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

The DNA was extracted 3 times with water-saturated isoamyl alcohol to remove the ethidium bromide, and the DNA was recovered by repeated ethanol precipitation.

Southern Hybridization. DNA samples were digested with restriction endonuclease(s) and fractionated by electrophoresis in 0.7% agarose gels. The DNA fragments were transferred to nitrocellulose bidirectionally (Smith & Summers, 1980). The DNA was immobilized and hybridized to radiolabeled DNA as described (Rixon et al., 1983).

Screening of Recombinant Plasmids. Colony hybridization with Whatman 541 paper as the support medium was performed as described (Gergen et al., 1979; Wallace et al., 1981). cDNA probes were labeled to high specific radioactivity by the method of Maniatis (Maniatis et al., 1975) as described (Rixon et al., 1983). Hybridization of the filters with nick-translated bovine cDNA was performed as described (Southern, 1975).

Screening of Human Genomic Library. The human genomic library was kindly supplied by Dr. Tom Maniatis (Maniatis et al., 1978). It contained random fragments of human liver DNA generated by partial digestion with restriction endonucleases *Alu*I and *Hae*III that were cloned into λ phage Charon 4A. The library was propagated in *E. coli* strain LE392, and the plaques were screened by plaque hybridization (Benton & Davis, 1977) as modified by Woo (1979). Plaques were recovered and diluted for rescreening until plaque purification was achieved.

Preparation of Phage DNA. DNA from recombinant phage was prepared by the plate lysate method (Maniatis et al., 1982). Approximately 10^6 plaque-forming units of phage were plated in each of 20 plates and incubated for about 8 h until confluent lysis occurred. Buffer containing 10 mM Tris-HCl, pH 7.5, and 20 mM $MgCl_2$ was added to each plate and incubated at 4 °C overnight. The buffer containing released phage was recovered, briefly centrifuged to remove debris, and digested with DNase I and RNase A (both at 1 μg/mL) at 37 °C for 1 h. The phage were precipitated by the addition of poly(ethylene glycol) (PEG 6000, purchased from Calbiochem) to a final volume of 10% and incubated on ice for about 2 h (Yamamoto et al., 1970). Phage were recovered by sedimentation (6000*g* for 30 min) and purified by centrifugation over a cesium chloride step gradient (specific density 1.3–1.7) at 5 °C for 1 h. Phage were dialyzed against several changes of buffer containing 0.1 M Tris-HCl, pH 7.4, and 0.3 M NaCl. Phage DNA was extracted by digestion of the phage with proteinase K in the presence of 1% sodium dodecyl sulfate, followed by phenol–chloroform extraction and ethanol precipitation.

DNA Sequence Analysis. DNA fragments were sequenced by the method of Maxam & Gilbert (1980) as described (Rixon et al., 1983).

Electron Microscopy. Human liver poly(A) RNA was kindly provided by Dr. George Long in our laboratory. Hybrid molecules of the human genomic DNA for the β chain of fibrinogen and human liver poly(A) RNA were formed in the presence of 70% formamide. The hybrid molecules were spread onto a hypophase, transferred onto a colodion-coated grid, stained and rotary shadowed, and examined by electron microscopy (Leicht et al., 1982).

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

Results

Cross-Species Hybridization. Human liver DNA was digested to completion with several restriction enzymes, and the fragments fractionated according to size by electrophoresis, employing a 0.7% agarose gel. The fragments were transferred to nitrocellulose and hybridized to a radiolabeled cDNA fragment coding for the β chain of bovine fibrinogen. By reducing the temperature of hybridization from 68 to 58 °C, the bovine cDNA readily hybridized to specific human genomic DNA fragments (Figure 1). This selective hybridization indicated that the bovine cDNA could be used under conditions of reduced stringency as a hybridization probe for the identification of the cDNA and genomic sequences for the β chain of human fibrinogen.

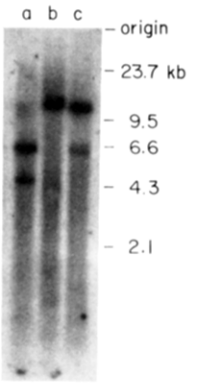

FIGURE 1: Autoradiograph of Southern hybridizations of human liver DNA with a cDNA fragment coding for the β chain of bovine fibrinogen. Human liver DNA was digested with *Eco*RI (lane a), *Bam*HI (lane b), or *Bgl*II (lane c) and fractionated by electrophoresis in a 0.7% agarose gel. DNA was transferred to nitrocellulose (Smith & Summers, 1980) and hybridized to radiolabeled bovine cDNA as described under Experimental Procedures.

