Characterization of a cDNA coding for human factor VII

(blood coagulation/DNA sequence/vitamin K-dependent proteins)

Frederick S. Hagen*, Charles L. Gray*, Patrick O'Hara*, Francis J. Grant*, Gena C. Saari*, Richard G. Woodbury*[†], Charles E. Hart*, Maggie Insley*, Walter Kisiel[‡], Kotoku Kurachi[§], and Earl W. Davie[§]

*ZymoGenetics, Inc., 2121 North 35th Street, Seattle, WA 98103; ‡Blood Systems Research Foundation Laboratory, Department of Pathology, University of New Mexico School of Medicine, Albuquerque, NM 87131; and §Department of Biochemistry, University of Washington, Seattle, WA 98195

Contributed by Earl W. Davie, December 20, 1985

ABSTRACT Factor VII is a precursor to a serine protease that is present in mammalian plasma. In its activated form, it participates in blood coagulation by activating factor X and/or factor IX in the presence of tissue factor and calcium. Clones coding for factor VII were obtained from two cDNA libraries prepared from poly(A) RNA from human liver and Hep G2 cells. The amino acid sequence deduced from the cDNAs indicates that factor VII is synthesized with a prepro-leader sequence of 60 or 38 amino acids. The mature protein that circulates in plasma is a single-chain polypeptide composed of 406 amino acids. The amino acid sequence analysis of the protein and the amino acid sequence deduced from the cDNAs indicate that factor VII is converted to factor VII_a by the cleavage of a single internal bond between arginine and isoleucine. This results in the formation of a light chain (152 amino acids) and a heavy chain (254 amino acids) that are held together by a disulfide bond. The light chain contains a y-carboxyglutamic acid (Gla) domain and two potential epidermal growth factor domains, while the heavy chain contains the serine protease portion of the molecule. Factor VII shows a high degree of amino acid sequence homology with the other vitamin K-dependent plasma proteins.

Factor VII is a plasma glycoprotein that participates in the coagulation process leading to the generation of fibrin (1). It is synthesized in the liver where it requires vitamin K for the formation of approximately 10 γ -carboxyglutamic acid residues that are present in the amino-terminal region of the protein (2). These amino acid residues bind to calcium ions and are involved in the interaction of the protein with phospholipid vesicles. One residue of β -hydroxyaspartic acid has also been identified in bovine factor VII (3). The biological function of the latter amino acid is not known.

Factor VII is a single-chain glycoprotein ($M_r \approx 50,000$) that is secreted into the blood where it circulates in a zymogen form (4-7). Factor VII is converted to factor VII_a by factor X_a (5, 8), factor XII_a (6, 9–11), factor IX_a (11), or thrombin (5) by minor proteolysis. This results in the formation of an enzyme composed of two polypeptide chains ($M_r \approx 20,000$ and 30,000) that are held together by a disulfide bond. The light chain of factor VII_a contains the y-carboxyglutamic acid residues, while the heavy chain contains the catalytic or serine protease portion of the molecule. In the presence of tissue factor and calcium ions, factor VII_a then converts factor X to factor X_a by limited proteolysis, and the latter enzyme in turn converts prothrombin to thrombin in the presence of factor Va, calcium ions, and phospholipid. Factor VII_a will also convert factor IX to factor IX_a in the presence of tissue factor and calcium (12). The relative physiological importance of the activation of factor X versus the activation of factor IX by factor VII_a is not clear (13).

Deficiencies in factor VII activity are rare, affecting roughly 200-400 individuals in the United States (14). Factor VII_a, however, may be useful in treating patients who have developed inhibitors to factor VIII (15). Accordingly, the preparation of this trace plasma protein by the techniques of recombinant DNA offers a potential source of this coagulation factor that also would be free of viruses, such as those causing hepatitis or acquired immune deficiency syndrome.

This paper describes the isolation and characterization of several clones coding for factor VII employing two different λ gt11 expression libraries prepared from poly(A) RNA isolated from human liver and Hep G2 cells. The isolation of a cDNA that encodes the entire factor VII protein sequence makes it possible to express factor VII in a mammalian expression system that synthesizes vitamin K-dependent proteins (16–18).

MATERIALS AND METHODS

Human $\lambda gt11$ cDNA Libraries. A $\lambda gt11$ cDNA library, containing approximately 14 million phage, prepared from poly(A) RNA from human liver was kindly provided by Savio L. C. Woo (19). A second $\lambda gt11$ cDNA library was prepared from poly(A) RNA from Hep G2 cells by an adaptation of the methods of Gubler and Hoffman (20) and Toole *et al.* (21). Twelve million recombinant phage were generated from 5 μg of poly(A) RNA. Approximately 99.7% of the phage were recombinants based on β -galactosidase activity.

