The EMBO Journal vol.6 no.9 pp.2565–2572, 1987

Organization of the fibronectin gene provides evidence for exon shuffling during evolution

Ramila S.Patel, Erich Odermatt[1], Jean E.Schwarzbauer[2] and Richard O.Hynes

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[1]Present address: COBI, Walliserwerke Lonza AG, CA 3930 Visp, Switzerland
[2]Present address: Princeton University, Department of Biology, Princeton, NJ 08544, USA

Communicated by F.E.Baralle

We report the organization of the two ends of the rat fibronectin gene which encode the type I and II repeating units of the protein. We show that each of these modular structural units is encoded by a separate exon. Homologous type I and II repeats are known to occur in tissue plasminogen activator, factor XII and a bovine seminal plasma protein. Comparison of these sequences and the exon structures of the fibronectin and tissue plasminogen activator genes indicates that exons encoding type I and type II repeats have reassorted during evolution. We also report analyses of the extreme 5′ and 3′ ends of the fibronectin gene including the promoter region and the exon encoding the prepro sequence of fibronectin and we show that the gene is transcribed from a single initiation site to a single polyadenylation site. These data provide information pertinent to the transcriptional regulation of the gene, the alternative splicing of the primary transcript and the structure of the primary translation product.

Key words: fibronectin/gene structure/exon shuffling/promoter

Introduction

Fibronectins are high mol. wt glycoproteins involved in cell adhesion, morphology and migration (Hynes and Yamada, 1982; Yamada, 1983; Hynes, 1986). They are composed of a series of homologous repeating units of three types. Type I and II repeats are disulfide-bonded units 45–50 amino acids long, while type III repeats are 90 amino acids long and lack disulfides (Hynes, 1985; Kornblihtt *et al.*, 1985; Skorstengaard *et al.*, 1986). This repeating structure raises questions concerning the exon structure of the gene(s) encoding fibronectins. A single gene encodes all known forms of fibronectin (Kornblihtt *et al.*, 1983; Tamkun *et al.*, 1984). This gene is large and is made up of multiple small exons (Hirano *et al.*, 1983; Hynes, 1985). The question is whether these exons correspond with the repeating structural units of the protein.

We have previously reported that the type III repeats are indeed encoded by repeating patterns of exons (Odermatt *et al.*, 1985). Most are encoded by pairs of exons, but at three positions where alternative splicing of the primary transcript is known to occur, this pattern is altered (Vibe-Pedersen *et al.* 1984; Odermatt *et al.* 1985; Oldberg and Ruoslahti, 1986; Schwarzbauer *et al.*, 1987b). The exon structure of the fibronectin gene in the regions encoding type I and II repeats is of particular interest since homologous repeats occur in several other proteins (Banyai *et al.*, 1983; Esch *et al.*, 1983; Baker, 1985; McMullen and Fujikawa, 1985) suggesting the possibility that these repeating structural units might have reassorted among different genes during evolution by shuffling of exons in the manner proposed by Gilbert (1978). This possibility can only be addressed by detailed analysis of the gene structure.

We report here the sequences of all the type I and II repeats of rat fibronectin and show that they are each encoded by single exons, providing strong support for the exon shuffling hypothesis. In the course of this study we also determined the sequences of the 5′ and 3′ ends of the rat fibronectin gene and we show that the 71-kb gene is probably transcribed from a single initiation site to a single poly(A) addition site. Therefore, the alternative splicing which generates multiple different fibronectin subunits occurs within a unique transcript.