Isolation of cDNA for β Chain of Human Fibrinogen. Approximately 6000 recombinant plasmids containing human liver cDNA inserts were screened with a portion of the cDNA coding for the β chain of bovine fibrinogen. A total of 148 positive clones were identified, and a DNA sequence analysis of one of these plasmids (pHIβ1)[2] confirmed that it coded for the β chain of human fibrinogen. Also, the presence of a *Bam*HI restriction site in the coding portion of the human β chain cDNA at amino acids 266–268 was identified.

In order to isolate a recombinant plasmid that extended farthest toward the 5′-end of the cDNA, the 148 positive clones were rescreened with a restriction fragment from the 5′-end of the bovine cDNA. This fragment corresponded to amino acid residues 74–171. With this cDNA probe, 28 of the colonies remained positive. DNA from the plasmids of each of the 28 colonies was then isolated (Birnboim & Doly, 1979) and digested with *Bam*HI and *Pst*I. This released the human cDNA insert from the pBR322 vector and cleaved the insert into two fragments. One of the fragments extended from the 5′-end of the cDNA insert to the *Bam*HI site and the other from the *Bam*HI site to the 3′-end of the cDNA. The identity of the fragment from the 5′-end was confirmed by Southern hybridization, employing the 5′-end of the bovine cDNA as probe. The plasmid carrying the longest insert for the β chain was named pHIβ2, and was mapped by restriction endonuclease digestion (Figure 2). The complete nucleotide sequence was then determined and the predicted amino acid sequence tabulated (Figure 3). Over 47% of the nucleotide sequence for plasmid pHIβ2 was determined on both strands, and over 70% of the sequence was confirmed in several cDNA clones.

[2] The nomenclature for pHIβ1 is as follows: p, plasmid; H, human; I, fibrinogen; β, β chain; 1, first plasmid identified. The nomenclature for λHIβ18K is as follows: λ, λ phage, H, human; I, fibrinogen; β, β chain; 18K, 18-kilobase insertion.

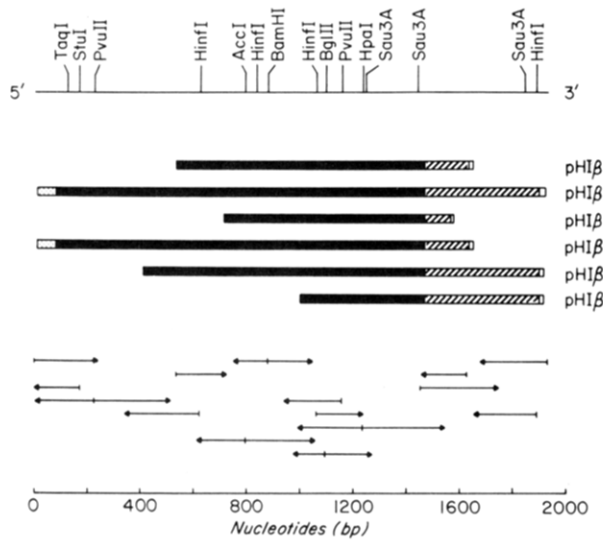

FIGURE 2: Restriction map and strategy for establishing sequence of cDNAs for the β chain of human fibrinogen. The overlapping bars below the restriction map represent the six different clones (pHIβ1 through pHIβ6). Noncoding sequences are represented by slashed bars, the signal peptide is represented by dotted bars, the coding regions are represented by solid bars, and poly(A) extensions are represented by open bars. Sites of labeling are indicated by vertical lines, and the arrows indicate the direction and extent of sequence determined. Base pairs are represented by bp.

The insert of plasmid pHIβ2 was 1932 nucleotides in length. This included 14 and 16 base pairs of poly(G) and poly(C) at the 5′- and 3′-ends, respectively. The cDNA included 66 nucleotides at the 5′-end coding for a portion or all of a leader sequence. The leader sequence was followed by 1383 base pairs coding for 461 amino acids present in the mature β chain, a stop codon of TAG, a noncoding region of 431 base pairs, and a poly(A) extension of 19 base pairs. The leader sequence was characteristic of most signal peptides and included a cluster of hydrophobic residues (six leucine residues), flanked by a hydrophilic lysine residue on either side, and an uncharged amino acid (serine) at the putative cleavage site of signal peptidase (Blobel et al., 1979). Whether the methionine residue at position −16 is the first residue in the signal peptide could not be established from these experiments.

