CLONING OF FIBRINOGEN GENES AND THEIR cDNA *

Dominic W. Chung, Mark W. Rixon, Benito G. Que, and Earl W. Davie

Department of Biochemistry
University of Washington
Seattle, Washington 98195

Comparisons of the amino acid sequences of the three chains of fibrinogen reveal that the three chains are highly homologous and suggest that they are derived from a common ancestral gene.[1] Subsequent studies on the biosynthesis of fibrinogen in cell-free systems provide evidence that each polypeptide is encoded by distinct messenger RNA species, confirming the notion that the synthesis of fibrinogen is the result of concerted expressions of three genes.[2-4] In order to understand the structure and regulation of these genes, we have employed recombinant DNA techniques for the cloning of the cDNA for human and bovine fibrinogen α and β chains.[2,5] Some of the properties of these cDNA clones and the isolation and characterization of the gene for the human β chain are described in this report.

Isolation of the Gene for the β Chain

The identification and characterization of a cDNA for the β chain of bovine fibrinogen enabled us to predict the complete amino acid sequence of the bovine fibrinogen β chain.[5] Amino acid sequence comparisons with the human β chain indicated that the overall homology was about 75%. This suggested that the nucleotide sequence homology between the human and bovine β chains was also very high. Under conditions of reduced stringency of hybridization, the bovine cDNA was found to hybridize to discrete fragments of human genomic DNA generated by various restriction endonucleases. This specific hybridization indicates that the bovine cDNA could be used as a hybridization probe for the isolation of the gene for the human fibrinogen β chain.

A human genomic DNA library containing random fragments of the human genome constructed by Maniatis and coworkers,[6] was screened using a radiolabeled cDNA from the bovine β chain of fibrinogen. Screening of the recombinant phage plaques was carried out by a modification of the method of Benton and Davis,[7] as described by Woo.[8] Seven out of two million recombinant phages hybridized specifically with the probe. These phages were purified by plaque dilution and the DNA extracted for analyses. Restriction endonuclease mapping and Southern hybridization [9] indicated that they contained varying amounts of human DNA, including a common segment that contained nucleotide sequences homologous to the bovine cDNA probe. This segment of DNA, which was presumed to contain the gene for the β chain of human fibrinogen, was about 10 kilobases in length. It was contained within four

* This work was supported in part by Research Grant HL 16919 from the National Institutes of Health.

0077-8923/83/0408-0449 $01.75/0

Eco RI restriction endonuclease sites. Further hybridization analyses on this segment using more restricted regions of the bovine β chain cDNA defined the direction of the coding sequences.

Electron Microscopic Heteroduplex Mapping

The overall structure of the gene for the β chain of human fibrinogen was visualized by electron microscopic heteroduplex mapping studies, performed in collaboration with Dr. Myles Mace at Baylor College of Medicine in Houston. Cloned genomic DNA containing the gene for the β chain was denatured and hybridized to total human liver mRNA under conditions which favored DNA-RNA hybridization to that of DNA-DNA reassociation. The hybrids were then examined by electron microscopy. Intervening sequences, which are not represented in mature mRNA, remained as single-stranded loops connecting segments of duplex structural sequences. Results from these experiments indicated seven apparent intervening sequences. Using the appropriate single-stranded and double-stranded DNA standards, we estimated the sizes of the eight structural segments, or exons, to be 0.15, 0.2, 0.2, 0.2, 0.15, 0.3, 0.3, and 0.4 kilobases in length, and the sizes of the seven intervening sequences, or introns, were about 1.5, 0.2, 0.7, 0.3, 0.6, 0.1, and 0.4 kilobases in length. A schematic representation of this structure is shown in Figure 1. We were unable to discern any nonhybridized poly-A protrusions at either ends of the hybrid molecules. Hence, the orientation of the gene could not be assigned from these studies.

