Complementary DNA Sequence of Lamprey Fibrinogen β Chain†

V. L. Bohonus, R. F. Doolittle,* M. Pontes, and D. D. Strong‡
Department of Chemistry, University of California, San Diego, La Jolla, California 92093
Received April 21, 1986; Revised Manuscript Received June 13, 1986

ABSTRACT: The cDNA sequence of the β chain of lamprey fibrinogen has been determined. To that end, an oligonucleotide probe was synthesized that corresponded to an amino acid sequence from the carboxy-terminal region of the lamprey fibrinogen β chain. The insert actually began with residue 3 of the fibrin β chain; it ran through to a terminator codon following the carboxy-terminal residue at position 443 and then continued for an additional 606 nucleotides of noncoding sequence to its 3′ end. The inferred amino acid sequence was verified by comparison with assorted cyanogen bromide fragments isolated from the β-chain protein, including two carbohydrate-containing peptides that corresponded to segments containing the carbohydrate-attachment consensus sequence. Overall, the lamprey chain is 49% identical with the β chain from human fibrinogen. This is the same degree of resemblance as was found for the lamprey and human γ chains. Moreover, the principal regions of conservation are the same in both the β and γ chains. Differences and similarities in the physiological behavior of the two fibrinogens are assessed in terms of the observed amino acid replacements.

Fibrinogen, the principal protein of blood clotting, is found in the blood of all vertebrates. The protein has been characterized from numerous mammals, and its properties have been found to be similar among them. In the case of primitive vertebrates, however, certain unique aspects have been reported, even though the principal features of the molecule appear to have been conserved. For example, when lamprey fibrinogen is clotted with a mammalian thrombin, only the fibrinopeptide B is removed (Doolittle, 1965; Cottrell & Doolittle, 1976). Lamprey thrombin, on the other hand, clots lamprey fibrinogen by removing both fibrinopeptides A and B. The lamprey fibrinopeptides themselves are noteworthy in being the smallest and longest fibrinopeptides, respectively. Also, the lamprey fibrin β chain is known not to have a histidine at position 2, a residue that others have reported to be essential for polymerization (Shimuzu et al., 1983; 1986).

On the other hand, lamprey fibrinogen has much in common with its mammalian counterparts, including the fundamental six-chain structure ($\alpha_2\beta_2\gamma_2$). The transformation into fibrin is inhibited by synthetic peptide analogues beginning with the sequence Gly-Pro-Arg, which corresponds to the amino terminus of the fibrin α chain (Laudano & Doolittle, 1980). The protein also binds β-chain analogues of the Gly-His-Arg type, even though its own β chain lacks the histidine residue.

Because of the ancient divergence, lamprey fibrinogen represents a unique material for exploring the evolution of fibrin formation. In this regard, we recently reported the construction of a lamprey liver cDNA library and, in our initial report, presented a complete cDNA sequence for the lamprey fibrinogen γ chain (Strong et al., 1985). We now report a cDNA sequence for the lamprey fibrin β chain. Although the cloned insert begins at fibrin residue 3, we had previously determined the amino acid sequence for the amino-terminal portion of the protein by conventional means (Cottrell & Doolittle, 1976); as a result we have in hand a complete sequence for the lamprey fibrinogen β chain, including its 36-residue fibrinopeptide B.

† This investigation was supported by a grant from the National Institutes of Health (HL26873).
‡ Present address: Loma Linda University, Loma Linda, CA.

Materials and Methods

Most of the methods and materials employed in this study, including the preparation of a lamprey liver cDNA library in pBR322, were described in detail in a previous article (Strong et al., 1985). The library was made up of about 6500 clones ordered on eight 96-well microtiter plates; it was screened by colony hybridization after transfer to Whatman 541 paper (Gergen et al., 1979) with an oligonucleotide based on an amino acid sequence deduced from small cyanogen bromide fragments from the lamprey fibrinogen β chain. The probe was synthesized manually by the phosphate triester method as described by Itakura and co-workers (Broka et al., 1980; Ito et al., 1982); it was labeled with [γ-^{32}P]ATP through the action of T4 polynucleotide kinase. Hybridizations were performed at 45–50 °C in Sears Seal-A-Meal bags with 100 ng of probe per filter (sp act. = 10^8 cpm/μg) as described by Wallace et al. (1981). Filters were washed at 40–50 °C at the same ionic strength as used in hybridization (0.9 M NaCl) prior to autoradiography.

