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

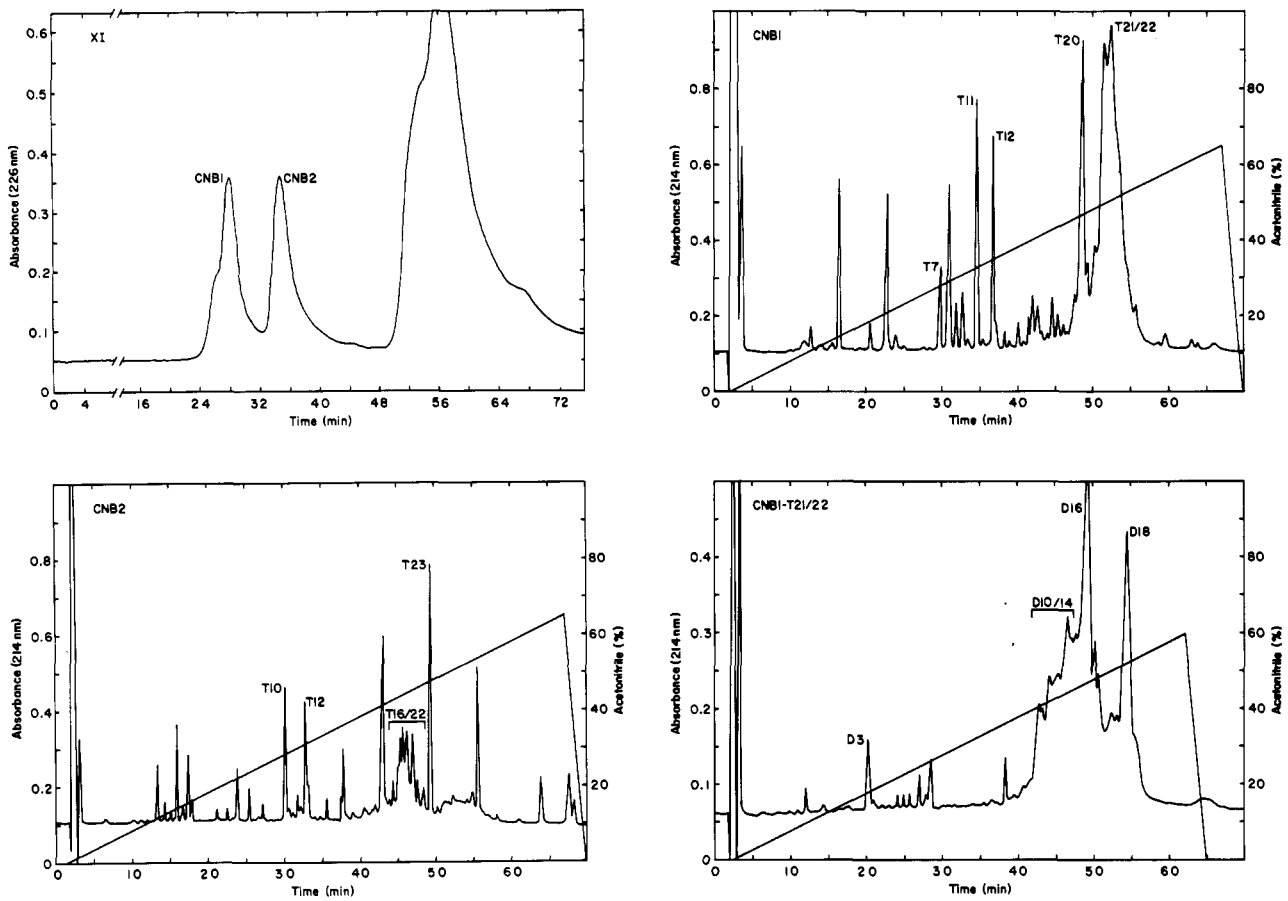

FIGURE 1: Fractionation of peptides derived from human factor XI. (Top left panel) Gel filtration of the cyanogen bromide digest of the protein on two TSK-G3000SW columns connected in tandem. The effluent was monitored for peptides at 226 nm, and fractions were collected manually at a flow rate of 1 mL/min. Peptides CNB1 and CNB2 contained Cys residues as detected by the fluorescence assay. (Top right panel) Fractionation of the tryptic digest of CNB1 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 a linear gradient was employed for elution, as described under Experimental Procedures. The concentration of acetonitrile is indicated on the ordinate. Peptides CNB1-T7, -T11, -T12, -T20, and -T21/22 contained Cys residues, as detected by the fluorescence assay or sequence analysis. (Lower left panel) Separation of peptides from the tryptic digestion of CNB2 employing a C-18 column. The conditions employed are the same as described in the top right panel. Peaks that are labeled contained Cys residues, as detected by the fluorescence assay. (Lower right panel) Separation of peptides obtained by acid cleavage of CNB1-T21/22 on a C-18 column.

tions as described in the preceding paper (McMullen et al., 1991).

Analytical Methods. Oxidation of a cystinyl peptide or intact factor XI was performed according to Hirs (1967), and the amino acid composition of the oxidized samples was determined by a Waters Picotag system as described by Bidlingmeyer et al. (1984). The molecular weight of one of the peptides (CNB1-T20) was determined by a Pharmacia Phast Gel system using an 8–25% gradient gel. The gel was run and silver-stained according to the supplier's instruction. Cystinyl peptides were detected by the procedure of Sueyoshi et al. (1985). Other methods used in this study are essentially the same as described in the preceding paper (McMullen et al., 1991).

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; D/T, acid cleavage followed by trypsin digestion; Th, thermolysin digestion; C, chymotrypsin digestion.

Results and Discussion