The amino acid composition for the mature protein predicted from the cDNA sequence of plasmid pHIβ2 was as follows: Asp_{28}, Asn_{32}, Thr_{22}, Ser_{31}, Glu_{30}, Gln_{26}, Pro_{22}, Gly_{42}, Ala_{24}, Val_{25}, Met_{15}, Ile_{16}, Leu_{28}, Tyr_{21}, Phe_{10}, Lys_{31}, His_7, Arg_{27}, $^1/_2Cys_{11}$, and Trp_{13}. This corresponds to a molecular weight of 52 279 for the polypeptide chain free of carbohydrate. This is equivalent to an approximate molecular weight of 54 779 with the addition of a carbohydrate side chain (M_r 2500). The carbohydrate attachment site of Asn-X-Thr occurs at Asn residue 364 on the β chain (Toddopfer-Petersen et al., 1976).

The amino acid sequence for the β chain as predicted from the cDNA is in good agreement with that reported by amino acid sequence analysis (Henschen & Lottspeich, 1977; Watt et al., 1979). A few minor differences were noted, however, including amino acid residues 115 and 116 that were identified as Phe-Gln by DNA sequence analysis and Gln-Phe by amino acid sequence analysis. The Phe-Gln sequence was also present in the homologous position in the β chain of bovine fibrinogen (Chung et al., 1981). The amide assignments for residues 202, 296, 297, 301, 405, and 432 agree with those reported by Henschen & Lottspeich (1977). In agreement with the amino acid sequence, the nucleotide sequence from plasmid pHIβ5 predicted a proline for amino acid residue 162. However, the nucleotide sequence from plasmid pHIβ2 contains a single nucleotide change of G to C at nucleotide 564. This indicates that amino acid residue 162 was alanine in pHIβ2. This suggests that the two alleles of the gene for the β chain may be polymorphic in this position. Alternatively, this difference may be the result of an error by reverse transcriptase during the construction of the cDNA library. The 3′ noncoding region was rich in A and T (69%) and included the putative polyadenylation signal of ATTAAA (Proudfoot & Brownlee, 1976). This sequence was located 14 base pairs upstream from the poly(A) tail.

Heterogeneity in Noncoding Region at 3′-End. In screening for the longest cDNA for the β chain of human fibrinogen, the positive clones were examined by digestion with a combination of *Pst*I and *Bam*HI. Approximately two-thirds of the cDNA clones for the β chain of human fibrinogen were shown to have significantly shorter 3′ noncoding regions. Nucleotide sequence analyses on six of these clones indicated that they contained noncoding regions of 98 and 167 base pairs followed by a poly(A) extension. Putative polyadenylation signals of ATTAAA and ATTATA were found 15 and 13 base pairs upstream from the poly(A) extensions in plasmids pHIβ3 and pHIβ4, respectively. Comparison of these nucleotide sequences with the noncoding region of pHIβ2 and that of the gene described below showed that the noncoding sequences are collinear with the gene sequence. This suggests that the cDNAs with shorter lengths in their 3′ noncoding region might be generated by early termination of transcription.

Isolation of Gene for β Chain of Human Fibrinogen. A human genomic library was screened with the bovine cDNA as a hybridization probe. These experiments were performed prior to the availability of a human cDNA probe. Seven positive phage out of 2×10^6 recombinant phage were identified. Restriction endonuclease mapping and Southern hybridization of one of the λ phage, designated λHIβ18K, indicated that the gene for the β chain of human fibrinogen was interspersed with intervening sequences and the entire coding region extended approximately 10 kilobases in length. Additional hybridization analyses of the gene using fragments of cDNA for the β chain of bovine fibrinogen defined the direction of the coding sequences. A partial restriction map of this region is shown in Figure 4. The overall organization of the gene was then examined by electron microscopy by forming hybrid molecules between the cloned genomic DNA of λHIβ18K and mRNA prepared from human liver. The DNA was thermally denatured and hybrids were formed under conditions favoring RNA–DNA hybridization and not DNA–DNA reassociation. A representative hybrid molecule and a corresponding line drawing are shown in Figure 5. Seven apparent intervening sequences were identified. The largest intervening sequence was about 1300 base pairs in length and was located toward one end of the molecule. Nucleotide sequence analysis of the genomic DNA as presented below showed that this large intervening sequence was located at the 5′-end of the gene.

