Amino Acid Sequence of the β Chain of Human Fibrinogen†

K. W. K. Watt, T. Takagi,‡ and R. F. Doolittle*

ABSTRACT: The β chain of human fibrinogen contains 461 amino acid residues, 15 of which are methionines. The calculated molecular weight, independent of a single carbohydrate cluster, is 52230. In this regard, we have isolated and characterized all 16 cyanogen bromide fragments. In one case (CN1), we have concentrated on a disputed portion of a previously reported fragment. The arrangement of the cyanogen bromide peptides was deduced by the use of overlap fragments obtained from the tryptic digestion of modified and unmodified β chains and from digestions with staphylococcal protease, as well as by considerations involving the plasmic digestion products of fibrinogen. In one case two adjacent fragments were aligned by homology with the corresponding segments of the γ chain. The homology of the β chain with the γ chain is especially strong over the course of the carboxy-terminal two-thirds of the sequence. Neither of these chains appears to be homologous with the α chain in these regions. With a few minor exceptions, the sequence reported in this article is in agreement with data reported by other groups in Stockholm and Munich.

Vertebrate fibrinogen molecules are composed of three pairs of disulfide-linked nonidentical polypeptide chains ($\alpha_2\beta_2\gamma_2$). In the case of human fibrinogen, the α, β, and γ chains have molecular weights of 65000, 55000, and 47000, respectively (McKee et al., 1966). The corresponding number of amino acids in each is approximately 625, 460, and 420. The β and γ chains also contain carbohydrate clusters which contribute to their overall molecular weights (Pizzo et al., 1972; Gaffney, 1972). During the past 5–10 years, major amino acid sequencing efforts on human fibrinogen and its fragments have been under way in several different laboratories. In the case of the β chain, the first major contribution to the sequence was reported by Blombäck & Blombäck (1972) who reported most of the sequence of a 115-residue fragment isolated from a cyanogen bromide digest of fibrinogen. The fragment corresponded to the amino terminal of the β chain and contained the fibrinopeptide B. The size of the fragment was subsequently amended to 118 residues, all but two residues of which were positioned (Blombäck et al., 1976). Takagi & Doolittle (1975a) reported the sequence of the carboxy-terminal cyanogen bromide peptide, noting that it was homologous to the corresponding segment of the γ chain. We also reported the sequence of the peptide immediately adjacent to the amino-terminal cyanogen bromide fragment which is found in the plasmin-derived moiety known as fragment E (Takagi & Doolittle, 1975b). Collen et al. (1975) reported the first ten residues of fragment D, and Henschen & Lottspeich (1976) published the sequence of a 42-residue segment which overlapped and connected the fragment E and D sequences noted above. Töpfer-Peterson et al. (1976) then completed the sequence of a 21-residue carbohydrate-containing section.

Recently we reported the amino acid sequence of most of the cyanogen bromide fragments of the β chain and suggested an arrangement for them based primarily on observed homologies with α and/or γ chains (Watt et al., 1977). Subsequently, Lottspeich & Henschen published two preliminary notes (1977a,b) which also reported the characterization of the cyanogen bromide peptides and another dealing with the homology of β and γ chains (Henschen & Lottspeich, 1977a). Recently these same authors have published a sequence for the segment corresponding to residues 119–461 (Henschen & Lottspeich, 1977b) which, combined with Blombäck's data for residues 1–118, would complete the sequence. At the time, we had submitted a communication reporting a tentative sequence, noting that several ambiguities remained because of not having all overlaps in hand (Watt et al., 1978).

In this article we report the amino acid sequence of the β chain, complete with supporting data. In almost all cases we note agreement with the sequences reported by Blombäck et al. (1976) and by Henschen & Lottspeich (1977b), although in one significant case a difference is noted (at residues 108–114, as reported by Blombäck et al., 1976). There are also two cases involving amides where we differ from Henschen & Lottspeich (1977b); in a few other cases involving amides we have not yet been able to make firm assignments.

Experimental Section

Materials and Methods. Almost all the materials and methods used in this study have been described in detail in previous publications from this laboratory, including the preparation of fibrinogen, the isolation of the individual chains, the production of plasmin degradation products, the cyanogen bromide recipe employed, enzyme digestion conditions, gel filtration and paper electrophoresis conditions, and amino acid analyses, end-group, stepwise degradation, and solid-phase sequencing procedures (Doolittle et al., 1967, 1977, 1977a,b; Takagi & Doolittle, 1975a,c).

Amino Acid Composition of β Chains. During the past several years we have analyzed several different batches of β chains after total acid hydrolysis for 24, 48, and 72 h (5.7 N HCl, 108 °C). A set of average values recalculated for a chain length of 461 residues is listed in Table I. In general, the agreement with previous reports (McKee et al., 1966; Cartwright & Kekwick, 1971; Henschen & Edman, 1972) is good. Moreover, the agreement with the final derived sequence is excellent (Table I).

Isolation of Cyanogen Bromide Fragments. CNBr[1] fragmentation was routinely conducted on 50–60-mg lots of the β chain (ca. 1 μmol). The chains were dissolved in 70% formic acid (5–7 mg/mL), and solid CNBr was added (10 mg/mL). Digestions were for 16–24 h at room temperature.

† From the Department of Chemistry, University of California, San Diego, La Jolla, California 92093. *Received June 1, 1978.* Supported by U.S. Public Health Service Grants HE-12759, HE-18576, and GM-17702. A preliminary account of this work was presented at the VIth International Congress on Thrombosis and Hemostasis, June 27, 1977 (Watt et al., 1977).

‡ Present address: Biological Institute, Tohoku University, Sendai, Japan.

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

0006-2960/79/0418-0068$01.00/0

Table I: Amino Acid Composition of the β Chain of Human Fibrinogen[a]

amino acid	av values[a]	from sequence
aspartic acid[b]	60.4	60
threonine	21.1	22
serine	31.7	31
glutamic acid[b]	57.7	56
proline	21.8	23
glycine	43.7	42
alanine	25.6	23
cysteine	10.0	11
valine	24.6	25
methionine	14.1	15
isoleucine	14.7	16
leucine	27.3	28
tyrosine	19.3	21
phenylalanine	9.9	10
histidine	7.4	7
lysine	31.4	31
arginine	27.5	27
tryptophan[c]	12.5	13
	460.7	461

[a] Calculated on the basis of 461 residues. The values given are the averages of duplicate runs of 24-, 48-, and 72-h acid hydrolysates. 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 a scaled down modification of the method of Spies & Chambers (1949).

