Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 7179–7182, October 1987
Genetics

Identification of a third region of cell-specific alternative splicing in human fibronectin mRNA

(cell and plasma fibronectins/internal homologies/differential mRNA processing)

ALEJANDRO GUTMAN AND ALBERTO R. KORNBLIHTT*

Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI–CONICET), Vuelta de Obligado 2490, 1428 Buenos Aires, Argentina

Communicated by Osvaldo A. Reig, May 22, 1987 (received for review January 8, 1987)

ABSTRACT **We describe here a third region of variability in human fibronectin (FN) due to alternative RNA splicing. Two other positions of alternative splicing have been reported previously (ED and IIICS). The third region involves a 273-nucleotide exon encoding exactly one 91-amino acid repeat of type III homology, located between the DNA- and the cell-binding domains of FN, which is either included in or excluded from FN mRNA. The two mRNA variants arising by an exon-skipping mechanism are present in cells known to synthesize the cellular form of FN. However, liver cells, which are the source of plasma FN, produce only messengers without the extra type III sequence. Therefore, the region described here resembles, both structurally and functionally, the previously described ED (for extra domain) region, located toward the C terminus of the molecule, between the cell- and heparin- (hep 2) binding domains. We conclude that both the extra type III repeat (named EDII) and ED represent sequences restricted to cellular FN. Combination of all the possible patterns of splicing in the three regions described to date may generate up to 20 distinct FN polypeptides from a single gene.**

Fibronectins (FNs) are high molecular weight glycoproteins found in plasma, on cell surfaces, and in extracellular matrices. By binding macromolecules such as collagen, fibrin, heparin, and themselves, as well as cells and bacteria, FNs play an important role in various contact processes within vertebrates, such as cell attachment and spreading, cell migration, embryonic development, wound healing, hemostasis, opsonization, and oncogenic transformation (for reviews see refs. 1–3).

Cell FN (cFN) and plasma FN (pFN) are two different products of a single gene (4). cFN is synthesized by fibroblasts and other cell types, which secrete it into the extracellular matrix, where it assembles in a fibrillar form. In contrast, pFN is made by hepatocytes (5) and circulates as a soluble molecule in plasma. Both forms are disulfide-linked dimers of ≈250-kDa subunits, taken from a pool of similar but not identical polypeptides. Sequencing studies have revealed that each polypeptide has three different types of internal repeats (homology types I, II, and III, which are, on average, 40, 60, and 90 amino acid residues long, respectively) (6, 7). The primary structural differences between the subunits are the result of a complex pattern of alternative splicing of the precursor mRNA. To date, two regions of alternative splicing have been identified. The IIICS (for type III connecting segment) region is located between the penultimate and the last type III repeats; it is encoded by a single exon, which is subdivided to yield three mRNA variants in rats (8) and up to five variants in humans (7). Apparently, variability in this region explains the difference in size between larger and smaller subunits of both cFN and pFN (9). The second region, named ED for extra domain, involves the inclusion in or exclusion from the mRNA of a 270-nucleotide exon coding for exactly one unit of type III homology (10). The ED sequence was identified in mRNA from fibroblasts and several cell lines, but it could not be detected in liver mRNA, leading to the prediction that the extra type III unit was a unique domain of cFNs (11). This was recently confirmed by the use of antibodies raised against ED sequences: the antibodies recognize cFN but not pFN (9, 12).

This report describes the existence of a third region of alternative splicing in human pre-mRNA, similar in structure and cell-specificity to the ED region. The third alternative splicing position gives rise to two further variants, which may contribute to the generation of up to 20 polypeptides from a single gene.

MATERIALS AND METHODS

Cell Cultures. The human teratocarcinoma cell line 2102Ep (13) was kindly provided by C. E. Graham (Department of Zoology, Oxford). The carcinosarcoma cell line HS578T (14) was cultured in Dulbecco's modified Eagle's medium containing 10% fetal calf serum.

