Gross, E., & Witkop, B. (1962) *J. Biol. Chem. 237*, 1856.
Lorand, L. (1972) *Ann. N.Y. Acad. Sci. 202*, 6.
Lorand, L., & Jacobsen, A. (1964) *Biochemistry 3*, 1939.
Lorand, L., & Ong, H. H. (1966) *Biochemistry 5*, 1747.
McKee, P. A., Mattock, P., & Hill, R. L. (1970) *Proc. Natl. Acad. Sci. U.S.A. 66*, 738.
Pisano, J. J., Finlayson, J. S., Peyton, M. P., & Nagai, Y. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 770.
Strong, D. D., Watt, K. W. K., Cottrell, B. A., & Doolittle, R. F. (1979) *Biochemistry* (first of three papers in this issue).
Takagi, T., & Doolittle, R. F. (1975) *Biochemistry 14*, 5149.

Amino Acid Sequence Studies on the α Chain of Human Fibrinogen. Overlapping Sequences Providing the Complete Sequence†

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

ABSTRACT: The complete amino acid sequence of the α chain of human fibrinogen has been determined. It contains 610 amino acid residues and has a calculated molecular weight of 66 124. The chain has 10 methionines, and fragmentation with cyanogen bromide yields 11 peptides [Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Friezner, S. J., Hucko, J. T., & Takagi, T. (1977) *Biochemistry 16*, 1703]. The arrangement of the 11 fragments was determined by the isolation of peptide overlaps from plasmic and staphylococcal protease digests of fibrinogen and/or α chains. In addition, certain of the cyanogen bromide fragments, preliminary reports of whose sequences have appeared previously, have been reexamined in order to resolve several discrepancies. The α chain is homologous with the β and γ chains of fibrinogen, although a large repetitive segment of unusual composition is absent from the latter two chains. The existence of this unusual segment divides the sequence of the α chain into three zones of about 200 residues each that are readily distinguishable on the basis of amino acid composition alone.

The α chain is the largest of the three nonidentical subunits that constitute the vertebrate fibrinogen molecule ($\alpha_2\beta_2\gamma_2$). In the case of human fibrinogen, the α-chain molecular weight has been found to be 63 500 by ultracentrifugation (McKee et al., 1966) and 70 000 by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (McKee et al., 1970). We have now completed the amino acid sequence of the chain; it contains 610 amino acid residues and has a calculated molecular weight of 66 124.

Although the completion of the sequence of the largest cyanogen bromide fragment (Strong et al., 1979) actually represented the final step in the determination, there are a number of data bearing on the arrangement of the CNBr[1] fragments and details of sequences reported in brief and preliminary communications that have not been published previously. Also, there are some discrepancies with other reports that need resolving. In this article we provide the final data from this laboratory bearing on the amino acid sequence of the α chain of human fibrinogen. The sequence has provided a wealth of information about the structure, function, and evolution of the chain itself and the fibrinogen molecule as a whole, including features dealing with fibrin stabilization and fibrinolysis.

Experimental Section

Materials and Methods. All the procedures used in this study have been fully described in previous publications from this laboratory (for a listing of these references, see the preceding two articles in this issue).

Amino Acid Composition of α Chains and CNBr Fragments. In an earlier article (Doolittle et al., 1977a), we published the results of amino acid analysis of human fibrinogen α chains and compared them with results reported from other laboratories. Although the agreement with the final sequence was quite good (we had noted 611 residues), the serine value was low and values for the other amino acids were slightly inflated as a result. Accordingly, we undertook another set of analyses, taking care to use only α_2 chains, which extend all the way to the authentic carboxy terminus of the chain (Cottrell & Doolittle, 1976). The results are in excellent agreement with the reported sequence (Table I).

In that earlier publication (Doolittle et al., 1977a) we also reported the amino acid compositions of the 11 CNBr peptides based on the sequences observed to that point. In this article we now present the actual amino acid compositions determined on the individual purified fragments and compare them with the number of residues expected from the observed sequences (Table II). Although the agreement is generally excellent, a few notable exceptions do occur. In particular, fragment CNI contains two more alanine residues than were identified in the sequence operations.

Arrangement of CNBr Fragments. Previously, we had suggested an arrangement for 10 of the 11 CNBr fragments (Doolittle et al., 1977a), based, for the most part, on inferences derived from earlier and concurrent studies on the plasmic digestion of fibrinogen (Takagi & Doolittle, 1975a,b; Doolittle et al., 1977b). The position of the 11th peptide was subsequently established (Cottrell & Doolittle, 1978), also as the result of isolating an overlapping plasmic peptide from fibrinogen. In this report we summarize the data used as a basis for CNBr fragment arrangement (Table III), including for the first time the results of digesting α chains with staphylo-

† 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.

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

[1] Abbreviations used: CNBr, cyanogen bromide; Dns, 5-(dimethylamino)naphthalene-1-sulfonyl; PhNCS, phenyl isothiocyanate.

0006-2960/79/0418-5410$01.00/0

Table I: Amino Acid Composition of the α Chain of Human Fibrinogen

amino acid	residues/molecule	
	av[a] values	from sequence
CM-cysteine	7.4	8
aspartic acid[b]	63.5	64
threonine	50.7	47
serine	85.6	87
glutamic acid[b]	63.7	62
proline	34.1	35
glycine	67.0	69
alanine	22.4	22
valine	29.5	28
methionine	9.7	10
isoleucine	17.0	17
leucine	29.2	29
tyrosine	9.2	9
phenylalanine	18.2	19
histidine	14.9	15
lysine	38.8	39
arginine	39.4	40
tryptophan[c]	11.1	10
total:	611.4	610

[a] Calculated on the basis of 610 residues. The values are given the averages of duplicate runs of 24-, 48-, and 72-h acid hydrolyses. Serine and threonine values were extrapolated back to zero time; only 48- and 72-h results were used for valine and isoleucine. [b] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [c] Measured by the modified method of Spies & Chambers (1949).

coccal protease (Houmard & Drapeau, 1972). The enzyme, which cleaves preferentially on the carboxy side of glutamic acid residues, and to a lesser extent next to aspartic acid, yielded several key overlaps that confirm our previously suggested arrangement.