Table II: Sixteen CNBr Fragments Isolated from the β Chain of Human Fibrinogen

G-50 pool[a]	amino-terminal sequence	no. of residues	no. of cysteines	location in chain
I	PCA-Gly-Val···[b]	118	3	1-118
II	Tyr-Leu-Leu···	72	0	119-190
III	Asn-Thr-Glu···	63	1	243-305
IVA	Glu-Tyr-Cys···	34	4	191-224
IVB	Glu-Asp-Trp···	40	0	315-354
VA	Phe-Phe-Ser···	53	2	374-426
VB	Gly-Glu-Asn···[c]	6	0	362-367
VIA	Gly-Pro-Thr···	9	0	306-314
VIB1	Asn-Gly-Ala···	7	0	355-361
VIB2	Tyr-Leu-Ile···	18	1	225-242
VIB3	Ser-Met	2	0	451-452
VIC1	Thr-Ile-His···	6	0	368-373
VIC2	Ala-Lys-His···	12	0	427-438
VID	Lys-Ile-Arg···[d]	9	0	453-461
VIF	Arg-Lys-Met	3	0	448-450
VII	Asn-Trp-Lys···	9	0	439-447
		461	11	

[a] Roman numeral indicates pool designation from G-50 column (Figure 1). [b] Starting from amino-terminal peptide of β chain. [c] Carbohydrate-containing peptide. [d] Carboxy-terminal peptide (no homoserine).

After dilution and freeze-drying, the preparations were dissolved in 10% acetic acid and passed over a Sephadex G-50 column (Figure 1). In some early experiments (Takagi & Doolittle, 1975a) we divided the column effluent into eleven pools. Subsequently we revised the procedure to seven reproducible pools (Figure 1). The Roman numeral designation for all fragments refers to the G-50 pool in which it originated.

Pools I–III were purified further by rechromatography on either G-100 or G-50 columns. Pools IV–VII were purified by low-voltage paper electrophoresis at pH 6.5 and/or 2.0. The amino acid compositions of the 16 purified CNBr fragments are tabulated in Tables III and IV and compared with the integral values obtained from sequencing. The yields of the various peptides vary depending on how many steps were necessary for the particular purification. Also, the best compositions obtained are shown; in fact most of these fragments were isolated on numerous occasions, and several different G-50 columns are represented in the tables.[2]

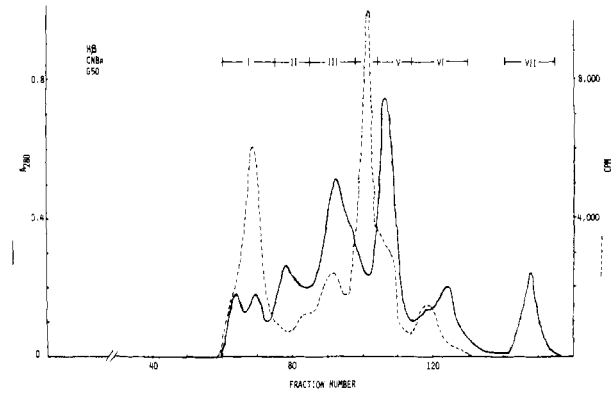

FIGURE 1: Gel filtration (Sephadex G-50) of CNBr digest of carboxymethylated β chains from human fibrinogen. Column = 2.5 × 170 cm, solvent = 10% acetic acid, flow rate = 60 mL/h, fraction size = 5 mL. Solid line = absorbance at 280 nm; broken line = radioactivity due to carboxymethylation.

Table III: Amino Acid Compositions of Six Large CNBr Fragments from the β Chain of Human Fibrinogen

	residues/mole of peptide[a]					
	CNI[b]	CNII	CNIII	CNIVA	CNIVB	CNVA
CM-cysteine[c]	2.5 (3)	–	1.0 (1)	3.7 (4)	–	1.5 (2)
aspartic acid[d]	12.6 (12)	11.1 (11)	10.4 (12)	1.4 (1)	6.1 (6)	9.2 (10)
threonine	3.6 (3)	2.3 (2)	4.6 (5)	3.2 (3)	1.9 (2)	3.7 (3)
serine	8.6 (9)	6.5 (7)	2.5 (2)	2.8 (3)	1.1 (1)	3.9 (3)
glutamic acid[d]	17.3 (17)	12.7 (13)	6.7 (6)	6.1 (6)	4.2 (4)	4.3 (3)
proline	9.7 (11)	1.6 (1)[h]	2.3 (2)	2.5 (2)	–	3.0 (2)
glycine	9.5 (10)	–	9.7 (10)	3.6 (3)	4.9 (5)	8.5 (7)
alanine	11.6 (13)	1.9 (1)[h]	1.0 (1)	–	4.2 (4)	2.1 (2)
valine	6.8 (6)	4.7 (6)	3.4 (3)	3.2 (3)	2.3 (3)	–
isoleucine	2.5 (2)	3.7 (4)	2.0 (2)	2.9 (3)	1.1 (1)	–
leucine	9.2 (9)	9.9 (10)	3.1 (3)	–	1.5 (1)	1.1 (1)
tyrosine	2.6 (2)	3.5 (3)	3.3 (3)	0.9 (1)	3.0 (3)	3.8 (5)
phenylalanine	2.6 (3)	–	2.0 (2)	–	1.1 (1)	1.3 (2)
histidine	1.8 (2)	1.1 (1)	–	–	1.1 (1)	0.9 (1)
lysine	6.9 (6)	7.4 (7)	4.4 (4)	1.8 (2)	4.6 (5)	2.0 (2)
arginine	8.5 (9)	3.6 (4)	2.8 (3)	1.8 (2)	1.1 (1)	4.9 (4)
tryptophan[e]	–	+ (1)	++ (3)	–	+ (1)	+ (5)
homoserine[f]	0.5 (1)	0.8 (1)	0.7 (1)	1.1 (1)	0.7 (1)	0.8 (1)
total	(118)	(72)	(63)	(34)	(40)	(53)
yield,[g] %	41	40	35	26	27	22

[a] Values given are for 24-h total acid hydrolysis. The values in parentheses are the number of residues determined by sequencing. [b] The Roman numeral designation pertains to the original G-50 pool from which the material originated (Figure 1). [c] CM-cysteine = carboxymethylcysteine. The number of residues was determined independently by the amount of radioactivity present in the peptide as a result of a alkylation with [^{14}C]iodoacetic acid. [d] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [e] Tryptophan presence detected qualitatively after acid hydrolysis as a peak on single-column amino acid analysis which appears in the region preceding arginine. [f] Homoserine values include homoserine lactone. [g] Yields are approximate and vary with the number of steps involved in the purification. [h] Even after repeated rechromatography CNII preparations contained more proline and alanine than could be accounted for by the sequence experiments. CNI is very rich in proline and alanine, however, and even a slight contamination could raise the CNII composition to the observed levels.

