Complete Sequence of the Lamprey Fibrinogen α Chain†

Y. Z. Wang,‡ J. Patterson, J. E. Gray, C. Yu, B. A. Cottrell, A. Shimizu,§ D. Graham, M. Riley, and R. F. Doolittle*

Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093

Received June 6, 1989; Revised Manuscript Received August 1, 1989

ABSTRACT: The complete amino acid sequence of the lamprey fibrinogen α chain has been determined by a combination of peptide sequencing and cDNA and genomic cloning. The chain, which has an apparent molecular weight by dodecyl sulfate–polyacrylamide gel electrophoresis of ca. 100 000, is composed of 961 amino acid residues and has a calculated molecular weight of 96 722. It is distinguished by a large number of 18-residue repeats in a region where mammalian fibrinogens have 13-residue repeats. The data are in accord with our previous finding that the lamprey α chain has a distinctive amino acid composition, almost half the residues being glycine, serine, or threonine. The chain differs from mammalian α chains in that there are no cysteines in the carboxy-terminal half, and thus no intrachain loop, nor are there any RGD sequences in the lamprey α chain. Taken together with previous data on the sequences of the β and γ chains, the findings bear significantly on our understanding of fibrin formation. The α chain also provides an interesting case of structural convergence during evolution.

Lampreys and hagfish, the only extant representatives of the cyclostome fish, are the most primitive true vertebrates; they are also the most primitive creatures known to have a blood clotting scheme in which the central action is the thrombin-catalyzed conversion of fibrinogen to fibrin (Doolittle & Surgenor, 1962). Over the years, we have been characterizing lamprey fibrinogen with an eye to finding what is similar to or different from mammalian fibrinogens, the intent being to identify the underlying features of fibrin formation and fibrinolysis (Doolittle et al., 1962). In this regard, similar aspects include three nonidentical polypeptide chains linked by a series of disulfide bonds (Doolittle, 1965a), the thrombin-catalyzed release of fibrinopeptides (Doolittle, 1965b), factor XIII induced fibrin stabilization (Doolittle & Wooding, 1974), and inhibition of polymerization by Gly-Pro-Arg-type peptides (Laudano & Doolittle, 1980). The significant differences include a very large fibrinopeptide B that contains carbohydrate (Doolittle & Cottrell, 1974) and a vastly oversized α chain (Doolittle, 1973; Murtaugh et al., 1974; Doolittle & Wooding, 1974) with an unusual amino acid composition (Doolittle et al., 1976). It is important to bear in mind that lamprey fibrinogen can be clotted by mammalian thrombins (Doolittle et al., 1962) and that mammalian fibrinogens are clotted by lamprey thrombin (Doolittle, 1965a).

We have previously reported the full sequences of lamprey fibrinogen β and γ chains by a combination of protein sequencing and cDNA cloning (Strong et al., 1985; Bohonus et al., 1986). Our early efforts to obtain a full-length cDNA clone for the α chain were unsuccessful, however, and eventually we had to resort to a genomic library. In the end, genomic and cDNA sequence data together, the latter aided by the advent of the polymerase chain reaction (Saiki et al., 1985), allowed us to infer the entire amino acid sequence of the lamprey fibrinogen α chain. We have also characterized a number of α-chain peptides that support the DNA sequence data.

Although portions of the lamprey sequence can be readily aligned with mammalian α-chain sequences, vast regions are radically different. Indeed, some of the distinctive features of the α chain are likely the basis of apparent anomalies in molecular weight measurements reported for the parent fibrinogen molecule (Gladner et al., 1981). These unusual features may also bear on the mechanism of fibrin formation and the behavior of fibrinogen in the presence of other proteins and cells. The α chain also provides an interesting example of convergent evolution at the molecular level.

Materials and Methods

Many of the methods and materials used in this study have been reported in previous articles from this laboratory (Doolittle et al., 1976; Strong et al., 1985). We have also leaned heavily on standard cloning manuals [e.g., see Maniatis et al. (1982) and Berger (1987)].

Peptide Isolation and Characterization. Lamprey α chains were isolated after full reduction and alkylation of fibrin B with a radioactive agent (iodo[^{14}C]acetic acid). The freeze-dried chains were subjected to cyanogen bromide degradation and fragments separated by gel filtration on a Sephadex G-100 column equilibrated with 10% acetic acid. Small peptides were purified further, either by paper electrophoresis or by HPLC, and their amino-terminal residues were determined by dansylation (Gray, 1972). Larger fragments were digested further with trypsin and peptides isolated by HPLC. All peptides were analyzed on a Spinco Model 121 automatic amino acid analyzer. Some were sequenced by the thioacetylation procedure (Mross & Doolittle, 1977); others were analyzed on the automatic gas-phase sequencer operated by the Department of Biology Microsequencer Facility, as were intact lamprey fibrinogen α chains.

