Olender, D. L., Zurawski, G., & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5524-5528.
Porschke, D., & Eigen, M. (1971) *J. Mol. Biol. 62*, 361-381.
Price, J. V., & Cech, T. R. (1988) *Genes Dev. 2*, 1439-1447.
Price, J. V., Engberg, J., & Cech, T. R. (1987) *J. Mol. Biol. 196*, 49-60.
Pyle, A. M., McSwiggen, J., & Cech, T. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8187-8191.
Reed, R., & Maniatis, T. (1986) *Cell 46*, 681-690.
Rochaix, J.-D., Rahire, M., & Michel, F. (1985) *Nucleic Acids Res. 13*, 975-984.
Sogin, M. L., & Edman, J. C. (1989) *Nucleic Acids Res. 17*, 5349-5359.
Somasekhar, M. B., & Mertz, J. E. (1985) *Nucleic Acids Res. 13*, 5591-5609.
Sugimoto, N., Kierzek, R., & Turner, D. H. (1988) *Biochemistry 27*, 6384-6392.
Suh, E. R., & Waring, R. B. (1990) *Mol. Cell. Biol. 10*, 2960-2965.
Szostak, J. W. (1986) *Nature 322*, 83-86.
Tuerk, C., Gauss, P., Thormer, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Brody, E. N., & Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364-1368.
Veldman, G. M., Klootwijk, J., de Regt, V. C. H. F., Planta, R. H., Branlant, C., Krol, A., & Ebel, J.-P. (1981) *Nucleic Acids Res. 9*, 6935-6952.
Waring, R. B., Towner, P., Minter, S. J., & Davies, R. W. (1986) *Nature 321*, 133-139.
Williams, A. P., Longfellow, C. E., Freier, S. M., Kierzek, R., & Turner, D. J. (1989) *Biochemistry 28*, 4283-4291.
Woodson, S. A., & Cech, T. R. (1989) *Cell 57*, 335-345.
Zaug, A. J., & Cech, T. R. (1986) *Science 231*, 470-475.
Zaug, A. J., Kent, J. R., & Cech, T. R. (1984) *Science 224*, 575-578.
Zaug, A. J., Been, M. D., & Cech, T. R. (1986) *Nature 324*, 429-433.

Location of the Disulfide Bonds in Human Plasma Prekallikrein: The Presence of Four Novel Apple Domains in the Amino-Terminal Portion of the Molecule†

Brad A. McMullen, Kazuo Fujikawa, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received September 24, 1990; Revised Manuscript Received November 19, 1990

ABSTRACT: The location of 16 of the 18 disulfide bonds in human plasma prekallikrein was determined by amino acid sequence analysis of cystinyl peptides produced by chemical and enzymatic digestions. A unique structure, named the apple domain, was established for each of the four tandem repeats in the amino-terminal portion of the molecule. The apple domains (90 or 91 amino acids) contain 3 highly conserved disulfide bonds linking the first and sixth, second and fifth, and third and fourth half-cystine residues present in each repeat. The fourth tandem repeat contains an extra disulfide bond that forms a second small loop within the apple domain. The carboxyl-terminal portion of plasma prekallikrein containing the catalytic region of the molecule was found to have disulfide bonds located in positions similar to those of other serine proteases.

Plasma prekallikrein, a precursor of plasma kallikrein, circulates in blood as an equimolar complex with high molecular weight kininogen (Mandle et al., 1976). It participates in the generation of kinin (Werle, 1955; Margolis, 1958; Laake & Vennerod, 1974) and may also play a role in the intrinsic pathway of blood coagulation (Weupper, 1973; Saito et al., 1974; Weiss et al., 1974) and fibrinolysis (Laake & Vennerod, 1974; Ogston et al., 1969; Mandle & Kaplan, 1977; Ichinose et al., 1986). Plasma prekallikrein (M_r 85 000) is converted to plasma kallikrein by factor XIIa by the cleavage of a single internal Arg-Ile peptide bond (Griffin & Cochrane, 1979; Heimark et al., 1980). Plasma kallikrein is a serine protease that is composed of a heavy chain (371 amino acids) and a light chain (248 amino acids) held together by a disulfide bond(s) (Chung et al., 1986).

The amino acid sequence of plasma prekallikrein is 58% identical with factor XI, a plasma protein that is converted to factor XIa in the early phase of the blood coagulation cascade. The amino-terminal regions of both plasma prekallikrein and factor XI also contain 4 tandem repeats of 90 (or 91) residues that show a high degree of amino acid sequence identity (Chung et al., 1986; Fujikawa et al., 1986). Furthermore, each of the four tandem repeats contains six half-cystine (Cys)[1] residues that are highly conserved, being present in the same position in each of the four tandem repeats in both proteins. Also, the positions of the six Cys residues in the two proteins do not align with the Cys residues of any known protein domain containing disulfide bonds. These data suggest that plasma prekallikrein as well as factor XI may have three unique disulfide bonds in each of their four tandem repeats. Plasma prekallikrein contains two additional Cys residues in the fourth repeat, while each of the first and fourth

† This work was supported in part by Grant HL 16919 from the National Institutes of Health. A preliminary report was presented at the Fourth Annual Symposium of the Protein Society held in San Diego, CA, Aug 11-15, 1990.

[1] Abbreviations: Cys, half-cystine; PTH, phenylthiohydantoin.

0006-2960/91/0430-2050$02.50/0

repeats of factor XI have one additional Cys residue. These additional Cys residues in each protein are also linked by disulfide bonds since neither protein contains a free sulfhydryl group.

The carboxyl-terminal region of plasma prekallikrein and factor XI contains the serine protease portion of each protein. This region in the two proteins also shows considerable amino acid sequence identity and is homologous to pancreatic trypsin.

