Nucleotide Sequence of the Gene for the γ Chain of Human Fibrinogen†

Mark W. Rixon,‡ Dominic W. Chung, and Earl W. Davie*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received October 5, 1984

ABSTRACT: A human genomic DNA library was screened for the gene coding for the γ chain of fibrinogen by using a human cDNA for the γ chain as a hybridization probe. The gene was identified in three overlapping recombinant λ bacteriophage, and its sequence, including the immediate 5′ and 3′ flanking regions, was determined. The DNA sequence analysis revealed the presence of 10 exons coding for 411 amino acids present in the mature protein and a signal sequence of 26 amino acids. Two 30 base pair (bp) direct repeats of 93% identity were found 468 bp upstream from the transcription initiation site. The DNA sequence of the gene for the γ chain of human fibrinogen showed considerable sequence homology with a partial sequence reported for the gene for the γ chain of rat fibrinogen.

Fibrinogen is a plasma glycoprotein that participates in the final stage of blood coagulation [for a recent review see Doolittle (1984)]. Fibrinogen (M_r 340 000) consists of pairs of three different polypeptide chains, designated α, β, and γ (McKee et al., 1966).[1] Thrombin releases fibrinopeptides A and B from the amino termini of the α and β chains, respectively, converting fibrinogen to fibrin monomers. The fibrin monomers then polymerize and are covalently cross-linked by factor XIIIa to form a tough insoluble fibrin clot (Doolittle, 1973, 1975).

The complete amino acid sequence for the three chains of human fibrinogen has been determined (Henschen & Lottspeich, 1977; Lottspeich & Henschen, 1977; Doolittle et al., 1979; Henschen et al., 1979; Watt et al., 1979) and shown to exhibit a high degree of amino acid sequence homology. This homology has led to the proposal that the three chains are descendants of a common ancestral gene (Doolittle et al., 1979; Doolittle, 1980; Henschen et al., 1980). The β and γ chains show the most homology, with an estimated time of divergence of 600 million years ago. The α chain shows the least amount of homology, having diverged from the other two chains about 1 billion years ago (Doolittle, 1980).

Fibrinogen is synthesized in hepatic parenchymal cells from three separate mRNA species (Nickerson & Fuller, 1981; Chung et al., 1980). A net increase in circulating levels of fibrinogen can be observed upon the induction of an acute-phase state (Koj, 1974). The acute-phase state develops in response to tissue damage, inflammation, or stress, in which the plasma levels of specific glycoproteins are increased. It has also been shown that glucocorticoid stimulation of cultured chicken embryo hepatocytes results in an increase in fibrinogen synthesis and secretion (Grieninger et al., 1978). Furthermore, Crabtree & Kant (1982a) have demonstrated a coordinate increase of liver mRNA levels for the three chains of rat fibrinogen following intravenous injection of Malayan pit viper venom, a defibrination agent. Thus, the expression of the three chains of fibrinogen is coordinately regulated.

A complete understanding of fibrinogen evolution and expression requires a knowledge of the genetic fine structure of its three genes. In this paper, we report the DNA sequence of the gene coding for the γ chain of human fibrinogen and compare it with a partial sequence of the gene for the γ chain of rat fibrinogen (Crabtree & Kant, 1982b; Fowlkes et al., 1984).

Experimental Procedures

Materials. DNA restriction endonucleases were purchased from Bethesda Research Laboratories, New England Biolabs, or Amersham. DNA modification enzymes were purchased from Bethesda Research Laboratories, Collaborative Research, or New England Nuclear. Radiolabeled nucleotides (^{32}P and ^{35}S) were obtained from New England Nuclear or Amersham. Nitrocellulose was purchased from Schleicher & Schuell, Keene, NH.

Screening of Human Genomic Library. A λ Charon 4A bacteriophage library containing human fetal liver DNA was kindly provided by Dr. Tom Maniatis (Maniatis et al., 1978). The library was grown in *Escherichia coli* strain LE 392, and the phage were screened by the plaque hybridization procedure of Benton & Davis (1977) as modified by Woo (1979). Positive phage were recovered and diluted for further hybridization screening until plaque purified. Phage DNA was prepared as described (Chung et al., 1983a).

DNA Sequence Analysis. The DNA sequence was determined by the chemical degradation method of Maxam & Gilbert (1980) and the dideoxy chain termination method of Sanger et al. (1977), utilizing the M13 cloning system developed by Messing et al. (1981). Either ^{32}P-labeled or ^{35}S-labeled deoxyadenosine 5′-triphosphate (dATP) was used in conjunction with the buffer gradient gels described by Biggin et al. (1983). DNA fragments to be sequenced by using the M13 dideoxy system were either cloned directly into the replicative form of M13mp10 or M13mp11 by restriction enzyme digestion and ligation or end deleted prior to ligation by the action of exonuclease *Bal*31 as described by Poncz et al. (1982). Approximately 75% of the DNA was sequenced on both strands, and 90% was sequenced more than once. Computer-assisted analysis of DNA sequences was accomplished by using the DNA sequence programs of Staden (1977). The homology matrix programs of Pustell & Kafatos (1982) were employed for comparison of DNA sequences, using range = 10, scale = 0.95, compression = 20, and minimum value plotted = 70%.

† This work was supported in part by Research 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: Fred Hutchinson Cancer Research Center, Seattle, WA 98104.

[1] The three chains of fibrinogen (factor I) have also been termed Aα, Bβ, and γ (Blomback, 1969).

0006-2960/85/0424-2077$01.50/0

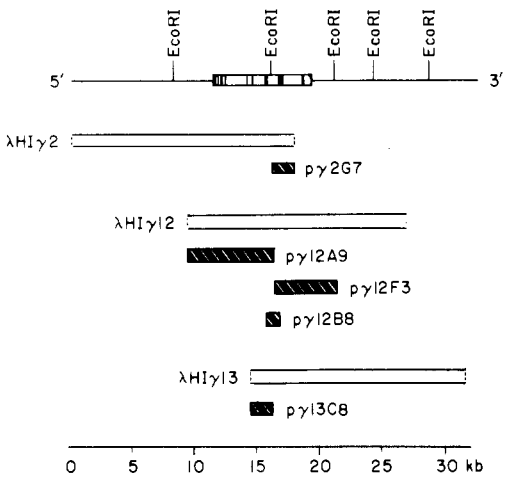

FIGURE 1: *Eco*RI map of recombinant λ phage containing genomic DNA coding for the γ chain of human fibrinogen. The location of the gene relative to the *Eco*RI sites is shown by the open bar with the dark regions (top line), indicating the locations of the 10 exons. The λ phage genomic DNA inserts (λHIγ2, λHIγ12, λHIγ13) are indicated by the open bars. The broken lines at the ends of the inserts represent the locations of the *Eco*RI linkers. *Eco*RI fragments subcloned into pBR322 for DNA sequence analysis are indicated by the solid slashed bars.

mRNA 5'-End Mapping. The primer extension method of Ghosh et al. (1978), as modified by Luse et al. (1981), was employed to map the 5' end of the transcript from the gene. A 1.5-kb fragment that encompasses the 5' end of the gene was obtained by digestion of pγ12A9[2] with *Hin*dIII and *Pvu*II. This fragment was then digested with *Ava*II and end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (Maxam & Gilbert, 1980). The labeled fragment was further digested with *Sst*I, and a 32-bp primer fragment was purified and recovered from a 3.5% polyacrylamide gel (Maniatis et al., 1975). This fragment was thermally denatured and then annealed to 50 μg of total human liver RNA. Primer extension was carried out with 9 units of AMV[3] reverse transcriptase (J. Beard, Life Sciences) at 30 °C. The extension products were analyzed on a DNA sequencing gel, and the size was determined by comparison with a DNA sequencing ladder.

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

RESULTS

A recombinant λ bacteriophage library containing human fetal liver DNA was screened for the gene coding for the γ chain of fibrinogen. A fragment from a cDNA coding for the γ chain of human fibrinogen (pHIγ2) (Chung et al., 1983b) was used as the hybridization probe. This probe contained 967 bp of DNA coding for amino acids 122–411 and included 94 bp of 3' noncoding sequences. Approximately 2 × 10^6 bacteriophage plaques were screened. Eight positive phage were plaque purified, and their DNA was isolated for further characterization.

Three different human genomic DNA fragments were identified following *Eco*RI restriction enzyme digestion and Southern blotting (Southern, 1975). An *Eco*RI restriction enzyme map indicated an overlap between the three different phage inserts, with λHIγ12 containing the entire gene (Figure 1).

The *Eco*RI fragments that hybridized to the cDNA probe (pγ2G7, pγ12A9, pγ12F3, and pγ13C8 in Figure 1) were subcloned into the *Eco*RI site of plasmid pBR322. Restriction enzyme sites in the *Eco*RI subcloned fragments were determined, and they aided in the development of the sequencing strategy shown in Figure 2. The entire nucleotide sequence of the gene for the γ chain of fibrinogen is shown in Figure 3. The nucleotide numbering of the gene was designated by assigning the proposed transcription start site (see below) as nucleotide 1. The DNA sequence across the internal *Eco*RI site (nucleotide position 4587) was determined directly from a subcloned *Hae*III fragment (pγ12B8) of the recombinant phage λHIγ12. The locations of all the 6-bp recognition restriction enzyme sites predicted from the DNA sequence, except for one, were confirmed by analytical digestion and gel electrophoresis. The exception was a predicted *Xba*I site located at nucleotide position 4749. Digestion of subclones pγ2G7 and pγ12F3 with *Xba*I failed to show cleavage at this position while the nucleotide sequence determined from both strands by Maxam and Gilbert sequencing and Sanger chain termination sequencing predicted its existence. Resistance to cleavage at this particular *Xba*I site can be explained by the fact that it contains a portion of the *dam* methylase recognition sequence GATC (Geier & Modrich, 1979). Methylation of the adenine in this sequence on both strands results in inhibition of cleavage by *Xba*I (McClelland, 1981).

