Proc. Nat. Acad. Sci. USA
Vol. 72, No. 8, pp. 3082–3086, August 1975
Biochemistry

Bovine factor X_1 (Stuart factor): Amino-acid sequence of heavy chain

(blood coagulation/homology to serine protease)

K. TITANI, K. FUJIKAWA, DAVID L. ENFIELD, LOWELL H. ERICSSON, KENNETH A. WALSH, AND HANS NEURATH

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

Contributed by Hans Neurath, June 9, 1975

ABSTRACT **The amino-acid sequence of the heavy chain of bovine blood coagulation factor X_1 (Stuart factor) isolated before and after activation has been determined. Sequence analysis was performed on fragments obtained by cleavage with cyanogen bromide and by tryptic digestion. Comparison of the complete sequence with those of other hepatic and pancreatic serine proteases demonstrates homology of the heavy chain of activated factor X_1 (factor X_{1a}) with the B chain of bovine thrombin as well as with bovine trypsin, chymotrypsins A and B, and porcine elastase. The activation peptide cleaved near the amino terminus by a protease from Russell's viper venom differs in both size and sequence from those of other serine proteases. With three exceptions, all of the residues which are important in the catalytic functions of trypsin and chymotrypsin occur in corresponding loci in the heavy chain of factor X_a. These findings suggest that the three-dimensional structure of the heavy chain is similar to that of the pancreatic serine proteases and that these enzymes have evolved from a common ancestral gene.**

Factor X (Stuart factor) is the zymogen of a protease which upon activation by a complex of activated factor IX (factor IX_a) and factor VIII catalyzes the conversion of prothrombin to thrombin (1). Bovine factor X is a glycoprotein with a molecular weight of 55,100. It is composed of two polypeptide chains ("heavy" and "light") held together by one or more disulfide bonds. The intact protein can be separated chromatographically into two components (factor X_1 and X_2) which are identical in molecular weight and amino-acid composition (2–4). Upon activation by the complex of factors IX_a and VIII in the presence of calcium and phospholipid, by factor VII and tissue factor in the presence of calcium, by a protease from Russell's viper venom, or by trypsin, glycopeptides are split, either near the amino or both the amino and carboxyl ends of the heavy chain, and an enzymatically active serine protease, factor X_a, is formed (5–8). The heavy chain of factor X_a is homologous to pancreatic and certain hepatic serine proteases, as evidenced by similarities of the amino-terminal sequence and the sequence surrounding the serine and histidine residues of the active site (9).

The sequence of the light chain of bovine factor X_1 (10), for which the pancreatic proteases have no counterpart, is homologous to the amino-terminal sequences of bovine prothrombin and factor IX (11). All three of these factors are vitamin K dependent, and this dependency appears to be expressed in the post-transcriptional synthesis of γ-carboxyglutamyl residues (12, 13) in the amino-terminal portions of the zymogens.

This communication presents the amino-acid sequence of the heavy chain of bovine factor X_1, and an examination of the structural homology of this chain with those of other serine proteases.

Abbreviation: CHO, carbohydrate.

METHODS

Factor X_1 was isolated from bovine plasma and converted to the active form (factor X_{1a}) by incubation with a protease from Russell's viper venom (4, 5). Factors X_1 and X_{1a} were reduced with dithioerythritol, and *S*-alkylated with iodoacetic acid or 4-vinylpyridine, and the chains were then separated as described (4, 5).

Cyanogen bromide fragments were prepared and separated as described (9). For clarity, fractions I through VI and A through E of the previous paper (9) are designated herein as CB I through CB VI and CB A through CB E. Fraction III was further separated into two fragments, CB III-1 and CB III-2, by recycling gel filtration on a Sephadex G-50 column in 9% formic acid. Acid cleavage of an aspartyl-proline bond in the activation peptide was carried out by the method of Fraser *et al.* (14). Tryptic digests of the whole heavy chain were prepared before and after succinylation of α- and ε-amino groups (15), and the peptides were separated by combinations of gel filtration and ion exchange chromatography.

