residues −17 to 37 in the prepro leader sequence and light chain correspond to exon II. Residues 55–254 in the heavy chain correspond to the last exon (exon VIII).

The nucleotide sequences at the intron/exon splice junctions follow the GT/AG rule and agree with the consensus sequences published by others (Breathnach & Chambon, 1981; Mount, 1982; Nevins, 1983). The introns vary in size and range from about 950 nucleotides (intron C) to about 7400 nucleotides (intron B) and include both splice junction types I and 0 (Table II).

The relationship of the position of the introns in the primary structure of human factor X and its tentative domains is depicted in Figure 2. The complete amino acid sequence for human factor X and its prepro leader sequence was deduced from partial (Leytus et al., 1984) and full-length (unpublished data) cDNAs. The protein coding sequence in the seven exons agrees well with the cDNA sequence with one exception. The cDNA sequence that was described earlier (Leytus et al.,

Table I: Intron–Exon Splice Junction Sequences in Gene for Human Factor X[a]

intron	splice junction sequences: exon | 5′ intron 3′ | exon
A	..... CCTGCCTTCCAG | TG TTC ATC (−17 Leu Phe Ile)
B	GAC AAG ACG | GTAAGGGCTGGG.....TTTTCCTTTTAG | AAT GAA TTC (37 Asp Lys Thr ; 38 Asn Glu Phe)
C	TAC AAA G | GTCAGTATTTTT.....CTCTCTTTGCAG | AT GGC GAC (46 Tyr Lys Asp ; 46 Asp Gly Asp)
D	GAA TTA T | GTAGGTTCCTCT.....CTCCTTTGGCAG | TC ACA CGG (84 Glu Leu Phe ; 84 Phe Thr Arg)
E	CCC ACA G | GTAGGAGGCACG.....TCTTTCTTTCAG | GG CCC TAC (128 Pro Thr Gly ; 128 Gly Pro Tyr)
F	CCC TGG CAG | GTAACAGTAGGA.....CGTCTGTCACAG | GCC CTG CTC (15 Pro Trp Gln ; 16 Ala Leu Leu)
G	AGG GTA G | GTAAGTGACCAA.....CGTCTGTCCCAG | GG GAC CGG (55 Arg Val Gly ; 55 Gly Asp Arg)
consensus sequence[b]	$^{C}_{A}$AG | GT$^{A}_{G}$AGT ..... $^{TT}_{CC}$N$^{T}_{C}$AG | G

[a] Numbers above the amino acids at the intron/exon boundaries are the same as those in Figure 2. [b] From Mount (1982).

Table II: Comparison of Location, Splice Junction Type, and Size of Introns in Genes for Human Factor X, Factor IX, and Protein C

intron	protein	location (amino acid)[a]	splice junction type[b]	size (bp)[c]
A	factor X	−17	I	?
	factor IX	−17	I	6206
	protein C	−19	I	1263
B	factor X	37/38	0	7400
	factor IX	37/38	0	188
	protein C	37/38	0	1462
C	factor X	46	I	950
	factor IX	47	I	3689
	protein C	46	I	92
D	factor X	84	I	1800
	factor IX	85	I	7163
	protein C	92	I	102
E	factor X	128	I	2900
	factor IX	128	I	2565
	protein C	137	I	2668
F	factor X	15/16	0	3400
	factor IX	15/16	0	9473
	protein C	15/16	0	873
G	factor X	55	I	1700
	factor IX	54	I	668
	protein C	55	I	1129