In this paper, the disulfide bonds in human plasma prekallikrein have been examined. The data indicate the presence of a novel disulfide bond structure in each of the four tandem repeats in the protein. These structures have been called apple domains for convenience since they can be drawn in the shape of an apple. These same structures have also been identified in human factor XI [see McMullen et al. (1991)].

Experimental Procedures

Materials. Plasma prekallikrein was isolated as described previously (Tait & Fujikawa, 1987) and was homogeneous by SDS–polyacrylamide gel electrophoresis. Ammonium 7-fluoro-2,1,3-benzoxadiazole-4-sulfonate was purchased from Wako Pure Chemicals, Osaka, Japan. Tri-*n*-butylphosphine was obtained from Aldrich. Trypsin treated with *N*-tosyl-L-phenylalanine chloromethyl ketone and chymotrypsin treated with *N*-tosyl-L-lysine chloromethyl ketone were purchased from Worthington. *Staphylococcus aureus* V8 protease was obtained from Miles. The TSK-G2000SW column (7.5 × 600 mm) was purchased from LKB-Produkter AB, Bromma, Sweden, while the μBondapak C-18 (4.9 × 300 mm) and Ultrapore C-3 columns (4.6 × 75 mm) were purchased from Waters and Altex, respectively. Reagents for amino acid sequence analysis were obtained from Applied Biosystems, Inc. All other chemicals were of the highest grade commercially available.

Chemical and Enzymatic Digestion. Cleavage of plasma prekallikrein at methionyl bonds was performed by dissolving 1 mg of the pure protein in 500 μL of 70% formic acid containing 1 mg of cyanogen bromide, and incubating the mixture at room temperature for 24 h. The sample was then dried under a stream of nitrogen prior to fractionation. Cleavage at aspartyl bonds was carried out by heating the sample for 4 h at 110 °C in 500 μL of 2% formic acid under argon. Digestions with trypsin and chymotrypsin were performed at 37 °C in 0.1 M ammonium formate, pH 6.5, at an enzyme-to-substrate weight ratio of 1/100 and 1/30, respectively. Incubation time for the digestion with trypsin was 18 h, while the incubation time with chymotrypsin was 4 h. Digestions with *Staphylococcus aureus* V8 protease were performed in 0.1 M ammonium acetate, pH 4.5, at 37 °C for 18 h at enzyme-to-substrate weight ratio of 1/30. In total, 2 mg of protein was used.

Purification of Peptides. Cyanogen bromide fragments were first fractionated by gel filtration on two TSK-G2000SW columns connected in tandem. The columns were previously equilibrated with 6 M guanidine hydrochloride in 1% phosphoric acid, pH 3.2. After application of the sample, the columns were eluted with the same solution at a flow rate of 1 mL/min. The effluent was monitored for peptides by the absorbance at 226 nm. Further purification of the cyanogen bromide fragments and peptides derived from enzymatic cleavage was carried out by reverse-phase HPLC using either a μBondapak C-18 column or an Ultrapore C-3 column connected to a Waters HPLC system. The gradient employed consisted of 0.1% trifluoroacetic acid as the mobile phase and 0.08% trifluoroacetic acid in 80% acetonitrile as the mobile phase modifier. The flow rate was 1.5 mL/min, and the effluents were monitored by the absorbance at 214 nm.

Sequence Analysis and Detection of Cystinyl Peptides. Amino acid sequence analysis of selected peptides was performed with an Applied Biosystems Model 477A pulse-liquid protein sequencer. The phenylthiohydantoin (PTH)-amino acids were separated on a Model 120A analyzer connected to the protein sequencer. DiPTH-cystine was identified as previously reported by Marti et al. (1987) and Hojrup and Magnusson (1987), while cystinyl peptides were detected by the procedure of Sueyoshi et al. (1985), and samples representing 2.5–20% of the total material were employed for assay. Fluoresence intensities were measured at 520 nm by a Perkin-Elmer Model LS-5 fluorescence spectrophotometer with excitation at 385 nm.

Peptide Nomenclature. The pooled fractions and peptides were identified by letters to indicate the type of cleavage performed. Numbers were employed to show the fraction or peptide positions relative to the other peaks in the chromatograms. The nomenclature used was as follows: CNB, cyanogen bromide cleavage; D, acid cleavage at an aspartyl peptide bond; T, trypsin digestion; E, *S. aureus* V8 protease digestion; E/C, *S. aureus* V8 protease followed by chymotrypsin digestion.

Results

Human plasma prekallikrein was cleaved at methionine residues with cyanogen bromide, and the digest was separated into three major peaks by gel filtration on the two TSK-G2000SW columns connected in tandem. Two of the peaks, CNB1 and CNB2, were found to contain cystinyl peptides by the fluorometric assay (left panel, Figure 1). CNB2 was further fractionated by reverse-phase HPLC on a C-18 column, and three major peaks were obtained (middle panel, Figure 1). CNB2-3 was found to contain Cys residues and was subjected to amino acid sequence analysis. It was completely sequenced with the detection of a single PTH-amino acid in each cycle, except for positions 2 and 8. The sequence of CNB2-3 corresponds to amino acid residues 31–52 in the first tandem repeat of plasma prekallikrein (Table I, Figure 2) (Chung et al., 1986). In the second cycle of the sequence analysis, no PTH-amino acid was identified. DiPTH-cystine, however, was found in the eighth cycle. These two positions correspond to the Cys residues at positions 32 and 38 in the amino acid sequence of plasma prekallikrein (Figure 2). Thus, a disulfide bond forming a small loop of seven amino acid residues was present in the first tandem repeat.