Results

Isolation and characterization of genomic clones

We have previously reported the isolation of two genomic clones containing the 3′ 25 kb of the rat fibronectin gene (Tamkun *et al.*, 1984). These clones were used to isolate a series of overlapping clones covering 80 kb (Figure 1a). The locations of exons were identified by restriction enzyme mapping, Southern blot hybridization using cDNA probes and DNA sequencing. Figure 1 shows the structures of two clones from the 5′ end of the gene (Figure 1b) and one from the 3′ end (Figure 1c). The structure

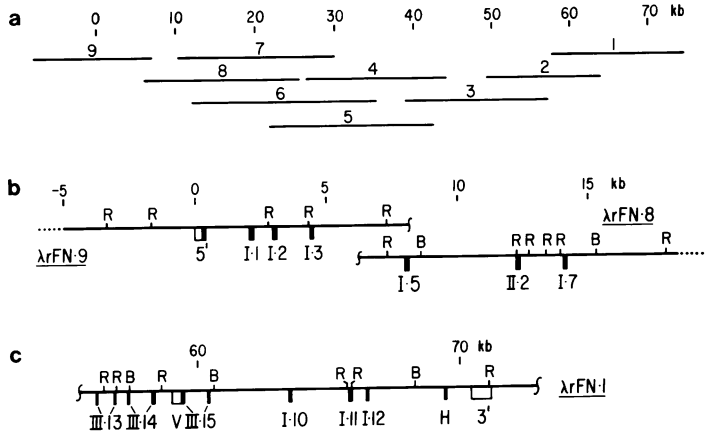


Fig. 1. Organization of genomic clones. (**a**) Nine overlapping clones which encompass the ~70-kb rat fibronectin gene. (**b** and **c**) Maps are shown for three clones covering ~30 kb at the 5′ end (λrFN-9 and λrFN-8) and ~17 kb at the 3′ end (λrFN-1). The scale is marked in kb from the transcription initiation site. All *Eco*RI (R) and *Bam*HI (B) sites are shown. Only those exons which have been sequenced are shown (see text). λrFN-9 and 8 contain additional exons encoding four more type I repeats and one more type II (not shown). The sequences of the exons at the 5′ end of λrFN-1 have been reported elsewhere (Tamkun *et al.*, 1984; Odermatt *et al.*, 1985).