Sequence Analysis of Genomic DNA. The genomic DNA cloned into λHIβ18K was subcloned into pBR322. Nucleotide sequences of two of these subclones containing the 5′- and 3′-ends of the gene and those around most of the intron–exon junctions were then determined by the strategy shown in Figure 4. The nucleotide sequence corresponding to the 5′-end of the gene is shown in Figure 6. Nucleotides coding for the signal peptide and the first eight amino acid residues of fibrinopeptide B constitute the first exon. The first exon was followed by the large intervening sequence of about 1300 base pairs. Three potential initiation methionine codons were identified at the

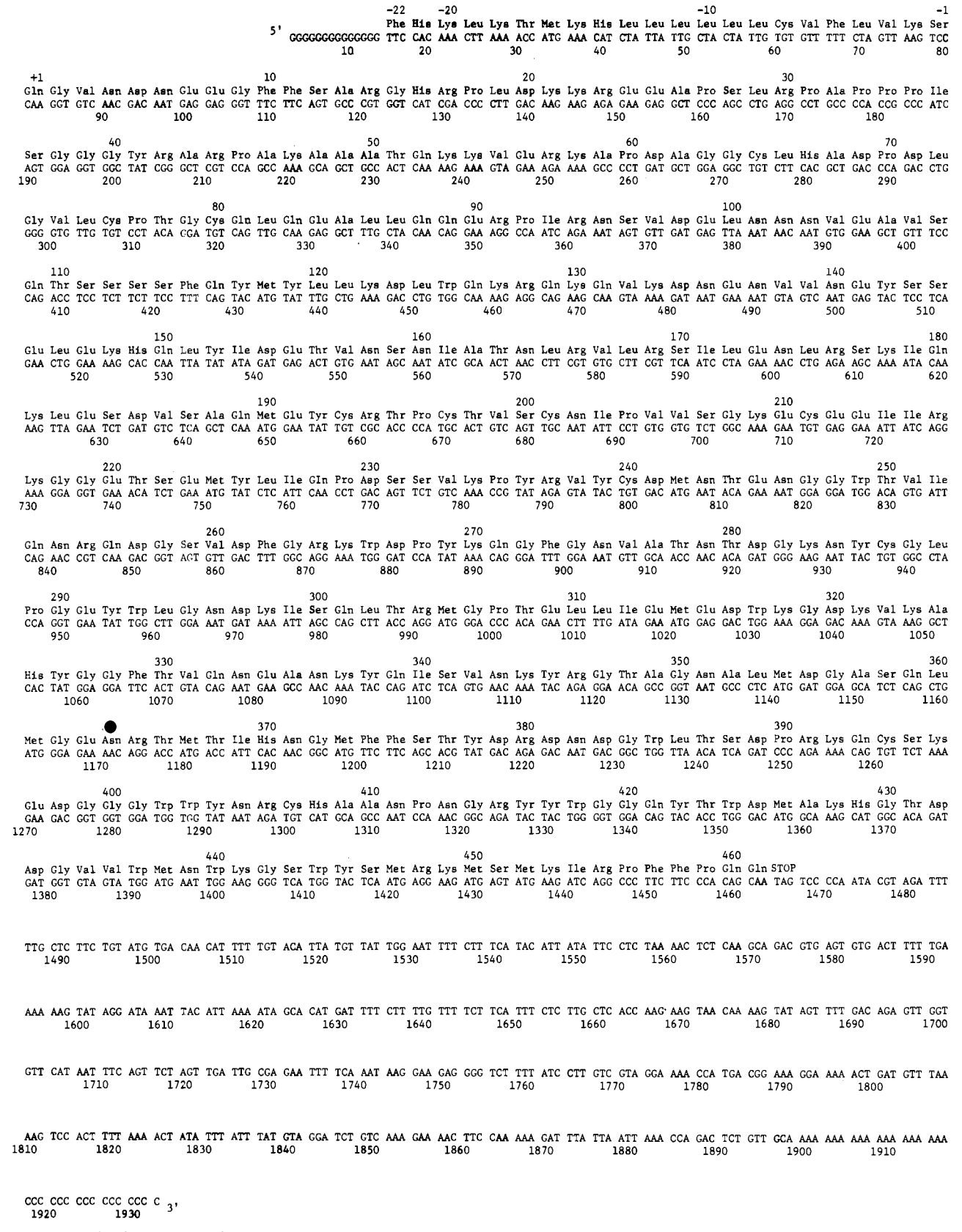

FIGURE 3: Nucleotide sequence for insert of pHIβ2. The nucleotide sequence of the coding strand and the corresponding predicted amino acid sequence are shown. Amino acid residues −1 to −22 include a portion or all of a signal peptide for the β chain. The attachment site for carbohydrate on Asn-364 is indicated by (♦).