RNA Preparation. Total RNA was extracted from confluent cell monolayers and from a human liver biopsy sample by the guanidine hydrochloride method (15).

DNA Clones. Isolation of the human genomic clone λFN4 containing 15.8 kilobases (kb) of the ≈50-kb FN gene is described elsewhere (10), as are cDNA clones pFH2 (16) and pFH154 (11).

Subcloning. A total *Eco*RI digest of λFN4 was cloned in pAT153/*Pvu* II/8 (17). Two subclones were isolated: pFN4E4.4, which has a 4.4-kb insert that hybridized to a cDNA probe of pFH154 spanning the C-terminal third of the type III repeat in line 21 (7); and pFN4E0.65, which has a 0.65-kb insert that hybridized to a cDNA probe of pFH154 spanning the N-terminal third of the type III repeat in line 22 (7). A 2.4-kb *Sac* I–*Eco*RI fragment from pFN4E4.4 was subcloned in M13mp18, yielding clone M13FN4S-E2.4. From the latter, a shorter clone named M13FN4Δ0.9, with a deletion starting from the *Sac* I end, was obtained by using the Cyclone System kit (International Biotechnologies, New Haven, CT). Two other subclones from pFN4E4.4 were obtained in the pAT vector for sequencing purposes: pFN4H0.5 and pFN4H0.95. All these subclones are depicted in Fig. 1*B*.

DNA Sequencing. DNA sequences were determined by the procedures of Maxam and Gilbert (18) and Sanger *et al.* (19).

S1 Nuclease Mapping. Two probes were prepared: A 220-base-pair (bp) *Hin*fI–*Hae* III fragment from pFN4H0.95 labeled at its *Hin*fI end and a 135-bp *Ava* I–*Eco*RI fragment from pFH2 labeled at its *Ava* I end. Probes were labeled by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: FN, fibronectin; cFN, cell FN; pFN, plasma FN; ORF, open reading frame.

*To whom reprint requests should be addressed.