In this regard, three separate digestions of α chains were prepared over the course of the study. The first of these was initially passed over a Sephadex "G-37" column, the second was passed over a conventional Sephadex G-25 column, and the third was passed over Sephadex G-50. Many of the data presented here were derived from the Sephadex "G-37" column (Figure 1), which was prepared by mixing equal volumes of 1:1 suspensions of Sephadex G-25 and Sephadex G-50 and pouring the mixed slurry into a column. Although some settling occurs during the process, as indicated by an increasing "whiteness" from top to bottom, the column was still the most effective of the three employed in separating the staphylococcal protease peptides. In all instances, peptides were purified further by paper electrophoresis (pH 6.5 and/or 2.0) and, where needed, by paper chromatography (1-butanol–acetic acid–water, 4:1:5). The amino acid compositions of seven overlap peptides (containing nine methionines) are listed in Table IV.

Sequences of Overlap Peptides. Different approaches were employed for the determination of the sequences of the overlap peptides depending on their size, composition, and the relative importance of the overlap. In this regard, peptides SP3F2, SP4G2, and SP3G1 were attached to aminoethylaminopropyl glass beads with a water-soluble carbodiimide reagent and successfully degraded over their entire lengths by the thioacetylation procedure (Mross & Doolittle, 1977). Peptides SP3E3 and SP3J4b, on the other hand, were digested with trypsin, and the expected subpeptides were isolated, analyzed, and subjected to end-group analysis by the Dns procedure (Gray, 1972).

Ambiguous Amides. On several occasions, peptides isolated from the staphylococcal protease digestions were at odds with prior amide assignments made on the basis of mobilities of tryptic peptides. Thus, partial cleavage with the staphylococcal protease was observed next to location α-234, a residue that we previously reported as asparagine (Takagi & Doolittle, 1975b), indicating the fractional existence of aspartic acid at that location. A similar situation was evident at Asn-212. The ambiguities may be the result of partial deamidation at these locations, either in vivo or during the isolation of the peptides. Peptides isolated from staphylococcal protease digests also led us to reverse the amide assignments in the carboxy-terminal peptide of HαCNIVA, such that α-143 is Gln and α-146 is Asp. A staphylococcal protease cleavage of α-137 also suggested some deamidation of Gln-137.

Table II: Amino Acid Compositions of 11 Cyanogen Bromide Fragments from the α Chain of Human Fibrinogen

	residues/molecule of peptide[a]										
	CNI	CNIII	CNIVA	CNIVB	CNVA	CNVB	CNVIA	CNVIB	CNVIC	CNVIIA	CNVIIB
CM-cysteine[b]	1.8 (2)	0.7 (2)			2.2 (4)						
aspartic acid[c]	20.2 (19)	7.8 (7)	8.4 (9)	3.9 (3)	7.2 (7)	9.3 (10)	3.1 (3)	4.5 (4)	2.1 (2)		
threonine	27.7 (29)		1.4 (1)	6.4 (6)		2.7 (3)	1.0 (1)	4.8 (5)	1.7 (2)		
serine	41.0 (46)	4.3 (4)	4.4 (4)	17.5 (18)	4.0 (5)	3.3 (3)	1.2 (1)	3.3 (3)	3.1 (3)		
glutamic acid[c]	21.0 (20)	9.2 (9)	8.6 (8)	8.8 (8)	5.1 (5)	4.6 (4)	3.6 (3)	2.1 (1)	3.2 (3)	1.2 (1)	
proline	16.7 (18)	1.7 (2)		1.8 (2)	3.5 (3)		4.4 (5)	3.2 (3)	1.0 (1)	0.8 (1)	
glycine	42.7 (43)	1.7 (1)	2.5 (1)	5.5 (6)	6.7 (7)	1.5 (1)	1.4 (1)	5.9 (6)	3.3 (3)		
alanine	7.1 (5)	2.8 (3)	2.3 (2)	1.3 (1)	3.0 (3)		1.1 (1)	2.4 (3)	4.1 (4)		
valine	9.5 (10)	3.7 (3)	5.1[g] (6)	1.3 (1)	1.7 (3)	1.3 (1)	2.9 (3)		0.9 (1)		
isoleucine	2.4 (2)	5.0 (5)	3.4[g] (4)	1.6 (2)		3.0 (3)		0.9 (1)			
leucine	4.0 (3)	7.5 (8)	6.0 (6)	0.5 (1)	1.0 (1)	4.5 (4)	2.9 (3)	3.1 (3)			
tyrosine	2.3 (2)	1.1 (1)	1.1 (1)	2.8 (3)	0.7 (1)	0.6 (1)					
phenylalanine	2.8 (2)		1.1 (1)	4.5 (5)	2.5 (2)	1.8 (2)	1.2 (1)	4.7 (6)			
histidine	4.4 (4)	0.9 (1)	1.0 (1)	1.9 (2)	0.7 (1)	0.8 (1)		1.8 (2)	2.8 (3)		
lysine	10.4 (10)	5.6 (6)	4.6 (4)	4.7 (5)	2.2 (2)	4.7 (5)	3.5 (4)	1.6 (1)	2.1 (2)		
arginine	13.8 (13)	7.2 (7)	6.2 (7)	3.0 (3)	3.6 (4)	1.1 (1)		1.9 (2)	2.0 (2)		1.0 (1)
tryptophan	6.0[d] (7)				++[e] (2)		+[e] (1)				
homoserine[f]	0.9 (1)	0.6 (1)	0.5 (1)	0.5 (1)	0.5 (1)	0.6 (1)	0.6 (1)	0.6 (1)		0.5 (1)	0.4 (1)
total:	(236)	(60)	(56)	(67)	(51)	(40)	(28)	(41)	(26)	(3)	(2)
residue no.:	241–476	148–207	92–147	518–584	1–51	52–91	208–235	477–517	585–610	236–238	239–240
amino terminal:	Glu	Lys	Glu	Leu	Ala	Lys	Lys	Asp	Ala	Pro	Arg