Screening cDNA Libraries. Synthetic oligonucleotides for probes and primers were synthesized by the UCSD Peptide Oligonucleotide Synthesis Facility. Originally, we probed a lamprey liver library carried in the plasmid pBR322 which we described in a previous publication (Strong et al., 1985). In the course of the work, however, it was necessary to prepare new libraries, some of which were quite restrictive in that they

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

* To whom correspondence should be addressed.

‡ Present address: Department of Biology, University of Science & Technology of China, Hefei, China.

§ Present address: Sumitomo Metal Industries, Ltd., Toin University, Yokohama, Japan.

0006-2960/89/0428-9801$01.50/0

Table I: Peptides Isolated from Lamprey α Chains

1[a]	DDISLRGPRLTEQ(R)SAGQ(G)S(C)SAS...
2[b]	LPALPDFGTWDM
3[c]	(T)(T)TFTGSAQGGSWSTGGSTATNTG...
4[d]	(I,N,M,P,L,Q,Q,L,T,L,L,Q,H,N,I)K
5[d]	(L,S,S,S,S,S,S,S,T)R
6[d]	(S,S,L,D)M
7[d]	(G,S,C,A)R
8[d]	(T,E,P,N,T,G,S,A,Q,G,G,S,W,S,T,G,G)R

[a] Amino-terminal sequence of (purified) reduced and alkylated α chain. [b] Used for fashioning original oligonnucleotide probes. [c] First 24 residues of a 36-residue tryptic peptide. [d] Amino acid composition of isolated peptide corresponds to observed DNA sequence.

were primed with oligonucleotides based on genomic sequence information. In the end, we also used the polymerase chain reaction (Saiki et al., 1988) directly on fresh lamprey liver cDNA. Primers for these experiments were based on genomic DNA data.

Genomic Library. A lamprey genomic library was prepared with a partial *Sau*3A digest of lamprey DNA cloned into the *Bam*HI site of a λ phage 2001 vector (Stratagene). Phage were plated on *Escherichia coli* grown on appropriate solid media, lifts taken, and papers probed with suitable cDNA inserts. Hybridizing plaques were picked, plaque-purified, amplified, and then subcloned into pBR322 after suitable restriction enzyme digestion. Sequences determined from some of these genomic clones were subsequently used as the basis for the synthesis of additional oligonucleotides that were used to screen freshly prepared cDNA libraries primed with the same oligonucleotides.

DNA Sequencing. In the early stages of this study, we used Maxam–Gilbert sequencing (Maxam & Gilbert, 1980), but subsequently we switched to the chain-termination method of Sanger et al. (1977) in conjunction with the single-stranded bacteriophage M13 (Messing, 1983). More recently, the availability of Sequenase (U.S. Biochemicals) permitted double-strand sequencing (Kraft et al., 1988). In some instances, sets of nested deletions were generated in M13 clones by the 3′ to 5′ exonuclease activity of T4 DNA polymerase after digestion with an appropriate restriction enzyme in the presence of an annealed oligonucleotide according to the procedure described by Dale et al. (1985).

Results

Peptide Sequencing. Early in the project, a 12-residue cyanogen bromide peptide was isolated and its sequence completely determined by the thioacetylation procedure (Table I). A suitable subsequence was chosen from it to fashion two degenerate oligonucleotides for probing cDNA libraries. Other peptides, and in particular all those containing radioactivity as a result of alkylation of cysteine residues, were isolated and their amino acid compositions determined. In one case, a 36-residue peptide was subjected to gas-phase sequencing, and 24 cycles were performed (Table I). Purified reduced and alkylated α chains were also examined by this procedure; the results confirmed and extended a 10-cycle manual degradation we had reported previously (Cottrell & Doolittle, 1976).

cDNA Clones from Ordered Libraries. Approximately 4000 ordered pBR322 clones were screened with two oligonucleotide probes, both of which were 21-mers and 32-fold-degenerate. Nine clones hybridized with both probes, the longest insert of which had an insert of about 1200 base pairs. This insert, denoted 3D (Figure 1), was sequenced in its entirety by the Maxam–Gilbert procedure. A second overlapping clone, 4E, was found to have 490 base pairs. The two clones together amounted to 1422 base pairs, 792 of which coded for the