Screening of the λ gt11 cDNA Libraries. Human factor VII was prepared as described (22), and rabbit antibodies (23) or monoclonal antibodies (24) were prepared and affinity purified by the method of Canfield and Kisiel (23). The purified monoclonal or polyclonal antibodies were labeled with ¹²⁵I and used to screen filter blots as described by Young and Davis (25, 26). The λ gt11 libraries were also screened with nick-translated factor VII cDNA (27).

DNA Sequence Analysis. The cDNAs were sequenced by the dideoxy chain termination method (28) using M13 universal primers and factor VII-specific oligonucleotide primers, and by the chemical degradation method of Maxam and Gilbert (29). Oligonucleotides were synthesized on an Applied Biosystems (Foster City, CA) oligonucleotide synthesizer.

Preparation of Human Factor VII for Amino Acid Sequence Analysis. Human factor VII was partially purified through the DEAE-Sepharose step essentially as described by Broze and Majerus (6). Factor VII from the DEAE-Sepharose step was then purified to homogeneity by immunoaffinity chromatog-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: Gla, γ -carboxyglutamic acid.

Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195.

Biochemistry: Hagen et al.

raphy using a column containing murine monoclonal antibody to human factor VII coupled to activated CH-Sepharose (Pharmacia). The affinity column was equilibrated with 0.05 M Tris·HCl buffer, pH 7.5, containing 0.15 M NaCl, and the adsorbed factor VII was eluted from the column with the same Tris-buffered saline containing 3 M NaSCN. Factor VII_a was prepared by incubating factor VII with human factor X_a (1:1000, enzyme:substrate weight ratio) for 30 min in the presence of 5 mM CaCl₂ and dilute human brain phospholipids. Following complete activation, factor VII_a was separated from factor X_a and phospholipids by immunoaffinity chromatography under the same conditions used to isolate factor VII.

Protein Sequence Analysis. Human factor VII_a (2 mg) was reduced and carboxymethylated by the method of Crestfield *et al.* (30), and the heavy and light chains were separated by HPLC (Varian C₁₈ Micropak column) employing an acetonitrile/H₂O gradient containing 0.1% trifluoroacetic acid. The light chain of factor VII_a was then digested with chymotrypsin overnight at 37°C in 0.1 M (NH₄)HCO₃, pH 7.8, with an enzyme-to-substrate ratio of 1 to 100. The digest was then fractionated by HPLC into about 12 peaks, as described above, and each peak was subjected to Edman degradation using a gas-phase protein sequenator, model 470A (Applied Biosystems).

RESULTS AND DISCUSSION

A radiolabeled monoclonal antibody to human factor VII was used to screen 6 million phage from a λ gt11 expression library containing cDNA inserts prepared from poly(A) RNA from human liver (25, 26). A positive clone (λ HVII2115) was isolated and plaque purified, and its cDNA insert was sequenced by the dideoxy chain termination method (Fig. 1). The cDNA insert was shown to code for factor VII by comparison of the predicted amino acid sequence with that



FIG. 1. Partial restriction maps and sequencing strategy for the cDNA inserts in clones λ HVII2115, λ HVII1923, λ HVII2463, λ HVII565, and λ HVII474. The extent of sequencing is shown by the length of each arrow, and the direction of the arrow indicates the strand that was sequenced. The heavy arrows indicate where Maxam and Gilbert sequencing was performed. The hatched bars indicate the 5'- and 3'-noncoding regions in each cDNA insert. The nucleotide numbers of the ends and break points of the cDNA inserts are indicated.

determined by amino acid sequence analysis of the human protein (22). Also, comparison with the known amino acid sequence of the amino-terminal portions of factor VII (22), factor IX (31), and factor X (32) indicated that this cDNA started with the DNA sequence encoding amino acid 36 of the mature factor VII (nucleotide 321, Fig. 2) and lacked the coding sequence for the amino-terminal region of the protein and a leader sequence. This clone also lacked two nucleotides (at positions 1005 and 1106, Fig. 2) that evidently arose from errors by reverse transcriptase (33). These nucleotides were present in other cDNA clones and in factor VII genomic DNA clones (P.O., unpublished data).

One other factor VII cDNA clone (λ HVII1923) was obtained from the human liver λ gt11 library when 14 million phage were screened with a nick-translated fragment of clone λ HVII2115. This probe included nucleotides 320–635 (Fig. 2). The cDNA insert in λ HVII1923 contained nucleotides 515–2441 (Fig. 2) and did not contain the two nucleotide deletions of clone λ HVII2115.

