Goebel, W. F., Olitsky, P. K., & Saenz, A. C. (1946) *J. Exp. Med. 87*, 445.
Habeeb, A. F. S. A. (1966) *Anal. Biochem. 14*, 328.
Haynes, R., & Feeney, R. E. (1968) *Biochemistry 7*, 2879.
Haynes, R., Osuga, D. T., & Feeney, R. E. (1967) *Biochemistry 6*, 541.
Hunter, M. J., & Ludwig, M. L. (1972) *Methods Enzymol. 25*, 585.
Jentoft, N., & Dearborn, D. G. (1979) *J. Biol. Chem. 254*, 4359.
Kalnitsky, G., Hummel, J. P., & Dierks, C. (1959) *J. Biol. Chem. 234*, 1512.
Kato, I., Kohr, W. J., & Laskowski, M., Jr. (1978). *FEBS Symp. 47*, 197–206.
Krysteva, M. A., & Dobrev, I. D. (1977) *Eur. J. Biochem. 74*, 501.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265.
Ludwig, M. L., & Hunter, M. J. (1967) *Methods Enzymol. 11*, 595.
Maekawa, K., & Kushibe, M. (1954) *Bull. Chem. Soc. Jpn. 27*, 277; *Chem. Abstr. 49*, 9698*e*.
Maekawa, K., & Kushibe, M. (1955) *Bull. Agric. Chem. Soc. Jpn. 19*, 28; *Chem. Abstr. 50*, 14022*f*.
Means, G. E. (1977) *Methods Enzymol. 47*, 469.
Means, G. E., & Feeney, R. E. (1968) *Biochemistry 7*, 2192.
Means, G. E., & Feeney, R. E. (1971) *Chemical Modification of Proteins*, Holden-Day, San Francisco.
Means, G. E., Congdon, W. I., & Bender, M. L. (1972) *Biochemistry 11*, 3564.
Shugar, D. (1952) *Biochim. Biophys. Acta 8*, 302.
Sklarz, B. (1967) *Q. Rev., Chem. Soc. 21*, 3.
Skoog, D. A., & West, D. M. (1976) *Fundamentals of Analytical Chemistry*, 3rd ed., Holt, Rinehart and Winston, New York.
Slobodian, E., Mechanic, G., & Levy, M. (1962) *Science 135*, 441.
Smith, F., & Unrau, A. M. (1959) *Chem. Ind. (London)*, 881.
Wang, D., Wilson, G., & Moore, S. (1976) *Biochemistry 15*, 660.

Amino Acid Sequence Studies on the α Chain of Human Fibrinogen. Complete Sequence of the Largest Cyanogen Bromide Fragment†

D. D. Strong, K. W. K. Watt,‡ B. A. Cottrell, and R. F. Doolittle*

ABSTRACT: The largest fragment produced by complete cyanogen bromide digestion of the α chain of human fibrinogen contains 236 residues and has a calculated molecular weight of 23 949. The complete amino acid sequence of the fragment was determined by the isolation of peptides generated by plasmin, trypsin (including digestion of citraconylated material), staphylococcal protease, and chymotrypsin. In addition, some key subfragmentation was achieved by selective chemical cleavage at tryptophan residues. The fragment has an unusual amino acid composition, more than half of its residues being glycine, serine, threonine, and proline. There are very few nonpolar residues, although 7 of the α-chain's 10 tryptophans occur in this fragment. The fragment contains 2 cysteine residues located 30 residues apart which are connected by an intrachain disulfide bond in the native molecule. The tryptophans occur with a definite periodicity that highlights a series of 13-residue homology repeats. The fragment also contains the two principal α-chain cross-linking sites.

Vertebrate fibrinogen molecules are composed of three pairs of nonidentical polypeptide chains ($\alpha_2\beta_2\gamma_2$). Invariably, the largest of these are the α chains, which range from 60 000 to 80 000 in molecular weight, depending on the particular species (Doolittle, 1973). In the case of human fibrinogen, the α chains have molecular weights of about 65 000 (McKee et al., 1966, 1970). The chains contain 10 methionines and give rise to 11 unique fragments upon digestion with cyanogen bromide (Doolittle et al., 1977a). Of these, complete sequences have been reported for the 10 smallest (Blombäck et al., 1972; Takagi & Doolittle, 1975; Cottrell & Doolittle, 1976, 1978; Doolittle et al., 1977a,b; Lottspeich & Henschen, 1978a,b; Doolittle et al., 1979a). Only the largest fragment has remained undetermined. In this article we report the complete amino acid sequence of that fragment, designated HαCNI. It contains 236 residues and has a calculated molecular weight of 23 949. The sequence determination was complicated by a number of unusual features of this part of the α chain. It has a very distinctive amino acid composition, more than half of the residues being serine, glycine, threonine, and proline. As a result, many of the enzymatically derived peptides have similar properties and compositions, and the purification and proof of uniqueness were a formidable challenge. Moreover, the task was further complicated by a number of anomalous cleavages, including the partial tryptic cleavages of two different arginyl–proline bonds, chymotryptic action adjacent to certain threonyl residues, and staphylococcal protease cleavage at a particular seryl residue.