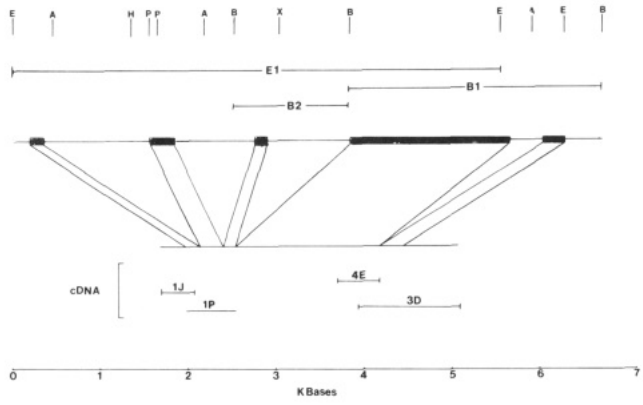

FIGURE 1: cDNA and genomic clones used to determine the sequence of lamprey fibrinogen α chain. Major restriction enzyme sites in genomic DNA at top (E = *Eco*RI, A = *Ava*I, H = *Hin*dIII, P = *Pvu*I, B = *Bam*HI, and X = *Xma*III).

carboxy-terminal 263 amino acids and terminator, and 630 of which corresponded to 3′ noncoding sequence. The ordered libraries were rescreened repeatedly with nick-translated inserts from the clones 3D and 4E, but no new useful clones were identified.

Genomic Library. A lamprey genomic library was constructed in a λ vector and screened with the nick-translated inserts from the two cDNA clones, 3D and 4E. Two positive plaques were identified among the more than 100 000 recombinants examined. The positives were plaque-purified and amplified, after which they were digested with suitable restriction enzymes for subcloning in pBR322. In the case of a *Bam*HI digest, two fragments, denoted B1 and B2, were found to have 2828 and 1235 base pairs, respectively. In another case, a 5549 base pair fragment was successfully cloned into pBR322 after digestion with *Eco*RI. It was found to include the smaller *Bam*HI fragment (B2) and to overlap the larger (B1). B2 (1235 base pairs) was completely sequenced by the Maxam–Gilbert procedure. The other two fragments were sequenced by a combination of Maxam–Gilbert and dideoxy-M13 procedures. All told, the three fragments spanned 6640 base pairs, including five exons encoding 857 amino acids (Figure 1). Translation into amino acid sequences allowed two of the exons at the 5′ end to be easily identified by matching with mammalian α-chain sequences; the intron–exon junctions were obvious. In the case of the other introns, boundaries were conjectured on the basis of potential splice sites (Shapiro & Senapathy, 1987), after which primers were synthesized for the PCR generation of cDNA sequences. The latter allowed the unequivocal identification of all sites. The largest exon contained 1830 base pairs and included 23 sets of a 54 base pair repeat.

Other cDNA Clones. All told, the genomic clones covered about 90% of the full length of the α chain, including all but the last nine amino acids at the carboxy terminus and 95 amino acid residues at the amino terminus. Repeated screening of the genomic library with fragments from either the 5′ or the 3′ ends failed to provide any further positives. Accordingly, we prepared fresh lamprey liver mRNA and primed it for reverse transcriptase with an oligonucleotide (21-mer) based on an obvious coding region from the 5′ end of the genomic E1 clone. The cDNA was cloned into pBR322, and after being annealed, the plasmid was transformed into *E. coli* HB-101 cells. Plated cells were screened with the same oligonucleotide used as a primer, and several hybridizing clones were identified. One of these, denoted 1J, was found to contain a 364-base insert corresponding to the amino terminus of the lamprey α

chain; it also included 15 base pairs corresponding to the last five residues of a putative signal sequence. Another cDNA sequence, 1P, was generated by PCR with primers based on the genomic sequence (Figure 1). All told, cDNA sequences covered 1057 nucleotides at the 5′ end and 1422 nucleotides at the 3′ end (Figure 1) and accounted for 610 of the 961 amino acids in the mature protein. In the end, only a region including 19 of the 18 amino acid residue repeats was based solely on genomic sequence; a major segment of this region was determined independently by direct amino acid sequencing (Table I).

Inferred Amino Acid Sequences. The amino acid sequences inferred from the combined cDNA and genomic DNA clones were in good agreement with expectations based on earlier biochemical characterization of lamprey α chains, including overall amino acid composition (Doolittle et al. (1976), dodecyl sulfate gel electrophoresis, cyanogen bromide and tryptic peptides, and the chemical fragmentation of the end-labeled protein (Jue & Doolittle, 1985).

Discussion

The lamprey α-chain sequence represents the last step in the determination of the complete amino acid sequence of lamprey fibrinogen, the γ-chain (Strong et al., 1985) and β-chain (Bohonus et al., 1986) sequences having been reported previously. The β- and γ-chain sequences are each about 50% identical with their mammalian counterparts; the α chain, on the other hand, differs considerably. In this regard, the species differences parallel the evolution of the subunits themselves. Thus, β and γ chains, whether from the lamprey or mammals, are about 35% identical with each other but are less than 20% identical with comparable portions of α chains. Nonetheless, the lamprey α-chain sequence is readily aligned with the corresponding mammalian sequences over the amino-terminal portion (Figure 4) and, more tentatively, in the carboxy-terminal region (Figure 5). It is the central region of the α chain where the most radical differences exist.