Minipreps of plasmids from hybridization-positive clones were prepared, and the plasmid DNA was treated with appropriate restriction enzymes before electrophoresis on agarose gels (Maniatis et al., 1982). Larger preparations of two β-chain-positive clones were grown (Norgard et al., 1979) and the plasmids purified by RPC-5 Analog chromatography (Thompson et al., 1983). In some cases fragments were obtained upon isolation of the insert after digestion with PstI and preparative electrophoresis on agarose gels. DNA restriction fragments were labeled either by the T4 kinase procedure (5′-label) or through the action of terminal transferase (3′-label). DNA sequencing was accomplished by the method of Maxam and Gilbert (1980) with some additional modification (Rubin & Schmid, 1980).

The amino acid sequences of several small cyanogen bromide peptides isolated from the lamprey fibrin β chain were determined by stepwise degradation after attachment to glass beads (Doolittle, L., et al., 1977).

Results

Peptide Sequences and Back-Translated Probe. A number of small cyanogen bromide peptides were isolated by a com-

0006-2960/86/0425-6512$01.50/0

```
GGG GGG AGG CCC CTG CCG AGT GGC ACC AGG GTG AGG AGG CCA CCG CTG CGC CAC CGG CGT CTG GCT CCG GGC GCC GTT ATG AGC CGC GAC CCA CCA GCG TCC CCA
      Arg Pro Leu Pro Ser Gly Thr Arg Val Arg Arg Pro Pro Leu Arg His Arg Arg Leu Ala Pro Gly Ala Val Met Ser Arg Asp Pro Pro Ala Ser Pro
        3                                                                                                   20

AGG CCG CAG GAG GCC CAG AAG GCG ATC CGC GAC GAG GGG GGC TGC ATG CTG CCC GAG AGC GAC CTG GGT GTG CTC TGC CCA ACG GGC TGC GAG CTG CGC GAG GAG
Arg Pro Gln Glu Ala Gln Lys Ala Ile Arg Asp Glu Gly Gly Cys Met Leu Pro Glu Ser Asp Leu Gly Val Leu Cys Pro Thr Gly Cys Glu Leu Arg Glu Glu
               40                                                                                                  60

CTC CTC AAG CAG CGC GAC CCG GTG CGC TAC AAG ATC TCC ATG CTC AAG CAG AAC CTC ACC TAC TTC ATC AAT AGC TTC GAC CGC ATG GCC TCC GAC TCC AAC ACG
Leu Leu Lys Gln Arg Asp Pro Val Arg Tyr Lys Ile Ser Met Leu Lys Gln Asn Leu Thr Tyr Phe Ile Asn Ser Phe Asp Arg Met Ala Ser Asp Ser Asn Thr
                                80                                  CHO                                                 100

CTC AAG CAG AAC GTG CAG ACG TTG CGC AGA CGC CTC AAC TCT CGA AGC AGC ACC CAC GTG AAC GCG CAG AAG GAG ATC GAG AAC CGC TAC AAG GAG GTG AAG ATC
Leu Lys Gln Asn Val Gln Thr Leu Arg Arg Arg Leu Asn Ser Arg Ser Ser Thr His Val Asn Ala Gln Lys Glu Ile Glu Asn Arg Tyr Lys Glu Val Lys Ile
                                           120                                                                                 140

CGC ATC GAG TCA ACG GTG GCC GGC TCG CTG CGC TCC ATG AAG TCC GTG CTG GAG CAC CTG CGA GCC AAA ATG CAG CGC ATG GAG GAG GCC ATC AAG ACG CAG AAG
Arg Ile Glu Ser Thr Val Ala Gly Ser Leu Arg Ser Met Lys Ser Val Leu Glu His Leu Arg Ala Lys Met Gln Arg Met Glu Glu Ala Ile Lys Thr Gln Lys
                                                    160

GAG CTG TGC AGC GCG CCT TGC ACC GTC AAC TGC CGC GTG CCC GTC GTG TCC GGA ATG CAC TGT GAG GAC ATC TAT CGA AAC GGT GGC CGA ACA AGT GAA GCG TAC
Glu Leu Cys Ser Ala Pro Cys Thr Val Asn Cys Arg Val Pro Val Val Ser Gly Met His Cys Glu Asp Ile Tyr Arg Asn Gly Gly Arg Thr Ser Glu Ala Tyr
               180                                                                                                 200