Table IV: Amino Acid Compositions of Ten Small CNBr Fragments from the β Chain of Human Fibrinogen

	residues/mole[a]									
	CNVB[b]	CNVIA	CNVIB1	CNVIB2	CNVIB3	CNVIC2	CNVIC1	CNVID	CNVIF	CNVII
CM-cysteine[c]				0.5 (1)						
aspartic acid[d]	1.1 (1)	–	1.0 (1)	2.4 (2)		1.9 (2)	1.2 (1)			0.9 (1)
threonine	0.9 (1)	1.0 (1)	–	–		0.9 (1)	1.1 (1)			
serine	–		1.1 (1)	2.0 (2)	1.1 (1)					1.9 (2)
glutamic acid[d]	1.1 (1)	2.5 (2)	1.1 (1)	1.2 (1)				2.0 (2)		
proline		1.1 (1)		1.9 (2)				2.1 (2)		
glycine	1.2 (1)	1.5 (1)	1.2 (1)			2.1 (2)	1.2 (1)			1.2 (1)
alanine			0.8 (1)			0.9 (1)				
valine				1.8 (2)		1.0 (2)				
isoleucine		1.0 (1)		0.9 (1)			0.6 (1)	0.6 (1)		
leucine		2.0 (2)	0.9 (1)	0.9 (1)						
tyrosine				2.2 (3)						0.8 (1)
phenylalanine								1.9 (2)		
histidine	0.3 (0)					0.9 (1)	1.0 (1)			
lysine				1.0 (1)		1.0 (1)		1.5 (1)	1.3 (1)	0.9 (1)
arginine	0.4 (1)			0.9 (1)				0.9 (1)	1.0 (1)	
tryptophan[e]						+ (1)				++ (2)
homoserine[f]	0.1 (1)	0.4 (1)	0.4 (1)	0.5 (1)	1.0 (1)	0.4 (1)	0.8 (1)		0.7 (1)	1.3 (1)
total	(6)	(9)	(7)	(18)	(2)	(12)	(6)	(9)	(3)	(9)
yield,[g] %	22	12	15	19	13	19	11	26	10	12

[a] Values given are for 24-h total acid hydrolysis. The values in parentheses are the number of residues determined by sequencing. [b] The Roman numeral part of the designation pertains to the G-50 pool from which the material was originally purified (Figure 1). [c] CM-cysteine = carboxymethylcysteine. The number of residues was also determined independently by the amount of radioactivity present in the peptide as a result of previous alkylation with [^{14}C]iodoacetic acid. [d] Aspartic and glutamic acid values include asparagine and glutamine, respectively. [e] Tryptophan presence detected qualitatively after acid hydrolysis as a peak on single-column analysis in the region preceding arginine. [f] Homoserine values include homoserine lactone. [g] Yields are approximate and vary with the number of steps involved in the purification.

In the following paragraphs brief descriptions and summaries are given of the data obtained in sequencing each of the CNBr fragments. Complete data, including amino acid compositions of all subpeptides and details of stepwise degradations, are available in the form of supplementary material. (See paragraph at end of paper regarding supplementary material.)

Cyanogen Bromide Fragments

CNI. The first material to emerge from the G-50 column was rechromatographed on Sephadex G-100; a symmetrical radioactive peak was well separated from a small amount of larger undigested material. Its amino acid composition was consistent with the 118-residue sequence reported by Blombäck et al. (1976), including the presence of three cysteine residues (Table II). The fragment had no detectable amino terminus either by the Dns procedure (Gray, 1972) or by thioacetylation (Mross & Doolittle, 1977), an anticipated observation because of the pyrrolidonecarboxylic acid known to exist at the amino terminus of the β chains. Digestion of CNI with trypsin followed by gel filtration on Sephadex G-50 and paper electrophoresis gave rise to the readily recognized fibrinopeptide B (14 residues) and eight other peptides, as well as free arginine and lysine. A peptide of about 23–24 residues and lacking arginine and lysine but containing homoserine was identified as the carboxy-terminal segment of the fragment. This peptide served as the starting material for an intensive reinvestigation of the carboxy-terminal sequence, since ambiguity has plagued that region since its original description (Blombäck & Blombäck, 1972). Accordingly, the tryptic peptide was further digested with staphylococcal protease (Houmard & Drapeau, 1972), and three peptides were readily separated by paper electrophoresis at pH 6.5. The largest of these contained 12 or 13 residues, including homoserine. The amino-terminal section of the peptide was twice sequenced by the Dns–PhNCS procedure (Gray, 1972). Also, on three separate occasions it was attached to glass beads by the homoserine method (Horn & Laursen, 1972) and degraded by the thioacetylation procedure. The data consistently yielded a sequence beginning Ala-Val-Ser-Gln-Thr. These results differ from the sequence reported by Blombäck et al. (1976) for this region, their corresponding order being Ala-Val-Gln(Thr,Ser). Unfortunately, in none of our several efforts were we able to sequence across the "polyserine bridge" to the carboxy-terminal tetrapeptide. Thus, Blombäck et al. (1976) reported a string of four serines after the two residues in parentheses (experimental details were not provided), but our stepwise degradation technology was unable to verify this section. It seemed especially important to establish that the peptide actually contained 13 residues (as based on the report by Blombäck et al., 1976), since the length of this peptide affects the count of residues thought to exist in the "coiled coils" interdomainal segment (Doolittle et al., 1977). As it stands now, the connection stretching between disulfide rings is exactly 111 residues in the α and γ chains but is 112 residues in the β chain if the peptide in question actually contains 13 residues. Accordingly, we undertook a series of amino acid analyses (18-, 24-, and 36-h hydrolysis times at 108 °C) in an attempt to verify the exact length of the peptide (Table V). Although the uncertainty in establishing the exact size of a peptide containing four to five serines cannot be overstated, our analyses indicate that the correct number may actually be four and not five. As a result, we have put the reported fifth serine (it would be residue β-114) in parentheses for the time being. Clearly a stepwise degradation reaching all the way across to the terminal residues will be necessary to establish the point with finality.

Verification of the Remainder of CNI. The remaining seven tryptic peptides (i.e., other than the well-characterized fibrinopeptide B and the carboxy-terminal peptide discussed in the preceding paragraphs) were all characterized with respect to amino acid analysis and amino-terminal identification. All were found to be in agreement with the sequence reported by Blombäck & Blombäck (1972). Moreover, a set of six sta-

[2] Altogether more than 2 g of β chain was used in these studies.

