Lilley, D. M. J. (1981) *Nucleic Acids Res. 9*, 1271–1289.
Lilley, D. M. J., & Pardon, J. F. (1979) *Annu. Rev. Genet. 13*, 197–233.
Mace, H. A. F., Pelham, H. R. B., & Travers, A. A. (1983) *Nature (London) 304*, 555–557.
Maquat, L. E., Kinniburgh, A. J., Rachmilewitz, E. A., & Ross, J. (1981) *Cell (Cambridge, Mass.) 27*, 543–553.
Mathis, D., Oudet, P., & Chambon, P. (1980) *Prog. Nucleic Acid Res. Mol. Biol. 24*, 1–55.
Mizuuchi, K., Mizuuchi, M., & Gellert, M. (1982) *J. Mol. Biol. 156*, 229–243.
Morris, N. R. (1976) *Cell (Cambridge, Mass.) 9*, 627–632.
Müntz, K. (1982) in *Nucleic Acids and Proteins in Plants I* (Boulter, D., & Parthier, B., Eds.) pp 505–545, Springer-Verlag, West Berlin.
Murray, M. G., & Thompson, W. F. (1977) *Year Book—Carnegie Inst. Washington 76*, 255–259.
Murray, M. G., Peters, D. L., & Thomspon, W. F. (1981) *J. Mol. Evol. 17*, 31–42.
Schon, E., Evans, T., Welsh, J., & Efstratiadis, A. (1983) *Cell (Cambridge, Mass.) 35*, 837–848.
Sinden, R. R., Broyles, S. S., & Pettijohn, D. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1797–1801.
Singleton, C. K., Klysik, J., Stirdivant, S. M., & Wells, R. D. (1982) *Nature (London) 299*, 312–316.
Slightom, J. L., Sun, S. M., & Hall, T. C. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1897–1901.
Smith, R. D., Seale, R. L., & Yu, J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5505–5509.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Spiker, S. (1984) *Adv. Genet. 22*, 145–208.
Spiker, S., Murray, M. G., & Thompson, W. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 815–819.
Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M., & Weintraub, H. (1980) *Cell (Cambridge, Mass.) 20*, 451–460.
Sun, S. M., Mutschler, M. A., Bliss, F. A., & Hall, T. C. (1978) *Plant Physiol. 61*, 918–923.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
Thompson, W. F., & Murray, M. G. (1981) in *The Biochemistry of Plants* (Marcus, A., Ed.) Vol. 6, pp 1–81, Academic Press, New York.
Walbot, V. (1973) *New Phytol. 72*, 479–483.
Weisbrod, S. (1982) *Nature (London) 297*, 289–295.
Wu, C. (1980) *Nature (London) 286*, 854–860.

γ and γ' Chains of Human Fibrinogen Are Produced by Alternative mRNA Processing†

Dominic W. Chung and Earl W. Davie*

ABSTRACT: cDNAs and the genomic DNA coding for the γ and γ' chains of human fibrinogen have been isolated and characterized by sequence analysis. The cDNAs coding for the γ and γ' chains share a common nucleotide sequence coding for the first 407 amino acid residues in each polypeptide chain. The predominant γ chain contains an additional four amino acids on its carboxyl-terminal end (residues 408–411). These four amino acids, together with the 3′ noncoding sequences, are encoded by the tenth exon. Removal of the ninth intervening sequence following the processing and polyadenylation reactions yields a mature mRNA coding for the predominant γ chain. The less prevalent γ' chain contains 20 amino acids at its carboxyl-terminal end (residues 408–417). These 20 amino acids are encoded by the immediate 5′ end of the ninth intervening sequence. This results from an occasional processing and polyadenylation reaction that occurs within the region normally constituting the ninth intervening sequence. Accordingly, the gene for the γ chain of human fibrinogen gives rise to two mRNAs that differ in sequence on their 3′ ends. These mRNAs code for polypeptide chains with different carboxyl-terminal sequences. Both of these polypeptides are incorporated into the fibrinogen molecule present in plasma.