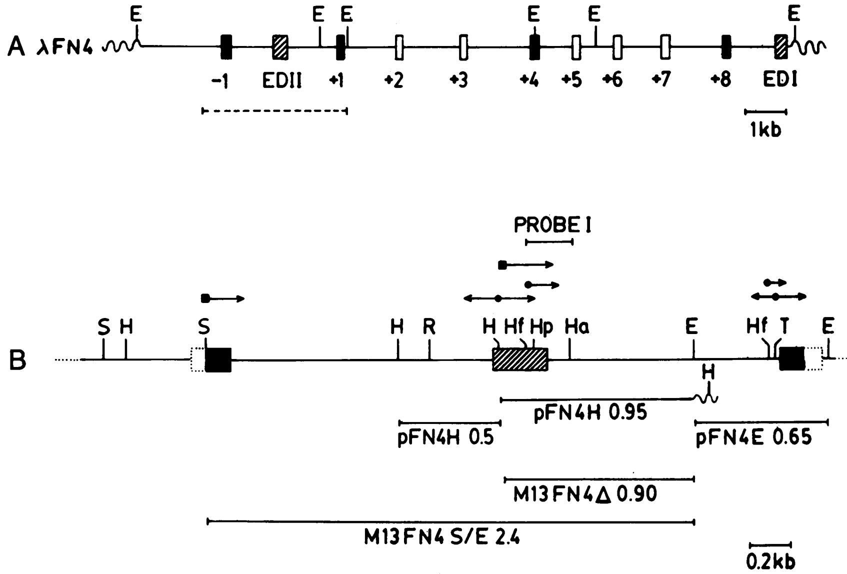

FIG. 1. (*A*) Map of the human FN gene in the region covered by clone λFN4. Wavy lines, vector sequences; E, *Eco*RI sites; empty boxes, exons whose exact position and length have not been determined; black and hatched boxes, exons whose exact position and length are known; hatched boxes, exons involved in alternative splicing. Exons were numbered with respect to EDII. (*B*) Expanded map of the part of the gene underlined by a broken line in *A* and length of the subclones obtained. The strategy for sequencing is shown with arrows. Arrows starting with ● indicate Maxam and Gilbert (18) sequencing; arrows starting with ■ indicate Sanger *et al.* (19) sequencing. Values listed after the subclone designations represent lengths in kb. E, *Eco*RI; H, *Hin*dIII; Ha, *Hae* III; Hf, *Hin*fI; Hp, *Hpa* II; R, *Rsa* I; S, *Sac* I; T, *Taq* I. The restriction enzyme sites used for sequencing are not all the sites in the sequence. Wavy line, vector sequence.

filling in with the Klenow fragment of DNA polymerase I and [α-^{32}P]dATP and [α-^{32}P]dCTP, respectively. Probes were hybridized to 40 μg of either cell line or liver total RNA in formamide at 52°C. Hybridization and subsequent steps were as described (11).

RESULTS

An Exon Absent from cDNA Clones. Southern blot analyses of restriction enzyme digests (results not shown) of λFN4, a human FN genomic clone (10), using several cDNA probes, allowed us to draw a map of a part of the human FN gene, shown in Fig. 1*A*. λFN4 covers 15.8 kb towards the middle of the human FN gene, with coding regions that extend from one of the type III repeats having the free cysteine (located between the DNA- and cell-binding domains) to the ED exon. Several subclones were obtained and analyzed, mainly in the area where an exon absent from existing human cDNA clones has been found in the rat FN gene (24). Mapping and sequence analysis of the subclones depicted in Fig. 1*B* showed the existence of an open reading frame (ORF) within the 2.7-kb intron of λFN4. Comparison of the ORF amino

A

```
        V  P  I  S  D  T  I  I  P
.....GTCCCTATCTCTGATACCATCATCCCAGgtaatagaaaat.........gctttattttcctctttttccctctatttcctttgcctccccctccctttt
                                                                 -60                 -40                 -20
                   E  V  P  Q  L  T  D  L  S  F  V  D  I  T  D  S  S  I  G  L  R  W  T  P  L  N  S  S  T  I  I
gctttgtaactcaatagAGGTGCCCCAACTCACTGACCTAAGCTTTGTTGATATAACCGATTCAAGCATCGGCCTGAGGTGGACCCCGCTAAACTCTTCCACCATTATT
                  1                 20                   40                   60                   80
 G  Y  R  I  T  V  V  A  A  G  E  G  I  P  I  F  E  D  F  V  D  S  S  V  G  Y  Y  T  V  T  G  L  E  P  G  I
GGGTACCGCATCACAGTAGTTGCGGCAGGAGAAGGTATCCCTATTTTTGAAGATTTTGTGGACTCCTCAGTAGGATACTACACAGTCACAGGGCTGGAGCCGGGCATTG
     100                  120                  140                  160                  180                  200
D  Y  D  I  S  V  I  T  L  I  N  G  G  E  S  A  P  T  T  L  T  Q  Q  T                                        *
ACTATGATATCAGCGTTATCACTCTCATTAATGGCGGCGAGAGTGCCCCTACTACACTGACACAACAAACGGgtgaattttg.........taggaaaattacattaac
                  220                  240                  260
                               A  V  P  P  P  T  D  L  R
tgagtcttccttcttgttcgaaaccagCTGTTCCTCCTCCCACTGACCTGCGA.....
```

B

```
ED II  EVPQLTDLSFVDITDSSIGLRWTPLNSSTIIGYRITVVAAGEGIPIFEDFVDSSVGYYTVTGLEPGIDYDISVITLINGGESAPTTLTQQT
        :   ::: :  ::: :: : : :        : :::::          : :         ::::: ::  : :::  :    :: :     : :
       -ØD-PTDL-FT-ØTDTSØ-ØSW-PP----ØTGYRØT--P-----P--E---P----S-TØTGL-PG-EY-ØSØYAØ-D--ES-PØ-G-Q-T
        :: :  : :: :   :: : :  :     :  ::::                             : :: :: :: ::: :: :  :: :: : : :
ED I   NIDRPKGLAFTDVDVDSIKIAWESPQGQ-VSRYRVTYSSPEDGIHELFPAPDGEEDTAELQGLRPGSEYTVSVVALHDDMESQPLIGTQST
```