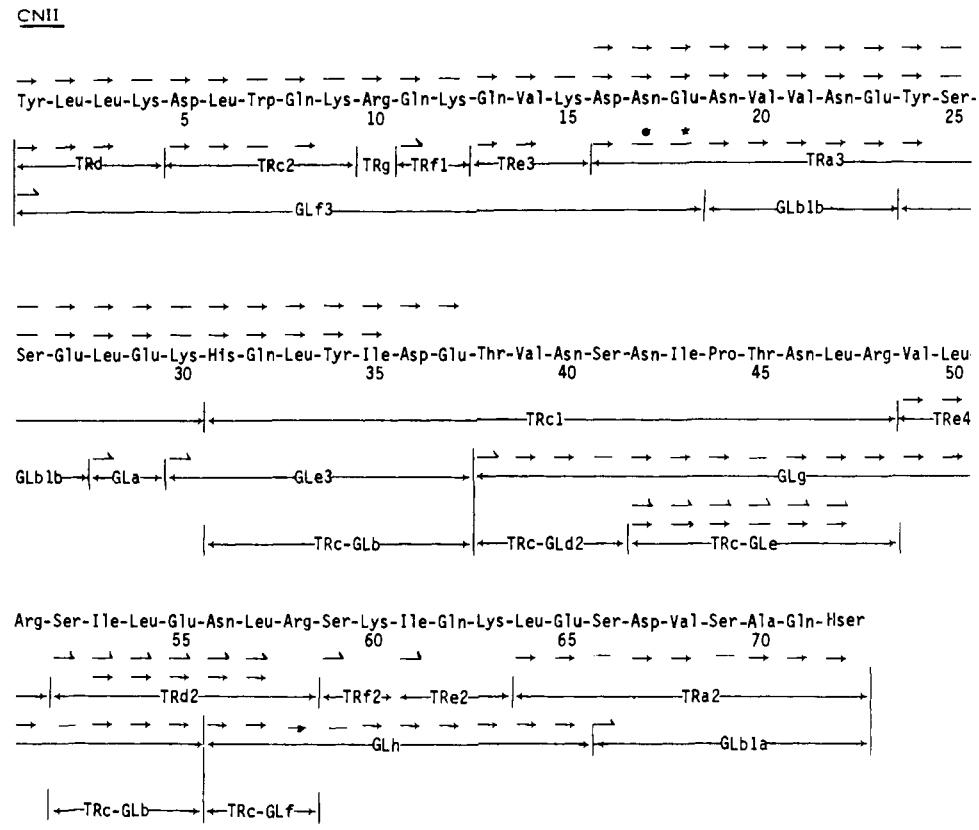

FIGURE 2: Summary of data used to establish sequence of fragment CNII. Full arrows = solid-phase sequencer run; half arrows = Dns identification. TR = trypsin; TH = thermolysin; GL = staphylococcal protease peptide. * = residues lost during workup.

phylococcal protease peptides from CNI was similarly characterized. Several of the latter were also attached to glass beads and degraded by the thioacetylation procedure.

Finally, CNI was isolated from fibrin (as opposed to fibrinogen) β chains. This material, which lacks the fibrinopeptide B and thus begins at residue 15, was attached to glass beads and subjected to thioacetylation stepwise degradation through 28 successful cycles. As a result of these studies, all but a few of the residues in CNI were independently verified.

CNII. The second pool from the G-50 column (Figure 1) was purified by rechromatography on the same G-50 column. The amino acid composition of the purified 72-residue fragment (now designated IIb) is listed in Table III, and a summary of the data used to obtain its sequence is depicted in Figure 2. The fragment was attached to amino glass beads by the lysine-coupling method (Laursen et al., 1972) and successfully degraded through 35 cycles in a solid-phase sequencer (Doolittle et al., 1977). Its amino terminus begins with the sequence Tyr-Leu-Leu (Table II). All the tryptic and staphylococcal protease peptides were isolated and characterized. The carboxy-terminal tryptic peptide was attached to beads by the homoserine method, and the lysine-containing peptides were attached by the lysine-coupling procedure. Other peptides were coupled to beads by the water-soluble carbodiimide procedure (Mross & Doolittle, 1977). Most of the peptides were also resequenced by the Dns–PhNCS procedure (Figure 2). The tryptic subpeptides of the staphylococcal protease peptides and vice versa were also isolated and characterized with regard to amino acid composition and amino terminals.

Fragment CNII includes the carboxy-terminal portion of the plasmin-derived fibrinogen fragment E and the amino-terminal portion of fragment D. Thus, a plasmin cleavage leading to the carboxy terminus of Eβ occurs at Lys-122 and another giving rise to the amino terminus of Dβ at Lys-133.

Table V: Amino Acid Composition of βCNITR2bGL6 Calculated for 12 and 13 Residues[a]

	hydrolysis time[b]				residues/mole of peptide[c]	
	18 h	24 h	30 h	mean	res/12	res/13
aspartic acid	5.9	6.9	5.7	6.2		
threonine	14.5	15.8	12.4	15.2	1.09	1.18
serine	56.1	56.4	60.9	57.7	4.12	4.49
glutamic acid	28.1	31.2	29.5	29.6	2.11	2.30
proline	3.6	4.2	3.5	3.9		
glycine	6.6	7.3	6.8	6.9		
alanine	16.0	17.5	15.8	16.4	1.17	1.28
valine	13.1	15.8	14.6	14.5	1.04	1.13
isoleucine	1.8	2.6	2.3	2.2		
leucine	3.0	3.5	3.5	3.3		
tyrosine	9.6	9.9	8.5	9.3	0.66	0.72
phenylalanine	10.7	11.9	12.0	11.5	0.82	0.89
histidine	–	–	–	–	–	–
lysine	2.1	2.7	3.1	2.6		
arginine	1.1	1.3	1.2	1.2		
homoserine[d]	10.2	11.0	10.6	10.6	0.75	0.82
					(12)	(13)

[a] The proposed sequence of βCNITR2bGL6 is A-V-S-Q-T-S-S-S(S)Q-F-Y-M. [b] Values given are in nanomoles and are the average of two determinations in each case. [c] Calculated on the basis of ratios of underlined values. [d] Sum of homoserine and homoserine lactone.

As such, data from both of these moieties were helpful in reconstructing the sequence. In particular, a solid-phase sequencer run on the β chain of fragment D (Doolittle et al., 1977b) yielded successful identifications through 21 degradative cycles for the segment corresponding to residues 16–37 of CNII.