The gene for the γ chain of fibrinogen contains 10 exons that code for 411 amino acids present in the mature protein plus a leader sequence of 26 amino acids. The leader sequence is encoded in exon I while the codon for the first amino acid of the mature protein starts exon II. The position of each exon, their lengths, and amino acid coding capacity along with the position and length of each intron are shown in Table I. The 5' and 3' splice junction sequences for each intron and their respective splice junction types (Sharp, 1981) are shown in Table II. All of the intron junction sequences are similar to the consensus sequences recently summarized by Mount (1982).

The assignment of the transcription initiation site in the gene for the γ chain was based on 5'-end mapping of liver RNA employing the primer extension method of Ghosh et al. (1978). The primer was generated by 5'-end labeling at the *Ava*II site, located at nucleotide position 59, followed by digestion with *Sst*I. This double-stranded fragment was 32 bp long. The 5' end of the primer was labeled at nucleotide position 62 (*Ava*II site), and its 3'-hydroxyl end was located at nucleotide position 31 (*Sst*I site). The primer fragment was thermally denatured and allowed to anneal with total human liver RNA. DNA extension was carried out with AMV reverse transcriptase and unlabeled deoxynucleotides. The resulting DNA extension products were analyzed under denaturing conditions on a DNA sequencing gel (Figure 4). Lane 5 shows the reverse transcriptase extension of the 32-bp primer without added RNA, and lane 6 shows the primer upon addition of RNA. The lengths of the major extension products range from 55 to 59 nucleotides when compared to a DNA sequencing ladder (lanes 1–4). Fainter bands, however, were also visible with lengths of 60 and 61 nucleotides. These extensions are equivalent to 23–29 bases beyond the primer. This corresponds to the gene sequence GGTGACA (Figure 3). Since duplicate experiments yielded identical results, the variability seen in the lengths of the extension products is most likely due to partial degradation of the liver RNA on the 5' end. This makes it difficult to assign the exact nucleotide where transcription initiation begins. Since adenosine 5'-triphosphate is the predominant nucleotide in eukaryotic mRNA that is

[2] The nomenclature for recombinant plasmids and bacteriophage is as follows: p, plasmid; H, human; I, fibrinogen (factor I); γ, γ chain; no., number of isolate; λ, λ phage.

[3] Abbreviations: AMV, avian myeloblastosis virus; bp, base pair(s); kb, kilobase(s).

Table I: Location and Size of Exons and Introns in the Gene for the γ Chain of Human Fibrinogen

exon	nucleotide positions	nucleotide length	amino acids[a]	intron	nucleotide positions	nucleotide length
I	1–129	129	−26 to −1[b]	A	130–225	96
II	226–270	45	1–15	B	271–459	189
III	460–643	184	16–76	C	644–762	119
IV	763–856	94	77–108	D	857–2463	1607
V	2464–2594	131	109–151	E	2595–2897	303
VI	2898–3031	134	152–196	F	3032–4010	979
VII	4011–4195	185	197–258	G	4196–5678	1483
VIII	5679–5956	278	259–350	H	5957–7594	1638
IX	7595–7764	170	351–407	I	7765–8306	542
X	8307–8525	219	408–411			

[a] Amino acids coded for by each exon. [b] Amino acids −26 to −1 refer to signal peptide.

Table II: Intron–Exon Splice Junction Sequences and Splice Junction Types in the Gene for the γ Chain of Human Fibrinogen

intron	splice junction sequences: exon | 5′	intron	3′ | exon	splice junction type[a]
A	GCA | GTAAGT	-----	TTTTAG | T	0
B	TTC | GTAAGT	-----	TTTCAG | G	0
C	CAA | GTGAGA	-----	TTACAG | A	I
D	TCG | GTAAGG	-----	TTGTAG | A	II
E	AAG | GTAACT	-----	CTCTAG | A	I
F	AAG | GTAATT	-----	AATTAG | A	0
G	CAG | GTACTG	-----	TCTCAG | T	II
H	AAG | GTATGT	-----	TTTTAG | G	I
I	CAG | GTCAGA	-----	TCACAG | G	0
consensus sequence[b]	$\begin{matrix}C\\A\end{matrix}$AG | GT$\begin{matrix}A\\G\end{matrix}$AGT	-----	$\begin{matrix}TT\\CC\end{matrix}N\begin{matrix}T\\C\end{matrix}$AG | G	

[a] From Sharp (1981). [b] From Mount (1982).

involved in the capping reaction (Corden et al., 1980), the first A in the nucleotide sequence AGGTGACA (Figure 3) was tentatively assigned as nucleotide 1.

Proposed "TATA" and "CCAT" sequences (Benoist et al., 1980) are located at nucleotide positions −24 and −57, respectively (Figure 3). The poly(A) addition site, as determined from the cDNA sequence for the γ chain, is at nucleotide position 8524 (Chung et al., 1983b). A summary of the codon usage for the γ chain has been reported previously (Chung et al., 1983b).

The size of the mature γ-chain mRNA free of its poly(A) tail was calculated as 1564 nucleotides in length as deduced from the gene sequence. The addition of a poly(A) tail of about 200 nucleotides (Mendecki et al., 1972) results in a total length of about 1760 nucleotides for the mature mRNA. This length is in good agreement with the estimated size of 1800 nucleotides reported for the size of the human γ-chain mRNA (Inman et al., 1983).

A tandem repeat of 30 bp was identified at positions −468 to −409. In these two repeats, there were two base pair differences between the two sequences. Similar repeated sequences, however, are not present in the 5′ flanking sequences in the gene for the γ chain of rat fibrinogen (Fowlkes et al., 1984).

The gene for the γ chain of human fibrinogen contains a single-copy repeat of exon IX sequences within the eighth intron. The locations of the repeat sequences are indicated in Figure 3 by the dashed underline. The repeat is from exon IX at nucleotide positions 7620–7777 and is located in the eighth intron at nucleotide positions 6577–6721. This same repeat was also observed by Fornace et al. (1984), although in their report they incorrectly designated exon IX as exon VII