Fibrinogen is a plasma glycoprotein that participates in the final phase of blood coagulation. It is composed of three pairs of nonidentical polypeptide chains, designated α (M_r 66 000), β (M_r 52 000), and γ (M_r 46 500) (McKee et al., 1966), and these chains are held together by interchain and intrachain disulfide bonds (Blomback & Blomback, 1972).

When isolated from plasma, fibrinogen is heterogeneous in charge and size (Mosesson et al., 1972; Henschen & Edman, 1972; Mosher & Blout, 1973). This heterogeneity is due in part to the presence of a minor variant form of the γ chain, called γB (Francis et al., 1980) or γ' (Wolfenstein-Todel & Mosesson, 1980, 1981). The γ' chain has a higher molecular weight than the γ chain, and in human fibrinogen, γ' constitutes about 11% of the total γ-chain population (Mosesson et al., 1972). In rodents, the γ' may be as high as 30% of the total γ-chain population (Legrele et al., 1982). Wolfenstein-Todel & Mosesson (1980) have shown that the differences between the human γ and γ' chains resides in the carboxyl-terminal region of the polypeptide chain. Amino acid sequence analysis has demonstrated that the carboxyl-terminal four amino acid residues of the regular γ chain are substituted by a peptide of 20 amino acids in the γ' chain (Wolfenstein-Todel & Mosesson, 1981). This difference in amino acid sequence accounts for the size and charge heterogeneity of the γ' chain. Accordingly, it was proposed that alternative splicing of the precursor mRNA might be responsible for the generation of the γ' chain (Wolfenstein-Todel & Mosesson, 1981). This proposal was supported by the results of Crabtree & Kant

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received April 25, 1984.* This work was supported in part by Research Grants HL 16919 and HL 28598 from the National Institutes of Health. D.W.C. is an Established Investigator of the American Heart Association.

0006-2960/84/0423-4232$01.50/0

(1982), who demonstrated that the mature mRNA for the rat γ′ chain is longer than that of the γ chain by about 500 nucleotides. Furthermore, the additional nucleotides are identical in length and in sequence with the seventh intron of the gene for the rat γ chain. On the basis of these observations, Crabtree and Kant concluded that the seventh intron was not always removed during mRNA processing and resulted in the generation of the longer mRNA retaining the entire seventh intron. Translation of this mRNA could give rise to the γ′ chain in rat liver.

By use of a cDNA specific for the γ chain (Chung et al., 1983a), the genomic DNA for the γ chain of human fibrinogen has been isolated and characterized. In this report, the isolation and characterization of cDNAs coding for the γ′ chain are described. Analyses of these cDNA and genomic sequences indicate that, in human liver, the γ′ chain is produced by alternative processing and polyadenylation rather than failure to remove the last intervening sequence during the formation of mature mRNA.

Materials and Methods

Screening of Genomic Library. A human genomic library, containing random fragments of human liver DNA fragments generated by partial *Alu*I and *Hae*III digestion and cloned into λ phage Charon 4A, was kindly provided by Dr. Tom Maniatis. Recombinant phage containing human γ-chain sequences were identified by hybridization to radiolabeled cDNA for the human γ chain (Chung et al., 1983a). The library was propagated in *Escherichia coli*, strain LE392, and the phage plaques were screened by plaque hybridization (Benton & Davis, 1977), as modified by Woo (1979). Recombinant phage were plaque purified and mapped, and segments were subcloned into plasmid pBR322.

DNA Sequence Determinations. Nonrandom DNA sequence determination in M13 phage was performed as described by Poncz and co-workers (Poncz et al., 1982). A set of progressive deletions of a segment of DNA coding for the 3′ end of the human γ-chain gene was generated by digestion with exonuclease *Bal*31. The family of overlapping DNA segments was cloned into M13mp11 in an orientation such that the deleted ends were situated immediately next to the primer site. The entire nucleotide sequence for the DNA fragment was established from sequences obtained from these overlapping clones by the chain terminator method of Sanger et al. (1977). DNA sequence was recorded and edited with the computer programs of Staden (1977).