Experimental Section

Materials and Methods. Almost all the materials and methods used in this study have been described in detail in

† From the Department of Chemistry, University of California, San Diego, La Jolla, California 92093. *Received July 18, 1979.* Supported by U.S. Public Health Service Grants HE-12,759, HE-18,576, and GM-17,702. A preliminary account of this work was presented at the 11th International Congress of Biochemistry, Toronto, July 12, 1979 (Strong et al., 1979).

‡ Present address: Department of Medicine, Harvard Medical School, Boston, MA.

0006-2960/79/0418-5399$01.00/0

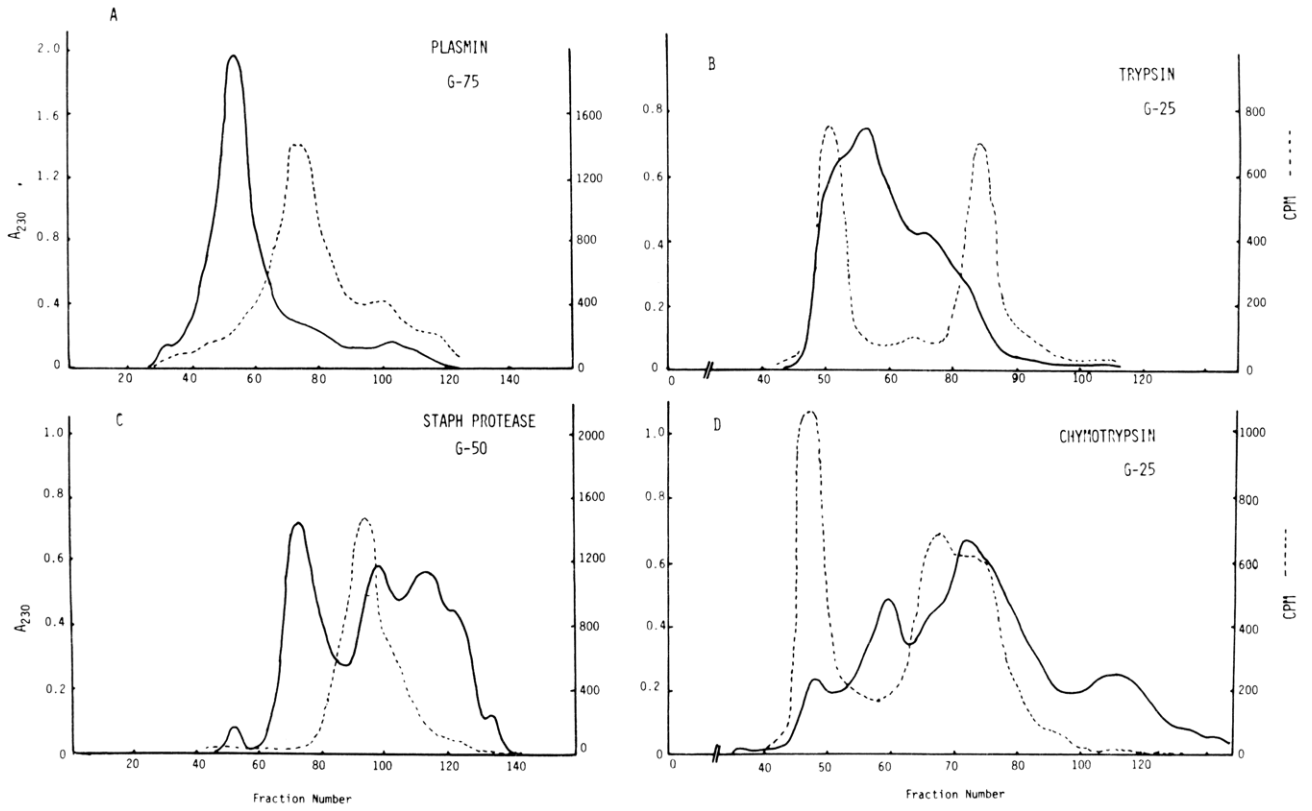

FIGURE 1: Gel filtration of (A) plasmic, (B) tryptic, (C) staphylococcal protease, and (D) chymotryptic digests of HαCNI. In each case approximately 1 μmol (25 mg) of material was digested and then applied directly to the column: 2.5 × 90, 2.5 × 145, 2.5 × 140, and 2.5 × 145 cm for plasmic, tryptic, staphylococcal protease, and chymotryptic digests, respectively. In all cases the columns were equilibrated and eluted with 0.1 M ammonium bicarbonate. The flow rate was the same in all studies (60 mL/h), and the fraction sizes were 5.0 mL except in the case of the plasmin digest, in which case the flow rate was 110 mL/min and the fraction size was 4.0 mL.

previous publications from this laboratory, including the preparation of fibrinogen (Doolittle et al., 1967), the isolation of the individual chains, the cyanogen bromide recipe employed, and the isolation and purification of the largest CNBr[1] fragment (CNIB) (Doolittle et al., 1977a), as well as enzyme digestion conditions, gel filtration procedures, paper electrophoresis, amino acid analyses, and end-group, stepwise degradation, and solid-phase sequencing procedures (Doolittle et al., 1977b; Takagi & Doolittle, 1975; Watt et al., 1979). In addition to the previously reported methods, we also utilized a chemical cleavage procedure for selective cleavage at tryptophanyl bonds (Ozols & Gerard, 1977). Molecular weight determinations were performed on appropriate sodium dodecyl sulfate–polyacrylamide slab gels.