To obtain the 5' end of the DNA coding for factor VII, a second cDNA library was constructed in λ gt11 from poly(A) RNA from Hep G2 cells. One million phage from this library were screened with a nick-translated fragment from clone λ HVII2115 (nucleotides 320–625, Fig. 2). Three additional clones were obtained from this library and characterized. The phage λ HVII2463 contained the largest factor VII cDNA insert with the sequence of nucleotides 1-2463 (Fig. 2). The other two clones contained sequences from the 5' end of the cDNA and lacked a stretch of nucleotides coding for part of the factor VII leader sequence present in λ HVII2463. These included nucleotides 100-165 for the insert in clone λ HVII565, and nucleotides 100–179 for the cDNA of clone λ HVII474 (Figs. 1 and 2). These two clones were evidently generated by an oligo(dT) priming event at the adenosine-rich region at position 638-649 (Fig. 2) during the first-strand synthesis of the cDNA. The deletion in λ HVII474 results in a frame shift for the remainder of the amino acids for factor VII and does not represent an authentic coding sequence for the protein. The absence of nucleotides 100-165 in clone λ HVII565 indicated that this clone coded for a leader peptide of 38 amino acids rather than a leader peptide of 60 amino acids that was predicted by clone λ HVII2463. When compared to the factor VII genomic sequence, these differences in cDNA sequences appeared to result from alternative splicing events (P.O., unpublished data). These analyses, therefore, predict that factor VII mRNA contains a sequence coding for a leader peptide of 60 or 38 amino acids, 406 amino acids of mature factor VII, 1026 nucleotides of 3'-noncoding sequence, and a poly(A) tail.

The leader sequence predicted by clone λ HVII2463 and clone λ HVII565 contained four arginine residues prior to the sequence of Ala-Asn-Ala-Phe-Leu, which is the aminoterminal sequence of human factor VII circulating in plasma. Since the bond between arginine and alanine is an unlikely cleavage site for signal peptidase (34), it is evident that factor VII is synthesized with a prepro-leader sequence analogous to the other vitamin K-dependent coagulation factors. Accordingly, it seems likely that signal peptidase cleaves the newly synthesized polypeptide at a peptide bond upstream from the Arg-Ala sequence. The bond between arginine and alanine would be cleaved by an additional processing protease(s) to yield the mature protein with an aminoterminal alanine (Fig. 3). Two potential initiation methionines are present in the prepro-leader sequence in clone λ HVII-2463. These occur at codon positions -60 and -26. The methionine at codon -60, however, is the most likely start site since only this methionine is followed by a hydrophobic region typical of signal peptides. The deletion of 66 nucleotides in clone λ HVII565 removes amino acids -18 to -39. A messenger with this sequence would code for a prepro-leader