[a] Based on averages of two or more analyses in each case. [b] CM-cysteine = (carboxymethyl)cysteine. [c] Aspartic acid and glutamic acid values include asparagine and glutamine values, respectively. [d] Tryptophan was determined by modification of the method of Spies & Chambers (1949). [e] Tryptophan was determined by fluorescence on paper. [f] Homoserine values include homoserine lactone. [g] Includes the Val–Ile bond.

Table III: Basis for Arrangement of Cyanogen Bromide Fragments

overlap no.	Met residue	peptide[a]	residues	sequence of overlap peptides
1	51	E α	1-78[b]	ADSG...CRMKG...YWK
		SP3E3	40-57	DWNYKCPSGCRMKGLIDE
2	91[e]	SP3F2	83-92	SHSLTTNIME
3	147	D α	111-199[c]	VSED...VDMKR...RDR
4	207	PL4B2b	207-219[d]	MKPVPNLVPGNFK
		SP3J4b	199-212	RQHLPLIKMKPVPD
5	235	fragment A	231-[d]	ALTDMPQMRMEL...
6	238	SP4H4e	229-241	WKALTDMPQMRME
7	240	SP4G2	235-241	MPQMRME
8	476[e]	SP6D2	475-477	AMD
9	517	SP3G1	503-520	TASTGKTFPGFFSPMLGE
10	584	PL3A2	584-610	MADE...RPV
		SP4H5b	579-587	SKSYKMADE

[a] SP, staphylococcal protease; PL, plasmin; D and E are also products of plasmin cleavage. [b] From Takagi & Doolittle (1975a). [c] From Doolittle et al. (1977b). [d] From Takagi & Doolittle (1975b). [e] The corresponding tryptic peptides containing these methionine residues were also isolated and characterized.

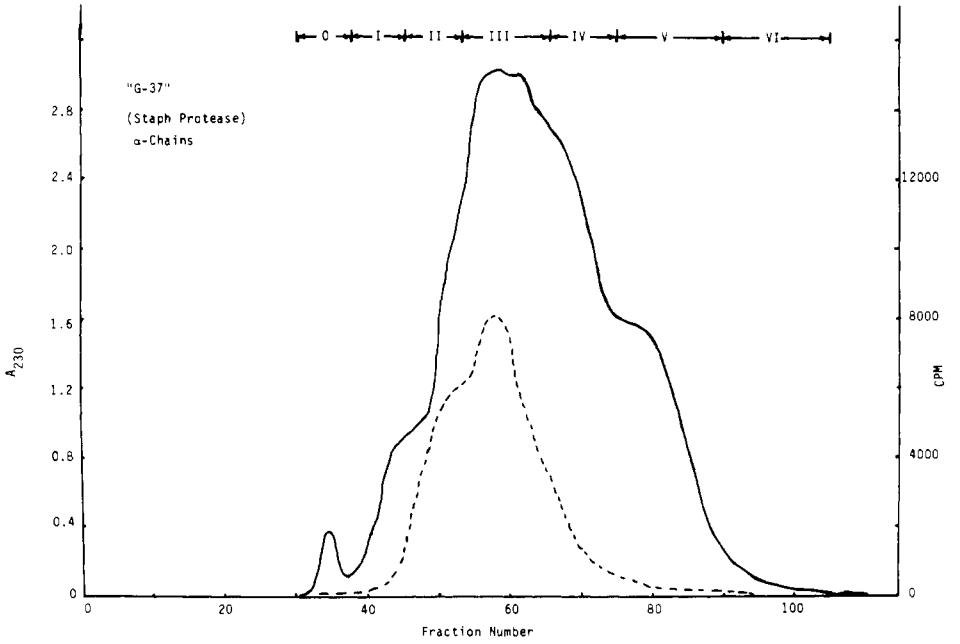

FIGURE 1: Gel filtration chromatography of a staphylococcal protease digest of human fibrinogen α chains (70 mg). The column (2.5 × 85 cm) was equilibrated and developed with 0.1 M ammonium bicarbonate. Flow rate, 60 mL/h; fraction size, 4.0 mL. Solid line, A_{230}; broken line, radioactivity from [^{14}C]carboxymethylated cysteines.