Purified factor XI was cleaved at the five methionine residues present in each identical subunit with cyanogen bromide, and the digest was separated into three major peaks by gel filtration on two TSK-G3000SW columns connected in tandem (top left panel, Figure 1). Two of the peaks, labeled CNB1 and CNB2, contained cystinyl peptides as shown by the fluorescence assay (Sueyoshi et al., 1985). Amino acid sequence analysis revealed that CNB1 contained the amino-terminal portion of the molecule and CNB2 contained the light chain or serine protease portion of factor XI. The two cyanogen bromide peptides were produced by the cleavage at Met_{358}, a residue located just after the 4 tandem repeats and 11 residues prior to the activation cleavage site (Figure 2).

CNB1 and CNB2 were then digested with trypsin in 0.1 M ammonium formate, pH 6.5, at a mass ratio of 1/100, and the fragments from each digest were separated by reverse-phase HPLC using a C-18 column. A number of peptides separated from the CNB1 digest including T7, T12, T20, and T21/22 contained Cys residues as shown by the fluorometric assay (top right panel, Figure 1). Sequence analysis of CNB1-T7 established a disulfide bond between Cys_{118} and Cys_{147} in the second tandem repeat (Table I, Figure 2). CNB1-T12 was found to have the same sequence as CNB1-T7, but continued beyond Arg_{121} indicating that a partial digestion of the peptide had occurred. When CNB1-T12 was cleaved with trypsin at a higher mass ratio (1/10), two peaks appeared corresponding to T7 and T11. Although CNB1-T11 was negative in the fluorescence assay, sequence analysis revealed a single sequence that linked Cys_{122} to Cys_{128} in the

FIGURE 2: Location of the disulfide bonds in the subunit of human factor XI. The four apple domains in the amino-terminal portion of the protein are labeled A_1, A_2, A_3, and A_4. The asterisk on Cys_{321} in the fourth apple domain identifies the disulfide bond between the two identical subunits. The Arg_{369}–Ile_{370} bond that is cleaved by factor XIIa during the activation reaction is marked by a curved arrow. The circled residues (His_{413}, Asp_{462}, and Ser_{557}) are members of the catalytic triad typical of serine proteases. The solid diamond on Asn_{473} indicates a carbohydrate site that was identified in the present studies. The four open diamonds on Asn residues at positions 72, 108, 335, and 432 are potential sites for the attachment of carbohydrate chains. The amino- and carboxyl-terminal residues are identified by -NH_2 and -COOH, respectively.

second repeat (Table I, Figure 2). The studies on the remaining two peaks, CNB1-T20 and T21/22, are described later.

