Suzuki, K., & Imahori, K. (1978b) J. Biochem. (Tokyo) 84, 1031–1039.
Tilby, M., Hindennach, I., & Henning, U. (1978) J. Bacteriol. 136, 1189–1191.
Weber, K., & Osborn, M. (1969) J. Biol. Chem. 244, 4406–4412.
Yamamoto, H., Nishida, K., Beppu, T., & Arima, K. (1978) J. Biochem. (Tokyo) 83, 827–834.

Amino Acid Sequence of the Light Chain of Bovine Factor X_1 (Stuart Factor)†

David L. Enfield,‡ Lowell H. Ericsson, Kazuo Fujikawa, Kenneth A. Walsh, Hans Neurath, and Koiti Titani*

ABSTRACT: The detailed proof of the amino acid sequence of the 140 residues (16 193 daltons) of the light chain of bovine factor X_1 (Stuart factor) is presented. Sequence analyses were performed on fragments obtained after chemical cleavage of asparagine–glycine and tryptophanyl peptide bonds and after various enzymatic digestion procedures. Twelve γ-carboxyglutamyl residues are clustered in the amino-terminal 39 residues and 13 half-cystine residues are found in the carboxyl-terminal 91 residues, suggesting two domains in the light chain, one exceptionally anionic and the other extensively cross-linked by disulfides.

Factor X (Stuart factor) is the zymogen of a protease which participates in the middle phase of a series of activation reactions which regulate blood coagulation (Davie & Fujikawa, 1975). The zymogen is activated by factor IX_a in the presence of calcium, phospholipid, and factor VIII (Lundblad & Davie, 1964, 1965; Biggs & MacFarlane, 1965; Hougie et al., 1967; Barton, 1967; Osterud & Rapaport, 1970; Fujikawa et al., 1974b) in the intrinsic pathway or by a tissue factor and activated factor VII (Williams & Norris, 1966; Nemerson & Pitlick, 1970; Osterud et al., 1972; Jesty & Nemerson, 1974; Fujikawa et al., 1974b) in the extrinsic pathway. It can also be activated under nonphysiological conditions by various proteases such as trypsin and a protease from Russell's viper venom (RVV-X) (Fujikawa et al., 1972b). Together with factor V, calcium, and phospholipid, factor X_a catalyzes the conversion of prothrombin to thrombin (Papahadjopoulos & Hanahan, 1964; Barton et al., 1967; Jobin & Esnouf, 1967; Jesty & Esnouf, 1973; Suttie & Jackson, 1977).

Factor X is a glycoprotein of molecular weight 55 000, consisting of a light chain having a molecular weight of 17 000 and a heavy chain having a molecular weight of 38 000 (Fujikawa et al., 1972a; Jackson, 1972). These chains are linked by a single disulfide bond (Titani et al., 1975). Bovine factor X can be fractionated chromatographically into two components (factors X_1 and X_2) possessing similar chemical and biological properties (Jackson & Hanahan, 1968). The protein contains ~10% carbohydrate associated with the heavy chain. During activation in the absence of phospholipid, a single peptide bond is cleaved from the amino terminus of the heavy chain, releasing a large glycopeptide (Fujikawa et al., 1972b, 1974a; Radcliffe & Barton, 1973; Jesty et al., 1974). During activation in the presence of phospholipid, another glycopeptide is also released from the carboxyl terminus of the heavy chain (Jesty et al., 1974), but the cleavage of the second peptide bond is unrelated to the activation process (Fujikawa et al., 1975). The heavy chain of activated factor X_{1a} is homologous to trypsin and other mammalian serine proteases, including the B chain of thrombin and the heavy chain of factor IX_a (Titani et al., 1972, 1975; Enfield et al., 1974).

The light chain of factor X has no counterpart in the zymogens of the pancreatic serine proteases. A preliminary analysis of its structure revealed that the amino-terminal region is homologous to the corresponding regions of prothrombin, factors VII and IX, protein C and protein S, the other vitamin K dependent plasma proteins (Fujikawa et al., 1974a; Kisiel & Davie, 1975; Fernlund et al., 1978; Stenflo & Jonsson, 1979). The 12 residues of γ-carboxyglutamic acid, an amino acid first identified in prothrombin (Stenflo et al., 1974; Nelsestuen et al., 1974; Magnusson et al., 1974), are located in the amino-terminal portion of the light chain of factor X (Bucher et al., 1976; Thorgersen et al., 1978).

This communication presents the detailed proof of the amino acid sequence of the light chain of factor X_1, as determined largely by automated Edman degradation of fragments in a sequenator. A preliminary account of this work has been published (Enfield et al., 1975).

Materials and Methods

TPCK-trypsin, α-chymotrypsin, and carboxypeptidases A and B were obtained from Worthington. Before use, chymotrypsin was treated with α-*N*-tosyllysine chloromethyl ketone to inactivate trypsin. Thermolysin was a product of Calbiochem. Aminopeptidase M was purchased from Rohm and Haas. A glutamyl-specific protease from *Staphylococcus aureus* was a gift from Dr. Gabriel Drapeau (University of Montreal). 4-Vinylpyridine monomer (practical grade) from J. T. Baker or Aldrich was further purified by vacuum distillation and stored at −20 °C in the dark. Reagent grade pyridine was redistilled from ninhydrin. Glycinamide hydrochloride was purchased from Sigma, and its purity was confirmed by high-voltage paper electrophoresis at pH 3.75. Hydroxylamine hydrochloride was from Merck and guanidine

† From the Howard Hughes Medical Institute Laboratory at the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received September 6, 1979.* Supported by research grants from the National Institutes of Health (GM 15731 and HL 11857).

* Address correspondence to this author at the University of Washington. K.T. is an Investigator of the Howard Hughes Medical Institute.
‡ D.L.E. was an Associate Investigator of the Howard Hughes Medical Institute. Present address: U.S. Public Health Service Hospital, Department of Medicine/Hematology, Seattle, WA 98114.

0006-2960/80/0419-0659$01.00/0

hydrochloride from Heico. BNPS-skatole[1] and reagents used in the Sequencer were obtained from Pierce.

Bovine factor X_1 was isolated and purified as described previously (Fujikawa et al., 1972a). The protein was reduced and S-pyridylethylated, and the two chains were separated as described (Fujikawa et al., 1972b).

Chemical Cleavage. Asparagine–glycine bonds were cleaved by hydroxylamine using a modification of the procedures described by Balian et al. (1971). A solution of 102 mg of S-pyridylethylated light chain in 25 mL of 6 M guanidine hydrochloride and 2 M hydroxylamine hydrochloride was prepared at ambient temperature with sufficient 4.5 M lithium hydroxide added to yield a pH of 9.0. During the reaction, the pH was maintained by the addition of 4.5 M LiOH. On the basis of an initial experiment, a 4-h reaction time was found to be optimum.

The pyridylethylated light chain (42.5 mg) was cleaved at tryptophanyl residues by the method of Omenn et al. (1970) except that the protein (10 mg/mL) was treated with BNPS-skatole (50-fold molar excess over tryptophan) in the dark in 80% acetic acid for 4 h at room temperature. Excess reagent was removed by extracting 3 times with 2 volumes of 1-chlorobutane. The aqueous layer was diluted 10-fold with water and lyophilized. The peptides were fractionated by chromatography on a column (1.5 × 88 cm) of Sephadex G-75, equilibrated, and eluted with 5% acetic acid or with 9% formic acid containing 6 M urea. Only a single peptide peak was observed.

