Volume 16 Number 8 1988 Nucleic Acids Research

Sequence analysis and *in vivo* expression show that alternative splicing of ED-B and ED-A regions of the human fibronectin gene are independent events

Giovanni Paolella, Claire Henchcliffe, Gianfranco Sebastio and Francisco E.Baralle

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received December 15, 1987; Revised and Accepted March 11, 1988

ABSTRACT

The structure of two alternatively spliced regions, ED-A and ED-B, of human fibronectin gene, was determined, in order to show whether any similarity was present between the two. Although some interesting features are present in each, no obvious common structure or sequence homology was found. Functional analysis of the alternative splicing events was carried out by transient expression in Hela cells. A hybrid gene was constructed by inserting the ED-B region into the third exon of the human α1-globin gene. The transfected hybrid gene is expressed and produces, in Hela cells, two alternatively spliced RNAs, showing a pattern very similar to that observed for the endogenous fibronectin gene in fibroblasts. Cotransfection of this gene with a similar gene containing the ED-A region, shows that no interference is present between the two alternative splicing processes.

INTRODUCTION

Fibronectin is a high molecular weight glycoprotein present in the extracellular matrix, as well as in plasma and in other body fluids. It is a dimer composed of two identical or slightly different molecules, linked by disulfide bridges (for reviews see 1, 2). Fibronectin monomers have a molecular weight ranging between 250 and 280 KDaltons and show a remarkable heterogeneity in the amino acid sequence. Functionally it is possible to distinguish two major types of fibronectin molecules: cellular fibronectin, which is insoluble and is present in the extracellular matrix, and is mainly involved in connecting the cell membrane to the basal lamina, and plasma fibronectin, which is found in soluble form in plasma and in other body fluids.

All the various fibronectin monomers, despite the sequence variations, derive from a single gene either by alternative splicing of a common precursor or from posttranslational modifications of the protein (1, 2). Three sites of alternative processing have been described up to now both in human and rat: the ED region (3, 4), now called ED-A, the IIICS region (5), and, more recently, the ED-B region originally described by Hynes et al. (6) and by our group in man (7). The IIICS region shows a complex splicing pattern, giving rise to five different mRNAs. The ED-

A and the ED-B regions, showing two possible splicing patterns each, contribute additional variability, leading to a theoretical number of 20 different messenger RNAs out of a single precursor. Many of the possible combinations have already been shown in vivo. The alternatively spliced RNAs are not equally represented in all the adult cells, but show different tissue specific distribution (8, 6), and some further variations have been observed in tumour cells (7) and in foetal tissues(1).

ED-A and ED-B regions, unlike the IIICS one, show a remarkable similarity. In both cases the alternative splicing causes the insertion of an additional peptide 91 amino acids long in the middle of the protein, without introducing frameshifts. These extra domains are inserted in a repetitive region consisting of 11 homologous domains, called type III homologies, and the inserted peptides are type III homologies themselves. The gene organization is also very similar, and the alternative splicing involves exon skipping in both cases. ED-A and ED-B exons are both absent in liver fibronectin mRNA, and both are present in fibroblasts and in cells from several other tissues, although at very different levels.

Because of the striking similarities between the two systems, it is possible that similar signal sequences are responsible for the specific splicing patterns in both regions. Moreover, since tissue specific mRNAs are produced, even starting from a common precursor, they may derive from the interaction between the precursor mRNA and cell type specific trans-acting factors; it is possible that one or more factors are common to both systems. On the other hand, although the tissue specific splicing patterns of ED-A and ED-B are similar, still there are important differences and it is possible that the two alternative splicing events are unrelated, each of them occurring independently of the other.

In this paper we report the gene structure of the ED-A and the ED-B region of human fibronectin and we show that, although some interesting characteristics are present in each region, they are not common to both. We also describe the expression of the ED-B region in transfected Hela cell and show that the information contained in the sequence is able to accurately reproduce the same splicing patterns as occurring in vivo in fibroblasts. Furthermore we show that no interference is present between the two systems when ED-A and ED-B genes are transfected together in Hela cells.