Characterization of the Gene for the β Chain of Human Fibrinogen

To further characterize the gene for the β chain, the cloned genomic DNA was excised from the recombinant phage by restriction endonuclease *Eco* RI and the contiguous fragments subcloned into the single *Eco* RI site of the

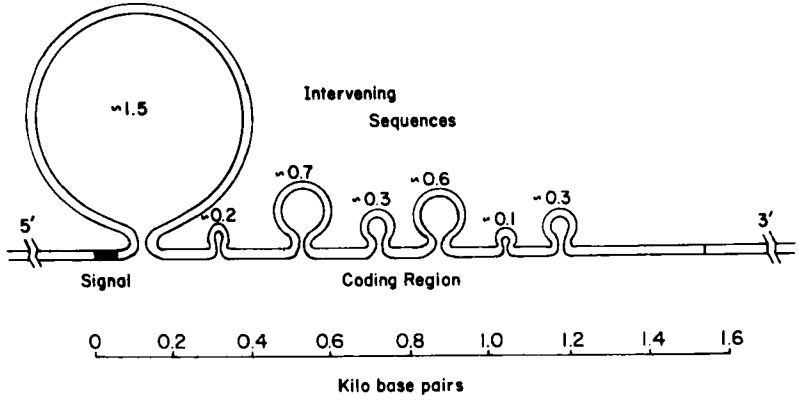

Figure 1. Schematic representation of the genomic DNA for the β chain of human fibrinogen as observed by electron microscopic heteroduplex mapping.

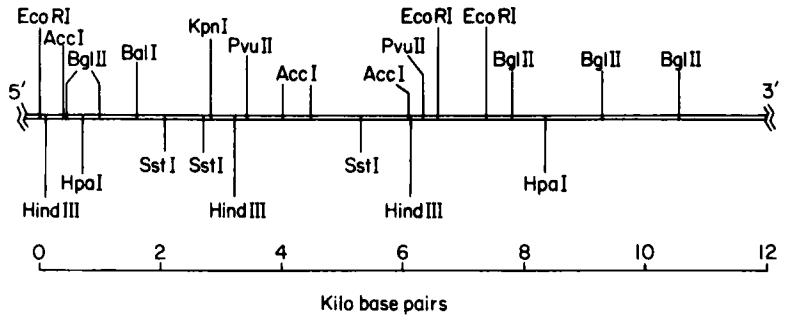

FIGURE 2. Restriction endonuclease map of the gene for the β chain of human fibrinogen. The locations of restriction enzyme cleavage sites with six-base recognition sequences are shown.

plasmid vector pBR322. Each of the subcloned *Eco* RI fragments was characterized by digestion with restriction endonucleases. The positions of restriction enzyme cleavage sites were determined by estimating the lengths of DNA segments generated by digestion to completion by a single enzyme or a combination of two or more enzymes. In some cases, partial digests of end-labeled DNA fragments were also performed. A partial restriction map showing cleavage sites of restriction enzymes with six-base recognition sequences is shown in FIGURE 2.

Using the detailed restriction map, the nucleotide sequence of the gene for the β chain was determined by the partial chemical degradative method of Maxam and Gilbert.[10] To date, over 65% of the nucleotide sequence of the gene has been determined. Specifically, these results confirm that the cloned DNA in fact codes for the β chain and that the cloned fragments contain the entire gene. The positions of the seven intervening sequences have been located and the direction of the gene was established. The size and distribution of these intervening sequences are in good agreement with those observed by electron microscopic heteroduplex mapping. The nucleotide sequence for the amino terminal and its surrounding region is shown in FIGURE 3. The first intervening sequence occurs between amino acid residues 8 and 9 within fibrinopeptide B. The size of this intervening sequence, determined from the restriction map, was about 1.5 kilobases in length, and it is the largest of the seven intervening sequences. The first amino acid of the β chain, as predicted from the genomic sequence, is glutamine which presumably undergoes cyclization to form pyroglutamic acid, which is found in the β chain of mature plasma fibrinogen.

Three potential translation initiation codons (ATG), located 16, 27 and 30 amino acids immediately upstream from the NH_2-terminal glutamine residue, have been identified. The predicted amino acid sequence for this region shows striking similarities to that of a typical signal peptide or leader sequence (FIG. 3). It contains a cluster of hydrophobic residues (six leucines) flanked by hydrophilic residues on either side, and an uncharged amino acid, in this case serine, at the putative cleavage site of signal peptidase. Comparison of the size of the mature β chain with its precursor synthesized in a cell-free translation system indicates that the length of the signal peptide is about 20 to 30

amino acids in length. This is in agreement with the predicted maximal length of 30 amino acids determined from the genomic DNA sequence.