Enzymatic Cleavage. Digestion with trypsin or chymotrypsin was performed at pH 8.0, 37 °C, for 2–6 h by using an enzyme/substrate molar ratio of 1:50. Digestion with thermolysin took place at pH 7.0 and 50 °C (enzyme/substrate molar ratio of 1:100) for 1 h and digestion with *S. aureus* protease (Houmard & Drapeau, 1972) at pH 7.9 and 37 °C for 4 h (enzyme/substrate weight ratio of 1:40). Subdigestion of a "core" peptide (see Results) with pepsin took place at 23 °C in 10% formic acid for 15 min (enzyme/substrate molar ratio of 1:200). Digestion with carboxypeptidases A and B (Ambler, 1967) and aminopeptidase M (Jackson & Hirs, 1970) followed published procedures.

Sequence Determination. Automated Edman degradations were accomplished with a Beckman Sequencer (Model 890B) as previously described (Edman & Begg, 1967; Hermodson et al., 1972). The sequences of small peptides were determined manually by the subtractive method (Shearer et al., 1967) or by a modification of the direct method of Peterson et al. (1972).

Analytical Techniques. Amino acid analyses were performed on Beckman Model 120C or Durrum Model D-500 amino acid analyzers after acid (Moore & Stein, 1963) or alkaline (Hugli & Moore, 1972) hydrolyses. Amidated residues of small peptides were identified by digestion with aminopeptidase M or by examination of peptide mobility during paper electrophoresis (2000 V) at pH 6.5. Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate followed the method of Weber & Osborn (1969). Sephadex column effluents were monitored at 280 nm with an LKB Uvicord II (Model 8300).

Results

The amino acid sequence of the reduced and S-pyridylethylated light chain of bovine factor X_1 was largely determined by automated Edman degradation of the intact polypeptide chain and of five large peptide fragments. Since the light chain is devoid of methionyl residues (Table I), fragments were generated by specific cleavages at asparagine–glycine, arginyl, glutamyl, and tryptophanyl peptide bonds. From each mixture of fragments only peptides useful for the sequence determination were selectively isolated and analyzed. The final sequence and a summary of the logistic of its determination are shown in Figure 1.

Table I: Amino Acid Composition of Factor X_1 Light Chain and of Core Peptides[a]

amino acid	whole light chain	calcd from sequence (Figure 1)	tryptic core	chymotryptic core	peptic core
Asp	13.9	15	4.2 (4)	4.0 (4)	3.0 (4)
Thr[b]	6.1	6	1.2 (1)	1.1 (1)	1.0 (1)
Ser[b]	10.8	11	2.8 (3)	2.5 (3)	1.9 (2)
Glu	27.0	27	14.9 (14)	13.9 (14)	14.1 (14)
Pro	2.4	2	0	0	0
Gly	14.2	14	1.5 (1)	1.2 (1)	1.2 (1)
Ala	6.2	6	4.0 (4)	3.8 (4)	4.0 (4)
1/2-Cys[c]	15.2	15	2.2 (2)	2.0 (2)	2.2 (2)
Val[d]	5.1	5	2.0 (2)	1.7 (2)	1.9 (2)
Ile[d]	2.1	2	0	0	0
Leu	6.8	7	4.0 (4)	3.4 (4)	3.9 (4)
Tyr	3.0	3		0.6 (1)	
Phe	8.1	8	3.0 (3)	2.8 (3)	3.0 (3)
Lys	7.0	7	2.2 (2)	1.4 (2)	1.0 (1)
His	3.0	3	0	0	0
Arg	8.0	8	2.3 (2)	2.0 (2)	2.0 (2)
Trp[e]	1.1	1	ND (1)	ND (1)	ND (0)
total residues	140.0	140	(43)	(44)	(40)

[a] Values are expressed as residues per molecule. Values in parentheses correspond to the sequence of residues 1–43 (tryptic core), 1–44 (chymotryptic core), and 1–40 (peptic core). [b] By extrapolation to zero time. [c] Determined as *S*-(pyridylethyl)cysteine (Hermodson et al., 1973). [d] Taken from 96-h hydrolysis. [e] Determined after alkaline hydrolysis (Hugli & Moore, 1972).

Amino-Terminal Sequence. Sequencer analysis of the intact *S*-pyridylethyl protein provided the first 40 residues (Figure 1), except for those at positions 35, 36, and 39, which were not identified. Furthermore, 10 glutamyl residues within the first 32 residues were only tentatively identified because their yields were lower than expected (Figure 2). This was particularly striking at positions 6 and 7. The residues at positions 6, 7, 14, 16, 19, 20, 25, 26, 29, 32, 35, and 39 have subsequently been identified as γ-carboxyglutamyl residues (Bucher et al., 1976; Thogersen et al., 1978). These acid-labile amino acid residues would be expected to be decarboxylated slowly to form glutamyl residues during the Edman degradation.

Cleavage at Asparagine–Glycine Bonds. Large fragments were generated by specific cleavage with hydroxylamine. The reaction mixture was fractionated by gel filtration (Figure 3) into four major fractions containing fragments of the following approximate molecular weights, as estimated by sodium dodecyl sulfate gel electrophoresis: A, 17000; B, 11000; C, 6000; D, 3500.

Fraction A contained only intact light chain, indicating incomplete cleavage. Fraction B contained two fragments (Figure 3) in a molar ratio of 95:5 as estimated by sequenator analyses. The sequence of the major component began with Ala-Asn-Ser-Phe-, identical with that of the whole protein. The sequence of the minor component began with Gly-Ile-Gly-Asp-. This minor sequence was also found at the amino

[1] Abbreviation used: BNPS-skatole, bromine adduct of 2-[(2-nitrophenyl)sulfenyl]-3-methylindole (Omenn et al., 1970).

