Human Plasma Prekallikrein, a Zymogen to a Serine Protease That Contains Four Tandem Repeats[†]

Dominic W. Chung, Kazuo Fujikawa, Brad A. McMullen, and Earl W. Davie*

Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received November 21, 1985; Revised Manuscript Received January 10, 1986

ABSTRACT: The amino acid sequence of human plasma prekallikrein was determined by a combination of automated Edman degradation and cDNA sequencing techniques. Human plasma prekallikrein was fragmented with cyanogen bromide, and 13 homogeneous peptides were isolated and sequenced. Cyanogen bromide peptides containing carbohydrate were further digested with trypsin, and the peptides containing carbohydrate were isolated and sequenced. Five asparagine-linked carbohydrate attachment sites were identified. The sequence determined by Edman degradation was aligned with the amino acid sequence predicted from cDNAs isolated from a λgt11 expression library. This library contained cDNA inserts prepared from human liver poly(A) RNA. Analysis of the cDNA indicated that human plasma prekallikrein is synthesized as a precursor with a signal peptide of 19 amino acids. The mature form of the protein that circulates in blood is a single-chain polypeptide of 619 amino acids. Plasma prekallikrein is converted to plasma kallikrein by factor XII$_a$ by the cleavage of an internal Arg–Ile bond. Plasma kallikrein is composed of a heavy chain (371 amino acids) and a light chain (248 amino acids), and these 2 chains are held together by a disulfide bond. The heavy chain of plasma kallikrein originates from the amino-terminal end of the zymogen and is composed of 4 tandem repeats that are 90 or 91 amino acid residues in length. These repeat sequences are also homologous to those in human factor XI. The light chain of plasma kallikrein contains the catalytic portion of the enzyme and is homologous to the trypsin family of serine proteases.

Plasma prekallikrein is a glycoprotein that participates in the surface-dependent activation of blood coagulation, fibrinolysis, kinin generation, and inflammation. The contact activation reactions are initiated when plasma is exposed to negatively charged surfaces (Ratnoff, 1966; Margolis, 1958; Niewiarowski & Prou-Wartelle, 1959). These reactions involve factor XII, plasma prekallikrein, factor XI, and high molecular weight kininogen. Contact activation is initiated by the binding of factor XII to a negatively charged surface (Revak et al., 1974; Fujikawa et al., 1980; Bock et al., 1981; Tans & Griffin, 1982), which leads to the activation of plasma prekallikrein. Plasma kallikrein then activates factor XII in a reciprocal reaction (Revak et al., 1977; Cochrane et al., 1973; Griffin & Cochrane, 1979). This amplifies the activation of the surface-dependent coagulation reactions (Heimark et al., 1980; Kurachi & Davie, 1977; Bouma & Griffin, 1977; Mannhalter et al., 1980). The rates of these activation reactions are greatly enhanced by the presence of high molecular weight kininogen, which participates as a cofactor in these reactions (Meier et al., 1977; Griffin & Cochrane, 1976). α-Factor XII$_a$ then converts factor XI to factor XI$_a$. Plasma kallikrein can also liberate the vasoactive peptide bradykinin from high molecular weight kininogen (Nagasawa & Nakayasu, 1973; Habal et al., 1974). In addition, plasma kallikrein has been shown to convert prorenin to renin in vitro and could play a role in the renin–angiotensin system (Sealey et al., 1979).

Plasma prekallikrein is synthesized in the liver and secreted into the blood as a single polypeptide chain with an apparent molecular weight of 88 000 (Mandle & Kaplan, 1977; Bouma et al., 1980; Heimark & Davie, 1981). It is present in plasma

at a concentration of 35–45 μg/mL and circulates as a noncovalent complex with high molecular weight kininogen (Mandle et al., 1976). Plasma prekallikrein is converted to plasma kallikrein by factor XII$_a$ by the cleavage of an internal peptide bond in the single-chain precursor molecule. This results in the formation of a serine protease composed of a heavy chain and a light chain, and these two chains are held together by a disulfide bond(s). The active site or catalytic domain of plasma kallikrein is associated with the light chain of the molecule (Mandle & Kaplan, 1977). The heavy chain contains the binding site for high molecular weight kininogen and is required for the surface-dependent procoagulant activity of plasma kallikrein (van der Graaf et al., 1982). Plasma kallikrein differs from tissue kallikrein in its site of synthesis, molecular weight, structure, and substrate specificity and is encoded by a gene that is different from those of tissue kallikreins (Shine et al., 1983).

The primary structure of two of the proteins involved in contact activation (factor XII and high molecular weight kininogen) has been determined (Fujikawa & McMullen, 1983; McMullen & Fujikawa, 1985; Takagaki et al., 1985). In this paper, the primary structure of plasma prekallikrein, determined by a combination of protein sequencing and cDNA sequencing techniques, is reported. In the following paper (Fujikawa et al., 1986), the primary sequence of human factor XI is also reported. This protein shows considerable sequence homology with plasma prekallikrein.

EXPERIMENTAL PROCEDURES

Preparation of Human Plasma Prekallikrein. Plasma prekallikrein was purified from 4.5 L of fresh frozen human plasma. The initial steps in the purification procedure were the same as those previously described up to the carboxymethyl-Sephadex (CM-Sephadex)[1] step (Heimark & Davie,

---

1981). Fractions containing plasma prekallikrein activity from the CM-Sephadex column were pooled and dialyzed against 20 L of 0.05 M Tris-HCl buffer, pH 7.5, containing 0.15 M NaCl. Diisopropyl fluorophosphate (DFP) was added to the dialyzed sample to a final concentration of 0.1 mM, and the sample was then applied to a column of high molecular weight kininogen–Sepharose (2.5 × 10 cm). The column was first washed with 200 mL of 0.05 M Tris-HCl buffer, pH 7.5, containing 0.15 M NaCl followed by 250 mL of 0.5 M NaCl in 0.05 M Tris-HCl buffer, pH 7.5, containing 0.1 mM DFP. Bound protein was eluted with 0.5 M guanidine hydrochloride in 0.05 M Tris-HCl buffer, pH 7.5. Eluted fractions were pooled and immediately dialyzed against 0.02 M sodium acetate buffer, pH 5.2, containing 0.15 M NaCl. The dialyzed sample was concentrated to 3 mL by ultrafiltration with the use of an Amicon concentrator containing a PM-10 membrane (Amicon) and applied to a column of Sephadex G-150 (2.5 × 95 cm) previously equilibrated with 0.05 M sodium acetate buffer, pH 5.2. Fractions containing plasma prekallikrein were pooled and concentrated as described previously. The final preparation of plasma prekallikrein showed a doublet corresponding to molecular weights of 81 000 and 76 000 on an SDS–polyacrylamide gel (Weber & Osborn 1969). After reduction, a major band of M_r 85 000, with minor bands corresponding to the heavy chain (M_r 52 000), and light chains (M_r 44 000 and 40 000) of plasma kallikrein were observed. This indicated that partial activation of prekallikrein occurred during the isolation procedure. The extent of activation varied, however, from one preparation to another. Approximately 10–15 mg of plasma prekallikrein was obtained from 4.5 L of plasma by using this procedure.