α-Chain Molecular Weights. It is well-known that the molecular weights of α chains from different species are much more variable than are those of β and γ chains. In a survey of 26 mammalian species, for example, α-chain molecular weights ranged from 60000 to 80000, the largest being found in equines (Doolittle, 1973). In an even more extreme situation, lamprey α chains exhibited apparent molecular weights on dodecyl sulfate–polyacrylamide gels electrophoresis of about 100000 (Doolittle, 1973). Subsequently, Murtaugh et al. (1974) and Gladner et al. (1981), on the basis of ultracentrifuge observations, offered some unusual explanations of the lamprey apparent α-chain molecular weight. Thus, Murtaugh et al. (1974) suggested that these chains might actually be of less than ordinary size but covalently cross-linked at the fibrinogen stage, thereby yielding a double molecular weight on gels. They were driven to this conjecture because analytical ultracentrifugation studies indicated that lamprey fibrinogen sedimented only slightly faster than mammalian fibrinogen (Doolittle et al., 1962). Subsequently, Gladner et al. (1981) suggested that the anomaly would be resolved if lamprey fibrinogen had only a single α chain, the subunit composition being $\alpha_1\beta_2\gamma_2$ instead of the $\alpha_2\beta_2\gamma_2$ reported by others (Doolittle, 1965a).

That neither of those suggestions has any basis is now shown by the full sequence. Although the calculated molecular weight of the lamprey α chain is 96722, its chain length, which is what matters in gel electrophoresis, corresponds to an apparent molecular weight of about 106000, in close agreement to the value of 105000 reported by Doolittle and Wooding (1974). Its unusual amino acid composition, more than half the residues being glycine, serine, or threonine, leads to a lower than usual average weight per residue, as well as to an unusual partial specific volume (Doolittle et al., 1976). That the number of α chains must be the same as the numbers of β and γ chains is clear from the conserved positions of cysteine residues, all of which are involved in interchain linkage, as well as from previously reported yield studies (Doolittle et al., 1976).

Cysteine Residues and Disulfide Arrangement. Human and rat fibrinogens, and probably those of all other mammals, contain 31 disulfide bonds, three of which hold the two halves of the molecule together. Of the remaining 28, half are interchain and half intrachain. Lamprey fibrinogen has only 28 cysteines in each of its halves, but each of the 28 is at the same position as a cysteine in the mammalian protein. Of the three that are not present, one involves a γ–γ bond holding the two halves together (Strong et al., 1985). The other two not present in lamprey are found in mammalian fibrinogen in the distal portion of α chains. In mammals, it is known that these 30-residue disulfide loops can become intermolecularly linked through a mixed disulfide to α_1-antitrypsin (Laurell & Thulin, 1975). The absence of these two cysteines precludes such a phenomenon in lamprey fibrinogen.

High-Repeat Structures. The central region of the human fibrinogen α chain contains ten 13-residue imperfect tandem repeats, most of which contain a characteristic tryptophan residue (Doolittle et al., 1979). Lamprey α chains also contain a series of tandem repeats typified by a tryptophan residue (Figure 3). Moreover, the general amino acid composition of the lamprey repeats is remarkably similar to the human type. In this case, the repeats, of which there are 23 all told, are 18 residues each and must have evolved independently. The fact that many of the lamprey repeats are identical with each other, or nearly so, indicates that the responsible genetic events must have occurred quite recently. In both settings, lamprey and mammal, the tandem repeats occur in a part of the α chain that is an exposed, flexible protuberance readily removed by a variety of proteases (Huseby et al., 1970). Inspite of the obvious independent origins, the remarkable similarity in amino acid compositions—rich in both glycine and serine and typified by one tryptophan per repeat—marks this as an unusual case of convergent evolution. The functional role of these apparently flexible regions remains to be demonstrated, although in mammals it has been shown that they can facilitate polymerization (Medved' et al., 1985).

α-Chain Cross-Linking. In both mammals and lampreys, the transglutaminase factor XIII induces intermolecular γ-glutamyl (ϵ-lysine) cross-links between the abutting carboxy termini of γ chains. In both systems, also, the same cross-links can be incorporated between α chains, large molecular weight multimers being formed in the limit. In human fibrinogen, the α-chain acceptor residues have been reported to be two glutamines in the region covered by the tandem repeats (Cottrell et al., 1979). Indeed, human α chains have only three glutamines among the 400 residues composing the carboxy-terminal two-thirds of the chain. The donor residues are thought to be some or all of numerous lysine residues that occur in the carboxy-terminal region (Cottrell et al., 1979; Sobel et al., 1988). Glutamine availability is surely not the limiting feature in lamprey α chains, however, since 19 of the 23 lamprey repeats have a glutamine residue (Figure 3), although none exist among the last 180 carboxy-terminal residues. That region of the chain, like mammalian α chains, contains a substantial number of lysines (Figure 5).