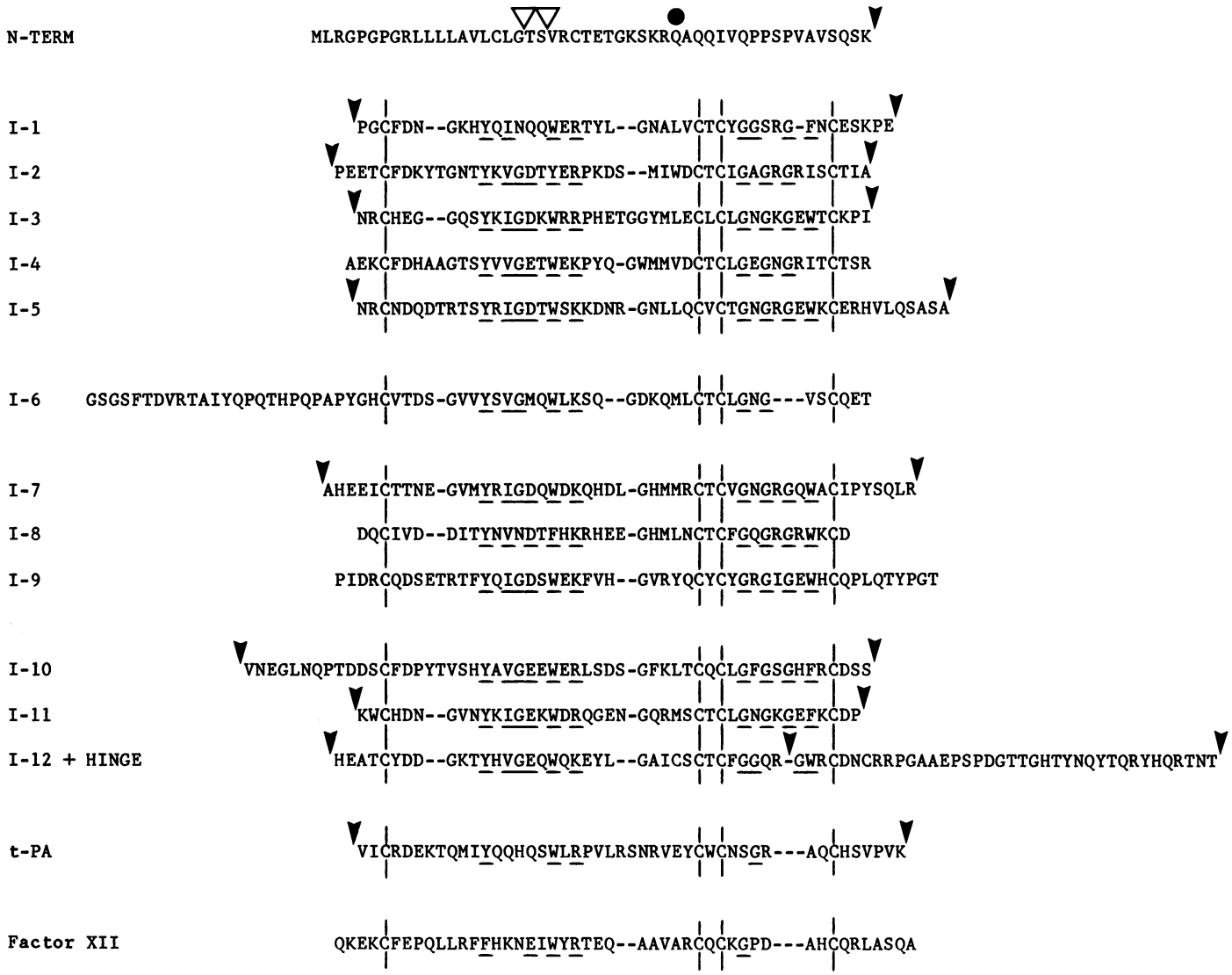


Fig. 2. Type I repeats in various proteins. Amino acid sequences of the 12 type I repeats of rat fibronectin. Single type I homologies from tissue plasminogen activator (t-PA; Ny *et al.*, 1984) and factor XII (McMullen and Fujikawa, 1985) are included for comparison. The sequences are given in the single letter code. Homologous cystine residues are marked by vertical lines and other highly conserved residues are underlined. The NH_2 terminus of the mature protein is marked by a solid circle and potential sites of signal sequence cleavage (see text) are marked by open arrowheads. Positions of introns are marked by closed arrowheads where known; they fall precisely between repeats except for the intron separating the I-12 and hinge exons (see text). Fibronectin type I repeats which were not sequenced directly from genomic clones were sequenced as cDNAs derived from genomic clones by passage through retroviral vectors (see Materials and methods). The two type II repeats (Figure 3) lie between repeats I-6 and I-7.

Fig. 3. Type II repeats in various proteins. The amino acid sequences of the two type II repeats of rat fibronectin are aligned with homologous repeats in factor XII (McMullen and Fujikawa, 1985) and a bovine seminal plasma protein (PDC-109; Esch *et al.*, 1983). The fibronectin and PDC repeats are each contiguous. Homologous cystines are marked by vertical lines and other conserved residues by underlining. The positions of the introns which separate the fibronectin type II repeats from each other and from the adjacent type I repeat (I-7) are marked by arrowheads. One fibronectin type II repeat (II-2) is encoded precisely by an exon. The same is probably true of the other (II-1) and is suggested to be true for those in the other proteins (see text). Fibronectin repeat II-1 was sequenced in a cDNA clone derived from genomic clones by passage through a retrovirus vector (see Materials and methods).