[a] The numbers used to define the location of the introns refer to amino acid residues in the sequences of factor X (Figure 2), factor IX (Anson et al., 1984; Yoshitake et al., 1985), and protein C (Foster et al., 1985). [b] From Sharp (1981). [c] The size of the introns in the gene for factor X was estimated from a restriction enzyme map of the gene.

```
                                                                         CAGGGACACAGTACTCGGCCACACC

   1  M   G   R   P   L   H   L   V   L   L   S   A   S   L   A   G   L   L   L   L   G   E   S   L   F   I   R   R   E   Q
  26 ATG GGG CGC CCA CTG CAC CTC GTC CTG CTC AGT GCC TCC CTG GCT GGC CTC CTG CTG CTC GGG GAA AGT CTG TTC ATC CGC AGG GAG CAG

  31  A   N   N   I   L   A   R   V   T   R   A   N   S   F   L   E   E   M   K   K   G   H   L   E   R   E   C   M   E   E
 116 GCC AAC AAC ATC CTG GCG AGG GTC ACG AGG GCC AAT TCC TTT CTT GAA GAG ATG AAG AAA GGA CAC CTC GAA AGA GAG TGC ATG GAA GAG

  61  T   C   S   Y   E   E   A   R   E   V   F   E   D   S   D   K   T   N   E   F   W   N   K   Y   K   D   G   D   Q   C
 206 ACC TGC TCA TAC GAA GAG GCC CGC GAG GTC TTT GAG GAC AGC GAC AAG ACG AAT GAA TTC TGG AAT AAA TAC AAA GAT GGC GAC CAG TGT

  91  E   T   S   P   C   Q   N   Q   G   K   C   K   D   G   L   G   E   Y   T   C   T   C   L   E   G   F   E   G   K   N
 296 GAG ACC AGT CCT TGC CAG AAC CAG GGC AAA TGT AAA GAC GGC CTC GGG GAA TAC ACC TGC ACC TGT TTA GAA GGA TTC GAA GGC AAA AAC

 121  C   E   L   F   T   R   K   L   C   S   L   D   N   G   D   C   D   Q   F   C   H   E   E   Q   N   S   V   V   C   S
 386 TGT GAA TTA TTC ACA CGG AAG CTC TGC AGC CTG GAC AAC GGG GAC TGT GAC CAG TTC TGC CAC GAG GAA CAG AAC TCT GTG GTG TGC TCC

 151  C   A   R   G   Y   T   L   A   D   N   G   K   A   C   I   P   T   G   P   Y   P   C   G   K   Q   T   L   E   R   R
 476 TGC GCC CGC GGG TAC ACC CTG GCT GAC AAC GGC AAG GCC TGC ATT CCC ACA GGG CCC TAC CCC TGT GGG AAA CAG ACC CTG GAA CGC AGG

 181  K   R   S   V   A   Q   A   T   S   S   S   G   E   A   P   D   S   I   T   W   K   P   Y   D   A   A   D   L   D   P
 566 AAG AGG TCA GTG GCC CAG GCC ACC AGC AGC AGC GGG GAG GCC CCT GAC AGC ATC ACA TGG AAG CCA TAT GAT GCA GCC GAC CTG GAC CCC

 211  T   E   N   P   F   D   L   L   D   F   N   Q   T   Q   P   E   R   G   D   N   N   L   T   R   I   V   G   G   Q   E
 656 ACC GAG AAC CCC TTC GAC CTG CTT GAC TTC AAC CAG ACG CAG CCT GAG AGG GGC GAC AAC AAC CTC ACC AGG ATC GTG GGA GGC CAG GAA

 241  C   K   D   G   E   C   P   W   Q   A   L   L   I   N   E   E   N   E   G   F   C   G   G   T   I   L   S   E   F   Y
 746 TGC AAG GAC GGG GAG TGT CCC TGG CAG GCC CTG CTC ATC AAT GAG GAA AAC GAG GGT TTC TGT GGT GGA ACC ATT CTG AGC GAG TTC TAC

 271  I   L   T   A   A   H   C   L   Y   Q   A   K   R   F   K   V   R   V   G   D   R   N   T   E   Q   E   E   G   G   E
 836 ATC CTA ACG GCA GCC CAC TGT CTC TAC CAA GCC AAG AGA TTC AAG GTG AGG GTA GGG GAC CGG AAC ACG GAG CAG GAG GAG GGC GGT GAG

 301  A   V   H   E   V   E   V   V   I   K   H   N   R   F   T   K   E   T   Y   D   F   D   I   A   V   L   R   L   K   T
 926 GCG GTG CAC GAG GTG GAG GTG GTC ATC AAG CAC AAC CGG TTC ACA AAG GAG ACC TAT GAC TTC GAC ATC GCC GTG CTC CGG CTC AAG ACC

 331  P   I   T   F   R   M   N   V   A   P   A   C   L   P   E   R   D   W   A   E   S   T   L   M   T   Q   K   T   G   I
1016 CCC ATC ACC TTC CGC ATG AAC GTG GCG CCT GCC TGC CTC CCC GAG CGT GAC TGG GCC GAG TCC ACG CTG ATG ACG CAG AAG ACG GGG ATT

 361  V   S   G   F   G   R   T   H   E   K   G   R   Q   S   T   R   L   K   M   L   E   V   P   Y   V   D   R   N   S   C
1106 GTG AGC GGC TTC GGG CGC ACC CAC GAG AAG GGC CGG CAG TCC ACC AGG CTC AAG ATG CTG GAG GTG CCC TAC GTG GAC CGC AAC AGC TGC

 391  K   L   S   S   S   F   I   I   T   Q   N   M   F   C   A   G   Y   D   T   K   Q   E   D   A   C   Q   G   D   S   G
1196 AAG CTG TCC AGC AGC TTC ATC ATC ACC CAG AAC ATG TTC TGT GCC GGC TAC GAC ACC AAG CAG GAG GAT GCC TGC CAG GGG GAC AGC GGG

 421  G   P   H   V   T   R   F   K   D   T   Y   F   V   T   G   I   V   S   W   G   E   G   C   A   R   K   G   K   Y   G
1286 GGC CCG CAC GTC ACC CGC TTC AAG GAC ACC TAC TTC GTG ACA GGC ATC GTC AGC TGG GGA GAG GGC TGT GCC CGT AAG GGG AAG TAC GGG

 451  I   Y   T   K   V   T   A   F   L   K   W   I   D   R   S   M   K   T   R   G   L   P   K   A   K   S   H   A   P   E
1376 ATC TAC ACC AAG GTC ACC GCC TTC CTC AAG TGG ATC GAC AGG TCC ATG AAA ACC AGG GGC TTG CCC AAG GCC AAG AGC CAT GCC CCG GAG

 481  V   I   T   S   S   P   L   K   *
1466 GTC ATA ACG TCC TCT CCA TTA AAG TGA GATCCCACTC-poly(A)
```