Cyanogen bromide fragment CNB2 contained the entire serine protease portion of factor XI in addition to a short segment of 11 amino acids (residues 359–369, Figure 2), at the amino-terminal end. By analogy with other serine proteases, it was highly likely that this short segment contained an interchain disulfide bridge between the heavy chains (Cys_{362}) and light chains (Cys_{482}) of factor XIa. Accordingly, it appeared likely that CNB2 contained 10 Cys residues that formed 5 disulfide bonds (Fujikawa et al., 1986). Fluorometric analysis of the fractions from the trypsin digest of CNB2 showed that T10, T12, and T23 contained Cys residues (lower left panel, Figure 1) while peptides T16 through T22 gave a week signal by the fluorescence analysis. Sequence analysis of CNB2-T10 showed a double sequence and established a disulfide bond between Cys_{362} and Cys_{482} (Table I, Figure 2). Similarly, a double sequence of CNB2-T12 paired Cys_{527} to Cys_{542}. A single sequence was seen for CNB2-T23 and established the third disulfide bond between Cys_{398} and Cys_{414}. The loop formed by this disulfide bond contained His_{413} of the catalytic triad. Peaks T16 through T22 contained a fragment(s) that held the remaining two disulfide bonds in the catalytic chain. By analogy with the known disulfide bond structures of other serine proteases such as trypsin (Kauffman, 1965) and the B-chain of plasmin (Wiman, 1977), these two linkages were assigned between Cys_{496}–Cys_{563} and Cys_{553}–Cys_{581} (Figure 2). In this region of the protein, there are four Trp residues which may have been partially oxidized during the treatment with cyanogen bromide and/or subsequent handling of the peptide(s). Accordingly, it appears likely that partially oxidized tryptophan residues resulted in peptide fragments that elute at several places from the column (T16 through T22). Unfortunately, fractions T16–T22 were also contaminated with several other fragments. Attempts to isolate clean peptides after further digestion resulted in poor recovery. Therefore, these two disulfide linkages were not determined.

Four of the disulfide bonds in the first and third tandem repeats in the amino-terminal region of the protein were established from the fragments produced by the acid cleavage of CNB1-T21/22 (lower right panel, Figure 1). Sequence analysis of CNB1-T21/22-D16 paired Cys_{28} to Cys_{58} and Cys_{32} to Cys_{38} in the first repeat (Table I, Figure 2). CNB1-T21/22-D18 gave a single sequence that began at Ala_{205} and continued beyond the third Cys at position 218. Trypsin, at a mass ratio of 1/10, was used to cleave between the disulfide bonds in this fragment. After separation of the tryptic digest by HPLC (data not shown), two cystinyl peptides were isolated. CNB1-T21/22-D18-T4 gave a double sequence

Table I: Peptides Isolated and Sequenced for the Identification of the Disulfide Bonds in Human Factor XI

Peptide	Amino Acid Sequence[a]	Position of the Disulfide Bond(s)
CNB1-T21/22-D3	T C F E G G X	11-Cys
CNB1-T21/22-D16	I T T V F T P S A K Y C Q V V X T Y H P R X L L F T F T A P T R W F T X V L K	28-58 and 32-38
CNB1-T18-D10/14-T20-Th6	I S A X N K S X A	92-175
CNB1-T7	S A Q E C Q E R N I X L L K	118-147
CNB1-T11	X T D D V H C H F F T Y A T R	122-128
CNB1-T18-D10/14-T20-Th9	L A X L Q S X R	182-265
CNB1-T21/22-D18-T4	A F V C G R N L X L L K	208-237
CNB1-T21/22-D18-T6	I X T H H P G X L F F T F F S Q E(W)P K	212-218
CNB1-T20-D/T12	H S I P V F C H S S F Y H L X K	273-356
CNB1-T20-D/T6	S H E A X Q K X Y L K	299-328
CNB1-T20-D/T16-C2	L C T N A V R X Q F F	303-309
CNB2-T10	D N E X T T K L E T T V(N)Y T D S Q R P I X L P S K G D R	362-482
CNB2-T23	H L X G G S I I G N Q W I L T A A H X F Y G V E	398-414
CNB2-T12	A K I P L V T N E E C Q K R I X A C Y R	527-542