Amino acid analyses were performed on Beckman (model 120C) and Durrum (model D-500) amino-acid analyzers according to the method of Spackman *et al.* (16). Automatic sequence analyses were performed on the Beckman Sequencer (model 890B) by the method of Edman and Begg (17) as modified by Hermodson *et al.* (18), and on the Sequemat by the method of Laursen (19). Column chromatography of glycopeptides was monitored by reaction with both ninhydrin and phenol–sulfuric acid (20). Hexose, hexosamine, and *N*-acetyl-neuraminic acid were analyzed by the methods of Dubois *et al.* (21), Elson and Morgan as described by Gardell (22), and Warren (23), respectively.

RESULTS

The amino-acid sequences of seven cyanogen bromide fragments of the heavy chains (H_1 and H_{1a}) of bovine factors X_1 and X_{1a}, and of several smaller fragments produced by chemical or enzymatic subfragmentation were determined largely by automatic sequence analysis as shown diagrammatically in Fig. 1. The sequences of the remaining portions of the fragments were determined by conventional methods on small peptides. The cyanogen bromide fragments were aligned with the aid of four tryptic peptides containing methionine (residues 109–126, 193–201, 206–240, and 279–281), which were isolated from the whole heavy chain, H_1.

Tryptic digestion of fragment CB II (37°, pH 8.0 for 1 hr) yielded two large peptides, T-1 and T-2. Peptide T-1 was identical to the activation peptide (residues 1–51) except for

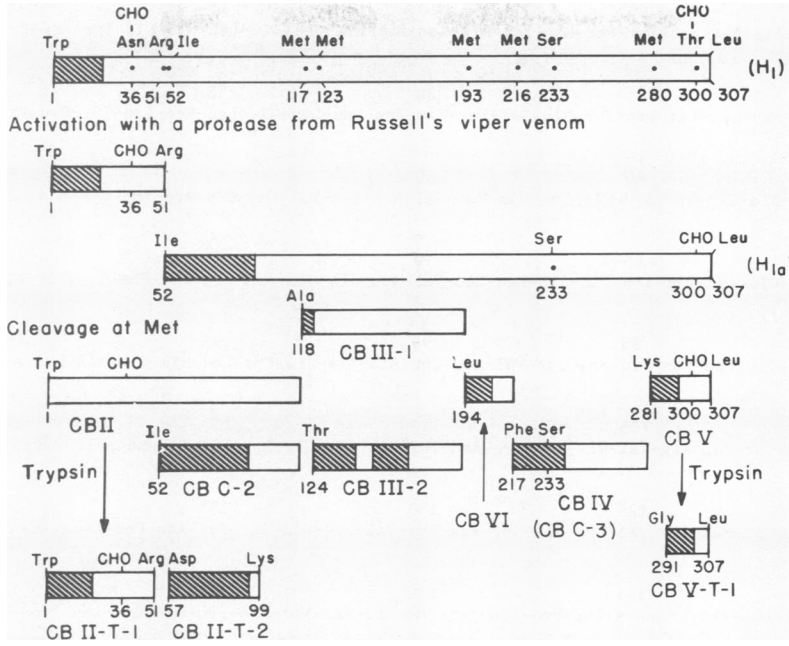

FIG. 1. Diagrammatic summary of fragments generated from the heavy chains of bovine factors X_1 and X_{1a} for sequenator analyses. The top bar represents the heavy chain (H_1) of factor X_1 and the residues that are important for its fragmentation, in addition to those at the amino- and carboxyl-termini, at the active site, and those to which carbohydrate (CHO) is attached. The hatched sections of horizontal bars indicate the portions of sequence determined by sequenator analyses. The sequence of the section in the interior of CB III-2 was determined on a tryptic peptide (residues 152–171) isolated from the whole heavy chain.

a slightly lower content of *N*-acetyl-neuraminic acid which appeared to be caused by partial loss under the acidic conditions of cleavage with cyanogen bromide. Most of the carbohydrate of the heavy chain H_1 (4) was recovered in the activation peptide (30), and was entirely associated with a tetrapeptide, Gly-Leu-Asn-Arg (residues 34–37). This peptide was obtained by digestion of the activation peptide with chymotrypsin followed by prolonged digestion with trypsin (37°, pH 8.0 for 2 days). The attachment of carbohydrate to Asn-36 was assumed from the results of sequence analysis.