MATERIALS AND METHODS

Chemicals were from Sigma, Analar and BDH. Restriction enzymes and other DNA modifying enzymes were from Amersham or Biolabs and were used according to the manufacturers instructions. All common DNA manipulations were according to established techniques (9).

Sequencing of the ED-A and ED-B regions of the human fibronectin gene.

M13 recombinant subclone MA10 (7), containing the ED-B region of human fibronectin gene, was completely sequenced. The sequence of the ED-A region was determined from two subclones, constructed by inserting into M13mp19 the PstI and the BstEII fragment from the human fibronectin gene, respectively containing the two introns flanking the ED-A exon, complete with the relative splice junctions. A nested set of partial deletions was produced by unidirectional digestion with Exo III, followed by S1 treatment and religation (10). A few subclones were also produced by using convenient restriction sites. The sequence of the subclones so produced was then determined by the method of Sanger (11). The sequence was almost completely determined on both strands. Some restriction fragments were also sequenced according to Maxam and Gilbert (12).

Subcloning and transient expression in cells.

Clone Bra 7 was constructed by inserting the ED-B and part of the neighbouring exons, together with the relative introns, in the third exon of the human α1-globin gene. The ED-B region was excised from clone MA10 as a 2.8 kb Sac I-Bam HI fragment and cloned in pSV-ED-a1W (13), in the BstEII site. All the ends were made blunt by treatment with the Klenow fragment of DNA polymerase I. Bra 11 is the same pSV-ED-alfa1W carrying the M13 polylinker in the BstEII site. FNED (14) is

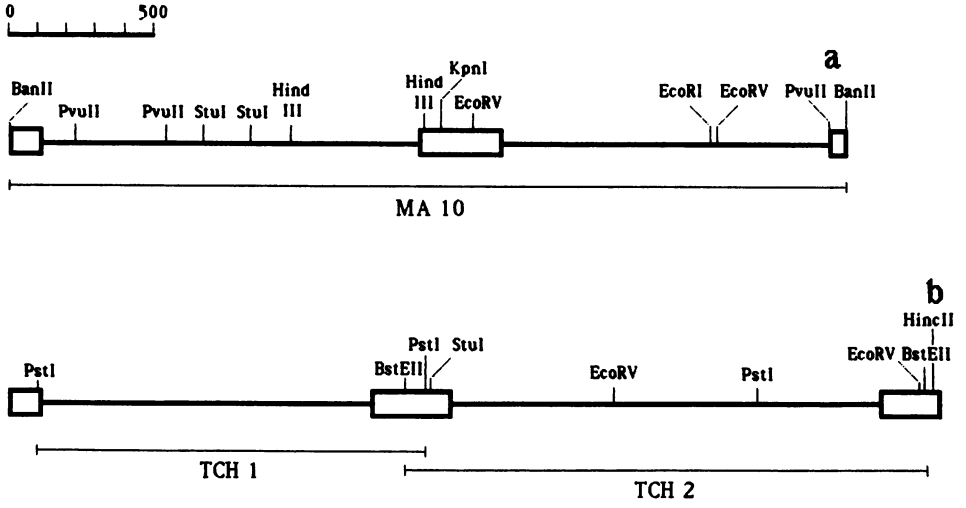
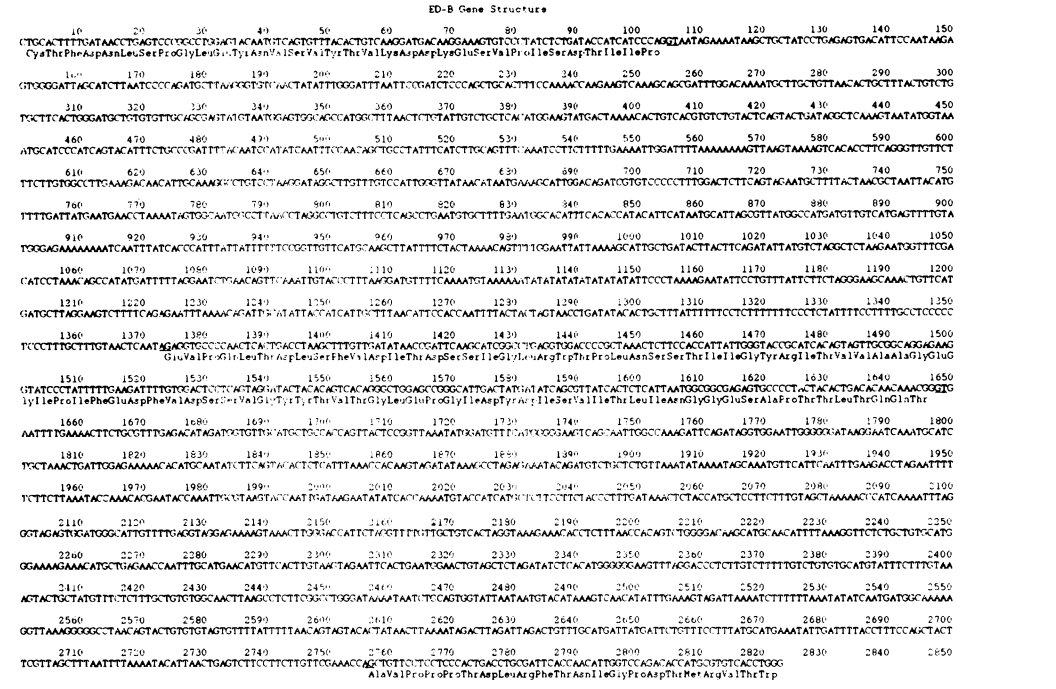


