MULTIPLE GENE DUPLICATION IN THE EVOLUTION OF PLASMINOGEN. FIVE REGIONS OF SEQUENCE HOMOLOGY WITH THE TWO INTERNALLY HOMOLOGOUS STRUCTURES IN PROTHROMBIN

Hendrik CLAEYS*, Lars SOTTRUP-JENSEN, Maria ZAJDEL,
Torben E. PETERSEN and Staffan MAGNUSSON**
Department of Molecular Biology, University of Aarhus, DK-8000 Århus C, Denmark

Received 29 September 1975
Revised version 19 November 1975

1. Introduction

In order to investigate the substrate specificity of urokinase (EC 3.4.99.26) and other plasminogen activators we isolated and sequenced a chymotryptic fragment of 38 residues from the activation cleavage site region of plasminogen [1]. The fragment overlapped the last 28 residues in the heavy chain of plasmin (EC 3.4.21.7) (derived from the N-terminal 3/4 of the single chain plasminogen molecule) to the previously known sequence [2] of the first 10 residues in the light chain (derived from the C-terminal 1/4 of the plasminogen). The fragment thus contained the $Arg_{approx.\ 600}$–Val bond which is cleaved on activation of plasminogen to plasmin [3]. In the N-terminal region of this chymotryptic fragment residues 2–11 showed 70% sequence identity with two sequences in prothrombin (residues 138–147 and 243–252) [1], which constitute the C-terminal parts of the two internally homologous kringle† region (residues 62–144 and 167–249) [4] in the 'pro' part (residues 1–274) of the prothrombin single chain (582 residues) [5]. These two 83-residue kringle regions in prothrombin show 31 sequence identities, including the 6 Cys-residues which are paired 1–6, 2–4, 3–5 in both kringle regions [4,5]. Since investigation of other chymotryptic peptides from the heavy chain part of plasminogen gave evidence for more extensive homology with the kringle regions of prothrombin [1] it was decided to determine the amino acid sequence of the heavy chain. Wiman and Wallén have determined the sequence of residues 1–81 [6] in human plasminogen. The present paper gives partial sequence evidence for the rest of the heavy chain (residues 76–approx. 587) of human plasmin deduced from the chymotryptic digest of reduced, carboxymethylated plasminogen B [1], from a chymotryptic digest of reduced, carboxamidomethylated plasminogen (mixture of A and B), and from a tryptic digest of reduced, carboxymethylated, citraconylated heavy chain of plasmin.

2. Materials and methods

The purification of plasminogen [7,8] from fresh frozen human plasma, the assay of plasminogen [8], the enzymes used for digestion, the inhibitors used to control activation or to stop digestion, the methods used to isolate peptides by chromatography, gel filtration and high voltage paper electrophoresis, the detection of peptides by Cd-ninhydrin and by specific stains for Arg, His/Tyr, Trp, and for Cys-residues labelled with either [^{14}C]iodoacetate or [^{14}C]iodoacetamide, methods used for qualitative and quantita-

*On leave of absence from the Laboratory of Blood Coagulation, Medical Research Department, University of Leuven, Belgium.

