

Proclotting Enzyme from Horseshoe Crab Hemocytes

cDNA CLONING, DISULFIDE LOCATIONS, AND SUBCELLULAR LOCALIZATION*

(Received for publication, June 27, 1990)

Tatsushi Muta‡, Ryuji Hashimoto§, Toshiyuki Miyata‡§, Hitoshi Nishimura‡, Yoshihiro Toh§, and Sadaaki Iwanaga‡§¶

From the ‡Department of Molecular Biology, Graduate School of Medical Science and the §Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Proclotting enzyme is an intracellular serine protease zymogen closely associated with an endotoxin-sensitive hemolymph coagulation system in limulus. Its active form, clotting enzyme, catalyzes conversion of coagulogen to insoluble coagulin gel. We present here the cDNA and amino acid sequences, disulfide locations, and subcellular localization of proclotting enzyme. The isolated cDNA for proclotting enzyme consists of 1,501 base pairs. The open reading frame of 1,125 base pairs encodes a sequence comprising 29 amino acid residues of prepro-sequence and 346 residues of the mature protein with a molecular mass of 38,194 Da. Three potential glycosylation sites for N-linked carbohydrate chains were confirmed to be glycosylated. Moreover, the zymogen contains six O-linked carbohydrate chains in the amino-terminal light chain generated after activation. The cleavage site that accompanies activation catalyzed by trypsin-like active factor B, proved to be an Arg-Ile bond. The resulting carboxyl-terminal heavy chain is composed of a typical serine protease domain, with a sequence similar to that of human coagulation factor XIa (34.5%) or factor Xa (34.1%). The light chain has a unique disulfide-knotted domain which shows no significant homology with any other known proteins. Thus, this proclotting enzyme has a mammalian serine protease domain and a structural domain not heretofore identified in coagulation and complement factors. Immunohistochemical studies showed that the proclotting enzyme is localized in large granules of hemocytes.

Limulus has a unique endotoxin (lipopolysaccharide, LPS)¹-sensitive hemolymph coagulation system thought to be one defense system against invading microorganisms (1). The Gram-negative bacteria contain in the outer membrane various LPS analogues. When limulus is infected by Gram-negative bacteria, such as *Vibrio cholerae*, at the time of injury, hemocytes circulating in the hemolymph recognize them and responses rapidly occur, including cell adhesion, aggregation, and degranulation (2). Following degranulation of the large

granules in hemocytes, hemolymph coagulation is initiated by the granular components (3). Several components associated with this coagulation reaction have been purified, and reconstitution experiments using the purified clotting factors suggested to us a cascade-type of reaction (4).

In the presence of LPS, a LPS-sensitive serine protease zymogen, factor C, is autocatalytically activated (5). The active factor C then activates zymogen factor B (6) to factor B, which subsequently activates proclotting enzyme (7) to clotting enzyme. The resulting clotting enzyme converts soluble coagulogen (8), an invertebrate fibrinogen-like substance, to an insoluble coagulin gel. Like the mammalian coagulation and complement systems, this cascade reaction is propagated by limited proteolysis. On the other hand, the hemocyte lysate is known to respond to (1,3)- β -D-glucan as well as to LPS (9). This (1,3)- β -D-glucan-sensitive system also activates the proclotting enzyme through a glucan-sensitive factor(s), factor G(s), resulting in the formation of coagulin gel. Because (1,3)- β -D-glucan is a major cell wall constituent of yeast and other fungi, this cascade reaction also seems to play an important role in biological defense mechanisms.

In 1985, the proclotting enzyme from the Japanese horseshoe crab *Tachypleus tridentatus* was purified (7). This protein is a single chain glycoprotein with molecular mass of 54 kDa. Upon activation by factor B, it is converted to a two-chain active form clotting enzyme composed of a light (L, 25 kDa) and a heavy (H, 31 kDa) chain (7). The clotting enzyme cleaves two sites in coagulogen and forms a coagulin gel. Thus, proclotting enzyme/clotting enzyme seems to be a prothrombin/ α -thrombin counterpart to invertebrates, since both catalyze a final step for gelation. Interestingly, the clotting enzyme has a substrate specificity similar to that of mammalian coagulation factor Xa and can catalyze the conversion of bovine prothrombin to α -thrombin at a reasonable rate (10).

We cloned a cDNA for the serine protease zymogen, factor C, an initiator of the coagulation cascade.² The structure of this protein revealed it to be a novel type of mosaic protein, hence a possible evolutionary relationship between the limulus coagulation and mammalian complement systems would have to be considered. We then investigated the structure of the proclotting enzyme in an attempt to acquire further information concerning the evolutionary origin of this protein and its biological role in the horseshoe crab coagulation system.

* This work was supported by a grant-in-aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed.

¹ The abbreviations used are: LPS, lipopolysaccharide; bp, base pairs; HPLC, high performance liquid chromatography; L, light; H, heavy; TPCK, L-(tosylamido 2-phenyl)ethyl chloromethyl ketone; PTH, phenylthiohydantoin.

² T. Muta, T. Miyata, Y. Misumi, F. Tokunaga, T. Nakamura, Y. Toh, Y. Ikehara, and S. Iwanaga, *J. Biol. Chem.*, submitted for publication.

EXPERIMENTAL PROCEDURES³

RESULTS

Partial Amino Acid Sequence—To determine a partial amino acid sequence, the *S*-pyridylethylated H and L chains derived from the clotting enzyme were separated by gel filtration followed by reverse-phase HPLC, as described under "Experimental Procedures." The isolated H chain was subjected to cyanogen bromide digestion, but the L chain was digested with trypsin, since it did not contain a methionine residue (7). The sequences of purified peptides derived from both chains were analyzed, in addition to the amino-terminal sequences of both chains. The sequences of these peptides are summarized in Table MI.

When the purified proclotting enzyme was subjected to amino acid sequence analysis, the recovery of the amino-terminal PTH-derivative was very low, although the partial amino-terminal sequence of the whole protein was identifiable, as shown in Table MI. This result suggests that almost all the amino-terminal residue of the intact protein is blocked. To confirm these observations, we examined a masked amino-terminal peptide in a trypsin digest of *S*-pyridylethylated proclotting enzyme (Fig. M1A). Among the 19 tryptic peptides obtained, the peptide Pe-PC-T21 was thought to be derived from the amino-terminal region, as its amino-terminal residue did not react with phenylisothiocyanate. As the amino acid composition (Table MII) indicated a 1 mol excess of glutamic acid, it was further digested with pyroglutamylpeptidase and the digest was fractionated by HPLC (Fig. M1B). The amino acid composition of the resulting peptide showed 1 mol less of glutamic acid (Table MII), and its amino-terminal sequence was identical to that obtained from sequence data on the intact protein (Table MIII). Based on these results, we conclude that the amino terminus of the intact protein is blocked by pyroglutamic acid.

cDNA Cloning—On screening 1.5 million clones of the hemocyte cDNA libraries with anti-clotting enzyme IgG, we found 23 positive clones. After sequence analyses, one was found to encode a partial amino acid sequence of the clotting enzyme and was named λ PC 17 (Fig. M2). Since this clone seemed to be derived from an *Eco*RI fragment of the cDNA, another preparation of the library was screened using λ PC 17 as a probe. The longest clone composed of 1,501 bp, named λ PC 11, was thus obtained (Fig. M2) and was then sequenced (Fig. 1). Following 216 bp of 5'-noncoding region, there is a long open reading frame encoding 375 amino acid residues, followed by a TAG termination codon and 157 bp of 3'-noncoding region, although both the poly(A) signal and poly(A) tail were missing.

Northern blot analysis of the hemocyte poly(A)⁺ RNA showed one band of about 1.5 kilobases that strongly hybridized with λ PC 17 (Fig. 2). Thus, the proclotting enzyme is indeed synthesized in the hemocyte and λ PC 11 contains almost the full length of the proclotting enzyme cDNA, apart from the poly(A) tail.