Screening of cDNA Clones. A collection of transformed *E. coli*, containing recombinant cDNAs coding for the human γ chain (Chung et al., 1983a), was screened by hybridization to a probe specific for the γ′ sequences. The collection of transformants was plated, transferred to Whatman 541 filter paper, amplified, and prepared for hybridization by the method of Gergen and co-workers (Gergen et al., 1979). The probe was a segment of DNA from the ninth intron of the human fibrinogen γ gene cloned into the single-stranded phage M13mp11. The recombinant phage template was labeled for use as a hybridization probe employing a synthetic hybridization probe primer by the method of Hu & Messing (1982).

Results

A recombinant λ phage genomic library, containing random fragments of human liver DNA (Maniatis et al., 1978), was screened for the gene of the γ chain of human fibrinogen. Overlapping phage containing sequences for the human γ chain were identified and isolated on the basis of hybridization to radiolabeled plasmid pHIγ2, a cDNA coding for the γ chain of human fibrinogen (Chung et al., 1983a). The gene for the γ chain was found to be encoded in two adjacent *Eco*RI fragments 7.5 kb and 5.7 kb in length. The coding region for the γ chain is interrupted by nine intervening sequences (D. W. Chung, M. W. Rixon, and E. W. Davie, unpublished results).

An analysis of the sequence for the gene for the γ chain indicated that the nucleotide sequences coding for the carboxyl-terminal 20 amino acid residues in the polypeptide for the γ′ chain (Wolfenstein-Todel & Mosesson, 1981) are located in the ninth intron of the gene. Figure 1 shows the nucleotide sequence of a segment of DNA from the 3′ end of the gene coding for the γ chain of human fibrinogen. The sequence was determined from a 1-kb *Hin*dIII restriction fragment that extends from a *Hin*dIII site in the eighth intron to a *Hin*dIII site in the tenth exon. The last four amino acid residues (408–411) of the γ chain (Ala-Gly-Asp-Val), together with the 3′ noncoding sequences, constitute the tenth exon. The ninth intron or intervening sequence interrupts the coding region between Gln-407 and Ala-408 and is 513 nucleotides in length. The immediate 5′ end of the ninth intron codes for the alternative carboxyl-terminal peptide of 20 amino acids of the γ′ chain. These amino acids replace the last four amino acids of the more abundant γ chain. Thus, the ninth intron is serving as an extension of the ninth exon in coding for the carboxyl-terminal region of the γ′ chain.

The above data suggest that the mRNA for the γ′ chain contains either a portion of the ninth intron or the entire ninth intron. In either event, translation for the γ′ chain would proceed from the ninth exon into the ninth intron and terminate at the first in-phase stop codon. In order to test for these two possibilities, it was necessary to isolate cDNA clones corresponding to the mRNA for the γ′ chains. In these experiments, a specific hybridization probe for the γ′ chain was constructed by subcloning a small fragment of DNA from the ninth intron (nucleotide residues 48–118 in Figure 1). This fragment was prepared by deleting all of the nucleotide sequences from the 3′ end of the *Hin*dIII fragment by exonuclease III and limited *Bal*31 digestion. This removes sequences downstream from nucleotide 118 of the ninth intron that are rich in AT. A fragment of 70 base pairs (nucleotides 48–118, underlined in Figure 1) was then generated by digestion with *Mst*II and made blunt ended by the fill-in reaction employing the Klenow fragment of *E. coli* polymerase I. It was then cloned into the *Sma*I site of M13mp11. Single-stranded recombinant phage DNA was then prepared, and the sequence of the insert was verified by dideoxy sequencing.