**To whom inquires may be sent.

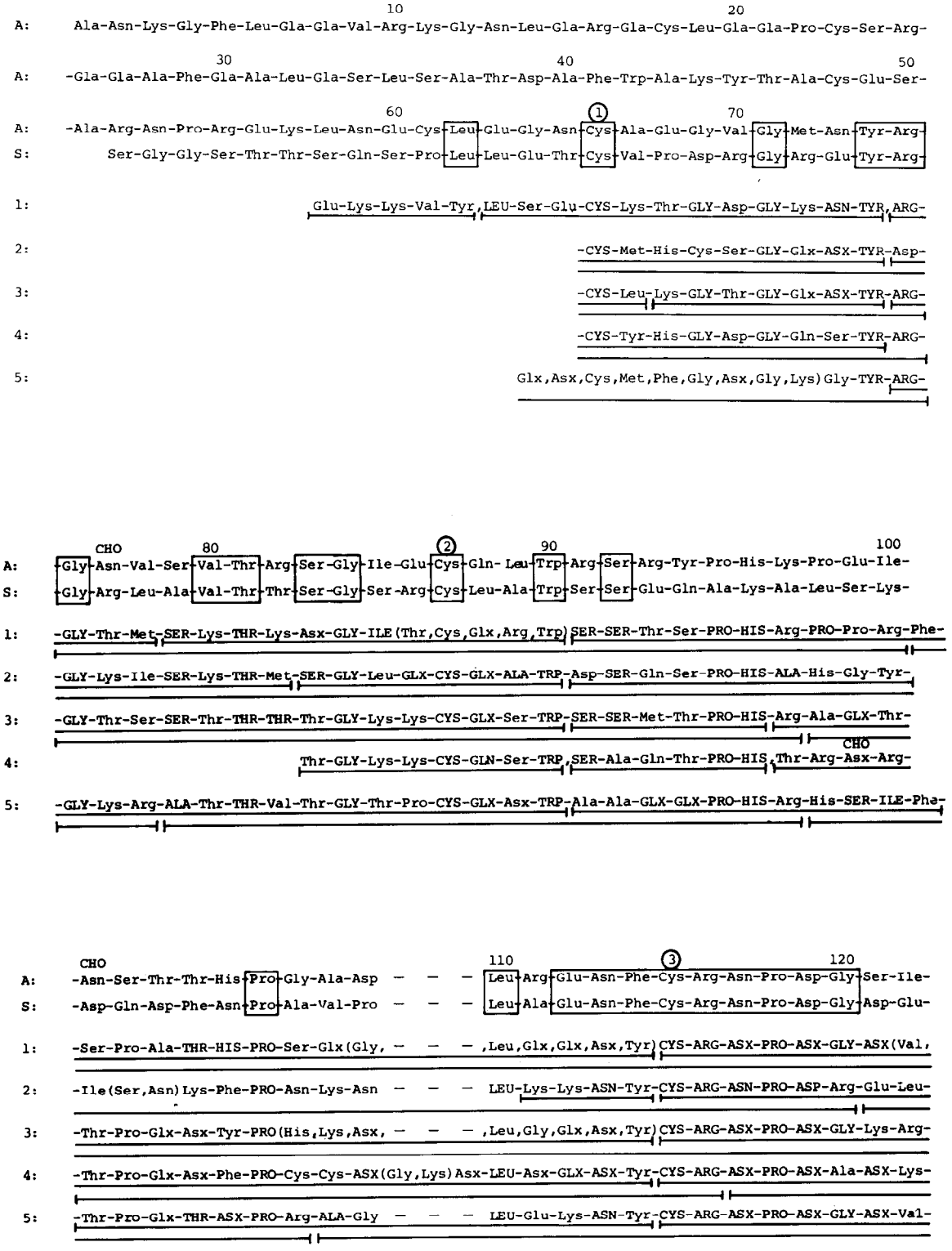
† Kringle: a classical shape of Scandinavian cake. Used as a tentative term for the two 83-residue internally homologous structures in prothrombin, because the two-dimensional representation of their disulfide-bridged structures resembles kringles [4,5].

North-Holland Publishing Company – Amsterdam

tive determination of amino acid composition, determination of sequence by the Dns-Edman method, identification of Dns-amino acid derivatives and for correlating peptide mobilities with charge and size were referenced elsewhere [1]. In addition an electrophoretic system with 60 mM NH_4HCO_3 adjusted with NH_3 to pH 8.9, running for 60–90 min at 1.5 kV was used for peptide separation. Trp-containing peptides were hydrolysed in toluene-sulphonic acid containing tryptamine [9]. Citraconylation of amino groups was used to prevent tryptic cleavage at Lys-residues [10]. To prepare the heavy chain of plasmin plasminogen was activated with urokinase (E:S 1:1000) in the presence of the Kunitz trypsin inhibitor (Trasylol, Baeyer) (I:S 1.2:1). Then the reduced, carboxymethylated, citraconylated chains were separated by gel filtration [11]. The tryptic peptides from the heavy chain derivative were purified by DEAE-cellulose chromatography [1], decitraconylated, and then separated by electrophoresis at pH 6.5, pH 2.1 and/or pH 8.9, if necessary followed by paper chromatography. In contrast to the other two digests the chymotryptic digest of reduced, carboxamidomethylated plasminogen was not completely soluble. The soluble fraction was subjected to initial separation in 10% acetic acid on Sephadex G-50. The larger peptides were then further purified by the NH_4HCO_3 gradient chromatography on DEAE-cellulose [1] and finally on paper. The insoluble fraction was digested with pepsin resulting in soluble peptides which were purified on paper after an initial Sephadex G-50 step. Detailed accounts of the chymotryptic and tryptic peptides will be given elsewhere and only the sequence results are presented here because there is now sufficient evidence to show that the heavy chain contains five regions of strong sequence homology with the two kringle regions in prothrombin.