FIGURE 1: Nucleotide sequence of cDNA coding for human factor X and predicted amino acid sequence. Nucleotides 1–95 were determined by sequencing the 5′ end of a full-length cDNA coding for factor X. This region includes 25 nucleotides of 5′ untranslated sequence and the nucleotide sequence coding for the signal peptide. Nucleotides 96–1502 include exons II–VIII of the gene for factor X. The sequence presented in this region was derived from analysis of genomic clones.

1984) lacked a stretch of nine nucleotides, corresponding to amino acids 51–53 of the heavy chain. These nucleotides were present in the full-length cDNA as well as in exon VII of the gene. They were also present in a cDNA coding for human factor X described by Fung et al. (1985).

The introns in the gene for factor X do not appear to occur in a random manner. Rather, the introns tend to fall between various functional domains of the protein (Figure 2). Exon II includes an apparent propeptide region of the leader sequence (Blanchard et al., 1985) and the γ-carboxyglutamic acid domain. Exon III includes a stretch of eight amino acids that link the γ-carboxyglutamic acid and growth factor domains. Exons IV and V each code for a potential growth factor domain, while exon VI covers a connecting region that includes the activation peptide. Exons VII and VIII include the catalytic domain typical of all serine proteases.

Because of the common modular protein structures of factor X, factor IX, and protein C, a comparison of the organization of their genes provides an excellent model for studying the evolutionary processes involved in the assembly of genes coding