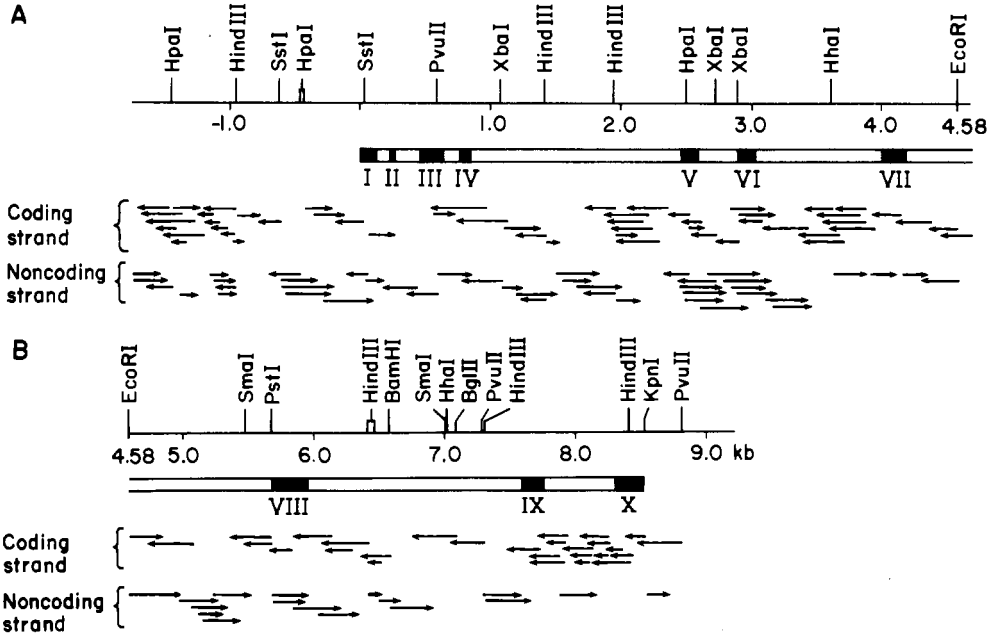

FIGURE 2: Partial restriction enzyme map of the gene for the γ chain of human fibrinogen. The DNA sequencing strategy is indicated by the length and direction of the arrows. The solid bars indicate locations of the exons. Panel A, nucleotides −1747 to 4587; panel B, nucleotides 4587–8817.

```
           CTACACA CTTCTTGAAG GCAAAGGCAA TGCTGAAGTC ACCTTTCATG TTCAAATCAT ATTAAAAAGT TAGCAAGATG TAATTATCAG TGTACTATGT  -1651
AAATCTTTGT GAATGATCAA TAATTACATA TTTTCATTAT ATATATTTTA GTAGATAATA TTTATATACA TTCAACATTC TAAATATAGA AAGTTTACAG AGAAAAATAA  -1541
AGCCTTTTTT TCCAATCCTG TCCTCCACCT CTGCATCCCA TTCTTCTTCA CAGAGGCAAC TGATTCAAGT CATTACATAG TTATTGAGTG TTAACTACAA CTATGTTAAG  -1431
TACAGCTATA TATGTTAGAT GCCGTAGCCA CAGAAATCAG TTTACAATCT AATGCAGTGG ATACAGCATG TATACATATA ATATAAGGTT GCTACAAATG CTATCTGAGG  -1321
TAGAGCTGTT TGAAAGAATA CTAATACTTA AATGTTTAAT TCAACTGACT TGATTGACAA CTGATTAGCT GAGTGGAAAA GATGGATGAG AAAGATTGTG AGACTTAATT  -1211
GGCTGGTGGT ATGGTGATAT GATTGACAAT AACTGCTAAG TCAGAGAGGG ATATATTAAG GAGGAGAAGA AAAGCAACAA ATCTGGTTTT GATGTGTTCA CTTTGTTATA  -1101
ATTATTGATT ATTTACTGAA TATGAATATT TATCTTTGTT TTTGAGTCAA TAAATATACC TTTGTAAAGA CAGAATTAAA GTATTAGTAT TTCTTTCAAA CTGGAGGCAT   -991
TTCTCCCACT AACATATTTC ATCAAAACTT ATAATAAGCT TGGTTCCAGA GGAAGAAATG AGGGATAACC AAAAATAGAG ACATTAATAA TAGTGTAACG CCCAGTGATA   -881
AATCTCAATA GGCAGTGATG ACAGACATGT TTTCCCAAAC ACAAGGATGC TGTAAGGGCC AAACAGAAAT GATGGCCCCT CCCCAGCACC TCATTTTGCC CCTTCCTTCA   -771
GCTATGCCTC TACTCTCCTT TAGATACAAG GGAGGTGGAT TTTTCTCTTC TCTGAGATAG CTTGATGGAA CCACAGGAAC AATGAAGTGG GCTCCTGGCT CTTTTCTCTG   -661
TGGCAGATGG GGTGCCATGC CCACCTTCAG ACAAAGGGAA GATTGAGCTC AAAAGCTCCC TGAGAAGTGA GAGCCTATGA ACATGGTTGA CACAGAGGGA CAGGAATGTA   -551
TTTCCAGGGT CATTCATTCC TGGGAATAGT GAACTGGGAC ATGGGGGAAG TCAGTCTCCT CCTGCCACAG CCACAGATTA AAAATAATAA TGTTAACTGA TCCCTAGGCT   -441
AAAATAATAG TGTTAACTGA TCCCTAAGCT AAGAAAGTTC TTTTGGTAAT TCAGGTGATG GCAGCAGGAC CCATCTTAAG GATAGACTAG GTTTGCTTAG TTCGAGGTCA   -331
TATCTGTTTG CTCTCAGCCA TGTACTGGAA GAAGTTGCAT CACACAGCCT CCAGGACTGC CCTCCTCCTC ACAGCAATGG ATAATGCTTC ACTAGCCTTT GCAGATAATT   -221
TTGGATCAGA GAAAAAACCT TGAGCTGGGC CAAAAAGGAG GAGCTTCAAC CTGTGTGCAA AATCTGGGAA CCTGACAGTA TAGGTTGGGG GCCAGGATGA GGAAAAAGGA   -111
ACGGGAAAGA CCTGCCCACC CTTCTGGTAA GGAGGCCCCG TGATCAGCTC CAGCCATTTG CAGTCCTGGC TATCCCAGGA GCTTACATAA AGGGACAATT GGAGCCTGAG     -1
                                                      -26
                                                      Met Ser Trp Ser Leu His Pro Arg Asn Leu Ile Leu Tyr Phe Tyr Ala
AGGTGACAGT GCTGACACTA CAAGGCTCGG AGCTCCGGGC ACTCAGACATC ATG AGT TGG TCC TTG CAC CCC CGG AAT TTA ATT CTC TAC TTC TAT GCT     99
                                     -1
Leu Leu Phe Leu Ser Ser Thr Cys Val Ala
CTT TTA TTT CTC TCT TCA ACA TGT GTA GCA GTAAGT GTGCTCTTCA CAAAACGTTG TTTAAAATGG AAAGCTGGAA AATAAAACAG ATAATAAACT AGTGA    200
                         +1                                        15
                         Tyr Val Ala Thr Arg Asp Asn Cys Cys Ile Leu Asp Glu Arg Phe
AATTT TCGTATTTTT TCTCTTTTAG TAT GTT GCT ACC AGA GAC AAC TGC TGC ATC TTA GAT GAA AGA TTC GTAAGTAGT TTTTATGTTT CTCCCTTTGT    299
GTGTGAACTG GAGAGGGGCA GAGGAATAGA AATAATTCCC TCATAAATAT CATCTGGCAC TTGTAACTTT TTAAAAACAT AGTCTAGGTT TTACCTATTT TTCTTAATAG    409
                                                     16
                                                     Gly Ser Tyr Cys Pro Thr Thr Cys Gly Ile Ala Asp Phe Leu Ser Thr
ATTTTAAGAG TAGCATCTGT CTACATTTTT AATCACTGTT ATATTTTCAG GGT AGT TAT TGT CCA ACT ACC TGT GGC ATT GCA GAT TTC CTG TCT ACT    507
Tyr Gln Thr Lys Val Asp Lys Asp Leu Gln Ser Leu Glu Asp Ile Leu His Gln Val Glu Asn Lys Thr Ser Glu Val Lys Gln Leu Ile
TAT CAA ACC AAA GTA GAC AAG GAT CTA CAG TCT TTG GAA GAC ATC TTA CAT CAA GTT GAA AAC AAA ACA TCA GAA GTC AAA CAG CTG ATA    597
                                                        76
Lys Ala Ile Gln Leu Thr Tyr Asn Pro Asp Glu Ser Ser Lys Pro
AAA GCA ATC CAA CTC ACT TAT AAT CCT GAT GAA TCA TCA AAA CCA A GTGAGAAAA TAAAGACTAC TGACCAAAAA ATAATAATAA TAATCTGTGA    692
                                                                               77
                                                                               Asn Met Ile Asp Ala Ala Thr Leu Lys Ser Arg
AGTTCTTTTG CTGTTGTTTT AGTTGTTCTA TTTGCTTAAG GATTTTTATG TCTCTGATCC TATATTACAG  AT ATG ATA GAC GCT GCT ACT TTG AAG TCC AGG    794
                                                                    108
Ile Met Leu Glu Glu Ile Met Lys Tyr Glu Ala Ser Ile Leu Thr His Asp Ser Ser Ile Arg
ATA ATG TTA GAA GAA ATT ATG AAA TAT GAA GCA TCG ATT TTA ACA CAT GAC TCA AGT ATT CG  GTAAGGATTT TTGTTTTAAT TTGCTCTGCA    886
AGACTGATTT AGTTTTTATT TAATATTCTA TACTTGAGTG AAAGTAATTT TTAATGTGTT TTCCCCATTT ATAATATCCC AGTGACATTA TGCCTGATTA TGTTGAGCAT    996
AGTAGAGATA GAAGTTTTTA GTGCAATATA AATTATACTG GGTTATAATT GCTTATTAAT AATCACATTG AAGAAAGATG TTCTAGATGT CTTCAAATGC TAGTTTGACC   1106
ATATTTATCA AAAATTTTTT CCCCATCCCC CATTTATCTT ACAACATAAA ATCAATCTCA TAGGAATTTG GGTGTTGAAA ATAAAATCCT CTTTATAAAA ATGCTGACAA   1216
ATTGGTGGTT AAAAAAATTA GCAAGCAGAG GCATAGTAAG GATTTTGGCT CCTAAAGTAA ATTATATTGA ATGTGGAGCA GGAAGAAACA TGTCTTGAGA GACTAAGTGT   1326
GGCAAATATT GCAAAGCTCA TATTGATCAT TGCAGAATGA ACCTGCATAG TCTCTTCCCT TCATTTGGAA GTGAATGTCT CTGTTAAAGC TTCTCAGGGA CTCATAAACT   1436
TTCTGAACAT AAGGTCTCAG ATACAGTTTT AATATTTTTC CCCAATTTTT TTTTCTGAAT TTTTCTCAAA GCAGCTTGAG AAATTGAGAT AAATAGTAGC TAGGGAGAAG   1546
TGGCCCAGGA AAGATTTCTC CTCTTTTTGC TATCAGAGGG CCCTTGTTAT TATTGTTATT ATTATTACTT GCATTATTAT TGTCCATCAT TGAAGTTGAA GGAGGTTATT   1656