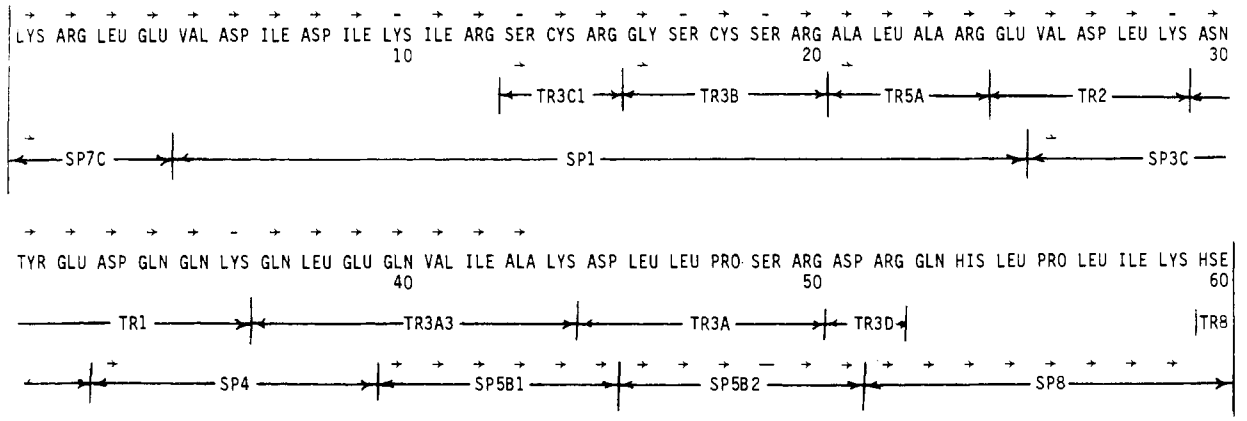

FIGURE 2: Summary of data used in completing the amino acid sequence of HαCNIII. (→) Successful identification from thioacetylation degradation; (⇀) identification by the Dns-PhNCS method. TR, trypsin; SP, staphylococcal protease.

HαCNIII. In our initial report on the characterization of CNBr peptides (Doolittle et al., 1977a), we successfully determined the first 43 residues of fragment HαCNIII by direct degradation using the thioacetylation procedure, but the remainder of the structure was presented in a composition form only. Moreover, many of the asparagine and glutamine residues were reported as Asx or Glx. Since that time, we completed the structure of the 60-residue peptide by isolating the staphylococcal protease and tryptic peptides of the fragment. A summary of the data is presented in Figure 2.

Resolution of Discrepancies. During the course of studies on the α chain of human fibrinogen, a number of discrepancies

Table IV: Amino Acid Compositions of Seven Methionine-Containing Peptides Isolated after Staphylococcal Protease Digestion of α Chains

	residues/molecule[a]						
	SP3E3[b]	SP3F2	SP3J4b[b]	SP4H4e	SP6D2	SP3G1c	SP4H5b
CM-cysteine	++[c] (2)						
aspartic acid[d]	3.0 (3)	2.1 (1)	1.4 (1)	1.2 (1)	1.1 (1)		1.5 (1)
threonine		1.5 (2)		1.3 (1)		3.0 (3)	
serine	0.6 (1)	2.1 (2)				2.1 (2)	1.7 (2)
glutamic acid[d]	1.1 (1)	1.4 (1)	1.2 (1)	2.0 (2)		1.2 (1)	1.2 (1)
proline	0.8 (1)		2.5 (3)	0.8 (1)		1.9 (2)	
glycine	1.8 (2)					3.1 (3)	
alanine				0.6 (1)	0.9 (1)	1.0 (1)	1.0 (1)
valine			1.1 (1)				
methionine	1.0 (1)	0.8 (1)	0.8 (1)	1.1 (3)	0.8 (1)	0.8 (1)	0.8 (1)
isoleucine	1.0 (1)	0.7 (1)	0.7 (1)				
leucine	1.0 (1)	0.8 (1)	2.4 (2)	0.5 (1)		1.3 (1)	
tyrosine	1.0 (1)						1.0 (1)
phenylalanine						2.7 (3)	
histidine		0.6 (1)	1.2 (1)				
lysine	2.0 (2)		1.8 (2)	0.9 (1)		1.0 (1)	1.9 (2)
arginine	0.7 (1)		1.0 (1)	0.6 (1)			
tryptophan[e]	+ (1)			+ (1)			
total:	(18)	(10)	(14)	(13)	(3)	(18)	(9)
amino terminal:[f]	Asp	Ser	Arg	(Trp)	Ala	Thr	Ser
methionine no.:	51	91	207	235, 238, 240	476	517	584
overlap no.:	1	2	4	5–7	8	9	10