Plasmin Digestion of HαCNI. Limited digestion of HαCNI with plasmin (Kabi, Stockholm) yields two principal subfragments that account for the entire original CNBr fragment. Routinely, 25 mg (ca. 1 μmol) of HαCNI was suspended in 1.2 mL of 0.1 M ammonium bicarbonate, and 100 μL of plasmin (in 50% glycerol) was added. Digestion was carried out for 2 h at 37 °C. The two subfragments were isolated by gel filtration on Sephadex G-75 (Figure 1A). As noted in a previous report (Doolittle et al., 1977a), the larger of the two subfragments contains all the strongly UV absorbing material, and the smaller contains all the radioactivity associated with the carboxymethylation of cysteines ([^{14}C]iodoacetic acid was used as an alkylating agent after reduction of all disulfide bonds in the native molecule). The homogeneity of PLI, and also HαCNI, was demonstrated on $NaDodSO_4$–polyacrylamide gel electrophoresis (Figure 2).

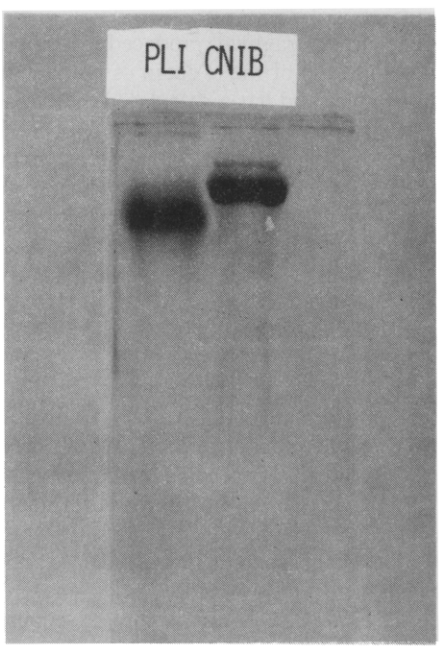

FIGURE 2: $NaDodSO_4$–polyacrylamide gel electrophoresis (12.5% acrylamide) of HαCNI and PLI. Running buffer = 0.1 M H_3PO_4–Tris, pH 7.0, 0.1% $NaDodSO_4$, and 0.19 M glycine. Running gel = 0.1 M H_3PO_4–Tris, pH 7.0, 12.5% acrylamide (1:10 bis(acrylamide)–acrylamide), 8 M urea, and 0.1% $NaDodSO_4$. Stacking gel (1 cm) = 5% acrylamide (1:20 bis(acrylamide)–acrylamide), 2 M urea, and 0.1% $NaDodSO_4$.

Sequence studies subsequently revealed that the larger subfragment (PLI) contains 184 residues and has a calculated molecular weight of 18 509. Its relatively high absorbance at

[1] Abbreviations used: CNBr, cyanogen bromide; Dns, 5-(dimethylamino)naphthalene-1-sulfonyl; PhNCS, phenyl isothiocyanate; TATG, (thioacetyl)thioglycolic acid; $NaDodSO_4$, sodium dodecyl sulfate.

Table I: Amino Acid Compositions of the Large Plasmic and Staphylococcal Protease Fragments from HαCNI

	residues/molecule[a]			
	CNIB	PLI	PLII	SPI
CM-cysteine[b]	1.8 (2)		1.0 (2)	
aspartic acid[c]	20.2 (19)	15.6 (16)	3.6 (3)	8.2 (8)
threonine	27.7 (29)	18.0 (18)	10.1 (11)	9.3 (9)
serine	41.0 (46)	38.7 (40)	6.8 (6)	19.7 (23)
glutamic acid[c]	21.0 (20)	14.8 (15)	5.7 (5)	3.0 (3)
proline	16.7 (18)	16.3 (16)	2.4 (2)	10.5 (10)
glycine	42.7 (43)	38.6 (38)	5.5 (5)	22.6 (21)
alanine	7.1 (5)	4.9 (4)	0.6 (1)	2.7 (2)
valine	9.5 (10)	4.4 (4)	5.2 (6)	
isoleucine	2.4 (2)	1.1 (1)	0.9 (1)	
leucine	4.0 (3)	2.7 (3)		
tyrosine	2.3 (2)	1.9 (2)		
phenylalanine	2.8 (2)	2.0 (2)		
histidine	4.4 (4)	2.9 (3)	1.2 (1)	
lysine	10.4 (10)	4.5 (4)	5.5 (6)	1.2 (1)
arginine	13.8 (13)	10.7 (11)	2.2 (2)	2.8 (3)
tryptophan[d]	6.0 (7)	7.5 (7)		4.1 (4)
homoserine[e]	0.9 (1)		0.6 (1)	
total:	(236)	(184)	(52)	(84)
residue no.:	1–236	1–184	185–236	24–107
amino terminal:	Glu	Glu	Thr	Thr/Ser

[a] Values are the average of 24-, 48-, and 72-h hydrolyses. Appropriate corrections have been used for partial destruction of serine and threonine. Only the 48- and 72-h values were used for valine and isoleucine. The values in parentheses are the numbers of residues determined by sequencing. [b] CM-cysteine = (carboxymethyl)cysteine. The number of residues was also determined by the amount of radioactivity present in the fragment as a result of previous alkylation with [^{14}C] iodoacetic acid. [c] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [d] Tryptophan was determined spectrophotometrically and/or by a micromodification of the Spies & Chambers (1949) method. [e] Homoserine values include homoserine lactone.

