Delbarre, A., Roques, B. P., & Le Pecq, J. B. (1977) *C. R. Hebd. Seances Acad. Sci., Ser. D 284*, 81–84.
Delbarre, A., Gaugain, B., Markovits, J., Vilar, A., Le Pecq, J. B., & Roques, B. P. (1981) in *Intermolecular Forces* (Pullman, B., Ed.) pp 273–283, D. Reidel Publishing Co., Dordrecht, Holland.
Gaugain, B., Barbet, J., Oberlin, R., Roques, B. P., & Le Pecq, J. B. (1978a) *Biochemistry 17*, 5071–5077.
Gaugain, B., Barbet, J., Capelle, N., Roques, B. P., & Le Pecq, J. B. (1978b) *Biochemistry 17*, 5078–5088.
Le Pecq, J. B., Le Bret, M., Barbet, J., & Roques, B. P. (1975) *Proc. Natl. Acad. Sci. U.S.A. 71*, 5078–5082.
Mandal, C., Kallenbach, N. R., & Englander, S. W. (1979) *J. Mol. Biol. 135*, 391–411.
Mandal, C., Englander, S. W., & Kallenbach, N. R. (1980) *Biochemistry 19*, 5819–5825.
Markovits, J., Roques, B. P., & Le Pecq, J. B. (1979) *Anal. Biochem. 94*, 259–264.
Markovits, J., Blanquet, S., Dessen, P., Roques, B. P., & Le Pecq, J. B. (1981a) *Biochem. Pharmacol. 30*, 1557–1562.
Markovits, J., Gaugain, B., Roques, B. P., & Le Pecq, J. B. (1981b) in *Intermolecular Forces* (Pullman, B., Ed.) pp 285–298, D. Reidel Publishing Co., Dordrecht, Holland.
Marquardt, D. W. (1963) *J. Soc. Ind. Appl. Math. 11*, 431–441.
McConnell, B., & Von Hippel, P. H. (1970) *J. Mol. Biol. 50*, 297–316.
Muller, W., & Crothers, D. M. (1975) *Eur. J. Biochem. 54*, 267–277.
Nakanishi, M., Kobayashi, M., Tsuboi, M., Takasaki, C., & Tamiya, N. (1980) *Biochemistry 19*, 3204–3208.
Olmsted, J., & Kearns, D. R. (1977) *Biochemistry 16*, 3647–3654.
Pelaprat, D., Delbarre, A., Le Guen, I., Le Pecq, J. P., & Roques, B. P. (1980) *J. Med. Chem. 23*, 1336–1342.
Record, M. T., Jr., Lohman, T. M., & De Hasseth, P. (1976) *J. Mol. Biol. 107*, 145–158.
Reinhardt, C., Roques, B. P., & Le Pecq, J. B. (1982) *Biochem. Biophys. Res. Commun. 104*, 1376–1385.
Roques, B. P., Barbet, J., Oberlin, R., & Le Pecq, J. B. (1976) *C. R. Hebd. Seances Acad. Sci., Ser. D 283*, 1365–1368.
Roques, B. P., Pelaprat, D., Le Guen, I., Porcher, G., Gosse, Ch., & Le Pecq, J. B. (1979) *Biochem. Pharmacol. 28*, 1811–1815.
Saucier, J. M., Festy, B., & Le Pecq, J. B. (1971) *Biochimie 53*, 973–980.
Schreier, A. A., & Baldwin, R. L. (1976) *J. Mol. Biol. 105*, 409–426.
Stryer, L. (1966) *J. Am. Chem. Soc. 88*, 5708–5712.
Teitelbaum, H., & Englander, S. W. (1975a) *J. Mol. Biol. 92*, 55–78.
Teitelbaum, H., & Englander, S. W. (1975b) *J. Mol. Biol. 92*, 79–92.
Wells, R. D., Larson, J. E., Grand, R. C., Shortle, B. E., & Cantor, C. R. (1970) *J. Mol. Biol. 54*, 465–497.
Woodward, C. K., & Hilton, B. D. (1979) *Annu. Rev. Biophys. Bioeng. 8*, 99–127.

Characterization of a Complementary Deoxyribonucleic Acid Coding for the α Chain of Human Fibrinogen†

Mark W. Rixon, Wai-Yee Chan,‡ Earl W. Davie,* and Dominic W. Chung

ABSTRACT: A human liver cDNA library was screened for the α chain of fibrinogen with a cDNA clone from the corresponding bovine molecule as a hybridization probe. Several human clones coding for the α chain were identified, and one of these was used to rescreen the entire cDNA library of 18 000 recombinants. Plasmids with the largest cDNAs were isolated, and their inserts were sequenced. The largest cDNA insert contained 2224 base pairs, including a noncoding region at the 5′-end followed by a region coding for a signal peptide of 19 (or 16) amino acids and a mature protein of 625 amino acids, a stop codon of TAG, another noncoding region, and a poly(A) tail at the 3′-end. Eight tandem repeats of 39 base pairs were observed starting with nucleotide 905 (amino acid residue 270) and ending with nucleotide 1213 (amino acid residue 372). The identity in the nucleotide sequence in the tandem repeats ranged from 72 to 95% when compared to a consensus sequence. The predicted amino acid sequence for the mature polypeptide chain was 15 amino acids longer at the carboxyl-terminal end than that of the α chain isolated and sequenced from plasma fibrinogen. This indicates that minor proteolysis has taken place on the carboxyl-terminal end of the α chains, and this modification has probably occurred during secretion or circulation of the protein in plasma.

Fibrinogen[1] (M_r 340 000) is a plasma protein that participates in the final phase of blood coagulation (Marder et al., 1982). Each molecule consists of two sets of three different polypeptide chains designated α, β and γ, with molecular weights of 66 000, 52 000, and 46 500, respectively (McKee et al., 1966). Fibrinogen is a glycoprotein containing four carbohydrate chains, including one on each of the β chains and one on each of the γ chains (Iwanaga et al., 1968; Töpfer-Petersen et al., 1976). The α chain of fibrinogen is free of carbohydrate (Pizzo et al., 1972). The three pairs of chains in fibrinogen are held together by disulfide bonds and

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received February 14, 1983.* This work was supported in part by Grants HL 16919 and HL 28598 from the National Institutes of Health. D.W.C. is an Established Investigator of the American Heart Association.
‡ Present address: Department of Pediatrics, University of Oklahoma, Oklahoma City, OK 73190.