3. Results and discussion

Among the 14 chymotryptic fragments showing extensive sequence homology with the kringle loops in prothrombin [1] were 5 different but very similar peptides containing 11 or 12 residues with identical N-terminal (Cys–Arg–Asx–Pro–Asx–) and C-terminal (–Pro–Trp) sequences. These peptides were isolated in approximately equimolar yields corresponding to the average for peptides from this digest, and showed 67–83% sequence identity with residues 115–126 in the A-fragment of prothrombin and with corresponding residues in the S-fragment. These five peptides can be seen in fig.1 and start with Cys-3 in lines 1–5. Combination of sequence evidence from the chymotryptic and tryptic peptides allows confident overlapping of four longer stretches of sequence containing these five peptides, namely 142 residues starting Arg–Gly on line 1, between Cys-1 and Cys-2, ending –Lys–Arg on line 2, between Cys-4 and Cys-5; 38 residues starting Ala–Glx on line 3, between Cys-2 and Cys-3 and ending –Arg–Trp between Cys-4 and Cys-5; 21 residues starting Cys–Arg– in line 4 with Cys-3 and ending –Arg–Trp between Cys-4 and Cys-5 and finally 91 residues starting Arg–Gly– in line 5, ending with the C-terminal –Arg of the heavy chain. In addition to these 291 residues we have overlapped three stretches of 31, 42 and 44–46 residues, respectively. They contain the homologous sequences around Cys-5 and Cys-6 on lines 2, 3 and 4 and continue beyond Cys-1 in lines 3, 4 and 5, respectively. Of the five remaining stretches accounting for 74 residues four are placed by homology, the fifth agrees with the known sequence for residues 76–80 of human plasminogen [6]. The hexosamine of the two carbohydrate prosthetic groups found so far are glucosamine (on the Asx–Arg–Thr– sequence on line 4) and galactosamine (on the Thr–Pro–Pro– sequence of line 3). The total number of residues accounted for in the present work is now 482–484. From the homology alignment in fig.1 one suspects that a stretch of seven residues is missing from line 4. Laursen and Lee [13] independently sequenced a stretch of 29 residues which confirmed part of our 142 residue stretch. In their [13] second stretch of 59 or 60 residues 39 residues confirm, with three differences, part of our 44–46 residue stretch. The remaining 21 residues of their [13] second stretch have not yet been found in our subdigests of peptides from that region but could be inserted if the two Lys–Lys– sequences between Cys-5 and Cys-6 one line 4 are not completely or partly identical. The last 30 residues of the heavy chain have recently been independently determined by Wiman and Wallén [14], and confirmed our sequence of this region [1] except for two amide differences. The total number of amino acids residues in the heavy chain including the first 75 residues [6]