Plasma prekallikrein was assayed by hydrolysis of the chromogenic substrate benzoyl-Pro-Phe-Arg-p-nitroanilide (Vega Chemical) after conversion of plasma prekallikrein to plasma kallikrein by β-factor XII$_a$. Human plasma prekallikrein–Sepharose was prepared by coupling 6 mg of purified protein to 1 g of activated CH-Sepharose (Pharmacia).

Preparation of High Molecular Weight Kininogen–Sepharose. Human high molecular weight kininogen was purified by the method of Kato et al. (1981). Two hundred milligrams of purified high molecular weight kininogen was coupled to 6 g of activated CH-Sepharose according to the manufacturer's instructions.

Preparation of Affinity-Purified Antibody to Human Plasma Prekallikrein. Antibody against human plasma prekallikrein was raised in rabbits by repeated injections of 0.5 mg of purified human plasma prekallikrein. Immunoglobulin from the immune serum was prepared according to Harboe and Ingild (1973). The immunoglobulin fraction (60 mg of protein) was dialyzed against 0.05 M Tris-HCl buffer, pH 7.5, containing 0.15 M NaCl and was applied to a small column of human plasma prekallikrein–Sepharose (4 mL) that had been previously equilbrated with the same buffer. After the column was washed extensively with 1 L of the equilibration buffer, the adsorbed antibody was eluted with 0.2 M glycine hydrochloride, pH 2.5. Aliquots of the eluate (1 mL) were collected and immediately neutralized with 1 M Tris-HCl buffer, pH 8.0. The pooled antibody was then dialyzed and concentrated simultaneously against 0.05 M Tris-HCl buffer, pH 7.5, containing 0.15 M NaCl by using a MicroProDiCon concentrator. Approximately 0.5 mg of affinity-purified an-

tibody to plasma prekallikrein was obtained. The affinity-purified antibody was radiolabeled with 0.2 mCi of Na^{125}I (Amersham) to a specific activity of 1 × 10^8 cpm/μg using Iodogen (Pierce).

Cleavage and Fractionation of Peptides from Human Plasma Prekallikrein. Seven milligrams of human plasma prekallikrein was reduced by dithiothreitol in the presence of 6 M guanidine hydrochloride and carboxymethylated by iodoacetate according to the method of Crestfield et al. (1963), as modified by McMullen and Fujikawa (1985). Cleavage by cyanogen bromide was performed as described by Titani et al. (1972). The peptides from the cyanogen bromide digest were separated on a Sephadex G-50 superfine column (1.6 × 95 cm) previously equilibrated with 5% HCOOH. Pooled peptide fractions from the Sephadex G-50 column were further purified by reverse-phase chromatography in a Waters HPLC system using an Altex Ultrapore C 3 column and a trifluoroacetic acid–acetonitrile solvent system as previously described (McMullen & Fujikawa, 1985). Additional purification of selected samples was performed on a Hamilton PRP-1 column with an ammonium bicarbonate–acetonitrile solvent gradient as described (McMullen & Fujikawa, 1985).

Trypsin Cleavage of Cyanogen Bromide Peptides. Cyanogen bromide peptides that contain carbohydrate were further digested with trypsin in 0.1 M NH$_4$HCO$_3$ at a mass ratio of 1:100. The digested peptides were separated by HPLC as described previously.

Analytical Methods. Homogeneous peptides were sequenced in a Beckman 890C sequencer as previously described (McMullen & Fujikawa, 1985; Edman & Begg, Brauer et al., 1975). PTH derivatives were identified by chromatography in two complementary reverse-phase HPLC systems. The lower limit of detection in these systems is approximately 20 pmol (Bridgen et al., 1976; Ericsson et al, 1977). Amino acid compositions of selected peptides were determined on a Dionex D 500 amino acid analyzer. The samples were hydrolyzed at 110 °C for 24 h in sealed, evacuated tubes containing 6 N HCl.

Screening of λgt11 Library. cDNAs coding for human plasma prekallikrein were isolated from a λgt11 expression library containing cDNA inserts prepared from human liver poly(A) RNA (Kwok et al., 1985). Recombinant plaques expressing fusion proteins with antigenic determinants of human plasma prekallikrein were identified with radiolabeled affinity-purified antibodies to plasma prekallikrein according to the method of Young and Davis (1983a,b), as modified by Foster and Davie (1984). The λgt11 library was also screened by the plaque hybridization technique of Benton and Davis (1977), using a cDNA for human plasma prekallikrein (λHPK19) labeled to high specific activity by nick translation (Maniatis et al., 1975). Positive clones were isolated and plaque purified, and phage DNA was prepared as described by Chung et al., (1983). cDNA inserts were released from recombinant phage DNA by digestion with the restriction enzyme *EcoRI* (Bethesda Research Laboratories), and the cDNA inserts were then subcloned into the *EcoRI* site of plasmid pBR322.

DNA Sequence Analysis. The nucleotide sequence of the cDNA was determined by the dideoxy chain terminator method of Sanger et al. (1977), using the M13mp18 and M13mp19 cloning and sequencing system (Messing et al., 1981; Norrander et al., 1983). Sequencing reactions were carried out in the presence of [^{35}S]dATPαS (Amersham). The reaction mixtures were electrophoresed in 6% denaturing polyacrylamide gels containing a buffer gradient as described

---

Table I: Amino Acid Sequences of Cyanogen Bromide Peptides of Plasma Prekallikrein

CNB1	G C L T Q L Y E N A F F R G G D V A
CNB2	Y T P N A Q Y C Q
CNB3	R C T F H P R C L L F S F L P A S S I N
CNB4	E K R F G C F L K D S V T G T L P K V H R T G A V S G H S L K Q C G H Q I S A C H R D I Y
CNB5	R G V N F - V S K V S S V E E C Q K R C T N/S N I R C Q F F S Y A T Q T F H K A E Y R N N C L L K Y S P G G T P T A I K V L S N V E
CNB6	N I F Q H L A F S D V D V A R V L T P D A F V C R T I C T Y H P N C L F F T F Y T N V W K I E S Q R N
CNB7	I R C Q F F T Y S L L P E D C K E E K C K C F L R L
CNB8	D G S P T R I A Y G T Q G S S G Y S L R L C N T G D N S V C T T K T S T R
CNB9	I V G G T - S S W G E W P W Q V S L Q V K L T A Q R H L C G G S L I G H Q W V L T A A
CNB10	V C A G Y K E G G K D A C K G D S G G P L V C K H N
CNB11	W R L V G I T S W G E G C A R R E Q P G V Y T K V A E Y
CNB12	D W I L E K T Q S S D G K A
CNB13	Q S

by Biggin et al. (1983). Nonrandom DNA sequencing by sequential *Bal*31 deletion was performed as described by Poncz et al. (1982) and modified by Yoshitake et al. (1985).