The single-stranded phage DNA template containing the 70 base pair fragment specific for the γ′ chain was labeled to high specific activity by the fill-in reaction employing the Klenow fragment of *E. coli* polymerase I using a synthetic hybridization probe primer according to the method of Hu & Messing (1982). The labeled template was then used to rescreen more than 300 cDNA clones previously identified as coding for the human γ chain (Chung et al., 1983a). The specificity of this probe was confirmed by the absence of hybridization to cDNAs for the γ chain, including pHIγ1 and pHIγ2. Ten recombinant plasmids specifically hybridized to this probe. The three longest plasmids, designated pHIγ′1, pHIγ′2, and pHIγ′3, were further characterized by restriction mapping and nucleotide sequencing.

Analyses of the genomic sequences for the γ chain predict that a cDNA for the γ′ chain would contain an *Mst*II restriction site and that this site would be absent in the cDNA for the γ chain. Furthermore, the inclusion of a part of the

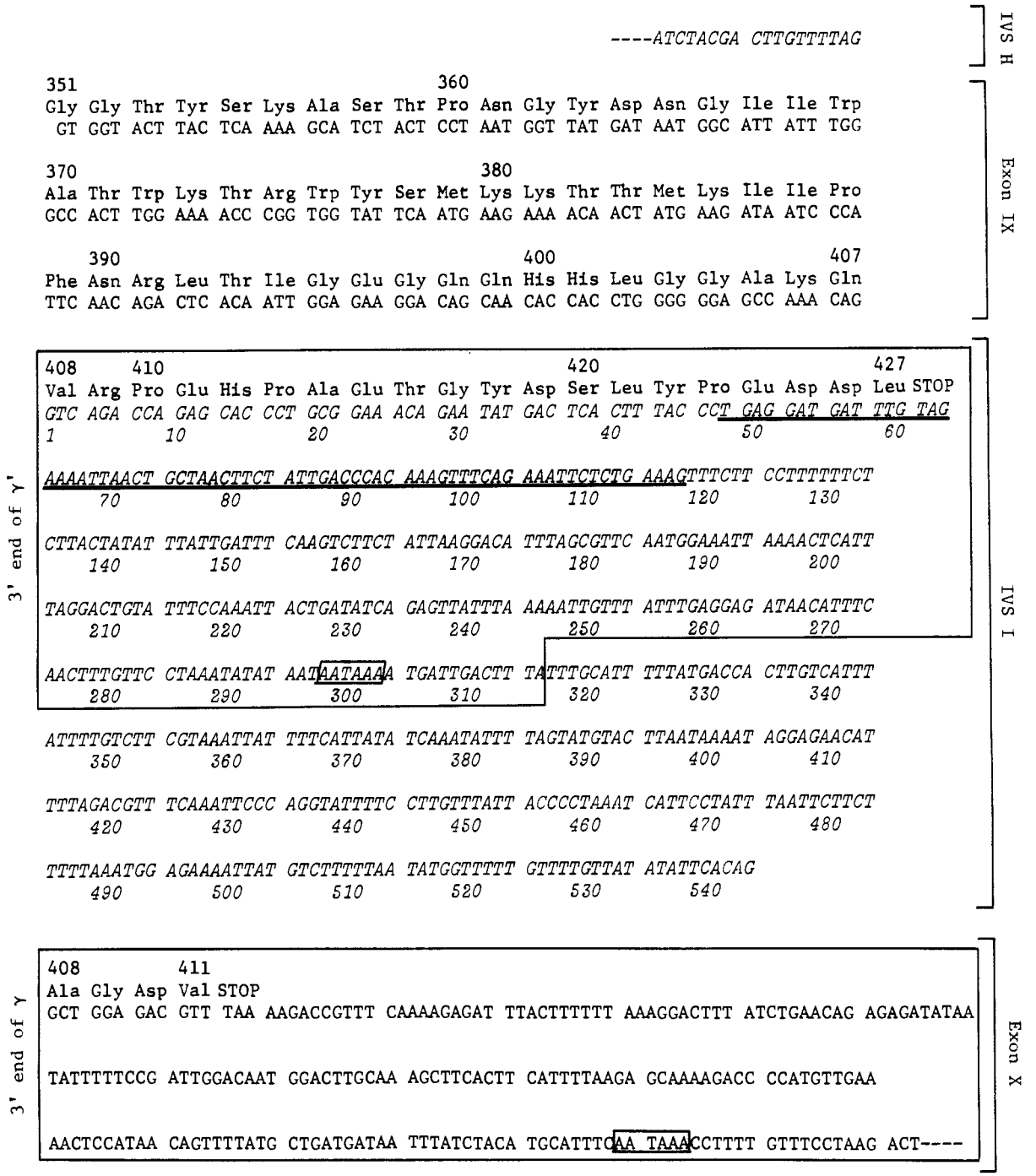

FIGURE 1: Nucleotide sequence from the 3′ end of genomic DNA for the γ chain of human fibrinogen. The sequence was determined from a subclone containing a 1-kb *Hin*dIII fragment extending from a *Hin*dIII site in the eighth intron to a *Hin*dIII site in the last (tenth) exon. Regions of the sequence that contain the 3′ ends of the γ and γ′ cDNAs are boxed. The underlined region within the intervening sequence represents a fragment that was cloned and used as a specific hybridization probe for the γ′ chain.

ninth intron in the γ′-chain cDNA would eliminate a *Bgl*I restriction site that is generated as a result of correct splicing of exons eight and nine. A restriction map of one of the longest plasmids (pHIγ′1) identified by the γ′ specific probe is shown in Figure 2 along with a restriction map of a cDNA coding for the γ chain (pHIγ2). Inserts from