A: Ala-Asn-Lys-Gly-Phe-Leu-Gla-Gla-Val-Arg-Lys-Gly-Asn-Leu-Gla-Arg-Gla-Cys-Leu-Gla-Gla-Pro-Cys-Ser-Arg-
A: -Gla-Gla-Ala-Phe-Gla-Ala-Leu-Gla-Ser-Leu-Ser-Ala-Thr-Asp-Ala-Phe-Trp-Ala-Lys-Tyr-Thr-Ala-Cys-Glu-Ser-
A: -Ala-Arg-Asn-Pro-Arg-Glu-Lys-Leu-Asn-Glu-Cys-Leu-Glu-Gly-Asn-Cys-Ala-Glu-Gly-Val-Gly-Met-Asn-Tyr-Arg-
S: Ser-Gly-Gly-Ser-Thr-Thr-Ser-Gln-Ser-Pro-Leu-Leu-Glu-Thr-Cys-Val-Pro-Asp-Arg-Gly-Arg-Glu-Tyr-Arg-
1: Glu-Lys-Lys-Val-Tyr,LEU-Ser-Glu-CYS-Lys-Thr-GLY-Asp-GLY-Lys-ASN-TYR,ARG-
2: -CYS-Met-His-Cys-Ser-GLY-Glx-ASX-TYR-Asp-
3: -CYS-Leu-Lys-GLY-Thr-GLY-Glx-ASX-TYR-ARG-
4: -CYS-Tyr-His-GLY-Asp-GLY-Gln-Ser-TYR-ARG-
5: Glx,Asx,Cys,Met,Phe,Gly,Asx,Gly,Lys)Gly-TYR-ARG-
A: Gly-Asn-Val-Ser-Val-Thr-Arg-Ser-Gly-Ile-Glu-Cys-Gln-Leu-Trp-Arg-Ser-Arg-Tyr-Pro-His-Lys-Pro-Glu-Ile-
S: Gly-Arg-Leu-Ala-Val-Thr-Thr-Ser-Gly-Ser-Arg-Cys-Leu-Ala-Trp-Ser-Ser-Glu-Gln-Ala-Lys-Ala-Leu-Ser-Lys-
1: -GLY-Thr-Met-SER-Lys-THR-Lys-Asx-GLY-ILE(Thr,Cys,Glx,Arg,Trp)SER-SER-Thr-Ser-PRO-HIS-Arg-PRO-Pro-Arg-Phe-
2: -GLY-Lys-Ile-SER-Lys-THR-Met-SER-GLY-Leu-GLX-CYS-GLX-ALA-TRP-Asp-SER-Gln-Ser-PRO-HIS-ALA-His-Gly-Tyr-
3: -GLY-Thr-Ser-SER-Thr-THR-THR-Thr-GLY-Lys-Lys-CYS-GLX-Ser-TRP-SER-SER-Met-Thr-PRO-HIS-Arg-Ala-GLX-Thr-
4: Thr-GLY-Lys-Lys-CYS-GLN-Ser-TRP,SER-Ala-Gln-Thr-PRO-HIS,Thr-Arg-Asx-Arg-
5: -GLY-Lys-Arg-ALA-Thr-THR-Val-Thr-GLY-Thr-Pro-CYS-GLX-Asx-TRP-Ala-Ala-GLX-GLX-PRO-HIS-Arg-His-SER-ILE-Phe-
A: -Asn-Ser-Thr-Thr-His-Pro-Gly-Ala-Asp - - - Leu-Arg-Glu-Asn-Phe-Cys-Arg-Asn-Pro-Asp-Gly-Ser-Ile-
S: -Asp-Gln-Asp-Phe-Asn-Pro-Ala-Val-Pro - - - Leu-Ala-Glu-Asn-Phe-Cys-Arg-Asn-Pro-Asp-Gly-Asp-Glu-
1: -Ser-Pro-Ala-THR-HIS-PRO-Ser-Glx(Gly, - - - ,Leu,Glx,Glx,Asx,Tyr)CYS-ARG-ASX-PRO-ASX-GLY-ASX(Val,
2: -Ile(Ser,Asn)Lys-Phe-PRO-Asn-Lys-Asn - - - LEU-Lys-Lys-ASN-Tyr-CYS-ARG-ASN-PRO-ASP-Arg-Glu-Leu-
3: -Thr-Pro-Glx-Asx-Tyr-PRO(His,Lys,Asx, - - - ,Leu,Gly,Glx,Asx,Tyr)CYS-ARG-ASX-PRO-ASX-GLY-Lys-Arg-
4: -Thr-Pro-Glx-Asx-Phe-PRO-Cys-Cys-ASX(Gly,Lys)Asx-LEU-Asx-GLX-ASX-Tyr-CYS-ARG-ASX-PRO-ASX-Ala-ASX-Lys-
5: -Thr-Pro-Glx-THR-ASX-PRO-Arg-ALA-Gly - - - LEU-Glu-Lys-ASN-Tyr-CYS-ARG-ASX-PRO-ASX-GLY-ASX-Val-