λ = 280 nm is due to seven tryptophans and two tyrosines. The smaller fragment is composed of 52 residues, none of which is tryptophan or tyrosine. The two cysteines of HαCNI are both found in this subfragment, however. The amino acid compositions of HαCNI, PLI, and PLII are listed in Table I. PLII contains homoserine and thus corresponds to the carboxy-terminal fragment. The amino-terminal residues of PLI and PLII, as determined by the DnsCl method (Gray, 1972), are glutamic acid and threonine, respectively.

Trypsin Digestion of HαCNI. Trypsin digestions were routinely conducted on HαCNI dispersed at 10 mg/mL in 0.1M ammonium bicarbonate containing 0.1 mg/mL TPCK-treated trypsin (Worthington). After 4 h at 37 °C, the digestion was allowed to continue for an additional 16 h at room temperature.

The initial purification of peptides was accomplished by gel filtration on Sephadex G-25 (Figure 1B) or Sephadex G-50 or combinations of the two. Most of the lysine-containing peptides were small and emerged in the retarded pool on Sephadex G-25. In contrast, with the exception of free arginine, arginine-containing peptides were larger and were found in the void volume of Sephadex G-25 preparations. The larger molecular weight pools also contained all the tryptophan-containing peptides. Radioactivity associated with the carboxymethylated cysteines was split into two peaks. It was subsequently determined that the smaller peptide was composed of four residues and was retained on Sephadex G-25; the other contained 15 residues and came near the front on Sephadex G-25 (Figure 1B). This latter peptide was also found to contain homoserine and corresponded to the carboxy-terminal peptide of HαCNI.

Final purification of tryptic peptides was achieved by paper electrophoresis at pH 6.5 and/or 2.0. Descending paper chromatography (1-butanol–acetic acid–water, 4:1:5) was employed when necessary. Peptides were sequenced by both a thioacetylation stepwise degradation procedure (Mross & Doolittle, 1977; Doolittle et al., 1977) and the DnsCl–PhNCS procedure (Gray, 1972). Also, tryptic peptides were fragmented further as necessary, by using either chymotrypsin, staphylococcal protease, or, on occasion, Nagarse and thermolysin.

A total of 21 peptides and (two) free arginines were purified from tryptic digests of HαCNI. Of these, four were anomalous in that they reflected partial cleavage at two different arginyl–proline bonds. As a result, two "expected" tryptic peptides (TR4 and TR5) were also found as their constituent subpeptides (TR4a and TR4b and TR5a and TR5b, respectively). Were it not for the partial cleavages, which amounted to about

Table II: Amino Acid Compositions of Seven Large Tryptic Peptides from HαCNI (Residues 1–166)

	residues/molecule[a]						
	TR1	TR2	TR3	TR4	TR5	TR4b	TR5b
aspartic acid[b]	1.2 (1)		2.9 (3)	5.3 (5)	6.0 (6)	1.2 (1)	3.2 (4)
threonine	1.1 (1)	2.8 (3)	1.3 (1)	7.2 (7)	4.7 (4)	1.8 (2)	2.2 (2)
serine		3.5 (4)	6.4 (7)	13.2 (15)	12.2 (13)	3.1 (3)	3.0 (2)
glutamic acid[b]	3.1 (3)	2.4 (2)		2.2 (2)	4.7 (5)	1.1 (1)	2.1 (2)
proline	1.0 (1)	1.3 (1)	1.9 (2)	6.5 (7)	5.4 (5)	1.8 (2)	2.5 (3)
glycine	2.3 (2)	3.7 (4)	5.6 (5)	16.7 (16)	11.8 (10)	3.0 (3)	3.7 (3)
alanine			0.9 (1)	2.0 (1)	2.1 (2)		
valine					2.3 (3)		1.6 (2)
isoleucine	0.8 (1)						
leucine	0.8 (1)						
tyrosine		0.9 (1)					
phenylalanine					1.7 (2)		0.7 (1)
histidine					1.4 (2)		
lysine				1.2 (1)			
arginine	1.9 (2)	1.3 (1)	1.0 (1)	1.8 (2)	2.7 (3)	1.0 (1)	1.1 (1)
tryptophan[c]			+ (1)	+ (3)	+ (3)	+ (1)	+ (1)
total:	(12)	(16)	(21)	(59)	(58)	(14)	(21)
residue no.:	1–12	13–28	29–49	50–108	109–166	95–108	146–166
amino terminal:	Glu	Gly	Asn	Asn	Gly	Pro[d]	Pro[d]

[a] Values are the average of three or more 24-h hydrolyses. Appropriate corrections have been made for partial destruction of serine and threonine. [b] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [c] Tryptophan was detected qualitatively after acid hydrolysis as a peak emerging after ammonia on a single-column amino acid analysis. [d] Note that the two proline-ending peptides represent partial splits of two other tryptic peptides.