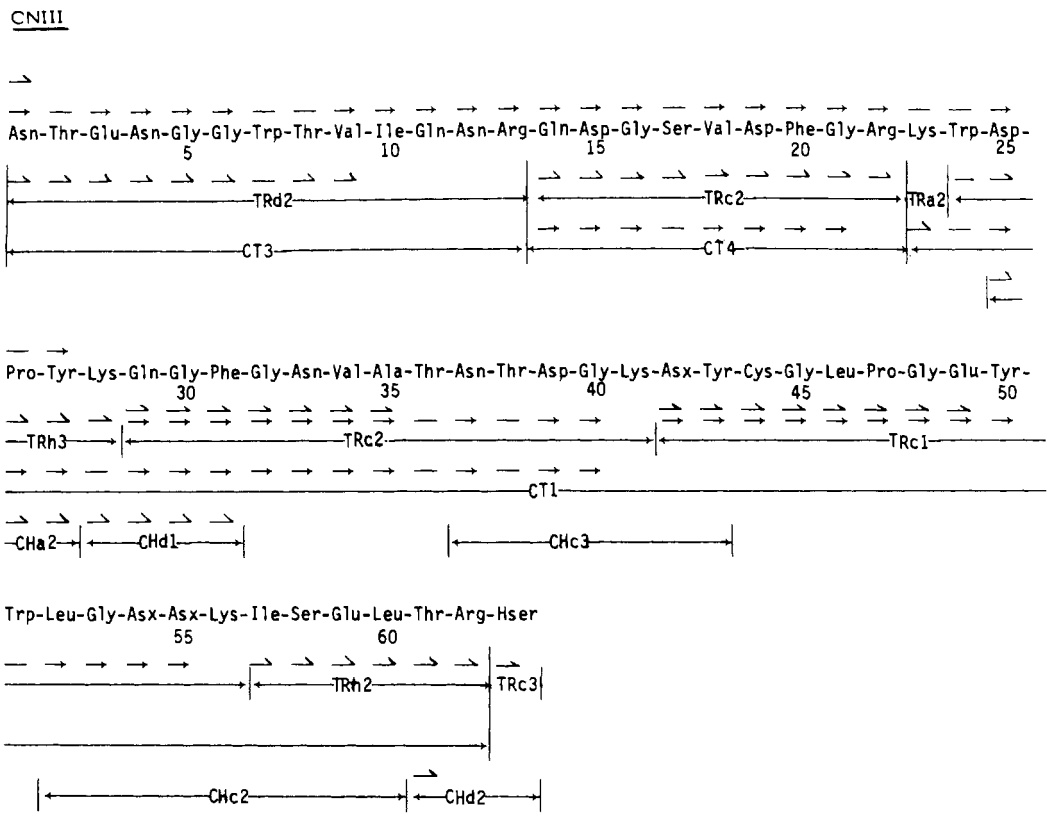

FIGURE 3: Summary of data used to establish sequence of fragment CNIII. See legend to Figure 2 for symbol designations; also, CH = chymotrypsin and CT = citraconyl-trypsin.

CNIII. Fragment CNIII contains 63 amino acid residues, including a single cysteine (Table III). The peptide was attached to beads by the lysine-coupling method and successfully degraded through 27 cycles. Digestion with trypsin gave rise to six subpeptides, as well as free lysine and free homoserine. The overlaps for those subpeptides not covered by the sequencer run were obtained by the isolation of another set of subpeptides after digestion with chymotrypsin (Figure 3). In addition, the fragment was citraconylated and subjected to limited tryptic digestion; three major peptides were isolated, two of which were attached to beads and subjected to stepwise degradation (Figure 3). Amides were established on the basis of the relative mobilities of the tryptic and chymotryptic subpeptides (Offerd, 1966).

CNIVA. The fourth pool prepared from the G-50 column (Figure 1) contained two CNBr fragments which were separable upon paper electrophoresis at pH 6.5, one moving toward the anode and the other slightly toward the cathode. The electronegative peptide, CNIVA, was found to contain 34 amino acid residues, including four cysteines (Table II). The peptide was attached to beads by the homoserine procedure and successfully degraded through 26 cycles. The positions of the four carboxymethylcysteines were confirmed by monitoring the release of radioactivity at each step. Digestion with trypsin yielded four principal subpeptides and free lysine, whereas digestion with staphylococcal protease gave rise to four subpeptides and free glutamic acid (Figure 4). Two of the tryptic peptides (TRb1 and TRd2) were attached to beads (the former by the lysine-coupling method and the latter by the homoserine procedure) and degraded by the thioacetylation procedure. In addition, one of the peptides obtained from the staphylococcal protease digestion was attached to beads by the water-soluble carbodiimide method and similarly cycled. The tryptic and staphylococcal protease peptides were also sequenced to various degrees by the

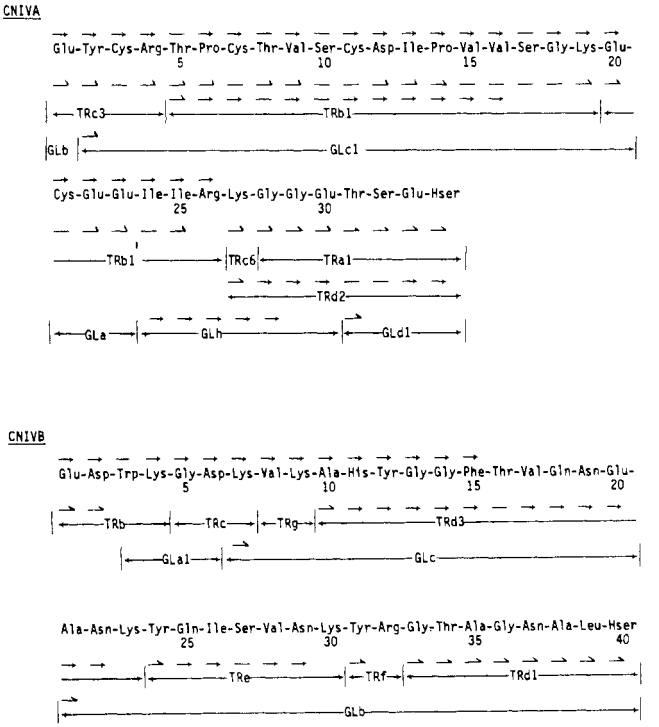

FIGURE 4: Summary of data used to establish sequences of fragments CNIVA (upper) and CNIVB (lower). See legends to Figures 2 and 3 for symbol designations.

Dns–PhNCS procedure (Figure 4). Amides were established on the basis of the electrophoretic mobilities of the tryptic and staphylococcal protease peptides, as well as on a consideration of the susceptibility to attack by the latter enzyme.