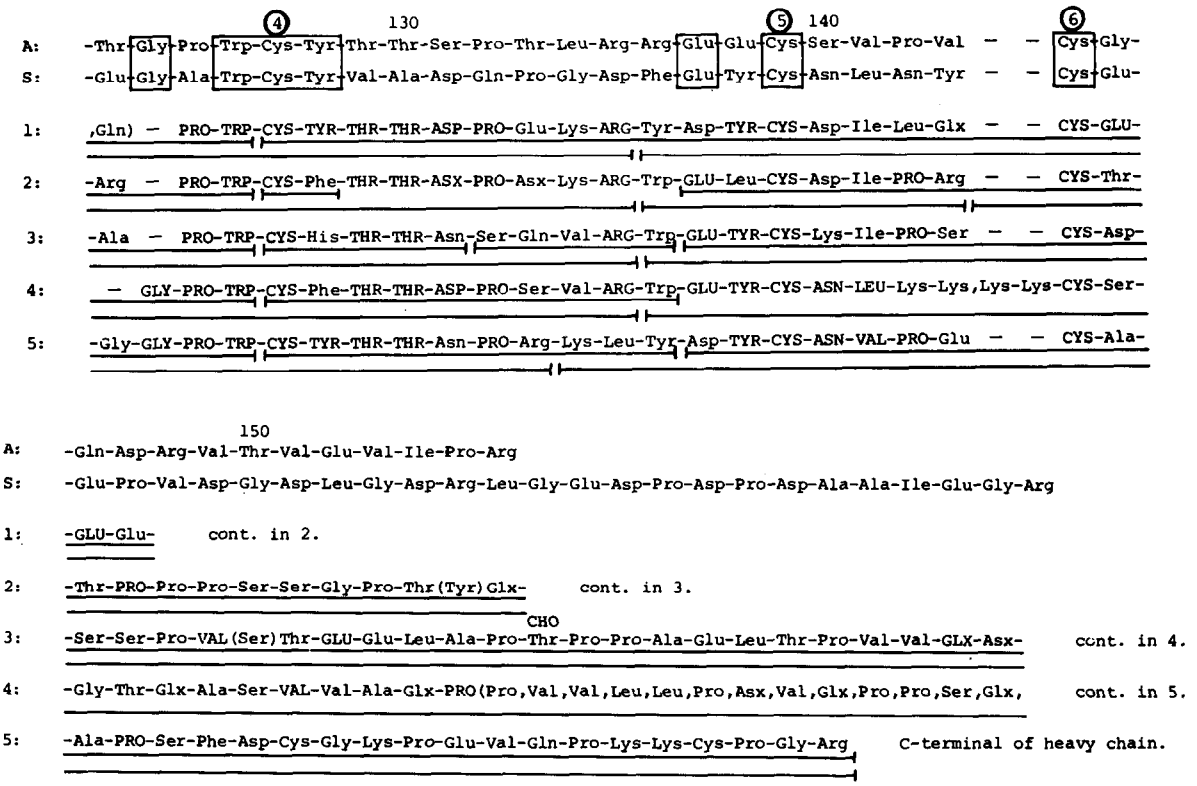


Fig.1. (A) Amino acid sequence of A-fragment from bovine prothrombin (residues 1–156) [4,5,12]. Gla: Tentative three-letter code for γ-carboxyglutamic acid [12]. (S) Sequence of S-fragment from prothrombin (residues 157–274) [4,5]. Boxed residues: Identical positions in A- and S-fragments. Cys-residues in kringle loops indicated by encircled numbers. Residues 275–582 of prothrombin are not shown. Lines 1–5: Partial sequence of heavy chain of human plasmin (residues 76–approx. 587 in plasminogen). Residues in three capital letters identical to corresponding residue in A- and/or S-fragment. Bracketed residues not yet sequenced. Commas outside brackets indicate non-established overlaps. Line 1 starts with Glu-76 and line 5 ends with the C-terminal –Arg of the heavy chain. The sequence is underlined twice. Upper set of lines: chymotryptic peptides. Lower set: citraconylated tryptic peptides (only larger variants). CHO: attached carbohydrate.

and the extra 21 residues [13] would be about 587 residues. The alignment shown in fig.1 (lines 1–5) is in reasonable agreement with the amino acid compositions, carbohydrate distribution and N-terminal sequences (total of 52 residues) of the CNBr-fragments recently described by Wiman and Wallén [15].

Our present data on the sequence show that the heavy chain contains certainly three (marked 1, 2 and 5 in fig.2) and most probably five regions of extensive sequence homology with the two kringle regions in prothrombin. Despite the fact that no information on the location of the disulphide bridges in plasminogen is yet available we see no good reason to expect that this extensive sequence homology is not reflected in close tertiary structure similarity. In order to facililtate visual comparison with the prothrombin kringle regions [4,5] the heavy chain sequence data have therefore been arranged (see fig.2) as five kringle regions with connecting strands of different size. The extent of sequence identity with either or both of the corresponding positions in the prothrombin kringles is approx. 55% in those parts of the heavy chain kringle regions that we have sequenced until now. This indicates beyond reasonable doubt that a partial gene 'quintiplication' has occurred in the evolution of plasminogen which involves the same basic structure that has been duplicated in the evolution of prothrombin, proving a common ancestry between the non-serine protease parts of prothrombin and plasminogen.