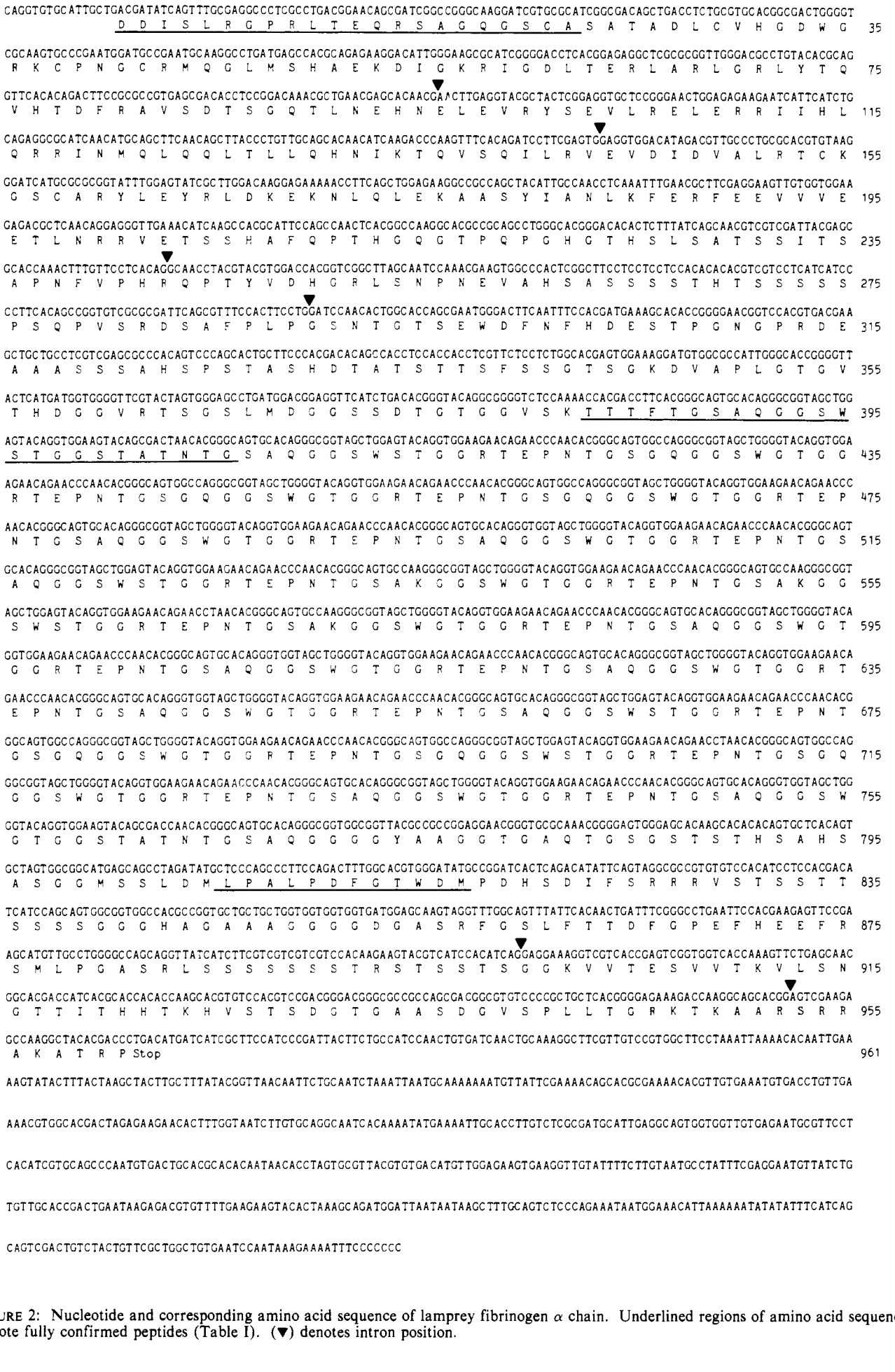

FIGURE 2: Nucleotide and corresponding amino acid sequence of lamprey fibrinogen α chain. Underlined regions of amino acid sequence denote fully confirmed peptides (Table I). (▼) denotes intron position.

```
        Lamprey 18-Res. Repeat        Human 13-Res. Repeat

        ...LMDGGSSDTGTGGVSKTTT      ...PGGNEIIRGGSTSYGTGSE

   1        FTGSAQGGSWSTGGSTAT              TESPRNPSSAGSW
   2        NTGSAQGGSWSTGGRTEP              NSGSSGPGSTGNR
   3        NTGSGQGGSWGTGGRTEP              NPGSSGTGSGATW
   4        NTGSGQGGSWGTGGRTEP              KPGSSGPGSTGSW
   5        NTGSGQGGSWGTGGRTEP              NSGSSGTGSTGNQ
   6        NTGSAQGGSWGTGGRTEP              NPGSPRPGSTGTW
   7        NTGSAQGGSWGTGGRTEP              NPGSSERGSAGHW
   8        NTGSAQGGSWSTGGRTEP              TSESSVSGSTGQW
   9        NTGSAKGGSWGTGGRTEP              HSESGSFRPDSPG
  10        NTGSAKGGSWSTGGRTEP                SGNARPNDPNW
  11        NTGSAKGGSWGTGGRTEP
  12        NTGSAQGGSWGTGGRTEP
  13        NTGSAQGGSWGTGGRTEP
  14        NTGSAQGGSWGTGGRTEP
  15        NTGSAQGGSWGTGGRTEP
  16        NTGSAQGGSWSTGGRTEP
  17        NTGSGQGGSWGTGGRTEP
  18        NTGSGQGGSWSTGGRTEP
  19        NTGSGQGGSWGTGGRTEP
  20        NTGSAQGGSWGTGGRTEP
  21        NTGSAQGGSWGTGGSTAT
(22)        NTGSAQGGGGYAAGGTGA
(23)        QTGSGSTSTHSAHSASGG

            MSSLDMLPALPDFGTWDMPDHS...       GTFEEVSGNVSPGTRREYHT...
```

FIGURE 3: Comparison of lamprey α-chain 18-residue repeat (left) with human α-chain 13-residue repeat (right). The lamprey sequence shown begins at the upper left with Leu-367 and ends with Ser-821; the human sequence shown begins with Pro-245 and ends with Thr-411 at the lower right.

```
                 ** *   *      **** *       *      *  **  *** **** **      *
Human       ADSG EGDFLAEGGGVRGPRVVERH   QSAcKDSDWPFcSDEDWNYKcPSGcRMKGLIDEVNQDF
Rat         ADTGTTSEFIEAGGDIRGPRIVERQ   PSQcKETDWPFcSDEDWNHKcPSGcRMKGLIDEANQDF
Lamprey               DDISLRGPRLTEQRSAGQGScASATADLcVHGDWGRKcPNGcRMQGLMSHAEKDI