[a] Conditions: 24-h hydrolysis; 108 °C; 5.7 N HCl. [b] Sum of compositions of four constituent tryptic peptides. [c] Determined by radioactivity of [^{14}C]carboxymethylated cysteine. [d] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [e] Qualitative determination by fluorescence. [f] Amino terminals were determined by the Dns method (Gray, 1972).

```
  1 Ala Asp Ser Gly Glu Gly Asp Phe Leu Ala Glu Gly Gly Gly Val Arg Gly Pro Arg Val Val Glu Arg His Gln
 26 Ser Ala Cys Lys Asp Ser Asp Trp Pro Phe Cys Ser Asp Glu Asp Trp Asn Tyr Lys Cys Pro Ser Gly Cys Arg
 51 Met Lys Gly Leu Ile Asp Glu Val Asn Gln Asp Phe Thr Asn Arg Ile Asn Lys Leu Lys Asn Ser Leu Phe Glu
 76 Tyr Gln Lys Asn Asn Lys Asp Ser His Ser Leu Thr Thr Asn Ile Met Glu Ile Leu Arg Gly Asp Phe Ser Ser
101 Ala Asn Asn Arg Asp Asn Thr Tyr Asn Arg Val Ser Glu Asp Leu Arg Ser Arg Ile Glu Val Leu Lys Arg Lys
126 Val Ile Gln Lys Val Gln His Ile Gln Leu Leu Gln Lys Asn Val Arg Ala Gln Leu Val Asp Met Lys Arg Leu
151 Glu Val Asp Ile Asp Ile Lys Ile Arg Ser Cys Arg Gly Ser Cys Ser Arg Ala Leu Ala Arg Glu Val Asp Leu
176 Lys Asn Tyr Glu Asp Gln Gln Lys Gln Leu Glu Gln Val Ile Ala Lys Asp Leu Leu Pro Ser Arg Asp Arg Gln
201 His Leu Pro Leu Ile Lys Met Lys Pro Val Pro Asn Leu Val Pro Gly Asn Phe Lys Ser Gln Leu Gln Lys Val
226 Pro Pro Glu Trp Lys Ala Leu Thr Asp Met Pro Gln Met Arg Met Glu Leu Glu Arg Pro Gly Gly Asn Glu Ile
251 Thr Arg Gly Gly Ser Thr Ser Tyr Gly Thr Gly Ser Glu Thr Glu Ser Pro Arg Asn Pro Ser Ser Ala Gly Ser
276 Trp Asn Ser Gly Ser Ser Gly Pro Gly Ser Thr Gly Asn Arg Asn Pro Gly Ser Ser Gly Thr Gly Ser Gly Ala
301 Thr Trp Lys Pro Gly Ser Ser Gly Pro Gly Ser Thr Gly Ser Trp Asn Ser Gly Ser Ser Gly Thr Gly Ser Thr
326 Gly Asn Gln Asn Pro Gly Ser Pro Arg Pro Gly Ser Thr Gly Thr Trp Asn Pro Gly Ser Ser Glu Arg Gly Ser
351 Ala Gly His Trp Thr Ser Glu Ser Ser Val Ser Gly Ser Thr Gly Gln Trp His Ser Glu Ser Gly Ser Phe Arg
376 Pro Asp Ser Pro Gly Ser Gly Asn Ala Arg Pro Asn Asp Pro Asn Trp Gly Thr Phe Glu Glu Val Ser Gly Asn
401 Val Ser Pro Gly Thr Arg Arg Glu Tyr His Thr Glu Lys Leu Val Thr Ser Lys Gly Asp Lys Glu Leu Arg Thr
426 Gly Lys Glu Lys Val Thr Ser Gly Ser Thr Thr Thr Thr Arg Arg Ser Cys Ser Lys Thr Val Thr Lys Thr Val
451 Ile Gly Pro Asp Gly His Lys Glu Val Thr Lys Glu Val Val Thr Ser Glu Asp Gly Ser Asp Cys Pro Glu Ala
476 Met Asp Leu Gly Thr Leu Ser Gly Ile Gly Thr Leu Asp Gly Phe Arg His Arg His Pro Asp Glu Ala Ala Phe
501 Phe Asp Thr Ala Ser Thr Gly Lys Thr Phe Pro Gly Phe Phe Ser Pro Met Leu Gly Glu Phe Val Ser Glu Thr
526 Glu Ser Arg Gly Ser Glu Ser Gly Ile Phe Thr Asn Thr Lys Glu Ser Ser Ser His His Pro Gly Ile Ala Glu
551 Phe Pro Ser Arg Gly Lys Ser Ser Ser Tyr Ser Lys Gln Phe Thr Ser Ser Thr Ser Tyr Asn Arg Gly Asp Ser
576 Thr Phe Glu Ser Lys Ser Tyr Lys Met Ala Asp Glu Ala Gly Ser Glu Ala Asp His Glu Gly Thr His Ser Thr
601 Lys Arg Gly His Ala Lys Ser Arg Pro Val
```

FIGURE 3: Complete amino acid sequence of human fibrinogen α chain.

have arisen in reports from different laboratories, and we have now reexamined the regions where differences have been observed. The numbering used in this discussion refers to the complete α-chain sequence given in Figure 3. Discrepancies are taken up in order from the amino terminal.