GTACAGAAAT TGCCTAAGAC AAGGTAGAGG GAAAACGTGG ACAAATAGTT TGTCTACCCT TTTTTACTTC AAAGAAAGAA CGGTTTATGC ATTGTAGACA GTTTTCTATC   1766
ATTTTTGGAT ATTTGCAAGC CACCCTGTAA GTAACTACAA AAGGAGGGTT TTTACTTCCC CCAGTCCATT CCCAAAGCTA TGTAACCAGA AGCATTAAAG AAGAAAGGGG   1876
AAGTATCTGT TGTTTTATTT TACATACAAT AACGTTCCAG ATCATGTCCC TGTGTAAGTT ATATTTTAGA TTGAAGCTTA TATGTATAGC CTCAGTAGAT CCACAAGTGA   1986
AAGGTATACT CCTTCAGCAC ATGTGAATTA CTGAACTGAG CTTTTCCTGC TTCTAAAGCA TCAGGGGGTG TTCCTATTAA CCAGTCTCGC CACTCTTGCA GGTTGCTATC   2096
TGCTGTCCCT TATGCATAAA GTAAAAAGCA AAATGTCAAT GACATTTGCT TATTGACAAG GACTTTGTTA TTTGTGTTGG GAGTTGAGAC AATATGCCCC ATTCTAAGTA   2206
AAAAGATTCA GGTCCACATT GTATTCCTGT TTTAATTGAT TTTTTGATTT GTTTTTCTTT TTCAAAAAGT TTATAATTTT AATTCATGTT AATTTAGTAA TATAATTTTA   2316
CATTTTCCTC AAGAATGGAA TAATTTATCA GAAAGCACTT CTTAAGAAAA TACTTAGCAG TTTCCAAAGA AAATATAAAA TTACTCTTCT GAAAGGAATA CTTATTTTTG   2426
                                            109
                                            Tyr Leu Gln Glu Ile Tyr Asn Ser Asn Asn Gln Lys Ile Val Asn Leu Lys Glu Lys
TCTTCTTATT TTTGTTATCT TATGTTTCTG TTTGTAG A TAT TTG CAG GAA ATA TAT AAT TCA AAT AAT CAA AAG ATT GTT AAC CTG AAA GAG AAG    2521
                                                                          151
Val Ala Gln Leu Glu Ala Gln Cys Gln Glu Pro Cys Lys Asp Thr Val Gln Ile His Asp Ile Thr Gly Lys
GTA GCC CAG CTT GAA GCA CAG TGC CAG GAA CCT TGC AAA GAC ACG GTG CAA ATC CAT GAT ATC ACT GGG AAA G GTAACTGA TGAAGGTTAT    2612
ATTGGGATTA GGTTCATCAA AGTAAGTAAT GTAAAGGAGA AAGTATGTAC TGGAAAGTAT AGGAATAGTT TAGAAAGTGG CTACCCATTA AGTCTAAGAA TTTCAGTTGT   2722
CTAGACCTTT CTTGAATAGC TAAAAAAAAC AGTTTAAAAG GAATGCTGAT GTGAAAAGTA AGAAAATTAT TCTTGGAAAA TGAATAGTTT ACTACATGTT AAAAGCTATT   2832
```

```
                                                                                152
                                                                                Asp Cys Gln Asp Ile Ala Asn Lys Gly Ala Lys Gln
TTTCAAGGCT GGCACAGTCT TACCTGCATT TCAAACCACA GTAAAAGTCG ATTCTCCTTC TCTAG  AT TGT CAA GAC ATT GCC AAT AAG GGA GCT AAA CAG   2932

Ser Gly Leu Tyr Phe Ile Lys Pro Leu Lys Ala Asn Gln Gln Phe Leu Val Tyr Cys Glu Ile Asp Gly Ser Gly Asn Gly Trp Thr Val
AGC GGG CTT TAC TTT ATT AAA CCT CTG AAA GCT AAC CAG CAA TTC TTA GTC TAC TGT GAA ATC GAT GGG TCT GGA AAT GGA TGG ACT GTG   3022
        196
Phe Gln Lys
TTT CAG AAG GTAATTTTT TCCCCACCAT GTGTATTTAA TAAATTCCTA CATTGTTTCT GCCATATGGC AGATACTTTT CTAAGCACCT TGTGAACCGT AGCTCATTTA   3130

ATCCTTGCAA TAGCCCTAAG AGGAAGGTAC TTCTGTTACT CCTATTTACA GAAAAGGAAA CTGAGGCACA CAAGGTTAAA TAACTTGCCC AAGACCACAT AACTAATAAG   3240

CAACAGAGTC AGCATTTGAA CCTAGGCAGT ATAGTTTCAG AGTTTGTGAC TTGACTCTAT ATTGTACTGG CACTGACTTT GTAGATTCAT GGTGGCACAT AATCATAGTA   3350

CCACAGTGAC AAATAAAAAG AAGGAAACTC TTTTGTCAGG TAGGTCAAGA CCTGAGGTTT CCCATCACAA GATGAGGAAG CCCAACACCA CCCCCCACCA CCCCACCACC   3460

ATCACCACCC TTTCACACAC CAGAGGATAC ACTTGGGCTG CTCCAAGACA AGGAACCTGT GTTGCATCTG CCACTTGCTG ATACCCACTA GGAATCTTGG CTCCTTTACT   3570

TTCTGTTTAC CTCCCACCAC TGTTATAACT GTTTCTACAG GGGGCGCTCA GAGGGAATGA ATGGTGGAAG CATTAGTTGC CAGACACCGA TTGAGCAATG GGTTCCATCA   3680

TAAGTGTAAG AATCAGTAAT ATCCAGCTAG AGTTCTGAAG TCGTCTAGGT GTCTTTTTAA TATTACCACT CATTTAGAAT TTATGATGTG CCAGAAACCC TCTTAAGTAT   3790

TTCTCTTATA TTCTCTCTCA TGATCCTTGC AGCAACCCTA AGAAGTAACC ATCATTTTTC CTATTTGATA CATGAGGAAA CTGAGGTAGC TTGGCCAAGA TCACTTAGTT   3900

GGGAGTTGAT AGAACCAGTG CTCTGTATTT TTGACAAAAT GTTGACAGCA TTCTCTTTAC ATGCATTGAT AGTCTATTTT CTCCTTTTGC TCTTGCAAAT GTGTAATTAG   4010
197
Arg Leu Asp Gly Ser Val Asp Phe Lys Lys Asn Trp Ile Gln Tyr Lys Glu Gly Phe Gly His Leu Ser Pro Thr Gly Thr Thr Glu Phe
AGA CTT GAT GGC AGT GTA GAT TTC AAG AAA AAC TGG ATT CAA TAT AAA GAA GGA TTT GGA CAT CTG TCT CCT ACT GGC ACA ACA GAA TTT   4100

Trp Leu Gly Asn Glu Lys Ile His Leu Ile Ser Thr Gln Ser Ala Ile Pro Tyr Ala Leu Arg Val Glu Leu Glu Asp Trp Asn Gly Arg
TGG CTG GGA AAT GAG AAG ATT CAT TTG ATA AGC ACA CAG TCT GCC ATC CCA TAT GCA TTA AGA GTG GAA CTG GAA GAC TGG AAT GGC AGA   4190
    258
Thr Ser
ACC AG  GTACTGTTTT GAAATGACTT CCAACTTTTT ATTGTAAAGA TTGCCTGGAA TGTGCACTTT CCAACTATCA ATAGACAATG GCAAATGCAG CCTGACAAAT   4295

GCAAACAGCA CATCCAGCCA CCATTTTCTC CAGGAGTCTG TTTGGTTCTT GGGCAATCCA AAAAGGTAAA TTCTATTCAG GATGAATCTA AGTGTATTGG TACAATCTAA   4405

TTACCCTGGA ACCATTCAGA GTAATAGCTA ATTACTGAAC TTTTAATCAG TCCCAGGAAT TGAGCATAAA ATTATAATTT TATCTAGTCT AAATTACTAT TTCATGAAGC   4515

AGGTATTATT ATTAATCCCA TTTTATAGAT TAACTTGCTC AAAGTCACAT TGCTGATAAG TGGTAGAGGT AGAATTCAGA CTCAAGTAGT TTAACTTTAG AGCCTGTCCT   4625

CTTAACAACT ATCCTGGTTG AAAAGCAAAT ACAGCCTCTT CAGACTTCTC AGTGCCTTGA TGGCCATTTA TTCTGTCAAA TCATGAGCTA CCCTAAAAGT AAACCAGCTA   4735

GCTCTTTTGA TGATCTAGAG GCTTCTTTTT GCTTGAGATA TTTGAAGGTT TTAAGCATTG TTACCTAATT AAAATGCAGA AAAATATCCA ACCCTCTTGT TATGTTTAAG   4845

GAATAGTGAA ATATATTGTC TTCAAACACA TGGACTTTTT TTTATTGCTT GGTTGGTTTT TAATCCAGAA AGTGCTATAG TCAGTAGACC TTCTTCTAGG AAAGGACCTT   4955

CCATTTCCCA GCCACTGGAG ATTAGAAAAT AAGCTAAATA TTTTCTGGAA ATTTCTGTTC ATTCATTAAG GCCCATCCTT TCCCCCACTC TATAGAAGTG TTGTCCACTT   5065

GCACAATTTT TTCCAGGAAA GAATCTCTCT AACTCCTTCA GCTCACATGC TTTGGACCAC ACAGGGAAGA CTTTGATTGT GTAATGCCCT CAGAAGCTCT CCTTCTTGCC   5175

ACTACCACAC TGATTTGAGG AAGAAAATCC CTTTAGCACC TAACCCTTCA GGTGCTATGA GTGGCTAATG GAACTGTACC TCCTTCAAGT TTTGTGCAAT AATTAAGGGT   5285

CACTCACTGT CAGATACTTT CTGTGATCTA TGATAATGTG TGTGCAACAC ATAACATTTC AATAAAAGTA GAAAATATGA AATTAGAGTC ATCTACACAT CTGGATTTGA   5395

TCTTAGAATG AAACAAGCAA AAAAGCATCC AAGTGAGTGC AATTATTAGT TTTCAGAGAT GCTTCAAAGG CTTCTAGGCC CATCCCGGGA AGTGTTAATG AGCTGTGGAC   5505