The deduced amino acid sequence from the first ATG initiation codon encoded all the partial amino acid sequences noted in the protein-based study (Fig. 1). The amino-terminal

³ Portions of this paper (including "Experimental Procedures," Figs. M1–M6, and Tables MI–MVIII) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press. The authors are also prepared to honor requests for copies of raw data (including HPLC patterns, the yields, the amino acid compositions, and the sequence data on all the peptides derived from proclotting enzyme).

ATACAGGCTACAAACATCTCATCAGACGCTACCTGGTGTGTTAAATCCCGAAAGCTTACGAGAACAGTCA	78
AACCTAGCTTGTGGTCCCGGACTGCTGCTGGAAGTAGTTCACAAATTCGGGCTTTATAATCAACCTAAA	153
CAAAACCGCTCTGAACTTTACCGTTCACATCCACCGAGGTAGCGGGTCTGCTCAGCAGATGTTGGTGAAT	228
N V F S I L C F P L L M S V V R C S T L S R Q R R	-26
AACGTGTTTCTACTGTGTTCCCACTCTTGATGTCTGGTGTAGATGCTAGTCTCAGCAGACGCTAGA	303
Q F V F P D E E E L C S N R F T E E G T C K N V L	-1
CAGTTGTTTCCCTGACGAGGAAGACTTGTCTCAACCGATTCTTCAAGGAAGAACATGCAAAATGTCTTG	378
GATTCTAGAAATCTTTACAAAAAATGATTATATTACTCAAGAAATCAATATCCGCTTTGAAGGCATAACA	453
D C R I L L Q K N D Y N L L K E S I C G F E G I T	50
CCCAAGTTTGTGTCGAAATCAAGCCATGTAATTCAGTACACAGGCACTCCAGAAACCTACGACTGAA	528
P K V C C P K S S H V I S S T Q A P P E T T T T E	75
CGCCACCAAAACAGATACCAACCAATCTCTGAGGTGTGGAAATTCACAATCTACCACTACGAGTATT	603
R P P K Q I P P N L P E V C G I H N T T T T R I I	100
GGAGCTCGGAAGCACTATTGAGGCTGGCGGTGATGCTGCTCTACATAAAGCAAGGAGTACAGAGT	678
G G R E A P I G A M P W M T A V V I K Q G G I R S	125
GTCAGTGTGGGCGCACTTCTCACTAACAGGCACTGATTACAGTCTCGCACTGTGTTGTAACAGTCAGGA	753
V Q C G A L V T N R H V I T A S C V V N S A G	150
ACAGATGTGATGCTGATGATTCTCGGTCTGCTGGTGAACACAAATTTATACAGTACCGATGAGTCTG	828
T D V M P A D V F S V R L G E H N L Y S T D D S	175
AATCAATAGATTTCGACTTACGTCGGTGAACATCAAGCACTTGTACTCGGCACTATTGAATGATCATC	903
N P I D E A V T S V K H H E H F V L A T Y L N I	200
GCAATTCATAGCTTAAATGACACAGTTACGTTTACAGACAGAAATTCGACCCATTTGTCTACCTTATGCTAAGTTG	978
A I L T L W D T V T F T D R I R P I C L P Y R K L	225
AGATAGCATGATCTAGCAATGAGAAACCGTTTATCACTGGATGGGGAACACAGCATTTAACGGCCATCTAGT	1053
R Y D D L A M R K P F I T G W G T T A F N G P S S	250
GCAGTGTGAGAGAGTACAGTTACCAATATGGGAACACAGGCTGTAGACAGGCTACGAGAGGATTAAAT	1128
A V L R E V Q L P I W E H E A C R Q A Y E K D L N	275
ATTACAAAGCTGTATATGTGCTGCTTTCAGATGGCGGAAGGATGCTTGCAGGCTGATTCGAGGTCCA	1203
I T N V Y M C A G F A D G G K D A C Q G D C G P	300
ATGATGTTGCTGTAAACCGGAGGTTTATCTCATTGGAATGTGTTTCGGAAGAAATGCCATGCTT	1278
M M L P V K T G E F Y L I G I V S F G K K C A L P	325
GGATTCTGCGGTTTACACAAAGTGACAGAGTTTATGATTGGATGCAACATATGGTGTAGCACTAAAGCT	1353
G F P G V Y T K V T E F L D W I A E H M V *	346
GTGAATGACTCATTTTCAATATTTAAACCACTACGCTGTAGCAAAAGCACTGAAGAATCAAGCGGGGACAA	1428
AACCACTTTTTTTTTTTTTTCTCAATAGTCTACCATGTATTGAAACCAACATGAACATCAATAA	1501

FIG. 1. Nucleotide and deduced amino acid sequences of proclotting enzyme. Nucleotide (upper) and deduced amino acid sequences (lower) are shown. The amino acid sequences are numbered from the amino terminus of the mature protein. The underlined amino acid residues were confirmed by protein sequencing. The cleavage sites accompanied by maturation and activation are shown by arrowheads. The potential attachment sites for *N*-linked carbohydrate chains are indicated by closed diamonds. The active triads of the serine protease domain are circled.

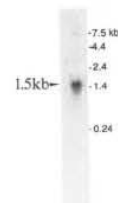


FIG. 2. Northern blot analysis of proclotting enzyme mRNA. Five μ g of poly(A)⁺ RNA from the hemocytes was separated by 1.0% agarose gel electrophoresis, transferred to a GVHP membrane, and hybridized with the λ PC 17 insert. See "Experimental Procedures" for details. kb, kilobase.

residue of the mature proclotting enzyme proved to be the 30th glutamine residue of the deduced sequence, in the form of pyroglutamic acid. Thus, the mature protein is composed of 346 amino acids, with calculated molecular mass of the protein portion 38,194 Da. From the amino-terminal sequence of the H chain (Table MI), the cleavage site for zymogen activation was identified to be an Arg⁹⁸-Ile⁹⁹. Therefore, the L chain is composed of 98 residues and the H chain of 248 residues. The amino acid composition of the proclotting enzyme, in addition to those of the L and H chains of clotting enzyme, agreed well with the sum of the total residues deduced from the cDNA sequence (Table MIV).

Three potential glycosylation sites, one in the L chain and two in the H chain, for *N*-linked carbohydrate chains (Asn-X-Ser/Thr) were present in the sequence. The carbohydrate

chains attached to these three asparagine residues will be described later.

Disulfide Locations—There are 14 cysteine residues in the sequence of the mature protein. When this protein was *S*-pyridylethylated, without reduction, *S*-pyridylethylcysteine was not detected in the amino acid analysis (data not shown). Thus, all these cysteines are linked by intramolecular disulfide bridges. Although the locations of disulfide linkages in the H chain could be assigned by analogy with those in trypsin, the linkages of 7 cysteine residues in the L chain could not be presumed, since there was no homologous protein. Thus, we decided to determine the complete disulfide linkage locations of this protein.

The purified proclotting enzyme and clotting enzyme were first digested, respectively, with lysyl endopeptidase, without reduction. All the procedures for the determination of disulfide locations, including enzyme digestions and separations of peptides, were carried out below pH 6.5, to avoid disulfide-exchange reactions. The digests were separated by reverse-phase HPLC (Fig. M3, A–C), and the isolated peptides were assigned based on the amino acid compositions (Table MV) and the sequence data (Table MVI), referring to the entire amino acid sequence predicted from the cDNA sequence for proclotting enzyme. These peptides were termed K1 to K18 from the amino terminus of the entire protein, as shown in Fig. 3.

Peptide K9 contains 2 cysteines, indicating a linkage between Cys¹²⁸ and Cys¹⁴⁴. The disulfide containing peptides K7–K10 and K14–K17 were also obtained, respectively, as a single peptide. Thus, the results indicate a Cys⁹⁹ and Cys²¹⁹ linkage in K7–K10, which holds together the L and H chains, and a Cys²⁹³ and Cys³²² linkage in K14–K17. Because the sequence linked between K11 and K12 was identified to be Lys-Pro, this bond was not cleaved by the enzyme. The peptide K11–K12–K13 was recovered as a single peptide, hence the Cys²⁶⁶ and Cys²⁸² are linked to each other.