After CNB1 was desalted on a C-3 column, it was digested with trypsin in 0.1 M ammonium acetate buffer, pH 6.5. This slightly acidic buffer was used for the tryptic and chymotryptic digestions to prevent possible rearrangements of the disulfide bridges in the peptide fragments (Ryle & Sanger, 1955; Spackman et al., 1960). The tryptic digest was then separated on a C-3 column, and five fractions, CNB1-T1, -T3, -T4, -T5, and -T6, were pooled (right panel, Figure 1). Four of the fractions, CNB1-T1, -T3, -T4, and -T6, were pooled and rechromatographed either on a C-18 or on a C-3 column (top and bottom panels, Figure 3).

CNB1-T1-16 (top left panel, Figure 3) gave two complete sequences in essentially equal amounts starting with Val_{112} and Asn_{145}. Only PTH-Ser was detected in the third cycle, and diPTH-cystine was identified in the seventh cycle. These results indicated that Cys_{118} was linked to Cys_{147} in the second tandem repeat (Table I, Figure 2). Similarly, a double sequence of CNB1-T3-1 (top right panel, Figure 3) indicated that Cys_{529} was paired with Cys_{544} in the catalytic region of the protein (Table I, Figure 2). CNB1-T3-2 also gave a double

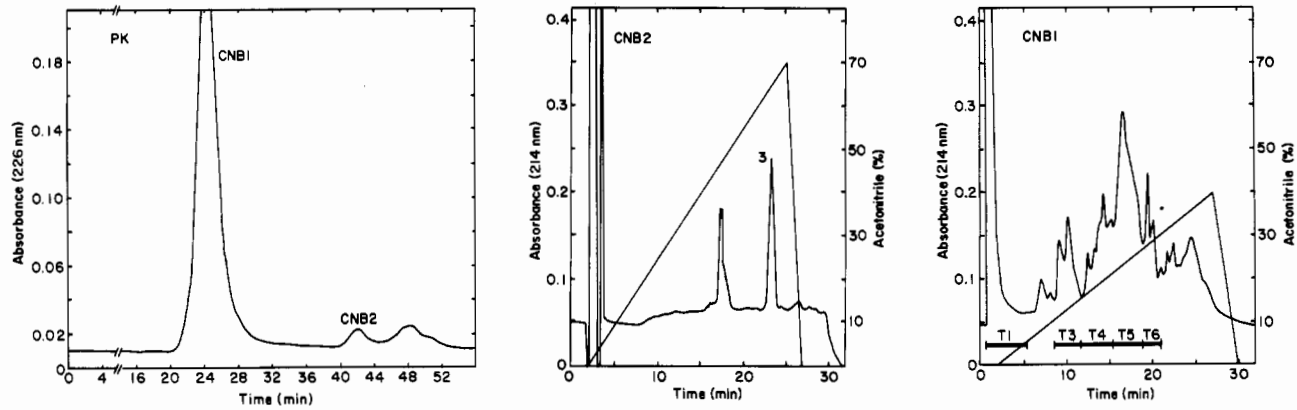

FIGURE 1: Fractionation of peptides derived from human plasma prekallikrein. (Left panel) gel filtration of the cyanogen bromide digest of the protein on two TSK-G2000SW columns connected in tandem. The effluent was monitored for peptides at 226 nm, and the fractions were collected manually. The flow rate was 1 mL/min. Two of the peaks were labeled CNB1 and CNB2, since they contained Cys residues as detected by the fluorescence assay. (Middle panel) separation of CNB2 on a C-18 column connected to a HPLC. After application of the sample, the column was washed for 2 min with 0.1% trifluoroacetic acid, and then a linear gradient was employed, as described under Experimental Procedures. The concentration of acetonitrile is indicated on the ordinate. Peptide CNB2-3 contained Cys residues, as detected by the fluorescence assay. (Right panel) initial separation of the tryptic digest of CNB1. The CNB1 digest was fractionated on an Altex C-3 column by employing the same conditions described for the middle panel. The five fractions indicated by solid bars were found to contain Cys residues by the fluorescence assay and were pooled individually.