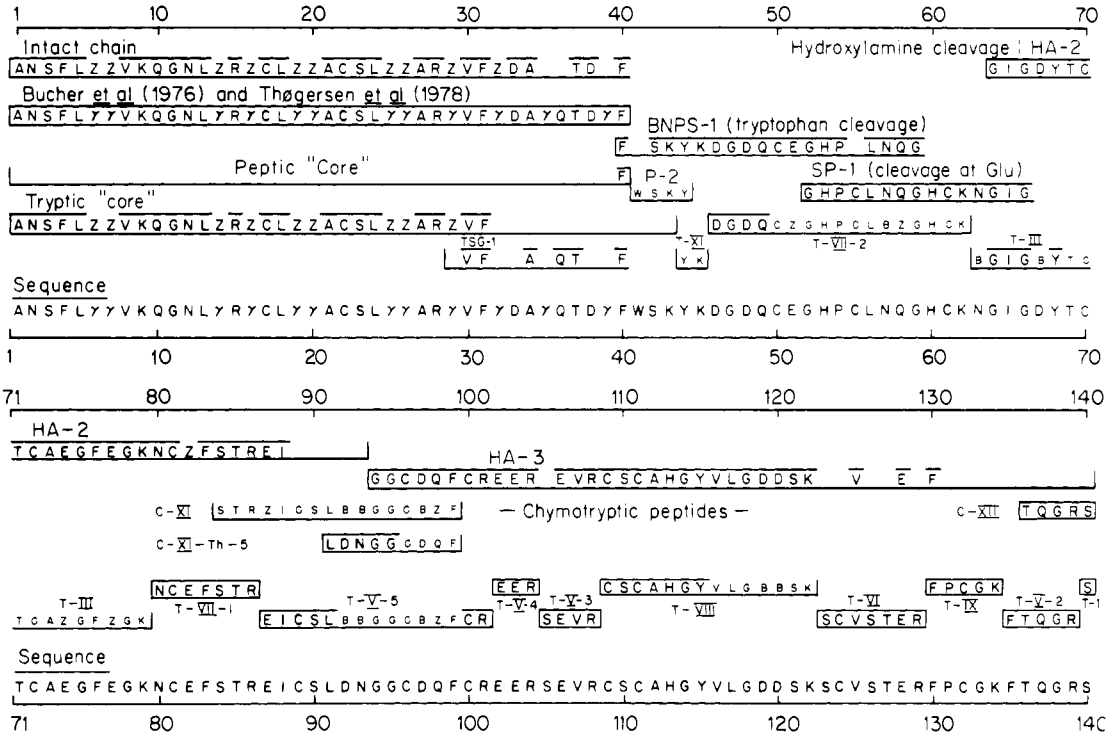

FIGURE 1: Summary of the proof of structure of the light chain of factor X_1. The one-letter code within the bars designates amino acid residues in that peptide. Large capital letters indicate residues identified after Edman degradation or by exopeptidase digestion. Small capital letters indicate placement by composition. The length of each bar indicates the number of residues of the peptide analyzed. An open end at the right indicates that the fragment is actually longer than the bar would indicate. Enclosed tops of the bar indicate that portion of the sequence which is proven; gaps in the upper enclosure indicate portions of sequence not identified. Cystine residues were identified as *S*-(pyridylethyl)cysteine. γ-Carboxyglutamyl residues are those identified by Bucher et al. (1976) or Thogersen et al. (1978) and are specified by the symbol γ. Other one-letter code abbreviations are B (aspartic acid or asparagine) and Z (glutamic acid, glutamine, or γ-carboxyglutamic acid). The remaining conventional abbreviations are listed in Figure 7.

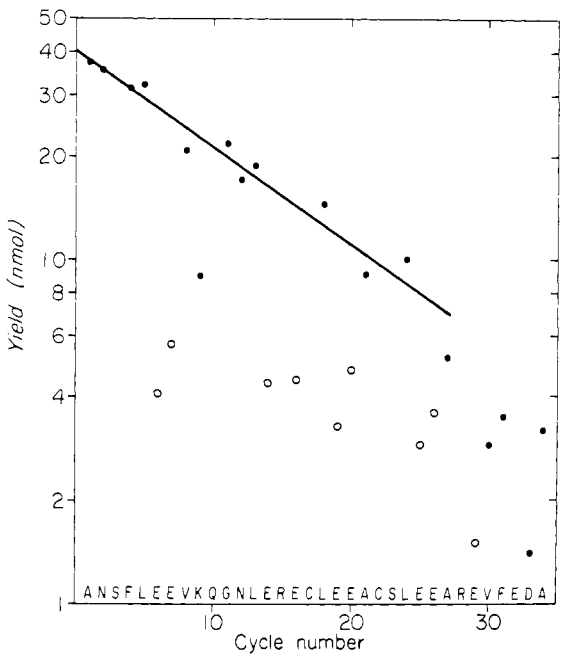

FIGURE 2: Yield of the phenylthiohydantoin at each step of the automated Edman degradation of the amino-terminal region of the light chain of factor X_1. The sequence is indicated at the bottom. Circled points were first identified as glutamyl residues in low yield but later identified by Bucher et al. (1976) and Thogersen et al. (1978) as γ-carboxyglutamyl residues.

terminus of the single component in fraction D. Sequenator analysis of fraction C revealed an 85:15 molar ratio of two fragments. The sequence of the major one began with Gly-Gly-Cys-Asp- and that of the minor one with the amino terminus of the light chain, Ala-Asn-Ser-Phe-. The major sequence was extended without further purification of the fragment as illustrated in Figure 1 for HA-3.

There apparently are two Asn-Gly bonds in the light chain, in the sequences Asn-Gly-Ile-Gly- and Asn-Gly-Gly-Cys-, but neither bond is completely cleaved under the conditions applied. If we denote the three fragments of complete cleavage HA-1, HA-2, and HA-3 (in order starting with the amino terminus), then segments overlapping HA-1/HA-2 and HA-2/HA-3 would be expected. Indeed, taking into account the sequences and molecular weights of fragments in fraction B, this fraction appears to contain both of these overlapping segments with HA-1-2 as the major component (Figure 3). Since segment HA-2-3 is a minor component, the data suggest that the bond in HA-2-3 is cleaved much more readily than that in HA-1-2. Furthermore, segment HA-2-3 must have the amino-terminal sequence Gly-Ile-Gly-. In accord with this reasoning, fraction D contained pure fragment HA-2 (beginning with Gly-Ile-Gly-). Fraction C contained both fragment HA-1 (in low concentration) and fragment HA-3 (the major component, beginning with Gly-Gly-Cys-). Thus, the approximate molecular weights of fragments HA-1, HA-2, and HA-3 are 6000, 3500, and 6000, respectively, and those of segments HA-1-2 and HA-2-3 are 9500 each. It should be noted that the sodium dodecyl sulfate gel electrophoresis data are in approximate accord with these conclusions.

The results of the sequence analyses of fraction D (fragment HA-2) and of the major peptide in fraction C (fragment HA-3) are recorded in Figure 1 in the positions finally established in the whole sequence. For clarity, the ultimately assigned residue numbers will be used in the following presentation of data documenting the alignment of the various fragments in the sequence of the light chain.

Table II: Amino Acid Composition of Tryptic Peptides Derived from Factor X_1 Light Chain[a]

	T-I, 140[e]	T-II, 1–9	T-III, 63–79	T-V-2, 135–139	T-V-3, 105–108	T-V-4, 102–104	T-V-5, 87–101	T-VI, 123–129	T-VII-1, 80–86	T-VII-2, 46–62	T-VIII, 109–122	T-IX, 130–134	T-XI, 44–45
Asp		1.04	1.37				2.82		1.00	3.00	2.15		
Thr			1.69	0.97				1.04	0.98				
Ser	1.00	0.91			0.95		0.88	1.79	0.94		1.82		
Glu		2.25	2.10	1.00	1.14	1.99	2.16	1.06	1.06	3.16			
Pro										1.08		0.86	
Gly			3.90	1.06			1.95			2.83	2.15	1.00	
Ala		0.87	0.94								0.94		
1/2-Cys[b]			2.16				2.91	0.99	1.03	3.24	2.20	0.96	
Val		1.14			1.01			0.92			0.94		
Ile			0.99				0.96						
Leu		0.97					1.00			0.96	1.00		
Tyr			0.96								0.92		0.95
Phe		1.01	1.11	0.95			1.01		1.00			0.91	
His										1.93	1.03		
Lys		1.00	1.04							0.96	1.25	0.98	1.06
Arg				1.00	1.00	1.08	1.00	1.03	1.00				
methods of purifn	*c*	*c*	*c*	*d*	*d*	*d*	*d*	*c*	*d*	*c*	*c*	*c*	*c*
total res (by sequence)	1	9	17	5	4	3	15	7	7	17	14	5	2
yield (%)	37	11	38	33	18	35	23	27	14	10	31	44	17

[a] All values are from 24-h hydrolysates. Prior to hydrolysis none contained tryptophan, as indicated by negative tests with Ehrlich's reagent. [b] Determined as *S*-(pyridylethyl)cysteine. [c] Isolated directly by chromatography on a column of Spinco AA-15 (Bradshaw et al., 1969); roman numeral corresponds to the pooled fractions in Figure 4. [d] Isolated from the indicated pooled fractions in Figure 4 by subfractionation on a column of Bio-Rad AG1W-X2 resin (Wikler et al., 1970). [e] Residue number(s) (see Figure 1).