TAC ATC CAG CCG GAC CTC TTC TCG GAG CCC TAC AAG GTC TTC TGC GAC ATG GAG AGC CAC GGA GGA GGC TGG ACG GTC GTC CAG AAC CGA GTG GAT GGC AGC TCC
Tyr Ile Gln Pro Asp Leu Phe Ser Glu Pro Tyr Lys Val Phe Cys Asp Met Glu Ser His Gly Gly Gly Trp Thr Val Val Gln Asn Arg Val Asp Gly Ser Ser
                                220                                                                                 240

AAC TTC GCA AGA GAC TGG AAC ACG TAC AAG GCT GAA TTT GGA AAC ATT GCC TTC GGT AAC GGA AAG TCC ATC TGC AAC ATT CCA GGC GAG TAC TGG CTG GGC ACG
Asn Phe Ala Arg Asp Trp Asn Thr Tyr Lys Ala Glu Phe Gly Asn Ile Ala Phe Gly Asn Gly Lys Ser Ile Cys Asn Ile Pro Gly Glu Tyr Trp Leu Gly Thr
                                           260                                                                                 280

AAA ACC GTT CAC CAG CTG ACT AAG CAG CAC ACG CAG CAG GTG CTG TTC GAC ATG AGC GAC TGG GAG GGC AGC AGC GTG TAC GCT CAG TAC GCC AGC TTC CGG CCG
Lys Thr Val His Gln Leu Thr Lys Gln His Thr Gln Gln Val Leu Phe Asp Met Ser Asp Trp Glu Gly Ser Ser Val Tyr Ala Gln Tyr Ala Ser Phe Arg Pro
                                                    300

GAG AAC GAG GCG CAG GGC TAC CGC CTC TGG GTG GAG GAT TAC AGC GGG AAC GCG GGC AAC GCG CTG CTG GAA GGA GCC ACG CAG CTC ATG GGC GAC AAC CGC ACC
Glu Asn Glu Ala Gln Gly Tyr Arg Leu Trp Val Glu Asp Tyr Ser Gly Asn Ala Gly Asn Ala Leu Leu Glu Gly Ala Thr Gln Leu Met Gly Asp Asn Arg Thr
               320                                                                                 340                 CHO

ATG ACC ATC CAC AAC GGC ATG CAG TTC AGC ACG TTC GAC CGA GAC AAT GAC AAC TGG AAC CCC GGC GAT CCA ACC AAG CAC TGC TCG CGC GAG GAC GCT GGT GGT
Met Thr Ile His Asn Gly Met Gln Phe Ser Thr Phe Asp Arg Asp Asn Asp Asn Trp Asn Pro Gly Asp Pro Thr Lys His Cys Ser Arg Glu Asp Ala Gly Gly
                                360                                                                                 380

TGG TGG TAC AAT CGC TGC CAT GCG GCC AAT CCC AAC GGA CGC TAC TAC TGG GGC GGC ATC TAC ACC AAG GAG CAG GCG GAC TAC GGC ACG GAC GAC GGC GTC GTG
Trp Trp Tyr Asn Arg Cys His Ala Ala Asn Pro Asn Gly Arg Tyr Tyr Trp Gly Gly Ile Tyr Thr Lys Glu Gln Ala Asp Tyr Gly Thr Asp Asp Gly Val Val
                                           400                                                                                 420

TGG ATG AAC TGG AAG GGC TCC TGG TAT TCT ATG CGC CAA ATG GCC ATG AAG CTT CGA CCC AAG TGG CCC TAG CCG TTT TTC TGT CAA TTA AAA AAA ATA CTC TCA
Trp Met Asn Trp Lys Gly Ser Trp Tyr Ser Met Arg Gln Met Ala Met Lys Leu Arg Pro Lys Trp Pro Stop
                                                    440         443

AGA GCC CTT TCG ATG CGT TGA CGG GTT GTT CCA CGT CTT TGT CGG CAT GAC CCC AAA AAA CAT CGG AGA ACC ATA CAG CAC AAC CTT GAA AAC CTA CGT AGA AGC

ATG CTT TCA TAA GGC CCC ACG AAA TTT GCA TGG GAT TCT TTT TTT AAA AAT CAG TTC GAA CTT CCA TCA GAA TTA TTG TGT TCT TTG ATA GGG TAC AGG ACC GTT

CTG GTT TGC TGC AGG GTT TTG GAG TTG TAG AAA CTG AAT GCA TAG AAT TAT ATT GTG GGA TAT GAT TAA AAT GTG GCC GTC AAT GAG ATA CTC AGT ATG TCA ATT

TGA TGT GAC AAG TGA CCA AGC TGT CCG CAA TTA TGT AGG CTA AAA CAA AGA GTG CAG TGA GTA AAG CTA ACA ATC ATG CCT TGT TTC CCG CAA AGC CAT CTC TGT

TGC ATT TTG AAT TCC TAT TAA AAA TAA CGC TTA CAT CAA AAG TTA AAA TAA ATC AAT ACA AAA GGG CTT TGT TTG GAG TTG TAC TGT CAG CGC CTT GAA AAA TAA

TGC ATT TTA AAT TAA GTG TTA AAA GAA AGC CAG TGC ATT TCT GTT CTA CCC CCC CCC CCC CCC
```