The other cysteine containing peptide derived from the L chain, K1–K2–K4–K5, was isolated as a single peak, as shown in Fig. 3, A and B. Three individual amino-terminal sequences of K2, K4, and K5 could be detected in this peptide, as based on the sequence data, since K1 has a masked amino-terminal residue. This cysteine-containing peptide was further digested with trypsin, and the digest was separated by HPLC, and sequenced. Through these procedures, K4 was identified to be linked with the carboxyl-terminal part of K1 (K1–K2–K4–K5–T1) (data not shown). Thus, a linkage between Cys²¹ and Cys⁴⁴ was demonstrated. Among the remaining 4 cysteine residues to be determined, two exist in peptide K5, sequence -Cys-Cys-. A possible combination for the disulfide bridges is

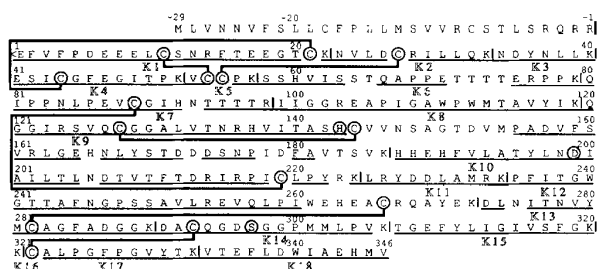


FIG. 3. Primary structure of the proclotting enzyme. Lysyl endopeptidase peptides derived from clotting enzyme are numbered K1 to K18 from the amino terminus. Underlined amino acid residues were confirmed by protein sequencing. Half-cysteines are shown by circles and disulfides between them are indicated by solid lines. Active site triads are shown by shadowed circles.

Cys¹¹-Cys⁵⁴ and Cys²⁷-Cys⁵⁵ or Cys¹¹-Cys⁵⁵ and Cys²⁷-Cys⁵⁴. To determine the correct combination, peptide K1–K2–K4–K5 was digested with endoproteinase Asp-N. The digest was separated by HPLC (data not shown), and peptide K1–K2–K4–K5–D2 was subjected to sequence analysis (Table MVII). Here, a relatively large amount of the sample was used to identify PTH-cystine. In the possible linkages between Cys¹¹-Cys⁵⁴ and between Cys²⁷-Cys⁵⁵, PTH-cystine would be identified at the third cycle, and in the other case, in the second cycle. The results showed that only the third and not the second cycle clearly yielded PTH-cystine (Fig. M4, A–D). Thus, we conclude that the linkages are between Cys¹¹ and Cys⁵⁴ and between Cys²⁷ and Cys⁵⁵, respectively.

Carbohydrate Chains—Amino acid analyses of the lysyl endopeptidase peptides indicated that three peptides, K6, K7–K10, and K11–K12–K13, contain hexosamines (Table MV). K6 was rich in galactosamine and had 1 serine and 5 threonine residues that appeared to be modified by *O*-linked carbohydrate chains, since these PTH-derivatives could not be detected in the sequence analyses (Tables MI and MIV). K11–K12–K13 had one potential *N*-linked glycosylation site, whereas K7–K10 had two. In the sequence analyses, all cycles corresponding to these 3 asparagines yielded no PTH derivative (Tables MI and MIV), thereby indicating that these residues are indeed glycosylated. K7–K10, which contained two glycosylated asparagines, was digested with trypsin and the resulting carbohydrate containing peptides were separated by HPLC (data not shown). Two peptides, K7–K10–T4 and K7–K10–T8, were isolated and both contained *N*-acetylglucosamine and *N*-acetylgalactosamine (data not shown). Four carbohydrate-containing peptides, K6, K7–K10–T4, K7–K10–T8, and K11–K12–K13, were obtained. In the HPLC profile, K6 and K7–K10–T4 were seen to have an apparent heterogeneity, presumably due to the carbohydrate chains, although it has been reported that the proclotting enzyme contains no sialic acid (7). The component sugar compositions of the four peptides were analyzed by derivatization with 2-aminopyridine. Under the conditions used (in 4 M trifluoroacetic acid at 100 °C for 3 h), hexosamines could not be quantitatively determined. However, all the peptides were confirmed to contain both *N*-acetylglucosamine and *N*-acetylgalactosamine (data not shown), in accord with the result of amino acid analysis. Monosaccharides derived from the *N*-linked carbohydrate chains in K7–K10–T4, K7–K10–T8, and K11–K12–K13 showed similar compositions rich in mannose and fucose. Galactose was also detected in K7–K10–T4 and K7–K10–T8. On the other hand, monosaccharides in K6 were distinct. Galactose and fucose are the major monosaccharides and K6 contains no mannose. A total of at least six *O*-linked carbohydrate chains could be found in the entire protein.

Subcellular Localization—To investigate the subcellular localization of proclotting enzyme in the hemocytes, an immunohistochemical technique was used. The hemocytes were first fixed with a mixture of paraformaldehyde and glutaraldehyde and then sections of these cells were stained with the purified anti-clotting enzyme IgG and gold particle-conjugated second antibody. When these preparations were analyzed by electron microscopy, numerous gold particles were observed in the granules (Fig. 4). Although two types of granules were apparent, large but less dense granules and smaller but dense ones, the gold particles were exclusively present in the former. This type of large granule is secreted from cells by stimulation with LPS (2). Thus, the proclotting enzyme is localized in the large granules and is apparently released through exocytosis on LPS stimulation.

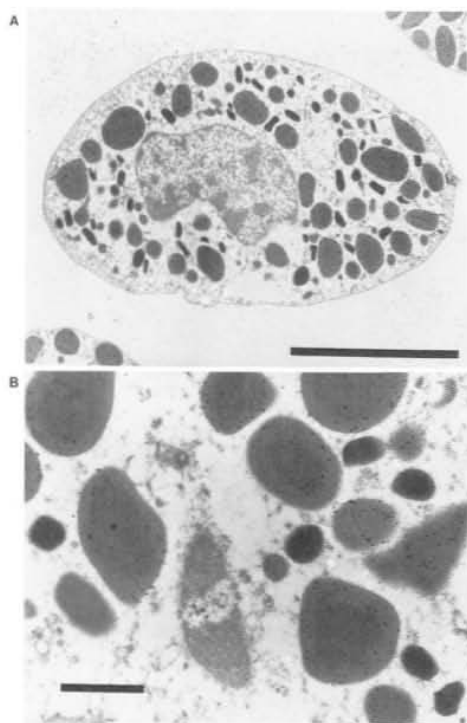


FIG. 4. Localization of proclotting enzyme in the cell. Hemocytes were fixed, stained with anti-clotting enzyme IgG and gold-particle-conjugated second antibody, and then scanned. Numerous gold particles are present in the larger but less dense granules. The bars are 5 (A) and 1 μ m (B), respectively.

DISCUSSION

From the above cDNA and protein-based studies, the nucleotide sequence and primary structure of proclotting enzyme was deduced, as shown in Figs. 1 and 3. The isolated cDNA consists of 1,501 bp, which is almost the full-length cDNA based on the results of Northern blotting. Although there is no termination codon and two ATG codons exist in the same frame of the upstream 5'-noncoding region, the first ATG codon is presumed to be a translation initiation codon, based on the following criteria; first, this clone has been shown to involve a full-length cDNA sequence, and second, the amino acid sequence from the first ATG codon shows characteristics of a signal sequence. Thus, the 29-amino acid residues (-29 to -1 in Fig. 1) preceding the amino-terminal glutamine residue of the mature protein seem to constitute a prepro-sequence. Using a method for predicting a signal sequence cleavage site (20), the most likely site cleaved by a signal peptidase is between Ser⁻¹³ and Val⁻¹² and the second possibility is between Ser⁻⁵ and Arg⁻⁴. The hydropathy profile calculated by the method of Kyte and Doolittle (21) indicates that this presequence region is highly hydrophobic (Fig. M5). The sequence between the mature protein and the signal peptide is assumed to consist of 12 (or 5) residues and appears to represent a propeptide region of the immature protein. In fact, the carboxyl-terminal sequence in this region shows the sequence -Arg-X-Arg-Arg (amino acid -4 to -1), similar to that found in mammalian serine protease zymogens, such as prothrombin (22), factors VII (23), IX (24), X (25), protein C (26), and complement factors C3 (27), C4 (28), and C5 (29). The presence of this sequence strongly suggests that an enzyme which proteolytically processes the dibasic cleavage site may also exist in the invertebrate hemocyte. The propeptide

region mentioned above may function to prevent unexpected activation of the proclotting enzyme, the result being clot formation within the cell.