Figure 1. Genomic organization of (a)ED-B and (b)ED-A regions of the human fibronectin gene. The M13mp18 subclones used for the sequencing are indicated by the thin lines.


ED-B Gene Structure
ED-A Gene Structure

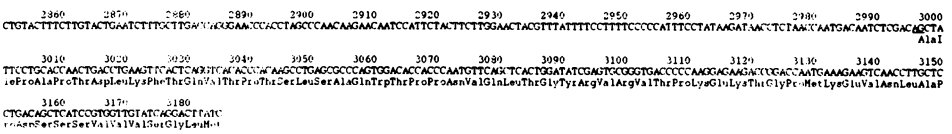


Figure 2. Nucleotide sequence of the ED-B and ED-A regions of the human fibronectin gene. The GT and AG nucleotides at the junctions are underlined. In the ED-A sequence the two converging arrows indicate a 16 nucleotide palindromic sequence; two Alu-type repetitive sequences are also underlined.

a similar construct containing the human ED-A region cloned in the same site of the α1-globin gene. pB5'SV (15), an expression vector containing the human beta globin gene, was usually cotransfected with the other clones.

Double stranded plasmid DNA was introduced into Hela cells by calcium phosphate coprecipitation as previously described (16). The cells were harvested after two days and the RNA was extracted from the transfected cells. Northern blots were then hybridised to probes specific for α-globin RNA, fibronectin RNA and ED-B exon, radioactively labeled by nick-translation (17), or by primer extension on single strand M13 recombinant DNA followed by digestion to recover the insert. Clones M13 FN and M13 ED are M13 subclones produced during the sequencing of the ED-B region and cover from the left end to the Stu I site and between positions 1399-1683 (see fig. 1 and 2).

Analysis of the splicing patterns by oligonucleotide probes

The accuracy of the splicing was verified by hybridizing the RNA produced in transfected cells to a set of oligonucleotides complementary to the cDNA expected on the basis of the sequence of the ends of the ED-B and the flanking exons in the genomic clone. The oligonucleotides were designed as 18mers covering nine nucleotides on each side of the splice junction; oligonucleotides n. 1 and 3 respectively reveal the removal of the two introns upstream and downstream from the ED-B exon; oligo n. 2 corresponds to the removal of both introns together and so only hybridize to RNA lacking the ED-B exon. The oligonucleotides were labeled with gamma-ATP in presence of T4-polinucleotide kinase and hybridized to northern blots. Washing at room temperature gave no hybridization either to unspliced or to otherwise spliced RNAs. More stringent wash at 50° C (5° below Tm) removed every additional background.

RESULTS

Sequence of ED-B region

The map of recombinant clone MA10, containing the ED-B region of

fibronectin gene, is shown in fig. 1a. It contains the coding sequence corresponding to nucleotides 3610-3780 of fibronectin mRNA and the ED-B exon, together with the corresponding introns.