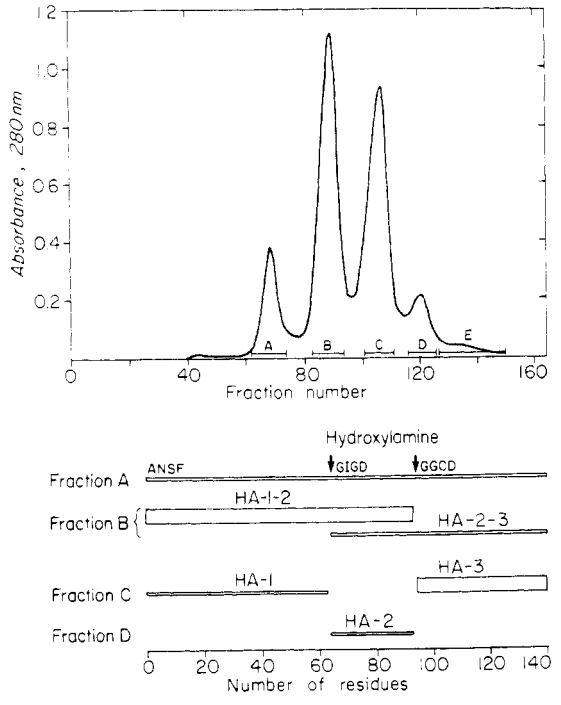

FIGURE 3: (Upper panel) Fractionation of a hydroxylamine digest of pyridylethylated factor X_1 light chain (102 mg). The column (2.5 × 115 cm) of Sephadex G-75 was equilibrated and eluted with 9% formic acid; fractions of 3.4 mL were collected at a flow rate of 10 mL/h. Fractions were pooled as indicated and lyophilized. (Lower panel) Interpretation of the composition of each of the four pooled fractions (see text). The length of each bar is proportional to the number of residues in the fragment. Thick bars indicate high yield; thin bars indicate low yield. Cleavage of the Asn–Gly bond at residues 93–94 (HA-1-2) appears to be much more extensive than that at residues 63–64 (HA-2-3). This is suggested by the relative thickness of the bars. The amino-terminal tetrapeptide sequences of the intact chain and of fragments HA-2 and HA-3 are indicated in one-letter code (see Figure 7).

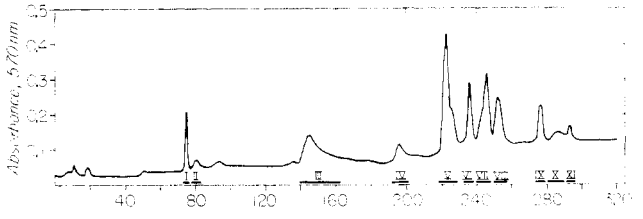

FIGURE 4: Fractionation of the acid-soluble tryptic peptides of the S-pyridylethylated light chain of factor X_1. The digest (4.5 μmol) was chromatographed on a column (2 × 15 cm) of Spinco AA-15 resin equilibrated, developed, and monitored as described previously (Bradshaw et al., 1969). The fractions were pooled as indicated and the peptides recovered by rotary evaporation. Compositions of the peptides are given in Table II, except for the minor peptides derived from pools II and IV and an overlapping peptide in pool X.

Enzymatic Digests of the S-Pyridylethylated Light Chain. Tryptic digests of the S-pyridylethylated light chain consistently yielded a core peptide which precipitated at pH 4 (Table I) and 13 acid-soluble peptides (Table II, Figure 4), 12 of which contained lysine or arginine (Table III). One fragment (T-1) contained only serine and no basic residue and is tentatively assigned to the carboxyl terminus of the protein. The other soluble peptides (excluding T-II) together account for five of the seven lysyl residues and six of the eight arginyl residues of the whole light chain. The remaining lysyl and arginyl residues were found in the acid-insoluble core (Table I). This core is rich in γ-carboxyglutamyl residues and resistant to tryptic digestion. The core peptide in Table I was in fact generated by shorter digestion (1 h) than that described under Materials and Methods for the soluble peptides (6 h) since the longer digestion partially cleaved T-II from the amino terminus of the core (Table II). Sequencer analysis of the core peptide derived by the 1-h digestion placed 31 residues in a sequence (Figure 1) which coincided with the amino-terminal sequence of the intact light chain (Figure 2).

A chymotryptic digest of the S-pyridylethylated light chain yielded several peptides, two of which (C-XI and C-XII) were useful in the structural proof. These are listed in Table IV.

These peptides, together with the sequence analyses of HA-2, HA-3, and the intact chain, permit the deduction of the sequence of the light chain. For simplicity of presentation,

Table III: Tryptic Peptides Derived from Factor X_1 Light Chain[a]

	technique used to prove sequence	residue assignment (Figure 2)
peptides containing Arg		
T-V-2, Phe-Thr-Gln-Gly-Arg	*b*, *c*	135–139
T-V-3, Ser-Glu-Val-Arg	*b*	105–108
T-V-4, Glu-Glu-Arg	*b*	102–104
T-V-5, Glu-Ile-Cys-Ser-Leu-(Asx_3,Gly_2,Cys,Phe,Glx)-Cys-Arg[d]	*c*	87–101
T-VI, Ser-Cys-Val-Ser-Thr-Glu-Arg	*c*	123–129
T-VII-1, Asn-Cys-Glu-Phe-Ser-Thr-Arg	*c*	80–86
peptides containing Lys		
T-II, (Ala,Asx,Ser,Phe,Leu,Glx_2,Val,Lys)	ND[f]	1–9
T-III, X-Gly-Ile-Gly-X-(Tyr,Asx_2,Gly_2,Thr_2,Cys_2,Ala,Glx_2,Phe,Lys)	*b*	63–79
T-VII-2, Asp-Gly-Asp-Gln-(Asx,Gly_2,Glx_2,Cys_3,His_2,Pro,Leu,Lys)	*b*	46–62
T-VIII, Cys-Ser[e]-Cys-Ala-His-Gly-Tyr-(Val,Leu,Gly,Asx_2,Ser,Lys)	*b*	109–122
T-IX, Phe-Pro-Cys-Gly-Lys	*b*	130–134
T-XI, (Tyr,Lys)	ND[f]	44–45
other		
T-1, free Ser		140

[a] Methods of purification and compositions are given in Table II. [b] Direct Edman (manual). [c] Subtractive Edman (manual). [d] Cys-Arg was placed at the carboxyl terminus by analysis of this dipeptide cleaved from T-V-5 by chymotrypsin. [e] Serine was placed by the composition of a peptide (Cys_2,Ser,Ala,His,Gly,Tyr) derived from chymotryptic cleavage of tryptic peptide T-VIII. [f] ND, not determined.

Table IV: Amino Acid Composition of Selected Chymotryptic (C) and Peptic (P) Peptides Derived from Factor X_1 Light Chain[a]

	C-XI, 84–99[h]	C-XI-Th-5, 91–99	C-XI-Th-5,[b] 91–99	C-XII, 136–140	P-1,[c] 42–44	P-2,[c] 41–44	P-3,[c] 41
Asp	2.82	3.00	1.63				
Thr	1.33			0.87			
Ser	1.88			0.95	0.86	0.93	
Asn			0.74				
Gln			0.74				
Glu	2.08	1.07		1.02			
Gly	2.06	2.03	1.58	1.00			
1/2-Cys[d]	2.09	1.26	ND				
Ile	0.93						
Leu	0.99	0.94	1.00				
Tyr					1.00	1.00	
Phe	1.10	1.04	0.74				
Lys					0.91	0.93	
Arg	1.10			0.91			
Trp						0.96	1.00
method of purifn	*e*	*f*		*e*	*g*	*g*	*g*
total res (by sequence)	16	9		5	3	4	1
% yield	47	9		46	11	37	21

[a] Values are expressed as residues per molecule. [b] Aminopeptidase M digest. [c] Peptic peptides were derived from the chymotryptic core (Table I). [d] Determined as *S*-(pyridylethyl)cysteine. [e] Isolated by chromatography on a column of Spinco AA-15 resin (Bradshaw et al., 1969). [f] Isolated by subfractionation of pooled fractions on a column of Bio-Rad AG1W-X2 (Wikler et al., 1970). [g] Fractionated on a column of Sephadex G-25 "Superfine" (Figure 6A). [h] Residue number(s) (see Figure 2).

the strategy is discussed beginning with a central region (surrounding fragment HA-2) and proceeding toward both ends (Figure 1).