Table III: Amino Acid Compositions of Small Tryptic Peptides from HαCNI (Residues 168–236)

	residues/molecule[a]											
	TR7	TR8	TR9	TR10	TR11	TR12	TR13	TR15	TR16	TR17	TR18	TR19
CM-cysteine								0.5 (1)				0.6 (1)
aspartic acid			1.0 (1)							1.3 (1)		2.1 (2)
threonine	1.1 (1)	1.3 (1)			0.9 (1)		4.7 (5)		1.7 (2)	0.7 (1)	0.9 (1)	1.2 (1)
serine		1.2 (1)					1.9 (2)	2.2 (2)				2.4 (2)
glutamic acid	2.4 (2)			1.2 (1)		1.0 (1)					1.1 (1)	2.7 (3)
proline										1.1 (1)		1.1 (1)
glycine			1.0 (1)		1.0 (1)		1.4 (1)			2.5 (2)		1.8 (1)
alanine												1.1 (1)
valine		0.7 (1)					1.1 (1)		1.2 (1)	0.3[b] (1)	0.9 (1)	1.4 (2)[c]
isoleucine										0.7[b] (1)		
leucine		0.6 (1)		0.9 (1)								
tyrosine	0.7 (1)											
phenylalanine												
histidine	0.9 (1)									1.2 (1)		
lysine	1.0 (1)	1.2 (1)	1.1 (1)		1.1 (1)	1.0 (1)		1.2 (1)	1.1 (1)	1.2 (1)	1.0 (1)	
arginine				0.9 (1)			0.9 (1)					
homoserine												0.5 (1)
total:	(6)	(5)	(3)	(3)	(3)	(2)	(10)	(4)	(4)	(9)	(4)	(15)
residue no.:	168–173	174–178	179–181	182–184	185–187	188–189	190–199	201–204	205–208	209–217	218–221	222–236
amino terminal:	Glu	Leu	Gly	Glu	Thr	Glu	Val	Ser	Thr	Thr	Glu	Glu

[a] See Tables I and II for explanation of values. [b] Val-Ile bond. [c] Val-Val bond.

50% when digestion was carried out as described, there would be 17 tryptic peptides and two (single) arginines (Tables II and III).

Tryptic peptides were also purified from digests of citraconylated (Atassi & Habeeb, 1972) HαCNI. The first five tryptic peptides from the amino terminus (TR1–TR5), amounting to 166 residues, were identical with the corresponding peptides isolated from underivatized material (Table II). The single lysine that occurs in the amino-terminal half of the fragment is adjacent to a proline residue and is not cleaved under either set of conditions. After removal of the citraconyl groups (pH 2.0; 37 °C; 4 h) and redigestion with trypsin, the remainder of the peptides found in digests of underivatized material were identified. Tryptic digestions were also carried out on PLI and PLII, as well as on the larger peptides isolated from staphylococcal protease and chymotryptic digestions.

Staphylococcal Protease and Chymotrypsin Digestion of HαCNI. Digestions of HαCNI with staphylococcal protease, on the one hand, and chymotrypsin, on the other hand, were undertaken under conditions similar to those described above for trypsin. The weight ratios (enzyme/substrate) were usually 1:50 for chymotrypsin and 1:40 for staphylococcal protease. Preparations were gel-filtered on either Sephadex G-50 (Figure 1C) or Sephadex G-25 (Figure 1D) or combinations of both. Further purification procedures followed the lines described for the purification of tryptic peptides. In the case of staphylococcal protease a large key fragment, designated SPI, was isolated in virtually pure form as a large molecular weight peak on Sephadex G-50 (Figure 1C); its amino acid composition is presented in Table I. Its amino terminal was invariably a mixture of threonine and serine, an observation which was subsequently found to reflect incomplete cleavage of the glutaminyl–serine bond at position 25 to 26.

Chymotrypsin digestion provided a large radioactive homoserine-containing fragment corresponding to the carboxy terminus. The material separated into two peptides upon electrophoresis at pH 6.5, the amino acid compositions of which were identical except for the presence of one additional residue each of threonine and lysine in one. The peptides contained 29 and 27 residues, the amino terminals being lysine and valine, respectively. Stepwise degradations revealed that the two peptides were identical except for the presence of an additional dipeptide (lysyl-threonine) present at the amino terminus of the 29-residue peptide. In both cases the chymotryptic splits were anomalous, one proving to be the cleavage of a threonyl–lysine bond (position 207 to 208) and the other a partial cleavage of a threonyl–valine bond (position 209 to 210).

The amino acid compositions and amino-terminal residues of 15 staphylococcal protease peptides and 15 chymotryptic peptides derived from HαCNI are provided in the supplementary material (see paragraph at end of paper regarding supplementary material). In addition, staphylococcal protease and chymotrypsin digestions were also undertaken on PLI, PLII, and several of the large tryptic peptides from HαCNI.

Chemical Cleavage at Tryptophanyl Bonds. HαCNI contains seven tryptophan residues (Table I). Accordingly, 31 mg of HαCNI was treated with 2.0 mL of a 1:1 mixture of heptafluorobutyric acid and formic acid containing 1.4 g of CNBr, according to the procedure described by Ozols & Gerard (1977) for the cleavage of tryptophanyl bonds. The digest was gel-filtered on Sephadex G-50 (equilibrated with 1% acetic acid). An 85-residue fragment with amino-terminal glycine was isolated from the first peak to emerge. For reasons that became apparent later, complete resolution of the remaining fragments was not possible because so many of them were the same size and had such similar amino acid compositions.

The 85-residue fragment was attached to aminoethylaminopropyl glass beads by using phenylene diisothiocyanate (Laursen et al., 1972) and subjected to 39 successful cycles of thioacetylation on an automatic sequencer (Doolittle et al., 1977). The results were especially significant in that the peptide overlaps the junction of PLI and PLII (Figure 3).