The DNA sequence of the area was determined and is shown in fig. 2, in which also the protein sequence derived from the translation of the exon is reported. The central exon is alternatively spliced(7) and codes for a 91 amino acid peptide. The insertion of this additional exon in the fibronectin mRNA does not cause frameshifts after the junction and only results in a 91 amino acid longer protein, with no other variations downstream. The amino acid sequence codes for a typical type III domain. The four intron-exon junctions involved in the alternative splicing of the ED-B exon are all contained in the sequenced region. Their sequence is essentially within the limits of the usual consensus sequences, all of them obeying the GT-AG rule. A few unusual features are present in the 5' splice junction in front of the ED, where the typical CAG is substituted by a less common TAG and an unusually long polypyrimidine tract was found, in which no adenine is present between position -43 and -12 of the junction.As a consequence no consensus for branch sites was found in this area.

Sequence of the ED-A region

The sequence of the ED-A region was completely determined and is shown in fig. 2. The genomic organization (16) is essentially the same as for ED-B, with the extra type III homology coded by a single exon. Two introns of similar length separate the ED-A exon from the neighbouring ones; each of them includes an Alu-type repeat, underlined in the figure. The intron-exon junctions are in agreement with the consensus sequences for eucariotic junctions. A palindromic sequence is present in the right intron, between position 1634 and 1649 (underlined in fig. 2). Two regions of internal homology are present in the ED-A exon. The first is shown in fig. 3a, and consists of two direct repeats between nucleotides 1364 and 1415 in the ED exon and of nucleotides 2982 and 3032, that is across the 3' junction of the downstream intron. This same junction is also strongly homologous to the corresponding ED-A junction, as shown in fig. 3b. Here 14 nucleotides out of 15 are identical and the neighbouring area is also conserved. The two areas of homology can be combined through the formation of stem-loop structures. One of the possible conformations is shown in fig 3c. The formation of this or analog stems could connect the two regions of homology, thus creating an RNA sequence very similar to that of the downstream junction.

Expression of ED-B region in Hela cells

We cloned the genomic fragment containing the ED-B region in the expression vector pSV-a1W (13) (fig. 4), thus constructing a hybrid gene in which the

