Sausville, E. A., Peisach, J., & Horwitz, S. B. (1976) *Biochem. Biophys. Res. Commun. 73*, 814-822.
Sausville, E. A., Peisach, J., & Horwitz, S. B. (1978a) *Biochemistry 17*, 2740-2746.
Sausville, E. A., Stein, R. W., Peisach, J., & Horwitz, S. B. (1978b) *Biochemistry 17*, 2746-2754.
Shirakawa, I., Azegami, M., Ishii, S., & Umezawa, H. (1971) *J. Antibiot. 24*, 761-766.
Strong, J. E., & Crooke, S. T. (1979) in *Bleomycin: Chemical, Biochemical and Biological Aspects* (Hecht, S. M., Ed.) pp 244-254, Springer-Verlag, New York.
Sugiura, Y. (1979) *Biochem. Biophys. Res. Commun. 90*, 375-383.
Sugiura, Y. (1980) *J. Am. Chem. Soc. 102*, 5208-5215.
Suzuki, H., Nagai, K., Yamaki, H., Tanaka, N., & Umezawa, H. (1969) *J. Antibiot. 22*, 446-448.
Takahashi, K., Yoshioka, O., Matsuda, A., & Umezawa, H. (1977) *J. Antibiot. 30*, 861-869.
Umezawa, H. (1971) *Pure Appl. Chem. 28*, 665-680.
Umezawa, H. (1973) *Biomedicine 18*, 459-475.
Umezawa, H. (1974) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 33*, 2296.
Umezawa, H. (1977) *Lloydia 40*, 67-81.
Umezawa, H. (1979) in *Bleomycin: Chemical, Biochemical and Biological Aspects* (Hecht, S. M., Ed.) pp 24-36, Springer-Verlag, New York.
Umezawa, H., Maeda, K., Takeuchi, T., & Okami, Y. (1966) *J. Antibiot., Ser. A 19*, 200-209.
Umezawa, H., Asakura, H., Oda, K., Hori, S., & Hori, M. (1973) *J. Antibiot 26*, 521-527.
White, R. E., & Coon, M. J. (1980) *Annu. Rev. Biochem. 49*, 315-356.

Lamprey Fibrinogen γ Chain: Cloning, cDNA Sequencing, and General Characterization†

D. D. Strong, M. Moore, B. A. Cottrell, V. L. Bohonus, M. Pontes, B. Evans, M. Riley, and R. F. Doolittle*

Department of Chemistry, University of California, San Diego, La Jolla, California 92093
Received June 22, 1984

ABSTRACT: A cDNA library from lamprey liver was constructed in pBR322 and screened with a synthetic mixed oligonucleotide probe, the sequence of which was based on a partial amino acid sequence of the lamprey fibrinogen γ chain determined by conventional procedures. Among the positive clones was one containing a 600-base insert that covered the carboxy-terminal third of the chain and another with a 1950-base insert that stretched more than full length. The two inserts were sequenced by the Maxam–Gilbert procedure. The DNA sequencing was corroborated by reference to the amino acid sequences of five cyanogen bromide peptides that compose the carboxy-terminal 130 amino acids, as well as to a number of tryptic peptides from elsewhere in the molecule. The clone with the smaller insert (6G) contained 594 nucleotides (not counting G and C tails), 435 of which are coding and correspond to residues 264–408 of the γ chain. The remaining 159 nucleotides included the terminator codon followed by a noncoding segment. The larger clone (2E) coded for 408 amino acids that could be readily aligned with the 411-residue human γ chain. A 24-residue signal peptide adjacent to the proposed amino terminal was also inferred. The amino acid sequence of the fibrinogen γ chain has been differentially conserved during evolution, the lamprey and human sequences being more than 70% identical in certain key regions but dropping to less than 25% in other sections, including the segment thought to be a part of the "coiled coils". Overall, the resemblance amounts to 50% identity. Of the 10 cysteines found in mammalian chains, 9 are at identical positions, but the tenth, which in mammalian fibrinogens is a part of the interdimeric bridging, is absent in the lamprey.

Vertebrate blood coagulation centers on the conversion of a soluble plasma protein, fibrinogen, into an insoluble gel, fibrin. In all species examined, from fish to mammals, the fibrinogen molecule consists of three pairs of nonidentical chains interconnected by a network of disulfide bonds. Because the lamprey is one of the most primitive of extant vertebrates, considerable attention has been focused on its fibrinogen (Doolittle, 1965; Murtaugh et al., 1974; Doolittle & Wooding, 1974; Cottrell & Doolittle, 1976). Although the fundamental plan of lamprey fibrinogen is the same as that found in mammals, there are some unique and interesting features, including the ability to clot upon the exclusive removal of its fibrinopeptide B, a moiety that in lampreys contains carbohydrate (Doolittle & Cottrell, 1974). There is no immunological cross-reactivity between lamprey and human fibrinogens, and the amino acid compositions of the various chains are recognizably different (Doolittle et al., 1976). Aspects which have been conserved throughout vertebrate evolution include the existence of two thrombin-released fibrinopeptides, a polymerization scheme that is inhibited by Gly-Pro-Arg peptides (Laudano & Doolittle, 1980), and a factor XIII catalyzed stabilization system (Doolittle & Wooding, 1974; Murtaugh et al., 1974).

We have pursued the study of lamprey fibrinogen over the years in the hope not only that clues would be revealed about the evolution of the molecule but also that the mechanisms of action involved in fibrin formation might be inferred from

† This work was supported by a grant from the National Institutes of Health (HL 26873).

0006-2960/85/0424-0092$01.50/0

constant or immutable features. Originally, it was our intention to determine the amino acid sequence of the protein by conventional procedures, and we were well under way with the project before the advantages of recombinant DNA procedures became so well-known, at which point we switched our strategy. We have now constructed a lamprey liver cDNA library and can report the cloning of the lamprey fibrinogen γ chain from its mRNA. The amino acid sequence data accumulated previously, amounting to most of the 130 residues at the carboxy terminus as well as numerous small tryptic peptides from scattered regions, were used for confirming the accuracy of the cDNA sequence, as well as for locating certain key features such as the amino terminus, carbohydrate attachment point, and fibrin cross-linking sites. These data were also essential for the construction of an oligonucleotide probe for identifying suitable clones.

Materials and Methods

Lampreys. Lampreys (*Petromyzon marinus*) were collected during their annual spawning runs from various New Hampshire rivers and streams. Blood was drawn from the caudal vein after severance of the tail; the anticoagulant employed was 0.1 M trisodium citrate, 1:10 by volume. Livers were removed at stream side and dropped into liquid nitrogen. After cross-country shipment on dry ice, livers were stored at −70 °C until use. Lamprey fibrinogen was prepared as described previously (Doolittle et al., 1976).

Protein Sequencing. Most of the materials and methods used in this study have been described in detail in previous publications from this laboratory, including isolation of the individual chains, cyanogen bromide (CNBr)[1] cleavage, gel filtration of the CNBr digests (in this case on Sephadex G-100), conditions for enzyme digestions, amino-terminal determinations, and the attachement of peptides to aminoethylaminopropyl glass beads (Takagi & Doolittle, 1975; Doolittle et al., 1976; 1977; Watt et al., 1979). Peptides were purified either by paper electrophoresis or by HPLC on a Beckman Model 332 liquid chromatograph equipped with either a 5-μm Ultrasphere-ODS (4.5 × 250 mm Altex) or a Vydac 214 TP510 reverse-phase column. In both cases, the starting solvent (A) was 0.05% F_3AcOH/H_2O; the gradient solvent (B) was acetonitrile (MCB, HPLC grade). Peptide sequences were determined mainly by sequencer runs on an automatic sequencer of our own design (Doolittle, L.et al., 1977) that employs a thioacetylation procedure with (thioacetyl)thioglycolic acid as the coupling agent (Mross & Doolittle, 1977). In some cases, a sequencing procedure using DABITC (Chang et al., 1978) was also used on peptides attached to glass beads in order to distinguish aspartic acid from asparagine and glutamic acid from glutamine. Digestion of the amino-terminal peptide with PCA peptidase was conducted according to the procedure described by Doolittle (1972).