[a] Sequence analyses of 14 peptides containing disulfide bonds are shown. The positions that were not identified but 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.

and established a linkage between Cys_{208} and Cys_{237} in the third repeat. A single sequence was seen when CNB1-T21/22-D18-T6 was examined and paired Cys_{212} to Cys_{218}.

It was previously suggested that the two subunits of factor XI were linked by two disulfide bonds between the heavy chains, since each of the first and fourth tandem repeats contained an extra Cys residue (Fujikawa et al., 1986). These linkages were either between the same repeats (first/first and fourth/fourth) or between the two different repeats (first/fourth and fourth/first). The former pair of disulfide bonds would form *parallel* and the latter an *antiparallel* structure between the two subunits (Fujikawa et al., 1986). CNB1-T21/22-D3 from the first repeat (lower right panel, Figure 1) contained one of the extra Cys residues. Sequence analysis of CNB1-T21/22-D3 gave a single sequence corresponding to residues 10–15 (Figure 2) with a diPTH-cystine peak appearing in the second turn. This suggested that CNB1-T21/22-D3 formed a dimer with itself. However, amino acid analysis of the oxidized sample gave the following molar composition: cysteic acid, 2.1; Glu, 1.0; Gly, 2.0; Thr, 1.0; Phe, 0.9. This composition agrees with that obtained from the peptide sequence except for the presence of 2 mol of Cys. These results indicate that Cys_{11} was bound to a single Cys residue. The presence of this Cys residue in both subunits was confirmed by the analysis of cysteic acid of oxidized intact factor XI without sample hydrolysis. In these experiments, 1.2 mol of cysteic acid was found per mole of subunit of factor XI.

The second extra Cys residue present in the fourth repeat was identified in CNB1-T20 (top right panel, Figure 1). Sequence analysis of this peptide gave a single sequence of His-Ser-Ile-Pro-Val beginning with His_{267}. Since no other sequence was observed, it was highly likely that CNB1-T20