TGGTTCACAT ATCTATTGCC TCTTGCCAGA TTTGCAAAAA ACTTCACTCA ATGAGCAAAT TTCAGCCTTA AGAAACAAAG TCAAAAATTC CAAGGAAGCA TCCTACGAAA   5615
                                                                               259
                                                                               Thr Ala Asp Tyr Ala Met Phe Lys Val Gly Pro Glu
GAGGGAACTT CTGAGATCCC TGAGGAGGGT CAGCATGTGA TGGTTGTATT TCCTTCTTCTCAG T ACT GCA GAC TAT GCC ATG TTC AAG GTG GGA CCT GAA   5715

Ala Asp Lys Tyr Arg Leu Thr Tyr Ala Tyr Phe Ala Gly Gly Asp Ala Gly Asp Ala Phe Asp Gly Phe Asp Phe Gly Asp Asp Pro Ser
GCT GAC AAG TAC CGC CTA ACA TAT GCC TAC TTC GCT GGT GGG GAT GCT GGA GAT GCC TTT GAT GGC TTT GAT TTT GGC GAT GAT CCT AGT   5805

Asp Lys Phe Phe Thr Ser His Asn Gly Met Gln Phe Ser Thr Trp Asp Asn Asp Asn Asp Lys Phe Glu Gly Asn Cys Ala Glu Gln Asp
GAC AAG TTT TTC ACA TCC CAT AAT GGC ATG CAG TTC AGT ACC TGG GAC AAT GAC AAT GAT AAG TTT GAA GGC AAC TGT GCT GAA CAG GAT   5895
                                                                            350
Gly Ser Gly Trp Trp Met Asn Lys Cys His Ala Gly His Leu Asn Gly Val Tyr Tyr Gln
GGA TCT GGT TGG TGG ATG AAC AAG TGT CAC GCT GGC CAT CTC AAT GGA GTT TAT TAC CAA G GTATGTTTTC CTTTCTTAGA TTCCAAGTTA   5986

ATGTATAGTG TATACTATTT TCATAAAAAA TAATAAATAG ATATGAAGAA ATGAAGAATA ATTTATAAAG ATAGTAGGGA TTTTATCATG TTCTTTATTT CAACTAAGTT   6096

CTTTGAAACT GGAAGTGGAT AATACCAAGT TCATGCCTAA AATTAGCCCT TCTAAAGAAA TCCACCTGCT GCAAAATATC CAGTAGTTTG GCATTATATG TGAAACTATC   6206

ACCATCATAG CTGGCACTGT GGGTTGTGGG ATCTCCTTTA GACATACAAC ATAAATGATC TGGATGGATT AACATTACTA CATGGATGCT TGTTGACACA TTAACCTGGC   6316

TTCCCATGAG CTTTGTGTCA GATACACGCA GTGAACAGGT GTTTGGAGGA ACAGAATAAA GAGAAGGCAA GCACTGGTAA GGGCAGGGGT TTGTGAAAGC TTGAGAGAAG   6426

AGACCAGTCT GAGGACAGTA GACACTTATT TTAGGATGGG GGTTGGATGA GGAGGCTATA GTTTGCTATA AGCTTGGAAT GGTTTGGAAC ACTGGTTTCA CTCACCTACC   6536

CAGCAGTTAT GTGTGGGGAA GCCTTACCGA TGCTAAAGGA TCCATGTTAC AATAATGGCA TTATTTGGAA ATCCCAGTGG TATTCCATGA ATAAAACCAC TATGAAGATA   6646

ATCCCACTCA ACAGACTCTC CGTTGGAGAA GGACAGCAAC ACCACCCTGG GAAAGCCAAA CAGTCAGACC AGACCTGTTT AGCATCAGTA GGACTTCCCT ACCATATCTG   6756

CTGGGTAGAT GAGTGAAACC AGTGTTCCAA ACCACTCCGG GCTTGTAGCA AACCATAGTC TCCTCATCTA CCAAGATGAG CAACCTTACC TCCTGATGTC CTAGCCAATC   6866

ACCAACTAGG AAACTTTGCA CAGTTTATTT AAAGTAACAG TTTGATTTTC ACAATATTTT TAAATTGGAG AAACATAACT TATCTTTGCA CTCACAAACC ACATAATGAG   6976

AAGAAACTCT AAGGGAAAAT GCTTGATCTG TGTGACCCGG GGCGCCATGC CAGAGCTGTA GTTCATGCCA GTGTTGTGCT CTGACAAGCC TTTTACAGAA TTACATGAGA   7086

TCTGCTTCCC TAGGACAAGG AGAAGGCAAA TCAACAGAGG CTGCACTTTA AAATGGAGAC ATAAAATAAC ATGCCAGAAC CATTTCCTAA AGCTCCTCAA TCAACCAACA   7196

AAATTGTGCT TTCAAATAAC CTGAGTTGAC CTCATCAGGA ATTTTGTGGC TCCTTCTCTT CTAACCTGCC TGAAGAAAGA TGGTCCACAG CAGCTGAGTC CGGGATGGAT   7306

AAGCTTAGGG ACAGAGGCCA ATTAGGGAAC TTTGGGTTTC TAGCCCTACT AGTAGTGAAT AAATTTAAAG TGTGGATGTG ACTATGAGTC ACAGCACAGA TGTTGTTTAA   7416
```

```
TAATATGTTT ATTTTATAAA TTGATATTTT AGGAATCTTT GGAGATATTT TCAGTTAGCA GATAATACTA TAAATTTTAT GTAACTGGCA ATGCACTTCG TAATAGACAG  7526
                                                                              351
                                                                              Gly Gly Thr Tyr Ser Lys Ala Ser Thr Pro Asn
CTCTTCATAG ACTTGCAGAG GTAAAAAGAT TCCAGAATAA TGATATGTAC ATCTACGACT TGTTTTAG  GT GGC ACT TAC TCA AAA GCA TCT ACT CCT AAT  7626

Gly Tyr Asp Asn Gly Ile Ile Trp Ala Thr Trp Lys Thr Arg Trp Tyr Ser Met Lys Lys Thr Thr Met Lys Ile Ile Pro Phe Asn Arg
GGT TAT GAT AAT GGC ATT ATT TGG GCC ACT TGG AAA ACC CGG TGG TAT TCC ATG AAG AAA ACC ACT ATG AAG ATA ATC CCA TTC AAC AGA  7716
                                                            407 408'
Leu Thr Ile Gly Glu Gly Gln Gln His His Leu Gly Gly Ala Lys Gln Val Arg Pro Glu His Pro Ala Glu Thr Glu Tyr Asp Ser Leu
CTC ACA ATT GGA GAA GGA CAG CAA CAC CAC CTG GGG GGA GCC AAA CAG GTC AGA CCA GAG CAC CCT GCG GAA ACA GAA TAT GAC TCA CTT  7806
                        427'
Tyr Pro Glu Asp Asp Leu STOP
TAC CCT GAG GAT GAT TTG TAG AAAATTAACT GCTAACTTCT ATTGACCCAC AAAGTTTCAG AAATTCTCTG AAAGTTTCTT CCTTTTTTCT CTTACTATAT  7907

TTATTGATTT CAAGTCTTCT ATTAAGGACA TTTAGCCTTC AATGGAAATT AAAACTCATT TAGGACTGTA TTTCCAAATT ACTGATATCA GAGTTATTTA AAAATTGTTT  8017
                                                                  ↓γ'
ATTTGAGGAG ATAACATTTC AACTTTGTTC CTAAATATAT AATAATAAAA TGATTGACTT TATTTGCATT TTTATGACCA CTTGTCATTT ATTTTGTCTT CGTAAATTAT  8127

TTTCATTATA TCAAATATTT TAGTATGTAC TTAATAAAAT AGGAGAACAT TTTAGAGTTT CAAATTCCCA GGTATTTTCC TTGTTTATTA CCCCTAAATC ATTCCTATTT  8237
                                                                      408         411
                                                                      Ala Gly Asp Val STOP
AATTCTTCTT TTTAAATGGA GAAAATTATG TCTTTTTAAT ATGGTTTTTG TTTTGTTATA TATTCACAG GCT GGA GAC GTT TAA AAGACCGTTT CAAAAGAGAT  8341

TTACTTTTTT AAAGGACTTT ATCTGAACAG AGAGATATAA TATTTTTCCT ATTGGACAAT GGACTTGCAA AGCTTCACTT CATTTTAAGA GCAAAAGACC CCATGTTGAA  8451
                                                                              ↓γ
AACTCCATAA CAGTTTTATG CTGATGATAA TTTATCTACA TGCATTTCAA TAAACCTTTT GTTTCCTAAG ACTAGATACA TGGTACCTTT ATTGACCATT AAAAAACCAC  8561

CACTTTTTGC CAATTTACCA ATTACAATTG GGCAACCATC AGTAGTAATT GAGTCCTCAT TTTATGCTAA ATGTTATGCC TAACTCTTTG GGAGTTACAA AGGAAATAGC  8671

AATTATGGCT TTTGCCCTCT AGGAGATACA GGACAAATAC AGGAAAATAC AGCAACCCAA ACTGACAATA CTCTATACAA GAACATAATC ACTAAGCAGG AGTCACAGCC  8781

ACACAACCAA GATGCATAGT ATCCAAAGTG CAGCTG                                                                                8817
```