RESULTS

Amino Acid Sequence of Cyanogen Bromide Fragments. Human plasma prekallikrein containing some plasma kallikrein was reduced, carboxymethylated, and digested with cyanogen bromide. The resulting peptides were fractionated according to size by gel filtration on a Sephadex G-50 superfine column (Figure 1). Pooled fractions were further fractionated by reverse-phase chromatography in an HPLC system, and 13 homogeneous peptides (designated CNB1–CNB13) were isolated. These peptides were sequenced by automated Edman degradation (Table I). The sequence of CNB1 was identical with the amino-terminal sequence of human plasma prekallikrein previously reported by Heimark and Davie (1981). The sequence of CNB9 indicated that this peptide was derived from the amino terminus of the light chain of plasma kallikrein. Its sequence differed from a previously reported sequence for the light chain in one position (residue 7), where a serine instead of an alanine was identified. The presence of a serine in this position was confirmed by cDNA sequences (see below). A dimorphic site was identified in the sequence of CNB5 (position 22) where both asparagine and serine, in a ratio of 7:3, were found.

Carbohydrate Attachment Sites. In sequencing the cyanogen bromide peptides, we did not detect any PTH-amino acid derivatives in the sixth cycles of CNB5 and CNB9 (Table I). Both of these residues were followed by the sequence of X-Ser, which suggested that they were asparagine residues with carbohydrate chains attached. In order to determine glycosylation of the asparagine residues in these peptides, and to locate other carbohydrate attachment sites, the cyanogen bromide peptides CNB5, CNB6, and CNB9 were further digested with trypsin. These peptides were shown to contain hexosamine by amino acid analysis. The resulting tryptic peptides were separated by HPLC, and five peptides containing hexosamine were isolated. The sequence of each of these
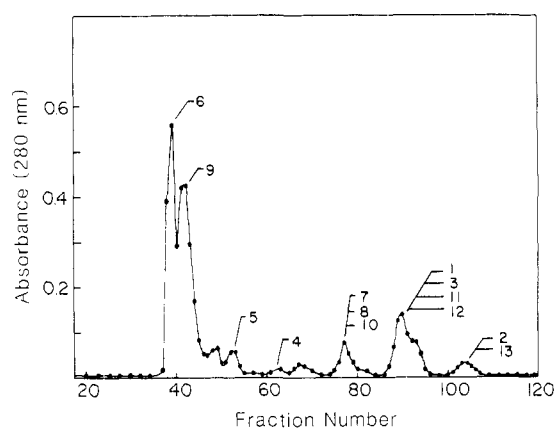
FIGURE 1: Fractionation of the cyanogen bromide peptides of human plasma prekallikrein by gel filtration. A mixture of plasma prekallikrein and plasma kallikrein (7 mg) was digested with cyanogen bromide and lyophilized. The peptides were dissolved in 1 mL of 5% formic acid and applied to a Sephadex G-50 superfine column (1.6 × 95 cm). Fractions (2 mL) were collected, and the absorbance at 280 nm was determined.

peptides was determined (Table II). Glycosylated asparagine residues were assigned to the positions indicated for the following reasons. First, the amino acid composition of each of these peptides indicated the presence of one additional Asp residue that was not accounted for in the sequence analysis. Second, hexosamine was present in each of these peptides. Third, in all cases, the potential attachment site was followed by the sequence of X-(T/S). Finally, the cDNA sequence as reported below agrees with these asparagine assignments. Accordingly, five Asn-linked carbohydrate attachment sites have been identified including three in the light chain and two in the heavy chain of plasma kallikrein. In the analysis of these peptides, no O-link carbohydrate attachment sites were encountered. All together, the sequence of 434 amino acid residues was determined in 16 nonoverlapping regions of the molecule. This corresponds to 70% of the entire protein sequence for human plasma prekallikrein.

Human Plasma Prekallikrein cDNAs. A λgt11 expression

Table II: Amino Acid Sequence of Carbohydrate-Containing Tryptic Peptides

			CHO
T1	from CNB5		G V N F(N) V S K
			CHO
T2	from CNB6		I Y P G V D F G G E E L(N) V T F V K
			CHO
T3	from CNB9		I V G G T(N)S S W G E W P W Q V S L Q V K
			CHO
T4	from CNB9		I Y S G I L(N)L S D I T K
			CHO
T5	from CNB9		L Q A P L(N)Y T E F Q K P I C L P S K

library containing human liver cDNA inserts (Kwok et al., 1985) was screened for human plasma prekallikrein by the immunological screening technique of Young and Davis (1983a,b). In these studies, affinity-purified rabbit antibody to human plasma prekallikrein was labeled with ^{125}I and used to identify recombinant phage expressing a fusion protein of β-galactosidase and human plasma prekallikrein. Twelve positives were initially identified from two million recombinants. On subsequent plaque purification, however, most of these recombinants stopped expressing the fusion product. Of these 12 recombinants, 1 (λHPK19) continued to express a fusion protein at low levels in about 10% of the phage progeny. This recombinant phage contained an insert of 700 base pairs, and its nucleotide sequence was determined. An analysis of the open reading frame showed that this clone encoded sequences identical with those of cyanogen bromide peptides CNB10, CNB11, and CNB12, and unequivocally identified it as a partial cDNA clone for human plasma prekallikrein. This insert was then used as a hybridization probe to rescreen the initial positive clones as well as the entire cDNA library. This screening identified all plasma prekallikrein clones, including those that were not inserted in the correct orientation and reading frame, for expression of a fusion protein containing plasma prekallikrein antigenic determinants. A total of 36 clones were isolated and plaque purified. A partial restriction map and sequence strategy for λHPK19 and the two longest cDNA inserts, λHPK31 and λHPK129, are shown in Figure 2. λHPK129 contains an insert of 2277 nucleotides. The complete nucleotide sequence and the predicted amino acid sequence derived from the cDNA insert in λHPK129 are shown in Figure 3. This clone contained 93 nucleotides of 5′ noncoding sequence followed by 57 nucleotides coding for a signal peptide of 19 amino acids and 1857 nucleotides coding for the mature plasma prekallikrein molecule of 619 amino acids. The 3′ noncoding region was 267 nucleotides in length and contained the polyadenylation signal AATAAA sequence (Proudfoot & Brownlee, 1976) located 13 nucleotides upstream from the poly(A) tail. A stop codon of TAA was present at nucleotides 49–51 in the cDNA sequence, indicating that the methionine codon beginning at nucleotide residue 94 was the correct initiation codon. Accordingly, the signal peptide for plasma prekallikrein is 19 amino acids in length and contains the typical hydrophobic residues which are involved in translocation and secretion of proteins across the rough endoplasmic reticulum. Sequences determined by Edman degradation of the cyanogen bromide and tryptic peptides are shown by the overlines in Figure 3. The five Asn-linked carbohydrate attachment sites are indicated in Figure 3 by the solid diamonds. Protein sequence analysis indicated that amino acid residue 124 was dimorphic, with asparagine and serine occurring at a ratio of 7:3. The cDNA inserts in both
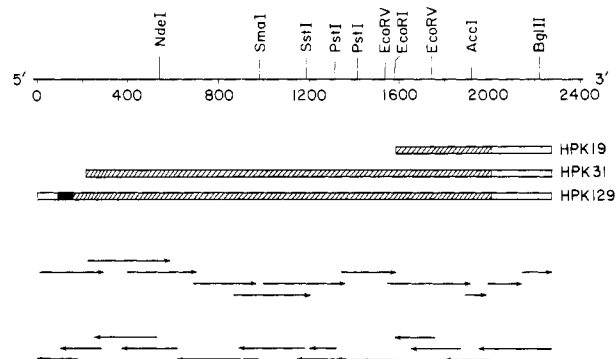
FIGURE 2: Restriction map and sequencing strategy of cDNAs for human plasma prekallikrein. The 5′ and 3′ noncoding sequences are shown by open bars and coding regions by slashed bars. The signal peptide is represented by a solid bar. The arrows indicate the direction and extent of the sequences that were determined.