                  **  *  *                          *       *  *  **  * *
Human       TNRINKLKNSLFEYQKNNKDSHSLTTNIMEILRGDFSSANNRDNTYNRVSEDLRSRIEVLKRKVIE
Rat         TNRINKLKNSLFDFQKNNKDSNSLTRNIMEYLRGDFANANNFDNTFGQVSEDLRRRIEILKRKVIE
Lamprey     GKRIGDLTERLARLGRLYTQVHTDFRAVSDTSGQTLNEHNELEVRYSEVLRELGRRIIHLQRRINM

                 *   **    *    * *****  * * *** *   **  *       *          *
Human       KVQHIQLLQKNVRAQLVDMKRLEVDIDIKIRScRGScSRALAREVDLKDYEDQQKQLEQVIAKDLL
Rat         KAQQIQVLQXXXXDQLIDMKRLEVDIDIKIRScKGScSRSVSREINLKDYEGQQKQLEQVIAKELL
Lamprey     QLQQLTLLQHNIKTQVSQILRVEVDIDVALRAcKGScARYL  EYRLDKEKNLQLEKAASYIANLK

                                    *       *    *
Human       PSRDRQHLPLIKMKPVPDLVPGNFKSQLQKVPPEWKALTDMPQMR
Rat         PAKDRQYLPAIKMSPVPDLVPGSFKSQLQEGPPEWKALTEMRQMR
Lamprey     FERFEEVVVEETLNRRVETSSHAFQPTHGQGTPQPGHGTHSLSAT
```

FIGURE 4: Alignment of lamprey α-chain amino-terminal sequence (residues 1–230) with corresponding human (residues 1–239) (Doolittle et al., 1979; Rixon et al., 1983) and rat (residues 1–240) (Crabtree et al., 1985) sequences.

```
                  *     *                                  *       *  *  *
Human       PEAMDL  G TLSGIGTLDGFRHRHPDEAAFFDTASTGKTFPGFFSP MLGEFVSETESRGSESG
Rat         GDGMDL  GLTHSFSGRLDELSRMHPELGSFYDSR          FGS LTSNF KEFGSKTSDSD
Lamprey     PDHSDIFSRRRVSTSSTTSSSSGGGHAGAAAGGGGDGASRFGSLFTTDFGPEFHEEFRSMLPGAS


                        *                *  *                       *        *
Human       IFTNTKESSSHHPGIAEFPSRGKSSSYSKQFTSSTSYNRGDSTFESKSYKMADEAGSEADHEG
Rat         IFTDIENPSSH   VPEFSSSSKTSTVRKQVT                KSYKMADEAASEAHQEG D
Lamprey     RLSSSSSSSTRSTSSTSGGKVVTESVVTKVLSNGTTITHHTKHVSTSDGTGAASDGVSPLLTGRK


                  *     *
Human       THSTKRGHAKSRPV
Rat         TRTTKRGRARTM
Lamprey     TKAARSRRAKATRP
```

FIGURE 5: Alignment of lamprey α-chain carboxy-terminal sequence (residues 818–961) with corresponding rat (residues 416–526) and human sequences (residues 475–610).