FIGURE 3: DNA sequence of the gene for the γ chain of human fibrinogen. The 30-bp repeats starting at nucleotide −468 are underlined. Potential "CCAT" and "TATA" sequences are boxed. The predicted amino acid translations of the exons are indicated above the DNA sequence. The γ'-specific carboxyl-terminal polypeptide is also underlined. Exon numbering is explained in Table I. The processing or polyadenylation signal sequences (Proudfoot & Brownlee, 1976) are boxed, and the site of polyadenylation is indicated by a vertical arrow for both γ and γ' mRNAs. The locations of the single-copy exon repeats are indicated by the dashed underlines.

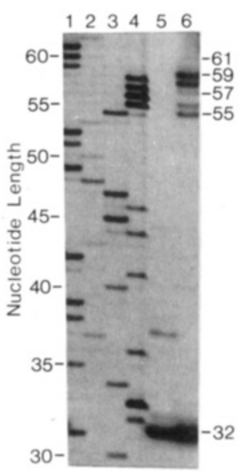

FIGURE 4: Primer extension reactions by reverse transcriptase employing human liver RNA. An end-labeled DNA primer from the gene for the γ chain of human fibrinogen was employed. Lanes 1–4, DNA sequencing ladder for size comparison; lane 5, primer extension reaction without RNA; lane 6, primer extension reaction with 50 μg of human liver RNA.

and the eighth intron as the sixth intron.

Discussion

Human genomic DNA containing sequences coding for the gene for the γ chain of fibrinogen has been isolated by using the corresponding human cDNA as a hybridization probe. The gene sequences were located within a 30-kb sequence of genomic DNA (Figure 1). The complete sequences of 10.5 kb of DNA encoding the gene revealed the presence of 10 exons coding for 411 amino acids present in the mature polypeptide and a leader sequence of 26 residues.

In genomic Southern blot experiments (unpublished results), DNA from nine unrelated individuals was digested with *Eco*RI and probed with radiolabeled cDNA for the human γ chain pHIγ2 (Chung et al., 1983b). Only two hybridizing bands, 7.8 and 5.1 kb in length, were observed in all nine cases. This pattern is consistent with the restriction map determined from the recombinant λ phage clones shown in Figure 1. This is also consistent with the conclusion that the gene for the human γ chain is present as a single copy in the haploid genome.

Table III: Differences in Nucleotides in the Gene and cDNAs Coding for the γ Chain of Human Fibrinogen

nucleotide no. in gene	nucleotide in gene	nucleotide in cDNA[a]	nucleotide in cDNA[b]
796	T	A	*c*
2971	G	A	*c*
7599	C	T	C
7747	C	C	T
7749	G	A	G
8405	CTTG	CTTGCTTG	CTTG
8492	T	T	C
8498	T	T	TT
8518	T	T	C
8524	T	T	C

[a] From Chung et al. (1983b). [b] From Kant et al. (1983). [c] The cDNA did not extend through this region.

Several differences were noted when the coding sequences in the gene were compared to the cDNA sequence determined by Chung et al. (1983b) as well as a partial cDNA sequence reported by Kant et al. (1983) (Table III). Nucleotide 796 was a T in the gene and an A in the cDNA. This difference was in the codon for amino acid residue 88 where the gene predicts an isoleucine and the cDNA predicts a lysine. Henschen et al. (1976) also reported a lysine at this position from their protein sequencing data. This suggests that the difference in the gene is due to polymorphism or a cloning artifact. Differences between the gene and the cDNAs at nucleotides 2971, 7599, and 7749 are in the third position of a codon and do not change the predicted amino acid. Several differences also occur in the 3′ noncoding sequence (Table III). At nucleotide position 8405 in the gene, the sequence CTTG occurs once. In our original cDNA sequence (Chung et al., 1983b), we erroneously reported a duplication of this sequence.

The 5′ end of the γ-chain transcript was mapped to nucleotide position 1 (Figure 3). This differs from the results

obtained from the cDNA sequence (Chung et al., 1983b), which had a 5′ end corresponding to position −20 in the gene sequence. The results obtained here are comparable with the results from the gene for the γ chain of rat fibrinogen (Fowlkes et al., 1984), which placed the 5′ end of its transcript at the nucleotide corresponding to position 8 in the human gene. Although no definitive explanation can be offered to explain this discrepancy between the human gene and cDNA, it is possible that the gene for the γ chain contains a weak upstream promoter similar to the human *c-myc* gene as described by Batey et al. (1983). Since the cDNA was selected on the basis of being the longest, it would not be surprising that a minor species was identified and characterized.

The positions of the 9 introns divide the gene into 10 segments that can roughly be assigned separate functions. Exon I codes for the signal peptide. The cysteines involved in linking the two γ chains together by disulfide bridges (Hoeprich & Doolittle, 1983) are encoded in exon II. Exon III contains the first disulfide ring (Doolittle et al., 1978), a portion of the coiled coil (Doolittle et al., 1978), and the carbohydrate attachment site at asparagine residue 52 (Iwanaga et al., 1968). Exon IV contains another portion of the coiled coil. Exon V contains the remaining portion of the coiled coil and the second disulfide ring. Exons VI, VII, and VIII all contain residues involved in forming the D domain. This region shows the highest degree of homology when compared to the β chain of fibrinogen (Henshen et al., 1980; Doolittle, 1980). Exon IX contains glutamine-398 and lysine-406, which are the two residues involved in the covalent cross-linking of the γ chains from separate fibrin monomers (Chen & Doolittle, 1970). Exons IX and X code for amino acids involved in fibrin polymerization (Olexa & Budzynski, 1981) and platelet receptor recognition (Kloczewiak et al., 1984).

The high degree of similarity between the rat and human cDNA sequences for the γ chain has been noted previously (Chung et al., 1983b). This similarity decreases somewhat upon comparison of their gene sequences. Figure 5 shows a comparison of the 5′ flanking sequences from the genes for the γ chain of human and rat fibrinogen (Fowlkes et al., 1984) in the form of a matrix, using the program of Pustell & Kafatos (1982). In this matrix 21 nucleotides from both sequences are compared for percent identity greater than or equal to 70. The identity is recorded as letters where A represents 98–100%, B represents 96–98%, etc. A considerable amount of homology is observed for about 300 nucleotides into the 5′ flanking reigion of the genes. Further upstream, regions of identity are also revealed by the letters that fall on the diagonal. Even as far 5′ as nucleotide −1460 in the human gene there is some sequence homology with the rat gene at position −1483. However, the greatest degree of homology exists only 300 bases upstream from the site of transcription initiation. The human gene contains a sequence nearly identical (87%) with a short conserved region proposed by Fowlkes et al. (1984) to be homologous in all three rat fibrinogen genes. The human gene, however, does not share the same degree of identity in the longer conserved region of homology found in the three rat genes for the α, β, and γ chains of fibrinogen. This long conserved region from the gene for the γ chain of rat fibrinogen does show about 50% identity with 38 nucleotides from the 30-bp repeats in the human gene (nucleotides −467 through −430). The rat gene, however, does not contain any obvious repeat structures homologous to the human 30-bp repeats.

Figure 6 illustrates a comparison between the human gene sequence and partial rat gene sequences (Crabtree & Kant,

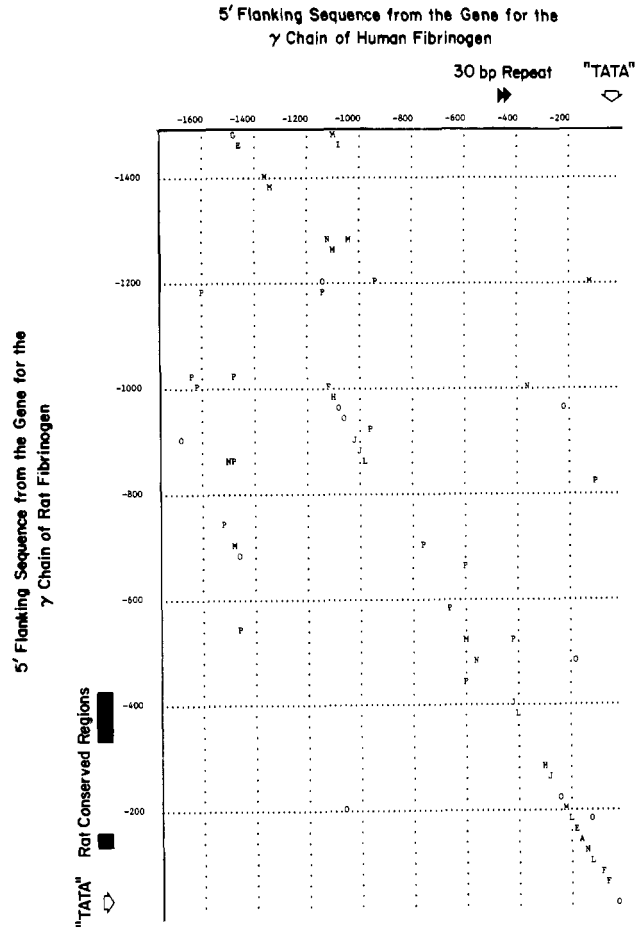

FIGURE 5: Comparison of 5′ flanking DNA sequences from the genes for the γ chains of human and rat fibrinogen. The locations of potential "TATA" sequences are shown with open arrows. The locations of the 30-bp repeats in the human gene are identified by the black arrowheads. The locations of the conserved regions found in the three rat fibrinogen genes (Fowlkes et al., 1984) are indicated by solid bars.