plasmid pHIγ1 contain a recognition sequence for *Mst*II but not for *Bgl*I or *Hin*dIII that is present in plasmid pHIγ2. The distance of the *Mst*II site to the 3′ end of the cDNA for the γ′ chain was aproximately 270 base pairs. These data show that the cDNA for the γ′ chain contains only a portion of the ninth intron. The lack of the *Bgl*I site confirms the absence of the predominant splicing of the eighth and ninth exons. The lack of the *Hin*dIII site indicates that sequences of the tenth exon are also absent. These results suggest that the cDNAs for the γ′ chain terminate within the ninth intron.

Nucleotide sequence analyses were then performed on plasmid pHIγ′1 to determine the exact 3′ end of the cDNA. This plasmid was found to be 1110 nucleotides in length. Its sequence was collinear with the sequence for the cDNA for the γ chain up to the nucleotide triplet of CAG. This triplet corresponds to amino acid residue 407 (the end of the ninth exon). The nucletide sequence continued, as predicted, into the ninth intron and terminated at nucleotide 315 of the ninth intron (Figure 1). Furthermore, it was polyadenylated at its 3′ end, and a putative processing or polyadenylation signal of AATAAA (Proudfoot & Brownlee, 1976) was identified 13 nucleotides upstream from the poly(A) tail. These data in-

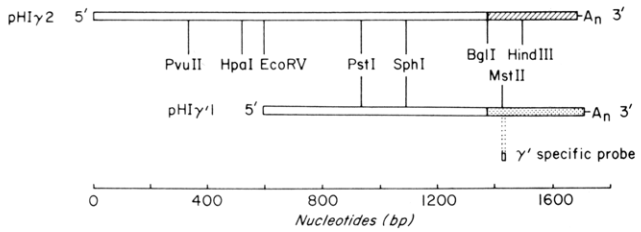

FIGURE 2: Restriction map for the 3′ ends of the cDNAs coding for the γ chain (pHIγ2) and the γ′ chain (pHIγ′1) of human fibrinogen. The slashed and dotted bars represent regions in the two cDNAs where the DNA sequences differ, and the open bars represent regions where the two DNA sequences are identical. The dotted lines below the restriction map of pHIγ′1 show the location of the DNA fragment that was used as a specific hybridization probe for cDNAs coding for the γ′ chain. The poly(A) tails are shown by A_n. The restriction map for pHIγ2 was from Chung et al. (1983a).