λHPK31 and λHPK129 coded for the more prevalent asparaginyl form. The nucleotide sequence of λHPK31 contained a T instead of a C at nucleotide 722. This changes amino acid residue 191 from an alanine to valine. λHPK19 contained a C instead of a T at nucleotide 1598, which changes Ile-483 to threonine. These changes, however, have not been observed in other clones that were analyzed and are probably cloning artifacts rather than true polymorphic differences. On the basis of the predicted amino acid sequence from λHPK129, the amino acid composition for plasma prekallikrein was calculated as follows: Asp_{23}, Asn_{29}, Thr_{48}, Ser_{51}, Glu_{30}, Gln_{32}, Pro_{26}, Gly_{50}, Ala_{25}, Val_{37}, Met_{11}, Ile_{32}, Leu_{46}, Tyr_{23}, Phe_{27}, Lys_{40}, His_{16}, Arg_{26}, Trp_{11}, and $^1/_2$-Cys_{36}. This corresponds to a molecular weight of 69 170 without carbohydrate and 79 545 with the addition of 15% carbohydrate.

The cleavage of plasma prekallikrein by factor XII_a involves limited proteolysis. During the activation reaction, the single-chain prekallikrein molecule is converted into a two-chain form containing a heavy chain and a light chain, and these two chains are linked by a disulfide bond(s). The light chain, or catalytic chain, begins with the sequence of Ile-Val-Gly-Gly-Thr-Asn-Ser-Ser-Trp (Heimark & Davie, 1981). This corresponds to cleavage of an Arg–Ile bond between residues 371 and 372 of the plasma prekallikrein molecule (shown by an arrowhead in Figure 3). Accordingly, the heavy chain of plasma kallikrein contains 371 amino acid residues and is derived from the amino terminus of the zymogen. The light chain contains 248 amino acids residues and is derived from the carboxyl terminus of the zymogen. The light-chain also contains the catalytic domain including the reactive site triad of His-415, Asp-464, and Ser-559. By analogy to the disulfide structure of the light chain (B chain) of plasmin, it is highly probable that Cy-484 in the light chain of plasma kallikrein

```
                                                     5'   GGT AGT AGC AAA TAT TCA AAT GAG AAC AGC TTG AAG ACC GTT
                                                          1                                      30

                                                  -19              Met Ile Leu Phe Lys Gln Ala Thr Tyr  -10 Ile Ser Leu Phe Ala Thr Val Ser  -1
        ***                                                        Met Ile                                                                   Cys
CAT TTT TAA GTG ACA AGA GAC TCA CCT CCA AGA AGC AAT TGT GTT TTC AGA ATG ATT TTA TTC AAG CAA GCA ACT TAT TTC ATT TCC TTG TTT GCT ACA GTT TCC TGT
            60                                  90                                  120                                         150

 1  Gly Cys Leu Thr Gln Leu Tyr Glu Asn Ala Phe Phe Arg Gly Gly Asp Val Ala Ser Met Tyr Thr Pro Asn Ala Gln Tyr Cys Gln Met Arg Cys Thr Phe His Pro
    GGA TGT CTG ACT CAA CTC TAT GAA AAC GCC TTC TTC AGA GGT GGG GAT GTA GCT TCC ATG TAC ACC CCA AAT GCC CAA TAC TGC CAG ATG AGG TGC ACA TTC CAC CCA
                        10                          180                  20                 210                   30                 240

    Arg Cys Leu Leu Phe Ser Phe Leu Pro Ala Ser Ser Ile Asn Asp Met Glu Lys Arg Phe Gly Cys Phe Leu Lys Asp Ser Val Thr Gly Thr Leu Pro Lys Val His
    AGG TGT TTG CTA TTC AGT TTT CTT CCA GCA AGT TCA ATC AAT GAC ATG GAG AAA AGG TTT GGT TGC TTC TTG AAA GAT AGT GTT ACA GGA ACC CTG CCA AAA GTA CAT
            40          270             50          300                 60          330                 70          360

    Arg Thr Gly Ala Val Ser Gly His Ser Leu Lys Gln Cys Gly His Gln Ile Ser Ala Cys His Arg Asp Ile Tyr Lys Gly Val Asp Met Arg Gly Val Asn Phe Asn
    CGA ACA GGT GCA GTT TCT GGA CAT TCC TTG AAG CAA TGT GGT CAT CAA ATA AGT GCT TGC CAT CGA GAC ATT TAT AAA GGA GTT GAT ATG AGA GGA GTC AAT TTT AAT
            80          390                 90          420                 100         450

    Val Ser Lys Val Ser Ser Val Glu Glu Cys Gln Lys Arg Cys Thr Asn Asn Ile Arg Cys Gln Phe Phe Ser Tyr Ala Thr Gln Thr Phe His Lys Ala Glu Tyr Arg
    GTG TCT AAG GTT AGC AGT GTT GAA GAA TGC CAA AAA AGG TGC ACC AAT AAC ATT CGC TGC CAG TTT TTT TCA TAT GCC ACG CAA ACA TTT CAC AAG GCA GAG TAC CGG
        110             480             120         510  Ser        130  540             140  570

    Asn Asn Cys Leu Leu Lys Tyr Ser Pro Gly Gly Thr Pro Thr Ala Ile Lys Val Leu Ser Asn Val Glu Ser Gly Phe Ser Leu Lys Pro Cys Ala Leu Ser Glu Ile
    AAC AAT TGC CTA TTA AAG TAC AGT CCC GGA GGA ACA CCT ACC GCT ATA AAG GTG CTG AGT AAC GTG GAA TCT GGA TTC TCA CTG AAG CCC TGT GCC CTT TCA GAA ATT
            150         600                 160         630                 170         660                 180  690

    Gly Cys His Met Asn Ile Phe Gln His Leu Ala Phe Ser Asp Val Asp Val Ala Arg Val Leu Thr Pro Asp Ala Phe Val Cys Arg Thr Ile Cys Thr Tyr His Pro
    GGT TGC CAC ATG AAC ATC TTC CAG CAT CTT GCG TTC TCA GAT GTG GAT GTT GCC AGG GTT CTC ACT CCA GAT GCT TTT GTG TGT CGG ACC ATC TGC ACC TAT CAC CCC
                    190  720                200         750                 210  780

    Asn Cys Leu Phe Phe Thr Phe Tyr Thr Asn Val Trp Lys Ile Glu Ser Gln Arg Asn Val Cys Leu Leu Lys Thr Ser Glu Ser Gly Thr Pro Ser Ser Ser Thr Pro