		T	CAACA	GGCA 10	GGG	GCAGC	CAC 1 20	FGCA(GAGAT	ГТ ТС 30	CATC	-60 Met ATG	Val GTC 40	Ser TCC	Gln CAG	Ala GCC 50	Leu CTC	Arg AGG	Leu CTC	Leu CTC 50	Cys TGC	-50 Leu CTT	Leu CTG 70	Leu CTT	61 y 666	Leu CTT 80	Gln CAG	G1y GGC	Cys TGC	Leu CTG 90	Ala GCT	-40 A1a GCA 1	Giy GGC 00	G1y GGG	Va l GT(
A1 <u>GC</u> 11	a Ly T A4 O	/s A1 Ng GC	a Ser C TCA 1	G1y GGA 20	Gly GGA	-30 Glu GAA 1	Thr ACA 130	Arg CGG	Asp GAC	Met ATG 140	Pro CCG	Trp TGG	Lys AAG 1	Pro CCG 50	Gly GGG	-20 Pro CCT	His <u>CAC</u> 160	Arg AGA	Val GTC	Phe TTC 170	Va1 GTA	Thr ACC	Gln CAG 18	G1u GAG 30	Glu GAA	-10 Ala GCC 1	His CAC 190	G1y GGC	Val GTC	Leu CTG 200	His CAC	Arg CGG	Arg CGC 21	Arg CGG 0	- 1 Arg CGC
A1 GC	1 a As C A# 220	sn A1 AC GC)	a Phe G TTC	Leu CTG 230	Glu GAG	Glu GAG	Leu CTG 24	Arg CGG 40	10 Pro CCG	Gly GGC	Ser TCC 250	Leu CTG	G1 u GAG	Arg AGG 260	Glu GAG	Cys TGC	Lys AAG 2	Glu GAG 70	20 Glu GAG	Gln CAG	Cys TGC 280	Ser TCC	Phe TTC	G1u GAG 290	Glu GAG	Ala GCC	Arg CGG 3(Glu GAG DO	30 Ile ATC	Phe TTC 3	Lys AAG 310	Asp GAC	Ala GCG	G1 u GAG 320	Arg AGG
Th AC	r Lj G A4	/s Le AG CT 330	40 u Phe G TTC	Trp TGG	Ile ATT 340	Ser TCT	Tyr TAC	Ser AGT 350	Asp GAT	G1y GGG	Asp GAC 36	Gln CAG 50	50 Cys TGT	Ala GCC	Ser TCA 370	Ser AGT	Pro CCA	Cys TGC 380	G1 n CAG	Asn AAT	61 y 666 39	Gly GGC 90	60 Ser TCC	Cys TGC	Lys AAG 100	Asp GAC	Gln CAG	Leu CTC 410	Gln CAG	Ser TCC	Tyr TAT 42	Ile ATC 20	70 Cys TGC	Phe TTC 4	Cys TGC 30
Le CT	u Pr C CC	ro Al CT GC 44	a Phe C TTC O	Glu GAG	G1y GGC 4	Arg CGG 50	80 Asn AAC	Cys TGT	Glu GAG 460	Thr ACG	His CAC	Lys AAG 470	Asp GAT	Asp GAC	G1n CAG 48	Leu CTG 30	90 Ile ATC	Cys TGT	Val GTG 190	Asn AAC	Glu GAG	Asn AAC 500	G1y GGC	Gly GGC	Cys TGT 51	Glu GAG IO	100 G1n CAG	Tyr TAC	Cys TGC 520	Ser AGT	Asp GAC	His CAC 530	Thr ACG	G1y GGC	Thr ACC
Ly AA 540	11 s Ar G CG	LO ng Se GC TC	r Cys C TGT 550	Arg CGG	Cys TGC	His CAC 560	Glu GAG	G1y GGG	Tyr TAC 57	Ser TCT 70	120 Leu CTG	Leu CTG	A1a GCA 580	Asp GAC	G1y GGG	Va1 GTG 590	Ser TCC	Cys TGC	Thr ACA 60	Pro CCC DO	130 Thr ACA	Val GTT	Glu GAA 510	Tyr TAT	Pro CCA	Cys TGT 620	G1y GGA	Lys AAA	Ile ATA 63	Pro CCT 30	140 Ile ATT	Leu CTA 6	Glu GAA 640	Lys AAA	Arc AGA
As AA 65	n A1 T GC	la Se CC AG	r Lys C AAA 6	Pro CCC 60	150 Gln CAA	G1y GGC	Arg CGA 670	Ile ATT	Val GTG	G1 y GGG 680	Gly GGC	Lys AAG	Val GTG 6'	Cys TGC 90	160 Pro CCC	Lys AAA	G1y GGG 700	Glu GAG	Cys TGT	Pro CCA 710	Trp TGG	Gln CAG	Val GTC 72	Leu CTG 20	170 Leu TTG	Leu TTG 7	Val GTG 730	Asn AAT	G1y GGA	Ala GCT 740	Gln CAG	Leu TTG	Cys TGT 75	G1y GGG 60	180 G1 GGC
Th AC	r Le C C1 760	eu Il Ig At D	e Asr C AAC	Thr ACC 770	Ile ATC	Trp TGG	Val GTG 71	Val GTC BO	190 Ser TCC	Ala GCG	A1a GCC 790	His CAC	Cys TGT	Phe TTC 800	Asp GAC	Lys AAA	Ile ATC 8	Lys AAG 10	200 Asn AAC	Trp TGG	Arg AGG 820	Asn AAC	Leu CTG	Ile ATC 830	Ala GCG	Val GTG	Leu CTG 84	G1y GGC 40	210 Glu GAG	His CAC E	Asp GAC 350	Leu CTC	Ser AGC	Glu GAG 860	His CAC
As GA	р G1 С G0	ly As GG GA 870	220 p G1u T GAG	Gln CAG	Ser AGC 880	Arg CGG	Arg CGG	Val GTG 890	A1a GCG	Gln CAG	Val GTC 90	Ile ATC 00	230 11e ATC	Pro CCC	Ser AGC 910	Thr ACG	Tyr TAC	Val GTC 920	Pro CCG	Gly GGC	Thr ACC 91	Thr ACC 30	240 Asn AAC	His CAC	Asp GAC 940	Ile ATC	Ala GCG	Leu CTG 950	Leu CTC	Arg CGC	Leu CTG 9f	His CAC 50	250 G1n CAG	Pro CCC و	Val GTO
Va GT	1 Le C C1	eu Th TC AC 98	r Asp T GAC	His CAT	Val GTG 9	Val GTG 90	260 Pro CCC	Leu CTC 1(Cys TGC 000	Leu CTG	Pro CCC	G1 u GAA 1010	Arg CGG	Thr ACG	Phe TTC 103	Ser TCT 20	270 G1u GAG	Arg AGG 1	Thr ACG 030	Leu CTG	Ala GCC	Phe TTC 1040	Val GTG	Arg CGC	Phe TTC 105	Ser TCA 50	280 Leu TŢG	Val GTC 1(Ser AGC 060	G1y GGC	Trp TGG	G1y GGC 1070	Gln CAG	Leu CTG	Lei CT(