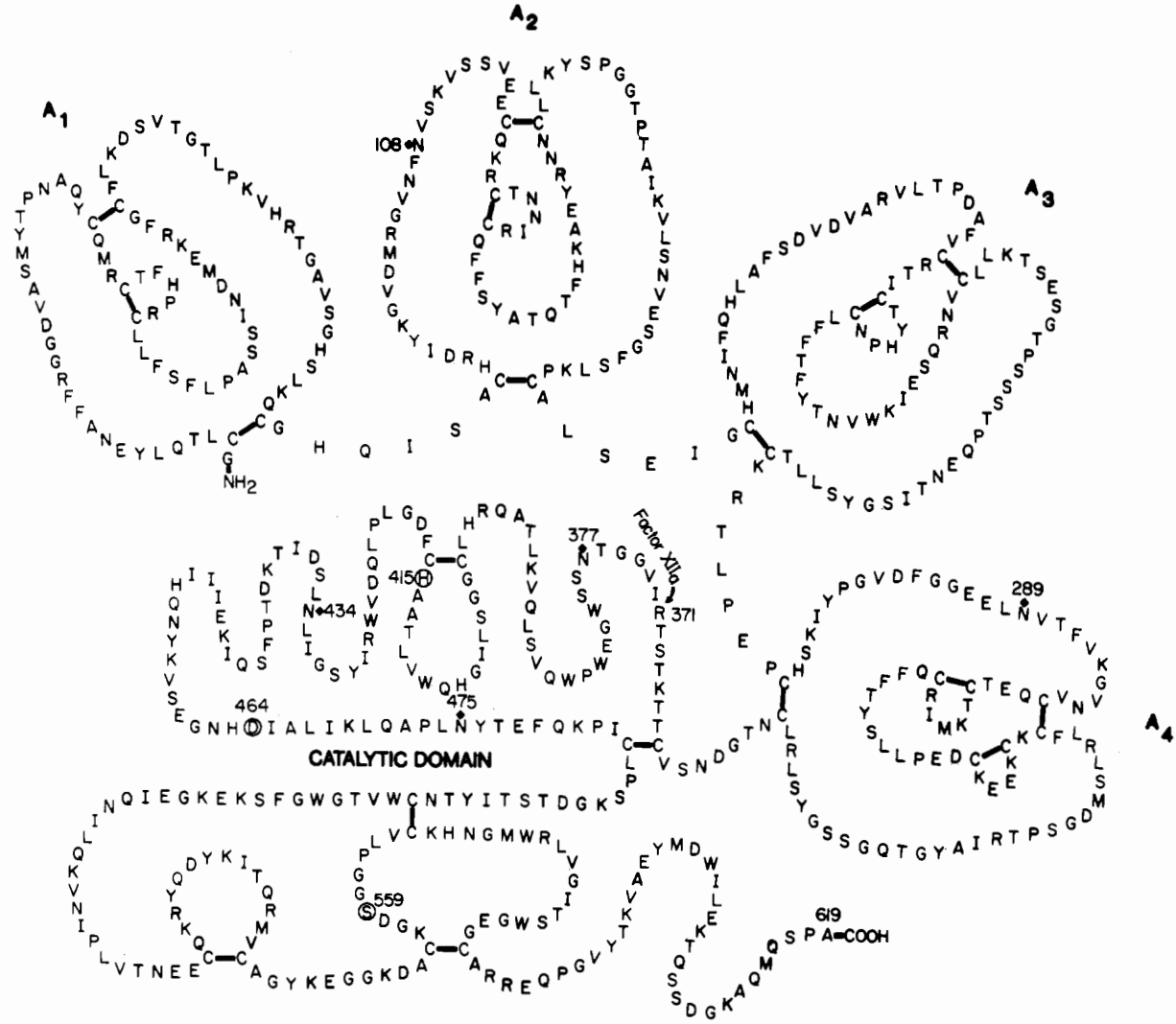

FIGURE 2: Location of the disulfide bonds in human plasma prekallikrein. The four apple domains in the amino-terminal portion of the molecule are labeled A_1, A_2, A_3, and A_4. The Arg_{371}/Ile_{372} bond that is cleaved by factor XIIa during the activation reaction is marked by a solid curved arrow. The circled residues (His_{415}, Asp_{464}, and Ser_{559}) are members of the catalytic triad characteristic of serine proteases. The Asn residues marked by solid diamonds (residues 108, 289, 377, 434, and 475) are attachment sites for carbohydrate chains. The amino- and carboxyl-terminal residues are identified by $-NH_2$ and -COOH, respectively.

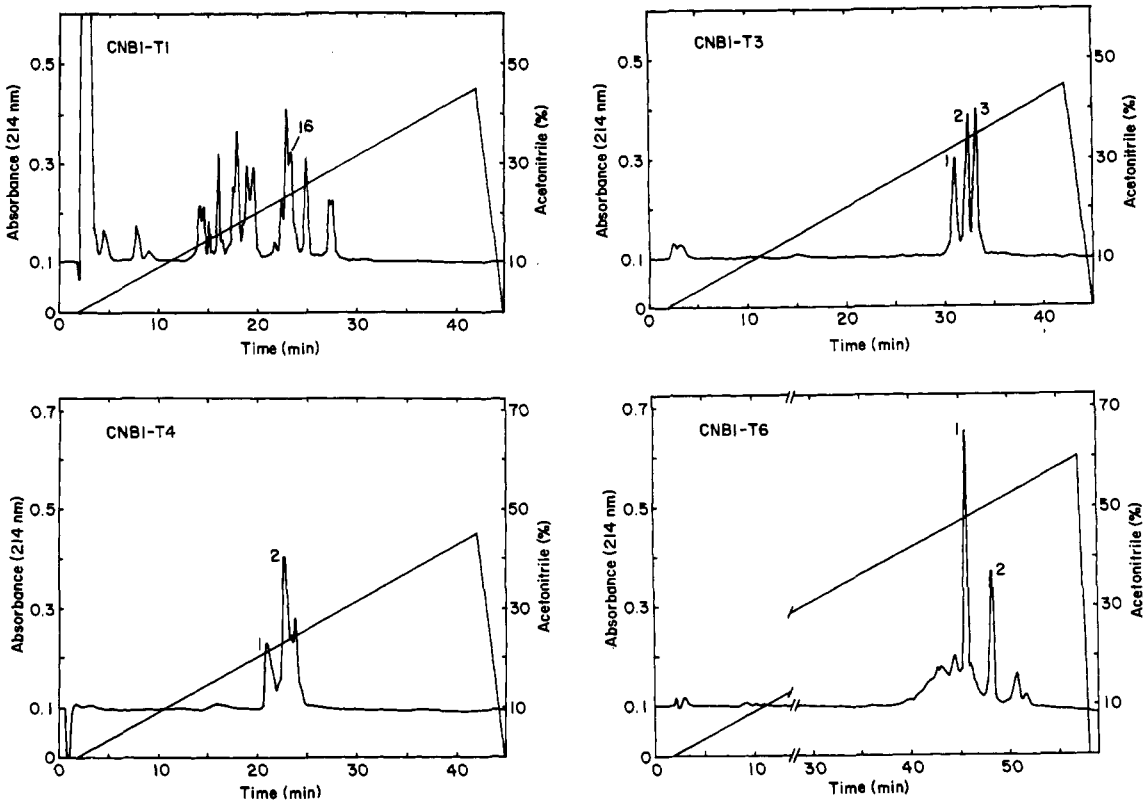

FIGURE 3: Rechromatography of selected tryptic fragments (CNB1-T1, -T3, -T4, and -T6 shown in the right panel, Figure 1) on C-18 or C-3 columns. CNBI-T1 (top left panel), CNBI-T3 (top right panel) and CNBI-T6 (lower right panel) were separated on a C-18 column, and CNBI-T4 was rerun on an Altex C-3 column (lower left panel). The peaks in each panel that are numbered contained Cys residues, as detected by the fluorescence assay.