The mature protein is composed of 346 amino acid residues with a molecular mass of 38,194. 85.5% of the total sequence was confirmed by protein sequencing, in which all the residues of the L chain were identified at the protein level, except for the carbohydrate-linked amino acid residues. The amino-terminal residue of the intact proclotting enzyme is masked by the pyroglutamyl residue. The purified material, however, contains a small amount of a molecular species with a second phenylalanine at the amino terminus. This suggests that the amino-terminal pyroglutamyl residue is partially removed during purification procedures and that a pyroglutamylpeptidase-like enzyme is present in horseshoe crab hemocytes.

The cleavage site associated with zymogen activation is Arg⁹⁸-Ile⁹⁹, the sequence of which connects the L and H chains. The sequence preceding this site has a unique structure, -Thr-Thr-Thr-Thr-Arg. This sequence agrees well with the fact that the horseshoe crab factor B shows an apparent affinity for the hydroxy amino acid at the P₂ site and efficiently hydrolyzes the synthetic peptide substrates *N*-tert-butoxycarbonyl-Met-Thr-Arg-4-methylcoumaryl-7-amide and benzoyl-Thr-Thr-Arg-4-methylcoumaryl-7-amide (6). The hydropathy profile shown in Fig. M5 indicates that the sequence around the cleavage site is relatively hydrophilic. Furthermore, the asparagine residue at the P₆ position found in the carboxyl-terminal portion of the L chain is glycosylated, which renders the region more hydrophilic (Fig. 1). Thus, the proclotting enzyme can be readily activated by trypsin (7).

The H chain of the proclotting enzyme is composed of a typical serine protease structure. To our knowledge, this is the first example of an invertebrate serine protease for which the disulfide locations have been determined (Fig. 3). The locations of four disulfide linkages in the H chain are the same as those in prothrombin (30) and factor IX (24). This clotting enzyme shows a substrate specificity similar to that of factor Xa (10). In fact, the entire sequence of the H chain corresponding to the serine protease region shows a close similarity to that of human factor Xa (34.1%), although the closest identity with the H chain is found in factor XIa (34.5%). It contains the His¹⁴³-Asp¹⁹⁹-Ser²⁹⁷ triad known to be the catalytic triad of serine proteases (Fig. M6). The substrate binding site corresponding to Ser¹⁸⁹, in chymotrypsin numbering, is aspartic acid, thereby indicating that this clotting enzyme has a typical trypsin-like specificity. Interestingly, the residue before His¹⁴³ has serine replacing alanine. The sequence around this histidine is strongly conserved in various serine proteases. However, some of the serine proteases from *Streptomyces griseus* are known to possess glycine at this position (Fig. M6). It is noteworthy that this enzyme does not contain the Ser-Trp-Gly sequence that is well conserved in the serine proteases and stabilizes the backbone of the substrate binding site. This sequence is replaced by Ser³¹⁷-Phe³¹⁸-Gly³¹⁹ in the clotting enzyme.

The L chain has no significant homology with any other protein thus far examined. The amino-terminal portion of the L chain consists of a novel disulfide-knotted structure, as shown in Fig. 5. This structure forms a clip-like shape consisting of a discrete domain. Actually, the clotting enzyme prepared by trypsin activation does not contain this portion and has Ser⁵⁸ as its amino terminus (Table MI). This would suggest that the carboxyl-terminal portion of the clip-like structure forms a hinge region susceptible to trypsin. The resulting smaller L chain may correspond to the L chain derivative (7). It seems likely that the disulfide-knotted do-

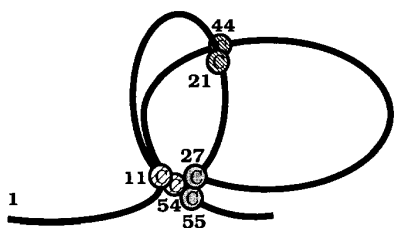


FIG. 5. Disulfide-knotted domain in the L chain of proclotting enzyme. The amino-terminal domain structure is schematically illustrated, based on its disulfide locations. Disulfide bridges are between Cys¹¹ and Cys⁵⁴, Cys²¹ and Cys⁴⁴, and Cys²⁷ and Cys⁵⁵.

main acts as a recognition site for factor $\bar{B}$, factor $\bar{G}$ or for coagulogen. Following this domain, a serine and threonine-rich sequence is highly glycosylated with *O*-linked carbohydrate chains. Furthermore, a proline-proline dimer sequence is conspicuously evident three times in the following region, a finding which suggests that the region forms an anomalous structure. The molecular mass of the L chain shows a large discrepancy in comparison with that estimated on sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and it is stained less strongly with Coomassie Brilliant Blue (7). We presume that these properties relate to the presence of carbohydrate chain clusters in the L chain.

As shown in Fig. 4, the proclotting enzyme is present in large granules in hemocytes, an observation in good agreement with findings that all the components required for coagulation of the hemocyte lysate are present in the granules (3). Various serine proteases in granules of mammalian mast cell and cytotoxic T cells have been noted but their functions remain to be clarified (31–33). These enzymes differ from the proclotting enzyme described here in molecular mass and primary structure.

Tai and Liu (34) reported the isolation of proclotting enzyme from the American horseshoe crab (*Limulus polyphemus*). This zymogen has a molecular mass of 150,000 and contains 43 γ -carboxyglutamic acids. It is activated directly by LPS, in a Ca^{2+} -dependent manner. However, our proclotting enzyme differs in terms of the apparent molecular mass, the amino acid composition, and the mode of activation (7). The present study confirms our reported data (7) and rules out the possibility that the proclotting enzyme is proteolytically derived from such a large precursor. Moreover, the propeptide region of this molecule does not contain the conserved residues for γ -carboxylation found in vitamin K-dependent clotting factors (35). A zymogen similar to the *Tachypleus* proclotting enzyme seems to exist in the hemocytes of *L. polyphemus*.

Acknowledgments—The initial part of this work was performed at the University of Washington in Seattle. We are grateful to Drs. E. W. Davie, K. Fujikawa, and B. A. McMullen for use of laboratory facilities and for their kind cooperation. We are also grateful to Dr. T. Nakamura for the kind gift of clotting enzyme and for valuable discussion, to Dr. K. Sueishi for preparing the antiserum, to Dr. Y. N. Han and A. Mizutani for technical assistance, to M. Ohara for comments on the manuscript, to S. Kajiyama and Y. Nishina for skillful analyses of the amino acids, and to N. Ueno for expert secretarial assistance.