The second intervening sequence occurs between amino acid residues 72 and 73, immediately preceding the disulfide knot region. The third intervening sequence occurs between amino acid residues 133 and 134, and is located in the middle of the coiled-coil region. This region is particularly sensitive to cleavage by plasmin. The fourth intervening sequence occurs between amino acid residues 209 and 210, about 12 amino acid residues past the coiled-coil domain. The remaining intervening sequences occur between

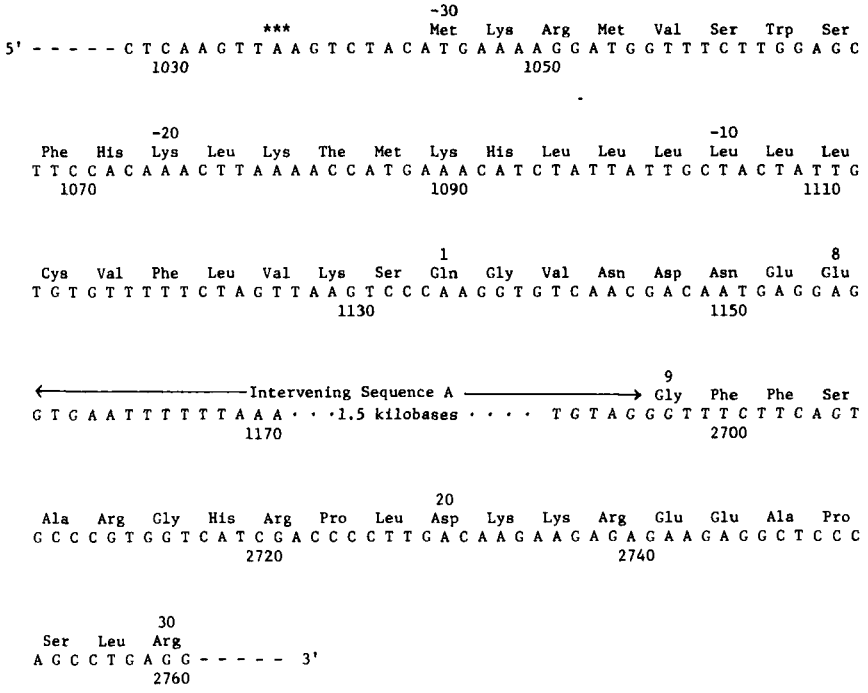

FIGURE 3. Nucleotide sequence of a portion of the gene for the β chain of human fibrinogen coding for the NH_2 terminal region of the polypeptide. The numbering of the nucleotide sequence refers to the distance from the 5' *Eco* RI site. *** indicates the position of an in-phase stop codon proximal to the potential initiation sites.

amino acid residues 247 and 248, 289 and 290, and 384 and 385. A number of the nucleotide sequences at the intron-exon junctions on the β chain gene have been determined. A shown in TABLE 1, they agree well with the consensus junction splice sequence derived from the analyses of similar junction sequences in other eucaryotic genes.[12] Intervening sequences can be classified according to their positions in the triplet codon they interrupt: type 0 intervening sequences interrupt the coding region between triplet codons; type I interrupt between the first and second nucleotide, and type II between the second and third nucleotide of the triplet codons. Intervening sequences of all three types are present in the gene for the β chain of human fibrinogen.

TABLE 1

NUCLEOTIDE SEQUENCE AT INTRON:EXON JUNCTIONS IN THE GENE FOR THE β CHAIN OF HUMAN FIBRINOGEN

Intervening Sequence	Location (AA residues)	Exon	Intron		Exon	Type
A	8–9	GAG	GTGAAT	AGGTTGTAG	GG	0
B	72–73	TTG	GTGGGT			0
C	133–134			T TTTTCCAG	AT	I
D	209–210			CATTTGCAG	AA	I
E	248–249	GAG	GTAAGC			I
F	289–290			TTCTT TTAG	GT	I
G	384–385	CTG	GTATGT	TTCTT TCAG	GT	II
Consensus Sequence *		C AG A	A GT AGT G	TTTTT T – AG CCCCC C	G–	–

* The consensus sequence for splice junctions as reported by Mount.[13]

It was proposed by Gilbert [13] that a possible function of intervening sequences may be to segregate a gene into structural regions corresponding to the structural domains of the protein. Analyses of the β globin and γ immunoglobulin gene support this notion. Although we presently do not know all the exact structural or functional domains on the β chain of fibrinogen, the positions of the first four intervening sequences are consistent with this hypothesis.