As GA 1080	29 AP AN AC CO	90 rg G1 GT GG	y Ala C GCC 1090	Thr ACG	Ala GCC	Leu CTG 1100	Glu GAG	Leu CTC	Met ATG 111	Val GTG 10	300 Leu CTC	Asn AAC 1	Va1 GTG 120	Pro CCC	Arg CGG	Leu CTG 1130	Met ATG	Thr ACC	Gln CAG 114	Asp GAC 40	310 Cys TGC	Leu CTG 1	Gln CAG 150	Gln CAG	Ser TCA	Arg CGG 1160	Lys AAG	Va1 GTG	Gly GGA 117	Asp GAC 70	320 Ser TCC	Pro CCA 11	Asn AAT 180	Ile ATC	Thi ACC
G1 GA 119	u Ty IG T/ IO	yr Me AC AT	t Phe G TTC 12	Cys TGT 200	330 A1a GCC	Gly GGC 1;	Tyr TAC 210	Ser TCG	Asp GAT	G1 y GGC 1220	Ser AGC	Lys AAG	Asp GAC 12	Ser TCC 30	340 Cys TGC	Lys AAG 1	G1 y GGG 240	Asp GAC	Ser AGT	G1 y GGA 1250	G1y GGC	Pro CCA	His CAT 120	Ala GCC 50	350 Thr ACC	His CAC 12	Tyr TAC 270	Arg CGG	Gly GGC	Thr ACG 1280	Trp TGG	Tyr TAC	Leu CTG 129	Thr ACG 10	360 G1) GG0
11 AT	e Va C G 1300	al Se TC AG D	r Trp C TGO	Gly GGC 1310	Gln CAG	G1y GGC	Cys TGC 13;	Ala GCA 20	370 Thr ACC	Val GTG 1	G1 y GGC 330	His CAC	Phe TTT	G1 y GGG 1340	Val GTG	Tyr TAC	Thr ACC 13	Arg AGG 50	380 Val GTC	Ser TCC 1	G1 n CAG 360	Tyr TAC	Ile ATC	Glu GAG 1370	Trp TGG	Leu CTG	Gln CAA 13	Lys AAG 80	390 Leu CTC	Met ATG 13	Arg CGC 390	Ser TCA	Glu GAG 1	Pro CCA 400	Are CG
Pr CC	o G A G	ly Va GA GT 1410	400 1 Leu C CTC) Leu CTG 1	Arg CGA 420	Ala GCC	Pro CCA	Phe TTT 1430	406 Pro CCC	STOP TAG	CCC 144	AGCA D	GCC	CTGG 1450	CCTG	TG G 1	AGAG 460	AAAG	C CA 14	AGGC 70	TGCG	TCG 148	AACT(0	GTC	CTGG(1490	CACC	AA A' 1	TCCC/ 500	ΑΤΑΤΑ	4 TTC 151	CTTC1 10	rgca	GTT# 1520	∖ATG€)	GG
TA 15	GAG GAG	GAGGG	CATO 1540	GGAG	igg A 1	GGGAO 550	GAGG	T GG(15(GGAG(50	GGAG	ACA 157	GAGA D	CAG	AAAC 1580	AGAG	AG A 1	GACA 590	GAGA	C AG	AGAG. 00	AGAC	TGA 161	GGGA(0	GAG /	ACTC1 1620	rgago	GA CA	ATGG/ 630	AGAG/	4 GAO 164	CTCA# 40	AGA	GAC1 1650	CCAA)	IGA
TT 16	CAA)	AGAGA	CTA/ 1670	TAGA	IGA C	ACAG/ 680	AGAT	G GA/ 16'	ATAG/ 90	AAAA	GAT 170	gaga D	GGC	AGAG 1710	GCAG	AC A	GGCG 720	CTGG	A CA	GAGG 30	GGCA	GGG 174	GAGT (0	GCC /	AAGG1 1750	TGTO	CC T(1	GGAGI 760	GCAG/	A CAG	3CCC/ 70	AGCT	GAG0 1780	:стсс)	;TT
AC 17	CTC	сстто	AGC0	AAGC	cc c	ACCT(810	GCAC	G TG/ 18	ATCT(20	GCTG	GCC 183	CTCA D	GGC	TGCT 1840	GCTC	TG C 1	CTTC 850	ATTG	С ТG 18	GAGA 60	CAGT	AGA 187	GGCA [.] O	TGA	ACAC# 1880	ACAT	GG A	TGCA 890	CACAG	: ACA 19(1CAC(00	SCCA	ATG0 1910	ACAC	:AC
AC	AGA	GATAT	GCA0		AC G	GATG(940	CACA	C AC/ 19	AGATO	GGTC	ACA 196	CAGA 0	GAT	ACGC 1970	AAAC	AC A	CCGA 980	TGCA	C AC	GCAC	ATẠG	AGA 200	TATG 0	CAC	ACAC# 2010	AGAT	GC A	CACA	CAGAT	F AT/ 20:	4CAC# 30	ATGG	ATG0 2040	CACG(:AC
A1 20	GCC/	AATGO	ACG0	ACAC	AT C	AGTG	CACA	C GG/ 20	ATGC/ 80	ACAG	AGA 209	TATG 0	CAC	ACAC 2100	CGAT	GT G 2	CGCA	CACA	C AG	ATAT 20	GCAC	ACA 213	CATG	GAT	GAGC/ 2140	ACAC/	AC A	CACC/ 150	AAGTO	5 CG(21(CACA0 50	CACA	CCG# 2170	NTGT/	ACA
C# 21	CAC	AGATO	CAC/	ACACA	- IGA T 2	GCACI	ACAC	A CCI 22	GATGO 10	CTGA	CTC 222	CATG 0	TGT	GCTG 2230	тсст	ст G 2	AAGG 240	CGGT	T GT	TTAG 50	стст	CAC 226	TTTT(0	CTG	GTTC1 2270	TAT	CC A	TTAT 280	CATCI	T TC/ 22	ACTT(90	CAGA	CAA1 2300	TCAC	SAA
GC 23	CATC	ACCAT	GCA1	IGGTG	igc g	AATG	CCCC	C AA	ACTCT 40	тссс	CCA 235	AATG 0	тат	TTCT 2360	ссст	TC G	CTGG 370	GTGC	C GG 23	GCTG 80	CACA	GAC 239	TATTI 0	ccc	CACC1 2400	TGCT	TC C	CAGC 410	TTCA	AA 24	7 <u>888</u> 0	CGGC	TGC(243(этст()	ст
 CC 24	GCA	CACCT	GTGG 2450	TGCC	тб с 2	CACC(460	CAAA/	A AA/ 24	4 444 70	AAAA	AAA 248	AAAA D																							