(*1*) *Residues 79–81 (HαCNVB).* Originally we reported

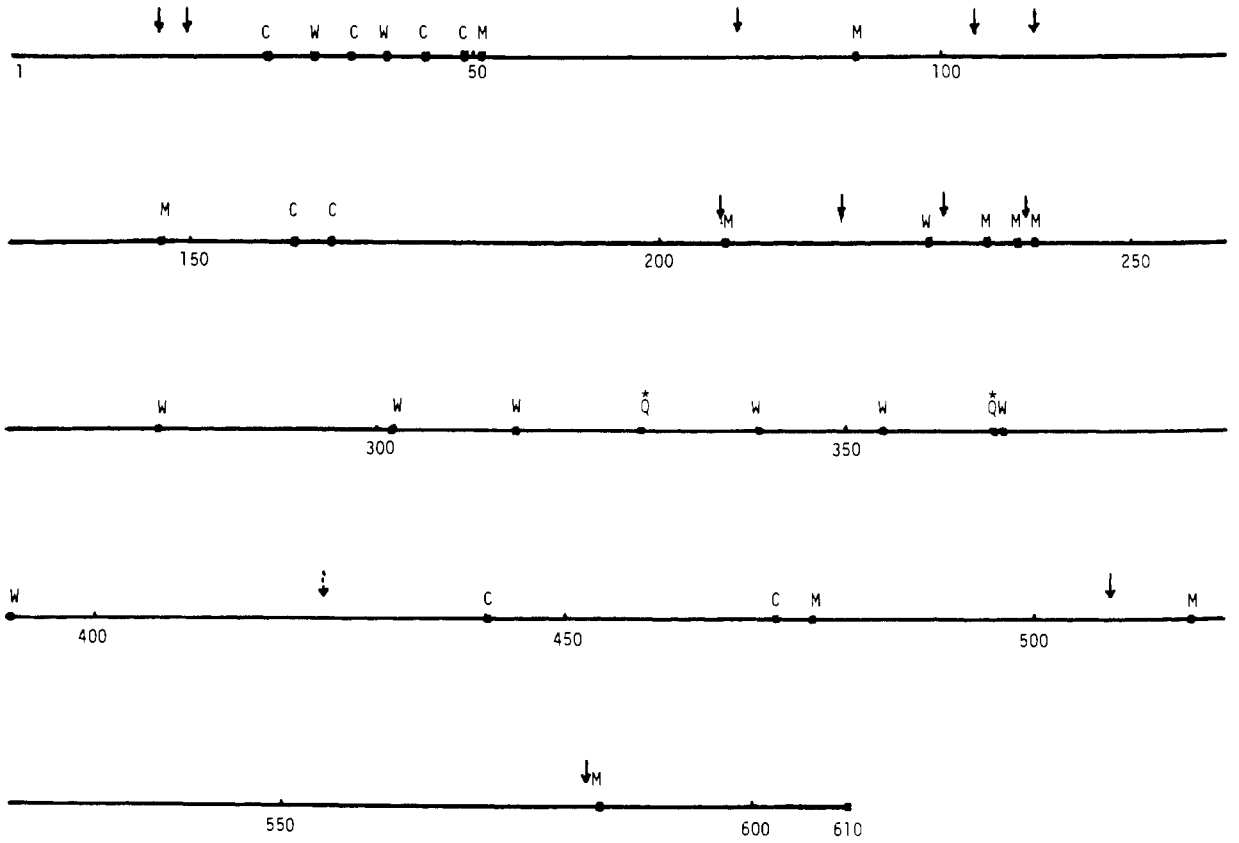

FIGURE 4: Schematic representation of human fibrinogen α chain showing locations of 10 methionines (M), 8 cysteines (C), 10 tryptophans (W), and two cross-linking acceptor sites (Q*). Arrows denote known cleavage points by plasmin-digesting fibrinogen (broken arrow represents plasmin digestion of HαCNI).

HαCNVB as containing 37 residues (Doolittle et al., 1977a). Gårdlund (1977), studying a CNBr peptide from fibrinogen, subsequently showed that this fragment actually contains 40 residues. In fact, we had missed a tryptic tripeptide, Asn-Asn-Lys. The peptide, which we have since isolated and sequenced, stains poorly with ninhydrin. The adjacent tryptic peptide begins with aspartic acid, and we had depended on a one-residue overlap with a 28-step sequencer run. Ironically, the sequencer run had given us a small amount of lysine at step 30, but in the absence of a confirming tryptic peptide we disregarded the result as "noise" from being too far into the degradation.

(*2*) *Residue 112* (*HαCNIVA*). Originally Collen et al. (1975), in a study on the amino-terminal sequences of fragment D, reported the sequence ...Arg-Val-Val-Glu... for the section that turns out to be α-chain residues 110–113. In our study, however, we found the sequence to be ...Arg-Val-Ser-Glu... (Doolittle et al., 1977b). The sequence that we reported has recently been confirmed by Lottspeich & Henschen (1978b).

(*3*) *Residues 196–199* (*HαCNIII*). We have reported the sequence for this region as ...Ser-Arg-Asp-Arg... (Doolittle et al., 1979). Lottspeich & Henschen (1978b) have reported the sequence as ...Arg-Ser-Asp-Arg. We feel confident about our reported sequence on the grounds that we isolated the tryptic peptide Asp-Arg from both HαCNIII and from the α-chain portion of fragment D (Doolittle et al., 1977b). Also, we have the corresponding staphylococcal protease peptides (Figure 2).

(*4*) *Residues 557–562* (*HαCNIVB*). In a report on the carboxy-terminal region of the α chain, Lottspeich & Henschen (1978a) noted that the penultimate CNBr peptide contained 67 residues. In an earlier article, Cottrell & Doolittle (1978) had reported 59 residues for that segment, although we pointed out that one overlap was missing and that the sequence at that junction had to be regarded as tentative. We have now isolated and sequenced the tryptic and staphylococcal protease peptides that confirm the sequence reported by Lottspeich & Henschen (1978a). Interestingly, we had reported the missing tryptic peptide in an earlier paper (Doolittle et al., 1977a), but, because of the difficulty in obtaining an overlap, decided it was a contaminant from elsewhere in the chain.

Discussion

In essence, the primary structure of the α chain of human fibrinogen was initiated by studies on the 16-residue fibrinopeptide A (Blombäck et al., 1962). The next significant advance came with the elucidation of the α-chain portion of the CNBr fragment of fibrinogen known as the "disulfide knot" (Blombäck et al., 1972), the sequence of which extended the known primary structure another 35 residues to α-51. Next, Takagi & Doolittle (1975a) determined the sequence of the adjacent 27 residues of the chain (α-52–α78) and thereby completed the structure of the α-chain segment of the plasmin-derived fragment E. At about the same time, scattered sequences from middle sections of the α chain were reported from several laboratories (Harfenist & Canfield, 1975; Takagi & Doolittle, 1975b; Hessel, 1975; Collen et al., 1975), and soon thereafter we reported the sequence of the carboxy-terminal CNBr fragment (Cottrell & Doolittle, 1976).