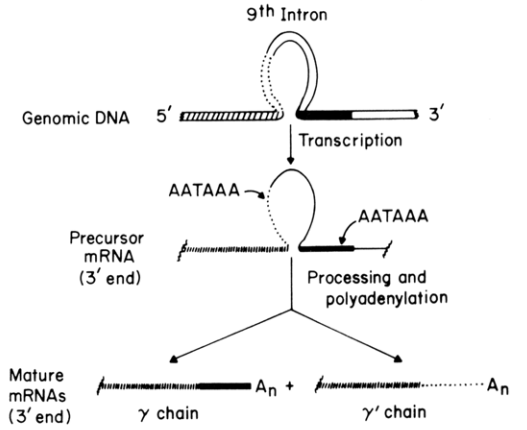

FIGURE 3: Mechanism for the formation of the mRNAs for the γ and γ′ chains of human fibrinogen. Only the 3′ ends for the genomic DNA, precursor mRNA, and mature mRNAs are shown. The ninth and tenth exons in the genomic DNA are shown by a slashed bar and a solid bar, respectively. Diagonal lines represent a coding region common to both chains in their mRNAs, while solid bars represent an additional coding region for the γ chain and dotted lines represent an additional coding region for the γ′ chain. The poly(A) tails are shown by A_n.

dicate that the γ′ chain in human fibrinogen results from an alternative processing and polyadenylation reaction that occurs in the ninth intervening sequence of the γ chain.

The restriction maps of the inserts from pHIγ′2 and pHIγ′3 are collinear with that of pHIγ′1 (data not shown). The inserts of pHIγ′2 and pHIγ′3 are 920 base pairs and 780 base pairs in length, respectively. Nucleotide sequence analysis indicates that pHIγ′2 is polyadenylated at its 3′ end at the same position as pHIγ′1. pHIγ′3 does not contain a poly(A) extension at its 3′ end, and the sequence stops at nucleotide 321 of the ninth intron.

Discussion

The present data provide evidence for two alternative processing and polyadenylation reactions leading to the formation of different mRNAs for the γ and γ′ chains of human fibrinogen (Figure 3). In this mechanism, an alternative processing and polyadenylation site of AATAAA located in the ninth intron is occasionally selected as well as the predominant polyadenylation site in the tenth exon. This leads to the termination and polyadenylation of the precursor mRNA within the ninth intron and removal of the 3′ end of the splice junction in the ninth intron. Maturation of this precursor molecule gives rise to an mRNA that codes for 20 different amino acids on the carboxyl-terminal end of the less prevalent γ′ chain.

This mechanism differs from that previously proposed for the γ′ chain in rat (Crabtree & Kant, 1982). The mRNA for the γ′ chain in rat liver apparently retains the entire last intron, and thus, its size is longer than the γ chain by the length of this intron. As shown in the present experiments, the difference in size between mRNAs for the γ and γ′ chains of human fibrinogen is only 90 nucleotides. Multiple processing and polyadenylation sites, however, do not rule out the possibility that in human liver some γ′ chains may also occur as a result of an occasional failure to splice out the ninth intron in the γ chain. Studies on the cDNAs for the β chain of human fibrinogen have shown that the mRNA for this polypeptide is polyadenylated in at least three discrete sites (Chung et al., 1983b). In this case, the polyadenylation sites are all located in the 3′ noncoding region of the gene and do not result in alternative splicing of the mRNA and an alteration in the amino acid sequence.

Present evidence indicates the existence of only one gene for the γ chain of human fibrinogen (D. W. Chung, M. W. Rixon, and E. W. Davie, unpublished results). Also, the two alleles coding for the γ chain in human DNA are coexpressed. This is exemplified in the case of a congenital fibrinogen abnormality, called γ Paris I, in which a change(s) affecting one of the alleles results in the replacement of about 50% of the normal γ chain by a functionally defective form. The γ Paris I fibrionogen, analogous to that in normal individuals, exhibits the typical γ/γ′ heterogeneity (Budzynski et al., 1974; Mosesson et al., 1976; Stathakis et al., 1978). These observations are consistent with the present evidence demonstrating the γ and γ′ chains result from a differential expression of a single gene for this polypeptide.