Oligonucleotide Probes. Oligodeoxyribonucleotide probes were synthesized by using the phosphotriester method developed by Itakura and co-workers (Broka et al., 1980; Ito et al., 1982; Miyoshi et al., 1980a,b). The four fully protected monomers were synthesized from commercially available deoxynucleosides (Vega) by using conventional methods (Narang et al., 1980; Broka et al., 1980). A library of fully protected dimers was then synthesized either by the method of Broka et al. (1980) or with the coupling reagent 1-(2,4,6-triisopropylbenzenesulfonyl)-3-nitro-1,2,4-triazole (de Rooij et al., 1979). Oligomers were assembled starting at the 3′ end by using 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole as the coupling reagent and commercially available (Bachem) starter resins (1% polystyrene resin with the starting dimethoxytrityl deoxynucleoside already attached). Detritylation was carried out with benzenesulfonic acid. Completed oligomers were cleaved from the resin with tetramethylguanidine *syn-p*-nitrobenzaldoximate; base-protecting groups were removed with concentrated ammonium hydroxide, and the material was purified on Sephadex G-50 equilibrated and eluted with 0.01 M triethylammonium bicarbonate, pH 7. The purified oligomers were then detritylated with 80% acetic acid, extracted with ether, and lyophilized (Miyoshi et al., 1980a; Ito et al., 1982). This lyophilized material was suitable for [γ-^{32}P]ATP labeling with T4 polynucleotide kinase.

Labeling and Purification of Oligonucleotides. Oligonucleotides were labeled at the 5′ end for sequencing or for use as hybridization probes by transfer of [^{32}P]phosphate from [γ-^{32}P]ATP in the presence of T4 polynucleotide kinase. The desired radiolabeled oligonucleotides were separated from [γ-^{32}P]ATP and shorter failure sequences by polyacrylamide slab gel electrophoresis using standard sequencing gels (Maxam & Gilbert, 1980). The gel slices containing the DNA of interest were eluted in water and eluates used directly for hybridization. The sequences of DNA probes were verified by a modified Maxam–Gilbert procedure, the important differing features of which were significantly longer reaction times and higher temperatures for the base modifications.

Cloning and Recombinant DNA Materials. All restriction endonucleases, *Escherichia coli* DNA polymerase I, terminal deoxynucleotidyl transferase, dG-tailed *Pst*I-cleaved pBR322, T4 kinase, bacterial and calf alkaline phosphatases, RPC-5 Analog, and oligo(dT)–cellulose were purchased from Bethesda Research Laboratories (Rockville, MD). AMV reverse transcriptase was obtained from J. Beard, Life Sciences (St. Petersburg, FL). Unlabeled deoxynucleotide triphosphates and oligo(dT) were purchased from P-L Biochemicals, Inc. (Milwaukee, WI), nuclease S1 was from Miles Laboratories (Kankakee, IL), and yeast RNA was from Worthington (Freehold, NJ). Proteinase K and calf intestine alkaline phosphatase were obtained from Boehringer Mannheim (Indianapolis, IN). Tetracycline was purchased from Calbiochem (San Diego, CA), nitrocellulose paper BA85 was from Schleicher & Schuell (Keene, NH), and [γ-^{32}P]ATP (>7000 Ci/mmol), [α-^{32}P]ddATP (>3000 Ci/mmol), and [^{3}H]dCTP (10–20 Ci/mmol) were from Amersham or ICN.

RNA Preparation. Total lamprey liver nucleic acid was prepared by standard phenol/chloroform/alcohol extraction procedures, including the treatment of tissue homogenates with proteinase K (Shields & Blobel, 1977). Total RNA was separated by precipitation with 3 M sodium acetate, pH 6 (Kern, 1975). Approximately 10 mg of RNA and 6 mg of high molecular weight DNA were obtained per g of liver. Polyadenylated RNA was separated from other nucleic acids by affinity chromatography on oligo(dT)–cellulose (Aviv & Leder, 1972). The yield of polyadenylated RNA was approximately 1.5% of the total nucleic acid applied to the column.

Preparation of Homopolymer-Tailed Double-Stranded cDNA. Double-stranded cDNA was prepared as described

[1] Abbreviations: CNBr, cyanogen bromide; F_3AcOH, trifluoroacetic acid; Tris, 2-amino-2-(hydroxymethyl)-1,3-propanediol; TATG, (thioacetyl)thioglycolic acid; DNS, 5-(dimethylamino)naphthalene-1-sulfonyl; PCA, pyrrolidonecarboxylic acid; DABITC, diaminobenzil isothiocyanate; HPLC, high-performance liquid chromatography; EDTA, ethylenediaminetetraacetic acid; $NaDodSO_4$, sodium dodecyl sulfate; nt, nucleotide.

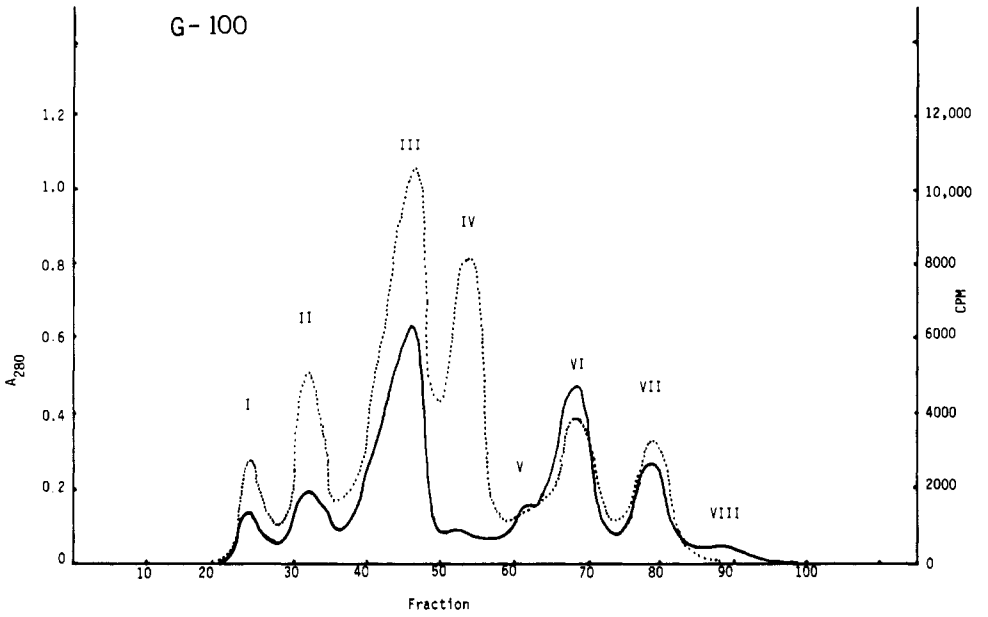

FIGURE 1: Gel filtration of the CNBr digest of 50 mg of lamprey fibrinogen chain. The material had been reduced and alkylated with iodo[^{14}C]acetic acid. The column (2.5 × 80 cm) was equilibrated and developed with 10% acetic acid. Pools I and II were incomplete digestion products and were not further characterized. Pools III–VII were characterized by procedures described in the text. Fraction size = 6 mL. The solid line denotes A_{280}; the dotted line indicates radioactivity.

previously by others (Wickens et al., 1978; Crabtree & Kant, 1981). The yield of double-stranded cDNA from 20 μg of RNA was 1.3 μg. After fractionation on Bio-Gel A-150 M, the 3′ ends of the double-stranded cDNA were "tailed" with ca. 15 dCMP residues as described by Michelson & Orkin (1982) except that the final [^{3}H]dCTP concentration was 60 μM and the reaction was carried out at 37 °C. Carrier yeast RNA was added prior to phenol extraction and ethanol precipitations. The tailed product was dissolved in annealing buffer (0.01 M Tris-HCl, pH 7.6, 0.1 M NaCl, and 0.001 M EDTA) to concentrations of 2–5 ng/μL.