sequence and established a disulfide bond between Cys_{122} and Cys_{128} in the second repeat. The sequence analysis of CNB1-T3-3 showed that Cys_{28} was paired with Cys_{58} in the first repeat. A disulfide bridge between Cys_{208} and Cys_{236} in the third repeat was found in the sequence of CNB1-T4-1 (lower left panel, Figure 3). Both CNB1-T6-1 and CNB1-T6-2 (lower right panel, Figure 3) gave a single sequence and established two disulfide bonds, one between Cys_{400} and Cys_{416} in the catalytic region and the other between Cys_{212} and Cys_{218} in the third tandem repeat (Figure 2).

CNB1-T4-2 (lower left panel, Figure 3), which contained three disulfide bonds, was shown to originate from the fourth repeat. It was sequenced completely and found to contain two subpeptides connected by disulfide bonds. The two peptides ran from Gly_{295} to Lys_{305} and from Cys_{309} to Arg_{331} (Figure 2). Inspection of the sequence revealed the presence of four Glu residues at positions 301, 319, 323, and 324 in the peptide. Accordingly, a digestion of CNB1-T4-2 with *S. aureus* V8 protease was carried out. An analytical run of the digest on a C-18 column showed that neither a new fragment nor a change of the column retention time of the starting material had occurred. Amino acid sequence analysis, however, showed the presence of a new N-terminus starting at Thr_{302}. This indicated that at least one peptide bond was cleaved. When treated with chymotrypsin, CNB1-T4-2 was partially digested and separated into four peaks on a C-18 column (top panel, Figure 4). Sequence analysis of CNB1-T4-2-E/C2 gave a double sequence and revealed a disulfide bond between Cys_{303} and Cys_{309} (Table I, Figure 2). CNB1-T4-2-E/C3 contained four Cys residues, but the precise location of the two disulfide bonds in this peptide was not directly established. The two disulfide bonds, however, were tentatively assigned by analogy to the Cys residues in repeats 1, 2, and 3 (see Discussion).

CNB1-T5 was digested further with *S. aureus* V8 protease, and the resulting peptides were resolved on a C-18 column (middle panel, Figure 4). CNB1-T5-E7 gave a double sequence that indicated a disulfide bond between Cys_{182} and Cys_{265} in the third repeat (Table I, Figure 2). Three sequences were observed in CNB1-T5-E8. Unlike the difficulty experienced with CNB1-T4-2-E/C3, it was readily possible to pair Cys_2 with Cys_{85} in the first repeat and Cys_{92} and Cys_{175} in the second repeat. Additional proof of the latter pairing was provided by the analysis of a small fragment isolated by digesting CNB1-T5-E9 with thermolysin (results not shown). The small difference in the retention time on the C-18 column of CNB1-T5-E8 and CNB1-T5-E9 was attributed to the cyclization of the latter at the N-terminal Gln residue to form a pyroglutaminyl residue.

The remaining disulfide bonds were found in the digest of CNB1-T5-E14. This peptide was treated with 2% formic acid, and the products were separated on a C-18 column (lower panel, Figure 4). Sequence analysis of CNB1-T5-E14-D2 showed that Cys_{273} was linked to Cys_{356} in the fourth repeat (Table I, Figure 2). In the catalytic region, Cys_{555} was found to be paired with Cys_{583} by analyzing CNB1-T5-E14-D3. A double sequence that was found in CNB1-T5-E14-D7 (lower panel, Figure 4) established a linkage between Cys_{498} and Cys_{565} in the catalytic domain. The interchain disulfide bond (Cys_{364} to Cys_{484}) between the heavy chain and the light chain of plasma kallikrein was finally found in the mixture of two peptides, CNB1-T5-E14-D4/5 (lower panel, Figure 4). These two peptides had the same amino acid sequence. A small difference in their retention time is probably due to a partial deglycosylation of the Asn-linked carbohydrate chain at position 475 by the acid treatment.

Discussion

In the present experiments, the location of the disulfide

Table I: Peptides Isolated and Sequenced for the Identification of the Disulfide Bonds in Human Plasma Prekallikrein

Peptide	Amino Acid Sequence[a]	Position of the Disulfide Bond(s)
CNB1-T5-E8	G C L T Q L Y E Q C G H Q I S A C H R S G F S L K P X A L S E	2-85 and 92-175
CNB1-T3-3	Y T P N A Q Y X Q HSer F G X F L K	28-58
CNB2-3	R X T F H P R C L L F S F L P A S S I N D HSer	32-38
CNB1-T1-16	V S S V E E C Q K R N N X L L K	118-147
CNB1-T3-2	C T N/S N I R C Q F F S Y A T Q T F H K	122-128
CNB1-T5-E7	I G X H HSer T S E S G T P S S S T P Q E N T I S G Y S L L T X K	182-265
CNB1-T4-1	V L T P D A F V C R N V X L L K	208-236
CNB1-T6-2	T I X T Y H P N C L F F T F Y T N V(W)K	212-218
CNB1-T5-E14-D2	T L P E P X H S K L X N T G	273-356
CNB1-T4-2-E/C3	G V N V X Q E (or) S L L P E D X K E E K C K X F L R (or)	299-326 or 328 and 321-326 or 328
CNB1-T4-2-E/C2	T C T K X Q F F T Y	303-309
CNB1-T5-E14-D4/5	N S V X T T K L Q A P L(N)Y T E F Q K P I X L P S	364-484
CNB1-T6-1	H L X G G S L I G H Q(W)V L T A A H X F D G L P L Q D V(W)R	400-416
CNB1-T5-E14-D7	T S T I Y T N C(W)V T G(W)G F S S G G P L V X K	498-565
CNB1-T3-1	V N I P L V T N E E C Q K R V X A G Y K	529-544
CNB1-T5-E14-D3	A X K G L V G I T S W G E G C A R	555-583

[a] Sequence analyses of 16 peptides containing disulfide bonds are shown. The positions that were not identified that are known from the published sequence to contain Cys residues are marked by "X". Residues identified as diPTH-cystine are marked by "C". The residues shown in parentheses are known from the published sequence but were not identified.