[1] The three chains in fibrinogen (factor I) have also been called Aα, Bβ, and γ (Blombäck, 1969). Plasmids were labeled as follows: p, plasmid; H, human; B, bovine; I, fibrinogen; α, α chain; 1, first plasmid identified.

0006-2960/83/0422-3237$01.50/0

form a trinodular structure linked by long slender filaments (Hall & Slater, 1959; Fowler & Erickson, 1979). The complete amino acid sequences of the three chains of human fibrinogen have been established by amino acid sequencing techniques (Henschen & Lottspeich, 1977; Lottspeich & Henschen, 1977; Doolittle et al., 1979; Henschen et al., 1979; Watt et al., 1979). The α chain contains 610 amino acids ending with a carboxyl-terminal valine. However, fibrinogens with different solubilities and chromatographic properties have been isolated and characterized (Mosesson et al., 1967), and proteolytic cleavage of the carboxyl terminus of the α chain has been proposed to account for some of these differences (Mosesson et al., 1972; Finlayson et al., 1972).

During the coagulation process, thrombin removes the amino-terminal fibrinopeptides A and B from both α and β chains by minor proteolysis, converting fibrinogen to fibrin monomers. The fibrin monomers then polymerize, leading to the formation of an insoluble fibrin clot (Doolittle, 1973, 1975).

The three chains of fibrinogen are synthesized in liver hepatic parenchymal cells (Forman & Barnhart, 1964) from three individual mRNA species (Nickerson & Fuller, 1981; Chung et al., 1980). They are processed, glycosylated, assembled, and eventually secreted into the circulating plasma as the mature fibrinogen molecule. cDNA clones for the α and β chains of bovine fibrinogen (Chung et al., 1981, 1982) and the α, β and γ chains of rat fibrinogen have been reported (Crabtree & Kant, 1981, 1982). The cDNA clone isolated for the α chain of bovine fibrinogen coded for the carboxyl-terminal 202 amino acids of the polypeptide chain (Chung et al., 1982). This report describes the isolation and characterization of cDNAs for the α chain of human fibrinogen, employing the bovine cDNA clone as a probe. The following two papers describe the properties of cDNAs coding for the β and γ chains of human fibrinogen, as well as the gene for the β chain of human fibrinogen (Chung et al., 1983a,b).

Experimental Procedures

Materials. DNA restriction endonucleases were purchased from Bethesda Research Laboratories, New England Biolabs, or Amersham. ^{32}P-Labeled nucleotides were obtained from New England Nuclear. Nitrocellulose (Schleicher and Schuell, Keene, NH) and Whatman 541 filters (Whatman Ltd., England) were used as hybridization solid supports.

Preparation of Hybridization Probes. Initial screening of the human cDNA library was carried out with a fragment of a bovine cDNA (pBIα1) that coded for the carboxyl-terminal 202 residues of the α chain of fibrinogen, 251 base pairs of noncoding nucleotides, and a poly(A) tail (Chung et al., 1982). The bovine cDNA fragment was prepared by digestion of pBIα1 with *Alu*I and *Taq*I restriction endonucleases and fractionation by electrophoresis in a 3.5% polyacrylamide gel (Maniatis et al., 1975a). DNA bands were visualized by either ethidium bromide staining or UV light shadowing on a fluorescent background. A band containing 161 base pairs corresponding to sequences coding for amino acid residues 422–477 in the α chain of human fibrinogen was cut out of the gel and the DNA recovered by electroelution (McDonnell et al., 1977). The human cDNA probe was prepared from plasmid pHIα2 by digestion with *Taq*I, and a fragment was isolated that coded for amino acid residues 162–252 in the α chain of human fibrinogen.

The cDNA probes were labeled to high specific radioactivity by the method of Maniatis et al. (1975b). The conditions were modified to include 40 μM calcium in addition to magnesium.

Screening of Recombinant Plasmids. A cDNA library prepared from human liver was kindly provided by Drs. S. L. C. Woo and T. Chandra of Baylor College of Medicine. It contained approximately 18 000 recombinants with cDNA inserts in the *Pst*I site of plasmid pBR322 (Chandra et al., 1983). The library was initially screened for the α chain of fibrinogen with the bovine cDNA fragment from pBIα1 and then with the human cDNA fragment from pHIα2. Individual recombinants were stored in 50% glycerol in microtiter dishes at −70 °C. Prior to screening, they were plated directly onto nitrocellulose filters, and the filters were then layered on L-agar containing tetracycline (12.5 μg/mL). Colonies were lysed by soaking the filters in 0.5 M sodium hydroxide, and the DNA was immobilized onto the filters by the method of Grunstein & Hogness (1975). Hybridization of the filters with nick-translated cDNA was performed as described (Southern, 1975), except in the cross-species hybridization where 0.09 M citrate–0.9 M sodium chloride, pH 7.0 (6× SSC), was used, and the prehybridization, hybridization, and washing were all carried out at 58 °C. In many experiments, Whatman 541 filters were employed rather than nitrocellulose since they were less expensive, could be used in multiple screenings, and were much easier to handle in large quantities (Gergen et al., 1979; Wallace et al., 1981).

DNA Sequence Determinations. DNA fragments were labeled at the 5′-end with [γ-^{32}P]ATP by T4 polynucleotide kinase (Richardson, 1965; Maxam & Gilbert, 1980). Labeling at the 3′-ends was performed with either fill-in synthesis using the Klenow fragment of *Escherichia coli* polymerase I in the presence of [α-^{32}P]dNTPs (Smith et al., 1979) or with cordycepin [α-^{32}P]triphosphate in the presence of terminal deoxynucleotide transferase and cobalt (Tu & Cohen, 1980). Labeled DNA was separated from unreacted nucleotide triphosphates by gel-filtration chromatography employing Ultrogel AcA54 (LKB). The labeled DNA was digested with an appropriate restriction endonuclease for end separation and fractionated by electrophoresis in polyacrylamide gel. Base-specific chemical modifications were carried out according to the method of Maxam & Gilbert (1980) with the following modifications: the amount of dimethyl sulfate was reduced to half in the G-specific reaction and the G+A reaction was performed at 37 instead of 0 °C. DNA sequence was recorded and edited with the computer programs of Staden (1977). The TMATRIX computer program for the dot–matrix analysis was written and kindly provided by Drs. W. Barnes and M. Zyda, Department of Biological Chemistry, Washington University School of Medicine, St. Louis, MO 63110.