Sequence Summary. A composite of the key data used to establish the complete sequence of HαCNI is depicted in Figure 3. Some of the significant features in establishing the 236-residue sequence include the following: (a) the overlap between PLI and PLII was established by a long sequencer run on a fragment obtained by chemical cleavage of a tryptophanyl bond, (b) the distribution of tryptic peptides in PLI and PLII is consistent with the results of tryptic digestion of citraconylated HαCNI, and (c) overlaps for all tryptic peptides

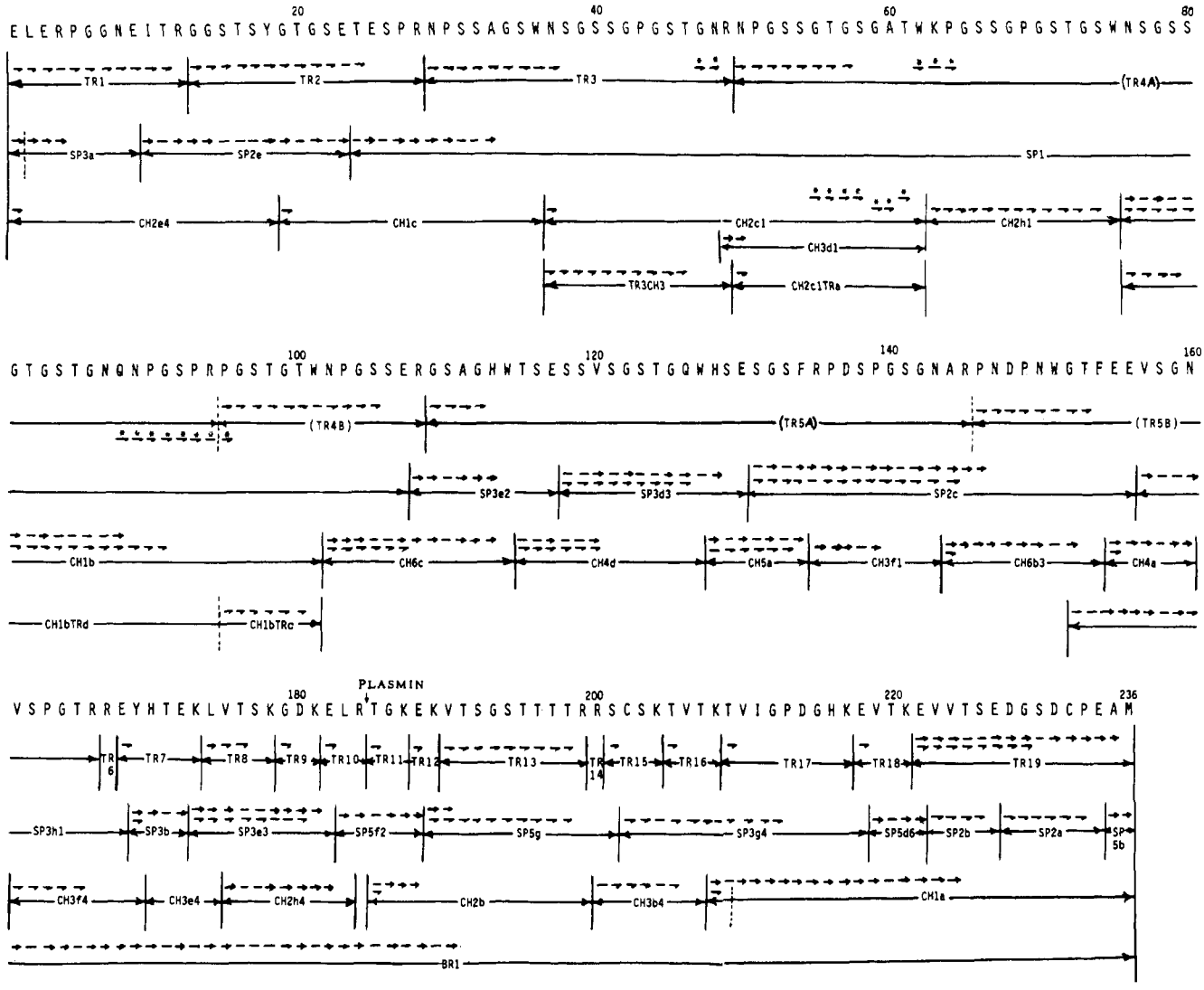

FIGURE 3: Composite of key data used in deducing the complete amino acid sequence of HαCNI. TR, trypsin; SP, staphylococcal protease; CH, chymotrypsin; PL, plasmin. (→) Successful identification after stepwise degradation by the TATG procedure; (⇀) identification by the Dns–PhNCS method; (—) stepwise degradation without positive identification; (*) identification on subpeptides derived by Nagarse or thermolysin digestion of designated peptide. Not shown are the results of stepwise degradations on HαCNI per se (10 successful steps) and on PLII (20 steps including identification of radioactive CM-Cys at position 202, step 18 of sequencer run).

were established with either staphylococcal protease or chymotryptic peptides or, in most cases, both. With rare exception, the amino acid composition of peptides corresponded perfectly with the determined sequence (Table III). In the case of peptide TR4 and its subpeptides covering the stretch 66–74, however, an extra alanine residue was consistently found, whereas no alanine was found by stepwise degradation. It is possible that one of the four serines in this stretch is, in whole or part, actually alanine.