The sequence of the activation peptide (or of CB II T-1) was largely determined on the intact peptide. The remainder of the sequence was derived from a peptide (residues 20–51) produced by cleavage of an aspartyl-proline bond, and from a chymotryptic peptide (residues 34–51). The remaining sequence (residues 52–117) of CB II was elucidated, first by sequenator analysis of fragments CB C-2 (9) and CB II T-2 (Fig. 1), which yielded the first 44 amino-terminal residues, and then by conventional analysis of peptides representing residues 96–117.

Methionyl-threonine bonds are known to resist cleavage with cyanogen bromide (24). Hence incomplete cleavage at residues 123–124 yielded two overlapping fragments, CB III-1 (residues 118–193) and CB III-2 (residues 124–193), which were of similar size and charge. They were separated by recycling gel filtration (five cycles) on a Sephadex G-50 column, prior to sequenator analysis. The remainder of the sequence of CB III was determined on tryptic and chymotryptic peptides isolated from a mixture of the two fragments.

Sequenator analysis yielded the amino-terminal sequence of 13 residues of fragment CB VI (residues 194–216). The sequence of the remaining portion of the fragment was determined conventionally on chymotryptic peptides.

Fragment CB IV comprises residues 217–280. Twenty-four amino-terminal residues were previously reported (9) for an identical fragment, CB C-3, of the heavy chain (H_{1a}). The remainder of the sequence of fragment CB IV was determined by conventional methods on nine chymotryptic peptides and 11 tryptic peptides isolated before and after succinylation of fragment CB IV.

Fragment CB V (residues 281–307), which represents the carboxyl-terminal portion of the heavy chain, contained a small amount of carbohydrate [hexosamine 0.92, hexose 2.7, and neuraminic acid (Neu) 0.48 mol/mol of peptide]. The sequence of the fragment through Pro-302 was determined with the sequencer on the intact fragment and on a tryptic peptide (residues 291–307). The carbohydrate was recovered in two chymotryptic peptides (residues 300–307), which were separated on Dowex 50-X2. These two peptides had the same amino-acid composition, but different carbohydrate content, i.e., CB V C-1 ($Thr_{0.86}$, $Pro_{4.09}$, $Val_{1.00}$, $Leu_{1.50}$, $hexosamine_{0.92}$, $hexose_{2.9}$, and $Neu_{0.94}$) and CB V C-2 ($Thr_{0.94}$, $Pro_{4.21}$, $Val_{1.00}$, $Leu_{1.59}$, $hexosamine_{0.91}$, $hexose_{2.4}$, and $Neu_{0.20}$). Subtractive Edman degradations established Thr-300 at their amino termini. Sequence analyses on the Sequemat of the peptides, attached to triethylenetetramine resin, established the remainder of the sequence but yielded no phenylthiohydantoin in the first turn. The absence of threonine phenylthiohydantoin is explained by the fact that only Thr-300 could be the attachment site for carbohydrate in these octapeptides. The carboxyl-terminal leucine was confirmed by carboxypeptidase A digestion. However, the low leucine content both in the peptides and upon digestion with carboxypeptidase A for 24 hr, 37°, pH 8.0 in 4 M urea (0.4 mol/mol of peptide) indicate that the carboxy-terminus is heterogeneous and ends with either Pro-306 or Leu-307.

The complete amino-acid sequence of 307 residues of the heavy chain (H_1) of factor X_1 is shown in Fig. 2. Carbohydrate units are attached to at least two residues, Asn-36 and Thr-300, but the possibility of a third site cannot be entirely excluded. The sequences around Asn-36 and Thr-300 are consistent with those, Asn(CHO)–X–Thr (or Ser) and Thr (or