CNIVB. The more basic of the two peptides found in pool IV (Figure 1) was fluorescent and stained for the presence of

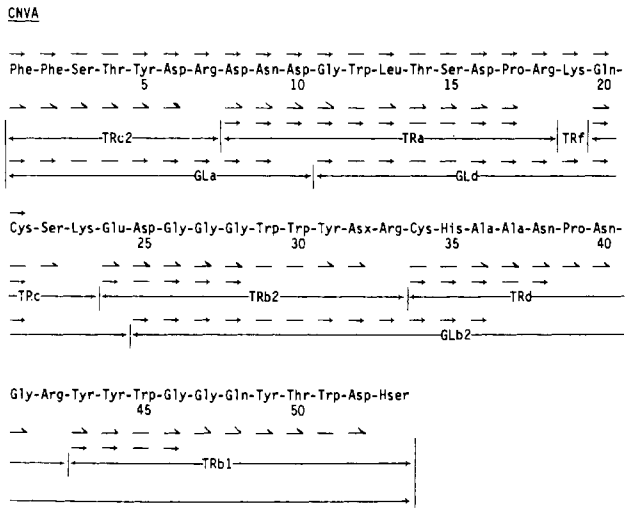

FIGURE 5: Summary of data used to establish sequence of fragment CNVA. See legends to Figures 2 and 3 for symbol designations.

histidine. Amino acid analysis indicated that the peptide contained 39 residues in addition to tryptophan. Subsequent characterization limited the number of tryptophans to one, the overall number of residues thus being 40 (Table III). The peptide was attached to beads by the homoserine method and carried through 15 successful steps. Tryptic digestion resulted in the isolation of seven subpeptides, the amino acid compositions of which added up to the overall fragment composition. The three largest of these were attached to beads and subjected to thioacetylation. In addition, a set of constituent subpeptides was isolated after digestion with staphylococcal protease (Figure 4). Finally, the fragment was citraconylated and then digested with trypsin. Two fragments were isolated, one arginine-containing peptide ($n = 32$) and an octapeptide-containing homoserine which corresponded to the carboxy terminus. The larger peptide was subsequently deblocked and retreated with trypsin, and six subpeptides were isolated and characterized (compare Figure 4). Amides were established on the basis of peptide mobilities and vulnerability to the staphylococcal protease.

CNVA. Fragment CNVA is a sparingly soluble peptide containing 53 residues, including five tryptophans and five tyrosines. The peptide also contains two cysteine residues (Table III). The peptide was attached to beads by the lysine-coupling procedure on four different occasions; the most successful of these runs carried through the 21st step, the carboxymethylcysteine residue at that position being identified by the radioactivity released. The fragment was digested with trypsin, and six subpeptides and free lysine were isolated. Digestion with staphylococcal protease produced three large peptides which overlapped the tryptic peptides. The tryptic and staphylococcal protease peptides were characterized in a variety of ways to establish the final sequence (Figure 5).

CNVB. Pool V (Figure 1) also contained another cyanogen bromide fragment, a hexapeptide with carbohydrate attached. Acid hydrolysis of the electrophoretically purified fragment revealed the presence of glucosamine. The peptide was sequenced both by thioacetylation after attachment to glass beads and by the Dns–PhNCS procedure (Figure 6). Glucosamine was found in the back-hydrolysate of residue 3 along with aspartic acid, indicating the precise position of attachment. The peptide also includes the sequence Asn-X-Thr, which is characteristic of glucosamine attachment points.

Amino acid analyses of the peptide always indicated the presence of approximately half a mole of histidine and less than a full mole of arginine, leading us to believe at first that polymorphism might exist in this peptide. Subsequently we found that the methionyl–threonine bond which junctions this peptide in the β chain does not fully cleave during the cyanogen bromide reaction, and the histidine (and some isoleucine) is evidently from the adjacent fragment (CNVIC1).

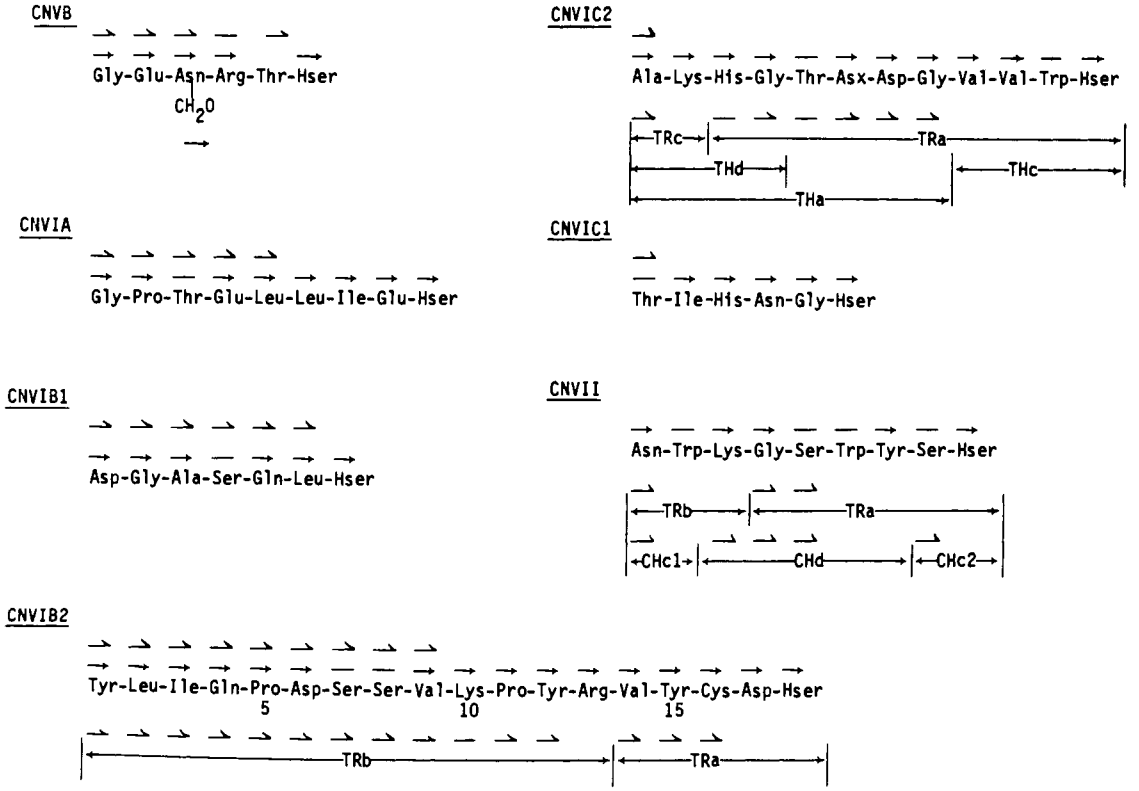

FIGURE 6: Summary of data used in establishing the sequences of seven small CNBr fragments. See legends to Figures 2 and 3 for symbol designations.

