Donor and acceptor splice signals within an exon of the human fibronectin gene: a new type of differential splicing

Karen Vibe-Pedersen°, Staffan Magnusson and Francisco E. Baralle*

Department of Molecular Biology and Plant Physiology, C.F. Møllers Alle 130, University of Aarhus, DK 8000 Aarhus C, Denmark and **Sir William Dunn School of Pathology, South Parks Road, Oxford University, Oxford OX1 3RE, England*

Received 29 August 1986

We have sequenced that area of a human fibronectin gene clone which codes for a connecting strand separating the last two areas of the type III homology. The gene has a complex exon with two 'AG' acceptor sites. One of these can be used (exon subdivision). In addition 93 basepairs inside the exon are sometimes spliced out as an intron. This is the third differential splicing found in the fibronectin gene transcript and it represents a new type of differential splicing.

Fibronectin gene clone *Differential splicing* *Nonhomologous connective strand*

1. INTRODUCTION

Fibronectin (FN) is a large glycoprotein (M_r 280000 per chain) which is a component of the extracellular matrix of e.g. basement membranes and it is also present in blood plasma, as a disulphide linked dimer. It is important for cell anchorage, cell motility [1,2] and for normal wound healing [3]. FN contains discrete domains with binding sites for fibrin, collagen, DNA, cell surfaces and heparin. The complete sequence of overlapping fibronectin cDNA clones from a human cell line has recently been published [4]. The amino acid sequence of bovine plasma fibronectin has been determined by peptide sequencing [5]. The sequence contains three types of internal homology [6]. Either 15 or 16 type III homologies constitute the central part of the polypeptide chain [4].

Previously two areas of differential splicing in the FN gene transcript have been described. A fraction of the FN mRNA molecules from cell cultures contains an extra internal fragment coding for one type III homology – the extra domain [4]. This fragment is absent from FN mRNA of liver which synthesizes plasma FN. Bovine plasma FN also does not contain the ED fragment [5]. The ED corresponds to one exon in the FN gene [7] and FN mRNA with and without the ED arise by tissue specific differential splicing of the exon skipping type [7,8].

Three types of FN mRNA have been isolated from a rat liver cDNA library. In a nonhomologous connecting strand, spacing the last two domains of type III homology the three clones code for 120, 95 and 0 amino acids, respectively [9]. This can explain the heterogeneity found in plasma FN, which appears as two bands on reduced SDS gels. Only the top band reacts with a monoclonal antibody against an epitope in the nonhomologous connecting area called the type III connecting strand [10]. The IIICS region of the rat FN gene is encoded by the 5'-end of a larger exon which also encodes the first third of the following type III unit. This exon has three 'AG' splice acceptor sites either of which can be used when splicing to the preceding 'GT' donor sequence [11]

Abbreviations: FN, fibronectin; ED, extra domain; IIICS, type III connecting strand; kbp, kilobasepair

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

(exon subdivision) giving sequences corresponding to the three rat cDNA clones described [9].

We have found cDNA clones for human FN in cDNA libraries of RNA from a human cell line [4,8] and from human liver [12]. In our cDNA clones the IIICS coding region lacks 93 nucleotides present in the 3'-end of the IIICS coding regions of all three rat clones. In order to elucidate this major difference between rat and human mRNA coding for otherwise very homologous proteins and find the splicing variation in this area of the human FN gene we have isolated the IIICS coding area of the human FN gene from a larger FN gene clone [7]. Recently a human FN cDNA clone was isolated with a IIICS region like the one found in rat FN cDNA [13]. In this paper we describe the gene structure of the IIICS coding region of the human FN gene and the splicing variations of the gene transcript.

2. MATERIALS AND METHODS

Enzymes were from New England Biolabs, Beverly, MA, USA unless otherwise mentioned.

A total *Bgl*II digest of the FN gene clone λFN5 [7] was ligated into the *Bam*HI site of the pAT 153/*Pvu*II/8 vector [14] and cloned in *E. coli* MC1061. The recombinant clones were transferred to filter paper (Whatman 541) [15] and hybridized to the inserts of M13 subclones from pFH1 [16] labelled with [α-^{32}P]ATP by primer extension using the Klenow fragment of DNA polymerase I [17]. Probe I was from *Alu*I 3214 to *Alu*I 3360, probe II from *Alu*I 3090 to *Alu*I 3214 (numbering as in [16]). Probe II spans sites for *Bgl*II and *Bst*EII. Washing conditions in the hybridization were 3 × SSC, 65°C. DNA from clones hybridizing to both probes (pFH5B2.3) and hybridizing only to probe II (pFH5B2.2) were analysed by restriction mapping and sequencing of restriction fragments (3'-end labelled by 'filling in' with Klenow polymerase I) by the Maxam and Gilbert method [18]. A subclone of a 1.0 kbp *Pvu*II fragment of pFH5B2.3 in the same vector was analysed in the same way.