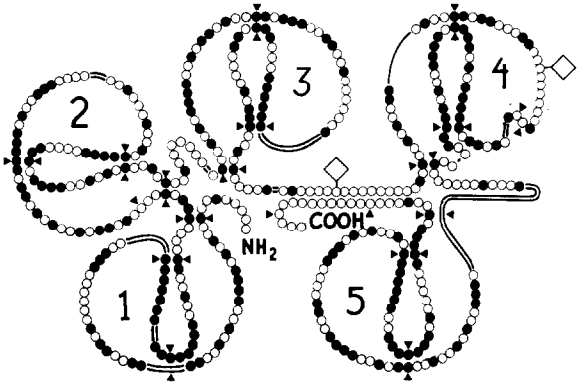


Fig.2. Partial sequence data from fig.1 of residues 76–approx. 587 of heavy chain of plasmin arranged as five kringle regions (marked 1–5) with 'connecting strands'. The kringles represent the sequences from Cys-1 to Cys-6 of lines 1–5 in fig.1. Circles: Sequenced residues. Filled circles: Identities with corresponding residue in A- and/or S-kringle of prothrombin. Double lines: Stretches of known composition, not sequenced. Single line: Stretch not yet found. Squares: Attached carbohydrate. Filled triangles point at Cys-residues.

Acknowledgements

We thank Lajla Brøns, Lene Kristensen, Lone Christensen, Ane Holm-Petersen and Margit Skriver for technical assistance and Hans Bennich for providing access to a Beckman 121M amino acid analyzer at the Biomedical Center, Univerity of Uppsala. L.S. and T.E.P. received fellowships from the Danish Science Research Council. The work was supported by the U.S. National Heart and Lung Institute (grant no. HL 16238 HEM). Dr J. Jørgensen, the Blood Group Laboratory, University Hospital, Århus, kindly provided human plasma.

References

[1] Sottrup-Jensen, L., Zajdel, M., Claeys, H., Petersen, T. E. and Magnusson, S. (1975) Proc. Nat. Acad. Sci. USA 72, 2577–2581.
[2] Robbins, K. C., Bernabe, P., Arzadon, L. and Summaria, L (1973) J. Biol. Chem. 248, 7242–7246.
[3] Robbins, K. C., Summaria, L., Hsieh, B. and Shah, R. J. (1967) J. Biol. Chem. 242, 2333–2342.
[4] Magnusson, S., Sottrup-Jensen, L., Petersen, T. E. and Claeys, H. (1975) in: Prothrombin and Related Coagulation Factors (Boerhaave Symposium, May 10–11, 1974), (Hemker, H. C. and Veltkamp, J., eds.), pp. 25–46, University Press, Leiden, Holland.
[5] Magnusson, S., Petersen, T. E., Sottrup-Jensen, L. and Claeys, H. (1975) in: Proteases and Biological Control (Cold Spring Harbor Conference, Sep. 10–15, 1974) (Reich, E., Rifkin, D. and Shaw, E., eds.), in the press, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[6] Wiman, B. and Wallén, P. (1975) Eur. J. Biochem. 50, 489–494.
[7] Deutsch, D. G. and Mertz, E. T. (1970) Science 170, 1095–1096.
[8] Claeys, H. and Vermylen, J. (1974) Biochim. Biophys. Acta 342, 351–359.
[9] Liu, T.-Y. (1972) in: Methods in Enzymology (Hirs, C. H. W. and Timasheff, S. N., eds.), Vol. 25, pp. 44–55. Academic Press, New York.
[10] Dixon, H. B. F. and Perham, R. N. (1968) Biochem. J. 109, 312–314.
[11] Summaria, L., Hsieh, B., Groskopf, W. R. and Robbins, K. C. (1967) J. Biol. Chem. 242, 5046–5052.
[12] Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Morris, H. R. and Dell, A. (1974) FEBS Lett. 44, 189–193.
[13] Laursen, R. A. and Lee, H.-M. (1975) FEBS Lett. 56, 70–72.
[14] Wiman, B. and Wallén, P. (1975) Thrombosis Research 7, 239–243.
[15] Wiman, B. and Wallén, P. Eur. J. Biochem., in the press.