FIGURE 1: Nucleotide sequence of lamprey fibrin β-chain cDNA and corresponding amino acid sequence. CHO = carbohydrate attachment sites. Numbers refer to amino acid location in fibrin β chain. A potential polyadenylation site near the 3′ end is underlined.

bination of gel filtration and paper electrophoresis and their amino acid compositions found. Selected peptides were attached to glass beads and their sequences determined, in whole or part, by stepwise degradation. Among these was a homoserine-free heptapeptide that was similar in sequence to the reported carboxy terminus of the human β chain (Takagi & Doolittle, 1975). This peptide in combination with a tripeptide and dipeptide allowed a 12-residue stretch to be aligned with the known sequence for the carboxy terminus of the human β chain (Henschen & Lottspeich, 1977; Watt et al., 1979).

human sequence ...(M) RKMSMKIRPFFPQQ

lamprey sequence ...(M) RQMAMKLRPKWP

A 17-mer mixed oligonucleotide probe was synthesized on the basis of a portion of this sequence:

Gln Met Ala Met Lys Le(u)

CAR ATG GCX ATG AAR YT

where R = both purine nucleotides, Y = both pyrimidine nucleotides, and X = a mixture of all four nucleotides. Actually, the anticomplement of the sequence was made:

5′-ARYTTCATXGCCATYTG-3′

Screening. Only two clones among the 6500 screened were found to hybridize strongly with the labeled probe. Both were about 2 kb in length. One, P610-2C, was found to begin at residue 3 of the fibrin β chain. The other contained only a small coding section amounting to the carboxy-terminal one-fifth of the chain, the remainder being composed of a very long 3′ noncoding region. Neither reached to a 3′ poly(A), although a potential polyadenylation sequence (AATAAA) was found to occur 499 nucleotides downstream from the terminator.

DNA Sequencing. The insert from the clone P610-2C was sequenced in its entirety; it contained 1932 nucleotides, independent of terminal poly(G) and poly(C) tails. The complete cDNA sequence of the insert from plasmid P610-2C is presented in Figure 1, as is the corresponding amino acid sequence. The sequence was corroborated by several lines of independent evidence. First, the inferred amino acid composition is in good agreement with the experimentally determined composition of the isolated fibrin β chain (Doolittle et al., 1976). Second, the inferred sequence matches residues 3–6 of the fibrin β chain as determined by direct stepwise degradation (Cottrell & Doolittle, 1976). Finally, we had sequenced several small cyanogen bromide peptides from

Table I: Codon Usage in Lamprey Fibrin β Chain

TTT (F) 1	TCT (S) 2	TAT (Y) 2	TGT (C) 1
TTC (F) 10	TCC (S) 10	TAC (Y) 17	TGC (C) 10
TTA (L) 0	TCA (S) 1	TAA (*) 0	TGA (*) 0
TTG (L) 1	TCG (S) 3	TAG (*) 1	TGG (W) 13
CTT (L) 1	CCT (P) 1	CAT (H) 1	CGT (R) 1
CTC (L) 10	CCC (P) 8	CAC (H) 9	CGC (R) 20
CTA (L) 0	CCA (P) 7	CAA (Q) 1	CGA (R) 7
CTG (L) 15	CCG (P) 7	CAG (Q) 20	CGG (R) 2
ATT (I) 2	ACT (T) 1	AAT (N) 4	AGT (S) 2
ATC (I) 12	ACC (T) 9	AAC (N) 23	AGC (S) 14
ATA (I) 0	ACA (T) 1	AAA (K) 2	AGA (R) 2
ATG (M) 17	ACG (T) 12	AAG (K) 20	AGG (R) 5
GTT (V) 2	GCT (A) 4	GAT (D) 3	GGT (G) 5
GTC (V) 6	GCC (A) 11	GAC (D) 20	GGC (G) 21
GTA (V) 0	GCA (A) 1	GAA (E) 3	GGA (G) 7
GTG (V) 15	GCG (A) 10	GAG (E) 24	GGG (G) 2