From this subclone insert fragments sequentially shortened from one by one digestion with exonuclease *Bal*31 were obtained in the following way (slightly modified from [19]): in the vector the *Pvu*II site is flanked by *Eco*RI and *Bam*HI sites. 20 μg of the recombinant clone was linearized with *Eco*RI, extracted with phenol, precipitated with EtOH, resuspended in 300 μl of 12 mM $CaCl_2$, 1 mM $MgCl_2$, 600 mM NaCl, 20 mM Tris-HCl, pH 8.1, 1 mM EDTA and shortened from both ends with 5 μl *Bal*31 (750 U/ml) at 30°C. Aliquots were removed at different times and immediately extracted with phenol. All fractions were precipitated with EtOH and finally digested with *Pvu*II to remove the vector from the inserts and with *Bam*HI to make one end of the vector fragment 'sticky'. The experiment was repeated using *Eco*RI and *Bam*HI in reverse order to get insert fragments shortened from the other end. The blunt-ended shortened fragments were cloned into M13 mp8 or mp9 (linearized with *Sma*I) by standard procedures [16]. The single-stranded DNA from the resulting clones and from M13 subclones of restriction fragments was sequenced by a modified dideoxy method essentially as in [20]: 200 ng of single-stranded template was dried down together with 20 μCi of [α-^{32}P]ATP (3000 Ci/mmol), 1 μl of 50 μM dGTP, 50 μM dTTP and resuspended in 2 μl 15-mer universal primer (1 pmol/μl, Biolab) and 1 μl RT buffer (RT buffer: 50 mM NaCl, 34 mM Tris-HCl, pH 8.3, 6 mM $MgCl_2$). After 10 min incubation at 55°C 0.5 μl reverse transcriptase (Life Sciences Inc., St. Petersburg, FL) (diluted to 1.25 units/μl with RT buffer) was added and the mixture incubated for 5 min at 55°C. Another 2 μl of diluted enzyme was added and 1 μl of this mixture was added to four tubes labelled G, T, C and A, each containing 1.5 μl of 250 μM dGTP, 250 μM dCTP, 250 μM dATP, 250 μM dTTP, 1 × RT buffer and 12.5 μM of the appropriate dideoxynucleotide, except for the G-tube which should have 8 μM ddGTP. These tubes were incubated at 55°C for 10 min. The content was mixed with 6 μl formamide dye, boiled and electrophoresed on 6% acrylamide, 7 M urea gels in the normal way [18].

*Sau*96I and *Hin*fI fragments (end labelled in their 3'-end) from the insert of pFH5B2.3 were eluted from an acrylamide gel run without urea and aliquots containing only 10000 cpm were used as probes when hybridizing to the insert of pFH1 spotted on 1 cm Millipore filters (0.45 μm). After washing the filters were counted in a liquid scintillator.

3. RESULTS AND DISCUSSION

We have isolated two plasmid clones pFH5B2.3 and pFH5B2.2 containing the area coding for the connecting strand between the last two domains of type III homology (the IIICS) area of the FN gene from the previously described FN gene clone λFN5 [7]. The insert of pFH5B2.3 is a 2.3 kbp *Bgl*II fragment of λFN5. It contains the area downstream from a *Bgl*II site in the differentially spliced IIICS region. The insert of pFH5B2.2 is a 2.2 kbp *Bgl*II fragment of λFN5 which contains the sequence from the same *Bgl*II site but in the upstream direction (fig.1).