REFERENCES

- Levin, J., and Bang, F. B. (1964) *Bull. Johns Hopkins Hosp.* **115**, 265–274
- Ornberg, R. L., and Reese, T. S. (1979) *Prog. Clin. Biol. Res.* **29**, 125–130
- Mürer, E. H., Levin, J., and Holme, R. (1975) *J. Cell. Physiol.* **86**, 533–542
- Iwanaga, S., Morita, T., Miyata, T., Nakamura, T., and Aketaga, J. (1986) *J. Protein Chem.* **5**, 255–268
- Nakamura, T., Morita, T., and Iwanaga, S. (1986) *Eur. J. Biochem.* **154**, 511–521
- Nakamura, T., Horiuchi, T., Morita, T., and Iwanaga, S. (1986) *J. Biochem. (Tokyo)* **99**, 847–857
- Nakamura, T., Morita, T., and Iwanaga, S. (1985) *J. Biochem. (Tokyo)* **97**, 1561–1574
- Miyata, T., Hiranaga, M., Umez, M., and Iwanaga, S. (1984) *J. Biol. Chem.* **259**, 8924–8933
- Morita, T., Tanaka, S., Nakamura, T., and Iwanaga, S. (1981) *FEBS Lett.* **129**, 318–321
- Nakamura, S., Morita, T., Harada-Suzuki, T., Iwanaga, S., Takahashi, K., and Niwa, M. (1982) *J. Biochem. (Tokyo)* **92**, 781–792
- Fujikawa, K., and McMullen, B. A. (1983) *J. Biol. Chem.* **258**, 10924–10933
- Hermanson, M. A., Ericsson, L. H., Neurath, H., and Walsh, K. A. (1973) *Biochemistry* **12**, 3146–3153
- Gross, E., and Witkop, B. (1962) *J. Biol. Chem.* **237**, 1856–1860
- Vieira, J., and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., and Hood, L. E. (1986) *Nature* **321**, 674–679
- Morohashi, K., Yoshioka, H., Sogawa, K., Fujii-Kuriyama, Y., and Omura, T. (1984) *J. Biochem. (Tokyo)* **95**, 949–957
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Lipman, D. J., and Pearson, W. R. (1985) *Science* **227**, 1435–1441
- Takemoto, H., Hase, S., and Ikenaka, T. (1985) *Anal. Biochem.* **145**, 245–250
- von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21
- Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Degen, S. J. F., and Davie, E. W. (1986) *Ann. N. Y. Acad. Sci.* **485**, 66–72
- Hagen, F. S., Gray, C. L., O'hara, P., Grant, F. J., Saari, G. C., Woodbury, R. G., Hart, C. E., Insley, M., Kisiel, W., Kurachi, K., and Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2412–2416
- Kurachi, K., and Davie, E. W. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 6461–6464
- Fung, M. R., Hay, C. W., and MacGillivray, R. T. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 3591–3595
- Foster, D. C., Yoshitake, S., and Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 4673–4677
- de Bruijn, M. H. L., and Fey, G. H. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 708–712
- Belt, K. T., Carroll, M. C., and Porter, R. R. (1984) *Cell* **36**, 907–914
- Lundwall, Å. B., Wetsel, R. A., Kristensen, T., Whitehead, A. S., Woods, D. E., Ogden, R. C., Colten, H. R., and Tack, B. F. (1985) *J. Biol. Chem.* **260**, 2108–2112
- Magnusson, S., Petersen, T. E., Sortrup-Jensen, L., and Claeys, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., and Shaw, E., eds) Vol. 28, pp. 123–149, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Trong, H. L., Parmelee, D. C., Walsh, K. A., Neurath, H., and Woodbury, R. G. (1987) *Biochemistry* **26**, 6988–6994
- Vanderslice, P., Craik, C. S., Nadel, J. A., and Coughley, G. H. (1989) *Biochemistry* **28**, 4148–4155
- Gershenfeld, H. K., and Weissman, I. L. (1986) *Science* **232**, 854–858
- Tai, J. Y., and Liu, T.-Y. (1977) *J. Biol. Chem.* **252**, 2178–2181
- Jorgensen, M. J., Cantor, A. B., Furie, B. C., Brown, C. L., Shoemaker, C. B., and Furie, B. (1987) *Cell* **48**, 185–191
- Fujikawa, K., Chung, D. W., Hendrickson, L. E., and Davie, E. W. (1986) *Biochemistry* **26**, 2417–2424
- Stenflo, J., and Fernlund, P. (1982) *J. Biol. Chem.* **257**, 12180–12190
- Leytus, S. P., Kurachi, K., Sakariassen, K. S., and Davie, E. W. (1986) *Biochemistry* **25**, 4855–4863
- Henderson, G., Krygsman, P., Liu, C. J., Davey, C. C., and Malek, L. T. (1987) *J. Bacteriol.* **169**, 3778–3784
- Epstein, D. M., and Wensink, P. C. (1988) *J. Biol. Chem.* **263**, 16586–16590

SUPPLEMENTAL MATERIAL

Proclotting Enzyme from Horseshoe Crab Hemocytes: cDNA CLONING, DISULFIDE LOCATIONS, AND SUBCELLULAR LOCALIZATION

by

T. Muta, R. Hashimoto, T. Miyata, H. Nishimura, Y. Toh, and S. Iwanaga

EXPERIMENTAL PROCEDURES

Materials — All DNA-modifying enzymes were obtained from Nippon Gene Co. (Toyama, Japan), Toyobo Co., LTD. (Osaka), Takara Shuzo Co. (Kyoto) or Bethesda Research Laboratories Life Technologies, Inc. (Gaithersburg, MD). [α - ^{32} P] dCTP and multi-prime labeling kit were from Amersham-Japan (Tokyo). Trypsin purified by benzamide-agarose column chromatography (11) was a kind gift from Dr. K. Fujikawa. L-tosylamide 2-phenyl ethyl chloromethyl ketone treated trypsin (TPCK-trypsin) was from Worthington Biochemical Corp. (Freehold, NJ). Lysyl endopeptidase from *Achromobacter lyticus* M497-1 was obtained from Wako Pure Chemical Industries, Ltd. (Osaka). Pyroglutamate carboxylase from *Bacillus amyloliquefaciens* and endoproteinase Asp-N from *Pseudomonas fragi* mutant were from Nacal Tesque, Inc. (Kyoto) and Boehringer Mannheim GmbH (Mannheim, FRG), respectively. All other chemicals were of the highest quality commercially available.

Proclotting enzyme was highly purified from the hemocyte lysates of *Tachyleptus tridentatus*, as described (7). The clotting enzyme activated by factor B or trypsin and anti-clotting enzyme IgG were kindly provided by Dr. T. Nakamura. λ gt11 cDNA libraries from hemocyte poly A⁺ RNA from *Tachyleptus tridentatus* were the same as that previously used³.

Partial Amino Acid Sequence — The purified clotting enzyme was reduced and S-pyridylethylated as described (12). The resulting S-pyridylethylated L and H chains were separated by gel filtration on a Sephadex G-100 (Pharmacia Fine Chemicals, Uppsala) column in 2.5% of formic acid, followed by reversed-phase high performance liquid chromatography (HPLC) on a column of UltraporeTM RPSC (4.6 x 75 mm) (AUX, Berkeley, CA). The S-pyridylethylated H chain was digested with cyanogen bromide (13) and the S-pyridylethylated L chain was digested with trypsin followed by lysyl endopeptidase digestion. These digests were purified by reversed-phase HPLC. An UltraporeTM RPSC column (4.6 x 75 mm) was used for the cyanogen bromide digest of H chain and a μ BondapakTM C18 (4.9 x 300 mm) (Waters) column for the trypsin digest of L chain. The purified peptides were subjected to amino-terminal sequence analyses.

Amino-terminal Sequence of Proclotting Enzyme — The S-pyridylethylated sample (2.4 nmol) was digested with TPCK-trypsin in 50 mM Tris-HCl, pH 8.0, with a molar ratio of enzyme to substrate of 1/100 at 37°C for 18 h. The digest was separated on a column of YMC-PACK A-302 S-5 120A ODS (4.6 x 150 mm) (Yamamura Chemical Laboratories Co., Ltd., Kyoto). The amino-terminal cysteine peptide, Pe-PC-T21 (652 pmol), was digested with pyroglutamate carboxylase (20 mU) in 0.1 M sodium phosphate, 10 mM EDTA and 10 mM β -mercaptoethanol, pH 8.0 at 37°C for 18 h. The resulting peptide free from pyroglutamate residue was purified on a column of Chemcosorb 7 ODS-H (2.1 x 150 mm) (Chemco Scientific Co., Ltd., Osaka) and subjected to amino acid and sequence analyses.