Central Region of the Light Chain (Residues 40–122). Cleavage of the S-pyridylethylated light chain with hydroxylamine yielded a 30-residue fragment (HA-2) which contains both isoleucyl residues of the light chain, one in the second position of HA-2 (Gly-Ile-Gly-) and the other in the 25th position (Arg-Glu-Ile-). These residues served as useful markers to locate other peptides overlapping each end of HA-2.

The isoleucyl residue near the amino terminus was recognized in a fragment (SP-1) generated by cleavage of fraction B (Figure 3) by the protease from *S. aureus*. The digest was fractionated by gel filtration in 9% formic acid (Figure 5); a fraction, pooled as indicated in Figure 5, was adjusted to pH 8 and then acidified to pH 4, and the resulting precipitate was discarded. The supernatant was subjected to sequenator analysis, which yielded a single continuous sequence of 15 residues, ending with Lys-*Asn*-Gly-Ile-Gly- (Figure 1), but with a marked drop in yield of the italicized asparagine-63.

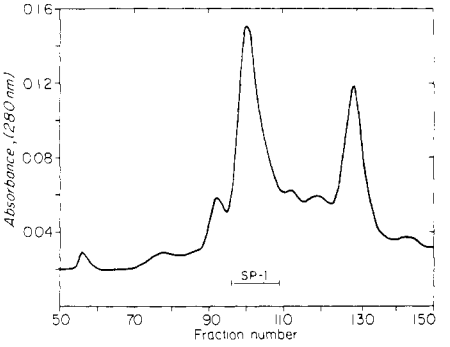

FIGURE 5: Fractionation of digest produced by cleavage of pooled fraction B (Figure 3) with *S. aureus* protease. The column (2.5 × 115 cm) of Sephadex G-50 was equilibrated and eluted with 9% formic acid. Fractions of 3.6 mL were collected at a flow rate of 27 mL/h. Fractions indicated by the bar (SP-1) were pooled and lyophilized.

This analysis extends the sequence toward the amino terminus from isoleucine-65 through the susceptible asparagine–glycine bond to glycine-52.

Of two tryptic peptides containing isoleucine (T-III and T-V-5), only the composition of T-III is consistent with the sequence of residues 63–79, although the aspartyl content is low (Table II). Therefore, T-III must contain isoleucine-65 and T-V-5 must include isoleucine-88.

The amino-terminal sequence of peptide T-V-5 proved to be Glu-Ile-Cys-Ser-Leu-, thus extending the sequence of HA-2 to leucine-91. The remainder of the peptide T-V-5 includes, by composition, eight residues which can be aligned with the amino-terminal octapeptide of HA-3. In fact, this is the only arginine-containing tryptic peptide which can fit this amino-terminal sequence. However, this alignment (Figure 1) leaves an unidentified Asx-Asx dipeptide between leucine-91 and glycine-94. This region was identified and the overlap confirmed by placement of an isoleucine-containing chymotryptic peptide (C-XI), the composition of which was consistent with that to be expected for the residues assigned to positions 84–99. Subdigestion of C-XI with thermolysin yielded a nonapeptide (C-XI-Th-5) with the sequence Leu-Asx-Asx-Gly-Gly-(Cys,Asx,Glx,Phe). Digestion of this peptide with aminopeptidase M (Table IV) released 1.63 aspartic acid residues and 0.74 asparagine residue per molecule. Since residue 97 is aspartic acid in HA-3 and since asparagine at residue 93 would be consistent with hydroxylamine cleavage, the residue at position 92 must be aspartic acid in the sequence Leu-Asp-Asn-Gly-. Since this sequence overlaps HA-2 and HA-3, a continuous sequence is provided from glycine-52 to lysine-122, except for an unidentified residue at position 105.

The sequence was further extended from glycine-52 toward the amino terminus by analysis of a product of cleavage of the single tryptophanyl bond in the light chain. Cleavage with BNPS-skatole yielded by gel filtration on Sephadex G-75 a single major component of ~12 000 daltons.

Sequenator analysis revealed, however, two new sequences in a ratio of 3:1 as follows: Ser-Lys-Tyr-Lys-Asp- and Phe-X-Ser-Lys-Tyr-. Since two of the three tyrosyl residues of the light chain have been placed elsewhere in different sequences (residues 68 and 115), these two new sequences must represent overlapping fragments which include the same tyrosyl residue. The single tryptophanyl residue can be placed preceding the major sequence on the basis of a peptic peptide P-2 (Table IV) with the composition (Trp,Ser,Lys,Tyr). Thus, the sequence Phe-Trp-Ser-Lys-Tyr-Lys-Asp- is established, leaving an anomalous cleavage point on the amino side of phenylalanine-40 (as discussed later). However, extended analysis of the major component provided a continuous sequence of 13 residues ending with Gly-His-Pro, which overlaps fragment SP-1. Thus, except for residue 105, a continuous sequence is established from phenylalanine-40 to lysine-122.

Carboxyl-Terminal Region of the Light Chain (Residues 105–140). Treatment of S-pyridylethylated light chain (46 nmol) with carboxypeptidase A in 4 M urea released 47 nmol of serine in 60 min. Similar digestion with a mixture of carboxypeptidases A and B released 60 nmol of serine and 43 nmol of arginine, indicating that the light chain terminated with arginylserine. A chymotryptic peptide (C-XII from the S-pyridylethylated light chain, Table IV) was found by subtractive Edman degradation to have the sequence Thr-Gln-Gly-Arg-Ser-. Since chymotrypsin normally cleaves peptide bonds at the carboxyl side of hydrophobic residues, it is very likely that this peptide has been derived from the carboxyl terminus of the whole light chain. This sequence overlaps that of the tryptic peptide T-V-2 (Phe-Thr-Gln-Gly-Arg, Table III), thus extending the carboxyl-terminal sequence to Phe-Thr-Gln-Gly-Arg-Ser.

At this point, all of the tryptic peptides can be considered. Of the eight arginyl residues, six were found in soluble peptides (Table III) and the other two are located in the amino-terminal core region (residues 15 and 28). Of the six soluble peptides, T-VII-1, T-V-5, T-V-4, and T-V-3 each contain arginine and together confirm the sequence between residues 80 and 108 (Figure 1); peptide T-V-3 identifies residue 105 as serine; peptide T-V-2 was placed next to the carboxyl-terminal serine, and only peptide T-VI (Ser-Cys-Val-Ser-Thr-Glu-Arg) remains to be placed. Similarly, of the seven expected lysyl peptides, two are present in the tryptic core but only one in the peptic core (Table I). Since the lysine in common is placed at residue 9 (and in T-II), the other should be the carboxyl terminus of the tryptic core. Four of the remaining five lysyl peptides are placed (see Figure 1) as follows: T-XI, T-VII-2, and T-III at residues 44–79 and T-VIII at residues 109–122. The only unplaced lysyl peptide is T-IX, which contains proline and cannot represent the carboxyl terminus of the tryptic core.

The two remaining tryptic peptides, T-IX and T-VI, can be placed between residues 123 and 134 on the basis of the partial sequence of HA-3. Following lysine-122, the sequence X-X-Val-X-X-Glu-X-Phe- has been observed (Figure 1). The placement of peptide T-VI at residues 123–129 is consistent with the location of valine-125 and glutamic acid-128, and the placement of peptide T-IX at residues 130–134 places the eighth and last phenylalanine at position 130 (the other seven have already been placed at residues 4, 31, 40, 76, 83, 99, and 135). Clearly these placements must be considered tentative pending confirmation of the overlap region around phenylalanine-40.