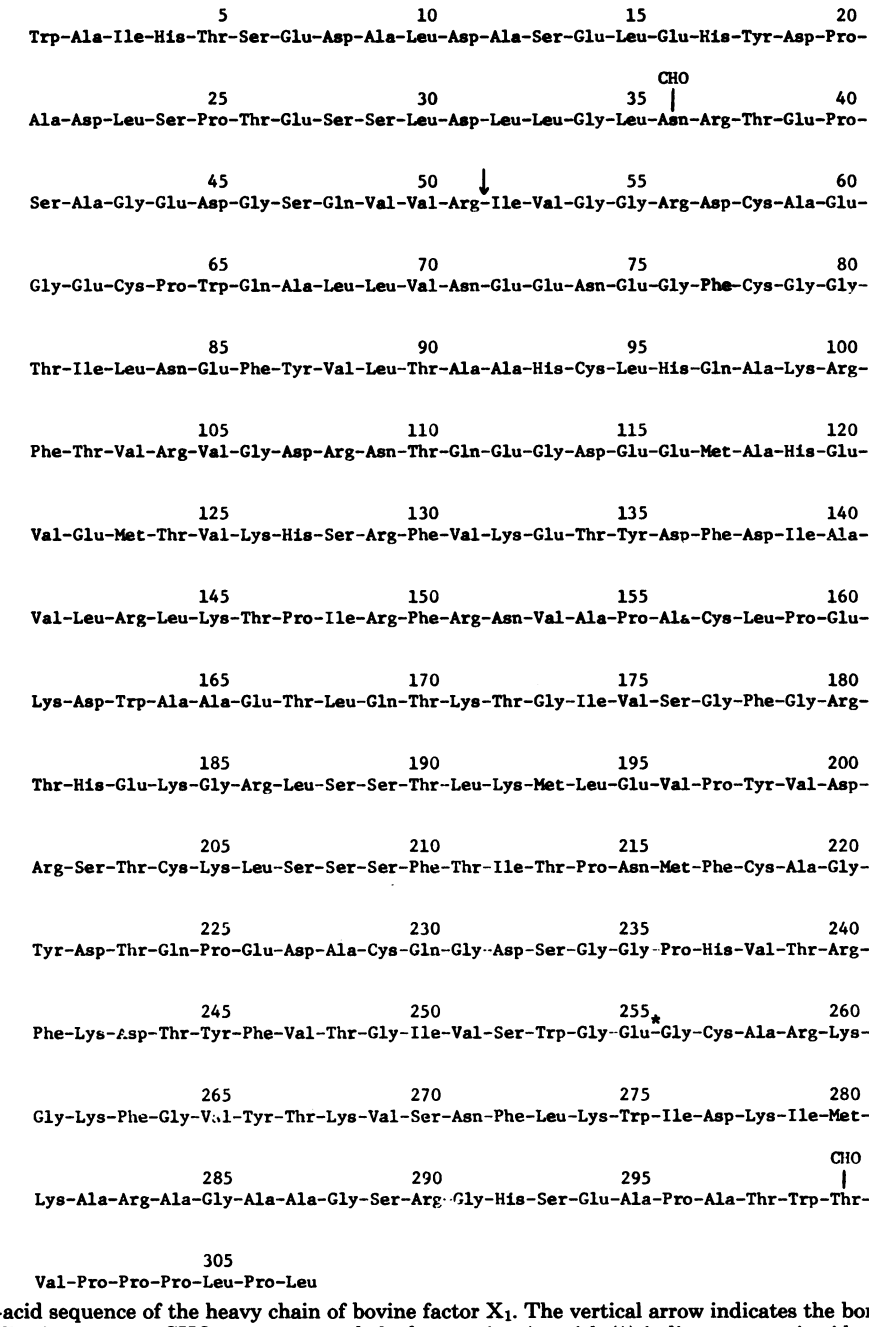

FIG. 2. The amino-acid sequence of the heavy chain of bovine factor X_1. The vertical arrow indicates the bond cleaved upon activation by a protease of Russell's viper venom. CHO represents carbohydrate units. Asterisk (*) indicates tentative identification. Overlap of Arg-151 to Asn-152 is also tentative.

Ser) (CHO)–X–X-Pro, proposed by Winzler (25) and Jollès *et al.* (26), respectively.

The total amino-acid composition of the heavy chain (H_1) calculated from the sequence (Asp_{18}, Asn_8, Thr_{25}, Ser_{20}, Glu_{25}, Gln_7, Pro_{16}, Gly_{27}, Ala_{25}, Cys_9, Val_{21}, Met_5, Ile_{10}, Leu_{23}, Tyr_7, Phe_{13}, Trp_6, Lys_{16}, His_9, and Arg_{17}) is in good agreement with that obtained by analysis (4). The calculated molecular weight of the peptide portion of the heavy chain (H_1) is 33,773. With carbohydrate added (approximately 10% of factor X_1 by weight) the heavy chain has a molecular weight of 39,300 which is in good agreement with the value (38,000) obtained by ultracentrifugal analysis (4).