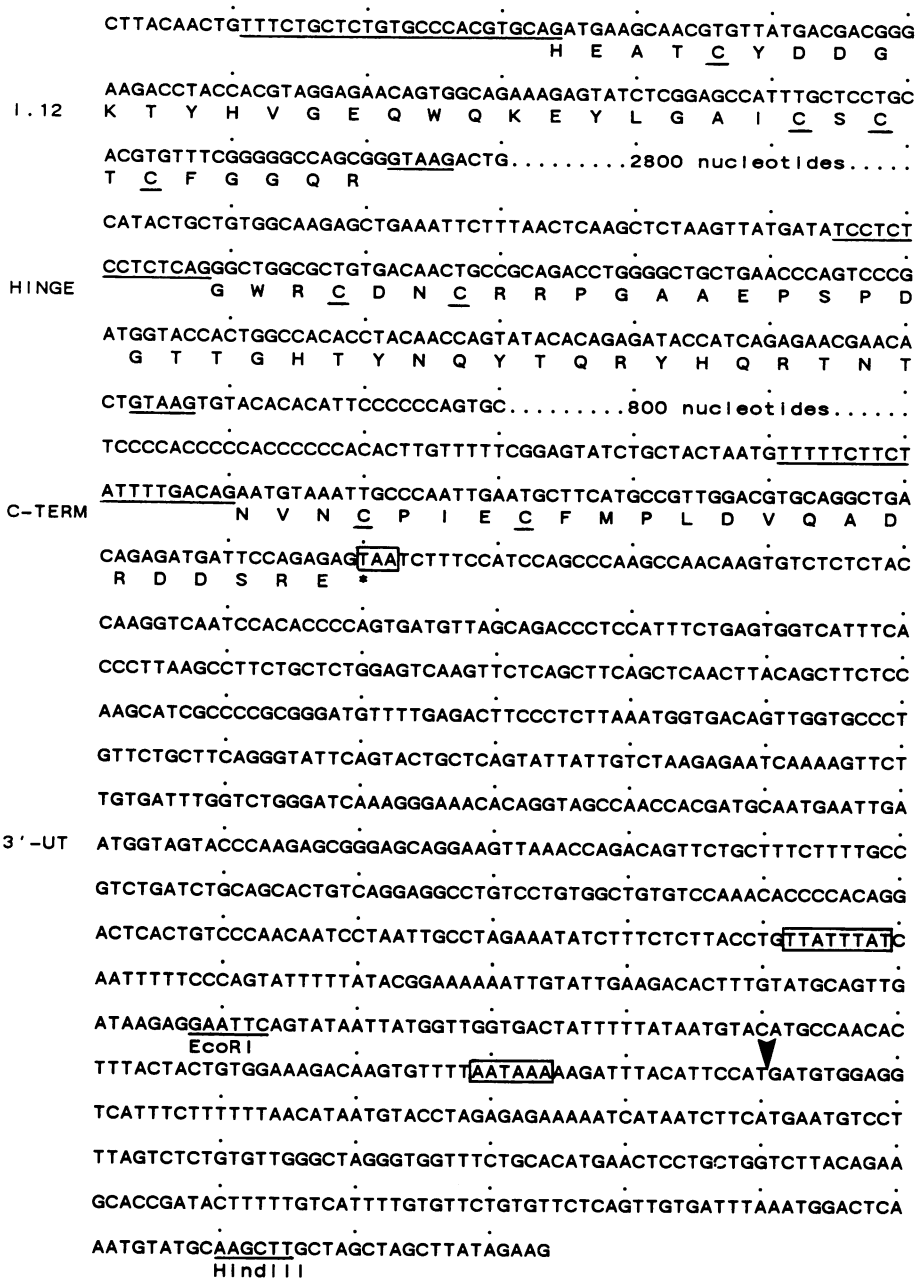


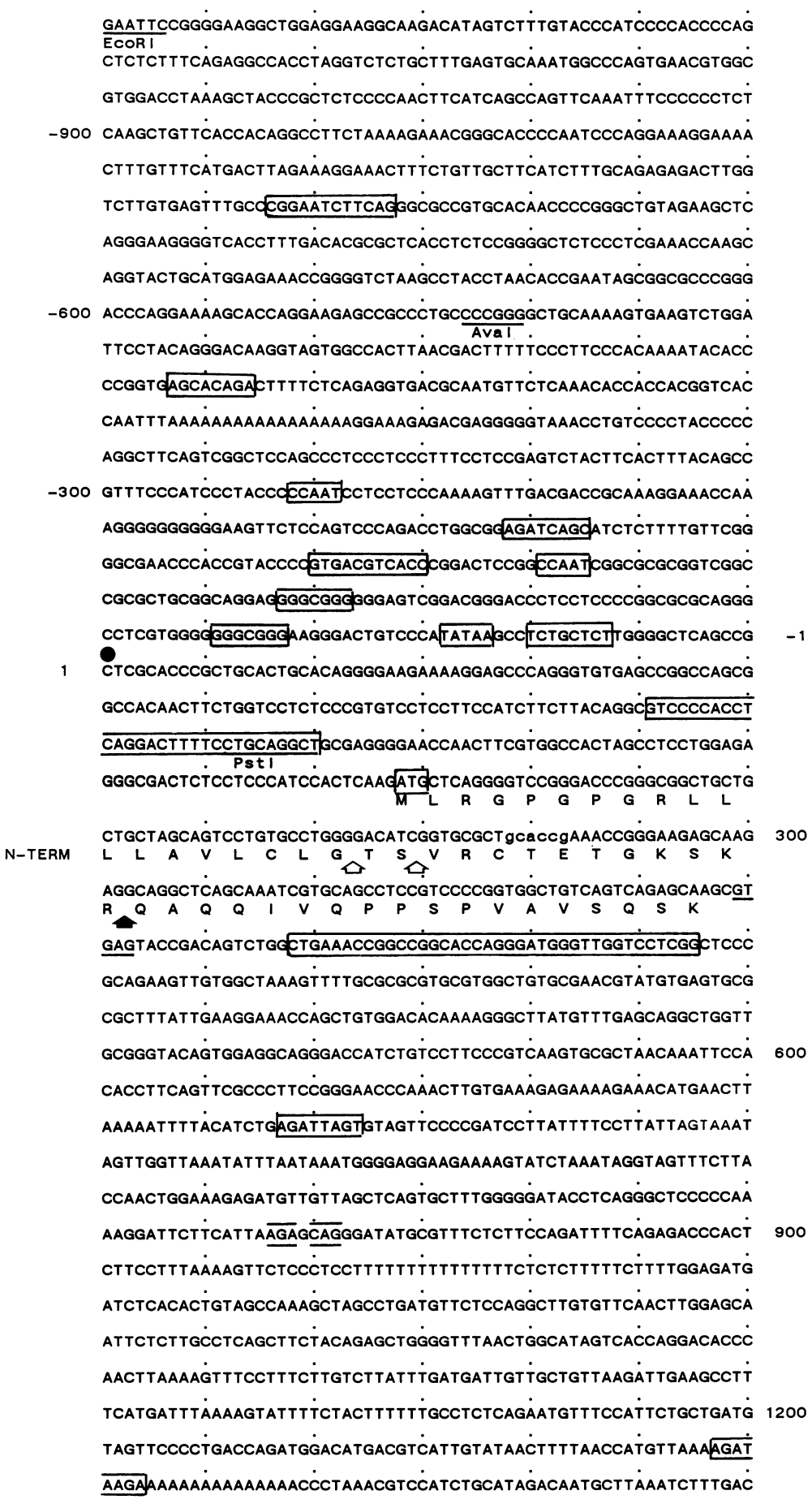
Fig. 4. Sequence of the 3′ end of the rat fibronectin gene. The sequence shown covers the exons encoding type I-12, the hinge segment, the C terminal disulfide-bonding segment and the 3′ untranslated region. Splice sites are underlined, the termination codon and poly(A) addition signal are boxed. The arrowhead marks the site of addition of poly(A) as determined by S1 nuclease (not shown) and RNase protection (Figure 6) experiments using probes extending between the *Hin*dIII and the *Eco*RI sites marked. The octanucleotide TTATTAT is common to the 3′ untranslated regions of several genes for proteins involved in inflammation (Caput *et al.*, 1986). The amino acids are given in the single letter code and cystine residues are underlined.

of the 5′ 6 kb of λrFN-1 has been reported previously (Tamkun *et al.*, 1984; Odermatt *et al.*, 1985) and will not be discussed further here. The rest of this clone contains five exons which encode three type I repeats, the C terminus of the protein and the 3′ untranslated region (see below).