Discussion

The largest CNBr fragment isolated from the α chain of human fibrinogen, designated HαCNI, occurs eighth from the amino terminus (Doolittle et al., 1977a, 1979a). The fragment is composed of 236 residues (Figure 4) and has a very unusual amino acid composition in that more than half of the residues are serine (46), glycine (43), threonine (29), and proline (18). There are relatively few nonpolar amino acids, although the fragment does contain seven tryptophan residues. These latter occur with a regular periodicity that suggests an evolutionary history of elongation by tandem duplications of a small portion of the α-chain gene. The homology repeat can be cast as an extended series of 13-residue repeats (Figure 5) or as a smaller

```
  1 Glu Leu Glu Arg Pro Gly Gly Asn Glu Ile Thr Arg Gly Gly Ser Thr Ser Tyr Gly Thr
 21 Gly Ser Glu Thr Glu Ser Pro Arg Asn Pro Ser Ser Ala Gly Ser Trp Asn Ser Gly Ser
 41 Ser Gly Pro Gly Ser Thr Gly Asn Arg Asn Pro Gly Ser Ser Gly Thr Gly Ser Gly Ala
 61 Thr Trp Lys Pro Gly Ser Ser Gly Pro Gly Ser Thr Gly Ser Trp Asn Ser Gly Ser Ser
 81 Gly Thr Gly Ser Thr Gly Asn Gln Asn Pro Gly Ser Pro Arg Pro Gly Ser Thr Gly Thr
101 Trp Asn Pro Gly Ser Ser Glu Arg Gly Ser Ala Gly His Trp Thr Ser Glu Ser Ser Val
121 Ser Gly Ser Thr Gly Gln Trp His Ser Glu Ser Gly Ser Phe Arg Pro Asp Ser Pro Gly
141 Ser Gly Asn Ala Arg Pro Asn Asp Pro Asn Trp Gly Thr Phe Glu Glu Val Ser Gly Asn
161 Val Ser Pro Gly Thr Arg Arg Glu Tyr His Thr Glu Lys Leu Val Thr Ser Lys Gly Asp
181 Lys Glu Leu Arg Thr Gly Lys Glu Lys Val Thr Ser Gly Ser Thr Thr Thr Thr Arg Arg
201 Ser Cys Ser Lys Thr Val Thr Lys Thr Val Ile Gly Pro Asp Gly His Lys Glu Val Thr
221 Lys Glu Val Val Thr Ser Glu Asp Gly Ser Asp Cys Pro Glu Ala Met
```

FIGURE 4: Complete amino acid sequence of largest CNBr peptide (HαCNI) derived from α chains of human fibrinogen. The 236 residues correspond to residues 241–476 in the intact α chain.

number of 39-residue duplications. The series of duplications accounted for certain technical difficulties in establishing the sequence. For example, many of the chymotryptic peptides were neutral 13-residue tryptophan-terminating peptides with very similar amino acid compositions. A similar set of peptides was generated by the chemical cleavage of HαCNI at tryptophanyl bonds.

```
1  ...Thr-Glu-Ser-Pro-Arg-Asn-Pro-Ser-Ser-Ala-Gly-Ser-Trp-
2     Asn-Ser-Gly-Ser-Ser-Gly-Pro-Gly-Ser-Thr-Gly-Asn-Arg-
3     Asn-Pro-Gly-Ser-Ser-Gly-Thr-Gly-Ser-Gly-Ala-Thr-Trp-
4     Lys-Pro-Gly-Ser-Ser-Gly-Pro-Gly-Ser-Thr-Gly-Ser-Trp-
5     Asn-Ser-Gly-Ser-Ser-Gly-Thr-Gly-Ser-Thr-Gly-Asn-Gln-
6     Asn-Pro-Gly-Ser-Pro-Arg-Pro-Gly-Ser-Thr-Gly-Thr-Trp-
7     Asn-Pro-Gly-Ser-Ser-Glu-Arg-Gly-Ser-Ala-Gly-His-Trp-
8     Thr-Ser-Glu-Ser-Ser-Val-Ser-Gly-Ser-Thr-Gly-Gln-Trp-
9     His-Ser-Glu-Ser-Gly-Ser-Phe-Arg-Pro-Asp-Ser-Pro-Gly-
10    Ser-Gly-Asn-Ala-Arg-          Pro-Asn-Asp-Pro-Asn-Trp...
```

FIGURE 5: Thirteen-residue homology repeat unit found in CNI region of human fibrinogen α chain.

The occurrence of several anomalous enzymatic cleavages also complicated the unravelment of the sequence. Trypsin fractionally, but reproducibly, cleaves the arginyl–proline bonds at 94 to 95 and 145 to 146. The arginyl–proline bond 135 to 136 is not attacked, however, nor is the lysyl–proline bond at 63 to 64. None of the other enzymes employed (plasmin, staphylococcal protease, or chymotrypsin) effected cleavage at the two unexpected sites attacked by trypsin. On the other hand, chymotrypsin repeatedly cleaved on the carboxyl side of threonine-207 and threonine-209. Staphylococcal protease split next to serine-201; there is a possibility that the serine is phosphorylated.[2]

Some other features of the amino acid composition are significant. For example, there are only two glutamines among the 236 residues, and both of them turn out to be acceptor residues for the factor XIII catalyzed incorporation of substitute cross-linking donors (Cottrell et al., 1979). The question arises as to whether these are the preferred acceptors because they are the *only* possible acceptors or they are the only glutamines because of the need for specificity in cross-linking.