DISCUSSION

The amino-acid sequence of the heavy chain of factor X_1 reported herein is complete except for the disulfide pairing of the nine half-cystine residues. Preliminary experiments indicate the presence of four intra-chain disulfide bonds between residues 58 and 63, 78 and 94, 204 and 218, 229 and 257, and one inter-chain bond between residue 157 and the light chain. Of these four intra-chain pairs, three are present in corresponding positions in all of the pancreatic serine proteases (27). The sequence of factor X_2 has not yet been completed, but partial sequence data (approximately 50% of the heavy chain) give no indications of differences. Jackson *et al.* (3) suggested that the two proteins differ in carbohydrate content, while Fujikawa *et al.* (4) found no difference in either amino-acid or carbohydrate composition of the two proteins.

As already discussed (11), the peptide portion of bovine factor X_1 (447 residues) is apparently smaller than that of prothrombin. Precursor proteins of the three vitamin K-de-

Table 1. Functional residues in hepatic and pancreatic serine proteases*

Function	Residue number		Enzyme Residue†					
	Chymotrypsin	Factor X_{1a} heavy chain	Factor X_{1a}‡ heavy chain	Thrombin B-chain	Trypsin	Chymotrypsin A	Chymotrypsin B	Elastase
"Charge-relay" system	57	93	His	His	His	His	His	His
	102	138	Asp	Asp	Asp	Asp	Asp	Asp
	195	233	Ser	Ser	Ser	Ser	Ser	Ser
	214	252	Ser	Ser	Ser	Ser	Ser	Ser
Ion-pair	16	52	Ile	Ile	Ile	Ile	Ile	Val
	194	232	Asp	Asp	Asp	Asp	Asp	Asp
H-bond to Gly-193	32	68	*Leu*	Met	Ser	Ser	Ser	Ser
	40	76	*Gly*	Leu	His	His	His	His
Orientation of Ser-195	193	231	Gly	Gly	Gly	Gly	Gly	Gly
	196	234	Gly	Gly	Gly	Gly	Gly	Gly
Substrate binding pocket	189	227	Asp	Asp	Asp	Ser	Ser	Ser
	190	228	*Ala*	Ala	Ser	Ser	Ser	Gly
	191	229	Cys	Cys	Cys	Cys	Cys	Cys
	192	230	Gln	Glu	Gln	Met	Met	Gln
H-bonded structure anti-parallel to substrate	214	252	Ser	Ser	Ser	Ser	Ser	Ser
	215	253	Trp	Trp	Trp	Trp	Trp	Phe
	216	254	Gly	Gly	Gly	Gly	Gly	Val
	226	264	Gly	Gly	Gly	Gly	Ala	Thr
P_2-site	99	135	*Tyr*	Leu	Leu	Ile	Val	Val
Identity with factor X_{1a} heavy chain (%)				47	40	34	33	33

* All enzymes are of bovine origin, except porcine elastase.
† All sequences except that of factor X_{1a} heavy chain are numbered according to bovine chymotrypsinogen A.
‡ Residues which differ in factor X_{1a} heavy chain from the corresponding residues in trypsin are printed in *italics*.

pendent blood coagulation factors, prothrombin and factors IX and X, are homologous and appear to have evolved from a common ancestral gene; however, only factor X_1 is composed of two peptide chains. Mattock and Esnouf (28) have claimed that bovine factor X may occur as a single polypeptide chain, which is subsequently cleaved to yield the two chains commonly isolated. These facts together suggest that bovine factor X may be synthesized as a single polypeptide and that a peptide segment is cleaved from the middle of the chain prior to or during isolation.

The heavy chain of factor X_a contains almost all of the functional residues identified in pancreatic serine proteases and hence, as discussed below, is the carrier of catalytic activity. The light chain strongly binds calcium, and probably phospholipid (which enhance the activity of factor X_a in the presence of factor V) and thus appears to regulate the activity of the heavy chain.