Transformation of E. coli. Approximately equimolar amounts of dGMP-tailed pBR322 (75 ng) and dCMP-tailed double-stranded cDNA (5 ng) were mixed in annealing buffer, heated to 65 °C for 10 min, moved to a 45 °C environment for 2 h, and cooled on ice for 1 h. The solution was adjusted to 0.1 M $CaCl_2$ just prior to transformation. Transformation of HB101 cells was performed according to the method of Dagert & Ehrlich (1979). Transformation was optimum when the bacteria were incubated in ice-cold 0.1 M $CaCl_2$ for 20 h. Recombinant clones were selected on L-agar plates containing 10 μg/mL tetracycline. Transformation efficiencies of 4 × 10^4 transformants per μg of hybrid DNA were obtained. All told, 2832 transformants were individually picked on to fresh L-agar tetracycline plates into ordered arrays. These recombinants were transferred to microtiter dishes and stored in 15% glycerol at −70 °C.

Cloning of recombinant DNA in *E. coli* was carried out in accordance with the guidelines for recombinant DNA research.

Screening of the cDNA Library. Colony hybridization on Whatman 541 paper was carried out as described by Gergen et al. (1979). Large ordered arrays (432 individual recombinants per paper) were screened in duplicate with the radiolabeled synthetic oligonucleotide probe. Hybridization was performed at 45–50 °C in Sears Seal-a-Meal bags with 90 ng of probe per filter (specific activity 6 × 10^7 cpm/μg) as described by Wallace et al. (1981). Filters were washed at 40–50 °C at low ionic strength prior to autoradiography.

Small quantities of plasmid DNA from hybridization-positive clones were prepared rapidly for screening as described by Maniatis et al. (1982). Plasmid DNA was digested with appropriate restriction endonucleases, electrophoresed on agarose gels, transferred to nitrocellulose paper, and hybridized with labeled synthetic oligonucleotides (Southern, 1975; Wallace et al., 1981). The hybridization and washing temperatures were 45 and 40 °C, respectively.

DNA Sequence Determination. Large quantities of plasmid DNA were prepared as described by Norgard et al. (1979); RNA was separated from plasmid DNA by RPC-5 Analog chromatography (Thompson et al., 1983). In some cases, the insert was released from the recombinant plasmid by digestion with *Pst*I and isolated from polyacrylamide or agarose gels (Maniatis et al., 1982). DNA restriction fragments were labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase (Richardson, 1965) or at the 3′ end by the action of terminal transferase and [γ-^{32}P]ddATP following the specifications of the manufacturer (Amersham). Fragments were sequenced by the Maxam–Gilbert procedure (Maxam & Gilbert, 1980; Rubin & Schmid, 1980). The following chemical modifications were employed: for the "G" reaction, 0.1% dimethyl sulfate for 4.5 min at 20 °C; "G + A" reaction, 67% formic acid for 4.5 min at 20 °C; "C + T" reaction, 67% hydrazine for 5 min at 20 °C; "C" reaction, 67% hydrazine/1.7 M NaCl; "A > C" reaction, 2 N NaOH for 5 min at 90 °C; "T" reaction, boil in H_2O for 2 min and add $KMnO_4$ to 26 μM. For fragments greater than 500 base pairs, all times were decreased by 1 min. Fragments were examined on 8% acrylamide gels (run times of 1.5, 3, and 4.5 h) and 20% acrylamide gels (1.5 h).

RESULTS

Amino Acid Sequencing of CNBr Fragments. Lamprey γ chains were subjected to CNBr treatment, and the digest was fractionated on Sephadex G-100 (Figure 1). The pools containing smaller peptides were purified further by paper electrophoresis or HPLC. Five such peptides accounting for 130 residues were isolated and characterized; their sequences were readily aligned with the carboxy-terminal third of the human chain. A summary of the data collected in determining the sequences of these peptides is presented in Figure 2.

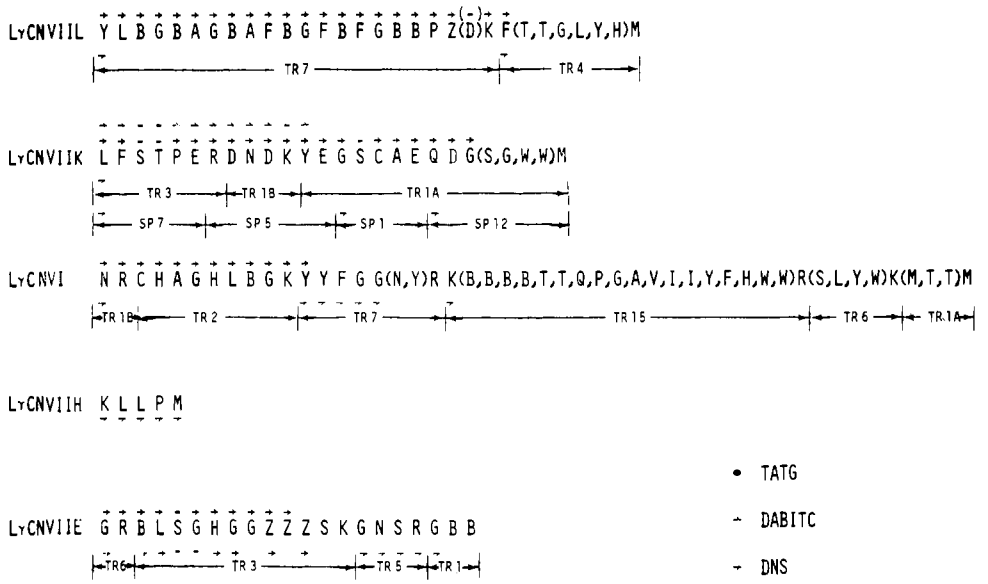

FIGURE 2: Summary of data used in determining partial amino acid sequences of five CNBr peptides (from pools V–VII in Figure 1). The peptides, which account for 130 amino acids total, are listed in the order in which they occur in an alignment with the human γ chain. Tryptic (TR) and staphylococcal protease (SP) subpeptides are set off by arrows.

Table I: Some Tryptic Peptides Isolated from Major CNBr Fragments[a]

pool CNIII			pool CNIV		
peptide	sequence	residue no.[b]	peptide	sequence[c]	residue no.[b]
T3f	I(S)PI(T)GK	146–152	T7a	QVR	1–3
T4j	AK	175–176	T7d	DLK	4–6
T7a	DWSYR	207–212	T11c	TAK	31–33
T5	LR	244–245	T11d	ITR	69–71
T3c	IBLTD(W)Z(H)R	246–256	T7c	B(SZG)R	70–83
T6b	YABYGHFK	257–264	T10	TVQK	84–87
T6a	L(SYF)hs	274–278			

[a] See Figure 1 for source of pools CNIII and CNIV. [b] Residue numbers refer to corresponding sequences identified from cDNA (Figure 6). [c] The amino terminus for sequence QVR is blocked, presumably because glutamine (Q) has cyclized to PCA.

Not every residue was absolutely determined in every peptide, only about three-fourths of the 130 residues being firmly established before these efforts were overtaken by the DNA sequencing results. The data were judged to be more than sufficient to ensure that all the DNA sequencing was being interpreted accurately, however. The DNA data also resolved a problem of a "missing" CNBr peptide when the sequence Met-Thr-Thr-Met was revealed at residues 379–382, the Met–Thr bond being resistant to CNBr fragmentation.