The two 5′ end clones, λrFN-8 and λrFN-9, cover >30 kb. The 3′ end of λrFN-8 contains exons encoding the first type III repeats (Schwarzbauer *et al.* 1987b). λrFN-9 contains the 5′ end of the gene (see below). Therefore, the clones shown in Figure 1 contain the exons encoding all 12 type I repeats (nine at the 5′ end and three at the 3′ end) and both type II repeats. The structure of the central 40 kb of the gene is presented in the following paper (Schwarzbauer *et al.*, 1987b).

Structure of exons encoding type I and II repeats

Suitable fragments of these three genomic clones were subcloned and sequenced. We have sequenced across all the exons shown in Figure 1. These exons encode 8 of the 12 type I repeats and 1 of the 2 type II repeats as well as other parts of the fibronectin mRNA which will be discussed below. As shown in Table I and Figures 2 and 3, each of these repeats is encoded by a separate exon. In most cases, the correspondence is very precise, with the introns falling exactly between repeats. Clearly, therefore, the gene arose by endoduplication of primordial exons encoding each of the structural units of fibronectin. Furthermore, comparison of the amino acid sequences of the type I and II repeats


EcoRI
AvaI
PstI
-900
-600
-300
1
-1
300
600
900
1200
N-TERM
M L R G P G P G R L L
L L A V L C L G T S V R C T E T G K S K
R Q A Q Q I V Q P P S P V A V S Q S K