5′-end corresponding to amino acid residues −16, −27, and −30. It is not known which of these three methionine residues is the actual start codon. The maximum length of the signal peptide was 30 amino acids, however, since a stop codon was found in phase with the coding sequence upstream from the methionine at position −30. By selective nucleotide sequence analyses, six additional intervening sequences have been identified, and these were located between amino acid residues 72 and 73, 133 and 134, 209 and 210, 248 and 249, 289 and 290, and 384 and 385 (data not shown). Accordingly, DNA

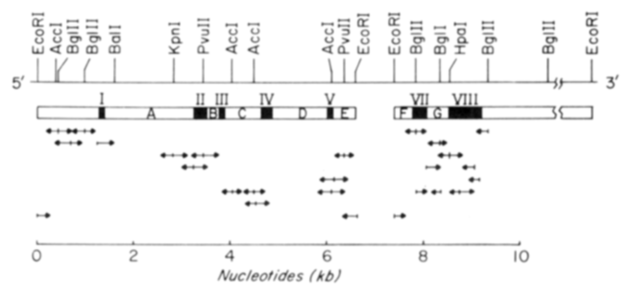

FIGURE 4: Restriction map of genomic DNA for phage λHIβ18K. Subclones containing the 5′- and 3′-ends of the gene for the β chain of human fibrinogen are represented by the boxed area. Exons are represented by solid bars, and introns are represented by open bars. Selective sequence analyses on the subclones are represented by the arrows. The sites of labeling are indicated by the vertical lines, and the direction and extent of sequence analyses are shown by the arrows. Nucleotides are in kilobases.

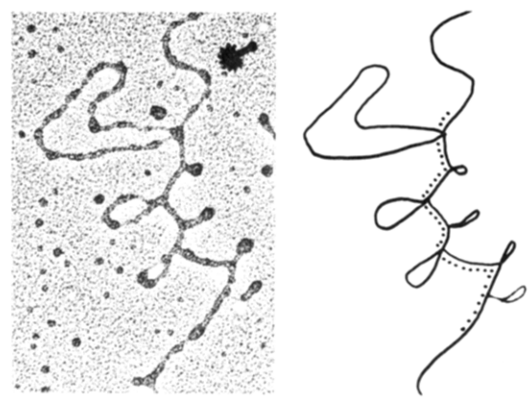

FIGURE 5: Electron micrograph and line drawing of a hybrid molecule containing genomic DNA for the β chain of human fibrinogen and human liver poly(A) RNA. Single-stranded human genomic DNA is shown by (—) and the mRNA is shown by (···) in the line drawing.

```
                                          -30         -27
                     STOP                 Met Lys Arg Met Val Ser Trp Ser Phe His
  5' .....CTCAAGT TAA GTC TAC ATG AAA AGG ATG GTT TCT TGG AGC TTC CAC
             1030        1040        1050        1060            1070

-20             -16                     -10                                 -1
Lys Leu Lys Thr Met Lys His Leu Leu Leu Leu Leu Leu Cys Val Phe Leu Val Lys Ser
AAA CTT AAA ACC ATG AAA CAT CTA TTA TTG CTA CTA TTG TGT GTT TTT CTA GTT AAG TCC
     1080        1090         1100        1110        1120             1130

 1                          8
Gln Gly Val Asn Asp Asn Glu Glu <-------------- Intervening Sequence A ---
CAA GGT GTC AAC GAC AAT GAG GAG GTGAATTTTTTAAA.....     ~ 1.3 kb
     1140        1150         1160     1170

                           9                                          20
---------------------------> Gly Phe Phe Ser Ala Arg Gly His Arg Pro Leu Asp Lys Lys
.......GCCTCTAAGGTTGTAG GGT TTC TTC AGT GCC CGT GGT CAT CGA CCC CTT GAC AAG AAG
                           2400        2410        2420        2430

                            30                                      40
Arg Glu Glu Ala Pro Ser Leu Arg Pro Ala Pro Pro Pro Ile Ser Gly Gly Gly Tyr Arg
AGA GAA GAG GCT CCC AGC CTG AGG CCT GCC CCA CCG CCC ATC AGT GGA GGT GGC TAT CGG
2440        2450        2460        2470        2480        2490

                            50                                      60
Ala Arg Pro Ala Lys Ala Ala Ala Thr Glu Lys Lys Val Glu Arg Lys Ala Pro Asp Ala
GCT CGT CCA GCC AAA GCA GCT GCC ACT CAA AAG AAA GTA GAA AGA AAA GCC CCT GAT GCT
2500        2510        2520        2530        2540        2550

                            70      72
Gly Gly Cys Leu His Ala Asp Pro Asp Leu <--- Intervening Sequence B -------
GGA GGC TGT CTT CAC GCT GAC CCA GAC TTG GTGGGTGCACTGAT.......................  3'
2560        2570        2580        2590        2600
```

FIGURE 6: Nucleotide sequence of a subclone of 5′-end of genomic DNA for the β chain of human fibrinogen. The signal sequence (16, 27, or 30 residues in length) is upstream from the Gln-Gly-Val sequence that is present in the mature polypeptide chain.

sequence analysis has confirmed the presence of seven intervening sequences that were previously identified in the electron microscope by the heteroduplex studies.