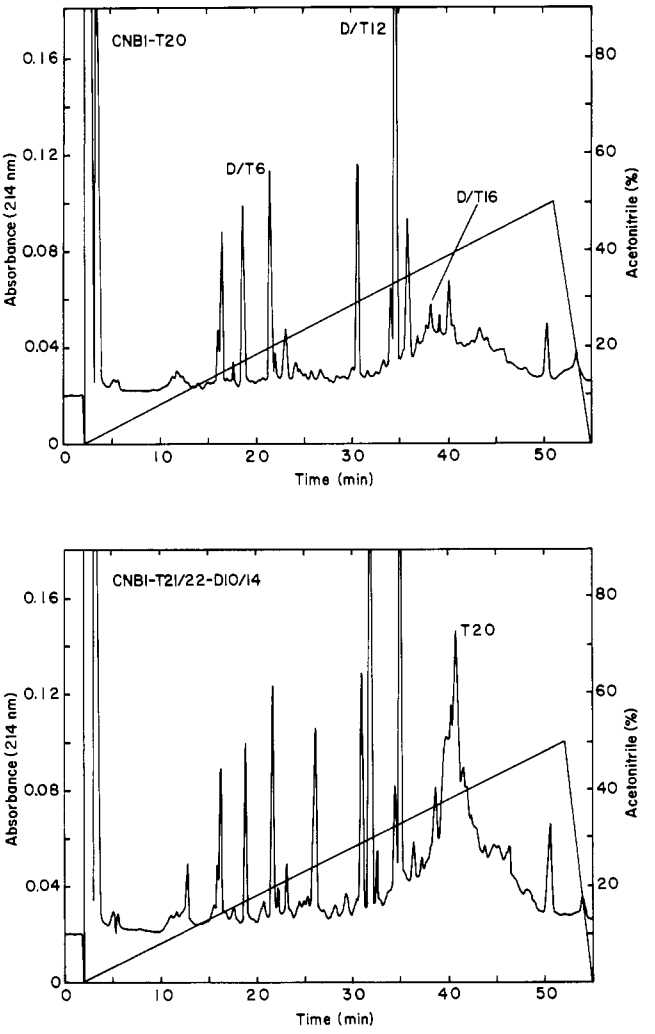

FIGURE 3: Separation of acid and/or trypsin digests on a C-18 column. (Top panel) Separation of peptides obtained by acid cleavage followed by trypsin digestion of CNB1-T20. (Bottom panel) Separation of peptides resulting from the digestion of CNB1-T21/22-D10/14 with trypsin.

contained the intact fourth repeat (His_{267} to Met_{358}, Figure 2) including the second extra Cys residue. Peptides were then isolated after cleaving CNB1-T20 with acid followed by trypsin digestion at a mass ratio of 1/10 to establish the location of the disulfide bonds in this fragment (top panel, Figure 3). Sequence analysis of CNB1-T20-D/T6 and D/T12 produced double sequences that paired Cys_{299} to Cys_{328} and Cys_{273} to Cys_{356}, respectively (Table I, Figure 2). Thus, four of the seven Cys residues in the fourth tandem repeat were identified. The remaining three Cys residues were found in CNB1-T20-D/T16. A chymotryptic digest of this fragment produced CNB1-T20-D/T16-C9 (data not shown), which showed that a disulfide bond existed between Cys_{303} and Cys_{309}. The other fragment that contained the last Cys at position 321 was isolated but was contaminated. However, it seemed likely that this Cys formed an interchain disulfide bond between the fourth repeats of the two subunits since the remaining six Cys residues were accounted for in other disulfide bonds. To determine if CNB1-T20 was involved in linking the subunits of factor XI together, an SDS–PAGE gel was run on reduced and unreduced samples of CNBI-T20 (Figure 4). A molecular weight of 19 500 was obtained for the unreduced sample and 11 000 for the reduced sample. These experiments clearly showed that the two fourth repeats formed a dimer. The molecular weight of 11 000 for the reduced sample agreed well

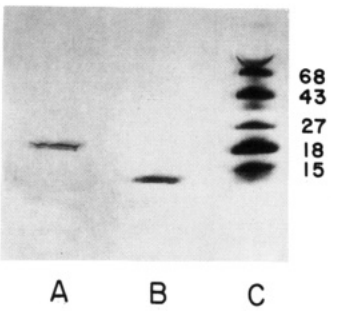

FIGURE 4: SDS–polyacrylamide gel electrophoresis of peptide CNB1-T20. Bands were visualized by silver staining. (A) Nonreduced sample; (B) reduced sample; (C) molecular weight standard under reducing conditions. The numbers refer to the molecular mass of the standard proteins in kilodaltons.

with the composition of the fourth repeat, which contained 91 residues (His_{267} through Met_{358}) and a potential carbohydrate chain at Asn_{335}. Plasma prekallikrein also has a Cys residue at the same position in the fourth tandem repeat, but this Cys forms a small peptide loop by its linkage to Cys_{326} (McMullen et al., 1991).

The final peptide fragment of interest, that held the first three tandem repeats together, was isolated from a shoulder appearing on the ascending side of CNB1-T21/22-D16 (lower right panel, Figure 1). Fractions D10–D14 were pooled, cleaved with trypsin at a mass ratio of 1/10, and separated by HPLC (bottom panel, Figure 3). Besides numerous fragments that were the same as those obtained in the digest of the fourth repeat (top panel, Figure 3), CNB1-T21/22-D10/14-T20was isolated. This fragment contained the Cys residues linking the first, second, and third repeats together. It was then digested with thermolysin, and the products were separated by HPLC. Twenty-seven were isolated of which the two major peaks, CNB1-T21/22-D10/14-T20-Th4 and Th6, paired Cys_{92} to Cys_{175} in the second repeat and Cys_{182} to Cys_{265} in the third repeat (Table I).