Containment. Experiments were performed in compliance with NIH Guidelines for Recombinant DNA Research.

Results

Approximately 4500 colonies from a human liver cDNA collection of 18 000 colonies were screened by the colony hybridization technique of Grunstein & Hogness (1975) employing a cDNA probe coding for the α chain of bovine fibrinogen (Chung et al., 1982). Fifty-seven positive colonies were identified on the basis of hybridization to the radiolabeled bovine cDNA probe. Small-scale preparations of plasmid DNA (Birnboim & Doly, 1979) for a number of these clones, followed by digestion with restriction endonuclease *Pst*I to release the cDNA insert, revealed that one of the plasmids, designated pHIα1, was approximately 1600 base pairs in length. A smaller cDNA (pHIα2), approximately 1000 base pairs in length, was also identified and shown by restriction enzyme mapping to contain additional nucleotides on the 5′-end of the insert (Figure 1). DNA sequencing by the method of Maxam & Gilbert (1980) demonstrated that plasmid pHIα1 contained 1571 base pairs of DNA coding for

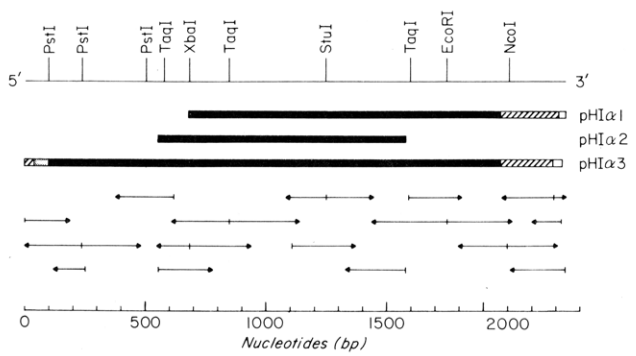

FIGURE 1: Restriction map and sequencing strategy for three cDNAs for the α chain of human fibrinogen. The bars below the restriction map represent the three clones (pHIα1, pHIα2, and pHIα3). The 5′- and 3′-noncoding regions are represented by slashed bars, the leader sequence is represented by a dotted bar, the coding regions are represented by solid bars, and the poly(A) tails are represented by open bars; the extent of sequencing is indicated by the length of each arrow, and the site of labeling is represented by the vertical lines.

amino acid residue 195 to the carboxyl-terminal end of the protein, in addition to a stop codon, 234 nucleotides of noncoding sequence, and a poly(A) tail. Plasmid pHIα2 contained 1017 base pairs of DNA coding for amino acid residues 153–491. From these data, it was estimated that 1.3% of the human liver cDNA library contained inserts coding for the α chain of fibrinogen.

To identify a clone with a larger cDNA insert, the *Taq*I–*Taq*I fragment from plasmid pHIα2 was prepared and used as a hybridization probe. This fragment contained DNA coding for amino acid residues 162–252 and was used to screen the entire cDNA library of 18 000 colonies on Whatman 541 filters for the colony hybridization. Forty-two additional positive colonies were identified by this screening. DNA was then isolated on a small scale from each of these plasmids and subjected to restriction enzyme digestion. The plasmid with the longest cDNA insert (pHIα3) was sequenced and shown to contain a cDNA coding for the α chain of human fibrinogen (Figure 2). The predicted amino acid sequence for this cDNA is also shown in Figure 2.

The cDNA insert of plasmid pHIα3 contained 2224 base pairs, including 11 G's and 12 C's on the 5′- and 3′-ends, respectively. A stop codon of TAG was found 84 base pairs upstream from the amino-terminal alanine present in the α chain of the mature protein. This places the potential initiator methionine codon at position −19 or −16 in the putative signal sequence. The rest of the amino acids in the leader sequence of the α chain share the characteristics of a typical signal sequence, including a hydrophobic core, a positively charged residue (arginine at position −15), and an uncharged residue (threonine at position −1) at the cleavage site for the signal peptidase (Blobel et al., 1979). The mature polypeptide was coded by 1875 base pairs corresponding to 625 amino acids followed by a stop codon of TAG. There were also 217 base pairs of 3′-noncoding in this plasmid and an apparent polyadenylation signal (AATAAA) 22 base pairs upstream from the poly(A) tail (Proudfoot & Brownlee, 1976).

The amino acid composition of the α chain of human fibrinogen predicted from the cDNA was as follows: Asp_{35}, Asn_{29}, Thr_{48}, Ser_{89}, Glu_{44}, Gln_{18}, Pro_{38}, Gly_{71}, Ala_{23}, Val_{28}, Met_{10}, Ile_{18}, Leu_{31}, Tyr_9, Phe_{19}, Lys_{40}, His_{16}, Arg_{41}, $^1/_2$-Cys_8, and Trp_{10}. This corresponds to a molecular weight of 67 625. Two amino acid sequences that are potential sites for carbohydrate side chains are present in the mature molecule. These two sequences (Asn-X-Ser) are present at asparagine residues 269 and 400, but neither is attached to carbohydrate in the mature molecule (Henschen et al., 1980a).

Table I: Differences among Amino Acid Sequences of the α Chain of Human Fibrinogen As Determined by Nucleic Acid and Protein Sequencing Techniques

amino acid residue	cDNA sequence	protein sequence 1[a]	protein sequence 2[b]
128	Glu	Gln	Glu
177	Asp	Asn	Asp
196–197	Ser–Arg	Ser–Arg	Arg–Ser
280	Ser	Ser	(Gly)[c]
285	Ser	Ser	(Gly)[c]
298–299	Gly–Thr	Ser–Gly	Gly–Thr
312	Ala	Thr	Thr
388	Asn	Asp	Asn
390	Asp	Asn	Asp

[a] Data from Doolittle et al. (1979). [b] Data from Henschen et al. (1980a). [c] Tentative assignment.

When the amino acid sequence as predicted from the cDNA was compared to that determined by protein sequencing techniques, some differences were observed (Table I). Most apparent was the fact that the nucleotide sequence predicted a polypeptide chain of 625 amino acids with a carboxyl-terminal proline, while the α chain isolated from plasma fibrinogen ends with a valine at amino acid residue 610 (Doolittle et al., 1979; Henschen et al., 1980a). The DNA sequence coding for the mature polypeptide was identical in plasmids pHIα1 and pHIα3, suggesting that an error by reverse transcriptase at a stop codon was unlikely during the construction of the cDNA library. Furthermore, these results are analogous to those observed for the α chain of bovine fibrinogen (Chung et al., 1982). In this case, there was also an additional 15 amino acids on the carboxyl end of the α chain.