    AAC TGC CTC TTC TTT ACA TTC TAT ACA AAT GTA TGG AAA ATC GAG TCA CAA AGA AAT GTT TGT CTT CTT AAA ACA TCT GAA AGT GGC ACA CCA AGT TCC TCT ACT CCT
            220         810                 230         840                 240  870             900

    Gln Glu Asn Thr Ile Ser Gly Tyr Ser Leu Leu Thr Cys Lys Arg Thr Leu Pro Glu Pro Cys His Ser Lys Ile Tyr Pro Gly Val Asp Phe Gly Gly Glu Glu Leu
    CAA GAA AAC ACC ATA TCT GGA TAT AGC CTT TTA ACC TGC AAA AGA ACT TTA CCT GAA CCC TGC CAT TCT AAA ATT TAC CCG GGA GTT GAC TTT GGA GGA GAA GAA TTG
                260     930                 270         960                 280  990

● 290 Asn Val Thr Phe Val Lys Gly Val Asn Val Cys Gln Glu Thr Cys Thr Lys Met Ile Arg Cys Gln Phe Phe Thr Tyr Ser Leu Leu Pro Glu Asp Cys Lys Glu Glu
    AAT GTG ACT TTT GTT AAA GGA GTG AAT GTT TGC CAA GAG ACT TGC ACA AAG ATG ATT CGC TGT CAG TTT TTC ACT TAT TCT TTA CTC CCA GAA GAC TGT AAG GAA GAG
    1020                    300  1050                310  1080                320  1110

    Lys Cys Lys Cys Phe Leu Arg Leu Ser Met Asp Gly Ser Pro Thr Arg Ile Ala Tyr Gly Thr Gln Gly Ser Ser Gly Tyr Ser Leu Arg Leu Cys Asn Thr Gly Asp
    AAG TGT AAG TGT TTC TTA AGA TTA TCT ATG GAT GGT TCT CCA ACT AGG ATT GCG TAT GGG ACA CAA GGG AGC TCT GGT TAC TCT TTG AGA TTG TGT AAC ACT GGG GAC
            330  1140                340  1170                350  1200                360  1230

    Asn Ser Val Cys Thr Thr Lys Thr Ser Thr Arg Ile Val Gly Gly Thr Asn Ser Ser Trp Gly Glu Trp Pro Trp Gln Val Ser Leu Gln Val Lys Leu Thr Ala Gln
    AAC TCT GTC TGC ACA ACA AAA ACA AGC ACA CGC ATT GTT GGA GGA ACA AAC TCT TCT TGG GGA GAG TGG CCC TGG CAG GTG AGC CTG CAG GTG AAG CTG ACA GCT CAG
                    370  1260                380  1290                390  1320

    Arg His Leu Cys Gly Gly Ser Leu Ile Gly His Gln Trp Val Leu Thr Ala Ala His Cys Phe Asp Gly Leu Pro Leu Gln Asp Val Trp Arg Ile Tyr Ser Gly Ile
    AGG CAC CTG TGT GGA GGG TCA CTC ATA GGA CAC CAG TGG GTC CTC ACT GCT GCC CAC TGC TTT GAT GGG CTT CCC CTG CAG GAT GTT TGG CGC ATC TAT AGT GGC ATT
            400  1350                410  1380                420  1410                430  1440

    Leu Asn Leu Ser Asp Ile Thr Lys Asp Thr Pro Phe Ser Gln Ile Lys Glu Ile Ile Ile His Gln Asn Tyr Lys Val Ser Glu Gly Asn His Asp Ile Ala Leu Ile
    TTA AAT CTG TCA GAC ATT ACA AAA GAT ACA CCT TTC TCA CAA ATA AAA GAG ATT ATT ATT CAC CAA AAC TAT AAA GTC TCA GAA GGG AAT CAT GAT ATC GCC TTG ATA
                440  1470                450  1500                460  1530

    Lys Leu Gln Ala Pro Leu Asn Tyr Thr Glu Phe Gln Lys Pro Ile Cys Leu Pro Ser Lys Gly Asp Thr Ser Thr Ile Tyr Thr Asn Cys Trp Val Thr Gly Trp Gly
    AAA CTC CAG GCT CCT TTG AAT TAC ACT GAA TTC CAA AAA CCA ATA TGC CTA CCT TCC AAA GGT GAC ACA AGC ACA ATT TAT ACC AAC TGT TGG GTA ACC GGA TGG GGC
    470             1560                480  1590                490  1620                500  1650

    Phe Ser Lys Glu Lys Gly Glu Ile Gln Asn Ile Leu Gln Lys Val Asn Ile Pro Leu Val Thr Asn Glu Glu Cys Gln Lys Arg Tyr Gln Asp Tyr Lys Ile Thr Gln
    TTC TCG AAG GAG AAA GGT GAA ATC CAA AAT ATT CTA CAA AAG GTG AAT ATT CCT TTG GTA ACA AAT GAA GAA TGC CAG AAA AGA TAT CAA GAT TAT AAA ATA ACC CAA
                510  1680                520  1710                530  1740                540  1770

    Arg Met Val Cys Ala Gly Tyr Lys Glu Gly Gly Lys Asp Ala Cys Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Lys His Asn Gly Met Trp Arg Leu Val Gly Ile
    CGG ATG GTC TGT GCT GGC TAT AAA GAA GGG GGA AAA GAT GCT TGT AAG GGA GAT TCA GGT GGT CCC TTA GTT TGC AAA CAC AAC GGA ATG TGG CGT TTG GTG GGC ATC
                550  1800                560  1830                570  1860

    Thr Ser Trp Gly Glu Gly Cys Ala Arg Arg Glu Gln Pro Gly Val Tyr Thr Lys Val Ala Glu Tyr Met Asp Trp Ile Leu Glu Lys Thr Gln Ser Ser Asp Gly Lys
    ACA AGC TGG GGT GAA GGC TGT GCC CGC AGG GAG CAA CCT GGT GTC TAC ACC AAA GTC GCT GAG TAC ATG GAC TGG ATT TTA GAG AAA ACA CAG AGC AGT GAT GGA AAA
                580  1890                590  1920                600  1950                610  1980

    Ala Gln Met Gln Ser Pro Ala STOP
                        619
    GCT CAG ATG CAG TCA CCA GCA TGA GAA GCA GTC CAG AGT CTA GGC AAT TTT TAC AAC CTG AGT TCA AGT CAA ATT CTG AGC CTG GGG GGT CCT CAT CTG CAA AGC ATG
                        2010                            2040                            2070

    GAG AGT GGC ATC TTC TTT GCA TCC TAA GGA CGA AAG ACA CAG TGC ACT CAG AGC TGC TGA GGA CAA TGT CTG CTG AAG CCC GCT TTC AGC ACG CCG TAA CCA GGG GCT
                            2100                            2130                            2160                            2190

    GAC AAT GCG AGG TCG CAA CTG AGA TCT CCA TGA CTG TGT GTT GTG [AAA TAA A]AT GGT GAA AGA TCA AAA AAA AAA    3'
                            2220                            2250            2277
```
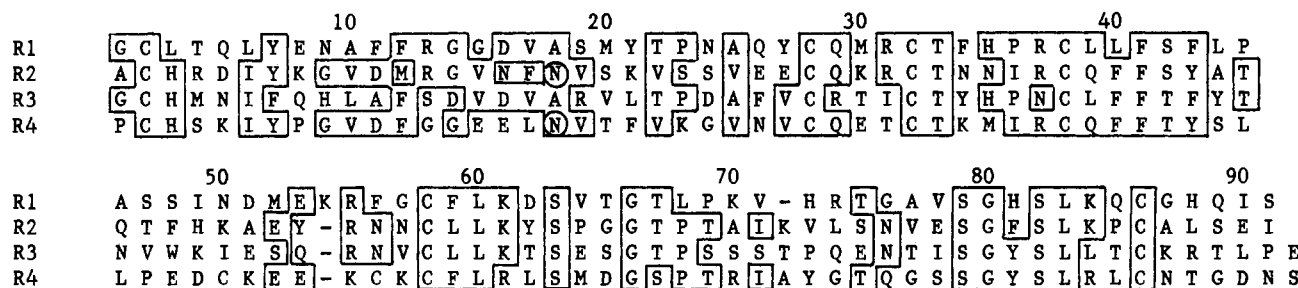
FIGURE 3: Nucleotide sequence of the cDNA insert of λPK129 coding for human plasma prekallikrein. The predicted amino acid sequence is shown above the DNA sequence. Carbohydrate attachment sites are indicated by solid diamonds. The solid arrowhead indicates the site of cleavage by factor XII$_a$ during the conversion of plasma prekallikrein to plasma kallikrein. The polyadenylation or processing sequence is boxed. Protein sequences determined by Edman degradation are overlined.