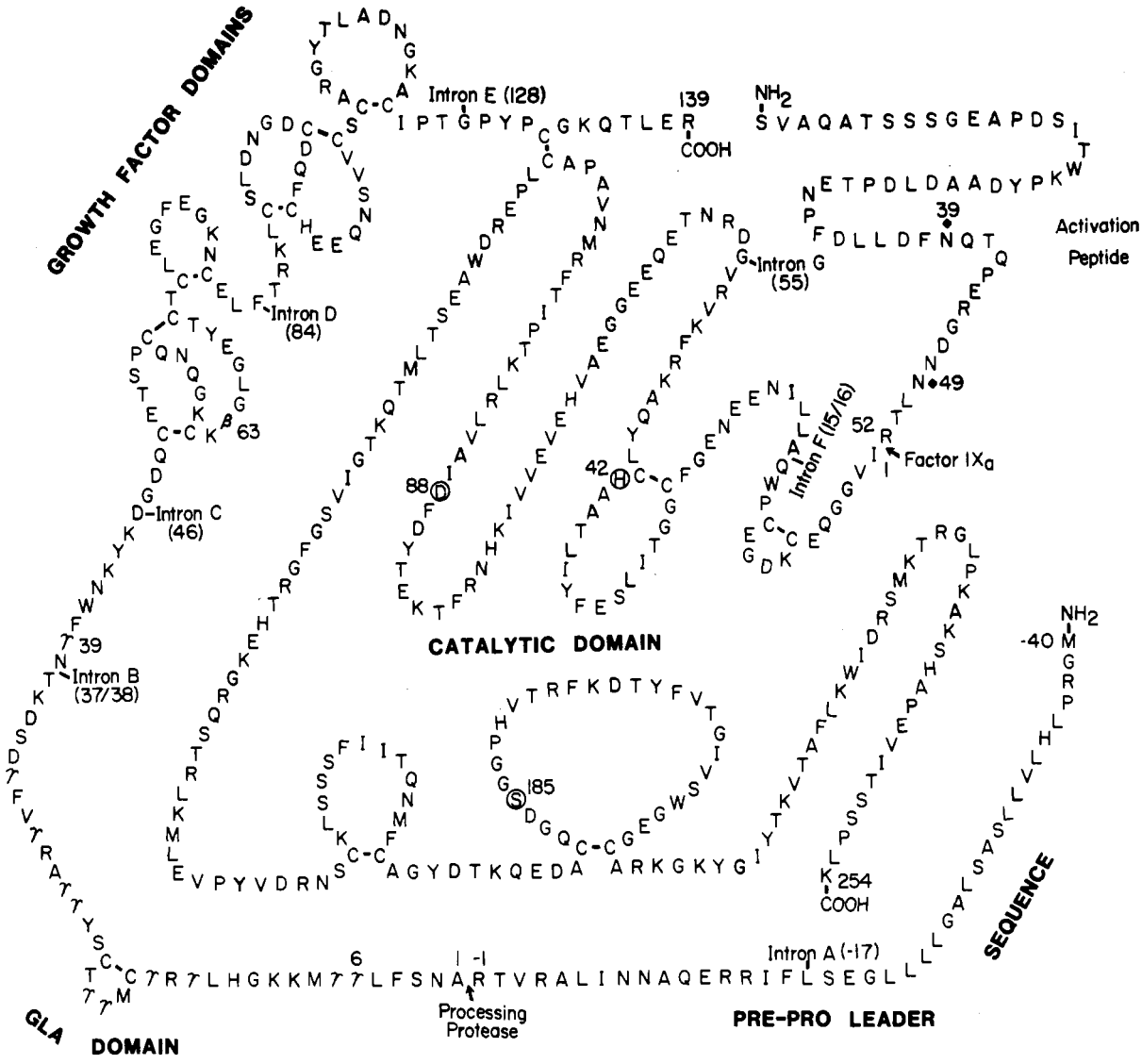

FIGURE 2: Amino acid sequence and tentative structure for human prepro factor X. The Arg-Lys-Arg tripeptide that connects the light and heavy chains during biosynthesis is not shown. The seven introns (A–G) and their location in the amino acid sequence are shown. The amino acid sequence was deduced from partial (Leytus et al., 1984) and full-length (unpublished data) cDNAs coding for human factor X. The numbering of the amino acids in this figure is different from that in Figure 1. The first amino acids in the light chain, activation peptide, and heavy chain start with number 1. The prepro leader sequence (numbered −40 to −1) is removed during biosynthesis by signal peptidase and a processing protease that hydrolyzes the Arg–Ala bond between −1 and 1. The Gla domain and potential growth factor domains are located in the light chain of factor X_a (residues 1–139). The activation peptide of 52 residues is released from factor X during its conversion to factor X_a by factor IX_a. The serine protease or catalytic domain of factor X contains 254 residues, including the three principal amino acids (His-42, Asp-88, Ser-185) participating in catalysis. The proposed disulfide bonds in factor X have been placed by analogy to those in bovine prothrombin (Magnusson et al., 1975) and epidermal growth factor (Savage et al., 1973). γ refers to γ-carboxyglutamic acid, and β refers to β-hydroxyaspartic acid.

for multidomain proteins. It has been suggested that genes coding for modular or multidomain proteins evolved by exploiting the splicing of nucleic acids to recruit and combine small segments of protein coding sequence (Gilbert, 1978; Blake, 1978, 1979; Lonberg & Gilbert, 1985). Detailed schemes have also been proposed to explain how exon shuffling could be involved in the evolution of plasma serine proteases (Patthy, 1985; Yoshitake et al., 1985; Foster et al., 1985). A comparison of the gene structures of factor X, factor IX, and protein C reveals that introns disrupt the coding sequences of these three proteins in analogous locations between the various functional domains. A comparison of intron locations, sizes, and splice junction types in the three genes is summarized in Table II. The slight differences in intron locations are completely predictable from the amino acid sequences and are due to minor variations in the sizes of corresponding domains in the three proteins. Despite the fact that their locations in the coding sequences are highly conserved, sizes of corresponding introns vary considerably and exhibit essentially no sequence homology. The first three introns in the gene for human prothrombin (Degen et al., 1983; Davie et al., 1983) also occur at locations analogous to the first three introns in the genes for factor X, factor IX, and protein C. However, in prothrombin, the γ-carboxyglutamic acid domain is followed by two kringle structures, which are unrelated in amino acid sequence to the potential growth factor domains of factor X, factor IX, and protein C. Accordingly, there appears to be little, if any, similarity between the organization of the gene for prothrombin and the organization of the genes for factor X, factor IX, and protein C following the first three introns.