FIG. 2. (*A*) Nucleotide sequence and deduced amino acid sequence of the exon encoding the extra type III repeat (EDII), its exon/intron boundaries, and the boundaries of the two flanking exons (−1 and +1). Intron sequences are in lowercase letters. Underlined sequences are splicing "donor" and "acceptor" sequences and a putative branch site. The two introns follow the rule of those separating type III repeats, falling between the first and the second bases of a codon (21). After the proline of exon −1, codon recomposition makes up an alanine if EDII is absent or a glutamate if EDII is present. ●, Predicted N-linked glycosylation site; *, adenine putatively involved in lariat formation. (*B*) Comparison of the amino acid sequence of EDII (top line) and EDI (bottom line) with the consensus amino acid sequence of the other 15 human type III repeats. Ø, hydrophobic residue (i.e., Phe, Ile, Leu, Met, and Val); A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Colons indicate identities or only hydrophobic residues.

acid sequence with the consensus sequence of 16 type III repeats previously sequenced (7) revealed that the ORF is an exon precisely encoding a full type III homology region (Fig. 2*B*). This repeat is absent from two human cDNA clones that cover the area (pFH111 and pFH154) as well as from any published sequence that we know of from other species. This observation indicates that, if the repeat is transcribed, it must be taken out from at least some of the FN mRNAs. These characteristics resemble those of the ED site, so we named this type III homology sequence and its exon EDII.

Expression of EDII in FN mRNA. Fig. 2*A* shows the nucleotide sequence and deduced amino acid sequence of the EDII exon, its intron/exon boundaries, and the boundaries of two flanking exons. The "donor" and "acceptor" sites of both introns flanking EDII fit the consensus sequences for splicing (20). A putative branch site has been identified 29 nucleotides upstream of the 3′ end of the 3′ intron (20). However, no sequences matching the consensus for branch sites have been identified within 72 nucleotides upstream of the 3′ end of the 5′ intron. Instead, an unusually long row of pyrimidines (−71 to −6 in Fig. 2*A*) precedes the "acceptor" site of the 5′ intron.

These results suggested that the EDII ORF might fulfill the sequence requirements to be spliced in within mature FN RNA. To confirm this hypothesis we performed S1 nuclease mapping experiments with a genomic probe labeled within the EDII sequence, hybridizing the probe with mRNA from different sources. The probe was the 220-nucleotide 3′-end-labeled antisense strand *Hin*fI–*Hae* III fragment of clone pFN4H0.95. When this probe was hybridized to RNA from WI38 human lung fibroblasts (not shown), HS578T cells (Fig. 3, lane 1) or teratocarcinoma cells (Fig. 3, lane 3), one S1-nuclease-resistant band (band A, 118 nucleotides) was observed, besides remnants of self-hybridized probe. Band A covers the distance between the *Hin*fI site and the base where the 3′ end of the EDII exon was predicted (G number 273 in Fig. 2*A*) on the basis of the size of other type III repeats and the presence of a "donor" consensus sequence. This was confirmed by running two lanes of a Maxam and Gilbert degradation (G+A and G) of the probe in parallel, as shown in Fig. 3. This result indicated the existence of FN mature RNA containing EDII sequences.