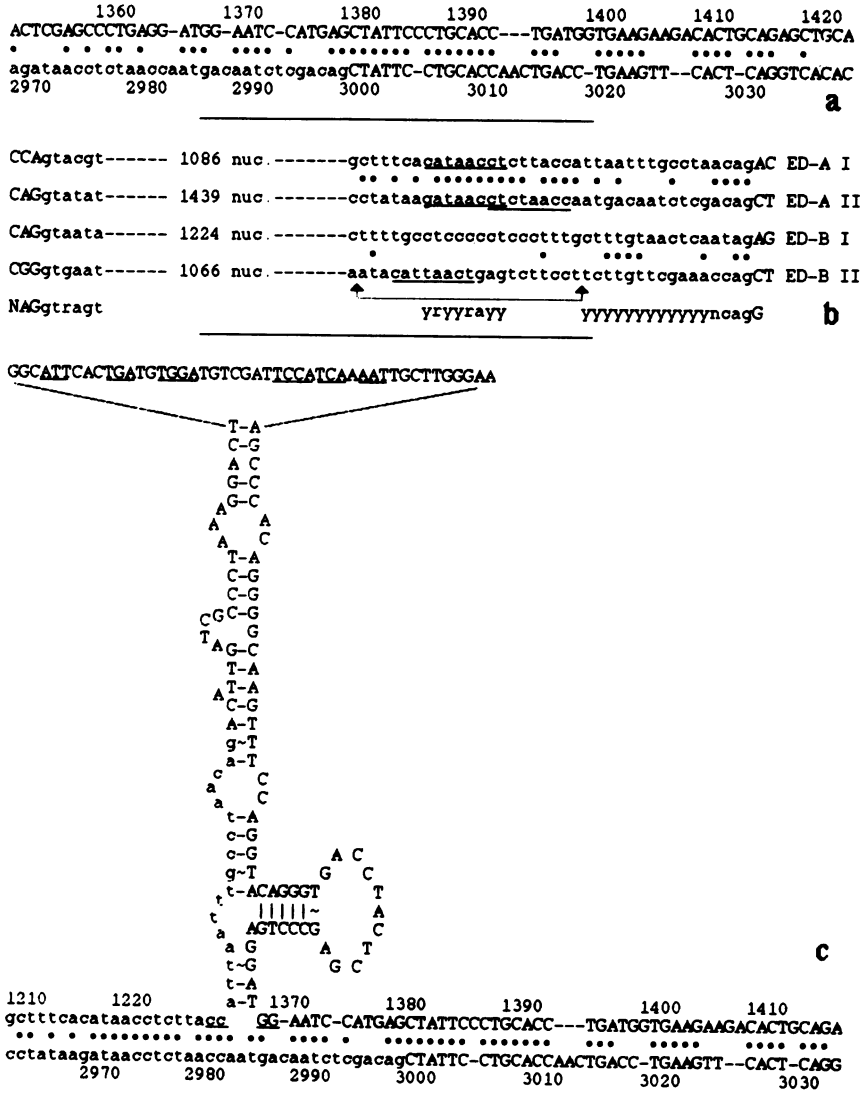


Figure 3. Analysis of the sequences of the ED-B and the ED-A regions of the human fibronectin gene. Upper and lower case characters respectively represent exon and intron sequences. (a) Two areas of the ED-A region, showing a high degree of homology, are compared. The dashes represent insertions introduced to maximize the homology. (b) The splice junction of the four introns are aligned and compared between them and to the consensus splice sequences for eucariotic genes. The area limited by the two arrows is where all the mapped branch sites are located. The possible branch sites are underlined. (c) A stem-loop secondary structure that could connect two areas of sequence homology. Both perfect base pairs(-) and GU pairs (~) are indicated. The dots represent the homology between the two areas after the formation of the stem. Additional complementary bases, not represented as secondary structure, are underlined.

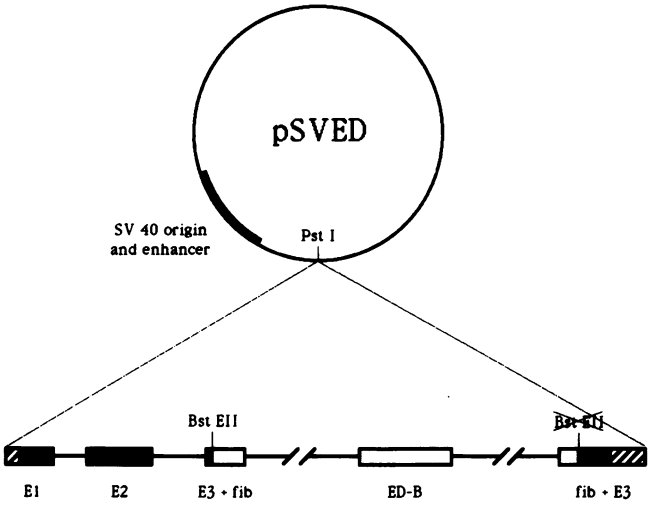


Figure 4. Schematic representation of the Bra7 clone, containing the fibronectin ED-B region (open boxes) inserted into the human α1-globin gene (closed boxes). Hatched boxes indicate non coding sequences of the α-globin gene.

fibronectin gene fragment is cloned in the third exon of the human α1-globin gene (Bra 7). This clone was introduced in Hela cells, by calcium phosphate coprecipitation. Transient expression of the clone gave rise to two different RNAs, which have been characterized by hybridization to M13 derived probes, specific for the fibronectin exon upstream from the ED-B and for the ED-B itself (M13 ED), as shown in fig. 5a and b. Both bands hybridize to the M13 FN probe, but only the upper one hybridizes to the ED-specific probe. Hybridization to the α-globin cDNA probe gives the same pattern as with the M13 FN probe. The accuracy of the splicing patterns was checked by hybridizing the northern to oligonucleotides 2 and 3, complementary to the RNA sequence produced by the splicing of the intron downstream from the ED-B, or the two introns flanking the ED-B. As shown in fig. 5, oligonucleotide 2 only hybridizes to the lower band, whereas the oligo n. 3 only hybridizes to the upper one. Another oligonucleotide, specific for the removal of the intron upstream from the ED-B, gives the same pattern as the oligonucleotide 3. S1 analysis was also partially carried out and is in agreement with the splicing patterns indicated (data not shown). The splicing patterns of the RNAs produced in the transfection experiment are so equivalent to the corresponding fibronectin mRNAs. Moreover, the relative amounts of the two RNAs are very similar to that found in fibroblasts. In fact the two forms are not equally represented, but the ED- is by far the predominant one, the ED+ form representing about 5% of the RNA produced.