Two major peptide fragments, CNIII and CNIV (Figure 1), accounting for the remaining 278 residues, were also characterized. Their apparent molecular weights as determined by $NaDodSO_4$ electrophoresis were 20 000 and 13 000, respectively. They were subjected to amino acid and amino-terminal analysis; the larger fragment contained carbohydrate. Neither of the major CNBr fragments revealed an amino terminus, leaving unsettled the problem of the blocked amino terminus of the lamprey γ chain noted previously (Doolittle et al., 1976). The DNA sequence subsequently revealed that the larger fragment has a tryptophan at its amino terminus. The two large CNBr fragments were digested with trypsin and several small peptides isolated and sequenced (Table I). In the case of the 13 000 molecular weight fragment, a ninhydrin-negative tryptic peptide was isolated which had the composition (Gln,Val)-Arg.

Fibrin Cross-Linking Sites. In some cases, lamprey fibrin was used which had been labeled at the γ-chain cross-linking site by the factor XIII catalyzed incorporation of a radioactive substitute donor, [^{14}C]glycine ethyl ester. The radioactivity was found to be exclusively associated with fragment CNVIIE. Stepwise degradation revealed that the bulk of the label resided on the glutamine at residue 10 of this peptide, which is the 12th residue from the carboxy terminus of the lamprey γ chain (Figure 2).

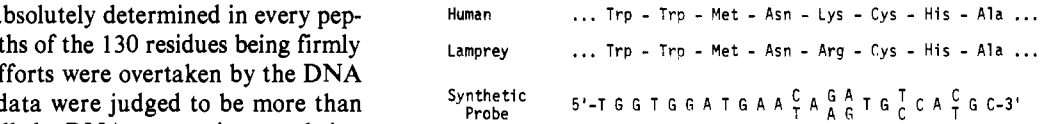

FIGURE 3: Amino acid sequence of portions of CNVIIK and CNVI aligned with a corresponding portion of the human sequence (Henschen & Lottspeich, 1977) and showing the sequence of the synthetic oligonucleotide probe. Note that not all possible codons for arginine were utilized in the mixed 23-mer.

Colony Screening and DNA Sequencing. Approximately 3000 transformants were screened with a synthetic mixed oligonucleotide probe which was 23 nucleotides in length and 32-fold degenerate (Figure 3). A total of 50 colonies were identified that bound the probe significantly under reasonably stringent colony hybridization conditions (Figure 4). These were rescreened by Southern blotting (Southern, 1975). In this regard, small quantities of *Eco*RI or *Pst*I restricted plasmid DNA were transferred to nitrocellulose filters after gel electrophoresis. Of these, 10 clones with internal *Eco*RI sites bound the probe. Large quantities of the two plasmids which showed the strongest binding (p420R6G and p423R2E) were prepared for restriction mapping and DNA sequence determination.

Clone p420R6G. The plasmid p420R6G contained a cDNA insert of approximately 600 base pairs. DNA from the plasmid

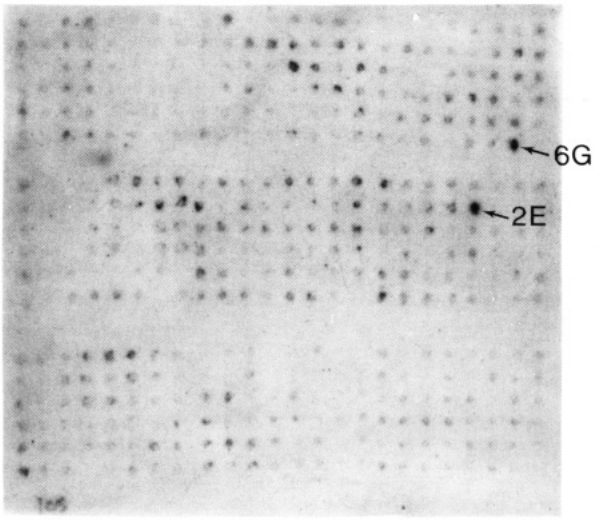

FIGURE 4: Filter paper screening of transformants with ^{32}P-labeled synthetic oligonucleotide 23-mer. Arrows denote clones R6G and R2E.

and the insert was digested with restriction enzymes singly or in combination to construct a detailed restriction map (Figure 5). On the basis of this map, several sites were chosen for end labeling and sequencing. An overlapping sequence of both strands of the cDNA insert was determined which corresponded to the carboxy-terminal 145 amino acids of the lamprey chain, the TGA stop codon, and an additional 156 nucleotides of the 3' noncoding region.

Clone p423R2E. During the period we were sequencing the clone 6G, we isolated a second clone that had a much larger insert. Digestion with *Pst*I released insert fragments of 1400, 380, and 180 nt, respectively. A number of subfragments were subsequently obtained by a variety of digests (Figure 5). Segments corresponding to the entire γ chain and an inferred signal peptide were sequenced, as well as most of the 5′ and 3′ noncoding regions (Figure 6). In the case of the latter, the 2E sequence was identical with that found for 6G, although it stretched some 300 bases further downstream, at which point it reaches the poly(A) segment of the message. All told, there were 432 amino acid codons from the logical initiator to the first terminator. The predicted amino acid composition for the 408-residue mature chain is in excellent agreement with the published analysis (Doolittle et al., 1976).

Identification of the γ-Chain Amino Terminus. Fibrinogen γ chains are ordinarily made as precursors with signal peptides that are released during the maturation process (Nickerson & Fuller, 1981). As such, the ultimate amino terminus cannot be determined from the cDNA sequence alone. The fact that the lamprey γ chain has a blocked amino terminus further complicated the situation. As noted above, a tryptic peptide had been isolated that had the composition (Gln, Val)-Arg. The peptide was neutral at pH 6.5 and gave no amino terminus by the dansyl reaction. The composition matches the tripeptidyl sequence Gln-Val-Arg at the expected location for the amino terminus as inferred from the cDNA sequence. It is likely that the block is due to cyclization of the glutamine after removal of the signal peptide, but we must report that digestion with PCA peptidase on two occasions failed to generate the expected products. The unlikely possibility that the peptide is acetylated cannot be excluded, however.

Initiator Codon and Signal Peptide. The inferred amino acid sequence contains a stretch of 19 nonpolar amino acids immediately before the glutamine residue thought to be the amino terminus. Moreover, a potential ATG initiator codon is situated appropriately in front of the presumed signal.

Carbohydrate Attachment Point. The lamprey γ chain amino acid sequence has only one potential asparaginyl carbohydrate attachment site, the sequence Asn-Phe-Thr oc-