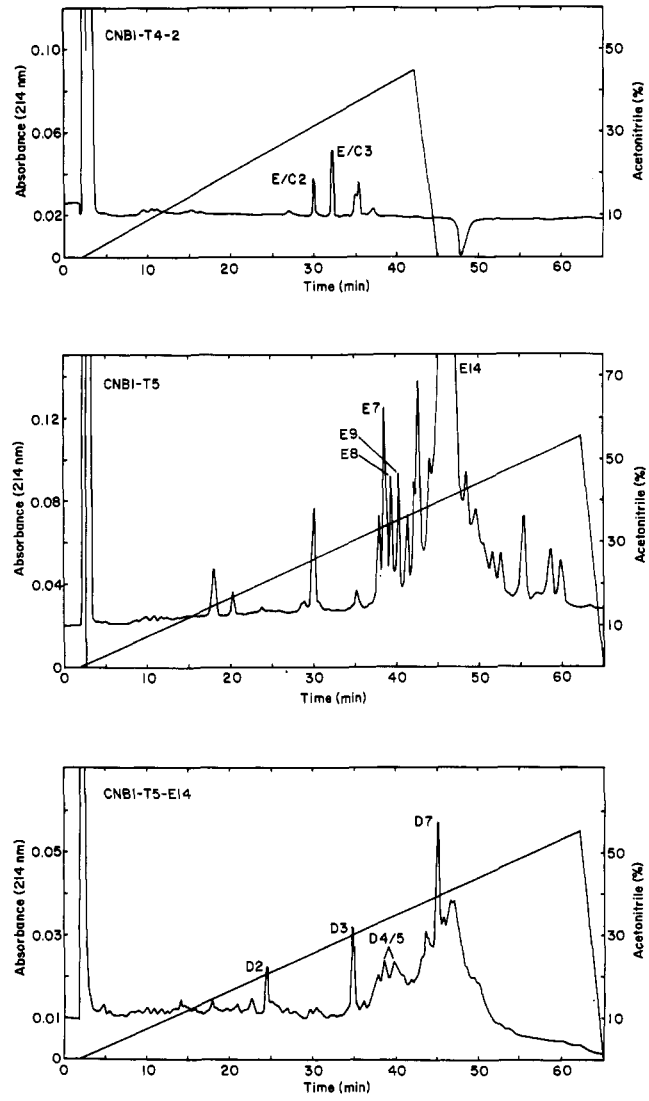

FIGURE 4: Fractionation of tryptic peptides of plasma prekallikrein on a C-18 column after digestion with *S. aureus* V8 protease, chymotrypsin, or acid. (Top panel) separation of peptides resulting from the digestion of CNB1-T4-2 with *S. aureus* V8 protease and chymotrypsin. (Middle panel) separation of peptides resulting from the digestion of CNB1-T5 with *S. aureus* V8 protease. (Lower panel) separation of peptides obtained by acid cleavage of CNB1-T5-E14.

bonds in human plasma prekallikrein was determined by sequence analysis of peptides produced by chemical and enzymatic digestions. Of the 18 disulfide bonds in plasma prekallikrein, 16 were unequivocally determined (Figure 2). The location of the two tentative disulfide bonds in the fourth tandem repeat is also shown in Figure 2. The 4 tandem repeats of 90 or 91 amino acids have characteristic disulfide bonds linking the first and sixth, second and fifth, and third and fourth Cys residues. These repeats have been called apple domains for convenience since they can be drawn in the shape of an apple.

Since each of the peptides that were used for the sequence analysis contained only one disulfide bond, no alternative pairing of disulfide bonds was possible in these peptides with the exception of CNB1-T5-E8 and CNB1-T4-2-E/C3. The determination of disulfide bonds in CNB1-T5-E8 was possible even though it contained four Cys residues. The detection of PTH-Gly and diPTH-cystine in the second cycle and PTH-Ala and diPTH-cystine in the ninth cycle of this peptide made it possible to establish the linkages between Cys_2 and Cys_{85} and between Cys_{92} and Cys_{175} (Table I).

The precise location of the two internal disulfide bonds in the fourth apple domain in plasma prekallikrein remains to be established. The peptide, CNB1-T4-2-E/C3, contained four Cys residues forming two disulfide bonds. However, sequence analysis was not successful in determining the exact linkage of these four Cys residues. This analysis did show, however, that Cys_{299} was not linked to Cys_{321} since no diPTH-cystine was observed in the seventh turn of the Edman degradation (see Table I). Unfortunately, the isolation of subfragments of this peptide could not be made, since CNB-T4-2-E/C3 was resistant to further digestion with trypsin and endopeptidase Lys-C. However, it is highly likely that Cys_{299} is linked to Cys_{328} since a disulfide bond is present in the three other apple domains in the same position (Figure 2). This suggests that the remaining two Cys residues at positions 321 and 326 form a disulfide bond leading to the second small loop of six amino acids in the fourth apple domain (Figure 2).