Preliminary cell-free transcription studies in HeLa cell extracts supplemented with polymerase II show that there is a strong promoter for initiation of transcription located about two to three hundred bases upstream from the putative signal peptide. The nucleotide sequence in this region is rich in A and T residues and several potential promoter sequences or TATA sequences have been located. Experiments are in progress to define the true promoter and transcription initiation sites.

```
                                        590
          Asp  Glu  Ala  Gly  Ser  Glu  Ala  Asp  His  Glu                Gly  Thr  His
Human  5' GATGAGGCCGGAAGTGAAGCCGATCATGAA------GGAACACAT
Bovine    GATGAGGCCGAAAGCCTAGAGGATCTTGGCTTTAAAGGAGCACAC
          Asp  Glu  Ala  Glu  Ser  Leu  Glu  Asp  Leu  Gly  Phe  Lys  Gly  Ala  His

               600                                                 610
          Ser  Thr  Lys  Arg  Gly  His  Ala  Lys  Ser  Arg  Pro  Val  Arg  Gly  Ile
Human  AGCACCAAGAGAGGTCATGCTAAATCTCGCCCTGTCAGAGGTATC
Bovine GGCACCCAGAAAGGCCATACTAAAGCTCGTCCTGCCAGAGGTATC
          Gly  Thr  Gln  Lys  Gly  His  Thr  Lys  Ala  Arg  Pro  Ala  Arg  Gly  Ile

                                        620                 625
          His  Thr  Ser  Pro  Leu  Gly  Lys  Pro  Ser  Leu  Ser  Pro  Stop
Human  CACACTTCTCCTTTGGGGAAGCCTTCCCCTGTCCCCCCTAGACTAAG
Bovine CACACTTCTCCTTTGGGGGAGCCCTACCCCTGACCCCCCTAGACTAAG
          His  Thr  Ser  Pro  Leu  Gly  Glu  Pro  Thr  Leu  Thr  Pro  Stop


Human  TTAAATATTTCTGCACAGTGTT-CCCATGGCCC 3'
Bovine TTCATTACTCCTGCAACGTGTCTCCCATAGCAC
```

FIGURE 4. Comparison of the nucleotide sequences around the termination codons in the cDNAs for the human and bovine α chain. Blanks are introduced to display maximum homology. The nucleotide sequence of the bovine α chain cDNA is taken from reference 2.

STUDIES ON THE cDNA FOR THE α CHAIN OF HUMAN FIBRINOGEN

A cDNA clone coding for the carboxyl 202 amino acids of the bovine α chain has been characterized.[2, 14] Unlike the β chain, the bovine α chain shares significantly less amino acid sequence homology with the corresponding region in the human α chain. A COOH-terminal extension of 15 amino acids has been identified. A hypervariable region that consists of 53 amino acids and shows no homology with the human chain is present in the bovine chain. This hypervariable region is flanked on both sides by regions of moderately good homology (70% and 56%, respectively). A portion of this cDNA, corresponding to a region of high homology was isolated and used as a hybridization probe for the human α chain. Preliminary hybridization experiments of this fragment with restriction fragments of total human genomic DNA

show discrete bands of hybridization, indicating the adequacy in specificity of this probe for the human α chain. However, because of the restricted size, the possibility of isolating the entire gene for the human α chain from the human genomic library is decreased. Alternatively, we have used this probe to isolate and identify a cDNA clone for the human α chain, which will then be used to isolate the human α chain gene. Accordingly, we have screened a collection of five thousand human cDNA clones prepared from total human liver mRNA. Fifty-six independent cDNA clones were positively identified. The inserts from these clones were mapped and the longest one we have sequenced to date is about 1.3 kilobases in length. It was observed that the codon for the COOH-terminal amino acid (Val 610) was not immediately followed by a stop codon. Instead, the mRNA apparently codes for an additional 15 amino acids, ending in proline as its COOH-terminal residue. As shown in FIGURE 4, this region is highly homologous to the putative bovine COOH-terminal extension predicted from its cDNA nucleotide sequence. These results suggest that the primary translation product of the α chain is probably longer than the mature form found in plasma. Fibrinogen subpopulations of varying degree of solubility have been isolated and characterized by Mosesson and coworkers.[15] Some of the heterogeneities have been accrued to proteolytic cleavage of the COOH-terminus of the α chain.[16-18]

SUMMARY

Cross-species hybridizations have enabled us to isolate and clone the gene for the β chain of human fibrinogen. Highlights of the gene for the β chain revealed by nucleotide sequence analyses, particularly in areas that have a direct bearing on defining the overall organization of the gene, have been presented. Nucleotide sequence determination has confirmed the presence of seven intervening sequences. The positions where several of these intervening sequences interrupt the coding region appear to be related to the functional domains of the polypeptide. A putative signal peptide has been identified. Studies on the cDNA for the human α chain indicate that the α chain polypeptide may be synthesized in a precursor form with a COOH-terminal extension of 15 amino acids as compared to the α chain present in the mature molecule found in plasma.