Total codons (including terminator) = 442

```
Human    GHRPLDK KREEAPSLRPAPPPISGGGYRARPAKAAATQKKVERKAPDAGGcLHADPDLG
Lamprey  GVRPLPSGTRVRRPPLR  HRRLAPGAVMSRDPPASPRPQEAQKAIRDEGGcMLPESDLG

Human    VLcPTGcQLQEALLQQERPIRNSVDELNNNVEAVSQTSSSSFQYMYLLKDLWQKRQKQVK
Lamprey  VLcPTGcELREELLKQRDPVRYKISMLKQNLTYFINSFDRMASDSNTLKQNVQTLRRRLN

Human    DNENVVNEYSSELEKHQLYIDETVNSNIPTNLRVLRSILENLRSKIQKLESDVSAQMEYc
Lamprey  SRSSTHVNAQKEIENRYKEVKIRIESTVAGSLRSMKSVLEHLRAKMQRMEEAIKTQKELc

Human    RTPcTVScNIPVVSGKEcEEIIRKGGETSEMYLIQPDSSVKPYRVYcDMNTENGGWTVIQ
Lamprey  SAPcTVNcRVPVVSGMHcEDIYRNGGRTSEAYYIQPDLFSEPYKVFcDMESHGGGWTVVQ

Human    NRQDGSVDFGRKWDPYKQGFGNVATNTDGKNYcGLPGEYWLGNDKISQLTRMGPTELLIE
Lamprey  NRVDGSSNFARDWNTYKAEFGNIA FGNGKSIcNIPGEYWLGTKTVHQLTKQHTQQVLFD

Human    MEDWKGDKVKAHYGGFTVQNEANKYQISVNKYRGTAGNALMDGASQLMGENRTMTIHNGM
Lamprey  MSDWEGSSVYAQYASFRPENEAQGYRLWVEDYSGNAGNALLEGATQLMGDNRTMTIHNGM

Human    FFSTYDRDNDGWLTSDPRKQcSKEDGGGWWYNRcHAANPNGRYYWGGQYTWDMAKHGTDD
Lamprey  QFSTFDRDNDNWNPGDPTKHcSREDAGGWWYNRcHAANPNGRYYWGGIYTKEQADYGTDD

Human    GVVWMNWKGSWYSMRKMSMKIRPFFPQQ
Lamprey  GVVWMNWKGSWYSMRQMAMKLRPKWP
```

FIGURE 2: Alignment of lamprey and human β-chain sequences. Dots denote identical residues.

lamprey fibrin β chains, and these were all identified by cDNA translation. The latter included two carbohydrate-containing peptides that match consensus sequences for asparagine-linked sugars. One of these is situated at Asn-88 and the other at Asn-348.

3′ Noncoding Region. Although the 3′ noncoding region in each of the two clones was extensive, no poly(A) region was found in either one. It has been reported that human β-chain cDNA is heterogeneous in the 3′ untranslated region (Chung et al., 1983). Morever, Kant et al. (1983) found a poly(A) region in a human β-chain clone only 85 nucleotides downstream from the terminator but no polyadenylation signal sequence. In our lamprey clone there is a consensus signal 499 nucleotides downstream from the terminator.

Codon Usage. The cDNA sequence of clone P620-2C revealed further evidence of an unusual codon preference in lampreys, there being an unusually high frequency of G or C at position 3 of most codons (Table I).

Discussion

The primary objective in determining the amino acid sequence of lamprey fibrinogen is to compare it with those of mammalian fibrinogens in order to find which features have been conserved over the course of 400 million years of vertebrate evolution. In this regard, a direct comparison of the lamprey fibrin β-chain amino acid sequence with the human fibrin β-chain sequence reveals that 218 of 443 comparable residues are identical (Figure 2). A single gap of one residue in the human chain and two small gaps (one and two residues, respectively) in the lamprey chain are required for this optimal alignment. In addition, the human β chain extends two residues further at the carboxy terminus. All ten cysteine residues occur at the same locations in both chains.