Recovery of the peptides was not quantitated in the present study since plasma prekallikrein is a trace protein and only small amounts of the purified protein were available for analysis. All cystinyl peptides that were used for the sequence analysis were major fragments, and no contradictory fragments were isolated. Indeed, all amino acid sequence analyses gave results that were in complete agreement with that previously determined by a combination of cDNA cloning and partial amino acid sequence analysis (Chung et al., 1986). This included the polymorphic site at residue 124 which contained

approximately 75% Asn and 25% Ser (CNB1-T3-2 in Table I).

The location of four internal disulfide bonds in the serine portion of plasma prekallikrein is very similar to those determined in other serine proteases, such as trypsin (Kauffman, 1965) and the B-chain of plasmin (Wiman, 1977). The interchain disulfide bond between positions 364 and 565 is also typical of that in other serine proteases (Kauffman, 1965; Wiman, 1977).

The amino-terminal chains of the various other serine proteases involved in blood coagulation and fibrinolysis consist of several different domains with characteristic disulfide bonds. Factor XII, t-PA (tissue-type plasminogen activator), and urokinase share common structures with disulfide bonds in the same positions as those present in kringle and growth factor domains (McMullen & Fujikawa, 1985). Factor XII also contains the type I and type II domains of fibronectin, while t-PA includes the type I domain of fibronectin in addition to two kringle domains. The vitamin K dependent plasma proteins (factors VII, IX, and X, protein C and protein S) contain two (or more) growth factor domains (Hedner & Davie, 1989). The heavy chains of these serine proteases have domains with binding functions that complement their enzyme activities. However, each domain does not have a specific functional role. For instance, the kringle structure was thought to be a binding domain to fibrin because it was present in fibrin binding proteins, such as plasminogen (Magnusson et al., 1976) and t-PA (Pennica et al., 1983). This domain, however, is also present in factor XII, urokinase, and prothrombin (Magnusson et al, 1975) as well as the hepatocyte growth factor (Nakamura et al., 1989), and these proteins are not known to be fibrin binding proteins.

Factor XI, which is highly homologous to plasma prekallikrein (Fujikawa et al., 1986), also has four apple domains in the amino-terminal region of the protein, analogous to plasma prekallikrein [see the following paper in this issue (McMullen et al., 1991)]. The two proteins also have other common biological characteristics in that both are substrates for factor XIIa (Griffin & Cochrane, 1979; Heimark et al., 1980; Bouma & Griffin, 1977; Kurachi & Davie, 1977) and bind to high molecular weight kininogen (Mandle et al., 1976). Furthermore, the binding of each protein to high molecular weight kininogen occurs through the amino-terminal portion of their molecule (van der Graaf, 1982, 1983), indicating a role for the apple domains in this reaction. Recently, Baglia et al. (1990) reported that the binding site of factor XI to HMW kininogen is located in the latter half of the first repeat (Phe_{56} to Ser_{86}), suggesting that the first apple domain is involved. The heavy chains of plasma kallikrein and factor XIa may also be important in the recognition of their protein substrates, i.e., factor XII, high molecular weight kininogen, and factor IX, since the light chains of plasma kallikrein and factor XIa retain their peptidase activity but lose their activity toward protein substrates when they participate in the blood coagulation cascade (van der Graf, 1982, 1983).

ACKNOWLEDGMENTS

We are grateful to Santosh Kumar and Drs. Harry Charbonneau and Joost C. M. Meijers for helpful discussions and to Lois Swenson and Dolly Williams for help in the preparation of the manuscript.

Registry No. Prekallikrein, 9055-02-1.

REFERENCES

Baglia, F. A., Jameson, B. A., & Walsh, P. N. (1990) *J. Biol. Chem. 265*, 4149–4154.

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

Chung, D. W., Fujikawa, K., McMullen, B. A., & Davie, E. W. (1986) *Biochemistry 25*, 2410–2417.

Fujikawa, K., Chung, D. W., Hendrickson, L. E., & Davie, E. W. (1986) *Biochemistry 25*, 2417–2424.

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

Hedner, U., & Davie, E. W. (1989) in *The Metabolic Basis of Inherited Diseases* (Scriver, C. R., Beaudet, A. L., Sly, W. S., & Valle, D., Eds.) 6th ed., pp 2107–2134, McGraw-Hill, New York.

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

Hojrup, P. & Magnusson, S. (1987) *Biochem. J. 245*, 887–892.

Ichinose, A., Fujikawa, K., & Suyama, T. (1986) *J. Biol. Chem. 261*, 3486–3489.

Kauffman, D. L. (1965) *J. Mol. Biol. 12*, 929–932.

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

Laake, K., & Vennerod, A. M. (1974) *Thromb. Res. 4*, 285–302.

Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Cleays, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) pp 123–149, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Dudek-Wojciechowska, G., & Cleays, H. (1976) in *Proteolysis and Physiological Regulation* (Robbins, D. W., & Brew, K., Eds.) pp 203–238, Academic Press, New York.

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

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

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

Marti, T., Rosselet, S. J., Titani, K., & Walsh, K. A. (1987) *Biochemistry 26*, 8099–8109.

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

McMullen, B. A., Fujikawa, K., & Davie, E. W. (1991) *Biochemistry* (following paper in this issue).

Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K. & Shimizu, S. (1989) *Nature 342*, 440–443.

Ogston, D., Ogston, C. M., Ratnoff, O. D., & Forbes, C. D. (1969) *J. Clin. Invest. 48*, 1786–1801.

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

Ryle, A. P., & Sanger, F. (1955) *Biochem. J. 60*, 535–540.

Saito, H., Ratnoff, O. D., & Donaldson, V. H. (1974) *Circ. Res. 34*, 641–651.

Spackman, D. H., Stein, W. H., & Moore, S. (1960) *J. Biol. Chem. 235*, 648–659.