The DNA sequence for the 3′-end of the gene and its flanking sequence is shown in Figure 7. This sequence starts with methionine-452. This figure also shows the noncoding

```
                       452                                         461
                    ---Met Lys Ile Arg Pro Phe Phe Pro Gln Gln STOP
                    ---ATG AAG ATC AGG CCC TTC TTC CCA CAG CAA TAG

TCCCCAATAC GTAGATTTTT GCTCTTCTGT ATGTGACAAC ATTTTTGTAC ATTATGTTAT
        10         20         30         40         50         60
                                            pHIβ3
                                            ↓
TGGAATTTTC TTTCATACAT TATATTCCTC TAAAACACAC AAGCAGACGT GAGTGTGACT
        70         80         90        100        110        120
                                                 pHIβ1  pHIβ4
                                                     ↓   ↓
TTTTGAAAAA AGTATAGGAT AAATTACATT AAAATAGCAC ATGATTTTCT TTTGTTTTCT
       130        140        150        160        170        180

TCATTTCTCT TGCTCACCAA GAAGTAACAA AAGTATAGTT TTGACAGAGT TGGTGTTCAT
       190        200        210        220        230        240

AATTTCAGTT CTAGTTGATT GCGAGAATTT TCAAATAAGG AAGAGGGGTC TTTTACCTTG
       250        260        270        280        290        300

TCGTAGGAAA ACCATGACGG AAAGGAAAAA CTGATGTTTA AAAGTTCACT TTTAAAACTA
       310        320        330        340        350        360

TATTTATTTA TGTAGGATCT GTCAAAGAAA ACTTCCAAAA AGATTTATTA ATTAAACCAG
       370        380        390        400        410        420
  pHIβ2, pHIβ5  pHIβ6
            ↓   ↓
ACTCTGTTGC AATAAGTTAA TGTTTTCTTG TTTTGTAACC ACACATTCAA TGAGTTAGGC
       430        440        450        460        470        480

TTTGCACTTG TAAGGAAGGA GAAGCGTGCA CAACCTCAAA TAGCTAATAA ACCGGTCTTG
       490        500        510        520        530        540

AATATTGAAG ATTTAAAATC TGACTCTAGG ACGGGCACGG TGGCTCACGA CTATAATCCC
       550        560        570        580        590        600

AACACTTTGG GAGGCCGAGG CGGGCGGTCA CAAGGTCAGG AGTTCAAGAC CAGCCTGACC
       610        620        630        640        650        660

AATATGGTGA AACCCCATCC TACTAAAAAT ACAAAAATTA GCCAGGCGTG GTGCTGATGG
       670        680        690        700        710        720

GAGAAAACAG ---
       730
```

FIGURE 7: Nucleotide sequence of a subclone of 3′-end of genomic DNA for the β chain of human fibrinogen. The vertical arrows show the polyadenylation sites in the cDNA inserts for plasmids 1–6. Potential polyadenylation signals (Proudfoot & Brownlee, 1976) are underlined.

regions for plasmids pHIβ1 through pHIβ6. These plasmids vary in the length of their 3′ noncoding regions. The arrows in this figure identify the site of polyadenylation in each of the three classes of cDNAs.

Discussion

A cDNA fragment for the β chain of bovine fibrinogen has been employed for the isolation and identification of the corresponding cDNA and gene for the β chain of human fibrinogen. The specificity in cross-species hybridization was attributed to the high degree of nucleotide sequence homology between the two species. Comparison of the nucleotide sequences of the human and bovine cDNAs shows an overall homology of about 85%. The longest stretch of identical sequence was 41 base pairs, and this region was located in the 3′-half of the molecule, starting with nucleotide 852 in the human cDNA. Nucleotide sequence determinations on the cDNA and the gene confirm the presence of a signal peptide and predict a maximum length of 30 amino acid residues. This is in good agreement with the data of Nickerson & Fuller (1981) and Uzan et al. (1981).

cDNAs for the β chain of human fibrinogen constitute about 2.6% of the human liver cDNA collection of 18 000 clones. This percentage is significantly higher than that found for either the α or the γ chains of fibrinogen, which were 1.3 and 1.7%, respectively (Rixon et al., 1983; Chung et al., 1983). The larger representation of the β chain may be the result of higher levels of mRNA for this polypeptide in liver. Unequal amounts of mRNA for coordinately regulated genes have been observed for the α and β globin genes (Lodish, 1971). Al-

ternatively, the larger number of cDNA clones for the β chain may be the result of higher efficiency in the cloning process for the β chain from liver mRNA.