The cDNA sequences of pHIα1 and pHIα3 were slightly different at their 3′-terminal ends. While the majority of the α chain cDNAs that were isolated contained only one internal *Eco*RI site at nucleotide 1745 (Figure 1), pHIα1 contained two *Eco*RI sites, the second being located at the extreme 3′-end of the cDNA insert. Sequencing of the cDNA insert for pHIα1 demonstrated that the poly(A) tract was located 17 base pairs downstream from the poly(A) tract in pHIα3 (Figure 3). This suggests that the DNA-dependent RNA polymerase may recognize more than one termination site, giving rise to mRNAs of different lengths on their 3′-ends. Alternatively, some nuclease activity or processing event at the 3′-end may result in some variation in the precise site of polyadenylation. Similar variations in the 3′-regions of the cDNAs coding for the β chains of human fibrinogen have also been observed (Chung et al., 1983a).

Watt and co-workers (Watt et al., 1979) reported the presence of 10 tandem repeats of 13 amino acids between residues 264 and 391 in the α chain of human fibrinogen. They postulated that a DNA duplication event was responsible for their origin. When the cDNA nucleotide sequence was determined, it was possible to make a detailed internal comparison of the cDNA by using computer-assisted reiterated searches. These results showed that there are eight well-defined tandem repeats of 39 base pairs in the α chain cDNA (Table II). These repeats start with nucleotide 905 (amino acid residue 270) and extend to nucleotide 1213 (amino acid residue 372). The degree of identity for each of the repeats ranged from 72–95% for the DNA and 54–92% for the amino acids when these sequences were compared to a consensus sequence.

Discussion

A cDNA (pHIα3) for the α chain of human fibrinogen has

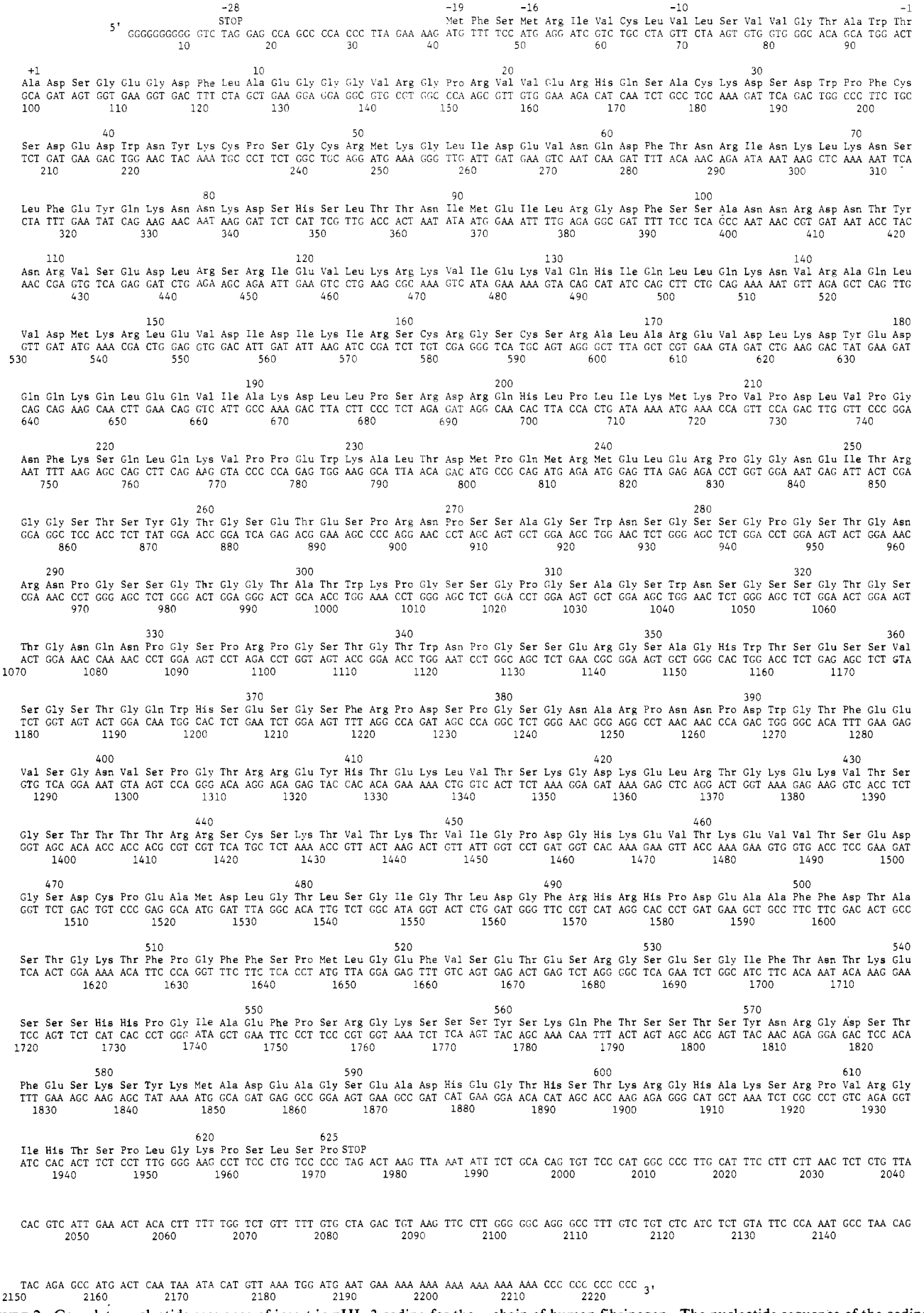

FIGURE 2: Complete nucleotide sequence of insert in pHIα3 coding for the α chain of human fibrinogen. The nucleotide sequence of the coding strand and the predicted amino acid sequence are also shown. Residues −1 to −19 (or −16) include the signal sequence, and residues 1–625 include the amino acids present in the mature α chain.