FIG. 2. Nucleotide sequence of the cDNA insert in clone λ HVII2463. Residues -60 to -1 represent the prepro-leader sequence for the insert in clone λ HVII2463. The boxed nucleotides and amino acids are those that are absent in clone λ HVII565, resulting in a prepro leader sequence of 38 amino acids. Residues 1 through 406 represent the sequence of the mature factor VII present in plasma. The apparent carbohydrate binding sites are shown with solid diamonds, and the cleavage site for the conversion of factor VII to factor VII_a is shown by a heavy arrow. The apparent polyadenylylation or processing signal is shown with lines above and below the nucleotides.

sequence with a size and hydrophobicity pattern similar to those of human prothrombin (35), protein C (36, 37), and factor IX (38, 39).

The amino-terminal region of human factor VII contains 10 glutamic acid residues that are probable sites for carboxylation. These residues are located in positions analogous to those for the γ -carboxyglutamic acid residues in the other vitamin K-dependent proteins and, therefore, most likely

constitute the γ -carboxyglutamic acid (Gla) domain of factor VII (Fig. 3). The Gla domain in the vitamin K-dependent proteins is involved in calcium binding and, by analogy to the other vitamin K-dependent proteins, the Gla domain in factor VII is probably required for its binding to the phospholipid-tissue factor complex during the coagulation process.

The Gla domain in factor VII is followed by two potential growth factor domains (Fig. 3). The function of these do-



FIG. 3. Amino acid sequence and tentative structure for prepro-factor VII. The prepro-leader sequence of 60 amino acids (numbered -60 to -1) or 38 amino acids is removed during biosynthesis by signal peptidase and a processing protease that cleaves the arginine-alanine (R-A) bond between -1 and 1. The boxed amino acids are absent in the cDNA insert in λ HVII565, resulting in a prepro leader sequence of 38 amino acids. The Gla domain and potential growth factor domains are located within residues 1-152 and constitute the light chain of factor VII_a. The site of cleavage in factor VII by factor X_a is shown by an arrow, and the amino acids where this cleavage occurs are circled. The serine protease or catalytic domain of factor VII_a contains 254 residues, including the three principal amino acids participating in catalysis. These three amino acids [histidine (H)-41, aspartic acid (D)-90, and serine (S)-192] are also circled. Two potential carbohydrate binding sites are shown by solid diamonds. The tentative disulfide bonds in factor VII have been placed by analogy to those in bovine prothrombin and epidermal growth factor. γ , γ -carboxyglutamic acid; β , β -hydroxyaspartic acid. The first amino acids in the light and heavy chains of factor VII_a start with number 1 and differ from the numbers shown in Fig. 2.

mains is unknown. They were first noted in factor X (40, 41) and are also present in factor IX and protein C (36). The first potential growth factor domain contains the sequence of Gly-Ser-Cys-Lys-Asp-Gln-Leu (starting with glycine-59). This sequence is nearly identical to that in bovine factor VII in which the aspartic acid is present as β -hydroxyaspartic acid (3). Furthermore, β -hydroxyaspartic acid is present in essentially the same position in factor IX, factor X, and protein C (3, 42). Thus, it is extremely likely that human factor VII also contains β -hydroxyaspartic acid at residue 63 in the first potential growth factor domain.