forms a disulfide bond with Cys-364 in the heavy chain.

The heavy chain of plasma kallikrein contains 4 tandem repeats of 90 (or 91) amino acids which are followed by a short connecting region of 9 amino acids. Figure 4 shows an alignment of the four tandem repeats where two or more identical amino acids occurring in corresponding positions are shown in boxes. Each of the four repeats is characterized by the presence of six conserved $^1/_2$-Cys residues. The fourth repeat contains two additional $^1/_2$-Cys residues which presumably form an additional disulfide bond within this repeat.

The two carbohydrate chains in the heavy chain are located in the second and fourth repeats where they are linked to asparagine residues in homologous positions within the repeat sequences (shown in Figure 4 by the circles). An analysis of these repeat sequences in plasma prekallikrein shows that no similar sequences occur in any other known proteins except human factor XI (Fujikawa et al., 1986).

DISCUSSION

The serine proteases involved in blood coagulation and fi-

```
                    10            20              30            40
R1   G C L T Q L Y E N A F F R G G D V A S M Y T P N A Q Y C Q M R C T F H P R C L L F S F L P
R2   A C H R D I Y K G V D M R G V N F N V S K V S S V E E C Q K R C T N N I R C Q F F S Y A T
R3   G C H M N I F Q H L A F S D V D V A R V L T P D A F V C R T I C T Y H P N C L F F T F Y T
R4   P C H S K I Y P G V D F G G E E L N V T F V K G V N V C Q E T C T K M I R C Q F F T Y S L

              50              60              70            80              90
R1   A S S I N D M E K R F G C F L K D S V T G T L P K V - H R T G A V S G H S L K Q C G H Q I S
R2   Q T F H K A E Y - R N N C L L K Y S P G G T P T A I K V L S N V E S G F S L K P C A L S E I
R3   N V W K I E S Q - R N V C L L K T S E S G T P S S S T P Q E N T I S G Y S L L T C K R T L P E
R4   L P E D C K E E - K C K C F L R L S M D G S P T R I A Y G T Q G S S G Y S L R L C N T G D N S
```

FIGURE 4: Alignment of the four tandem repeat sequences (R1–R4) present in human plasma prekallikrein. Two or more residues in homologous positions are boxed. R1 represents amino acid residues 1–90; R2, 91–180; R3, 181–271; R4, 272–362. Asparagine residues that are glycosylated are circled.

brinolysis differ from the digestive proteases in that they have large noncatalytic segments present in the amino-terminal regions of their molecules. The noncatalytic segments serve to mediate the binding of these proteases or their zymogens to other proteins or surfaces, and, through these interactions, determine the location and substrate specificity of these enzymes. These noncatalytic segments, which have also been referred to as regulatory regions, are often organized into recognizable domains (Patthy, 1985). Thus far, five types of domains have been identified, including kringle domains, vitamin K dependent calcium binding domains (or gla domains), growth factor domains, and type I and type II domains of fibronectin (Kurachi & Davie, 1982; Degan et al., 1983; Leytus et al., 1984; Fung et al., 1985; Malinowski et al., 1984; Pennica et al., 1983; McMullen & Fujikawa, 1985; Foster & Davie, 1984; Gunzler et al., 1982). These domains, including the kringle domain, growth factor domain, and type I and type II fibronectin domains, are present in other proteins (Sottrup-Jensen et al., 1978; Gray et al., 1983). This indicates that these domains have evolved in genes other than the serine protease family and have been transferred and fused to the serine protease domain during evolution by some translocation event.

The four tandem repeats that constitute the heavy chain of plasma kallikrein define a new type of domain that has not been observed in other proteases except for factor XI (Fujikawa et al., 1986). These tandem repeats are essential for the coagulant activity and neutrophil aggregation by plasma kallikrein (van der Graaf et al., 1982; Schapira et al., 1982). Cleavage of the heavy chain of plasma kallikrein by an unidentified protease(s) in acetone-activated plasma produces a modified form of plasma kallikrein, designated β-kallikrein. This enzyme shows reduced activity in the cleavage of high molecular weight kininogen and fails to elicit neutrophil aggregation and elastase release (Colman et al., 1985). The contribution of the heavy chain to the expression of coagulant activity can be attributed to the specific binding of the heavy chain to high molecular weight kininogen. This interaction is required prior to the proteolytic conversion of high molecular weight kininogen to an activated form that is composed of an amino-terminal heavy chain and a carboxyl-terminal light chain. This proteolytic modification of high molecular weight kininogen is associated with the expression of enhanced coagulant cofactor activity (Scott et al., 1984). The role of the heavy chain in neutrophil aggregation is unclear. It is suggested that the heavy chain might bind to neutrophil membrane receptors and that such binding is required prior to the expression of the kallikrein proteolytic activity that leads to neutrophil aggregation and elastase release (Colman et al., 1985).