In this study, the location of 16 of the 19 disulfide bonds present in subunits of factor XI has been established. Three disulfide bonds could not be determined because of the difficulty in obtaining adequate amounts of pure peptides to determine these bonds. By analogy with plasma prekallikrein and other serine proteases, it is clear that these bonds were located between Cys_2 and Cys_{85}, Cys_{497} and Cys_{563}, and Cys_{553} and Cys_{581}. In the heavy chain, the six conserved Cys residues in the four repeats were connected by disulfide bonds linking the first/sixth, second/fifth, and third/fourth Cys residues. Accordingly, these repeats form apple domains that are the same as those in plasma prekallikrein (McMullen et al., 1991). The final structure of the factor XI subunit is illustrated in Figure 2.

During the sequence analysis of CNB2-T10, no PTH-amino acid was observed in cycle 10 of this peptide. This is consistent with an apparent glycosylation of Asn at position 473. Beside this position, four other N-linked carbohydrate attachment sites (positions 72, 108, 335, and 432) were predicted from the cDNA analysis of human factor XI (Fujikawa et al., 1986). In the present studies, however, they were not determined since these sites were not present in any of the cystinyl peptides that were isolated.

Secreted plasma proteins usually do not contain free SH groups. In a congenitally abnormal plasma protein, antithrombin III_{Toyama} (Koide et al., 1984), a mutation of Arg_{47} to Cys resulted in an odd number of Cys residues as determined by amino acid sequence analysis of an abnormal peptide. It was suggested that this unpaired residue formed a disulfide bridge with another Cys residue since amino acid analysis of the protein showed the presence of 8 mol of Cys per mole of protein. Whether this Cys is a single amino acid or is found in a small peptide such as glutathione was not determined. This is analogous to Cys_{11} in factor XI that forms a disulfide bridge with a single Cys residue. This modification probably occurred after the translation of the protein and is most likely nonfunctional. It is not known whether dimerization through the fourth repeat is important for the functional activity of factor XI. Studies are presently in progress in our laboratory to further examine this question by expression of appropriate mutants of factor XI.

ACKNOWLEDGMENTS

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

Registry No. Factor XI, 9013-55-2.

REFERENCES

Baglia, F. A., Sinha, D., & Walsh, P. N. (1989) *Blood 74*, 244–251.

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

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93–104.

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

De La Cadena, R. A., Baglia, F. A., Johnson, C. A., Wenk, R. E., Amernick, R., Walsh, P. N., & Colman, R. W. (1988) *Blood 72*, 1748–1754.

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

Fujikawa, K., & Saito, H. (1988) *The Metabolic Basis of Inherited Disease* (Scriver, C. R., Beaudet, A. L., Sly, W. S., & Valle, D., Eds.) 6th ed., pp 2189–2206, McGraw-Hill, New York.

Hirs, C. H. W. (1967) *Methods Enzymol. 11*, 59–62.

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

Koide, T., Odani, S., Takahashi, K., Ono, T., & Sakuyragawa, N. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 289–293.

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

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

Naito, K., & Fujikawa, K. (1991) (submitted for publication).

Schmaier, A. H., Silverberg, M., Kaplan, A. P., & Colman, R. W. (1987) *Hemostasis and Thrombosis* (Colman, R. W., Hirsh, J., Marder, V. J., & Salzman, E. W., Eds.) 2nd ed., pp 18–38, J. B. Lippincott Co., Philadelphia.

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

Thompson, R. E., Mandle, R., Jr., & Kaplan, A. P. (1977) *J. Clin. Invest. 60*, 1376–1380.

van der Graaf, F., Greengard, J. S., Bouma, B. N., Kerbiriou, D. M., & Griffin, J. H. (1983) *J. Biol. Chem. 258*, 9669–9675.

Walsh, P. N. (1985) *Semin. Hematol. 22*, 178–186.

Wiman, B. (1977) *Eur. J. Biochem. 76*, 129–137.