Table II: Amino Acid and Nucleotide Sequences for Tandem Repeats in the α Chain of Human Fibrinogen[a]

tandem repeat no.														homology (%)
	270													
1	Pro	Ser	Ser	Ala	Gly	Ser	Trp	Asn	Ser	Gly	Ser	Ser	Gly	77
	CCT	AGC	AGT	GCT	GGA	AGC	TGG	AAC	TCT	GGG	AGC	TCT	GGA	92
	283													
2	Pro	Gly	Ser	Thr	Gly	Asn	Arg	Asn	Pro	Gly	Ser	Ser	Gly	92
	CCT	GGA	AGT	ACT	GGA	AAC	CGA	AAC	CCT	GGG	AGC	TCT	GGG	95
	296													
3	Thr	Gly	Gly	Thr	Ala	Thr	Trp	Lys	Pro	Gly	Ser	Ser	Gly	62
	ACT	GGA	GGG	ACT	GCA	ACC	TGG	AAA	CCT	GGG	AGC	TCT	GGA	85
	309													
4	Pro	Gly	Ser	Ala	Gly	Ser	Trp	Asn	Ser	Gly	Ser	Ser	Gly	85
	CCT	GGA	AGT	GCT	GGA	AGC	TGG	AAC	TCT	GGG	AGC	TCT	GGA	95
	322													
5	Thr	Gly	Ser	Thr	Gly	Asn	Gln	Asn	Pro	Gly	Ser	Pro	Arg	69
	ACT	GGA	AGT	ACT	GGA	AAC	CAA	AAC	CCT	GGA	AGT	CCT	AGA	79
	335													
6	Pro	Gly	Ser	Thr	Gly	Thr	Trp	Asn	Pro	Gly	Ser	Ser	Glu	85
	CCT	GGT	AGT	ACC	GGA	ACC	TGG	AAT	CCT	GGC	AGC	TCT	GAA	85
	348													
7	Arg	Gly	Ser	Ala	Gly	His	Trp	Thr	Ser	Glu	Ser	Ser	Val	54
	CGC	GGA	AGT	GCT	GGG	CAC	TGG	ACC	TCT	GAG	AGC	TCT	GTA	79
	361													
8	Ser	Gly	Ser	Thr	Gly	Gln	Trp	His	Ser	Glu		Ser	Gly	62
	TCT	GGT	AGT	ACT	GGA	CAA	TGG	CAC	TCT	GAA		TCT	GGA	72
consensus sequence	Pro	Gly	Ser	Thr	Gly	Asn	Trp	Asn	Pro Ser	Gly	Ser	Ser	Gly	
	CCT	GGA	AGT	ACT	GGA	AAC	TGG	AAC	T C CT	GGG	AGC	TCT	GGA	

[a] A gap was inserted in tandem repeat 8 for better alignment.

```
pHIα3   TGTATTCCCAAATGCCTAACAGTACAGAGCCATGACTC

pHIα1   TGTATTCCCAAATGCCTAACAGTACAGAGCCATGACTC


pHIα3   AATAAATACATGTTAAATGGATGAATGAAAAAAAAAAA

pHIα1   AATAAATACATGTTAAATGGATGAATGAATTCCTCTGA
                                    EcoRI

pHIα3   AAAAAAAAACCCCCCCCCCCC

pHIα1   AACTCTAAAAAAAAAAAACCCCCCCCCCCCCCCCCCCCC
```

FIGURE 3: Comparison of 3′-noncoding regions of pHIα3 and pHIα1. The apparent polyadenylation signals (Proudfoot & Brownlee, 1976) for pHIα3 and pHIα1 are shown in boxes. The additional *Eco*RI site in the extended 3′-noncoding region of pHIα1 is underlined.

been isolated and characterized from a human liver cDNA collection with a cDNA fragment coding for the α chain of bovine fibrinogen as the initial hybridization probe. The probe was chosen on the basis of amino acid homology between the amino acid sequence predicted from the bovine cDNA (Chung et al., 1982) and the amino acid sequence of the human protein. The bovine cDNA probe covered a region of 55 amino acids starting with amino acid residue 422. In this region, there is 65% identity between the human and bovine amino acid sequences. The longest stretch of amino acid identity was nine residues in a row (amino acid residues 435–443 in the human molecule). A comparison of the corresponding DNA in this region revealed a nucleotide identity of 73%. The longest uninterrupted DNA sequence that was identical was 20 nucleotides in a row (nucleotides 1493–1512).

The overall identity in nucleotides for the α chain was 57% when the bovine cDNA was compared to the human cDNA. This low homology is due primarily to the presence of a hypervariable region in the α chain that extends from amino acid 472 to 552 in the human molecule (Henschen et al., 1980b). The divergence of DNA sequence in this region is best illustrated by a matrix that compares the corresponding nucleotides in the human and bovine cDNAs. In Figure 4, identical nucleotides are represented by a dot when 8–10 nucleotides in a row are the same in each cDNA. Diagonal lines represent regions of homology, while shifts in the line result from insertions or deletions. The hypervariable region is identified by a discontinuity in the diagonal line. It is evident from the matrix that the two cDNAs share no significant homology in the hypervariable region. Beyond this region, the line of homology is displaced 78 nucleotides from the diagonal line, which indicates that the human cDNA contains 78 more nucleotides than the bovine molecule. Consequently, the human protein is 26 amino acids longer than the bovine molecule in the hypervariable region. Other small shifts in the diagonal line are located within the coding sequences and the 3′-noncoding sequences and are due to smaller insertions or deletions. The origin of the hypervariable region is not obvious from this comparison. Since there is no homology between the human and bovine molecules in this region, the event that created this region presumably occurred after the two genes diverged. One possible explanation is that it may involve differences in intervening sequence processing. Isolation and characterization of the gene for the α chain may help to answer this possibility.

The signal sequence predicted from the cDNA sequence was 19 or 16 amino acids long, depending on whether the methionine at −19 or −16 is the correct start site for the signal sequence. Accordingly, the molecular weight for the signal peptide is either 2435 or 2016. This size is consistent with the molecular weight of 2000 that was estimated by Uzan et al. (1981) in a cell-free translation system employing human liver mRNA. Nickerson & Fuller (1981) have also observed the presence of a signal sequence in the α chain of rat fibrinogen.