Carbohydrate Locations. The location of carbohydrate attachment points is another matter, however. It is well-known that the human β chain contains a single carbohydrate cluster located at Asn-350 (human fibrin numbering) (Henschen & Lottspeich, 1977; Watt et al., 1979), and carbohydrate is indeed found at the same place (Asn-348) in the lamprey β chain. In addition, however, carbohydrate is also found at residue-88, which is situated at the midpoint of the coiled-coil regions that connect the central and terminal globules that constitute fibrinogen molecules. Its presence was confirmed by characterization of the corresponding CNBr peptide, which was found to contain carbohydrate. The finding is of considerable interest since we had previously found that the carbohydrate cluster located in the equivalent location in mammalian γ chains is *absent* in the lamprey γ chain (Strong et al., 1985). Thus, lamprey and mammalian fibrinogens both have a carbohydrate cluster in the middle of the coiled coils, but in mammals it is on the γ chain and in lampreys on the β chain.

Resemblance Profiles. Although the lamprey fibrin β-chain sequence is 49% identical with that of the human chain, the degree of conservation varies considerably along the length of the chain. A computer-generated resemblance profile reveals that the region of the coiled coils (residues 66–177, lamprey fibrin numbering) is only 20–30% identical, whereas there are peaks of resemblance near the carboxy terminus that are greater than 80% identical (Figure 3). A similar profile had been obtained for a comparison of lamprey and human γ chains. It is not altogether surprising that the resemblance profiles for the two sets of chains follow similar courses, given the common ancestry of β and γ chains. In this regard, lamprey and human γ chains are 50% identical (205 identities among 408 comparable residues), indicating that β and γ chains have been evolving at very similar rates since the gene duplication that gave rise to their separate identities about 700 million years ago, a time based on the fact that in both lampreys and humans β chains and γ chains are 33% identical with each other. The chief evolutionary restraints likely have to do with folding and the necessity for the chains to interact with each other, as well as quasi-equivalent roles in fibrin formation. For reference purposes, it is interesting to note that lamprey hemoglobin is only 33% identical with mammalian globins and lamprey cytochrome *c* is 80% identical with mammalian cytochromes *c*.

Conserved Sequences and Fibrin Formation. In its simplest guise, fibrin formation depends on the uniform packing of individual units in a half-staggered overlapping polymer two molecules thick. The particular interactions leading to this regular system are thought to involve the amino-terminal segments of fibrin α chains and a complementary site situated somewhere in the carboxy terminus 109 amino acid residues of the γ chain [for a recent review, see Doolittle (1984)]. Binding studies with peptides corresponding to the amino-terminal segments of fibrin α and β chains, combined with observations on the common ancestry of β and γ chains, have led to the hypothesis that the carboxy-terminal regions of β chains ought to contain a complementary site for binding the