Tissue Specificity. When similar S1 nuclease analysis was carried out with human liver RNA, no protected bands were detected (Fig. 3, lane 5), indicating that liver produces mainly forms of FN without EDII. As a positive control, the same liver RNA preparation protected 93 nucleotides from a 135-bp *Ava* I–*Eco*RI cDNA probe from pFH2, coding for cFN pre- and pro-peptides (Fig. 3, lanes 7 and 8). The difference between the input probe and the S1-nuclease-resistant band corresponds to 42 bases of vector sequences that do not hybridize to FN mRNA. This control showed at the same time that the liver mRNA preparation had FN mRNA and that liver (plasma) FN has the same pre and pro sequences as cFN.

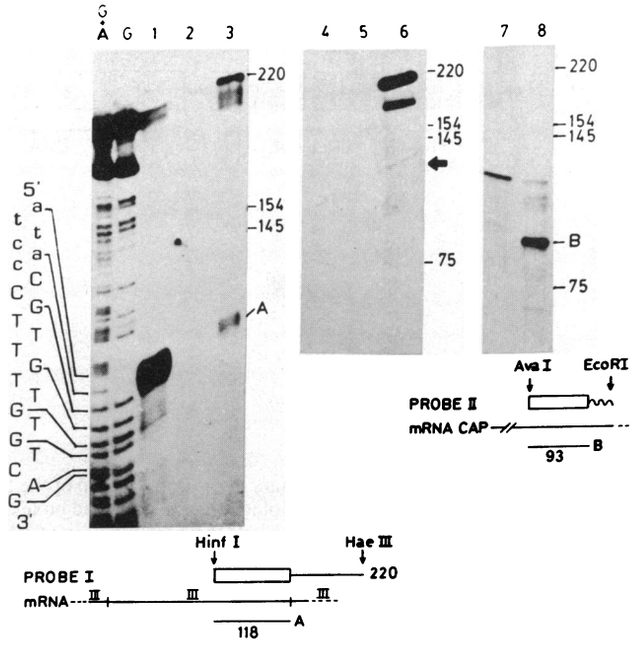

FIG. 3. S1 nuclease analysis of RNA·DNA hybrids. (*Left* and *Center*) Probe I (lanes 1–5) is the 220-bp *Hin*fI–*Hae* III fragment of pFN4H0.95 insert and it is shown in Fig. 1*B*. A sample of the probe was run in lane 6 and after G+A and G degradations of the Maxam and Gilbert procedure (18) to the left of lane 1. In both cases a spurious smaller band that did not interfere with the experiment could be seen. Hybridization was performed with human RNAs from the following sources: HS578T cells (lane 1), teratocarcinoma cells (lane 3), and liver (lane 5). Lanes 2 and 4 are controls without RNA in the hybridization mixture. (*Right*) Positive control for the presence of FN mRNA in the liver RNA preparation. Probe II, the 135-bp *Ava* I–*Eco*RI fragment of clone pFH2 (16), was hybridized to liver RNA (lane 8). Lane 7 is a control without RNA in the mixture. Probes, S1 nuclease-resistant products, and mRNA species are shown in diagrams below the figure. Correspondence between diagrams and gels is indicated by letters. The arrow indicates the expected position of band A in the experiment with liver RNA (lane 5). The autoradiograph on the left has been overexposed to show band A in lane 3.

DISCUSSION

This report provides evidence for a third position of variability due to alternative splicing in human FN. Variability is produced by the inclusion or exclusion of a full repeat of type III homology, named EDII, by virtue of its similarity to the previously described ED (10, 11). The EDII repeat is encoded by exactly one exon in the human genome and alternative splicing seems to occur by an exon-skipping mechanism (10, 21). Similarities between EDI and EDII regions are evident both structurally and functionally. In both cases full type III repeats are encoded by nonsplit (fused) exons. This is in contrast with the other nine type III repeats analyzed so far, in which two smaller exons are needed for each (Fig. 1*A* and refs. 10 and 21). Most interestingly, alternative splicing of the EDII exon seems to have the same cell specificity as that one of EDI. S1 nuclease mapping experiments shown in Fig. 3 revealed that the EDII is present in FN mRNAs from cells known to produce cFN, whereas it is not found in mRNAs from liver, the source of pFN, indicating that the type III repeat we describe here is restricted to cell surface FN, as happens with EDI.