Sueyoshi, T., Miyata, T., Iwanaga, S., Toyo'oka, T., & Imai, K. (1985) *J. Biochem. (Tokyo) 97*, 1811–1813.

Tait, J. F., & Fujikawa, K. (1987) *J. Biol. Chem. 262*, 11651–11656.

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

van der Graaf, F., Greengard, J. S., Bouma, B. N., Kerbiriou, D. M., & Griffin, J. H. (1983) *J. Biol. Chem. 258*, 9669–9675.
Weiss, A. S., Gallin, J. I., & Kaplan, A. P. (1974) *J. Clin. Invest. 53*, 622–633.
Werle, E. (1955) in *Polypeptides which Stimulate Plain Muscle* (Gaddum, J. H., Ed.) pp 20–27, E. and S. Livingstone, Edinburgh, Scotland.
Wiman, B. (1977) *Eur. J. Biochem. 76*, 129–137.
Wuepper, K. D. (1973) *J. Exp. Med. 138*, 1345–1355.

Location of the Disulfide Bonds in Human Coagulation Factor XI: The Presence of Tandem Apple Domains†

Brad A. McMullen, Kazuo Fujikawa, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received September 24, 1990; Revised Manuscript Received November 19, 1990

ABSTRACT: Factor XI is a plasma glycoprotein that participates in the blood coagulation cascade. Of the 19 disulfide bonds present in each of the subunits of the human protein, 16 were determined by amino acid sequence analysis of peptide fragments produced by chemical and enzymatic digestion. Four apple domains of 90 or 91 amino acids were identified in the tandem repeats present in the amino-terminal portion of each subunit of factor XI. The disulfide bonds in the carboxyl-terminal portion of the molecule were similar to those in the catalytic region of other serine proteases. The two identical subunits of factor XI were connected by a single disulfide bond at Cys_{321} linking each of the fourth apple domains while each of the Cys residues at position 11 in the first apple domains forms a disulfide bond with another Cys residue.

Factor XI (M_r 143 000) is a zymogen of a plasma serine protease (factor XIa) that participates in the early phase of the blood coagulation cascade (Walsh, 1985; Schmaier et al., 1987; Fujikawa & Saito, 1989). It circulates in blood as a equimolar complex with high molecular weight (HMW) kininogen (Thompson et al., 1977). Factor XI is a unique protease precursor in that it is composed of two identical subunits held together by a disulfide bond(s). During the conversion of factor XI to factor XIa, an internal Arg–Ile bond in each subunit is cleaved to produce two amino-terminal heavy chains and two carboxy-terminal light chains, and these four chains are held together by disulfide bonds. Factor XIa contains two catalytic sites, and each is inhibited by antithrombin III (Kurachi & Davie, 1977).

The primary structure of factor XI deduced by partial amino acid sequence analysis and cDNA sequence analysis showed that it is 58% identical with plasma prekallikrein (Fujikawa et al., 1986; Chung et al., 1986). Like plasma prekallikrein, the amino-terminal region of each subunit of factor XI contains 4 tandem repeats of 90 (or 91) amino acids while the carboxyl-terminal region contains the catalytic triad of His, Asp, and Ser that is characteristic of serine proteases. The amino-terminal region also contains the binding site to high molecular weight kininogen and its substrate, factor IX (van der Graaf et al., 1983; Baglia et al., 1989; De La Cadena et al., 1988). Recently, Baglia et al. (1990) reported that a synthetic peptide comprising a region of factor XI between residues Phe_{56} and Ser_{86} is important in the binding reaction since it inhibited the interaction of HMW kininogen with factor XI.

Each of the four amino-terminal repeats in the subunits of factor XI contain six half-cystine (Cys)[1] residues at highly conserved positions. The first and the fourth repeats also contain an additional Cys residue, and these residues are thought to be involved in interchain disulfide bonds linking the two identical subunits (Fujikawa et al., 1986). In order to clarify the structure of the tandem repeats in the molecule and the bonds holding the two subunits together, it was important to establish the location of the disulfide bonds present in this plasma glycoprotein. In the present report, the positions of 16 of the 19 disulfide bonds were established by amino acid sequence analysis of various peptides isolated from enzymatic and chemical digests of the protein.

Experimental Procedures

Materials. Human factor XI was purified by a modified method of Naito and Fujikawa (1991) using a high molecular weight kininogen peptide affinity column. Thermolysis was purchased from Boehringer Mannheim, and the TSK-G3000SW column was from LKB–Produkter AB. All other materials were the same as described in the preceding paper (McMullen et al., 1991).

Chemical and Enzymatic Digestions. The conditions employed for the protein and peptide digestions were essentially the same as those used in the preceding paper (McMullen et al., 1991) except for the following modifications. Acid cleavage at aspartyl bonds was conducted in 0.01 N HCl, pH 2.0, instead of 2% formic acid. Peptides were digested by trypsin at mass ratios of 1/100 or 1/10. The digestion of thermolysin was performed in 0.1 M ammonium formate, pH 6.5, containing 2 mM $CaCl_2$, by incubation at 60 °C for 1 h at a mass ratio of 1/30.

Purification of Peptides. The cyanogen bromide digest of factor XI was fractionated by using TSK-G3000SW columns connected to an LKB HPLC system. Other peptides were separated by a Waters μBondapak C-18 reverse-phase column connected to a Waters HPLC system under the same condi-

† This work was supported in part by Grant HL16919 from the National Institutes of Health. A preliminary report was presented at the Fourth Annual Symposium of the Protein Society held in San Diego, CA, Aug 11–15, 1990.

[1] Abbreviations: Cys, half-cystine; PTH, phenylthiohydantoin.

0006-2960/91/0430-2056$02.50/0