Sequence analysis from a *Bst*EII site 30 bp from the cloning site in pFH5B2.2 showed that the exon–intron boundary 5′ to the IIICS region is located at the same point as the one found in the rat gene [11]. The rest of the IIICS region is contained in a 1.0 kbp fragment obtained by digestion of pFH5B2.3 with *Pvu*II. This fragment has been completely sequenced. Fig.2 shows a sequence starting with the last 14 bp of the intron preceding the IIICS region and ending with 571 bp of the intron interrupting the sequence coding for the last domain of type III homology. This domain is located after the IIICS region. The bases 14–281 in fig.2 correspond to the IIICS region which we have previously found in the FN cDNA clone pFH1 [16]. Then follows a stretch of 93 bp which is homologous to the 3′-part of the rate IIICS region and present in the human FN cDNA clone FN421 [13], but which is absent from the human cDNA clones pFH1 from a human cell line and pFHL1 and pFHL8 from human liver [12]. The next exon coding for 36 amino acids of the last type III homology follows immediately after the 93 bp stretch.

The 93 bp sequence constitutes a short normal intron sequence with a 5′-GT donor and a 3′-AG acceptor sequence in accordance with the consensus splice sequences [25]. In the rat FN gene an AT sequence corresponds to this human GT donor sequence. This explains why this part is not spliced out in the rate gene transcript. The same must be true for the bovine FN gene, since the corresponding two bases are the last two of an asparagine codon [5].

The acceptor site in the rat sequence corresponding to the 3′-end of the human 93 bp stretch is used when that mRNA which codes for FN lacking the IIICS region is made [9]. We do not know at present whether the splice variant which skips the IIICS exon also occurs in the human FN.

The occurrence of cDNA clones with and without the 93 bp stretch indicates that it can be

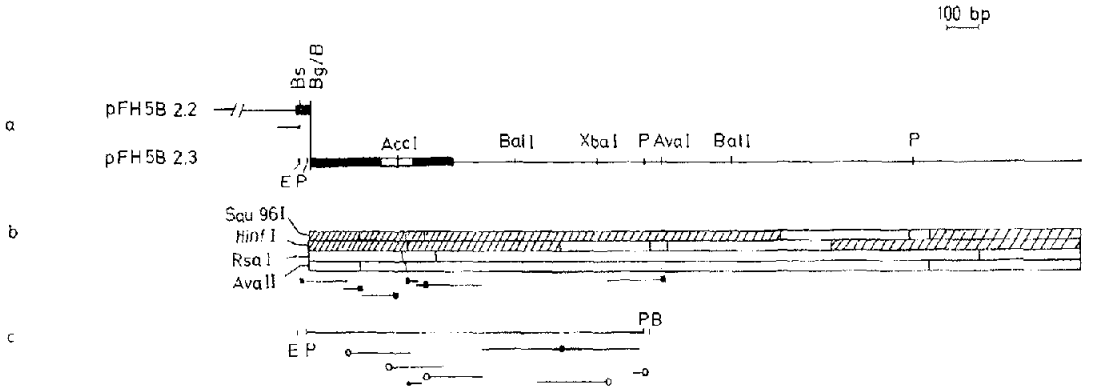

Fig.1. Restriction map of the pFH5B2.3 and part of the pFH5B2.2 clones, and the sequencing strategy. (a) Black boxes are exons, the open box illustrates a 93 bp intron which can also be part of an exon. (b) Shaded boxes illustrate *Sau*96I and *Hin*fI fragments containing exon sequences. (a and c) Restriction sites outside the horizontal line are from the vector. (c) The top line is a 1.0 kbp *Pvu*II restriction fragment subcloned for sequencing (see section 2). (a–c) The fragments which have been sequenced by the Maxam and Gilbert method are marked by lines with black squares. Sequences obtained by dideoxy sequencing starting at specific restriction sites are marked by lines with filled circles; those with open circles were obtained by sequential shortening with the exonuclease *Bal*31 (see section 2). Bs, *Bst*EII; Bg, *Bgl*II; B, *Bam*HI; E, *Eco*RI; P, *Pvu*II.