Amino-Terminal Region of the Light Chain (Residues 1–40). Combination of amino-terminal analyses of the light chain with the placements by Bucher et al. (1976) and Thogersen et al. (1978) of γ-carboxyglutamyl residues provides a continuous sequence of 40 residues (Figure 1). Repeated attempts to digest the reduced and alkylated light chain with either trypsin or chymotrypsin yielded large fragments ("core peptides") derived from the amino-terminal region which could be isolated by precipitation at pH 4. The tryptic and chymotryptic core peptides corresponded in composition to residues 1–43 and 1–44, respectively (Table I), and their amino-terminal sequences were identical with that of the intact light chain. The tryptic core peptide derived by extended (6 h) digestion was separated into two components (~6:1) by gel filtration on Sephadex G-50 in 0.1 M NH_4HCO_3. These corresponded in composition to residues 1–43 (65% yield, Table I) and 10–43 (10% yield, data not shown). Attempts to subdigest the core peptides with thermolysin or subtilisin failed to cleave the peptides further, but treatment of the chymotryptic core peptide with pepsin did yield upon gel filtration (Figure 6A) a peptic core peptide (residues 1–40, Table I), two small peptides (P-1 and P-2, Table IV), and free tryptophan. Digestion of the peptic core with carboxypeptidase A in 4 M urea released small amounts (0.1 equiv) of a single amino acid, phenylalanine, which must be residue 40. The amino acid composition of peptide P-2 was identical with that of peptide P-1 except that it contained tryptophan. The sum of the compositions of P-2 and of the peptic core was equal to the composition of the chymotryptic core peptide. Since the single tryptophan has already been tentatively placed in the sequence Phe-Trp-Ser-Lys-Tyr-44, these data place tryptophan at residue 41.

An attempt was made to confirm this sequence by modifying both carboxyl and ϵ-amino groups, followed by tryptic cleavage of the arginyl-28 peptide bond. The amino groups of the

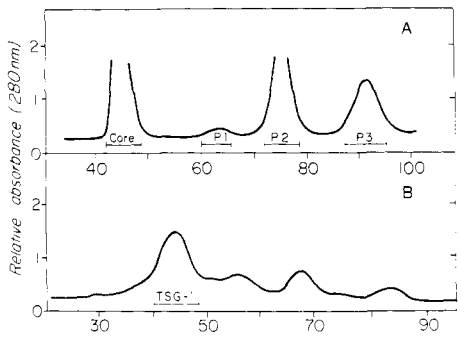

FIGURE 6: (A) Fractionation of peptic digest of chymotryptic core peptide. A column (1.5 × 88 cm) of Sephadex G-25 "Superfine" was equilibrated and eluted with 0.1 M ammonium bicarbonate, pH 7.9. Fractions of 1.2 mL were collected at a flow rate of 7.5 mL/h. Fractions were pooled as indicated and lyophilized. (B) Fractionation of the tryptic digest of succinylated, glycinamide-coupled light chain. A column (1.5 × 89 cm) of Sephadex G-50 "Superfine" was equilibrated and eluted with 9% formic acid. Fractions of 1.4 mL were collected at a flow rate of 14 mL/h. Fractions indicated by the bar (TSG-1) were pooled and lyophilized.

S-pyridylethylated light chain (106 mg) were succinylated by the method of Hermodson et al. (1973), and free carboxyl groups were blocked by reaction with glycinamide and water-soluble carbodiimide (Hoare & Koshland, 1967; Gibson & Anderson, 1972). Subsequent tryptic digestion followed by gel filtration (Figure 6B) yielded a large peptide (TSG-1) having the amino-terminal sequence X-Val-Phe-X-X-Ala-X-Gln-Thr-X-X-Phe-. Since the glycinamide derivatives of γ-carboxyglutamyl, aspartyl, and glutamyl residues would not be recognized by gas chromatography, this partial sequence is consistent with the sequence of residues 29–40. However, the analysis failed to identify residue 39 or to extend through tryptophan-41.

Residue 39 remains unidentified. The compositions of both the chymotryptic core and the peptic core (Table I) indicate glutamic acid or an acid-labile derivative thereof. Thogersen et al. (1978) identified this residue as γ-carboxyglutamic acid, thus completing the sequence.

Discussion

The light chain of bovine factor X_1 contains 140 amino acid residues in the sequence illustrated in Figure 7. This sequence is in accord with the amino acid composition (Table I), except for a small discrepancy in the aspartic acid content.

Of the 140 residues, 116 were identified by sequenator analyses either of the intact chain or of four key fragments, i.e., BNPS-1, SP-1, HA-2, and HA-3 (Figure 1). The first three of these fragments form an overlapping set comprising residues 42–93, and the sequenator analyses established the sequence of residues 42–88. The remaining strategy consisted of linking this overlapping set to the amino-terminal 40-residue sequence at one end and to the carboxyl-terminal fragment HA-3 at the other. The sequence of the carboxyl-terminal 20 residues was established by aligning 5 tryptic peptides. The sequences of certain regions are more convincingly established than those of others, and the weakest aspects of the structure should be reviewed.

Certain overlapping sequences are less firmly established than others. The alignment of T-V-5 and C-XI-Th-5 at leucine-91 is predicated upon the unique compositions of C-XI and T-V-5 and the amino-terminal sequence of HA-3. The alignment of T-VIII, T-VI, and T-IX (residues 109–134) is based upon both the partial sequence of HA-3 at residues 125, 128, and 130 and the independent placement of the carboxyl-terminal six residues. The reliability of these alignments

```
                5         10        15        20        25        30
  1  A N S F L Y Y V K Q G N L Y R Y C L Y Y A C S L Y Y A R Y V
 31  F Y D A Y Q T D Y F W S K Y K D G D Q C E G H P C L N Q G H
 61  C K N G I G D Y T C T C A E G F E G K N C E F S T R E I C S
 91  L D N G G C D Q F C R E E R S E V R C S C A H G Y V L G D D
121  S K S C V S T E R F P C G K F T Q G R S

                          COMPOSITION

          6 ALA A    6 GLN Q    7 LEU L   11 SER S
          8 ARG R    9 GLU E    7 LYS K    6 THR T
          6 ASN N   14 GLY G    0 MET M    1 TRP W
          9 ASP D    3 HIS H    8 PHE F    3 TYR Y
         15 CYS C    2 ILE I    2 PRO P    5 VAL V
         12 GLA Y

MOL. WT. = 16,193          NUMBER OF RESIDUES = 140
```

FIGURE 7: Amino acid sequence and composition of the light chain of bovine factor X_1 (Stuart factor).

is supported by the agreement between the composition of the whole light chain and that of the sum of the various tryptic peptides.

The placement of the amino-terminal 43 residues presented a more difficult analytical problem. Prior to the publication of the independent analyses of Bucher et al. (1976) and Thogersen et al. (1978), we had identified a continuous sequence of the first 34 residues (with tentative identifications of glutamic acid at residues 6, 7, 14, 16, 19, 20, 25, 26, 29, and 32) followed by X-X-Thr-Asp-X-Phe- (where X denotes an unidentified residue). The partial sequence of the glycinamide-modified peptide, TSG-1, placed glutamine at residue 36 but left residues 35 and 39 unidentified. The composition of the peptic core (Table I) indicated that residues 35 and 39 were glutamic acid or acid-labile derivatives thereof. Bucher et al. (1976) provided mass spectrometric identification of residue 35 as well as residues 6, 7, 14, 16, 19, 20, 25, 26, 29, and 32 as γ-carboxyglutamic acid. Since Katayama & Titani (1978) observed that BNPS-skatole cleaves at γ-carboxyglutamyl peptide bonds, the amino-terminal sequence and composition of fragment BNPS-1 suggest that residue 39 must be a γ-carboxyglutamyl residue. Thogersen et al. (1978) confirmed this fact, as well as the placements by Bucher et al. (1976) of the other 11 γ-carboxyglutamyl residues. In fact, the data of Thogersen et al. (1978) provide a continuous sequence from the amino terminus to the single tryptophan (residue 41) and fully confirm the validity of this portion of the sequence. This single tryptophan is also found in the peptic peptide P-2, which provided a key overlap to the fragment BNPS-1.