Screening — From 1.5 million clones screened with anti-clotting enzyme IgG, twenty-three positive clones were obtained. Positive clones were sequenced after being subcloned into pUC118 or 119 vectors. The insert of the clone, λ PC17, that encoded the partial sequences of clotting enzyme was labeled with [α - ^{32} P] dCTP by multi-prime labeling kit. Using this probe, 400,000 clones from the second libraries constructed with *Eco*RI adapter were screened. The longest clone, λ PC11, thus obtained was sequenced.

Nucleotide Sequencing — The inserts of positive clones were transferred into pUC118 and 119 (Takara Shuzo Co.). Single stranded DNA templates were prepared using MV1184 as host cells and M13K07 as helper phages (14). Nucleotide sequence determination was performed on an Applied Biosystems 370 A DNA sequencer using fluorescent oligonucleotide primers (15). SequenaseTM from United States Biochemical Corp. (Cleveland, OH) or Taq polymerase (Promega Corp., Madison, WI) were used for extension reactions.

Northern Blotting — Total RNA was extracted from the hemocytes (16) and poly A⁺ RNA was purified on an oligo dT cellulose column (17). Five μ g of poly A⁺ RNA was separated on 1.0% agarose gel and transferred to GVHP membrane (Nippon Millipore Kogyo K.K., Yonezawa, Japan). Marker RNA (Bethesda Research Laboratories Life Technologies, Inc.) was electrophoresed simultaneously and stained with ethidium bromide. The filter was hybridized with [^{32} P]- λ PC17 insert in 5 x SSC (1 x SSC = 15 mM sodium citrate and 150 mM NaCl, pH 7.0), 5 x Denhardt's solution (1 x Denhardt's solution = 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin, 0.02% Ficoll), 50 mM sodium phosphate, pH 6.8, and 50% formamide at 42°C overnight and then washed in 0.1 x SSC containing 0.1% sodium laurylsulfate at 65°C. The washed filter was subjected to autoradiography.

Determination of Disulfide Linkages — Proclotting enzyme (2.2 nmol) or clotting enzyme (12.8 nmol) was digested with lysyl endopeptidase in 0.1 M Tris-HCl, pH 6.5, containing 4 M urea with a weight ratio of enzyme to substrate of 1/100 at 37°C for 16 h. The digest was separated by reversed-phase HPLC on a column of YMC-PACK A-302 S-5 120A ODS (4.6 x 150 mm). The purified peptides were subjected to amino acid and sequence analyses. The peaks containing more than one peptide (K11-K12-K13 + K15) were rechromatographed on a column of Chemcosorb 7 ODS-H (2.1 x 150 mm). Peptide, K1-K2-K4-K5, was further digested with TPCK-trypsin in 50 mM ammonium formate, pH 6.5, containing 20 mM CaCl₂ with a molar ratio of enzyme to substrate of 1/20 at 37°C for 18 h or with endoproteinase Asp-N in 20 mM sodium phosphate, pH 6.5, with a weight ratio of enzyme to substrate of 1/40 at 37°C for 8 h. The digests were also separated by reversed-phase HPLC on a column of Chemcosorb 7 ODS-H (2.1 x 150 mm).

Amino Acid and Sequence Analysis — Amino acid analysis was performed on a PICO-TAC system (Waters, Millipore Corp., Milford, MA) or a Hitachi 18500 automatic analyzer. Amino acid sequence analyses were done using an Applied Biosystems 477A Protein Sequencer.

Homology Search — The amino acid sequence of the proclotting enzyme was compared with all entries in the NBRF Protein Sequence Database (release 23.0, December, 1989) with GENETYX system (Software Development Co., LTD, Tokyo) based on the algorithm of the computer program FASTP (18).

Component Sugar Analyses — The sugar containing peptides K7-K10 were digested with TPCK-trypsin in 50 mM Tris-HCl, pH 8.0, with a molar ratio of enzyme to substrate of 1/50 at 37°C for 16 h. The digest was separated by the reversed-phase HPLC on a column of Chemcosorb 7 ODS-H (2.1 x 150 mm). Component sugar analyses were performed on samples hydrolyzed in 4 M trifluoroacetic acid at 100°C for 3 h, using derivatizations with 2-aminopyridine, as described (19).

Subcellular Localization — A polyclonal antibody raised against clotting enzyme was purified before use, by immunoaffinity chromatography. The column was Affigel 15 (Nippon Bio-Rad Laboratories, Tokyo) coupled with the clotting enzyme treated with diisopropylfluorophosphate. The monospecific IgG was eluted with 0.1 M glycine-HCl, pH 2.8. Hemolymph was collected into sterilized tubes and immediately fixed with 4% paraformaldehyde and 1.2% glutaraldehyde in 0.1 M sodium cacodylate buffer (pH 7.4) containing 4.5 mM CaCl₂ and 0.1% sucrose at 0°C for 30 min. The hemocytes were precipitated by centrifugation and washed several times with 0.1 M sodium cacodylate (pH 7.4). The specimens were dehydrated with a gradual increase in methanol concentration and then embedded in LR Gold (Polaron Instruments Inc., Cambridge, MA), which was polymerized by ultraviolet irradiation. Ultrathin sections were prepared by using a diamond knife. Sections were mounted on a nickel grid and washed with distilled water. Following blocking with 10 mM sodium phosphate buffer (pH 7.2) containing 1% (w/v) bovine serum albumin and 0.5 M NaCl, the sections were incubated for 1 h with anti-clotting enzyme IgG (0.4 μ g/ml) in the same albumin containing buffer, at room temperature. The sections were washed 6 times with 10 mM sodium phosphate (pH 7.2) containing 0.5 M NaCl and incubated with 1.2 μ g/ml of goat anti-rabbit IgG gold Colloidal particles (20 nm) (E.Y. Labs, Inc., San Mateo, CA) at room temperature for 1 h. After washing with 10 mM sodium phosphate (pH 7.2) containing 0.5 M NaCl and then with distilled water, the sections were doubly stained with 2% uranyl acetate and lead tartrate and examined under an electron microscope (Hitachi H-500).

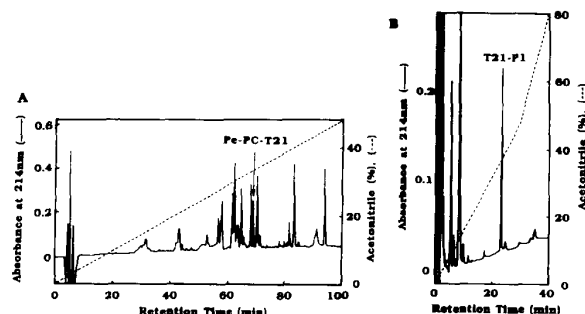


FIG. M1. Separation of trypsin digest of S-pyridylethylated proclotting enzyme (A) and purification of pyroglutamate carboxylase digest of amino-terminal peptide (B). (A) The trypsin digest was applied to a column of YMC-PACK A-302 S-5 120A ODS (4.6 x 150 mm). The peptide, Pe-PC-T21, derived from the amino-terminal portion of the proclotting enzyme is indicated by an arrow. (B) Pyroglutamate carboxylase digest of the amino-terminal peptide (Pe-PC-T21) was applied to a column of Chemcosorb 7 ODS-H (2.1 x 150 mm). The peptides were eluted with a linear gradient of acetonitrile containing 0.1% trifluoroacetic acid. The flow rate was 0.5 ml/min.