Consensus sequences for splicing are present in both introns flanking the EDII exon. However, only the 3′ intron presents a canonical branch point. The absence of a typical branch point and the presence of an unusually long row of pyrimidines preceding the "acceptor" site of the 5′ intron might be related to the particular mechanism of the EDII splicing. It has been reported that deletions of mammalian branch sites located 20–50 nucleotides upstream of the 3′ splice site slow the splicing reaction by a variable amount while other sequences are used as branch sites (20).

Concerning the primary structure of FN, the longest human FN polypeptide should be 2446 amino acids long, including EDII, EDI, and the 120-amino acid variant of the IIICS area, and it would have 17 type III repeats (Fig. 4). Thirty-one amino acids corresponding to signal and pro sequences should be added if immature FN is considered (16). If one assumes that all the permutations in the three regions of variability are possible, at least 20 different human FN polypeptides could be generated from a single gene in

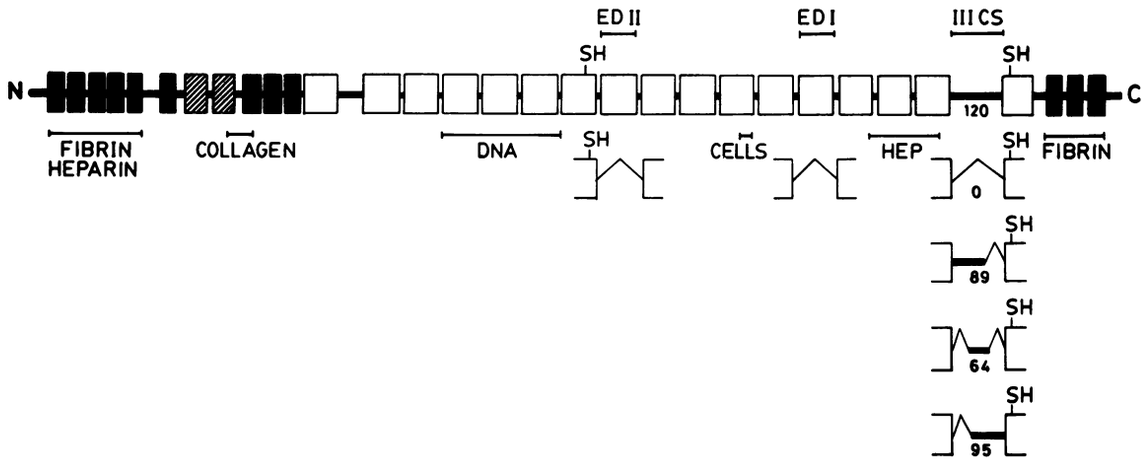

FIG. 4. Variations of the FN primary structure. The top of the figure represents the longest possible human FN pre-pro-polypeptide of 2476 amino acids. Black boxes, type I homology regions; hatched boxes, type II homology regions; empty boxes, type III homology regions. Binding regions are bracketed (HEP, heparin). All the possible variants that could arise by alternative splicing (EDII, EDI, and IIICS) are shown.

cells producing cFN, whereas 5 FN polypeptides should be detected in heptatocytes. The lack of cDNA clones that cover contiguously the three variability regions prevented assessing all the permutations. However, clones pFH111 and pFH154 (11) reflect the existence of two combinations of FN mRNA: $EDII^-$ EDI^+ and $EDII^-$ EDI^-. Preliminary results (not shown) indicate that the other two combinations ($EDII^+$ EDI^+ and $EDII^+$ EDI^-) also exist.