The conversion of factor VII to factor VII_a is due to the cleavage of the bond between arginine-152 and isoleucine-153 by proteases such as factor X_a , factor XII_a, or thrombin. This results in a new amino-terminal sequence of Ile-Val-Gly-Gly-for the heavy chain of factor VII_a (22) (Fig. 3). The heavy chain (254 amino acids) is attached to the light chain (152 amino acids) by a disulfide bond. This bond has been tentatively placed between cysteine-135 in the light chain and cysteine-110 in the heavy chain by analogy to the other

vitamin K-dependent proteins. The heavy chain of factor VII_a also contains the three principal residues involved in the catalytic activity of this serine protease, including histidine-41, aspartic acid-90, and serine-192 (Fig. 3). These residues are analogous to histine-57, aspartic acid-102, and serine-195 in the active site of chymotrypsin (43). Also, aspartic acid-186 in the heavy chain of factor VII_a is probably located in the bottom of the substrate binding pocket of the enzyme. An aspartic acid residue in this position is characteristic of the trypsin family of serine proteases that are specific for the hydrolysis of peptide bonds containing a basic amino acid. This sequence in the active site of factor VII_a is also consistent with the cleavage of a specific bond between arginine and isoleucine in factor X and between arginine and alanine and between arginine and valine in factor IX by this enzyme.

To examine whether an activation peptide is released from the carboxyl terminus of the light chain during the activation reaction, factor VII_a was reduced and carboxymethylated, and the light and heavy chains were separated by HPLC. The light chain was subjected to digestion with chymotrypsin, and the fragments were separated by HPLC and subjected to amino acid sequence analysis. One fragment gave a sequence of Glu-Lys-Arg-Xaa-Ala-Ser-Lys-Pro-Gln-Gly-Arg. This sequence is identical to that predicted by the cDNA starting with glutamic acid-142 and originates from the carboxylterminal end of the light chain of factor VII_a. These data provide strong evidence for the conclusion that the cleavage of a single internal bond between arginine and isoleucine in factor VII leads to the activation of the zymogen. Whether additional cleavages occur in vivo, however, cannot be ruled out.

The amino acid composition of human factor VII circulating in plasma as predicted from the cDNA was determined as follows: Asp₁₉, Asn₁₃, β-OH Asp₁, Thr₂₂, Ser₂₉, Glu₁₇, Gln₂₀, Gla10, Pro21, Gly37, Ala21, Val27, Met4, Ile17, Leu38, Tyr12, Phe13, Lys17, His12, Arg24, 1/2Cys24, and Trp8. The molecular weight for the protein free of carbohydrate was calculated as 45,512. This calculation assumes the presence of 10 Gla residues and 1 β -hydroxyaspartic acid residue in factor VII, as discussed. Human factor VII contains two potential amino acid sequences for attachment of carbohydrate chains to asparagine residues. These sequences (Asn-Ala-Ser and Asn-Ile-Thr) occur at asparagine-145 in the light chain and asparagine-170 in the heavy chain (Fig. 3). Addition of two carbohydrate chains would increase the molecular weight for the glycoprotein to about 50,000. Human factor VII is highly homologous with the other vitamin K-dependent proteins, such as prothrombin (25% identity) (35), factor IX (40% identity) (38), factor X (40% identity) (44), and protein C (40% identity) (36, 37). This is consistent with the concept that this family of plasma proteins has evolved from a common ancestral gene during evolution (45).

The authors would like to thank Dr. Michael Parker for his assistance in the DNA sequence and computer analyses during the early phase of these studies, and Ila McCullough, Margo Rogers, and Lois Swenson for secretarial assistance. This research was supported by a contract between Novo Industries and ZymoGenetics, a research grant from the National Institutes of Health (HL 16919) to E.W.D., and a research grant from Blood Systems, Inc., to W.K.

- Davie, E. W., Fujikawa, K., Kurachi, K. & Kisiel, W. (1979) 1. Adv. Enzymol. 48, 277–318.
- DiScipio, R. G., Hermodson, M. A., Yates, S. G. & Davie, 2. E. W. (1977) Biochemistry 16, 698-706.
- McMullen, B. A., Fujikawa, K. & Kisiel, W. (1983) Biochem. 3. Biophys. Res. Commun. 115, 8-14.
- Kisiel, W. & Davie, E. W. (1975) Biochemistry 14, 4928-4934.
- Radcliffe, R. & Nemerson, Y. (1975) J. Biol. Chem. 250, 5. 388-395.
- Broze, G. J. & Majerus, P. W. (1980) J. Biol. Chem. 255, 6. 1242-1247.
- Bajaj, S. P., Rapaport, S. I. & Brown, S. F. (1981) J. Biol. 7. Chem. 256, 253-259.
- Radcliffe, R. & Nemerson, Y. (1976) J. Biol. Chem. 251, 8. 4797-4802.
- Kisiel, W., Fujikawa, K. & Davie, E. W. (1977) Biochemistry 9. 16, 4189-4194.
- 10. Radcliffe, R., Bagdasarian, S., Colman, R. & Nemerson, Y. (1977) Blood 50, 611-617
- 11. Seligsohn, U., Osterud, B., Brown, S. F., Griffin, J. H. & Rapaport, S. I. (1979) J. Clin. Invest. 64, 1056-1065.
- Osterud, B. & Rapaport, S. I. (1977) Proc. Natl. Acad. Sci. 12 USA 74, 5260–5264.
- Jesty, J. & Silverberg, S. A. (1979) J. Biol. Chem. 254, 13. 12337-12345.
- 14. Roberts, H. R. & Zeitler, K. D. (1982) in Hemostasis and Thrombosis, eds. Colman, R. W., Hirsh, J., Marder, V. J. & Salzman, E. W. (Lippincott, Philadelphia), pp. 134-137.