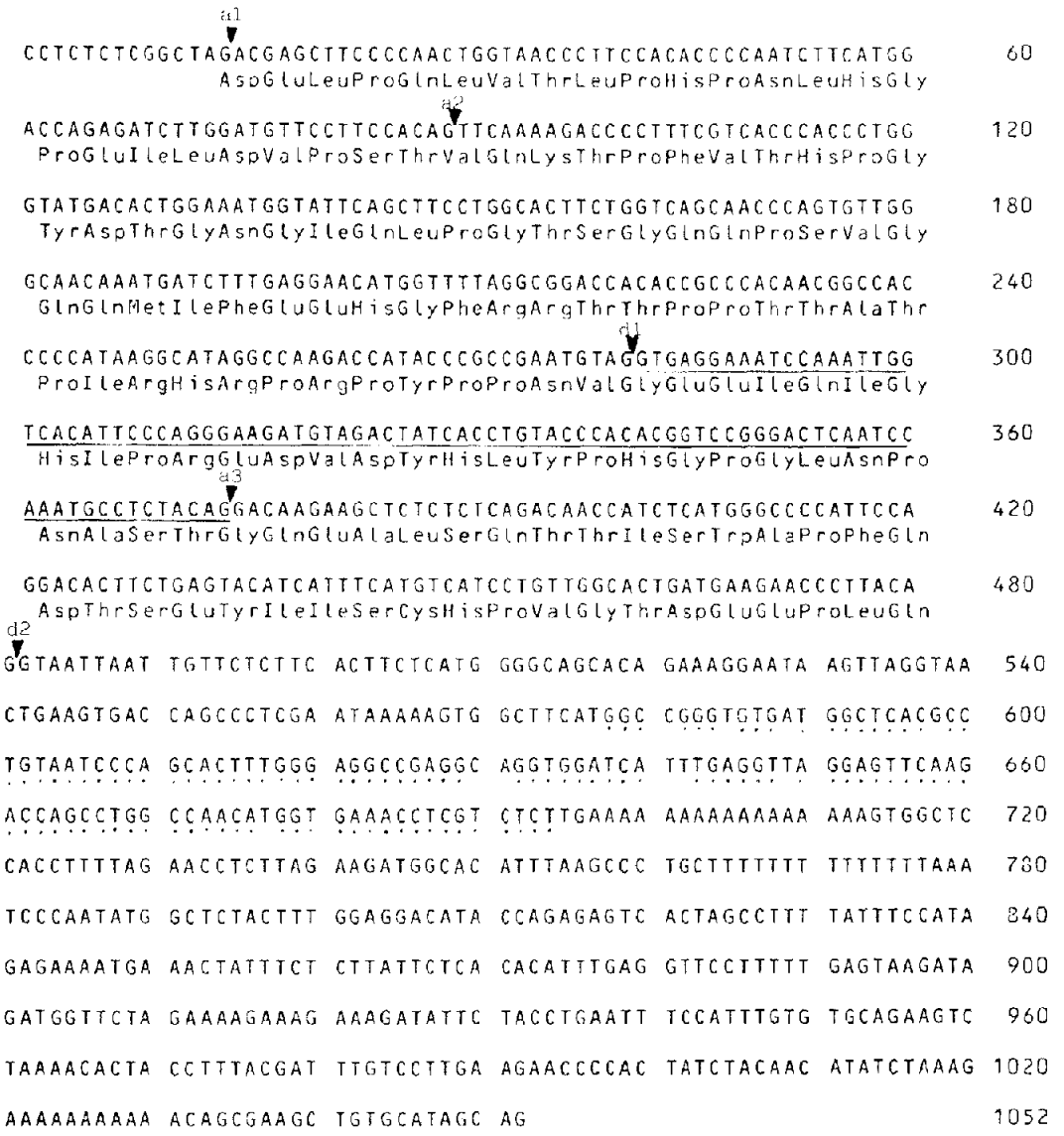

Fig.2. Sequence of the IIICS coding region plus part of the surrounding introns. Arrowheads are intron–exon junctions, a1, a2 and a3 are acceptor sites, d1 and d2 are donor sites. The 93 bp intron which can be part of an exon is underlined. All exon sequences are translated. Bases identical to those of the first monomer of the consensus sequences of the Alu family [24] are underdotted.

differentially spliced, being regarded either as part of an exon or as a short intron. The sequence of the 93 bp in the FN421 cDNA clone [13] and our gene clone are identical except for one silent G to A exchange which creates an extra cleavage site for the restriction enzyme *Hin*fI in our clone.

The sequence TGTACCCAC, the underlined A being 35 bp before the 3′-end of the 93 bp intron can form base pairs with the 5′-end of the intron in 6 of the 9 positions and it is therefore a putative recognition site for the so-called lariat formation [21]. The intron sequence as such does not give any obvious clue to why this intron is not always spliced out. The intron length itself could perhaps be the reason. Wieringa et al. [22] found a minimal intron length of 81 bp to be necessary for the cor-

rect splicing of their artificial mutants of the first intron of the β-globin gene. In addition the original first 6 and last 24 bases of the intron were found to be required.