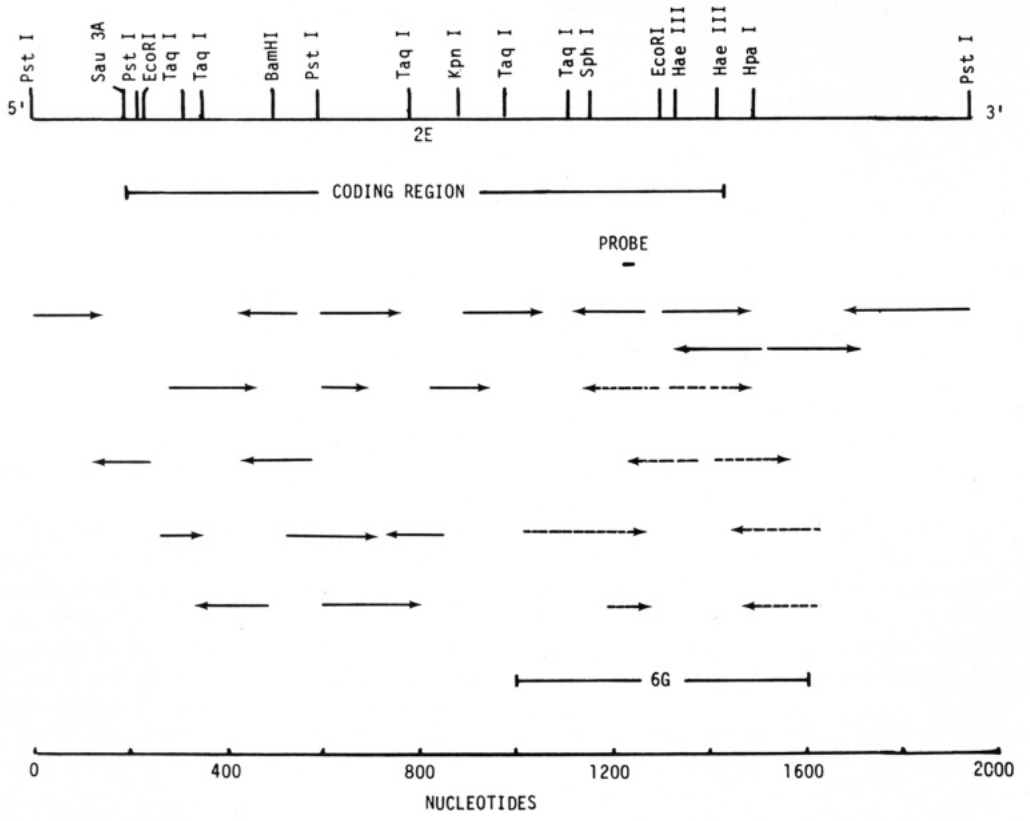

FIGURE 5: Overlapping fragments used to determine the DNA sequences of clones 2E (solid lines) and 6G (broken lines). The segment corresponding to the synthetic oligonucleotide probe is also marked.

```
                                                                                                                G GGG GGG GGG GGG GGG GGG GGC AGA GTC GAG TGA GGA

GGA ACA TCG CTG ACT AAC ACA GTT AAT CAA ACT GTT GGT TCG TGG CTG TGT ACG TGT GTA TAG TGC ACA CGG GGG TAC TGC ATT TGT CTG TTA TAC ATC GAC ACA

GGC AGA CGA GGC AAG AGT GAT ACA AAC ATG GGC AGG ATC GGG ACG CCC GTC TTC CTC GCC TTC TTG TCC GCG CTT ACG TGC TCC CTG CAG GTT CAC GCG CAG GTC
                                  Met Gly Arg Ile Gly Thr Pro Val Phe Leu Ala Phe Leu Ser Ala Leu Thr Cys Ser Leu Gln Val His Ala Gln Val
                                  -24             -20                                     -10                                    -1  +1

AGG GAC CTC AAG CAG TGC TCC AAT GAT CCC GAA TTC GGC CGC TAC TGC CCC ACG ACG TGC GGG GTG GCG GAC GTG CTG TCG AAG TAC GCC AAG GGC GTG GAC GAG
Arg Asp Leu Lys Gln Cys Ser Asn Asp Pro Glu Phe Gly Arg Tyr Cys Pro Thr Thr Cys Gly Val Ala Asp Val Leu Ser Lys Tyr Ala Lys Gly Val Asp Glu
                              10                                  20                                  30

GAC TCG AGT TTC ATC GAC TCG GTG CTC ACG CAG CTC GCG GCC AAG CAC GGC ATC GTG GAG GGC AAC GTG AAC ATC GTG AAC GAG GAC GTG CGA ATC ACG CGC GAC
Asp Ser Ser Phe Ile Asp Ser Val Leu Thr Gln Leu Ala Ala Lys His Gly Ile Val Glu Gly Asn Val Asn Ile Val Asn Glu Asp Val Arg Ile Thr Arg Asp
         40                                  50                                  60                                  70

GAG GCG CAG ATC ATC AAG GAC TCG GGC CAG AAG ACA GTG CAG AAA ATT CTG GAG GAG GTG CGG ATC CTG GAG CAG ATT GGA GTC AGC CAC GAC GCC CAA ATC CAG
Glu Ala Gln Ile Ile Lys Asp Ser Gly Gln Lys Thr Val Gln Lys Ile Leu Glu Glu Val Arg Ile Leu Glu Gln Ile Gly Val Ser His Asp Ala Gln Ile Gln
                              80                                  90                                  100

GAG CTG TCA GAG ATG TGG CGC GTG AAC CAG CAG TTC GTG ACG CGA CTG CAG CAG CAG CTC GTT GAC ATC CGG CAG ACG TGC TCA CGC TCC TGC CAG GAC ACG ACA
Glu Leu Ser Glu Met Trp Arg Val Asn Gln Gln Phe Val Thr Arg Leu Gln Gln Gln Leu Val Asp Ile Arg Gln Thr Cys Ser Arg Ser Cys Gln Asp Thr Thr
         110                                 120                                 130                                 140

GCC AAC AAG ATA TCG CCC ATC ACC GGG AAA GAC TGC CAA CAA GTG GTG GAT AAC GGG GGC AAG GAC AGC GGA CTC TAC TAC ATC AAA CCC CTC AAG GCC AAG CAG
Ala Asn Lys Ile Ser Pro Ile Thr Gly Lys Asp Cys Gln Gln Val Val Asp Asn Gly Gly Lys Asp Ser Gly Leu Tyr Tyr Ile Lys Pro Leu Lys Ala Lys Gln
                              150                                 160                                 170

CCT TTC CTG GTC TTC TGC GAA ATC GAG AAT GGC AAT GGC TGG ACC GTC ATC CAG CAT CGT CAC GAT GGC AGC GTG AAC TTT ACG CGC GAC TGG GTG TCG TAC CGC
Pro Phe Leu Val Phe Cys Glu Ile Glu Asn Gly Asn Gly Trp Thr Val Ile Gln His Arg His Asp Gly Ser Val Asn Phe Thr Arg Asp Trp Val Ser Tyr Arg
         180                                 190                                 200         CHO                     210

GAG GGC TTC GGG TAC CTG GCG CCG ACG CTC ACC ACT GAG TTC TGG CTC GGA AAC GAG AAG ATC CAC CTG CTG ACC GGG CAG CAG GCC TAC CGC CTG CGC ATC GAC
Glu Gly Phe Gly Tyr Leu Ala Pro Thr Leu Thr Thr Glu Phe Trp Leu Gly Asn Glu Lys Ile His Leu Leu Thr Gly Gln Gln Ala Tyr Arg Leu Arg Ile Asp
                              220                                 230                                 240

CTC ACC GAC TGG GAG AAC ACG CAC AGG TAC GCG GAC TAT GGA CAC TTC AAG TTG ACG CCG GAG TCT GAC GAG TAC CGC CTC TTC TAC TCC ATG TAC CTT GAC GGT
Leu Thr Asp Trp Glu Asn Thr His Arg Tyr Ala Asp Tyr Gly His Phe Lys Leu Thr Pro Glu Ser Asp Glu Tyr Arg Leu Phe Tyr Ser Met Tyr Leu Asp Gly
         250                                 260                                 270                                 280

GAT GCC GGC AAC GCC TTT GAT GGA TTC GAC TTT GGC GAT GAC CCG CAG GAT AAG TTC TAC ACG ACG CAC CTA GGC ATG CTG TTC TCT ACG CCG GAA CGC GAC AAC
Asp Ala Gly Asn Ala Phe Asp Gly Phe Asp Phe Gly Asp Asp Pro Gln Asp Lys Phe Tyr Thr Thr His Leu Gly Met Leu Phe Ser Thr Pro Glu Arg Asp Asn
                              290                                 300                                 310

GAC AAG TAC GAG GGC TCG TGC GCC GAG CAG GAC GGC TCG GGA TGG TGG ATG AAC CGG TGC CAC GCG GGG CAC CTC AAC GGA AAA TAC TAC TTC GGT GGA AAC TAC
Asp Lys Tyr Glu Gly Ser Cys Ala Glu Gln Asp Gly Ser Gly Trp Trp Met Asn Arg Cys His Ala Gly His Leu Asn Gly Lys Tyr Tyr Phe Gly Gly Asn Tyr
         320                                 330                                 340                                 350

CGC AAG ACC GAT GTG GAA TTC CCA TAT GAT GAT GGC ATC ATC TGG GCC ACG TGG CAC GAC CGC TGG TAT TCG CTC AAG ATG ACC ACC ATG AAG TTG CTG CCC ATG
Arg Lys Thr Asp Val Glu Phe Pro Tyr Asp Asp Gly Ile Ile Trp Ala Thr Trp His Asp Arg Trp Tyr Ser Leu Lys Met Thr Thr Met Lys Leu Leu Pro Met
                              360                                 370                                 380

GGC AGG GAC CTC TCT GGG CAT GGT GGC CAG CAG CAG AGC AAG GGC AAC AGC CGC GGG GAC AAC TGA GCA GGA ACT TGG CTC TCA CCT CCT CAT GAC ACG GCA GCG
Gly Arg Asp Leu Ser Gly His Gly Gly Gln Gln Gln Ser Lys Gly Asn Ser Arg Gly Asp Asn Stop
         390                         X           400  X                      408

AGA TGT TAA CTG CCT CGC CTG GTA TAC AGG AGG TCG TGG GTC CGA TCC CGA CCC TGA CGC CCG CTG CAT ATT TGG AGT TTG CAT GTT CTC CCC TGT TTT TCC CTC

CAA ATT TAG AAA CAT GCG TTG AGA GTA TTT GGC TGC TAT AAA AAA AAA ATG TCC AAC GCG ATA AGC CAA TTT GTT GAG CTG TCA TTG TGG CCA GTC ACT CGG AAG

AGC CAT GGC TAC GTC CCA CAT TCT CCA CCG CAC TGC ACC ACG GAT GGC GAC ACT AGA ATG ATA GTA CGA TAT ATC GTA TGA CAT AGG GTC CAC GAT TTA TCG TTC

CAT CCG TTT TAG AAA GTA TAG TCC CAT CGT GAT TAT CGT GGT TCG TGT TTT TTT GTT GCA TAT ATT TTA TTT ACT GAC AAA ATG GTT ATT AAA TAA ACG TAT TGC

CAT CAA AGT TCT CAA TAA AAA AAA AAA AAA AAA AAC CCC CCC CCC CCC CCC CCC CCC
```