An analysis of the degree of homology among the four repeats in plasma prekallikrein indicates that repeats 1 and

3, and repeats 2 and 4, share the highest degree of homology. This suggests that the time of divergence of repeat 3 from repeat 1, and repeat 4 from repeat 2, is very similar. This pattern of similarity is consistent with the proposal that the four tandem repeats in the heavy chain of plasma kallikrein resulted from two gene duplication events in which a single primordial repeat element was initially duplicated to give rise to two tandem repeats. Subsequently, the entire locus containing the two contiguous repeats was duplicated to give rise to the present four tandem repeats.

Plasma prekallikrein shows a high degree of identity (58%) with factor XI which also contains four homologous tandem repeats in the amino terminus of the molecule (Fujikawa et al., 1986). The high degree of identity between these two proteins extends beyond the tandem repeats into the catalytic domain and suggests the genes for these two proteins are very closely related and have diverged very recently from a common precursor. The difference coefficient between these two sequences ($K_{PK,FXI}$), uncorrected for multiple substitutions at the same amino acid position, was calculated to be 0.53 (Kimura & Ohta, 1971; Dickerson, 1971; Doolittle, 1979). If one assumes that amino acid replacements occur in the plasma prekallikrein and factor XI genes at the same rate as in the globin genes, the time that plasma prekallikrein and factor XI share a common ancestor can be estimated to be about 280 millions years ago (Doolittle, 1983). This is in good agreement with an estimated divergence time of 256 millions years ago, which is calculated by using the nucleotide substitution rate of 0.41 nucleotide per 100 codons per million years (Fitch, 1976; Wilson et al., 1977).

Homology alignments of the catalytic chain of plasma kallikrein with the corresponding chains of other serine proteases indicate that the light chain of plasma kallikrein is closely related to those of factor IX, factor XII, and protein C, and more distantly related to those of plasmin, urokinase, tissue plasminogen activator, and pancreatic kallikrein. The fact that the heavy chain of plasma prekallikrein has no structural similarity with these serine proteases suggests that the primordial gene for plasma prekallikrein and factor XI has probably diverged from a common ancestor of factor IX through a recombination event(s) that fused the catalytic chain to the heavy-chain repeat elements.

Purified human plasma prekallikrein consists of two similar forms with apparent molecular weights of 85 000 and 88 000. This is apparently due to heterogeneity in the light chain of the molecule. Thus, activation of plasma prekallikrein gives rise to two very similar forms of plasma kallikrein that contain light chains with molecular weights of either 36 000 or 33 000 (Mandle & Kaplan, 1977). These two forms are present in approximately equal amounts. In the course of sequencing plasma prekallikrein and its derived peptides, and in sequencing and mapping of the cDNAs for this molecule, no evidence of

heterogeneity in the sequence was observed that would account for the apparent difference in molecular weight of the light chains. In the determination of the carbohydrate attachment sites on CNB9, only tryptic peptides that contain hexosamine were purified and sequenced. These studies identified three asparagine residues on the light chain that were glycosylated. However, these results do not prove that every plasma prekallikrein molecule is necessarily glycosylated at all three sites. In addition, it is possible that this heterogeneity may be caused by differences associated with the carbohydrate chains attached to the light chain. In related studies, Hojima and co-workers (Hojima et al., 1985) have reported that treatment of prekallikrein with neuraminidase eliminated some of the heterogeneity in purified preparations of human plasma prekallikrein identified by isoelectric focusing. Additional studies of the gene copy number and structure may be helpful in determining the cause of this heterogeneity.

ACKNOWLEDGMENTS

REFERENCES

Benton, W. D., & Davis, R. W. (1977) Science (Washington, D.C.) 196, 188–192.

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

Bock, P. E., Srinivasan, K. R., & Shore, J. D. (1981) Biochemistry 20, 7258–7266.

Bouma, B. N., & Griffin, J. H. (1977) J. Biol. Chem. 252, 6432–6437.

Bouma, B. N., Miles, L. A., Beretta, G., & Griffin, J. H. (1980) Biochemistry 19, 1151–1160.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) Biochemistry 14, 3029–3035.

Bridgen, P. J., Cross, G. A. M., & Bridgen, J. (1976) Nature (London) 263, 613–614.

Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., & Davie, E. W. (1983) Biochemistry 22, 3244–3250.

Cochrane, C. G., Revak, S. D., & Wuepper, K. D. (1973) J. Exp. Med. 138, 1564–1583.

Colman, R. W., Wachtfogel, Y. T., Kucich, U., Weinbaum, G., Hahn, S., Pixley, R. A., Scott, C. F., de Agostini, A., Burger, D., & Schapira, M. (1985) Blood 65, 311–318.

Crestfield, A. M., Moore, S., & Stein, W. H. (1963) J. Biol. Chem. 238, 622–627.

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

Dickerson, R. E. (1971) J. Mol. Evol. 1, 26–48.

Doolittle, R. F. (1979) Proteins (3rd Ed.) 4, 1–118.

Doolittle, R. F. (1983) Ann. N.Y. Acad. Sci. 408, 13–26.

Edman, P., & Begg, G. (1967) Eur. J. Biochem. 1, 80–91.

Ericsson, L. H., Wade, R. D., Gagnon, J., McDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in Solid Phase Methods in Protein Sequence Analysis (Previero, A., & Coletti-Previero, M. A., Eds.) pp 137–142, Elsevier, Amsterdam.

Fitch, W. M. (1976) in Molecular Evolution (Ayala, F. J., Ed.) p 160, Sinauer, Sunderland, MA.

Foster, D., & Davie, E. W. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 4766–4770.

Fujikawa, K., & McMullen, B. A. (1983) J. Biol. Chem. 258, 10924–10933.

Fujikawa, K., Heimark, R. L., Kurachi, K., & Davie, E. D. (1980) Biochemistry 19, 1322–1330.

Fujikawa, K., Chung, D. W., Hendrickson, L. E., & Davie, E. W. (1986) Biochemistry (following paper in this issue).

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

Gray, A., Dull, T. J., & Ullrich, A. (1983) Nature (London) 303, 722–725.

Griffin, J. H., & Cochrane, C. G. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 2554–2558.

Griffin, J. H., & Cochrane, C. G. (1979) Semin. Thromb. Hemostasis 5, 254–273.

Gunzler, W. A., Steffens, G. J., Otting, F., Kim, S.-M. A., Frankus, E., & Flohe, L. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 1155–1165.

Habal, F. M., Movat, H. Z., & Burrowes, C. E. (1974) Biochem. Pharmacol. 23, 2291–2303.

Harboe, N., & Ingild, A. (1973) Scand. J. Immunol. 2, Suppl. 1, 161–164.

Heimark, R. L., & Davie, E. W. (1981) Methods Enzymol. 80, 157–172.

Heimark, R. L., Kurachi, K., Fujikawa, K., & Davie, E. W. (1980) Nature (London) 286, 456–460.

Hojima, Y., Pierce, J. C., & Pisano, J. J. (1985) J. Biol. Chem. 260, 400–406.

Kato, H., Nagasawa, S., & Iwanaga, S. (1981) Methods Enzymol. 80, 172–198.