The gene for the β chain is coded by approximately 10 kilobases of DNA, which includes about 80% noncoding sequences. As observed by heteroduplex studies and confirmed by selective nucleotide sequence analyses, the gene contains eight exons that are interrupted by seven intervening sequences. It has been proposed by Gilbert that intervening sequences may serve to segregate a gene into structural regions corresponding to structural domains in the protein (Gilbert, 1982). Analyses of the β globin and immunoglobulin genes support this hypothesis. Although all of the structural or functional domains on the β chain of fibrinogen are not known, the locations of several of the seven intervening sequences in the gene of the β chain are consistent with this hypothesis. For example, the first exon contains the signal peptide and a portion of the fibrinopeptide B. The second exon (amino acid residues 9–72) contains the thrombin cleavage site and a region that participates in fibrin polymerization. The third and fourth exons (amino acid residues 73–133 and 134–204) encode the connecting region between interdomainal structures. These include the half-cysteines that form the two disulfide rings and the sequence that participates in superhelix folding (Watt et al., 1979).

Preliminary cell-free transcription studies using subcloned fragments for the gene of the β chain and HeLa cell extracts supplemented with polymerase II (D. W. Chung and L. Beach, unpublished results) suggest that the promoter for the β chain is located about 200 nucleotides upstream from the putative signal peptide. Experiments are now in progress to identify the transcription initiation site.

Heterogeneity at the 3′-end of the cDNAs indicates the presence of at least three different polyadenylated mRNA subpopulations for the β chain. They differ in the length of their 3′ noncoding sequences that are collinear to the gene for the β chain. These polyadenylated mRNA species appear to terminate at sites preceded by sequences similar to or identical with the polyadenylation signal. In addition, they occur upstream from regions rich in T nucleotides (Lewin, 1980). Sequences in the 3′ flanking region of the gene beyond the termination of the cDNAs contain additional potential polyadenylation signal sequences that are not utilized. These results indicate that a polyadenylation signal alone may not be sufficient for transcription termination or polyadenylation.

Northern hybridization of fractionated human liver mRNA to radiolabeled cDNA indicates two distinct mRNA sizes for the β chain (data not shown). The major band was about 1600 nucleotides in length and appeared to correspond to mRNAs containing the two subpopulations with 98 and 167 base pairs of 3′ noncoding sequences. This band was diffuse, however, because it contained two mRNA species differing in length by less than 70 nucleotides. In addition, each mRNA species may be heterogeneous in the length of its poly(A) extension. A second minor band, about 1850 nucleotides in length, represented the mRNA species with a 3′ noncoding region of 431 base pairs. The relative abundance of mRNA in these subpopulations, as estimated by the intensity of hybridization, indicated that mRNAs with shorter 3′ noncoding regions (98 and 167 base pairs) were about 3 times more prevalent than those with 431 base pairs. The occurrence of these mRNAs may be the result of endonucleolytic processing of the mRNA prior to polyadenylation or variations in transcription termination. Studies on the transcription of adenovirus 2 suggest that poly(A) addition requires prior endonucleolytic cleavage of the transcript (Nevins & Darnell, 1978; Frazer et al., 1979; Nevins et al., 1980). Alternatively, multiple transcription termination sites in the mouse α-amylase gene and dihydrofolate reductase gene have been described. In the case of the α-amylase gene, comparison of the nucleotide sequence of the gene with that of the major and minor mRNA species shows that the gene contains two polyadenylation sites located 237 base pairs apart (Tosi et al., 1981). The dihydrofolate reductase gene, on the other hand, gives rise to four distinct polyadenylated mRNA species in about equal abundance (Setzer et al., 1980). Sequence analysis also demonstrated collinearity of the cDNA with the gene sequence (Setzer et al., 1982). No functional difference was observed among the different mRNA species.

At present, approximately 65% of the nucleotide sequence of the gene for the β chain has been completed. Experiments are in progress to complete the remaining nucleotide sequence for this gene. With the structure of the gene defined, comparisons can then be made with the genes for the α and γ chains. It will also be possible to make structural comparisons with genes for the β chain in individuals with fibrinogen deficiency or functionally defective β chains.

Acknowledgments

We thank Drs. Savio L. C. Woo and Margaret Leicht for helpful discussions and assistance and Drs. Roger Staden and Jon Herriott for their help in the computer programs for storing and analyzing DNA sequences.

Registry No. Fibrinogen (human liver clone pHIβ2 β-chain protein moiety reduced), 85882-44-6; fibrinogen (human liver β-chain precursor protein moiety reduced), 85882-41-3; DNA (human liver β-chain-specifying mRNA complementary), 85882-38-8.

References

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 181–182.

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.