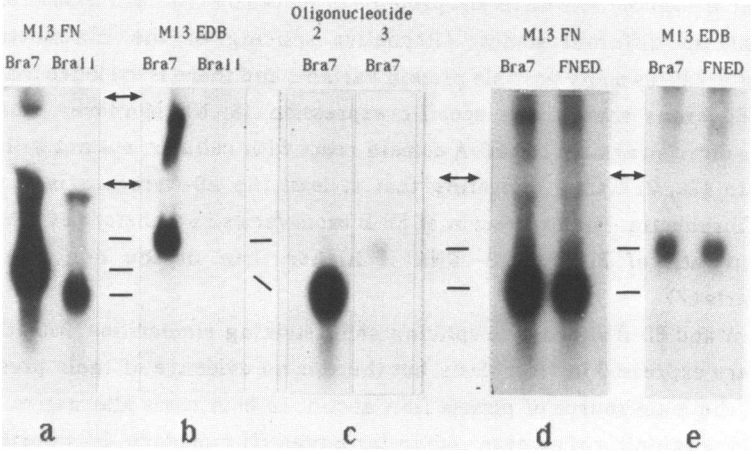


Figure 5. Northern blots of total RNA prepared from transfected Hela cells were hybridized to M13 probes covering the ED-B region. The arrows indicate the position of 28S and 18S RNAs. Clone M13 FN maps from the left end to the Stu I site and clone M13 ED between positions 1399-1683 (see fig. 1 and 2). Alternatively oligonucleotides specific for the ED- RNA (2) and ED+ RNA (3) were used. Hela cells were transfected with clone Bra 7 (ED-B/α-globin hybrid), clone Bra 11 (a control clone in which only the mp18 polylinker was inserted in the same site of α-globin gene; this hybridizes to the M13FN probe which also contains the polylinker sequence). Lanes marked FNED in panels d and e derive from the cotransfection of both Bra7 (ED-B) and FNED(ED-A) clones.

In order to explore whether there are any common factors between the ED-A and the ED-B systems, that can become limiting in excess of RNA substrate we transfected at the same time the ED-B clone and a similar one (FNED) containing the ED-A region (14). The RNA from the transfected cells was hybridized to probes specific of the ED-B region, as shown in fig. 5d-e. The presence of the ED-A precursor RNA induces no variations in the expression patterns of ED-B, thus showing that no interference between the two processes is appreciable in our system. ED-A expression patterns are also identical to the previously obtained ones with the ED-A containing clone (data not shown).

DISCUSSION

Alternative splicing of the precursor RNA, originally thought to be limited to a few genes, was subsequently demonstrated in a large number of genes and is now considered a widespread system of producing polipeptide variants from a single gene. In some cases alternative splicing happens in the non coding region of mRNAs, and the possible biological significance of the process is not clearly understood. In other cases, like calcitonin (18), immunoglobulin (19) and others, it results in the

production of functionally different proteins produced by different tissues, or by the same tissue in different stages. Alternative splicing of the fibronectin gene transcript produces many possible protein variants and there is evidence that at least the two ED exons show tissue specific expression (8, 6). Moreover monoclonal antibodies directed against the ED-A domain react with cellular, but not with plasma fibronectin (20, 21), thus indicating that at least the ED-A region is specific to cellular fibronectin. The expression of ED-B exon varies in transformed fibroblasts, where the ratio of ED-B+/ED-B- RNA is higher than in the non transformed counterparts (7).