It should be noted that the large fragments produced for sequenator analysis were in several cases not completely pure. The fractionation procedures did yield polypeptides of sufficient purity (usually at least 95%) for reliable sequence determination but not sufficient to give integral amino acid compositions. Especially in the case of large polypeptides, small amounts of contaminating fragments can cause major deviations from the expected compositions. Consequently, the proof of sequence has not included compositions of most of these fragments. Three core peptides, derived by exhaustive digestion with trypsin, chymotrypsin, or pepsin, did provide confirmation of the composition of residues 1–44.

Although the disulfide pairing between the 15 cysteine residues has not yet been established, there must be 7 intrachain disulfide bonds and 1 interchain bond linking the light chain to cysteine-157 of the heavy chain. This follows from observations that there are no free sulfhydryl groups in factor X and that all cysteine residues except residue 157 of the heavy chain are joined by disulfide bonds within that chain (Titani

et al., 1975). Only 2 half-cystine residues are found in the amino-terminal 49 residues of the light chain, and they form an intrachain bridge (unpublished observation). The remaining 13 are clustered within the next 83 residues, suggesting an extensive network of cross-links which could contribute significantly to the stability of this domain of the light chain.

A striking feature of the light chain is the cluster of 12 γ-carboxyglutamyl residues within a 34-residue region near the amino terminus. This region (residues 6–39) also contains 2 aspartyl, 1 lysyl, and 2 arginyl residues, or a total of 29 charged groups among 34 residues. At physiological pH, one would expect a net charge of −23 and ample provisions for ionic interactions with neighboring proteins or cations. In contrast, the remaining 106 residues carry 17 negative and 16 positive charges in a seemingly random distribution. These facts imply that this amino-terminal portion of the light chain forms a highly charged segment of the surface of an otherwise tightly cross-linked domain. The γ-carboxyl groups occur in four coagulation factors and two other plasma proteins as products of a vitamin K dependent carboxylation of specific glutamyl residues (Sadowski et al., 1976; Jones et al., 1977). This process can be prevented by coumarin treatment, resulting in a loss of the ability of the protein to bind calcium or phospholipid and to undergo physiological activation.

The present analysis contributes to the growing evidence of the homologous relationships among the zymogens of serine proteases involved in the blood coagulation cascade. It has already been shown that the heavy chain of factor X_1 contains sequences homologous with corresponding regions of the active site of factors XII_a, XI_a, IX_a, and VII_a and thrombin (Titani et al., 1972, 1975; Enfield et al., 1974; Magnusson et al., 1975; Kisiel et al., 1976, 1977; Fujikawa et al., 1977; Kurachi & Davie, 1977). The present sequence data of the light chain of factor X_1 permit comparison with the amino-terminal portions of their functionally related zymogens. Fujikawa et al. (1974a) have demonstrated homology among the first 15 residues of prothrombin, factor IX, and factor X_1. This homology was subsequently extended to factor VII (Kisiel & Davie, 1975), protein C (Kisiel et al., 1976), and protein S (DiScipio et al., 1977; Stenflo & Jonsson, 1979). Since these six proteins comprise all the known vitamin K dependent proteins of blood plasma, it is perhaps not surprising that the regions of these molecules which contain the γ-carboxyglutamyl residues should also be homologous. This similarity in sequence extends further into portions of the sequence of the light chain of factor X_1 and corresponding regions of factor IX and protein C (Enfield et al., 1974; Fernlund et al., 1978). A detailed comparison of these relationships is reported in a separate communication, together with the complete sequence of factor IX (Katayama et al., 1979).

Acknowledgments

The authors are grateful for the technical assistance of Dr. J. C. Mercier, Brita Moody, Santosh Kumar, Douglas Marshall, and Richard Granberg. We also thank Dr. E. W. Davie for helpful discussions. Dr. G. Drapeau (University of Montreal) generously supplied the *S. aureus* protease.

References

Ambler, R. P. (1967) *Methods Enzymol. 11*, 436.
Balian, G., Click, E. M., & Bornstein, P. (1971) *Biochemistry 10*, 4470.
Barton, P. G. (1967) *Nature (London) 215*, 1508.
Barton, P. G., Jackson, C. M., & Hanahan, D. J. (1967) *Nature (London) 214*, 923.
Biggs, R., & MacFarlane, R. G. (1965) *Thromb. Diath. Haemorrh., Suppl. 17*, 23.
Bradshaw, R. A., Babin, D. R., Nomoto, M., Srinivasan, N. G., Ericsson, L. H., Walsh, K. A., & Neurath, H. (1969) *Biochemistry 8*, 3859.
Bucher, D., Nebelin, E., Thomsen, J., & Stenflo, J. (1976) *FEBS Lett. 68*, 293.
Davie, E. W., & Fujikawa, K. (1975) *Annu. Rev. Biochem. 44*, 799.
DiScipio, R. G., Hermodson, M. A., Yates, S. G., & Davie, E. W. (1977) *Biochemistry 16*, 698.
Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80.
Enfield, D. L., Ericsson, L. H., Fujikawa, K., Titani, K., Walsh, K. A., & Neurath, H. (1974) *FEBS Lett. 47*, 132.
Enfield, D. L., Ericsson, L. H., Walsh, K. A., Neurath, H., & Titani, K. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 16.
Fernlund, P., Stenflo, J., & Tufvesson, A. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5889.
Fujikawa, K., Legaz, M. E., & Davie, E. W. (1972a) *Biochemistry 11*, 4882.
Fujikawa, K., Legaz, M. E., & Davie, E. W. (1972b) *Biochemistry 11*, 4892.
Fujikawa, K., Coan, M. H., Enfield, D. L., Titani, K., Ericsson, L. H., & Davie, E. W. (1974a) *Proc. Natl. Acad. Sci. U.S.A. 71*, 427.
Fujikawa, K., Coan, M. H., Legaz, M. E., & Davie, E. W. (1974b) *Biochemistry 13*, 5290.
Fujikawa, K., Titani, K., & Davie, E. W. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3359.
Fujikawa, K., Walsh, K. A., & Davie, E. W. (1977) *Biochemistry 16*, 2270.
Gibson, D., & Anderson, P. J. (1972) *Biochem. Biophys. Res. Commun. 49*, 453.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., & Walsh, K. A. (1972) *Biochemistry 11*, 4493.
Hermodson, M. A., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1973) *Biochemistry 12*, 3146.
Hoare, D. G., & Koshland, D. E., Jr. (1967) *J. Biol. Chem. 242*, 2447.
Hougie, C., Censon, K. W. E., & Biggs, R. (1967) *Thromb. Diath. Haemorrh. 18*, 211.
Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506.
Hugli, T. E., & Moore, S. (1972) *J. Biol. Chem. 247*, 2828.
Jackson, C. W. (1972) *Biochemistry 11*, 4873.
Jackson, C. M., & Hanahan, D. J. (1968) *Biochemistry 7*, 4506.
Jackson, R. L., & Hirs, C. H. W. (1970) *J. Biol. Chem. 245*, 624.
Jesty, J., & Esnouf, M. P. (1973) *Biochem. J. 131*, 791.
Jesty, J., & Nemerson, Y. (1974) *J. Biol. Chem. 249*, 509.
Jesty, J., Spencer, A. K., & Nemerson, Y. (1974) *J. Biol. Chem. 249*, 5614.
Jobin, F., & Esnouf, M. P. (1967) *Biochem. J. 102*, 666.
Jones, J. P., Gardner, E. J., Cooper, T. G., & Olson, R. E. (1977) *J. Biol. Chem. 252*, 7738.
Katayama, K., & Titani, K. (1978) *FEBS Lett. 95*, 157.
Katayama, K., Ericsson, L. H., Enfield, D., Walsh, K. A., Neurath, H., Davie, E. W., & Titani, K. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4990.
Kisiel, W., & Davie, E. W. (1975) *Biochemistry 14*, 4928.
Kisiel, W., Ericsson, L. H., & Davie, E. W. (1976) *Biochemistry 15*, 4893.
Kisiel, W., Fujikawa, K., & Davie, E. W. (1977) *Biochemistry 16*, 4189.
Kurachi, K., & Davie, E. W. (1977) *Biochemistry 16*, 5831.
Lundblad, R. L., & Davie, E. W. (1964) *Biochemistry 3*, 1720.