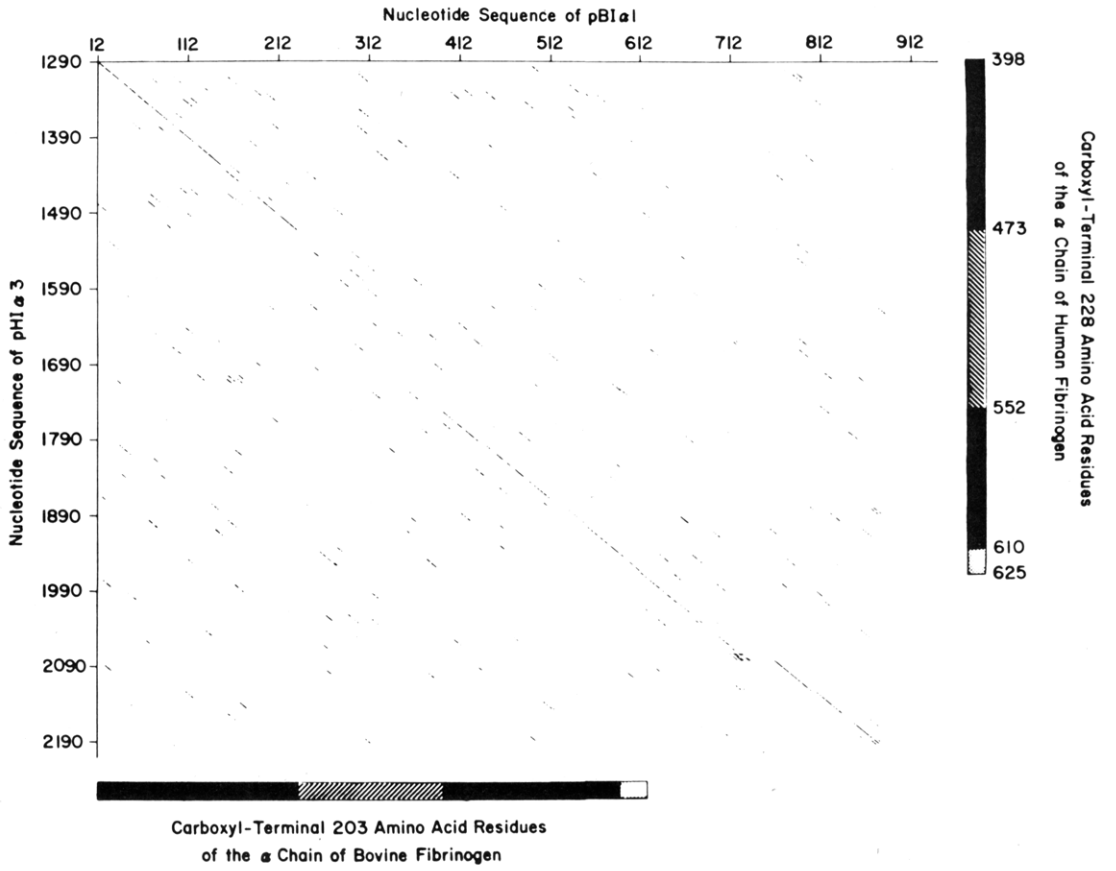

FIGURE 4: Comparison of cDNA for the α chain of bovine fibrinogen (pBIα1) with cDNA for the α chain of human fibrinogen (pHIα3) by a dot matrix. The bovine cDNA sequence is shown across the top axis, while the human cDNA sequence is shown on the left axis. Each dot represents nucleotide identity when 8–10 nucleotides in a row are identical. The corresponding protein sequences are indicated on the right axis (human) and the bottom axis (bovine). The bovine protein sequence was predicted from the cDNA in pBIα1 and was not numbered since the sequence for the amino-terminal two-thirds of the protein was not known. The solid bars represent the higher homology regions between the two molecules, while the slashed bars represent the hypervariable regions. The carboxyl-terminal extension of 15 amino acids is indicated by the dotted bars.

Table I shows the differences between the amino acids predicted from the cDNA and the amino acids as determined by protein sequence analysis for the α chain of fibrinogen (Doolittle et al., 1979; Henschen et al., 1980a). Most of the differences involve either amide assignments or inversion of dipeptides. One significant difference between the amino acid sequence predicted from the cDNA and that determined by protein sequence analysis is residue 312. Both Doolittle et al. (1979) and Henschen et al. (1980a) have reported a threonine at this position, while the cDNA sequence for plasmids pHIα2 and pHIα3 predicted an alanine. Since an error in the amino acid sequence assignment appears unlikely, it seems probable that this difference is due to minor polymorphism in the α chain.

The 15 amino acid carboxyl-terminal extension predicted by the cDNA sequence provides new insight into the structure of the fibrinogen molecule. The functional role of this carboxyl-terminal extension is not clear. Presumably, it was not observed by protein sequencing techniques (Doolittle et al., 1979; Henschen et al., 1980a) due to its low concentration or absence in plasma fibrinogen. It appears likely that the 15 amino acids are removed during assembly, during secretion, or during the circulation of fibrinogen in plasma by limited proteolysis.

The DNA sequence between nucleotides 905 and 1213 of pHIα3 contained eight homologous tandem repeats, each containing 36–39 base pairs. Five of the eight tandem repeats contained an *Sst*I restriction endonuclease site corresponding to an amino acid sequence of Gly-Ser-Ser. The restriction map for the α chain of rat fibrinogen contained two *Sac* I restriction endonuclease sites at approximately the same position (Crabtree & Kant, 1981). Since *Sac*I and *Sst*I both have the same DNA recognition sequence, it appears likely that the α chain of rat fibrinogen also contains internal tandem repeats. Cottrell et al. (1979) have identified Gln-328 and Gln-366 as the cross-linkage acceptor sites in the α chain. Gln-328 is the seventh amino acid in tandem repeat 5, and Gln-366 is the sixth amino acid in tandem repeat 8 (Table II). These are the only two glutamine residues in the eight tandem repeat structures. This suggests that the repeats have originated by some mechanism of gene duplication with two of the repeats evolving into cross-linkage sites.