ED-A and ED-B alternative splicing show striking similarities. Both the extra domains are expressed in fibroblasts, but there is no evidence of their presence in the liver, the main source of plasma fibronectin. In both cases alternative splicing results in the skipping of an exon coding for a type III homology. It is worth noting that the skipped type III homologies, unlike most of the others which are coded by two exons, are coded by a single exon. A third type III homology is encoded as a single exon in rat, but no alternative splicing has been observed in that area (6).

A construct containing the ED-A area of fibronectin gene fused to the α-globin gene, when transfected in Hela cells undergoes alternative splicing and generates two processed RNAs in equal amounts (16). A similar experiment was carried out with an analogous construct made with the corresponding ED-B area. The transfection of this hybrid gene (fig. 5a) also gives rise to two mRNAs, but the relative amounts are very different, with a ratio of about 1:20 between ED-B^+ and ED-B^- forms. The ratio between the two RNAs produced in this experiment resembles the situation found in vivo in fibroblasts, where the endogenous fibronectin gene produces a very similar splicing pattern. The accuracy of the splicing was checked by hybridization to oligonucleotides covering the cDNA junctions, and found to be correct in both RNAs (fig. 5c). Since the ED-A alternative splicing in Hela was also similar to the fibroblast pattern, transfected Hela cells should be considered a good model system for studying fibronectin alternative splicing. In both the ED-A and ED-B cases all the information to produce accurate alternative splicing is contained in a gene fragment of about 3 Kb, out of more than 70 Kb; this excludes the importance of long range interactions in the pre-mRNA in influencing alternative splicing. Moreover, there is no connection with fibronectin gene transcription initiation and/or termination processes. How this information is coded in the sequence is yet unknown. It is suggestive that both ED-A and ED-B, unlike the other type III homologies, are encoded by a single exon, but this may not be related to the alternative splicing mechanism since a third type III homology in rat, coded by a single exon does not seem to be alternatively spliced (6).

Sequence analysis of the two regions showed some unusual features in each of them. The 3' splice junction upstream from the ED-B has an unusually long polypyrimidine strech, between positions -67 and -11 of the junction, in which only one adenosine is found in position -44. As a consequence in the area between -15 and -37, in which all the mapped branch sites are found in other genes (23), there is no adenosine, and it is difficult to find a typical branch site, even allowing for mismacthes with the already highly degenerated consensus. However this junction is used in vivo, at least in some cases, so a functional branch site must be present, either in a different position, or in the usual position, but using a different nucleotide for the formation of the lariat structure. It is known that atypical branch sites can work (24) and that also other nucleotides can substitute for the adenosine in artificial genes (22), but in general this leads to a lower splicing activity.

The most interesting structural feature in the ED-A area is a direct repeat involving the ED-A exon and the downstream 3' splice junction, reported in fig. 4. A relatively large region (80 bp), may be aligned only by allowing for a few insertions. An even larger region of homology can be produced by the formation of the stem-loop secondary structure showed in fig. 3c. This structure could be related to the evolution of the gene or to other functions, but it is noteworthy that recently Mardon et al. (14) reported that a deletion in this same area (nuc. 1345-1430), induced an alteration of the splicing pattern, resulting in the ED-A exon being always treated as an intron, instead of being present in about half of the processed RNAs, as expected. Moreover the reinsertion of this fragment in the opposite orientation did not restore the activity (14), thus showing that the effect is sequence specific. Both the deletion and the inversion of this fragment result in the loss of the structure shown in fig. 3c.

The sequences of the ED-A and ED-B regions were compared to establish whether any analogy existed between them. No extensive sequence homology or other structural similarity was found, apart from the general intron-exon organization. Also the features previously discussed are specific to each sequence and are not present in the correspondent areas of the other, so if equivalent signal sequences exist, they must be functionally, but not structurally equivalent.

Since the alternative splicing patterns are tissue specific, they may derive from the interaction between the precursor RNA and cell specific factors. The similarities between ED-A and ED-B suggest that one or more cellular factors involved in the control of the splicing process may be in common to both. Clone Bra 7 was transfected in Hela cells, together with FNED (14), a similar clone containing the ED-A region inserted in the α-globin gene, to test whether the presence of the ED-A construct could affect the alternative splicing of the ED-B region. The transfected hybrid gene was transcribed and spliced exactly in the same way both in presence

and in absence of the ED-A construct, thus indicating that no important limiting factor necessary to the alternative splicing is shared by the two systems (see fig. 5d-e). Also the ratio between the two ED-A forms was not affected by the cotransfection (data not shown).