Certainly one of the major unresolved problems in fibrinogen–fibrin chemistry has to do with the exact nature of α-chain cross-linking, not only because of its impact on fibrin stabilization but also because its solution could reveal the whereabouts of the flexible α-chain appendages after fibrin polymerization has been completed.

Other Structure–Function Aspects. Neither lamprey nor mammalian α chains contain carbohydrate, even though both contain consensus sequences typical of asparaginyl-linked sugars. It is interesting to note, also, that lamprey α chains do not contain the RGD sequence thought to be involved in binding to various cell types. Both rat and human α chains contain this sequence in two different locations.

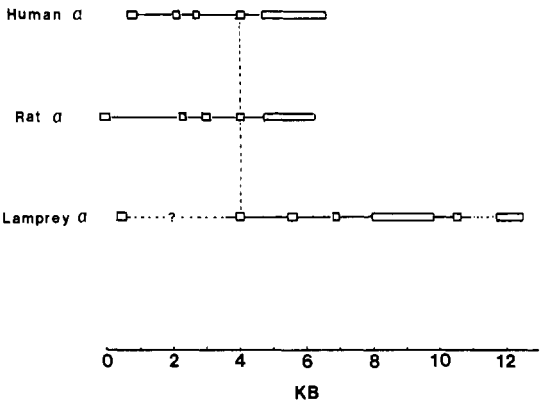

FIGURE 6: Comparison of exon/intron distribution of α-chain genes for lamprey, human (D. Chung, personal communication), and rat (Crabtree et al., 1985). The exons connected by the vertical dashed line are readily aligned in all three species (lamprey amino acid residues 97–143; rat residues 105–152; human residues 104–151). The 5′ region of the lamprey gene has not yet been sequenced but must contain at least one more exon encoding the amino-terminal region of the protein. More likely, it will contain three exons equivalent to the three found in rat and humans. The question mark (?) emphasizes the current uncertainty.

Genetic Aspects. In previous studies on four other lamprey gene products, we observed an extremely biased codon usage in which the third position of most codons was G or C. The α chain is not as distinctly biased in this regard, although if the 414 codons involved in the 23 tandem repeats are omitted from the analysis the bias is clearly evident. In this regard, a frame shift appears to have been involved in the original tandem duplication.

The exon–intron situation in lamprey α chains is both similar to and different from that in mammals. Thus, both rat and human α-chain genes are composed of only five exons separated by four introns (Figure 6). The lamprey α-chain gene is composed of seven or more exons, four additional introns being uncovered so far (Figure 6). Genomic clones covering the lamprey 5′ region have not yet been found. A full description of α-chain intron sequences and the general arrangement of lamprey fibrinogen genes will be reported elsewhere.

ACKNOWLEDGMENTS

We are indebted to Andrew P. Laudano for assistance over the years in collecting lampreys and particularly for providing the most recent batch of livers. We are also grateful to Neal G. Woodbury and Da-Fei Feng for computer programming and to Karen Anderson for assistance in preparing the manuscript.

REFERENCES

Berger, S. L. (1987) *Method Enzymol. 152*, 215.
Bohonus, V. L., Doolittle, R. F., Pontes, M., & Strong, D. D. (1986) *Biochemistry 25*, 6512–6516.
Cottrell, B. A., & Doolittle, R. F. (1976) *Biochim. Biophys. Acta 453*, 426–438.
Cottrell, B. A., Strong, D. D., Watt, K. W. K., & Doolittle, R. F. (1979) *Biochemistry 18*, 5405–5410.
Crabtree, G. R., Comeau, C. M., Fowlkes, D. M., Fornace, A. J., Jr., Malley, J. D., & Kant, J. A. (1985) *J. Mol. Biol. 185*, 1–19.
Dale, R. M. K., McClure, B. A., & Houchins, J. P. (1985) *Plasmid 13*, 31–40.
Doolittle, R. F. (1965a) *Biochem. J. 94*, 735–741.
Doolittle, R. F. (1965b) *Biochem. J. 94*, 742–750.
Doolittle, R. F. (1973) *Adv. Protein Chem. 27*, 1–109.