We are in the process of isolating the genes for the α and γ chains by a similar approach. We are hopeful that these studies will provide information as to how they are regulated and how they have undergone changes in the course of evolution.

REFERENCES

1. DOOLITTLE, R. F., K. W. K. WATT, B. A. COTTRELL, D. D. STRONG & M. RILEY. 1979. Nature (London) **280:** 464–468.
2. CHUNG, D. W., M. W. RIXON & E. W. DAVIE. 1982. The Biosynthesis of Fibrinogen and the Cloning of its cDNA. *In* Proteins in Biology and Medicine. R. A. Bradshaw, R. L. Hill, J. Tang, L. Chih-chuan, T. Tien-chin & T. Chen-lu, Eds. Academic Press. New York. pp. 309–328.
3. NICKERSON, J. M. & G. M. FULLER. 1981. Proc. Natl. Acad. Sci. USA **78:** 303–307.

4. CRABTREE, G. R. & J. A. KANT. 1981. J. Biol. Chem. **256:** 9718–9723.
5. CHUNG, D. W., M. W. RIXON, R. T. A. MACGILLIVRAY & E. W. DAVIE. 1981. Proc. Natl. Acad. Sci. USA **78:** 1466–1470.
6. MANIATIS, T., R. C. HARDISON, E. LAZY, J. LANER, C. O'CONNELL, D. QUON, G. K. SIM & A. EFSTRATIADIS. 1978. Cell **15:** 687–702.
7. BENTON, W. D. & R. W. DAVIS. 1977. Science **196:** 181–182.
8. WOO, S. L. C. 1979. Methods Enzymol. **68:** 389–395.
9. SOUTHERN, E. M. 1975. J. Mol. Biol. **98:** 503–517.
10. MAXAM, A. M. & W. GILBERT. 1980. Methods Enzymol. **65:** 499–560.
11. TAKAGI, T. & R. F. DOOLITTLE. 1975. Biochemistry **14:** 940–946.
12. MOUNT, S. M. 1982. Nuc. Acid Res. **10:** 459–472.
13. GILBERT, W. 1982. Nature (London) **271:** 501.
14. CHUNG, D. W., M. W. RIXON, R. T. A. MACGILLIVRAY & E. W. DAVIE. 1981. Thrombos. Haemostas. **46:** 103.
15. MOSESSON, M. W., N. ALKJAERSIG, B. SWEET & S. SHERRY. 1967. Biochemistry **6:** 3279–3287.
16. MOSESSON, M. W., J. S. FINLAYSON, R. A. UMFLEET & D. GALANAKIS. 1972. J. Biol. Chem. **247:** 5210–5219.
17. FINLAYSON, J. S., M. W. MOSESSON, T. J. BRONZERT & J. J. PISANO. 1972. J. Biol. Chem. **247:** 5220–5222.
18. MOSESSON, M. W., J. HAINFELD, R. H. HASCHEMEYER & J. WALL. 1981. Blood **58** (Suppl. 1): 222a.

DISCUSSION OF THE PAPER

G. GRIENINGER (*New York Blood Center, New York*): How can you do a transcription assay with HeLa cells? Do they make fibrinogen?

D. W. CHUNG: We used HeLa cell extracts to provide necessary enzymes and co-factors. We then added the cloned DNA containing the putative promoter regions to produce *in vitro* transcription. If you cut the DNA into a specific size you can see that it terminates from that site.

L. J. WANGH (*Brandeis University, Waltham, MA*): I wonder if Gerry Fuller might comment on whether or not in translation of rat fibrinogens the pre-α's might still have an extension on the COOH-terminus and would therefore be expected to be larger than the secreted α?

G. FULLER (*University of Texas Medical Branch, Galveston*): We did not notice it, but I think that we will go back and look at that question again.

WANGH: I might add that in our examination of *xenopus* fibrinogens the pre-α that we observe is the same size as the secreted α.