- Hedner, U. & Kisiel, W. (1983) J. Clin. Invest. 71, 1836-1841. 15.
- Anson, D. S., Austen, D. E. G. & Brownlee, G. G. (1985) 16. Nature (London) 315, 683-685.
- Busby, S., Kumar, A., Joseph, M., Halfpap, L., Insley, M., Berkner, K., Kurachi, K. & Woodbury, R. (1985) Nature (London) 316, 271-273.
- 18. De La Salle, H., Altenburger, W., Elkaim, R., Dott, K., Dieterle, A., Drillien, R., Cazenave, J.-P., Tolstoshev, P. & Lecocq, J.-P. (1985) Nature (London) 316, 268-270.
- Kwok, S. C. M., Ledley, F. D., Di Lella, A. G., Robson, 19. K. J. H. & Woo, S. L. C. (1985) Biochemistry 24, 556-561.
- Gubler, U. & Hoffman, B. J. (1983) Gene 25, 263-269. 20.
- Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, W. B., Coe, M. L., Knutson, G. J., Fass, D. N. & Hewick, R. M. (1984) Nature (London) 312, 342-347.
- Kisiel, W. & McMullen, B. A. (1981) Thromb. Res. 22, 22. 375-380.
- 23. Canfield, W. M. & Kisiel, W. (1982) J. Clin. Invest. 70, 1260-1272.
- Brown, J. P., Wright, P. W., Hart, G. E., Woodbury, R. G., 24 Hellstrom, K. E. & Hellstrom, I. (1980) J. Biol. Chem. 255, 4980-4983.
- Young, R. A. & Davis, R. W. (1983) Science 222, 778-782.
- Young, R. A. & Davis, R. W. (1983) Proc. Natl. Acad. Sci. 26. USA 80, 1194-1198
- Benton, W. D. & Davis, R. W. (1977) Science 196, 180-181. 27.
- 28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- Maxam, A. M. & Gilbert, W. (1980) Methods Enzymol. 65, 29. 499-560.
- Crestfield, A. M., Moore, S. & Stein, W. H. (1963) J. Biol. 30. Chem. 238, 622-627.
- Kurachi, K. & Davie, E. W. (1982) Proc. Natl. Acad. Sci. 31. USA 79, 6461-6464.
- McMullen, B. A., Fujikawa, K., Kisiel, W., Sasagawa, T., 32. Howald, W. N., Kwa, E. Y. & Weinstein, B. (1983) Biochemistry 22, 2875-2882.
- 33. Mott, J. E., Arsdell, J. V. & Platt, T. (1984) Nucleic Acids Res. 12, 4139-4152.
- Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., 34. Erickson, A. H. & Singappa, R. (1979) in Secretory Mechanisms, Society for Experimental Biology Symposium, eds. Hopkins, C. R. & Duncan, D. J. (Cambridge University Press, London), Vol. 33, pp. 9-36.
- 35. Degen, S. J. F., MacGillivray, R. T. A. & Davie, E. W. (1983) Biochemistry 22, 2087–2097. Foster, D. C., Yoshitake, S. & Davie, E. W. (1985) Proc.
- 36. Natl. Acad. Sci. USA 82, 4673-4677.
- 37. Beckmann, R. J., Schmidt, R. J., Santerre, R. F., Plutzky, J., Crabtree, G. R. & Long, G. L. (1985) Nucleic Acids Res. 13, 5233-5247.
- Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & 38 Kurachi, K. (1985) Biochemistry 24, 3736-3750.
- Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F. 39. Gould, J. A., Huddleston, J. A. & Brownlee, G. G. (1984) EMBO J. 3, 1053-1060.
- Dayhoff, M. O. (1978) Atlas of Protein Sequence and Struc-40. ture (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3.
- Doolittle, R. F., Feng, D. F. & Johnson, M. S. (1984) Nature 41. (London) 307, 558-560.
- 42. Drakenberg, T., Fernlund, P., Roepstorff, P. & Stenflo, J. (1983) Proc. Natl. Acad. Sci. USA 80, 1802-1806.
- 43. Sigler, P. B., Blow, D. M., Matthews, B. W. & Henderson, R. (1968) J. Mol. Biol. 35, 143-164.
- Leytus, S. P., Chung, D. W., Kisiel, W., Kurachi, K. & 44. Davie, E. W. (1984) Proc. Natl. Acad. Sci. USA 81, 3699-3702.
- Katayama, K., Ericsson, L. H., Enfield, D. L., Walsh, K. A., 45. Neurath, H., Davie, E. W. & Titani, K. (1979) Proc. Natl. Acad. Sci. USA 76, 4990-4994.