FIGURE 6: Nucleotide sequence of lamprey fibrinogen γ-chain cDNA and the corresponding amino acid sequence as based on sequences of plasmids R2E and R6G. CHO = carbohydrate site; X = putative cross-linking sites.

curring at positions 203–205. In line with this occurrence, CNBr fragment III, which corresponds to residues 113–278, yielded hexosamine upon analysis, as did a tryptic peptide derived from it which had the same composition as the inferred peptide comprising residues 198–206.

Discussion

The cloning of the lamprey γ chain is the first step of a general plan aimed at a complete characterization of fibrinogen from this primitive vertebrate. The primary goal of the study is to determine which features lamprey fibrinogen has in common with its mammalian counterparts. This information is valuable on two counts: first, inferences can be made about structure–function relationships by noting which sequences are conserved and which are not; second, the sequences of the most highly conserved regions are logical candidates for synthetic DNA probes for screening prevertebrates in a search for earlier ancestral genes. This latter approach is aimed at finding how this molecule was invented in the first place (Doolittle, 1983).

The lamprey γ chain has a differentially conserved structure relative to mammalian types, retaining more than 70% identity in certain regions and virtually none in others (Figure 7). The overall resemblance is approximately 50%. This stands in contrast to lamprey hemoglobin, which is only 33% identical with mammalian globin α chains (after the suitable introduction of gaps). Lamprey cytochrome *c*, on the other hand, is more than 80% identical with mammalian cytochrome *c* types. Cytochrome *c* is generally regarded as a very strongly conserved protein.

Comparison of Lamprey and Human γ Chains. All told, the lamprey chain has 205 identical residues of 408 possible comparisons with the human chain (Figure 7). It lacks one residue at either end, and there are two small gaps in its sequence and a single small one in the human sequence. Interestingly, 13 lysines in the human sequence are arginines in the lamprey.

It is instructive to compare the lamprey and human sequence section by section in order to appreciate the relationship between structure and function and the wide variation in the degree of sequence conservation.

Initiator Point and Signal Peptide. An ATG codon occurs at a position 24 residues in advance of the γ-chain amino terminus. The 24 residues include a 19-residue span of nonpolar amino acids and appear to be a typical signal peptide.

FIGURE 7: Alignment of lamprey and human γ-chain sequences. Dots denote identical residues. Arrows denote cross-linking acceptors; lines under and over residues are likely donor sites.

The corresponding signal peptides in rat (Crabtree & Kant, 1982) and human (Chung et al., 1983) fibrinogens are 26 residues in length.

Amino Terminus. The γ chains of mammalian fibrinogens invariably have tyrosine at the amino terminus (Blomback & Yamashina, 1958). In lamprey fibrinogen, however, the mature γ chains apparently has a cyclized glutamine (PCA) at this position, although attempts to unblock the chain with PCA peptidase were unsuccessful.

Central Domain Region. In mammals, the two halves of the fibrinogen dimer are joined by three disulfide bonds: one between the two α chains and two between the two γ chains. The two γ-chain cysteines occur at residues 8 and 9. In the lamprey, the equivalent of residue 8 is a glutamine; therefore, only a single disulfide bond connects the two chains. This result is somewhat contradictory to the report that in human fibrinogen the adjacent cysteines 8 and 9 are bridged in an antiparallel fashion to cysteines 9′ and 8′, respectively (Hoeprich & Doolittle, 1983).

"*Coiled Coils*". It is thought that the three major domains of vertebrate fibrinogen are connected by three-stranded coiled coils [see Doolittle (1984) for a review]. These regions of each chain are delineated by sets of interconnected cysteines denoted "disulfide rings". In lamprey γ chains, the four cysteines involved occur at exactly the same positions as in mammalian γ chains, the number of residues between the two braces being 110, the same as in mammalian fibrinogens. The sequence resemblance in this sector is very low, however, only 27 of the 110 residues being identical. Apparently the structural constraints on a three-stranded coiled coil are sufficiently relaxed that considerable variability is allowed, so long as the characteristic rhythmic polarity is maintained (Doolittle et al., 1978).

Carbohydrate Attachment Point. Fibrinogen γ chains invariably contain carbohydrate (more than a dozen species have been examined). In humans, the single asparaginyl attachment point is at residue 52, about one-third of the way along the coiled coil. Rat γ chains have a putative attachment sequence at the same place. Lamprey γ chains, however, do not. Instead, the cDNA sequences and peptide data both indicate the carbohydrate is attached at Asn-203, a residue that could be located at the most distal point of the terminal domain (Doolittle, 1981).

Terminal Domain Region. The terminal domains of fibrinogen, which in large part correspond to fragments D, mainly comprise two homologous segments of the β and γ chains. In this region, the lamprey and human γ chains have closely related sequences. The resemblance pattern is similar to that observed between β and γ chains (Figure 8). In mammals, the carboxy-terminal region of the γ chain has been implicated in a number of important events. The first of these to be brought to light was the reciprocal intermolecular cross-linking that occurs as a result of the action of factor XIII (Chen & Doolittle, 1971). The carboxy-terminal region is also known to be involved in calcium binding (Nieuwenhuizen & Haverkate, 1983), clumping of certain bacteria (Hawiger et al., 1982a; Strong et al., 1982), platelet aggregation (Hawiger et al., 1982b), and binding of antipolymerant peptides (Laudano & Doolittle, 1980). In this latter regard, other evidence has been reported that also indicates this region is a primary contact site for polymerization (Olexa & Budzynski, 1981).