1982b) downstream from the transcription initiation site in a matrix similar to Figure 5. Figure 6A shows a comparison of sequences from the transcription initiation site into intron C. The diagonal lines of letters indicate regions of homology that generally fall within the exon sequences and into the 5′ ends of the introns. The rest of the intron sequences do not score greater than 70% identity and for the most part appear to be unrelated, although scattered small areas of 50–60% identity are present. Intron A from the human gene does not contain any of the 32 alternating GA copolymers that are present in intron A from the rat gene. Figure 6B shows a comparison of sequences from the last intron and final exon of the human and rat genes. Again, there is a high degree of homology shown in the exon sequences. However, this matrix also illustrates the high degree of similarity present in the final intron. This is the intron that contains the sequences that code for the γ-chain variant, γ′ (Chung & Davie, 1984; Crabtree & Kant, 1982b). The human γ′ polypeptide is larger than the predominant γ chain (Francis et al., 1980; Wolfenstein-Todel & Mosesson, 1980, 1981) and is generated by alternative processing and polyadenylation of the γ gene transcript within the ninth intron. This gives rise to a read-through of the ninth exon/ninth intron junction sequence. A proposed mechanism for the formation of the γ′ chain of human fibrinogen has been described elsewhere (Chung & Davie, 1984). The high degree of homology conserved in the rat and human final introns suggests that evolutionary restraints have been placed on these sequences to hinder their

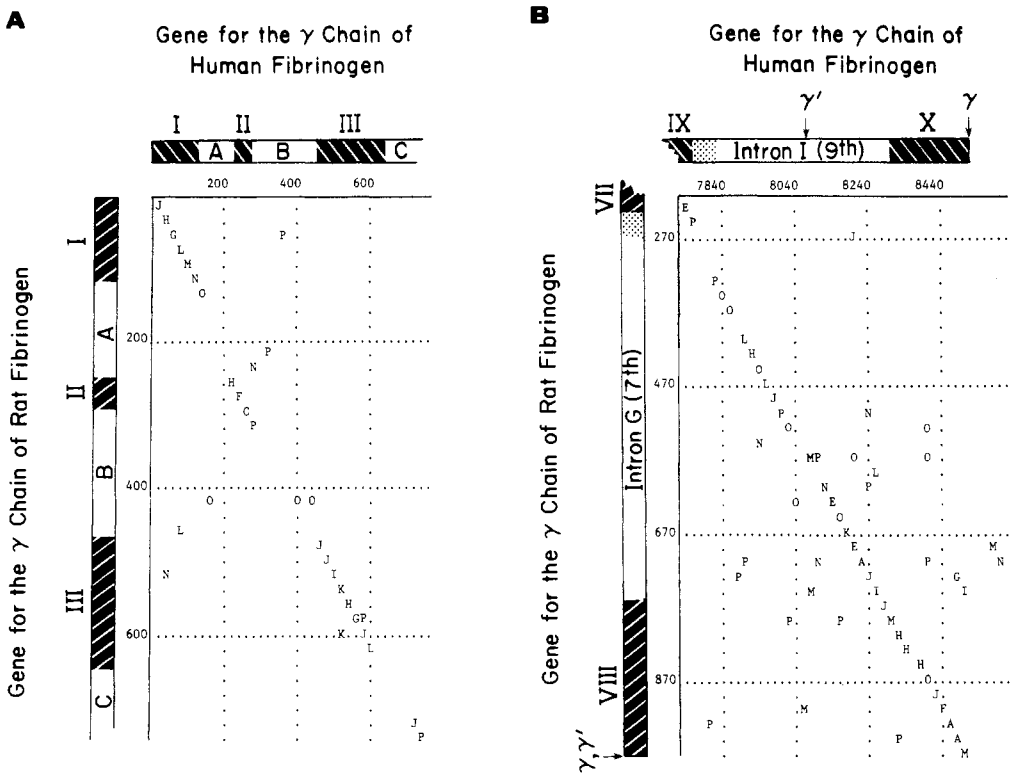

FIGURE 6: Comparison of DNA sequences from the genes for the γ chains of human and rat fibrinogen. Panel A shows a comparison of sequences from the proposed transcription initiation sites into intron C, and panel B shows a comparison of sequences from the penultimate exons to the polyadenylation sites. Slashed bars indicate exons, open bars indicate introns, and dotted bars indicate the locations of DNA coding for the γ′-specific polypeptides. Sites of poly(A) addition are indicated by arrows.

divergence as compared to the other intron sequences.

A difference does exist between the number of reported exons in the rat and human genes. The human gene has 10 exons while Crabtree & Kant (1982b) have reported the presence of 8 in the rat. The difference in the numbering of the human exons that Fornace et al. (1984) reported in describing the single-copy repeat of exon IX can be explained if they have numbered their human exons by analogy to the rat gene structure.

The splice junction sequences in the human gene for the γ chain of fibrinogen agree well with the proposed consensus sequences of Mount (1982) except for the 5′ splice junction for intron IX (Table II). This junction has the sequence GTC as the first three nucleotides. The ninth intron is not removed in the processing of the mRNA for the γ′ chain (Chung & Davie, 1984) and includes the processing and poly(A) addition signals. The rat gene also contains GTC at the 5′ splice site of its last intron. According to Mount (1982), only 3 of the 139 5′ splice junctions have a C in the third position. This 5′ splice junction sequence may be involved in the alternative processing that occurs in the generation of the γ′ mRNA molecule.

A recent report by McDevitt et al. (1984) describes DNA sequences that are required for the correct 3′ processing and polyadenylation of the adenovirus E2A transcript. Sequences approximately 35 nucleotides downstream of the poly(A) addition site appear to be necessary. These sequences can potentially form a stable stem–loop structure in the primary transcript with the AAUAAA conserved sequence (Proudfoot & Brownlee, 1976). Analysis of the DNA sequences in the gene for the γ chain of human fibrinogen, downstream of the poly(A) addition sites for the γ and γ′ transcripts, revealed similar sequences that have the potential for stable stem–loop structure formation. Figure 7 shows the sequences around each of the two poly(A) addition sites. The site of poly(A) addition

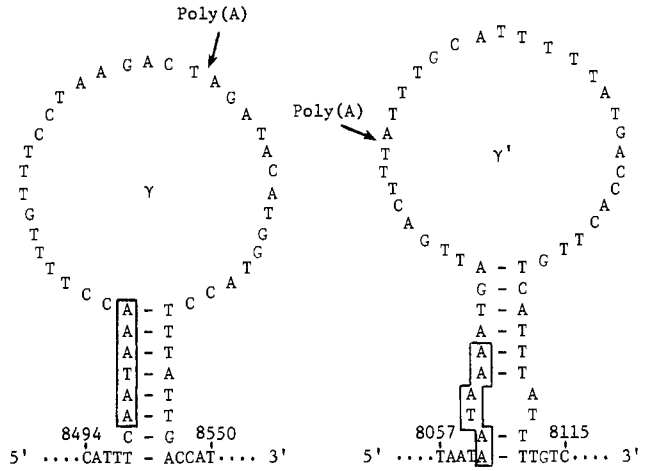

FIGURE 7: Potential stem–loop structures at the γ and γ′ poly(A) addition sites. The poly(A) addition sites are indicated by arrows. The processing or polyadenylation sequences of AATAAA are boxed. The numbers above the nucleotides correspond to those in Figure 3.

is indicated, and the Proudfoot & Brownlee (1976) consensus sequence is boxed. Nucleotides that are complementary and could potentially anneal are shown in the stem region. In the γ chain specific segment there is a potential for a stem of 8 nucleotides and a loop of 33, while in the γ′ chain specific segment there is a potential stem of 10 (containing 2 mismatches) with a loop of 31. Since the stem in the γ-chain segment has a longer uninterrupted hybrid (8 vs. 6), it may form a more stable structure. The stability of this stem–loop structure could be a factor in the choice of the γ chain specific site as the predominant site for 3′ processing and polyadenylation.

Fibrinogen and α_1-antitrypsin are acute-phase proteins whose plasma levels increase upon the induction of an

acute-phase response (Koj, 1974). If this increase is related to an increase in transcriptional activity, comparison of their respective gene sequences may reveal homologous DNA sequences that are involved. A comparison of 5′ flanking sequences from the gene for the γ chain of human fibrinogen with the gene for human α_1-antitrypsin (Long et al., 1984) did not, however, reveal any long areas of high sequence identity. Regions of 21–26 nucleotides in length with 61–67% sequence identity were observed. The significance of these sequence homologies is questionable since some of the same sequences are also present in the gene of human prothrombin, which is not an acute-phase protein (Degen et al., 1983; S. Degen, personal communication). Comparison of the DNA sequence from other acute-phase protein genes may aid in determining regions of potentially significant homology.

The determination of the nucleotide sequence of the gene for the γ chain of human fibrinogen will enable future comparisons to be made with the gene sequences of the α and β chains of human fibrinogen.

ACKNOWLEDGMENTS

We thank Drs. Wai-Yee Chan, Sandra Degen, Evan Sadler, Shinji Yoshitake, Steve Leytus, and Ko Kurachi for helpful discussions and Dr. Jon Herriott for providing the computer facilities used in storing and analyzing the DNA sequence.

REFERENCES

Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., & Leder, P. (1983) *Cell (Cambridge, Mass.) 34*, 779–787.

Benoist, C., O'Hare, K., Breathnach, R., & Chambon, P. (1980) *Nucleic Acids Res. 8*, 127–142.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–182.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Chen, R., & Doolittle, R. F. (1970) *Proc. Natl. Acad. Sci. U.S.A. 66*, 472–479.

Chung, D. W., & Davie, E. W. (1984) *Biochemistry 23*, 4232–4236.

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

Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., & Davie, E. W. (1983a) *Biochemistry 22*, 3244–3250.

Chung, D. W., Chan, W.-Y., & Davie, E. W. (1983b) *Biochemistry 22*, 3250–3256.

Corden, J., Wasylyk, B., Buchwalder, P., Sassone-Corsi, P., Kedinger, C., & Chambon, P. (1980) *Science (Washington, D.C.) 209*, 1406–1414.