Alternative processing has been observed for a number of other proteins such as immunoglobulins (Alt et al., 1980; Rogers et al., 1980; Early et al., 1980; Maki et al., 1981), vimentin (Capetanaki et al., 1983), fibronectin (Schwarzbauer et al., 1983), high and low molecular weight kininogens (Kitamura et al., 1983), calcitonin (Amara et al., 1982), and crystallin (King & Piatigorsky, 1983). In the cases of immunoglobulin heavy chains, calcitonin, and α-amylase (Young et al., 1981), alternative processing has been shown to be tissue specific. In addition to alternative splicing, the utilization of alternative polyadenylation sites, as we have shown here for the γ chain of human fibrinogen, provides a mechanism for variation in protein structure resulting from a single gene. Whether the γ and γ′ polypeptides in fibrinogen have different biological functions, however, remains to be determined.

Acknowledgments

We thank Dr. Jon Herriott for providing the computer facilities used in storing and analyzing DNA sequences. We thank Drs. Wai-Yee Chan and Mark Rixon for helpful discussion.

References

Alt, F. W., Bothwell, A. L. M., Knapp, M., Siden, E., Mather, E., Koshland, M., & Baltimore, D. (1980) *Cell* (*Cambridge, Mass.*) *20*, 293–301.

Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S., & Evans, R. M. (1982) *Nature* (*London*) *298*, 240–244.

Benton, W. D., & Davis, R. W. (1977) *Science* (*Washington, D.C.*) *196*, 181–182.

Blomback, B., & Blomback, M. (1972) *Ann. N.Y. Acad. Sci. 202*, 77–97.

Budzynski, A. Z., Marder, V. J., Menache, D., & Guillin, M. C. (1974) *Nature* (*London*) *252*, 66–67.

Capetanaki, Y. G., Ngai, J., Flytzanis, C. N., & Lazarides, E. (1983) *Cell (Cambridge, Mass.) 35*, 411–420.

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

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

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

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., & Hood, L. (1980) *Cell (Cambridge, Mass.) 20*, 313–319.

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

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

Henschen, A., & Edman, P. (1972) *Biochim. Biophys. Acta 263*, 351–367.

Hu, N., & Messing, J. (1982) *Gene 17*, 271–277.

King, C. R., & Piatigorsky, J. (1983) *Cell (Cambridge, Mass.) 32*, 707–712.

Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H., & Nakanishi, S. (1983) *Nature (London) 305*, 545–549.

Legrele, C. D., Wolfenstein-Todel, C., Hurbourg, Y., & Mosesson, M. W. (1982) *Biochem. Biophys. Res. Commun. 105*, 521–529.

Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W., & Tonegawa, S. (1981) *Cell (Cambridge, Mass.) 24*, 353–365.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, D. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–702.

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

Mosesson, M. W., Finlayson, J. S., & Umfleet, R. A. (1972) *J. Biol. Chem. 247*, 5223–5227.

Mosesson, M. W., Amrani, D. L., & Menache, D. (1976) *J. Clin. Invest. 57*, 782–790.

Mosher, D. F., & Blout, E. R. (1973) *J. Biol. Chem. 248*, 6896–6903.

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

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., & Wall, R. (1980) *Cell (Cambridge, Mass.) 20*, 303–312.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R., & Hynes, R. O. (1983) *Cell (Cambridge, Mass.) 35*, 421–431.

Staden, R. (1977) *Nucleic Acids Res. 4*, 4037–4051.

Stathakis, N. E., Mosesson, M. W., Galanakis, D. K., & Menache, D. (1978) *Thromb. Res. 13*, 467–475.

Wolfenstein-Todel, C., & Mosesson, M. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5069–5073.

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

Woo, S. L. C. (1979) *Methods Enzymol. 68*, 381–395.

Young, R. A., Hagenbuchle, O., & Schibler, U. (1981) *Cell (Cambridge, Mass.) 23*, 451–458.