Kimura, M., & Ohta, T. (1971) J. Mol. Evol. 1, 1–25.

Kurachi, K., & Davie, E. W. (1977) Biochemistry 16, 5831–5839.

Kurachi, K., & Davie, E. W. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6461–6464.

Kwok, S. C., Ledley, F. D., DiLella, A. G., Robson, K. J., & Woo, S. L. C. (1985) Biochemistry 24, 556–561.

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

Malinowski, D. P., Sadler, J. E., & Davie, E. W. (1984) Biochemistry 23, 4243–4250.

Mandle, R., Jr., & Kaplan, A. P. (1977) J. Biol. Chem. 252, 6097–6104.

Mandle, R., Jr., Colman, R. W., & Kaplan, A. P. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 4179–4183.

Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 1184–1188.

Mannhalter, C., Schiffman, S., & Jacobs, A. (1980) J. Biol. Chem. 255, 2667–2669.

Margolis, J. (1958) J. Physiol. (London) 144, 1–22.

McMullen, B. A., & Fujikawa, K. (1985) J. Biol. Chem. 260, 5328–5341.

Meier, H. L., Pierce, J. V., Colman, R. W., & Kaplan, A. P. (1977) J. Clin. Invest. 60, 18–31.

Messing, J., Crea, R., & Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309–321.

Nagasawa, S., & Nakayasu, T. (1973) J. Biochem. (Tokyo) 74, 401–403.

Niewiarowski, S., & Prou-Wartelle, O. (1959) Thromb. Diath. Haemorrh. 3, 593–603.

Norrander, J., Kempe, T., & Messing, J. (1983) Gene 26, 101–106.

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

Pennica, D., Holmes, W. E., Kohn, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heynecker, H. L., Goeddel, D. V., & Collen, D. (1983) Nature (London) 301, 214–221.

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

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Ratnoff, O. D. (1966) *Prog. Hematol. 5*, 204–245.

Revak, D. D., Cochrane, C. G., Johnson, A. R., & Hugli, T. E. (1974) *J. Clin. Invest. 54*, 619–627.

Revak, S. D., Cochrane, C. G., & Griffin, J. H. (1977) *J. Clin. Invest. 59*, 1167–1175.

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

Schapira, M., Despland, E., Scott, C. F., Boxer, L. A., & Colman, R. W. (1982) *J. Clin. Invest. 69*, 1199–1202.

Scott, C. F., Silver, L. D., Schapira, M., & Colman, R. W. (1984) *J. Clin. Invest. 73*, 954–962.

Sealey, J. E., Atlas, S. A., Laragh, J. H., Silvergerg, M., & Kaplan, A. P. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5914–5918.

Shine, J., Mason, A. J., Evans, B. A., & Richards, R. I. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 419–426.

Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T. E., & Magnusson, S. (1978) *Prog. Chem. Fibrinolysis Thrombolysis 3*, 191–209.

Takagaki, Y., Kitamura, N., & Nakanishi, S. (1985) *J. Biol. Chem. 260*, 8601–8609.

Tans, G., & Griffin, J. H. (1982) *Blood 59*, 69–75.

Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H., & Davie, E. W. (1972) *Biochemistry 11*, 4899–4903.

van der Graaf, F., Tans, G., Bouma, B. N., & Griffin, J. H. (1982) *J. Biol. Chem. 257*, 14300–14305.

Weber, K., & Osborn, M. (1969) *J. Biol. Chem. 244*, 4406–4412.

Wilson, A. C., Carlson, S. S., & White, T. J. (1977) *Annu. Rev. Biochem. 46*, 573–639.

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

Young, R. A., & Davis, R. W. (1983a) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194–1198.

Young, R. A., & Davis, R. W. (1983b) *Science (Washington, D.C.) 222*, 778–782.

Amino Acid Sequence of Human Factor XI, a Blood Coagulation Factor with Four Tandem Repeats That Are Highly Homologous with Plasma Prekallikrein[†]

Kazuo Fujikawa, Dominic W. Chung, Lee E. Hendrickson, and Earl W. Davie*

Department of Biochemistry, University of Washington, Seattle, Washington 98195

Received November 21, 1985; Revised Manuscript Received January 10, 1986

ABSTRACT: A λgt11 cDNA library prepared from human liver poly(A) RNA has been screened with affinity-purified antibody to human factor XI, a blood coagulation factor composed of two identical polypeptide chains linked by a disulfide bond(s). A cDNA insert coding for factor XI was isolated and shown to contain 2097 nucleotides, including 54 nucleotides coding for a leader peptide of 18 amino acids and 1821 nucleotides coding for 607 amino acids that are present in each of the 2 chains of the mature protein. The cDNA for factor XI also contained a stop codon (TGA), a potential polyadenylation or processing sequence (AACAAA), and a poly(A) tail at the 3′ end. Five potential N-glycosylation sites were found in each of the two chains of factor XI. The cleavage site for the activation of factor XI by factor XII$_a$ was identified as an internal peptide bond between Arg-369 and Ile-370 in each polypeptide chain. This was based upon the amino acid sequence predicted by the cDNA and the amino acid sequence previously reported for the amino-terminal portion of the light chain of factor XI. Each heavy chain of factor XI$_a$ (369 amino acids) was found to contain 4 tandem repeats of 90 (or 91) amino acids plus a short connecting peptide. Each repeat probably forms a separate domain containing three internal disulfide bonds. The light chains of factor XI$_a$ (each 238 amino acids) contain the catalytic portion of the enzyme with sequences that are typical of the trypsin family of serine proteases. The amino acid sequence of factor XI shows 58% identity with human plasma prekallikrein.

Factor XI (plasma thromboplastin antecedent) is a zymogen to a serine protease that participates in the early or contact phase of blood coagulation (Davie et al., 1979). Patients deficient in this plasma glycoprotein have a mild bleeding disorder or are asymptomatic (Rapaport et al., 1961; Ratnoff & Saito, 1979). The physical and chemical properties of factor XI have been determined with purified preparations isolated from human (Bouma & Griffin, 1977; Kurachi & Davie, 1977, 1981) and bovine plasma (Koide et al., 1976, 1977a; Kurachi et al., 1980). Factor XI is unique in that it is a zymogen to

a serine protease that is composed of two identical polypeptide chains linked by a disulfide bond(s). Its molecular weight has been reported to be between 125 000 and 160 000 for the dimer and between 55 000 and 63 000 for the monomer. Human factor XI contains 5% carbohydrate (Kurachi & Davie, 1977), whereas bovine factor XI contains 11% carbohydrate (Koide et al., 1977a).

Factor XI circulates in plasma as a complex with high molecular weight kininogen (HMW kininogen)[1] (Thompson et al., 1977). It is converted in vitro to factor XI$_a$ by factor

---

[1] Abbreviations: HMW kininogen, high molecular weight kininogen; SDS, sodium dodecyl sulfate; TBS, 50 mM Tris-HCl buffer, pH 7.5, containing 150 mM NaCl; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; kb, kilobase(s).