As described by Fujikawa *et al.* (29, 30), a heptadecapeptide containing carbohydrate (residues 291–307) is cleaved from the carboxyl-terminal portion of the heavy chain of bovine factor X during activation by factor IX_a in the presence of phospholipid. Of the various plasma and pancreatic serine proteases, only bovine factor X has this extended carboxyl-terminal portion (residues 284–307). This peptide, having an unusual content of proline and carbohydrate, appears to be unnecessary for the activity of factor X_a, since an alternative method of activation, by a protease from Russell's viper venom, leaves the peptide attached.

The activation peptide released from the amino terminus of the heavy chain (residues 1–51) differs in size and sequence from those of bovine trypsinogen and chymotrypsinogen, and appears to be only distantly related to the A-chain of bovine thrombin (31). The functional or physiological role of the two carbohydrate-containing activation and degradation peptides of factor X, if any, is not yet known. In analogous systems, the activation peptides of bovine trypsinogen (32) and complement C 3 (33) are reported to have pharmacological activity.

The sequence of the heavy chain of bovine factor X_{1a} (residues 52–283) is homologous to those of other plasma (31) and pancreatic serine proteases (27). Although a complete comparison among the plasma enzymes is not yet possible, the degrees of identity can be calculated by aligning the available sequences so as to yield maximum homology. As shown in Table 1, the heavy chain of factor X_{1a} and the B-chain of thrombin (31) show the highest degree of identity (47%) as might be expected from the similar functional role of the two proteins (34). This is of the same order as the identity between bovine trypsin and chymotrypsin A (46%). Among the pancreatic serine proteases, trypsin shows the highest degree of identity with the heavy chain of factor X_{1a} (40%), as might be expected from the common specificity of these two enzymes toward small substrates (35). Homology is less evident at the carboxyl-terminus, since the heavy chain of factor X_{1a} contains approximately 24 extra residues when oompared to the pancreatic enzymes. Table 1 also provides convincing evidence of homology when functional residues (36–38) are compared. Components of the "charge-relay system" of pancreatic serine proteases are all present in the corresponding loci of the heavy chain of bovine factor X_{1a} (His-93, Asp-138, Ser-233). A few changes occur among residues which form the substrate binding pocket of bovine trypsin and chymotrypsin A. These types of changes are also observed among pancreatic serine proteases and explain the different substrate specificities of these enzymes. Histidine-40 in chymotrypsinogen, which may have a special significance among the pancreatic enzymes, is replaced by glycine-76 in the heavy chain of factor X_{1a}. Nevertheless, these

results suggest that the three-dimensional structure of the heavy chain of factor X_{1a} is similar to those of the pancreatic serine proteases, and that its active site is similar to that of pancreatic trypsin. If, indeed, the conformation of the heavy chain of factor X_{1a} is similar to that of bovine trypsin, approximately 63% of the internal residues would occupy identical positions in these two proteins. Verification of these assumptions must await an x-ray analysis of the structure of bovine factor X_a.

The authors wish to thank Dr. E. W. Davie for valuable discussion and Dr. M. A. Hermodson for contribution to the initial phase of this work. Thanks are also due to Santosh Kumar, Richard Granberg, Douglas Marshall, and Judy Lane for excellent technical assistance. This work was supported in part by the National Institutes of Health (GM 15731 and HL 11857) and by the American Cancer Society (BC 91P). D.L.E. is an Associate Investigator of the Hughes Medical Institute.