The results of both the sequence data analysis and the transfection experiments thus show that, although the ED-A and the ED-B regions are very similar in gene organization, no further analogy is apparent between the two systems. In particular, although the tissue specific distribution of the splicing patterns is similar, it is unlikely that the alternative splicing depends on common factors, unless these are present in such high concentration that their availability is not limited, even in presence of the high amounts of specific RNA produced in transient expression experiments.

ACKNOWLEDGEMENTS

We thank K. Gould for providing the oligonucleotides. This study has been partially supported by research grants from Cancer Research Campaign and Delta Biotechnology; G.Paolella and G. Sebastio are on leave of absence, respectively, from Istituto di Scienze Biochimiche and Clinica Pediatrica, II Facoltà di Medicina, Universita di Napoli, Italy and are both recipient of long-term EMBO fellowships. C. Henchcliffe was supported by an MRC studentship.

REFERENCES

1. Hynes, R.O. (1985) Annu. Rev. Cell Biol., **1**, 67-90.
2. Owens, R.J., Kornblihtt, A.K. and Baralle, F.E. (1986) In McLean, N. (ed.), *Oxford Surveys on Eukariotic Genes* Oxford University Press, **3**, 141-160.
3. Kornblihtt, A.K., Vibe-Pedersen, K. and Baralle, F.E. (1984) Nucl. Acids Res., **12**, 5853-5868.
4. Odermatt, E., Tamkun, J.W., and Hynes, R.O. (1985) Proc. Natl. Acad. Sci. USA **82**, 6571-6575
5. Schwarzbauer, J.E., Tamkun, J.W., Lemischka, I.R. and Hynes, R.O. (1983) Cell **35**, 421-431.
6. Schwarzbauer, J.E., Patel, R.S., Fonda, D. and Hynes, R.O. (1987) EMBO J. **6**, 2573-2580
7. Zardi L., Carnemolla B., Siri A., Petersen T.E., Paolella G., Sebastio G. and Baralle F.E. (1987) EMBO J. **6**, 2337-2342
8. Kornblihtt, A.K., Vibe-Pedersen, K. and Baralle, F.E. (1984) EMBO J. **3**, 221-226
9. Maniatis, T., Fritsch, E.F. and Sambrook (1982) *DNA Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.
10. Henikoff, S. (1984) Gene **28**, 351-359.
11. Sanger, F., Nicklen, S. and Coulson, A.R. (1980) Proc. Natl Acad. Sci. USA **74**, 5463-5467.
12. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol., **65**, 499-580.
13. Higgs, D.R., Goodburn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983) Nature **306**, 398-400.
14. Mardon, H.J., Sebastio, G. and Baralle, F.E. (1987) Nucl. Acids Res. **15**, 7725-7733.
15. Grosveld, G.C., de Boer, E., Shewmaker, C.K. and Flavell, R.A. (1983) Nature **295**, 120-126.
16. Vibe-Pedersen, K., Kornblihtt, A.K. and Baralle, F.E. (1984) EMBO J. **3**, 2511-2516.
17. Rigby, P.W.J., Dieckmann, H., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. **113**, 237-251.

18. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982). Nature **298**, 240-244.
19. Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimore, D. (1980) Cell **20**, 293-301.
20. Paul, J.I., Schwarzbauer, J.E., Tamkun, J.W. and Hynes, R.O. (1986) J.Biol. Chem. **261**, 12258-12265
21. Carnemolla, B., Borsi, L., Zardi, L., Owens, R.J. and Baralle, F.E. (1987) FEBS **215**, 269-273
22. Hornig, H., Aebi, M. and Weissmann, C. (1986) Nature **324**, 589-591
23. Reed, R. and Maniatis, T. (1985) Cell **41**, 95-105.
24. Ruskin, B., Greene, J.M. and Green, M.R. (1985) Cell 41, 833-844.