Chung, D. W., MacGillivray, R. T. A., & Davie, E. W. (1980) *Ann. N.Y. Acad. Sci. 343*, 210–215.

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

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.

Frazer, N. W., Nevins, J. R., Ziff, E., & Darnell, J. E. (1979) *J. Mol. Biol. 129*, 643–656.

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

Gilbert, W. (1982) *Nature (London) 271*, 501.

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

Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Jr., Davie, E. W., & Woo, S. L. C. (1982)

Nature (London) 297, 655–659.
Lewin, B. (1980) in *Gene Expression*, Vol. 2, pp 824–826, Wiley-Interscience, New York.
Lodish, H. F. (1971) *J. Biol. Chem. 246*, 7131–7138.
Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184–1188.
Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–702.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, A Laboratory Manual*, pp 65–66, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Nevins, J. R., & Darnell, J. E. (1978) *Cell (Cambridge, Mass.) 15*, 1477–1493.
Nevins, J. R., Blanchard, J. M., & Darnell, J. E. (1980) *J. Mol. Biol. 144*, 377–386.
Nickerson, J. M., & Fuller, G. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 303–307.
Palmiter, R. D. (1974) *Biochemistry 13*, 3606–3615.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Radloff, R., Bauer, W., & Vinograd, J. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 1514–1521.
Rixon, M. W., Chan, W.-Y., Davie, E. W., & Chung, D. W. (1983) *Biochemistry* (first paper of three in this issue).
Setzer, D. R., McGrogan, M., Nunberg, J. H., & Schimke, R. T. (1980) *Cell (Cambridge, Mass.) 22*, 361–370.
Setzer, D. R., McGrogan, M., & Schimke, R. T. (1982) *J. Biol. Chem. 257*, 5143–5147.
Smith, G. E., & Summers, M. D. (1980) *Anal. Biochem. 109*, 123–129.
Southern, E. (1975) *J. Mol. Biol. 98*, 503–517.
Töpfer-Petersen, E., Lottspeich, F., & Henschen, A. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 357*, 1509–1513.
Tosi, M., Young, R. A., Hagenbüchle, O., & Schibler, U. (1981) *Nucleic Acids Res. 9*, 2313–2323.
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.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 381–395.
Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., & Treiber, G. (1970) *Virology 40*, 734–744.

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

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

ABSTRACT: A number of cDNAs coding for the γ chain of human fibrinogen have been isolated from a liver cDNA library by employing a synthetic nucleotide mixture as a probe. One of the positive clones was then employed to screen the entire cDNA library of 18 000 recombinants, yielding 320 positive clones for the γ chain. The largest cDNA was 1638 base pairs in length and contained 10 base pairs of poly(G) at the 5′-end followed by 71 base pairs of noncoding nucleotides. The next 78 base pairs coded for a leader sequence that was 26 amino acids in length and included a methionine start signal and a typical hydrophobic core. The following 1233 base pairs coded for 411 amino acids that are present in the mature protein followed by a stop codon of TAA, 207 base pairs of noncoding nucleotides, a poly(A) track of 15 base pairs, and 22 base pairs of poly(C). Specific regions of the cDNA of the γ chain were then compared with the cDNAs for the α and β chains of human fibrinogen.

The γ chain of human fibrinogen (M_r 46 500) consists of 411 amino acids (Lottspeich & Henschen, 1977; Henschen & Lottspeich, 1977) and a single carbohydrate chain linked to Asn-52 (Iwanaga et al., 1968). Two γ chains, together with two α and two β chains, are cross-linked by intrachain disulfide bonds and form a distended polydomainal structure containing three nodules. An internal portion of the γ chain (amino acid residues 24–134) is bounded by unique cystine residues and is located in the interdomainal connecting structure linking the three nodules (Doolittle et al., 1978). The carboxyl-terminal region of the γ chain participates in fibrin polymerization (Olexa & Budzynski, 1981) and in intermolecular covalent cross-linking catalyzed by factor $XIII_a$ (Chen & Doolittle, 1970).

The complete amino acid sequence of the γ chain of human fibrinogen has been determined by amino acid sequencing techniques (Lottspeich & Henschen, 1977). This chain contains the unique sequence of Trp-Trp-Met-Asn-Lys starting with Trp-334. From this amino acid sequence, it can be predicted that the mRNA for the γ chain should contain the tetradecamer sequence of 5′-UGG-UGG-AUG-AA$^{U}_{C}$-AA-3′. By using the approach of Wallace et al. (1981) and Suggs et al. (1981), a DNA mixture complementary to these two tetradecamers has been employed as a hybridization probe for the identification of cDNAs coding for the γ chain of human fibrinogen. The largest cDNA that was identified was then

† 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.

0006-2960/83/0422-3250$01.50/0