1. Davie, E. W. & Fujikawa, K. (1975) *Annu. Rev. Biochem.* **44,** 799–829.
2. Jackson, C. M. & Hanahan, D. J. (1968) *Biochemistry* **7,** 4506–4517.
3. Jackson, C. M. (1972) *Biochemistry* **11,** 4873–4882.
4. Fujikawa, K., Legaz, M. E. & Davie, E. W. (1972) *Biochemistry* **11,** 4882–4891.
5. Fujikawa, K., Legaz, M. E. & Davie, E. W. (1972) *Biochemistry* **11,** 4892–4899.
6. Fujikawa, K., Legaz, M. E. & Davie, E. W. (1974) *Biochemistry* **13,** 4508–4516.
7. Radcliffe, R. D. & Barton, P. G. (1973) *J. Biol. Chem.* **248,** 6788–6795.
8. Jesty, J. & Nemerson, Y. (1974) *J. Biol. Chem.* **249,** 509–515.
9. Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H. & Davie, E. W. (1972) *Biochemistry* **11,** 4899–4902.
10. Enfield, D. L., Ericsson, L. H., Walsh, K. A., Neurath, H. & Titani, K. (1975) *Proc. Nat. Acad. Sci. USA* **72,** 16–19.
11. Enfield, D. L., Ericsson, L. H., Fujikawa, K., Titani, K., Walsh, K. A. & Neurath, H. (1974) *FEBS Lett.* **47,** 132–135.
12. Stenflo, J. (1974) *J. Biol. Chem.* **249,** 5527–5535.
13. Stenflo, J., Fernlund, P., Egan, W. & Roepstorff, P. (1974) *Proc. Nat. Acad. Sci. USA* **71,** 2730–2733.
14. Fraser, K. L., Poulson, K. & Haber, E. (1972) *Biochemistry* **11,** 4974–4977.
15. Yaoi, Y., Titani, K. & Narita, K. (1964) *J. Biochem. Jap.* **56,** 222–229.
16. Spackman, D. H., Stein, W. H. & Moore, S. (1958) *Anal. Chem.* **30,** 1190–1206.
17. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1,** 80–91.
18. Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H. & Walsh, K. A. (1972) *Biochemistry* **11,** 4493–4502.
19. Laursen, R. A. (1971) *Eur. J. Biochem.* **20,** 89–102.
20. Hirs, C. H. W. (1967) in *Methods in Enzymology,* ed. Hirs, C. H. W. (Academic Press, New York), Vol. 11, pp. 411–413.
21. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) *Anal. Chem.* **28,** 350–356.
22. Gardell, S. (1957) in *Methods of Biochemical Analysis,* ed. Glick D. (Interscience Publishers, New York), Vol. 6, pp. 289–317.
23. Warren, L. (1959) *J. Biol. Chem.* **234,** 1971–1975.
24. Schroeder, W. A., Shelton, J. B. & Shelton, J. R. (1969) *Arch. Biochem. Biophys.* **130,** 551–556.
25. Winzler, R. J. (1973) in *Hormonal Proteins and Peptides,* ed. Li, C. H. (Academic Press, New York), Vol. 1, pp. 1–15.
26. Jollès, J., Schentgen, F., Alais, C., Fiat, A. M. & Jollès, P. (1972) *Helv. Chim. Acta* **55,** 2872–2883.
27. Shotten, D. M. & Hartley, B. S. (1970) *Nature* **225,** 802–806.
28. Mattock, P. & Esnouf, M. P. (1973) *Nature New Biol.* **242,** 90–92.
29. Fujikawa, K., Coan, M. H., Legaz, M. E. & Davie, E. W. (1974) *Biochemistry* **13,** 5290–5299.
30. Fujikawa, K., Titani, K. & Davie, E. W. (1975) *Proc. Nat. Acad. Sci. USA,* in press for September.
31. Magnusson, S. (1971) *The Enzymes,* ed. Boyer, P. D. (Academic Press, New York), 3rd Ed., Vol. 3, pp. 277–321.
32. Abita, J. P., Moulin, A. & Lazdunski, M. (1973) *FEBS Lett.* **34,** 251–255.
33. Hugli, T. E., Vallota, E. H. & Müller-Eberhard, H. J. (1975) *J. Biol. Chem.* **250,** 1472–1478.
34. Fujikawa, K., Coan, M. H., Enfield, D. L., Titani, K., Ericsson, L. H. & Davie, E. W. (1974) *Proc. Nat. Acad. Sci. USA* **71,** 427–430.
35. Esnouf, M. P. & Williams, W. J. (1962) *Biochem. J.* **84,** 62–71.
36. Stroud, R. M., Kay, L. M. & Dickerson, R. E. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 125–140.
37. Hartley, B. S. (1968) *Phil. Trans. R. Soc. Ser B* **257,** 77–87.
38. Segal, D. M., Powers, J. C., Cohen, G. H., Davies, D. R. & Wilcox, P. E. (1971) *Biochemistry* **10,** 3728–3738.