At that point we published a detailed characterization of all 11 CNBr fragments and suggested an arrangement for 10 of them (Doolittle et al., 1977a). The characterization included the complete sequences of several previously unreported fragments and partial sequences of the others. We also pres-

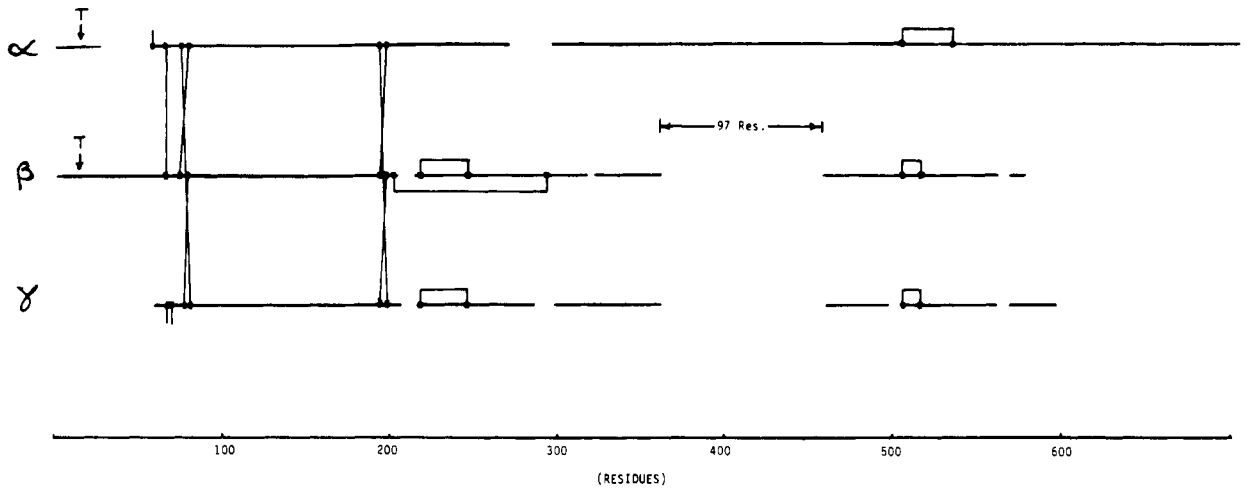

FIGURE 5: Alignment of α, β, and γ chains from human fibrinogen showing locations of disulfide attachments. The thrombin attack points for removal of the fibrinopeptides A and B are indicated with arrows and (T).

ented virtually the entire sequence of the α-chain portion of fragment D, which contains 89 or 95 residues, depending on the degree of plasmic degradation (Doolittle et al., 1977b). Since that time a number of reports have appeared from this laboratory and elsewhere, including two dealing with the carboxy-terminal quarter of the chain (Cottrell & Doolittle, 1978; Lottspeich & Henschen, 1978a), a disulfide loop region (Doolittle et al., 1978b; Henschen et al., 1978), and a report which mainly confirmed the previously reported sequence of the amino-terminal portion of the chain (Lottspeich & Henschen, 1978b). Recently, we published a brief communication covering about 95% of the sequence (Doolittle et al., 1979a), as well as a preliminary report of the entire sequence (Doolittle et al., 1979b). The most difficult part of the sequence determination proved to be the largest CNBr fragment, HαCNI (Strong et al., 1979), the structure of which, taken together with the overlapping sequences reported in this article, completes the α-chain sequence.

The unraveling of the primary structure of the α chain of human fibrinogen has revealed some interesting and unusual characteristics. For example, the first third of the sequence (approximately residues 1–200) has an ordinary amino acid composition. Much of this portion of the chain evidently exists as α helix in a three-stranded rope with the β and γ chains (Doolittle et al., 1978a). The next 200 residues (approximately residues 201–400), on the other hand, have a very unusual amino acid distribution typified by large numbers of serine, glycine, threonine, and proline residues. Seven tryptophans also occur in this segment. Moreover, the tryptophan residues occur with a definite periodicity that reveals a 13-residue homologous structure repeated 9 to 10 times through this central region. The last 200-residue section of the α chain is also rich in serines, threonines, and glycine, but—in contrast to the middle section—there is a significant contribution of nonpolar residues. A schematic depiction of the α chain denoting certain key residues, including methionines, cysteines, tryptophans, and the two glutamines which act as acceptors for substitute cross-link donors, is presented in Figure 4. Also shown are 12 known plasmin attack points, many of which, incidentally, proved useful in determining the arrangement of CNBr fragments.

As has been pointed out previously (Doolittle, 1976), all three of the nonidentical chains that comprise fibrinogen are homologous and have descended from a common ancestor. The α chain contains features which the β and γ chains do not, however, including the 13-residue homology repeat section. The three chains can be aligned readily at their cysteine residues (Figure 5) and homologous segments observed throughout their lengths if a 97-residue deletion in the β and γ chains (or an equivalent insertion in the α chain) is included (Doolittle et al., 1979b). The 97-residue section includes the two α-chain cross-link acceptor sites and may play a role in the interaction of fibrinogen with other macromolecules and platelets.

Acknowledgments

In addition to the authors of this article, a number of other persons in our group have made contributions to the determination of the α-chain sequence over the course of several years, including T. Takagi, K. G. Cassman, S. J. Friezner, J. Hucko, M. Riley, D. Trovato, and L. R. Doolittle. We acknowledge their collaboration and assistance and thank them sincerely. We also thank K. Anderson, M. Licciardi, and E. Ottoson for the burdensome typing required for sequence data presentation. We are also grateful to N. Woodbury for computer assistance.

References

Blombäck, B., Blombäck, M., Edman, P., & Hessel, B. (1962) *Nature (London) 193*, 883.

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

Collen, D., Kudryk, B., Hessel, B., & Blombäck, B. (1975) *J. Biol. Chem. 250*, 5808.

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

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

Doolittle, R. F. (1976) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 35*, 2145.