Crabtree, G. R., & Kant, J. A. (1982a) *J. Biol. Chem. 257*, 7277–7279.

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

Degen, S. J. F., MacGillivray, R. T. A., & Davie, E. W. (1983) *Biochemistry 22*, 2087–2097.

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

Doolittle, R. F. (1975) *Plasma Proteins (2nd Ed.) 2*, 109–161.

Doolittle, R. F. (1980) *Protides Biol. Fluids 28*, 41–46.

Doolittle, R. F. (1984) *Annu. Rev. Biochem. 53*, 195–229.

Doolittle, R. F., Goldbaum, D. M., & Doolittle, L. R. (1978) *J. Mol. Biol. 120*, 311–325.

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

Fornace, A. J., Jr., Cummings, D. E., Comeau, C. M., Kant, J. A., & Crabtree, G. R. (1984) *Science (Washington, D.C.) 224*, 161–164.

Fowlkes, D. M., Mullis, N. T., Comeau, C. M., & Crabtree, G. R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2313–2316.

Francis, C. W., Marder, V. J., & Martin, S. E. (1980) *J. Biol. Chem. 255*, 5599–5604.

Geier, G. E., & Modrich, P. (1979) *J. Biol. Chem. 254*, 1408–1413.

Ghosh, P. K., Reddy, V. B., Swinsoe, J., Lebowitz, P., & Weissman, S. M. (1978) *J. Mol. Biol. 126*, 813–846.

Grieninger, G., Hertzberg, K. M., & Pindyck, J. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5506–5510.

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

Henschen, A., Lottspeich, F., Sekita, T., & Warbinek, R. (1976) *Hoppe-Seyler's Z. Physiol. Chem. 357*, 605–608.

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

Henschen, A., Lottspeich, F., Topfer-Petersen, E., Kehl, M., & Timpl, R. (1980) *Protides Biol. Fluids 28*, 47–50.

Hoeprich, P. D., & Doolittle, R. F. (1983) *Biochemistry 22*, 2049–2055.

Inman, A. M. A., Eaton, M. A. W., Williamson, R., & Humphries, S. (1983) *Nucleic Acids Res. 11*, 7427–7434.

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

Kant, J. A., Lord, S. T., & Crabtree, G. R. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3953–3957.

Kloczewiak, M., Timmons, S., Lukas, T. J., & Hawiger, J. (1984) *Biochemistry 23*, 1767–1774.

Koj, A. (1974) *Struct. Funct. Plasma Proteins 1*, 73–131.

Long, G. L., Chandra, T., Woo, S. L. C., Davie, E. W., & Kurachi, K. (1984) *Biochemistry 23*, 4828–4837.

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

Luse, D. S., Haynes, J. R., VanLeeuwen, D., Schon, E. A., Cleary, M. L., Shapiro, S. G., Lingrel, J. B., & Roeder, R. G. (1981) *Nucleic Acids Res. 9*, 4339–4354.

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

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–701.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

McClelland, M. (1981) *Nucleic Acids Res. 9*, 5859–5866.

McDevitt, M. A., Imperiale, M. J., Ali, H., & Nevins, J. R. (1984) *Cell (Cambridge, Mass.) 37*, 993–999.

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

Mendecki, J., Lee, S. Y., & Brawerman, G. (1972) *Biochemistry 11*, 792–798.

Messing, J., Crea, R., & Seeburg, P. H. (1981) *Nucleic Acids Res. 9*, 309–321.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.

Nickerson, J. M., & Fuller, G. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 303–307.

Olexa, S. A., & Budzynski, A. Z. (1981) *Thromb. Haemostasis 46*, 161.

Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., & Surrey, S. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4298–4302.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 236*, 211–214.

Pustell, J., & Kafatos, F. C. (1982) *Nucleic Acids Res. 10*, 4765-4782.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Sharp, P. A. (1981) *Cell (Cambridge, Mass.) 23*, 643-646.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503-517.
Staden, R. (1977) *Nucleic Acids Res. 4*, 4037-4051.
Watt, K. W. K., Takagi, T., & Doolittle, R. F. (1979) *Biochemistry 18*, 68-76.
Wolfenstein-Todel, C., & Mosesson, M. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5069-5073.
Wolfenstein-Todel, C., & Mosesson, M. W. (1981) *Biochemistry 20*, 6146-6149.
Woo, S. L. C. (1979) *Methods Enzymol. 68*, 389-395.

Solvent Effects on the Stability of A_7U_7p†

David R. Hickey and Douglas H. Turner*
Department of Chemistry, University of Rochester, Rochester, New York 14627
Received July 16, 1984; Revised Manuscript Received November 15, 1984

ABSTRACT: The thermodynamics of double-helix formation were measured spectrophotometrically for A_7U_7 in water at 1 M NaCl and for A_7U_7p in a variety of solvent mixtures and salt. Comparison of the A_7U_7 results with calorimetric measurements indicates duplex formation involves intermediate states. For A_7U_7p between 0.06 and 0.55 M Na^+, $dT_m/d(\log [Na^+]) = 17.4$ °C, similar to the value of 19.6 °C for poly(A)·poly(U) [Krakauer, H., & Sturtevant, J. M. (1968) *Biopolymers 6*, 491-512]. At 1 M NaCl, the A_7U_7p duplex is most stable in 100% water. For 10 mol % solutions, the order for A_7U_7p duplex stability is ethylene glycol > glycerol > ethanol > 2-propanol > dimethyl sulfoxide > 1-propanol > formamide > *N*,*N*-dimethylformamide > urea > dioxane. Comparison of changes in stability and thermodynamic parameters with literature results for proteins suggests proteins and A_7U_7p interact differently with solvent. The results suggest hydrophobic bonding is not a major contributor to the stability of the A_7U_7p duplex. Comparisons with bulk solvent surface tension suggest the energy of cavity formation is also not a major contributor to duplex stability.

Solvent is thought to make important contributions to the stabilities of nucleic acids (Cantor & Schimmel, 1980; Bloomfield et al., 1974). It has been suggested that either classical hydrophobic bonding (Kauzmann, 1959; Tanford, 1973) or the energies of solvent cavities (Sinanoglu & Abdulnur, 1964, 1965; Sinanoglu, 1968, 1980, 1982) drive formation of double helices. The effects of solvent and the environment on stabilities of nucleic acids have implications for predicting the structures and properties of nucleic acids in both aqueous and partially aqueous environments. The latter are of increasing importance. For example, many RNA–protein complexes are being discovered (Kole et al., 1980; Stark et al., 1978; Lerner & Steitz, 1981; Walter & Blobel, 1982, 1983); a powerful new method for detecting sequence changes in DNA depends on denaturation by cosolvents (Lerman et al., 1984; Fischer & Lerman, 1983); many hybridization experiments are conducted on solid-phase supports. Despite the importance of understanding environmental effects on nucleic acids, there is relatively little experimental data available (Levine et al., 1963; Lowe & Schellman, 1972; Herskovits & Harrington, 1972; Herskovits & Bowen, 1974; Breslauer et al., 1978; Dewey & Turner, 1980; Freier et al., 1981; Albergo & Turner, 1981). This paper reports the thermodynamics of duplex formation by A_7U_7p in water and aqueous cosolvent mixtures. The results in water have implications for deriving thermodynamic parameters useful in predicting RNA structure (Tinoco et al., 1971, 1973; Nussinov et al., 1982; Nussinov & Tinoco, 1981; Pipas & McMahan, 1975; Salser, 1977; Zuker & Stiegler, 1981; Auron et al., 1982; Borer et al., 1974; Gralla & Crothers, 1973). The results from solvent perturbations suggest hydrophobic and solvent cavity effects are relatively unimportant in determining nucleic acid stability.

Materials and Methods

Synthesis of A_7U_7p. A_7U_7p was synthesized in three steps. All reactions were monitored by high-performance liquid chromatography (HPLC)[1] as outlined by Petersheim & Turner (1983).

Poly(U) (Sigma) at 15 mg/mL was dialyzed 4 times against 10 mM NaCl and 10 mM Tris, pH 7.5, for 12 h each. The first dialysis solution also contained 10 mM EDTA. The poly(U) was hydrolyzed with 0.9 M KOH (Borer, 1972; Martin et al., 1971) at 0 °C for 4 h and neutralized with concentrated $HClO_4$, producing $KClO_4$ precipitate. The supernatant and washings were combined, the pH was lowered to 3 with 1 M HCl, and the solution was incubated for 3 h at 37 °C to break cyclic phosphates. The solution was brought to pH 7, and the products were isolated. $p(Up)_7$ was prepared by incubating $(Up)_7$ with 15 units/mL T4 polynucleotide kinase for 6 h at 37 °C (Uhlenbeck & Cameron, 1977).

† This work was supported by National Institutes of Health Grant GM 22939.

[1] Abbreviations: ATP, adenosine 5′-triphosphate; BSA, bovine serum albumin; CD, circular dichroism; DEAE, diethylaminoethyl; DMF, *N*,*N*-dimethylformamide; Me_2SO, dimethyl sulfoxide; DSC, differential scanning calorimetry; EDTA, ethylenediaminetetraacetic acid; EtOH, ethanol; Form, formamide; HPLC, high-performance liquid chromatography; poly(U), poly(uridylic acid); 1-PrOH, 1-propanol; 2-PrOH, 2-propanol; TEAB, triethylammonium bicarbonate; TEACl, tetraethylammonium chloride; TMACl, tetramethylammonium chloride; Tris, tris(hydroxymethyl)aminomethane; UV, ultraviolet.

0006-2960/85/0424-2086$01.50/0