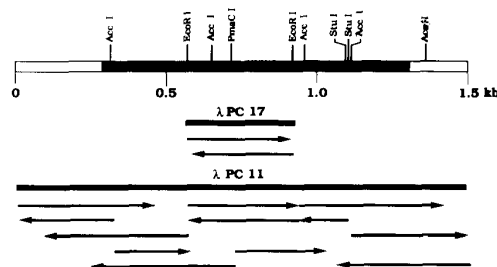


FIG. M2. Restriction map and sequence strategy of proclotting enzyme cDNA. A black box in the restriction map represents an open reading frame. The nucleotide sequence analysis was carried out with an Applied Biosystems 370 A DNA sequencer in the directions and length indicated by arrow, after being subcloned into pUC118 or 119 vector.

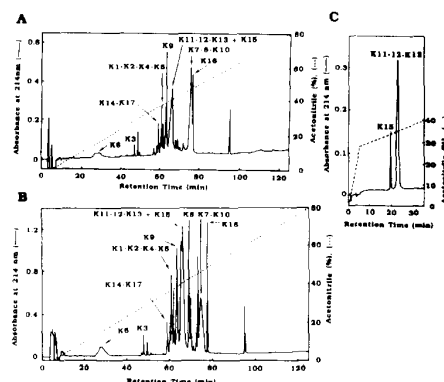


FIG. M3. Separation of lysyl endopeptidase digest of proclotting enzyme (A) and clotting enzyme (B) and rechromatography of K11-K12-K13 and K15 (C). Lysyl endopeptidase digests of proclotting enzyme (A) and clotting enzyme (B) were applied, respectively, to a column of YMC-PACK A-302 S-5 120A ODS (4.6 x 150 mm). The peak shown as K11-K12-K13 + K15 in Fig. M3B was rechromatographed on a column of Chemcosorb 7 ODS-H (2.1 x 150 mm). The peptides were eluted by a linear gradient of acetonitrile containing 0.1% trifluoroacetic acid. The flow rate was 0.5 ml/min.

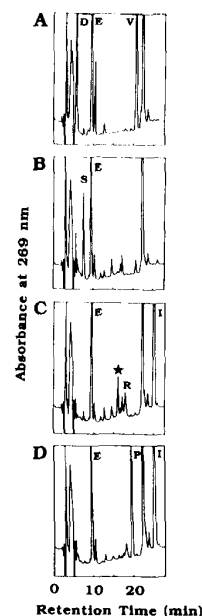


FIG. M4. Identification of PTH-cysteine generated after Edman degradation of peptide K1-2-4-5-D2. The peptide, K1-2-4-5-D2, was subjected to amino-terminal analysis by an Applied Biosystems 477A Protein Sequencer. The first cycle of Edman degradation (A) released PTH-derivatives of Asp, Glu, and Val. The second cycle (B) yielded PTH-Ser and PTH-Glu. In the third cycle (C), there was a peak at the retention time of PTH-cysteine indicated by an asterisk in addition to PTH-Glu, PTH-Arg, and PTH-Ile. The fourth cycle yielded PTH-Glu, PTH-Pro, and PTH-Ile, but the peak of PTH-cysteine had disappeared.

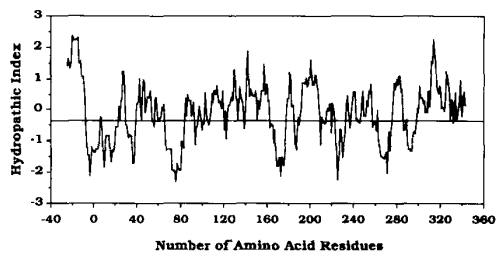


FIG. M5. Hydropathy profiles of the proclotting enzyme. The hydropathic index is the mean value of 9 successive residues, calculated by the method by Kyte and Doolittle (21).

Clotting Enzyme (Horseshoe Crab)	NRHVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Factor XIa (Human)	NRVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Factor Xa (Human)	EFVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Thrombin (Bovine)	NRVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Protein C (Bovine)	VSVVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Granzyme A (Rat)	KNVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Trypsin (Cattle)	PNVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Factor C (Horseshoe Crab)	EKVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Clr (Human)	NRVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Protease A (<i>Streptomyces griseus</i>)	VAVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
α -lytic protease (<i>Lysobacter enzymogenes</i>)	TKGVITAHICVYN	MDIALILLEN	CAGF	ADGGK	DACGSDSG
Chymotrypsin No.	57	102	195		

FIG. M6. Alignments of sequences around active sites of clotting enzyme and other serine proteases. Identical residues are boxed. The numbering is that of chymotrypsinogen. Sequences are from the following references: Factor XIa (36), Factor Xa (25), Thrombin (30), Protein C (37), Granzyme A (33), Trypsin (32), Factor C (38), Protease A (39) and α -lytic protease (40).

Protein	Sequence
Factor XIa (Human)	NRHVITAHICVYN
Factor Xa (Human)	EFVITAHICVYN
Thrombin (Bovine)	NRVITAHICVYN
Protein C (Bovine)	VSVVITAHICVYN
Granzyme A (Rat)	KNVITAHICVYN
Trypsin (Cattle)	PNVITAHICVYN
Factor C (Horseshoe Crab)	EKVITAHICVYN
Clr (Human)	NRVITAHICVYN
Protease A (<i>Streptomyces griseus</i>)	VAVITAHICVYN
α -lytic protease (<i>Lysobacter enzymogenes</i>)	TKGVITAHICVYN

Table MII
Amino acid compositions of the amino-terminal peptide derived from the proclotting enzyme and its pyroglutamylopeptidase digest^a

Amino acid	Pe-PC-T21	T21-P1
Asp	2.4 (2)	2.1 (2)
Glu	4.0 (4)	2.9 (3)
Ser	0.8 (1)	1.2 (1)
Gly	0.6	0.4
Arg	1.0 (1)	1.1 (1)
Pro	1.1 (1)	1.0 (1)
Tyr	0.4	0.8 (1)
Val	0.9 (1)	1.0 (1)
Pec	0.7 (1)	0.8 (1)
Leu	1.2 (1)	1.0 (1)
Phe	1.8 (2)	1.9 (2)
Total	14	13
Position	1-14	2-14

^aAnalysed with a PICO-TAG system on hydrolysates in 6 M HCl vapor at 110 °C for 20 h. Values in parentheses are from the sequence data.
^bS-pyridylethylcysteine.
^cNot determined.

Table MIII
Amino acid sequence of T21-P1

Cycle No.	Amino acid	Yield (pmol)
1	Phe	137
2	Val	118
3	Phe	147
4	Pro	104
5	Asp	46
6	Gly	80
7	X ^a	
8	Glu	112
9	Leu	71
10	Pec	16
11	Ser	25
12	Asn	28
13	X	

Repetitive 103.6
yield (%) (Phe1-Phe3)

^aNot identified.

Table MIV
Amino acid compositions of heavy and light chains of clotting enzyme and proclotting enzyme^a

Amino acid	Light chain ^b	Heavy chain ^b	Proclotting enzyme ^c
	Residues/molecule		
Cys	7.5 (7)	8.4 (7)	11.3 (14)
Asp	9.9 (3)	26.6 (16)	36.0 (19)
Asn	(6)	(9)	(15)
Thr	9.6 (12)	18.0 (18)	28.5 (30)
Ser	3.8 (6)	10.3 (11)	17.4 (17)
Glu	15.3 (10)	16.5 (10)	31.3 (20)
Gln	(3)	(5)	(8)
Pro	10.1 (10)	13.5 (13)	21.3 (23)
Gly	4.6 (4)	23.9 (23)	26.9 (27)
Ala	1.1 (1)	20.3 (20)	20.9 (21)
His	1.4 (2)	8.1 (8)	7.9 (10)
Val	4.8 (5)	18.9 (21)	25.1 (26)
Met	0 (0)	6.2 (7)	6.6 (7)
Ile	4.8 (6)	13.7 (17)	20.5 (23)
Leu	7.3 (7)	17.4 (17)	24.9 (24)
Tyr	1.0 (1)	8.8 (9)	11.0 (10)
Phe	4.6 (4)	11.3 (11)	15.6 (15)
Lys	7.0 (6)	11.3 (10)	19.5 (16)
Trp	0 (0)	3.6 (5)	5.0 (5)
Arg	4.2 (4)	11.2 (11)	15.3 (15)
Total	97.0 (97)	248.0 (248)	345.0 (345)
Glucosamine	2.4	5.5	9.1
Galactosamine	5.1	1.4	7.9

^aRecalculated from the data of Nakamura et al. (1). Values in parentheses are from the sequence deduced from cDNA.
^bCalculated on sample of a 24 h hydrolysate.
^cCalculated from extrapolated or average values estimated on samples of 24, 48 and 72 h hydrolysates.
^dIdentified as S-carboxymethylcysteine.
^eTaken from the value of a 72 h hydrolysate.