In rat and bovine FN the amino acid sequence Arg-Gly-Asp-Val appears in the transcribed 93 bp stretch. This is a putative second cell binding site [23]. However, in the human gene sequence and in the human cDNA clone which includes this part [13] the sequence is Arg-Glu-Asp-Val which does not bind cells. The functional importance of this sequence is therefore dubious.

Fig.2 also shows the first 571 nucleotides of the intron following the exon coding for 36 amino acids of the last unit of type III homology. The length of this intron has been estimated to be approx. 2 kbp by hybridization of restriction fragments to pFH1 (see section 2). It is remarkably rich in adenines and thymidines and includes stretches of 15 As, 14 Ts and 11 As. It contains a 150 bp stretch belonging to the so-called Alu family [24] starting 92 bp from the 5′-end or the intron (fig.2). Preliminary results suggest a 300 bp stretch belonging to the Alu family on the non-coding strand further downstream in the intron (not shown).

In conclusion we have described a short region of the FN gene with a complex splicing variation including exon subdivision of the 5′-end of an exon and a short intron which is not always removed. This is the third place where differential splicing has been observed so far in the human FN gene transcript and it represents a new type of differential splicing. This splicing pattern differs between species as a consequence of a GT-AT difference. This system should be useful for analysing the factors regulating differential splicing.

ACKNOWLEDGEMENTS

We thank C.M. Mikkelsen for excellent technical assistance and A.R. Kornblihtt and T.E. Petersen for valuable discussions. This work was supported by the Danish Cancer Society, Hindsgauls Foundation and the Danish Natural Science Research Council.

REFERENCES

[1] Yamada, K.M. (1983) Annu. Rev. Biochem. 52, 761–766.
[2] Hynes, R.O. (1985) Annu. Rev. Cell Biol. 1, 67–90.
[3] Knox, P., Crooks, S. and Rimmer, C.S. (1986) J. Cell Biol. 102, 2318–2323.
[4] Kornblihtt, A.R., Umezawa, K., Vibe-Pedersen, K. and Baralle, F.E. (1985) EMBO J. 4, 1755–1759.
[5] Skorstengaard, K., Jensen, M.S., Sahl, P., Petersen, T.E. and Magnusson, S. (1986) Eur. J. Biochem., in press.
[6] Petersen, T.E., Thøgersen, H.C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. and Magnusson, S. (1983) Proc. Natl. Acad. Sci. USA 80, 137–141.
[7] Vibe-Pedersen, K., Kornblihtt, A.R. and Baralle, F.E. (1984) EMBO J. 3, 2511–2516.
[8] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1984) EMBO J. 3, 221–226.
[9] Schwarzbauer, J.E., Tamkun, J.W., Lemischka, I.R. and Hynes, R.O. (1983) Cell 35, 421–432.
[10] Schwarzbauer, J.E., Paul, J.I. and Hynes, R.O. (1985) Proc. Natl. Acad. Sci. USA 82, 1424–1428.
[11] Tamkun, J.W., Schwarzbauer, J.E. and Hynes, R.O. (1984) Proc. Natl. Acad. Sci. USA 81, 5140–5144.
[12] Umezawa, K., Kornblihtt, A.R. and Baralle, F.E. (1985) FEBS Lett. 186, 31–34.
[13] Bernard, H.P., Kolbe, M., Weil, D. and Chu, M.L. (1985) Biochemistry 24, 2698–2704.
[14] Anson, D.S., Choo, K.H., Rees, D.J.G., Gianelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1984) EMBO J. 3, 1053–1060.
[15] Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979) Nucleic Acids Res. 7, 2115–2136.
[16] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1984) Nucleic Acids Res. 12, 5853–5868.
[17] Shoulders, C.C. and Baralle, F.E. (1982) Nucleic Acids Res. 10, 4873–4882.
[18] Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–569.
[19] Ponz, M., Solowiejczyk, D., Ballantine, M., Schwarz, E. and Surrey, S. (1982) Proc. Natl. Acad. Sci. USA 79, 4298–4302.
[20] Duncan, C.H. (1985) in: NEN Product News 4, no.3.
[21] Keller, W. (1984) Cell 39, 423–425.
[22] Wieringa, B., Hofer, E. and Weissman, C. (1984) Cell 39, 423–425.
[23] Pierschbacher, M.D. and Ruoslahti, E. (1984) Proc. Natl. Acad. Sci. USA 81, 5985–5988.
[24] Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065–1070.
[25] Mount, S.M. (1982) Nucleic Acids Res. 10, 459–473.