Doolittle, R. F., & Surgenor, D. M. (1962) *Am. J. Physiol. 203*, 964-970.
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., Oncley, J. L., & Surgenor, D. M. (1962) *J. Biol. Chem. 237*, 3123-3127.
Doolittle, R. F., Cottrell, B. A., & Riley, M. (1976) *Biochim. Biophys. Acta 453*, 439-452.
Doolittle, R. F., Watt, K. W. K., Cottrell, B. A., Strong, D. D., & Riley, M. (1979) *Nature 280*, 464-468.
Gladner, J. A., Lewis, M. S., & Chung, S. I. (1981) *J. Biol. Chem. 256*, 1772-1781.
Gray, W. (1972) *Methods Enzymol. 25*, 333-344.
Huseby, R. M., Mosesson, M. W., & Murray, M. (1970) *Physiol. Chem. Phys. 2*, 374-384.
Jue, R. A., & Doolittle, R. F. (1985) *Biochemistry 24*, 162-170.
Kraft, R., Tardiff, J., Krauter, K. S., & Leinwand, L. A. (1988) *BioTechniques 6*, 544-546.
Laudano, A. P., & Doolittle, R. F. (1980) *Biochemistry 19*, 1013-1019.
Laurell, C.-B., & Thulin, E. (1975) *Scand. J. Immunol. 4*, *Suppl. 2*, 7-12.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *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.
Medved', L. V., Gorkun, O. V., Manyakov, V. F., & Belitser, V. A. (1985) *FEBS Lett. 181*, 109-112.
Messing, J. (1983) *Methods Enzymol. 101*, 20-78.
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.
Rixon, M. W., Chan, W.-Y., Davie, E. W., & Chung, D. W. (1983) *Biochemistry 22*, 3237-3244.
Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985) *Science 230*, 1350-1354.
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Ehrlich, H. A. (1988) *Science 239*, 487-491.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Shapiro, M. B., & Senapathy, P. (1987) *Nucleic Acids Res. 15*, 7155-7174.
Sobel, J. H., Thibodeau, C. A., & Canfield, R. E. (1988) *Thromb. Hemostasis 60*, 153-159.
Strong, D. D., Moore, M., Cottrell, B. A., Bohonus, V. L., Pontes, M., Evans, B., Riley, M., & Doolittle, R. F. (1985) *Biochemistry 24*, 92-101.

Primary Structure of a Ribonuclease from Porcine Liver, a New Member of the Ribonuclease Superfamily

Jan Hofsteenge,* Renate Matthies, and Stuart R. Stone
Friedrich Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland
Received June 26, 1989

ABSTRACT: In most tissues, ribonucleases (RNases) are found in a latent form complexed with ribonuclease inhibitor (RI). To examine whether these so-called cytoplasmic RNases belong to the same superfamily as pancreatic RNases, we have purified from porcine liver two such RNases (PL_1 and PL_3) and examined their primary structures. It was found that RNase PL_1 belonged to the same family as human RNase U_s [Beintema et al. (1988) *Biochemistry 27*, 4530-4538] and bovine RNase K_2 [Irie et al. (1988) *J. Biochem. (Tokyo) 104*, 289-296]. RNase PL_3 was found to be a hitherto structurally uncharacterized type of RNase. Its polypeptide chain of 119 amino acid residues was N-terminally blocked with pyroglutamic acid, and its sequence differed at 63 positions with that of the pancreatic enzyme. All residues important for catalysis and substrate binding have been conserved. Comparison of the primary structure of RNase PL_3 with that of its bovine counterpart (RNase BL_4; M. Irie, personal communication) revealed an unusual conservation for this class of enzymes; the 2 enzymes were identical at 112 positions. Moreover, comparison of the amino acid compositions of these RNases with that of a human colon carcinoma-derived RNase, RNase HT-29 [Shapiro et al. (1986) *Biochemistry 25*, 7255-7264], suggested that these three proteins are orthologous gene products. The structural characteristics of RNases PL_1 and PL_3 were typical of secreted RNases, and this observation questions the proposed cytoplasmic origin of these RI-associated enzymes.

Ribonucleases (RNases)[1] that degrade RNA endonucleolytically via formation of nucleoside 2′,3′-cyclic phosphates can be found in almost any tissue and body fluid of mammals (Sierakowska & Shugar, 1977). The most widely studied members of this type of enzyme are the pancreatic RNases [Smyth et al., 1963; for a review, see Blackburn and Moore (1982) and Beintema (1987)]. In addition, the structures of four types of nonpancreatic RNases have been determined; in the bovine species, these correspond to seminal RNase (Suzuki et al., 1976), kidney RNase K_2 (Irie et al., 1988),

* To whom correspondence should be addressed.

[1] Abbreviations: BNPS-skatole, 2-[(2-nitrophenyl)sulfenyl]-3-bromo-3-methylindolenine; RI, ribonuclease inhibitor; RNase, ribonuclease; SDS, sodium dodecyl sulfate; TFA, trifluoroacetic acid.

0006-2960/89/0428-9806$01.50/0