Table MII
Amino acid compositions of the amino-terminal peptide derived from the proclotting enzyme and its pyroglutamylopeptidase digest^a

Amino acid	Pe-PC-T21	T21-P1
Asp	2.4 (2)	2.1 (2)
Glu	4.0 (4)	2.9 (3)
Ser	0.8 (1)	1.2 (1)
Gly	0.6	0.4
Arg	1.0 (1)	1.1 (1)
Pro	1.1 (1)	1.0 (1)
Tyr	0.4	0.8 (1)
Val	0.9 (1)	1.0 (1)
Pec	0.7 (1)	0.8 (1)
Leu	1.2 (1)	1.0 (1)
Phe	1.8 (2)	1.9 (2)
Total	14	13
Position	1-14	2-14

^aAnalysed with a PICO-TAG system on hydrolysates in 6 M HCl vapor at 110 °C for 20 h. Values in parentheses are from the sequence data.
^bS-pyridylethylcysteine.
^cNot determined.

Protein	Asp	Glu	Ser	Gly	Arg	Pro	Tyr	Val	Pec	Leu	Phe	Total
Factor XIa (Human)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Factor Xa (Human)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Thrombin (Bovine)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Protein C (Bovine)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Granzyme A (Rat)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Trypsin (Cattle)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Factor C (Horseshoe Crab)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Clr (Human)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
Protease A (<i>Streptomyces griseus</i>)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14
α -lytic protease (<i>Lysobacter enzymogenes</i>)	2.4 (2)	4.0 (4)	0.8 (1)	0.6	1.0 (1)	1.1 (1)	0.4	0.9 (1)	0.7 (1)	1.2 (1)	1.8 (2)	14

Table MVII
Amino acid sequences of 1991 endopeptidase peptides

Cycle No.	K1-K2-K4-K5				K3		K6		K7-K10			
	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)
1	X ^a	Asn 11.4	Glu 18.2	Val 67.1	Asn 56.6	Ser 45.6	Glu 171.9	Val 18.4	Val 18.4	Val 18.4	Val 18.4	Val 18.4
2		Val 21.0	Ser 22.7	X	Asp 22.8	Asp 22.8	Val 15.3	Val 15.3	Val 15.3	Val 15.3	Val 15.3	Val 15.3
3		Leu 40.9	Leu 40.9	Pro 36.2	Pro 36.2	Pro 36.2	Pro 36.2	Pro 36.2	Pro 36.2	Pro 36.2	Pro 36.2	Pro 36.2
4		Arg 6.0	X	X	Arg 6.0	Arg 6.0	Arg 6.0	Arg 6.0	Arg 6.0	Arg 6.0	Arg 6.0	Arg 6.0
5		Leu 23.4	Glu 26.6	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
6		Arg 2.9	Phe 86.7	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
7		Leu 29.6	Glu 32.1	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
8		Leu 29.6	Glu 32.1	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
9		Leu 29.6	Glu 32.1	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
10		Glu 10.1	Thr 15.3	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
11		Glu 8.5	Pro 22.7	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
12		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
13		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
14		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
15		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
16		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
17		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
18		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
19		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
20		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
21		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
22		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
23		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
24		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
25		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
26		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
27		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
28		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
29		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
30		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
31		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
32		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
33		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
34		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
35		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
36		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
37		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
38		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1
39		X	X	X	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1	Leu 191.6	Ser 4.1

Position 1-22 41-52 53-57 58-79 80-98 187-224
 Repetitive yield (%) 86.7 97.9 87.2 87.2 87.2 87.2
 (Leu1-Leu8) (Leu1-Leu8) (Leu1-Leu8) (Leu1-Leu8) (Leu1-Leu8) (Leu1-Leu8)
 Not identified.
 Two peptides have the same amino acids.
 Sum of yields from two peptides.

^aNot identified.
^bSum of yields from two peptides.

^cNot identified.
^dPTH-cysteine was detected.

^eNot quantitatively determined.

(continued)

Cycle No.	K8		K9		K11-K13				K14-K17				K15		K18	
	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)	Amino Yield acid (pmol)
1	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
2	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
3	Glu 36.8	Glu 36.8	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9
4	Glu 36.8	Glu 36.8	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9
5	Arg 6.0	X	Arg 6.0	X	Arg 6.0	X	Arg 6.0	X	Arg 6.0	X	Arg 6.0	X	Arg 6.0	X	Arg 6.0	X
6	Glu 24.5	Ser 13.3	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2
7	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2	Val 30.7	Asp 127.2
8	Glu 24.5	Ser 13.3	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2	Leu 413.5	Asp 127.2
9	X	Asp 127.2	X	Asp 127.2	X	Asp 127.2	X	Asp 127.2	X	Asp 127.2	X	Asp 127.2	X	Asp 127.2	X	Asp 127.2
10	Glu 36.8	Glu 36.8	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9	Asp 127.2	Leu 109.9
11	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
12	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
13	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
14	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
15	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
16	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8
17	Leu 95.1	Glu 31.7	Leu 733.0	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8	Asp 131.8

Position 99-115 120-186 225-272 273-290 321-333 387-420 444-466
 Repetitive yield (%) 87.2 87.2 87.2 87.2 87.2 87.2 87.2
 (Glu1-Glu8) (Glu1-Glu8) (Glu1-Glu8) (Glu1-Glu8) (Glu1-Glu8) (Glu1-Glu8) (Glu1-Glu8)
 Not identified.
 Two peptides have the same amino acids.
 Sum of yields from two peptides.

^aNot identified.
^bPTH-cysteine was detected.

^cNot quantitatively determined.

Table MVIII
Amino acid sequence of K1-K2-K4-K5-D2.

Cycle No.	K1-K2-K4-K5-D2				Amino acid (pmol)	Yield (%)	Amino acid (pmol)	Yield (%)	Amino acid (pmol)	Yield (%)	Amino acid (pmol)	Yield (%)
	Amino acid (pmol)	Yield (%)	Amino acid (pmol)	Yield (%)								
1	Asp ^a	(329) ^b	Asp ^a		Glu	278	Val	706				
2	Glu	187	X		Ser	121	X ^d					
3	Glu	240	Arg	31	Ile	463	Cys ^d					
4	Glu	248	Ile	517	X		Pro	292				
5	Leu ^a	(846) ^b	Leu ^a		Gly	340	Lys	133				
6	X		Leu	502	Phe	365	X					
7	Ser	91	Glu	284	Glu	225						
8	Asn	155	Lys	152	Gly	227						
9	Arg	66	X		Ile	328						
10	Phe	182	X		Thr	76						
11	Thr	114			Pro	137						
12	Glu	65			Lys	55						
13	Glu	87			X							
14	Gly	99										
15	Thr	50										
16	X											
17	Lys	36										
18	X											

Position 6-22 26-33 41-52 53-57
 Repetitive yield (%) 90.0 96.5
 (Glu2-Glu12) (Glu1-Glu7)

^aTwo peptides have the same amino acids.

^bSum of yields from two peptides.

^cNot identified.

^dPTH-cysteine was detected.

^eNot quantitatively determined.

Table MVIII
Neutral sugar compositions of glycopeptides derived from proclotting enzyme^a

Neutral sugar	Residues/molecule			
	K6	K7-K10-T4	K7-K10-T8	K11-12-K13