The EDII amino acid sequence shows one consensus sequence for N-linked glycosylation (Fig. 2*A*), increasing to 11 the maximum number of putative N-linked oligosaccharide chains per FN subunit. The control of the presence of this glycosylation site by alternative splicing may have functional implications. Although no specific functions can be assigned to the EDII and EDI sequences yet, their particular locations along the FN molecule seem interesting: EDII is between the DNA- and cell-binding domains, and EDI is between the cell- and heparin- (hep 2) binding domains (Fig. 4).

Increasing numbers of eukaryotic genes are now known to generate protein diversity by controlled mechanisms of alternative RNA processing (22). Among these, perhaps the most complex case is that of the rat skeletal fast troponin T gene, for which intricate combinatorial patterns of splicing generate a minimum of 10, and potentially 64, distinct troponin T isoforms (23). Two internal regions of differential processing contribute to troponin T diversity. Besides the FN gene, we know of no reported case in which three separate internal regions participate in the generation of diversity.

The definition of primary structural differences between cFN and pFN has proved to be useful in the design of clinical tests for tissue injury, as recently reported by Peters *et al.* (12). On the other hand, results reported here may contribute to the study of the molecular bases of cell-specific splicing, which are still poorly understood.

We thank Drs. M. Lucero and M. T. Franze for kindly providing the Cyclone System kit and Dr. S. Iapalucci for her valuable advice on its use. This work was partially supported by a grant from the Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina (CONICET). A.G. is a recipient of a research fellowship from CONICET and A.R.K. is a Career Investigator of the CONICET.

1. Yamada, K. M. (1983) *Annu. Rev. Biochem.* **52,** 761–799.
2. Hynes, R. O. (1985) *Annu. Rev. Cell Biol.* **1,** 67–90.
3. Owens, R., Kornblihtt, A. R. & Baralle, F. E. (1986) in *Oxford Surveys on Eukaryotic Genes*, ed. McLean, N. (Oxford Univ. Press, Oxford), Vol. 3, pp. 141–160.
4. Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3218–3222.
5. Tamkun, J. W. & Hynes, R. O. (1983) *J. Biol. Chem.* **258,** 4641–4647.
6. Petersen, T. E., Thøgersen, H. C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 137–141.
7. Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* **4,** 1755–1759.
8. Tamkun, J. W., Schwarzbauer, J. E. & Hynes, R. O. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5140–5144.
9. Paul, J. I., Schwarzbauer, J. E., Tamkun, J. W. & Hynes, R. O. (1986) *J. Biol. Chem.* **261,** 12258–12265.
10. Vibe-Pedersen, K., Kornblihtt, A. R. & Baralle, F. E. (1984) *EMBO J.* **3,** 2511–2516.
11. Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1984) *EMBO J.* **3,** 221–226.
12. Peters, J. H., Ginsberg, M. H., Bohl, B. P., Sklar, L. A. & Cochrane, C. G. (1986) *J. Clin. Invest.* **78,** 1596–1603.
13. Cossu, G., Andrews, P. W. & Warren, L. (1983) *Biochem. Biophys. Res. Commun.* **111,** 952–957.
14. Hackett, A. J., Smith, H. S., Springer, E. L., Owens, R. B., Nelson-Rees, W. A., Riggs, J. L. & Gardner, M. B. (1977) *J. Natl. Cancer Inst.* **58,** 1795–1800.
15. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
16. Gutman, A., Yamada, K. M. & Kornblihtt, A. R. (1986) *FEBS Lett.* **207,** 145–148.
17. Anson, D., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A. & Brownlee, G. G. (1984) *EMBO J.* **3,** 1053–1064.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
20. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
21. Oddermatt, E., Tamkun, J. W. & Hynes, R. O. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6571–6575.
22. Leff, S. E., Rosenfeld, M. G. & Evans, R. M. (1986) *Annu. Rev. Biochem.* **55,** 1091–1117.
23. Breitbart, R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V. & Nadal-Ginard, B. (1985) *Cell* **41,** 67–82.
24. Schwarzbauer, J. E., Patel, R. S., Fonda, D. & Hynes, R. O. (1987) *EMBO J.*, in press.