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., Goldbaum, D. M., & Doolittle, L. R. (1978a) *J. Mol. Biol. 120*, 311.

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

Doolittle, R. F., Cottrell, B. A., Strong, 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.

Gårdlund, B. (1977) *Thromb. Res. 10*, 689.

Gray, W. (1972) *Methods Enzymol. 25*, 333.
Harfenist, E. J., & Canfield, R. E. (1975) *Biochemistry 14*, 4110.
Henschen, A., Lottspeich, F., & Hessel, B. (1978) *Hoppe-Seyler's Z. Physiol. Chem. 359*, 1607.
Hessel, B. (1975) *Thromb. Res. 7*, 75.
Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506.
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. Biphys. 25*, 1.
Spies, J. R., & Chambers, D. C. (1949) *Anal. Chem. 21*, 1249.
Strong, D. D., Watt, K. W. K., Cottrell, B. A., & Doolittle, R. F. (1979) *Biochemistry* (first of three papers in this issue).
Takagi, T., & Doolittle, R. F. (1975a) *Thromb. Res. 7*, 813.
Takagi, T., & Doolittle, R. F. (1975b) *Biochemistry 14*, 5149.

Covalent Structure of Collagen: Amino Acid Sequence of α2-CB5 of Chick Skin Collagen Containing the Animal Collagenase Cleavage Site†

S. N. Dixit,* C. L. Mainardi,‡ J. M. Seyer, and A. H. Kang

ABSTRACT: The amino acid sequence of the 112 residues from the amino terminus of α2-CB5 from chick skin collagen was determined by automated sequential degradation of intact α2-CB5 and several chymotryptic and tryptic peptides. This segment of the peptide includes the site of the action of animal collagenases. As compared to the sequence around the α1 cleavage site, the α2 sequence is notable for the remarkable constancy of the residues to the amino side and the relative abundance of hydrophobic residues to the carboxyl side of the cleavage site, suggesting that these features are important in the recognition by the enzyme. The sequence of this region of the α2 chain is consistent with the Gly-X-Y triplet structure and the preference of certain residues for either the X or Y position in distribution. However, three of the six residues of leucine were found in the Y position rather than the X position. Leucine residues were found only once in the Y position in the α1(I) chain. This preference does not appear to hold in the α2 chain.

Comparisons of the primary structure of collagens from different species can provide useful information relating structure to molecular properties. The interstitial collagen type I, the major protein constituent of skin, bone, and tendon, exists as a triple-stranded helix composed of two α1(I) chains and one α2 chain, each containing over 1000 amino acid residues. The complete amino acid sequence of the α1(I) chain from chick skin collagen has now been established (Kang & Gross, 1970; Kang et al., 1975; Dixit et al., 1975a,b; Highberger et al., 1975), as have the amino acid sequences of substantial portions of rat and calf α1(I) (Hulmes et al., 1973; Gallop & Paz, 1975; Piez, 1976; Fietzek & Kuhn, 1967). Knowledge of the covalent structure of the α2 chain is still incomplete. Cleavage of the α2 chain with CNBr gives rise to three small peptides, α2-CB1 (14 residues), α2-CB0 (3 residues), and α2-CB2 (30 residues), and three large peptides, α2-CB3, α2-CB4, and α2-CB5, containing 338, 321, and 320 amino acid residues respectively (Kang et al., 1969a; Fietzek & Piez, 1969). The order of the CNBr peptides in the α chain has been determined to be CB1-0-4-2-3-5 (Vuust et al., 1970; Igarashi et al., 1970). The covalent structure of α2-CB1 and α2-CB0 of chick, rat, and calf and α2-CB2 of chick, calf, human, rabbit, rat, and guinea pig skin collagens has been reported (Kang & Gross, 1970; Kang et al., 1967; Fietzek et al., 1974a,b; Highberger et al., 1971). In addition, the complete amino acid sequence of α2-CB3 of chick skin (Dixit et al., 1977a,b) and α2-CB4 of calf skin (Fietzek & Rexrodt, 1975) has been published. Also the sequence of 45 amino-terminal residues of the rat skin α2-CB5 has been determined (Fietzek & Kuhn, 1973).

As a continuation of our systematic investigations on the primary structure of type I collagen from chick skin, we now report the amino acid sequence of the amino-terminal region of α2-CB5 of chick skin which contains the cleavage site of animal collagenases (Gross & Nagai, 1965; Gross et al., 1974). To our knowledge the covalent structure of the α2 chain around the cleavage site has not previously been determined.

Materials and Methods

Preparation of α2-CB5. Purified chick skin collagen was prepared from 3-week-old lathyritic white Leghorn chicks by neutral salt and acid extraction procedures described previously (Kang et al., 1969b). The α2 chain was isolated by CM-cellulose[1] chromatography and was subjected to CNBr cleavage in 0.1 N HCl at a concentration of 4 to 5 mg/mL at 37 °C for 4 h (Bornstein & Piez, 1966) with approximately twice the weight of CNBr. At the end of incubation, the digest was diluted 10-fold with cold water and lyophilized. The peptide

† From the Veterans Administration Medical Center and the Departments of Biochemistry and Medicine, University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104. *Received June 29, 1979; revised manuscript received September 14, 1979.* This work was supported by the Veterans Administration and by National Institutes of Health Grants AM-16506 and AM-20385.

‡ Research Associate of the Veterans Administration.

[1] Abbreviations used: CM-cellulose, carboxymethylcellulose; DEAE, diethylaminoethyl; EDTA, ethylenediaminetetraacetate; $NaDodSO_4$, sodium dodecyl sulfate; >PhNCS, phenylthiohydantoin.

0006-2960/79/0418-5416$01.00/0