A number of abnormal fibrinogen molecules with specific changes in the amino acid sequence of their α chains have been reported (Blombäck et al., 1968; Henschen et al., 1980a, 1982; Higgins & Shafer, 1981; Morris et al., 1981). The amino acid changes that have been reported thus far have occurred within the first 19 amino acid residues of the polypeptide chain (Table III). Accordingly, these amino acid substitutions often influence the release of fibrinopeptide A during the conversion of fibrinogen to fibrin by thrombin. Most of these abnormalities can be accounted for by a single base change in their gene involving both transversions and transitions (Table III). Fibrinogen Munich, however, requires two base changes in the replacement of Arg by Asn at residue 19. The availability of human cDNAs for the α chain, as well as the β and γ chains (Chung et al., 1983a,b), will now enable the study of other fibrinogen abnormalities at the gene level. Of particular in-

Table III: Comparison of Normal and Abnormal α Chains of Human Fibrinogen and the Proposed Nucleotide Changes[a]

normal fibrinogen		abnormal fibrinogen		
amino acid	codon	amino acid	codon	identification
Arg-19	A G [G]	Ser-19	A G [T/C]	Detroit[b]
Arg-19	A [G G]	Asn-19	A [A T/C]	Munich[c]
Arg-16	C [G] T	His-16	C [A] T	Sydney[d,e]
Arg-16	[C] G T	Cys-16	[T] G T	Zurich[d,f]
Asp-7	[G] A C	Asn-7	[A] A C	Lille[g]

[a] Nucleotides shown in boxes are those that have been changed by gene mutation. [b] From Blomback et al. (1968). [c] From Henschen et al. (1980a,b). [d] From Higgins & Shafer (1981) and Henschen et al. (1982). [e] Fibrinogens Petoskey, Bicetre, Louisville, Manchester, and New Albany also contain histidine in position 16 (Henschen et al., 1981). [g] From Morris et al. (1981).

terest are individuals with afibrinogenemia in which there is a complete absence of fibrinogen in their plasma.

Acknowledgments

We thank Drs. Sandra Degen, Douglas Malinowski, George Long, and Kotoku Kurachi for helpful discussions, Drs. W. Barnes and M. Zyda for providing the computer program for the dot-matrix analysis, and Drs. Roger Staden and Jon Herriott for their assistance in the computer programs for storing and analyzing DNA sequences.

Registry No. DNA (human liver clone pHIα3 fibrinogen α-chain-specifying mRNA complementary), 85882-40-2; fibrinogen (human liver clone pHIα3 α-chain precursor reduced), 85882-43-5; fibrinogen (human liver clone pHIα3 α-chain reduced), 85882-46-8.

References

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Blobel, G., Walter, P., Chang, C. N., Goldman, B. M., Erickson, A. H., & Lingappa, R. (1979) in *Secretory Mechanisms* (Hopkin, C. R., & Duncan, C. J., Eds.) Vol. 33, pp 9–36, Cambridge University Press, London.

Blombäck, B. (1969) *Thromb. Diath. Haemorrh., Suppl. 35*, 161–166.

Blombäck, M., Blombäck, B., Mammen, E. F., & Prasad, A. S. (1968) *Nature (London) 218*, 134–137.

Chandra, T., Stackhouse, R., Kidd, V. J., & Woo, S. L. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1845–1848.

Chung, D. W., MacGillivray, R. T. A., & Davie, E. W. (1980) *Ann. N.Y. Acad. Sci. 343*, 210–217.

Chung, D. W., Rixon, M. W., MacGillivray, R. T. A., & Davie, E. W. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1466–1470.

Chung, D. W., Rixon, M. W., & Davie, E. W. (1982) in *Proteins in Biology and Medicine* (Bradshaw, R. A., Hill, R. L., Tang, J., Liang, C. C., Tsao, T. C., & Tsou, C. L., Eds.) pp 309–328, Academic Press, New York.

Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., & Davie, E. W. (1983a) *Biochemistry* (second paper of three in this issue).

Chung, D. W., Chan, W.-Y., & Davie, E. W. (1983b) *Biochemistry* (third paper of three in this issue).

Cottrell, B. A., Strong, D. D., Watt, K. W. K., & Doolittle, R. F. (1979) *Biochemistry 18*, 5405–5410.

Crabtree, G. R., & Kant, J. A. (1981) *J. Biol. Chem. 256*, 9718–9723.

Crabtree, G. R., & Kant, J. A. (1982) *Cell (Cambridge, Mass.) 31*, 159–166.

Doolittle, R. F. (1973) *Adv. Protein Chem. 27*, 1–109.

Doolittle, R. F. (1975) in *The Plasma Proteins* (Putnam, F. W., Ed.) Vol. 2, pp 110–162, Academic Press, New York.

Doolittle, R. F., Watt, K. W. K., Cottrell, B. A., Strong, D. D., & Riley, M. (1979) *Nature (London) 280*, 464–468.

Finlayson, J. S., Mosesson, M. W., Bronzert, T. J., & Pisano, J. J. (1972) *J. Biol. Chem. 247*, 5220–5222.

Forman, W. B., & Barnhart, M. I. (1964) *JAMA, J. Am. Med. Assoc. 187*, 128–132.

Fowler, W. E., & Erickson, H. P. (1979) *J. Mol. Biol. 134*, 241–249.

Gergen, J. P., Stern, R. H., & Wensink, P. C. (1979) *Nucleic Acids Res. 7*, 2115–2136.

Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.

Hall, C. F., & Slayter, H. S. (1959) *J. Biophys. Biochem. Cytol. 5*, 11–16.

Henschen, A., & Lottspeich, F. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1643–1646.

Henschen, A., Lottspeich, F., & Hessel, B. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 1951–1956.

Henschen, A., Lottspeich, F., Southan, C., & Töpfer-Petersen, E. (1980a) in *Protides of Biological Fluids* (Peters, H., Ed.) pp 51–56, Pergamon Press, New York.

Henschen, A., Lottspeich, F., Töpfer-Petersen, E., Kehl, M., & Timpl, R. (1980b) in *Protides of Biological Fluids* (Peters, H., Ed.) pp 47–50, Pergamon Press, New York.

Henschen, A., Southan, C., Soria, J., Soria, C., & Samama, M. (1981) *Thromb. Haemostasis 46*, 103.

Henschen, A., Southan, C., Kehl, M., & Bode, W. (1982) *Hoppe-Seyler's Z. Physiol. Chem. 363*, 982.

Higgins, D. L., & Shafer, J. A. (1981) *J. Biol. Chem. 256*, 12013–12017.

Iwanaga, S., Blombäck, B., Gröndahl, N. J., Hessel, B., & Wallen, P. (1968) *Biochim. Biophys. Acta 160*, 280–283.