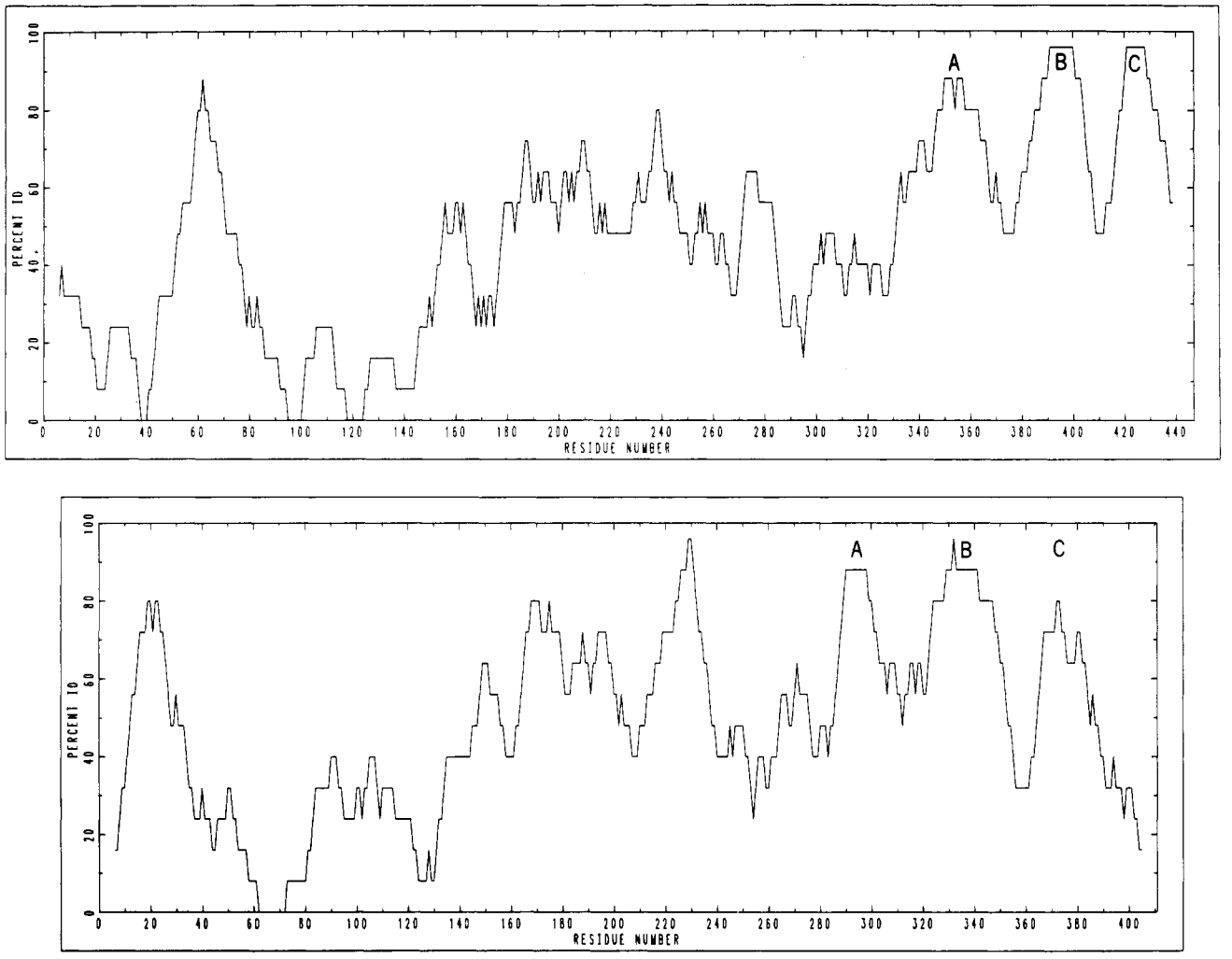

FIGURE 3: Resemblance profiles of lamprey and human fibrin β chains (top) and lamprey and human fibrinogen γ chains (bottom). A, B, and C denote equivalent and alignable sections of both β and γ chains. Note that β chains are longer than γ chains, having some additional residues at the amino terminus, but lacking some others at the carboxy terminus.

amino terminus of fibrin β chains (Doolittle & Laudano, 1980), thereby allowing the lateral involvement of the two molecule thick protofibrils. If this were the case, then it might be possible to gauge which residues contribute to these homologous sites on the basis of similarities and differences in lamprey and human β and γ chain sequences.

For example, one of the features that most distinguishes β chains from γ chains in their carboxy-terminal regions is the presence of carbohydrate in the former. The carbohydrate, which is present in lamprey fibrinogen β chains at precisely the same place as in humans, may prevent intermolecular interaction at a site that is homologous with the authentic polymerization site on the γ chain. Certainly there is good reason to think that the carbohydrate at β-350 (human fibrin numbering) is what limits degradation by plasmin in this region of the β chain. Similarly, the γ chain carboxy terminus is well-known to be the primary site of intermolecular cross-linking during fibrin stabilization. The equivalent residues, amounting to a skein of 19 amino acids, are absent in both lamprey and mammalian β chains.