Lundblad, R. L., & Davie, E. W. (1965) *Biochemistry 4*, 113.
Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Morris, J. R., & Dell, A. (1974) *FEBS Lett. 44*, 189.
Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) *Cold Spring Harbor Conf. Cell Proliferation 2*, 128.
Moore, S., & Stein, W. H. (1963) *Methods Enzymol 6*, 819.
Nelsestuen, G. L., Zytokovicz, T. H., & Howard, J. B. (1974) *J. Biol. Chem. 249*, 6347.
Nemerson, Y., & Pitlick, F. A. (1970) *Biochemistry 9*, 5100.
Omenn, G. S., Fontana, A., & Anfinsen, C. B. (1970) *J. Biol. Chem. 245*, 1895.
Osterud, B., & Rapaport, S. I. (1970) *Biochemistry 9*, 1854.
Osterud, B., Berre, A., Otnaess, A. B., Borklid, E., & Pyrdz, H. (1972) *Biochemistry 11*, 2853.
Papahadjopoulos, D., & Hanahan, D. J. (1964) *Biochim. Biophys. Acta 90*, 436.
Peterson, J. D., Nehrlich, S., Oyer, P. E., & Steiner, D. F. (1972) *J. Biol. Chem. 247*, 4866.
Radcliffe, R. D., & Barton, P. G. (1973) *J. Biol. Chem. 248*, 6788.
Sadowski, J. A., Esman, C. T., & Suttie, J. W. (1976) *J. Biol. Chem. 251*, 2770.
Shearer, W. T., Bradshaw, R. A., Gurd, F. R. N., & Peters, T. (1967) *J. Biol. Chem. 242*, 5451.
Stenflo, J., & Jonsson, M. (1979) *FEBS Lett. 101*, 377.
Stenflo, J., Ferlund, P., Egar, W., & Paepstorff, P. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2730.
Suttie, J. W., & Jackson, C. M. (1977) *Physiol. Rev. 57*, 1.
Thogersen, J. C., Petersen, T. E., Sottrup-Jensen, L., Magnusson, S., & Morris, H. R. (1978) *Biochem. J. 175*, 613.
Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H., & Davie, E. W. (1972) *Biochemistry 11*, 4899.
Titani, K., Fujikawa, K., Enfield, D. L., Ericsson, L. H., Walsh, K. A., & Neurath, H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3082.
Weber, K., & Osborn, M. (1969) *J. Biol. Chem. 244*, 4406.
Wikler, M., Titani, K., & Putnam, F. W. (1970) *J. Biol. Chem. 245*, 2158.
Williams, W. J., & Norris, D. G. (1966) *J. Biol. Chem. 241*, 1847.

Primary Structure of Chicken Liver Dihydrofolate Reductase†

A. Ashok Kumar, Dale T. Blankenship, Bernard T. Kaufman, and James H. Freisheim*

ABSTRACT: The complete covalent structure of dihydrofolate reductase from chicken liver is described. The S-carboxymethylated protein was subjected to cleavage by cyanogen bromide which produced five fragments. Fragment CB2 contained an internal homoserine residue which was not cleaved by cyanogen bromide. Sequences and ordering of the cyanogen bromide fragments were established by means of automated sequencer analyses of the fragments and from smaller peptides generated by proteolysis with trypsin and staphylococcal protease. The covalent structure of the single polypeptide chain comprises 189 residues of molecular weight 21 651. The chicken liver enzyme is homologous to that from L1210 cells and shows regions of homology to dihydrofolate reductases from *Streptococcus faecium*, *Escherichia coli*, and *Lactobacillus casei*. These homologous regions in the chicken liver enzyme are primarily related to conserved amino acid residues implicated in the binding of NADPH and methotrexate by bacterial dihydrofolate reductases.

Dihydrofolate reductase (EC 1.5.1.3) catalyzes the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate. The latter serves as a coenzyme for a number of one-carbon transfer reactions in purine and pyrimidine biosynthesis, including that of thymidylate (Blakely, 1969; Huennekens et al., 1971). The reductase appears to be the major intracellular receptor for the action of 4-amino analogues of folic acid, such as amethopterin (4-amino-10-methyl-4-deoxyfolate, methotrexate). Inhibition of the enzyme by methotrexate depletes the tetrahydrofolate pool, resulting in a decreased synthesis of thymidylate and, in turn, an inhibition of DNA synthesis. Thus, methotrexate has been employed extensively in the chemotherapeutic treatment of leukemias, lymphomas, psoriasis, and other clinical disorders (Bertino & Johns, 1972). In addition, a differential sensitivity to certain drugs, such as trimethoprim, by mammalian and bacterial reductases has lead to the development of a class of compounds with potent antibacterial activity (Burchall & Hitchings, 1965). The underlying molecular basis for these differential inhibitory effects must reside in differences in the three-dimensional architecture of the active centers of bacterial vs. animal dihydrofolate reductases. Thus, a knowledge of the complete primary and tertiary structure of dihydrofolate reductases from both bacterial and animal sources in the presence and absence of inhibitors, coenzymes, and substrates should aid in our understanding of these differential inhibitory effects. In addition, such knowledge may suggest approaches for the rational design of specific chemotherapeutic agents.

Previous results from this laboratory have lead to the elucidation of the complete amino acid sequence of dihydrofolate reductase from *Lactobacillus casei* (Freisheim et al., 1977, 1978; Bitar et al., 1977). This sequence information has aided in the resolution of the X-ray structure of the *L. casei* reductase–NADPH–methotrexate ternary complex at 2.5 Å as recently described by Matthews, Kraut, and co-workers (Matthews et al., 1978, 1979). Sequence homology alignments suggested by this laboratory (Freisheim et al., 1978; Bitar et al., 1977) have aided in implicating certain highly conserved

† From the Department of Biological Chemistry, College of Medicine, University of Cincinnati, Cincinnati, Ohio 45267 (A.A.K., D.T.B., and J.H.F.), and the National Institute of Arthritis, Metabolism, and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014 (B.T.K.). *Received July 16, 1979.* This research was supported by Grant CA 11666, National Cancer Institute, Department of Health, Education, and Welfare, and Grant CH80C from the American Cancer Society. A preliminary report of this work was presented at the 6th International Symposium on the Chemistry and Biology of Pteridines, La Jolla, CA (Freisheim et al., 1979).

0006-2960/80/0419-0667$01.00/0