Lottspeich, F., & Henschen, A. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 935–938.

Maniatis, T., Jeffrey, A., & van de Sande, H. (1975a) *Biochemistry 14*, 3787–3794.

Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975b) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.

Marder, V. J., Francis, C. W., & Doolittle, R. F. (1982) in *Hemostasis and Thrombosis* (Colman, R. W., Hirsh, J., Marder, V. J., & Salzman, E. W., Eds.) pp 145–163, Lippincott, Philadelphia, PA.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 560–580.

McDonnell, M. W., Simon, M. N., & Studier, F. W. (1977) *J. Mol. Biol. 110*, 119–146.

McKee, P. A., Rogers, L. A., Marler, E., & Hill, R. L. (1966) *Arch. Biochem. Biophys. 116*, 271–279.

Morris, S., Denninger, M.-H., Finlayson, J. S., & Menache, D. (1981) *Thromb. Haemostasis 46*, 104.

Mosesson, M. W., Alkjaersig, N., Sweet, B., & Sherry, S. (1967) *Biochemistry 6*, 3279–3287.

Mosesson, M. W., Finlayson, J. S., Umfleet, R. A., & Galanakis, D. (1972) *J. Biol. Chem. 247*, 5210–5219.

Nickerson, J. M., & Fuller, G. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 303-307.
Pizzo, S. V., Schwartz, M. L., Hill, R. L., & McKee, P. A. (1972) *J. Biol. Chem. 247*, 636-645.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211-214.
Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. U.S.A. 54*, 158-165.
Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., & Hall, B. D. (1979) *Cell (Cambridge, Mass.) 16*, 753-761.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503-517.
Staden, R. (1977) *Nucleic Acids Res. 4*, 4037-4051.
Töpfer-Petersen, E., Lottspeich, F., & Henschen, A. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 357*, 1509-1513.
Tu, C. D., & Cohen, S. N. (1980) *Gene 10*, 177-183.
Uzan, G., Ardaillou, N., Larrieu, M. J., Kahn, A., & Marguerie, G. (1981) *Thromb. Haemostasis 46*, 180.
Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879-894.
Watt, K. W. K., Takagi, T., & Doolittle, R. F. (1979) *Biochemistry 18*, 68-76.

Characterization of Complementary Deoxyribonucleic Acid and Genomic Deoxyribonucleic Acid for the β Chain of Human Fibrinogen†

Dominic W. Chung, Benito G. Que, Mark W. Rixon, Myles Mace, Jr., and Earl W. Davie*

ABSTRACT: A total of 148 cDNAs coding for the β chain of human fibrinogen have been identified from a human liver cDNA library employing a bovine cDNA as a probe. The largest cDNA insert contained 1932 base pairs cloned into the *Pst*I site of plasmid pBR322. This cDNA insert contained 66 base pairs coding for a portion or all of a signal sequence, 1383 base pairs coding for 461 amino acids in the mature protein, a stop codon of TAG, a noncoding region of 431 base pairs, and a poly(A) tail of 19 base pairs. Most of the cDNA inserts coding for the β chain were found to have a noncoding region of 98 or 167 base pairs rather than 431 base pairs at the 3'-end. The bovine cDNA for the β chain was also employed as a probe for screening a λ phage library containing human genomic DNA. Seven positive phage were identified. One of the phage, which contained the entire gene for the β chain of fibrinogen, was examined by electron microscopy, and portions of its DNA sequence are presented. Seven intervening sequences were identified in the gene for the β chain of human fibrinogen. The largest intervening sequence (approximately 1.3 kilobases) was found at the 5'-end of the gene and was located between amino acid residues 8 and 9, which are present in fibrinopeptide B. A sequence analysis of the 5'-end of the gene also indicated that the B chain of human fibrinogen contained a signal sequence of either 16, 27, or 30 amino acid residues.

The β chain of human fibrinogen (M_r 52 000) is composed of 461 amino acids and one carbohydrate chain. The complete amino acid sequence has been elucidated by Henschen and co-workers and Watt and co-workers (Henschen & Lottspeich, 1977; Watt et al., 1979). The β chain and the α and γ chains are synthesized in hepatic parenchymal cells where the expression of the individual genes for the three chains is coordinately controlled (Nickerson & Fuller, 1981; Chung et al., 1980, 1982; Crabtree & Kant, 1981). To further understand the structure, function, and regulation of synthesis of fibrinogen, we and others have used molecular cloning techniques for the isolation of cDNAs and the genes for the three chains. Complementary DNA for the α and β chains of bovine fibrinogen (Chung et al., 1981, 1982) and the α, β, and γ chains of rat fibrinogen has been isolated and partially characterized (Crabtree & Kant, 1981, 1982). In this paper, we describe the isolation and characterization of cDNAs and the gene for the β chain of human fibrinogen. Accompanying papers describe the isolation and characterization of cDNAs for the α and γ chains of human fibrinogen (Rixon et al., 1983; Chung et al., 1983).

Experimental Procedures

Enzymes. All restriction endonucleases were purchased from either Bethesda Research Laboratories or New England Biolabs and were used as recommended by the manufacturer. T4 polynucleotide kinase was purchased from Bethesda Research Laboratories, and the Klenow fragment of *Escherichia coli* polymerase I was purchased from P-L Biochemicals. All radioactive nucleotide triphosphates were supplied by New England Nuclear.

Isolation of Liver DNA. Human liver, pulverized in liquid nitrogen, was homogenized in 10 volumes of buffer containing 10 mM Tris-HCl,[1] pH 7.5, 1 mM EDTA, and 1% sodium dodecyl sulfate (Chung et al., 1980). The homogenate was digested with proteinase K (100 μg/mL) at 40 °C for 1 h. Total nucleic acid was isolated by phenol–chloroform extraction followed by repeated ethanol precipitation (Palmiter, 1974). High molecular weight DNA was separated from RNA by equilibrium density centrifugation in cesium chloride in the presence of ethidium bromide (Radloff et al., 1967).

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195 (D.W.C., B.G.Q., M.W.R., and E.W.D.), and the Department of Pathology, Baylor College of Medicine, Houston, Texas 77030 (M.M.). *Received February 14, 1983.* This work was supported in part by Grants HL 16919 and HL 28598 from the National Institutes of Health. D.W.C. is an Established Investigator of the American Heart Association.

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/83/0422-3244$01.50/0