We would cautiously extend this line of reasoning to a comparison of similarities and differences in sequence with the observed binding of peptide analogues by lamprey and human fibrinogen. The relevant considerations are as follows: (a) Homologous sites ought to exist in equivalent positions on homologous chains. (b) Lamprey and human fibrinogens both bind Gly-Pro-Arg and Gly-His-Arg type peptides. (c) Lamprey fibrinogen binds Gly-Val-Arg peptides but human fibrinogen does not. (d) The peptide binding sites are likely negatively charged pockets. (e) The Gly-Pro-Arg binding site is situated somewhere among the carboxy terminus 109 residues of the γ chain. (f) Gly-Pro-Arg-type peptides can effectively fit into Gly-His-Arg and Gly-Val-Arg sites, but not vice versa.

```
LB   ..VHQLTKQHT  QQVLFdMSdWEGSSVYAQYASFRPENEAQGYRLWVEdY SGNAGNALLEGA
HB   ..ISQLTRMGP  TELLIEMEdWKGdKVKAHYGGFTVQNEANKYQISVNKY RGTAGNALMdGA
LG   ..IHLLTGQQA  YRLRIdLTdWENTHRYAdYGHFKLTPESdEYRLFYSMYLdGdAGNAF dGF
HG   ..IHLISTQSAIPYALRVELEdWNGRTSTAdYAMFKVGPEAdKYRLTYAYFAGGdAGdAF dGF

LB     TQLMGdNRTMTIHNGMQFSTFdRdNdNWNPGdPTKHCSREdAGGWWYNRCHAANPNGRYY
HB     SQLMGENRTMTIHNGMFFSTYdRdNdGWLTSdPRKQCSKEdGGGWWYNRCHAANPNGRYY
LG   dFGddPQdKFYTTHLGMLFSTPERdNdKYEGS      CAEQdGSGWWMNRCHAGHLNGKYY
HG   dFGddPSdKFFTSHNGMQFSTWdNdNdKFEGN      CAEQdGSGWWMNKCHAGHLNGVYY

LB   WGGIYTKEQAdYGTddGVVWMNWKGSWYSMRQMAMKLRPKWP
HB   WGGQYTWdMAKHGTddGVVWMNWKGSWYSMRKMSMKIRPFFPQQ
LG   FGGNYRKTdVEFPYddGIIWATWHdRWYSLKMTTMKLLPMGRdLSGHGGQQQSKGNSRGDN
HG   QGGTYSKASTPNGYdNGIIWATWKTRWYSMKKTTMKIIPFNRLTIGEGQQHHLGGAKQAGDV
```

FIGURE 4: Four-way alignment of carboxy-terminal regions of lamprey and human β and γ chains. Dots denote positions with identical residues in all four chains. Aspartic acid residues are shown in lower case. The asterisk (*) identifies an asparagine (N) in the two β chains that has carbohydrate attached. The arrow denotes a plasmin-sensitive bond that removes a portion of the human γ chain essential to the binding of Gly-Pro-Arg-type peptides. The dashed lines (---) indicate two segments that may be involved in binding to amino-terminal "donor" sites.

These considerations led us to align the carboxy-terminal segments of all four sequences and assess the situation with regard to the conservation of aspartic and glutamic acid residues (Figure 4). In this regard, aspartic acid residues are much more prevalent in this region than are glutamic acids and are more rigorously conserved. In the end, it would appear that the putative homologous binding sites are most likely situated between residues Glu-339 and Asp-382 (lamprey fibrin β (chain numbering) a region containing several sets of regularly spaced aspartic acids. In particular, the segment delineated by Asp-363 and Asn-370 best suits all the constraints imposed by the binding studies.

Importance of the Coiled Coils. Early amino acid sequence data led to the proposal that the coiled-coil regions of fibrinogen must extend across equivalent 110-residue segments between similar braces of cysteines in all three chains (Doolittle, R., et al., 1977). There are two important aspects in maintaining the integrity of these interdomainal connections. First, the number of residues involved in all three chains must be kept constant, since any gap of more than a residue or two would destroy the synchrony of the cysteines involved in the disulfide rings. Second, the polarity of the side chains must be conserved to the extent that the formation of compounded α helices can form. Beyond that there appears to be little in the way of restraint on amino acid replacement, and as a result of these regions of both β and γ chains have experienced a rapid replacement rate since the divergence of lampreys and other vertebrate groups. There does seem to be one more requirement, however: a carbohydrate cluster must exist in the central region of the coiled coils. In mammalian fibrinogen, the carbohydrate is attached to the γ chain. In lampreys it is situated at an equivalent position on the β chain.

ACKNOWLEDGMENTS

We are grateful for much help and encouragement during the course of this work from Barbara Cottrell, David Graham, Margaret Moore, and Marcia Riley. We thank Karen Anderson for assisting in the preparation of the manuscript.

REFERENCES

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

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

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

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

Doolittle, R. F., & Laudano, A. P. (1980) *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., Takagi, T. (1977) *Biochemistry 16*, 1710–1715.

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

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

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*, 3853–3957.

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, New York.

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

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

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

Shimizu, A., Saito, Y., Matsushima, A., & Inada, Y. (1983) *J. Biol. Chem. 258*, 7915–7917.

Shimizu, A., Saito, Y., & Inada, Y. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 591–593.

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.

Takagi, T., & Doolittle, R. F. (1975) *Biochim. Biophys. Acta 386*, 617–622.

Thompson, J. A., Blakesley, R. W., Doran, K., Hough, C. J., & Well, 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.