Given all these important roles, it is not surprising that portions of this region are strongly conserved. This is not to say that significant changes have not occurred in this part of the sequence during the last 450 million years, however. For example, although lamprey γ chains are covalently cross-linked by factor XIII during fibrin stabilization (Doolittle & Wooding, 1974; Murtaugh et al., 1974), the detailed packing arrangement must be different from that in mammals. In mammals, the acceptor glutamine and the donor lysine are eight residues removed from each other, whereas in lamprey the sequence data reveal that the two partners are only four residues removed (Figure 2 and 6). It is known that in mammals the relationship is reciprocal, the lysine of one unit

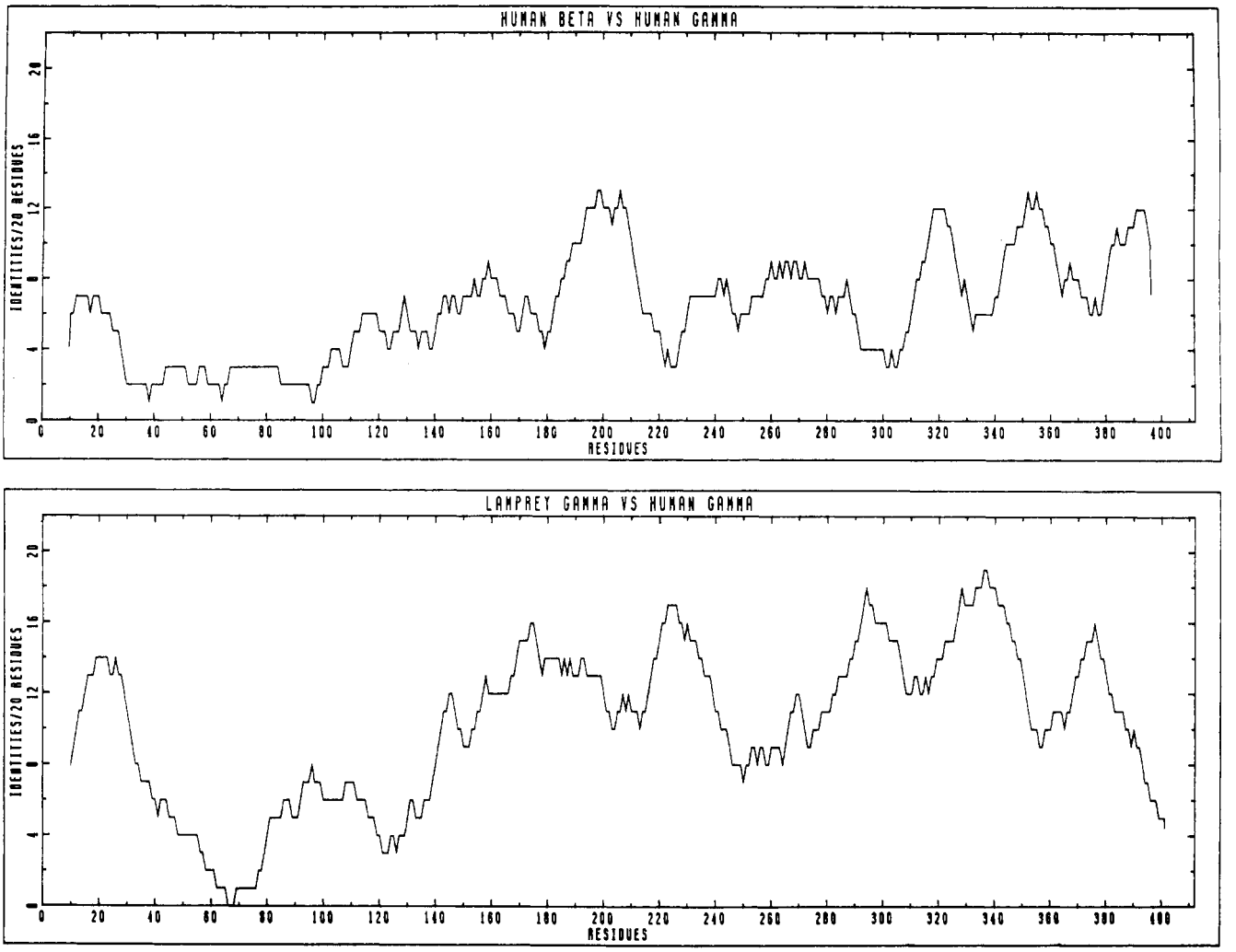

FIGURE 8: Resemblance profiles of human β and γ chains (top) and lamprey and human γ chains (bottom). Regions of low similarity occur in both situations over the course of the coiled coils. (In the β–γ chain comparison, the first 57 residues of the β chain were omitted.)

Table II: Codon Usage in Lamprey and Human γ Chains[a]

codon (AA)[b]	lamprey	human	codon (AA)	lamprey	human	codon (AA)	lamprey	human	codon (AA)	lamprey	human
TTT (F)	3	10	TCT (S)	3	8	TAT (Y)	3	15	TGT (C)	0	7
TTC (F)	16	10	TCC (S)	5	4	TAC (Y)	16	7	TGC (C)	10	4
TTA (L)	0	8	TCA (S)	2	7	TAA (-)	0	0	TGA (-)	1	0
TTG (L)	3	6	TCG (S)	9	1	TAG (-)	0	1	TGG (W)	10	11
CTT (L)	2	4	CCT (P)	1	7	CAT (H)	2	7	CGT (R)	1	0
CTC (L)	14	5	CCC (P)	6	1	CAC (H)	11	4	CGC (R)	13	1
CTA (L)	1	3	CCA (P)	1	4	CAA (Q)	3	11	CGA (R)	2	1
CTG (L)	13	7	CCG (P)	4	0	CAG (Q)	25	13	CGG (R)	3	2
ATT (I)	2	13	ACT (T)	1	12	AAT (N)	3	16	AGT (S)	1	7
ATC (I)	18	8	ACC (T)	8	7	AAC (N)	15	8	AGC (S)	5	3
ATA (I)	1	5	ACA (T)	2	9	AAA (K)	4	19	AGA (R)	0	6
ATG (M)	8	9	ACG (T)	17	1	AAG (K)	18	14	AGG (R)	4	1
GTT (V)	2	5	GCT (A)	0	12	GAT (D)	10	18	GGT (G)	3	10
GTC (V)	5	2	GCC (A)	11	8	GAC (D)	27	13	GGC (G)	20	6
GTA (V)	0	4	GCA (A)	0	8	GAA (E)	4	20	GGA (G)	8	14
GTG (V)	17	4	GCG (A)	8	0	GAG (E)	18	2	GGG (G)	7	9

[a] Human codon count from Chung et al. (1983). [b] AA, amino acid.

bonding to the glutamine of the neighbor and vice versa (Chen & Doolittle, 1971). The same can certainly be true of lamprey, although the actual distance between the two cross-links must be less.

The sequence differences at the very end of the chain—there are only 5 identities over the course of the last 21 residues—are also reflected in the observation that lamprey fibrinogen does not clump those strains of bacteria known to be clumped by most mammalian fibrinogens. On the other hand, lamprey fibrinogen does bind the same antipolymerant peptides as do mammalian fibrinogens (Laudano & Doolittle, 1980). It is reasonable to think that polymerization must involve the more highly conserved sectors, such as residues 318–352 (human numbering), where there are 30 identities among 35 residues, and residues 363–385, where there are 18 identities in 23 residues.

The carboxy termini of fibrinogen γ chains in mammals are known to exist in different forms as the result of differences in messenger splicing (Wolfenstein-Todel & Mosesson, 1981; Crabtree & Kant, 1982). Both of our clones correspond to

Table III: Base Composition of cDNA Sequences of Human and Lamprey γ Chains[a]

	% A	% G	% C	% T
human	32.5	21.7	19.4	26.4
lamprey	22.8	30.6	29.4	17.1

[a] Coding regions only.

the major mammalian type that is spliced, as opposed to a minor form that in mammals reads through into the adjacent intron. We have not identified any minor variant chains during our amino acid sequence studies, and it seems likely that, if they exist, they constitute only a small fraction.