Table VI: Basis for Arrangement of CNBr Fragments

overlap no.	Met res	peptide[a]	res	sequence of overlap peptides
1	118	Eβ	43-122	from fragment E[b]
2	190	CTIVc1	177-194	S(K,I,Q,K,L,E,S,D,V,S,A,Q,M,E,Y,C)R
		GLf1	184-191	S-D-V(S)A-Q-M-E
3	224	CTV	217-237	K-G-G-E(T,S)E-M-Y-L-I-Q-P-D···
		GLg2b	224-230	M-Y-L-I-Q-P-D
4	242	CTVIb	238-255	V-Y-C-D-M-N(T)E-N-G(···)R
		GLf2	242-245	M-N-T-E
5	305	GLf5	302-309	L-T-R-M-G-P(T)E
6	314	CTIA	305-346	M-G-P(T)E-L-L-I-E-M-E-D(···)R
		GLe4	314-316	M-E-D
7/8	354/361	CTIIIb1	347-365	G(T)A-G-N-A(L,M,D,G,A,S,Q,L,M,G,E,N)R
		GLTRe1	347-363	G(T)A-G-N-A-L-M-D-G-A(S,Q,L,M,G)E
9/10	367/373	GLc	364-383	N-R(T)M(T)I-H-N-G-M-F(F,S,T,Y,D,R,D,N)D
		CNVb	352-373	G-E-N-R(T)M(T)I-H-B-G(M)
11	426	TRVII	416-428	Y-Y(W)G-G-Q-Y(T,W)D-M-A-K
		CTIb	416-448	Y-Y(W)G-G-Q-Y(···)R
12	438	–	–	homology with γ chain
13	447	TRVIe	442-448	G-S(W)Y-S(M)R
14/15	450/452	TRIIIh	449-453	K-M(S)M-K

[a] CT = citraconyl-trypsin; TR = trypsin without prior modification of peptide; GL = peptide from staphylococcal protease digestion.
[b] From Takagi & Doolittle (1975a).

Pool VI Peptides. Pool VI (Figure 1) contains six cyanogen bromide fragments, ranging in size from 6 to 18 residues (Table IV). The largest of these (CNVIB2) was attached to beads by the homoserine method and successfully cycled through all 18 residues, including the homoserine. The two serine residues were confirmed by a nine-cycle degradation using the Dns–PhNCS method (Figure 6). The two constituent tryptic subpeptides were also characterized by the latter procedure.

The characterization of three other peptides in this pool was routine (Figure 6), and the details used in sequencing another (CNVID) have been reported by us previously (Takagi & Doolittle, 1975a). In the case of the sixth peptide we encountered some difficulty, however. This peptide (CNVIC1 in Figure 6 and Table II) occurs in rather low yield and is difficult to purify. In our original procedure, whereby we fractionated the cyanogen bromide digest into 11 pools (Takagi & Doolittle, 1975a), we found this threonine-ending hexapeptide in pool X. It was separated from another pool X peptide—now designated in VIB1—by electrophoresis at pH 2. When the peptide did not reappear in our later (streamlined) fractionation procedure, it was forgotten and/or attributed to a contaminating substance, until the recent report of Henschen & Lottspeich (1977b). Following that report, we reisolated the peptide, attached it to beads, and successfully identified the remaining five residues by the thioacetylation procedure.

On the notion that one reason for the relatively low yield might be incomplete cleavage of the methionyl–threonine bond (Shroeder et al., 1969), especially since we had detected histidine (and some isoleucine) in (what was reported to be) the adjacent cyanogen bromide fragment (Henschen & Lottspeich, 1977b), we reexamined the fragment CNVB by the thioacetylation procedure, continuing the degradation through 12 cycles. As anticipated, isoleucine and histidine were revealed at positions 8 and 9. As it happens, the two peptides have very similar amino acid compositions, each having one aspartic acid (after hydrolysis), one threonine, one glycine, and one homoserine (Table IV); they differ in that one has glutamic acid and arginine while the other has isoleucine and histidine.

Pool VI also contained two small peptides which are not shown in Figure 6. One of these was the tripeptide Arg-Lys-Met. It was readily sequenced by attaching it to beads by the homoserine method and degrading it by the thioacetylation procedure. The other was the neutral dipeptide Ser-Met (CNVIB3). The amino-terminal serine was determined by the Dns method. The peptide occurs in low yield, a fact which may also be related to the difficulty of cleaving homoseryl-hydroxy amino acid linkages (Schroeder et al., 1969). In fact, we originally isolated a basic peptide which contained equimolar amounts of serine, arginine, and lysine, as well as homoserine, in a yield greater than the dipeptide (Watt et al., 1978). The isolation of an incongruous overlapping tryptic peptide (TRVIE), combined with the report of Henschen & Lottspeich (1977b), led to a reexamination and the conclusion that Ser-Met is the true peptide.

Pool VII Peptides. The seventh pool from the G-50 column (Figure 1) contained a single cyanogen bromide fragment, a fluorescent nonapeptide (CNVII in Figure 6). The valyl–valine bond is very slow to hydrolyze. It was attached to beads by the homoserine method and sequenced by thioacetylation; the tryptic and chymotryptic subpeptides were also characterized (Figure 6).

Overlapping Sequences

For the most part, the arrangement of the 16 CNBr fragments was established by the isolation of key methionine-containing peptides from various enzymic digests of β chains, although the positions of the amino-terminal and carboxy-terminal fragments were self-evident. In this regard, batches of β chains were digested with staphylococcal protease, trypsin, and trypsin after the chains were first modified by citraconylation (Attasi & Habeeb, 1972). Altogether, 13 of the 15 methionines were accounted for in such overlapping peptides (Table VI). Most of the peptides were characterized by attachment to beads and subsequent stepwise degradation of thioacetylation; some others were characterized by the Dns–PhNCS procedure. The sequenced residues are listed in Table VI.

Another of the junctions between CNBr fragments was known from previous experiments on fragment E (Takagi & Doolittle, 1975b). Similarly, we have previously presented the case for the carboxy-terminal CNBr fragment's not only being homologous with the carboxy-terminal fragment of the γ chain (Takagi & Doolittle, 1975a) but also being the only fragment not containing homoserine. The arrangement of the two remaining peptides (CNVIC2 and CNVII), which involves

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

FIGURE 7: Complete amino acid sequence of the β chain of human fibrinogen. The chain contains 461 residues; a carbohydrate site is located at Asn-364.

a single junction (Met-438), was based on the strong homology with the corresponding γ-chain segments (Table VI).