Since serine proteases share considerable amino acid sequence homology, it is thought that most, if not all, are ultimately derived from a single ancestral gene. Through a series of events that probably included gene duplications, unequal

crossing over between homologous genes, crossing over between partially homologous genes, and random transpositions of genetic material, a number of different evolutionary pathways leading from the primordial gene were established from which a variety of different serine proteases have evolved. On the basis of similarities in protein structure and related physiological functions, certain serine proteases have further been grouped into distinct families. However, similarity in gene organization is a more definitive criterion for grouping serine proteases into particular families since it reveals at a more basic level their degree of relatedness. In order to trace the evolutionary history of a serine protease family, it is important to define the intron/exon boundaries in their genes and to correlate these with their structural or functional domains. Establishing that introns interrupt the coding sequence in the gene for factor X at essentially the same positions as they do in factor IX and protein C provides extremely strong evidence that vitamin K dependent plasma serine proteases truly constitute a distinct family of serine proteases. Furthermore, the vitamin K dependent plasma serine proteases apparently arose more recently through gene duplications since their gene organizations are very similar and have not had sufficient time to undergo major changes.

ACKNOWLEDGMENTS

We thank Drs. Barbara Schach, Shinji Yoshitake, Rei Asakai, and Dominic Chung for valuable discussions and Dr. Tom Maniatis for making available the human genomic library.

REFERENCES

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A., & Brownlee, G. G. (1984) *EMBO J. 3*, 1053–1060.

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.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

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

Blake, C. C. F. (1979) *Nature (London) 277*, 598.

Blanchard, R. A., Faye, K. L. K., Barnett, J. M., & William, B. (1985) *Blood 66*(Suppl. 1), 331A.

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

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

Davie, E. W., Degen, S. J. F., Yoshitake, S., & Kurachi, K. (1983) *Dev. Biochem. 25*, 45–52.

Degen, S. J. F., MacGillivray, R. T. A., & Davie, E. W. (1983) *Biochemistry 22*, 2087–2097.

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

DiScipio, R. G., Hermodson, M. A., & Davie, E. W. (1977b) *Biochemistry 16*, 5253–5260.

Foster, D. C., Yoshitake, S., & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4673–4677.

Fung, M. R., Hay, C. W., & MacGillivray, R. T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3591–3595.

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

Kisiel, W., Canfield, W. M., Ericsson, L. H., & Davie, E. W. (1977) *Biochemistry 16*, 5824–5831.

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

Leytus, S. P., Chung, D. W., Kisiel, W., Kurachi, K., & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3699–3702.

Lonberg, N., & Gilbert, W. (1985) *Cell (Cambridge. Mass.) 40*, 81–90.

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 123–249, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

Marlar, R. A., Kleiss, A. J., & Griffin, J. (1982) *Blood 59*, 1067–1072.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Micard, D., Sobrier, M. L., Couderc, J. L., & Dastugue, B. (1985) *Anal. Biochem. 148*, 121–126.

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

Nevins, J. R. (1983) *Annu. Rev. Biochem. 52*, 441–466.

Pan, L. C., & Price, P. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6109–6113.

Patthy, L. (1985) *Cell (Cambridge, Mass.) 41*, 657–663.

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

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Savage, C. R., Jr., Hash, J. H., & Cohen, S. (1973) *J. Biol. Chem. 248*, 7669–7672.

Sharp, P. A. (1981) *Cell (Cambridge. Mass.) 23*, 643–646.

Stenflo, J., & Suttie, J. W. (1977) *Annu. Rev. Biochem. 46*, 157–172.

Vehar, G. A., & Davie, E. W. (1980) *Biochemistry 19*, 401–410.

Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.

Woo, S. L. C. (1979) *Methods Enzymol. 68*, 389–395.

Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.