Coding Aspects and Molecular Evolution. Apart from the physiological aspects, there are some other features of the lamprey cDNA sequence that warrant comment. A tally of the base changes relative to the human cDNA sequence (Chung et al., 1983; Kant et al., 1983) and codon position yielded an ordinary distribution, the rate of change of the third position in codons being slightly less than twice that of the rate observed for the first two positions. The codon usage for lamprey γ chains, however, is vastly different from that employed in the human γ chain (Table II). It is decidely nonrandom and violates many of the trends reported for eukaryotic genes (Nussinov, 1981). For example, there is no bias against GC dinucleotides or codons containing that sequence, unlike the situation noted in the human sequence (Chung et al., 1983). The violation of this rule, which has been used as a partial explanation of a restricted usage for some arginine codons, was reflected in the performance of our synthetic oligonucleotide probe. The 23-mer we made was 32-fold degenerate but did not comprise all possible codons for arginine. The probe was designed with the notion that it could be used universally among the vertebrates, and since mammalian γ chains have lysine at position 338 (human numbering), whereas lampreys have arginine, the nucleotides were chosen with an eye to accommodating both situations. In fact, the arginine in the lamprey sequence turned out to be coded for by CGG, one of the least frequently observed arginine codons. Fortunately, the length and GC content of the probe were sufficient to compensate for the single mismatch that resulted. Other points of interest with regard to codon usage include the fact that all 10 cysteines use the codon TGC and among the 22 valine codons there are no occurrences of GTA.

Of the 1317 aligned nucleotide locations, 621 have changed; just less than half of these (294) occur at codon position 3. There is a remarkable shift in base composition as a result of these changes, the lamprey sequence having significantly less A and T than the human sequence (Table III). The directionality of these changes seems extraordinary, being especially odd in the third base position where it might have been expected to be random (Table IV). Thus, at those locations, there were 56 changes of A to G (human to lamprey direction), but only 5 G to A. Similarly, there were 75 T to C, but only 3 C to T. The directions of A/C and T/G changes are similarly biased in one direction. It remains to be seen whether this reflects a general feature of lamprey genes.

Added in Proof

The existence of PCA at the amino terminus of the mature γ chain has now been confirmed by its release by PCA peptidase after digestion of the terminal tripeptide with carboxypeptidase B.

Table IV: Nature of Individual Base Changes between Human and Lamprey γ-Chain cDNA Sequences[a]

	codon position			
	1	2	3	all
A → G	26	18	56	100
G → A	16	12	5	33
C → T	8	14	3	25
T → C	14	7	75	96
A → T	10	16	8	34
T → A	12	14	9	35
A → C	31	21	46	98
C → A	11	20	6	37
G → T	6	3	5	14
T → G	16	5	37	58
G → C	18	7	18	43
C → G	10	12	26	48
	178[b]	149[b]	294[b]	621[b]

[a] Sequences given in the human → lamprey direction. [b] Total.

Acknowledgments

We thank Barry Rothenberg for much assistance and advice during the early stages of this work, Andy Laudano for help in obtaining lampreys, Toni Claudio and Sandra Friezner-Degen for many helpful suggestions about DNA sequencing, Gary Cabirac for a line of HB101 cells, and Chris Broka for much advice in the area of oligonucleotide synthesis. We also thank Da-Fei Feng for computer programming.

References

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Blomback, B., & Yamashina, I. (1958) *Ark. Kemi 12*, 299–319.

Broka, C., Hozumi, T., Arentzen, T., & Itakura, K. (1980) *Nucleic Acids Res. 8*, 5461–5471.

Chang, J. Y., Brauer, D., & Wittman-Liebold, B. (1978) *FEBS Lett. 93*, 205–214.

Chen, R., & Doolittle, R. F. (1971) *Biochemistry 10*, 4486–4491.

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

Cottrell, B. A., & Doolittle, R. F. (1976) *Biochim. Biophys. Acta 453*, 426–438.

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

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

Dagert, M., & Ehrlich, S. O. (1979) *Gene 6*, 23–28.

De Rooij, J. F. M., Wille-Hazeleger, G., van Duersen, P. H., Serdijn, J., & van Boom, J. H. (1979) *Recl. Trav. Chim. Pays-Bas 98*, 537–548.

Doolittle, L. R., Mross, G. A., Fothergill, L. A., & Doolittle, R. F. (1977) *Anal. Biochem. 78*, 491–505.

Doolittle, R. F. (1965) *Biochem. J. 94*, 735–741.

Doolittle, R. F. (1972) *Methods Enzymol. 25B*, 231–244.

Doolittle, R. F. (1983) *Ann. N.Y. Acad. Sci. 408*, 13–26.

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

Doolittle, R. F., & Cottrell, B. A. (1974) *Biochem. Biophys. Res. Commun. 60*, 1090–1096.

Doolittle, R. F., & Wooding, G. L. (1974) *Biochim. Biophys. Acta 271*, 277–282.

Doolittle, R. F., & Laudano, A. P. (1981) *Protides Biol. Fluids 28*, 311–316.

Doolittle, R. F., Cottrell, B. A., & Riley, M. (1976) *Biochim. Biophys. Acta 453*, 439–452.

Doolittle, R. F., Cassman, K. G., Cottrell, B. A., Friezner, S. J., Hucko, J. T., & Takagi, T. (1977) *Biochemistry 16*, 1703–1709.

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

The Federal Register (1982) *The Guidelines for Research Involving DNA Molecules*, Vol. 47 (Aug), p 38048, NIH, Washington, DC.

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

Hawiger, J., Timmons, S., Strong, D., Cottrell, B. A., Riley, M., & Doolittle, R. F. (1982a) *Biochemistry 21*, 1407–1413.

Hawiger, J., Timmons, S., Kloczewiak, M., Strong, D. D., & Doolittle, R. F. (1982b) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2068–2071.

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

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

Ito, H., Ike, Y., Ikuta, S., & Itakura, K. (1982) *Nucleic Acids Res. 10*, 1755–1769.

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

Kern, H. (1975) *Anal. Biochem. 67*, 147–156.

Laudano, A. P., & Doolittle, R. F. (1980) *Biochemistry 19*, 1013–1019.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, A Laboratory Manual*, pp 1–545, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

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

Mickelson, A. M., & Orkin, S. H. (1982) *J. Biol. Chem. 257*, 14773–14782.

Miyoshi, K., Arentzen, R., Huan, T., & Itakura, K. (1980a) *Nucleic Acids Res. 8*, 5491–5504.

Miyoshi, K., Arentzen, R., Huan, T., & Itakura, K. (1980b) *Nucleic Acids Res. 8*, 5507–5517.

Mross, G. A., & Doolittle, R. F. (1977) *Mol. Biol. Biochem. Biophys. 25*, 1–20.

Murtaugh, P. A., Halver, J. E., Lewis, M. S., & Gladner, J. A. (1974) *Biochim. Biophys. Acta 359*, 415–420.

Narang, S. A., Brousseau, R., Hsiung, H. M., & Michniewicz, J. J. (1980) *Methods Enzymol. 65*, 610–620.

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

Nieuwenhuizen, W., & Haverkate, F. (1983) *Ann. N.Y. Acad. Sci. 408*, 92–96.

Norgard, M. V., Emigholz, K., & Monahan, J. J. (1979) *J. Bacteriol. 138*, 270–272.

Nussinov, R. (1981) *J. Mol. Biol. 149*, 125–131.

Olexa, S. A., & Budzynski, A. Z. (1981) *J. Biol. Chem. 256*, 3544–3549.

Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. U.S.A. 54*, 158–165.

Rubin, C. M., & Schmid, C. W. (1980) *Nucleic Acids Res. 8*, 4613–4619.

Shields, D., & Blobel, G. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2059–2093.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Strong, D. D., Laudano, A. P., Hawiger, J., & Doolittle, R. F. (1982) *Biochemistry 21*, 1414–1420.

Takagi, T., & Doolittle, R. F. (1975) *Biochemistry 14*, 5149–5156.

Thompson, J. A., Blakesley, R. W., Doran, K., Hough, C. J., & Wells, R. D. (1983) *Methods Enzymol. 100*, 368–399.

Wallace, R. B., Johnson, M. J., Hirose, T., Miyabe, 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.

Wickens, M. P., Buell, G. N., & Schimke, R. T. (1978) *J. Biol. Chem. 253*, 2483–2495.

Wolfenstein-Todel, C., & Mosesson, M. W. (1981) *Biochemistry 20*, 6146–6149.