The fragment contains two cysteine residues located 30 residues apart. Independent studies have shown that these cysteines are disulfide-linked in the native molecule (Takagi & Doolittle, 1975; Harfenist & Canfield, 1975). Interestingly, some faint homology can be discerned in the area of the disulfide loop with the smaller intrachain disulfide bonds that occur in the β and γ chains (Doolittle et al., 1978). With regard to the rest of the CNBr fragment, homology with the β and γ chains is apparent over the course of the first 60 residues, but then it abruptly disappears (Doolittle et al., 1979b). After a gap of 97 residues, the homology resumes at a low level and continues over a range of 70–80 residues, including the disulfide loops, until the carboxy termini of the β and γ chains are reached.

A computer scheme for predicting likely conformations (Garnier et al., 1978) indicated that HαCNI should be mostly random coil with some likelihood of reverse turns. In fact, analysis of a prototype 13-residue sequence based on the homology repeat indicated that the reverse turns ought to occur at regular intervals interspersed with the random coil. On the other hand, the plasmin-derived fragment, PLI (184 residues), is water-soluble, and examination by circular dichroism revealed a completely random coil structure.[3] The ready solubility of this fragment, in contrast to that of HαCNI, and the presence of two iodinatable tyrosine residues make PLI a useful agent for many studies involving the interaction of fibrinogen with other particles [see, e.g., Cottrell et al. (1979)].

[2] R. Jue and R. F. Doolittle, unpublished observations.

[3] The circular dichroism study was kindly conducted by M. Teintze.

Acknowledgments

We are grateful to Marcia Riley and Dennis Trovato for incomparable assistance in the isolation and analysis of many of the materials described in this work.

Supplementary Material Available

Experimental material including amino acid compositions of staphylococcal protease peptides from HαCNI (Table IS) and of chymotryptic peptides from HαCNI (Table IIS) (6 pages). Ordering information is given on any current masthead page.

References

Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 546.

Blombäck, B., Hessel, B., Iwanaga, S., Reuterby, J., & Blombäck, M. (1972) *J. Biol. Chem. 247*, 1496.

Cottrell, B. A., & Doolittle, R. F. (1976) *Biochem. Biophys. Res. Commun. 71*, 754.

Cottrell, B. A., & Doolittle, R. F. (1978) *Thromb. Res. 12*, 1135.

Cottrell, B. A., Strong, D. D., Watt, K. W. K., & Doolittle, R. F. (1979) *Biochemistry* (second of three papers in this issue).

Doolittle, L. R., Mross, G. A., Fothergill, L. A., & Doolittle, R. F. (1977) *Anal. Biochem. 78*, 491.

Doolittle, R. F. (1973) *Adv. Protein Chem. 27*, 1.

Doolittle, R. F., Schubert, D., & Schwartz, S. A. (1967) *Arch. Biochem. Biophys. 118*, 456.

Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Friezner, S. J., Hucko, J. T., & Takagi, T. (1977a) *Biochemistry 16*, 1703.

Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Friezner, S. J., & Takagi, T. (1977b) *Biochemistry 16*, 1710.

Doolittle, R. F., Cottrell, B. A., Strong, D. D., & Watt, K. W. K. (1978) *Biochem. Biophys. Res. Commun. 84*, 495.

Doolittle, R. F., Cottrell, B. A., Strong, D. D., & Watt, K. W. K. (1979a) *Thromb. Res. 14*, 787.

Doolittle, R. F., Watt, K. W. K., Cottrell, B. A., Strong, D. D., & Riley, M. (1979b) *Nature (London) 280*, 464.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97.

Gray, W. (1972) *Methods Enzymol. 25*, 333.

Harfenist, E. J., & Canfield, R. E. (1975) *Biochemistry 14*, 4110.

Laursen, R. A., Horn, M. J., & Bonner, A. G. (1972) *FEBS Lett. 21*, 67.

Lottspeich, F., & Henschen, A. (1978a) *Hoppe-Seyler's Z. Physiol. Chem. 359*, 1451.

Lottspeich, F., & Henschen, A. (1978b) *Hoppe-Seyler's Z. Physiol. Chem. 359*, 1611.

McKee, P. A., Rogers, L. A., Marler, E., & Hill, R. L. (1966) *Arch. Biochem. Biophys. 116*, 271.

McKee, P. A., Mattock, P., & Hill, R. L. (1970) *Proc. Natl. Acad. Sci. U.S.A. 66*, 738.

Mross, G. A., & Doolittle, R. F. (1977) *Mol. Biol., Biochem. Biophys. 25*, 1.

Ozols, J., & Gerard, C. (1977) *J. Biol. Chem. 252*, 5986.

Spies, J. R., & Chambers, D. C. (1949) *Anal. Chem. 21*, 1249.

Strong, D. D., Cottrell, B. A., Watt, K. W. K., & Doolittle, R. F. (1979) *Proc. Int. Congr. Biochem., 11th*, Abstract No. S98.

Takagi, T., & Doolittle, R. F. (1975) *Biochemistry 14*, 5149.

Watt, K. W. K., Takagi, T., & Doolittle, R. F. (1979) *Biochemistry 18*, 68.