Asparagines and Glutamines. The methodology we employed did not allow the unambiguous assignment of asparagine/asparatic acid and glutamine/glutamic acid in every case, and in two cases we note different results from those reported by Henschen & Lottspeich (1977b). Thus, we report aspartic acid at position β-202, whereas those authors find asparagine, and, similarly, we find glutamic acid at β-301 where they report glutamine. In both cases our identification is based on cleavages with the staphylococcal protease, which has a specificity known to be directed to side-chain free carboxylates (Houmard & Drapeau, 1972). On the other hand, we cannot exclude the possibility that the conflicting results are due to deamidation during peptide isolation. In five other cases—all involving asparagine/aspartic acid—we were unable to make a firm identification based either on peptide mobilities or on vulnerability to the staphylococcal protease.

Discussion

The completion of the amino acid sequence of the β chain represents another milestone on the way to establishing the complete structure of the human fibrinogen molecule. Already a number of important structure–function aspects can be appreciated. First, our early notions about the common ancestry of the nonidentical chains of fibrinogen (Doolittle, 1970, 1973; Takagi & Doolittle, 1975a) have been completely substantiated. Interestingly, homology among all three chains is evident in the amino-terminal thirds of the chains, the percent identity between any two of the chains running about 15%. On the other hand, there is little or no homology evident between the α chain, on the one hand, and the β and γ chains, on the other, when the carboxy-terminal two-thirds are compared. A comparison of the carboxy-terminal 142 residues of the α chain (Cottrell & Doolittle, 1978) with the corresponding regions of the β and γ chains failed to show any further relationship than was evident from a comparison of the carboxy-terminal CNBr fragments (Cottrell & Doolittle, 1976). The homology between the β and γ chains is especially strong in this region, however, the two chains being about 35% identical.

The strong resemblance between β and γ chains in their carboxy-terminal two-thirds—the portions which comprise fragment D—extends to the arrangement of their disulfide bonds. We have recently reported the pairing of the disulfide bonds in fragment D (Bouma et al., 1978) and noted that the situations in the β and γ chains are exactly equivalent. We are struck with the rather curious situation of a threefold symmetry, extending through the portions of the molecule which connect the central and terminal domains (Doolittle et al., 1978), and a projected twofold symmetry in the terminal domains per se, where the β and γ chains likely have equivalent three-dimensional structures.

According to the model we currently favor (Figure 7), fibrinogen comprises two terminal domains connected by three-stranded ropes to a smaller central domain. As such, the β chains can be divided into three sectors for discussion purposes: those 75 residues at the amino terminus which are thought to exist in the central region of the molecule, the next 122 residues which include two sets of disulfide rings delineating a 112-residue coiled-coil connector (Doolittle et al., 1978), and the remaining 264 residues which reside in each of the distal (terminal) domains.

The distribution of amino acids in the three different sectors is distinctive. Thus, 11 of the 12 β-chain tryptophans occur in the distal region. Interestingly, nine of these have exact counterparts at the equivalent positions in the γ chain. In contrast, 10 of the 23 prolines in the β chain and 11 of the 23 alanines occur among the first 75 residues which are a part of the central domain. The most predominant amino acid in the β chain is glycine (42 of 461 residues), and yet there are no glycines at all in the 112-residue connector, consistent with the α-helical nature of the region (Doolittle et al., 1978). Charge distribution seems to be relatively equitable throughout the chain, however, the sum of glutamic and aspartic acid residues in each sector being roughly equal to the sum of arginines and lysines.

In summary, the amino acid sequence of the β chain of human fibrinogen has been determined. Its most interesting

features are its extensive homology with the γ chain—and to a lesser extent with the α chain—and a nonrandom distribution of amino acids, the nature of which is consistent with a fibrinogen molecule in which two compact terminal domains are attached to a central region by coiled-coil connectors.

Acknowledgments

We are pleased to acknowledge the cooperation and assistance of Barbara Cottrell, Danielle Goldbaum, Marcia Riley, and Dennis Trovato.

Supplementary Material Available

Tables giving (a) amino acid compositions of peptides obtained by enzymatic digestion of β chains and CNBr fragments, (b) details of materials employed in solid-phase thioacetylation stepwise degradations, (c) elution profiles for column purification of various CNBr fragments and some key enzymatically derived peptides, and (d) residue by residue basis for sequence assignment (36 pages). Ordering information is given on any current masthead page.

References

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

Blombäck, B., & Blombäck, M. (1972) *Ann. N.Y. Acad. Sci. 202*, 77.

Blombäck, B., Hessel, B., & Hogg, D. (1976) *Thromb. Res. 8*, 639.

Bouma, H., III, Takagi, T., & Doolittle, R. F. (1978) *Thromb. Res. 13*, 557.

Cartwright, T., & Kekwick, R. G. O. (1971) *Biochim. Biophys. Acta 236*, 550.

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

Doolittle, R. F. (1970) *Thromb. Diath. Haemorrh., Suppl. 39*, 25.

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., Takagi, T., Watt, K., Bouma, H., Cottrell, B. A., Cassman, K. G., Goldbaum, D. M., Doolittle, L. R., & Friezner, S. J. (1977c) *Proc. 11th FEBS Meet. 47*, 163.

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

Gaffney, P. J. (1972) *Biochim. Biophys. Acta 263*, 453.

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

Henschen, A., & Edman, P. (1972) *Biochim. Biophys. Acta 263*, 351.

Henschen, A., & Lottspeich, F. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 357*, 1801.

Henschen, A., & Lottspeich, F. (1977a) *Thromb. Res. 11*, 869.

Henschen, A., & Lottspeich, F. (1977b) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1643.

Horn, M. J., & Laursen, R. A. (1972) *FEBS Lett. 36*, 285.

Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506.

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

Lottspeich, R., & Henschen, A. (1977a) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1521.

Lottspeich, R., & Henschen, A. (1977b) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1639.

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

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

Offerd, R. E. (1966) *Nature (London) 211*, 591.

Pizzo, S. V., Schwartz, M. L., Hill, R. L., & McKee, P. A. (1972) *J. Biol. Chem. 247*, 636.

Schroeder, W. A., Shelton, J. B., & Shelton, J. R. (1969) *Arch. Biochem. Biophys. 130*, 551.

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

Takagi, T., & Doolittle, R. F. (1975a) *Biochim. Biophys. Acta 386*, 617.

Takagi, T., & Doolittle, R. F. (1975b) *Thromb. Res. 7*, 813.

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

Töpfer-Peterson, E., Lottspeich, R., & Henschen, A. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 375*, 1509.

Watt, K., Goldbaum, D., Cottrell, B. A., & Doolittle, R. F. (1977) *Throm. Diath. Haemorrh. 38*, 26 (abstract).

Watt, K. W. K., Takagi, T., & Doolittle, R. F. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1731.