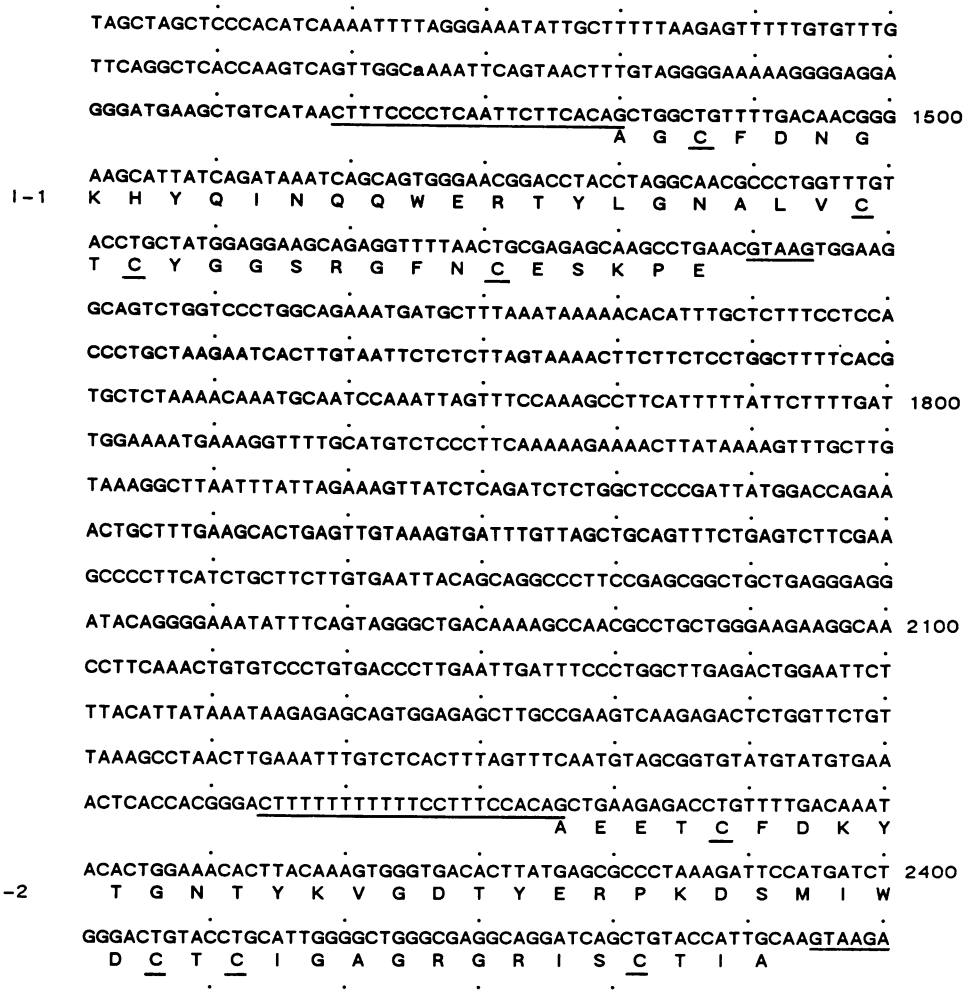


Fig. 5. Sequence of the 5′ end of the rat fibronectin gene. The sequence shown includes the typical type I exons encoding repeats I-1 and I-2 as well as the 5′ exon encoding the signal, pro- and N-terminal sequences. Splice sites are underlined, the initiation codon, TATAA and CCAAT boxes and other potential transcriptional control elements are boxed (see text). The probable transcription initiation site (solid circle, position 1) was determined by RNase protection experiments (Figure 6) using a probe extending between the *Ava*I and *Pst*I sites marked. The amino acid sequence is given in single letter code. The cystine residues in the type I repeats are underlined and potential signal (⇧) and pro sequence (⬆) processing cleavages in the N-terminal segment are marked by arrowheads (see text). The six nucleotides in lower case in the N terminal exon are potentially in error because of difficulties with sequencing this stretch, although the exon has been sequenced several times on both strands.

of fibronectin with those of other proteins (Figures 2 and 3) shows that homologous structural units occur in different proteins. This is almost certainly due to reassortment of exons encoding these repeats (see Discussion).

One type I repeat (I-12) of fibronectin differs from the rest in containing two extra cysteine residues to make six in all (Skorstengaard *et al.*, 1982). This type I repeat is encoded largely by a single exon but two of the cysteine residues fall in a second exon (Figures 2 and 4). This second exon also contains a short stretch of amino acid sequence which does not correspond with any of the homology types, is poorly conserved among species (Kornblihtt *et al.*, 1983; Schwarzbauer *et al.*, 1983) and is very susceptible to proteolysis. We refer to this as the hinge exon (Figures 2 and 4).

Exons encoding the 5′ and 3′ ends of fibronectin mRNA

The hinge exon is followed in the gene by a large exon which encodes the C-terminal segment of fibronectin, including the two interchain disulfide bonds, and the entire 3′-untranslated segment (Figure 4 and see below). Similarly, the exon encoding the first type I repeat does not contain the N-terminal segment of fibronectin (Figures 2 and 5). Approximately 1.2 kb 5′ of the type I-1 exon is an exon which encodes the N-terminus of mature fibronectin (QAQQ). The first methionine codon is 96 bases upstream of this. Analysis of the 32-amino acid segment between this methionine and the N-terminal glutamine suggests that it contains both a signal (pre) and a pro sequence (see Discussion).

In order to determine precisely the sites for initiation and termination of transcription we performed nuclease protection experiments using probes from the 5′ and 3′ ends of the gene hybridized with RNA preparations from various sources. In several experiments we observed protection of segments extending 137 bases 5′ of a *Pst*I site (Figure 5) and 101 ± 1 bases 3′ of an *Eco*RI site (Figure 4) which are respectively near the 5′ and 3′ ends of the gene. Representative results are shown in Figure 6. We did not detect protection of longer or shorter fragments using RNA preparations from liver, normal or transformed fibroblasts or astrocytes. Therefore, transcription appears to start at a single position which follows 25 bases after a TATAA box and ends 71 kb downstream 16 bases after a single AATAAA polyadenylation signal (see Figures 4–6 and Discussion). The sequence of the 3′ untranslated region is strikingly similar to those of human and bovine fibronectins (Kornblihtt *et al.*, 1983). Overall, the three species are 75–80% identical

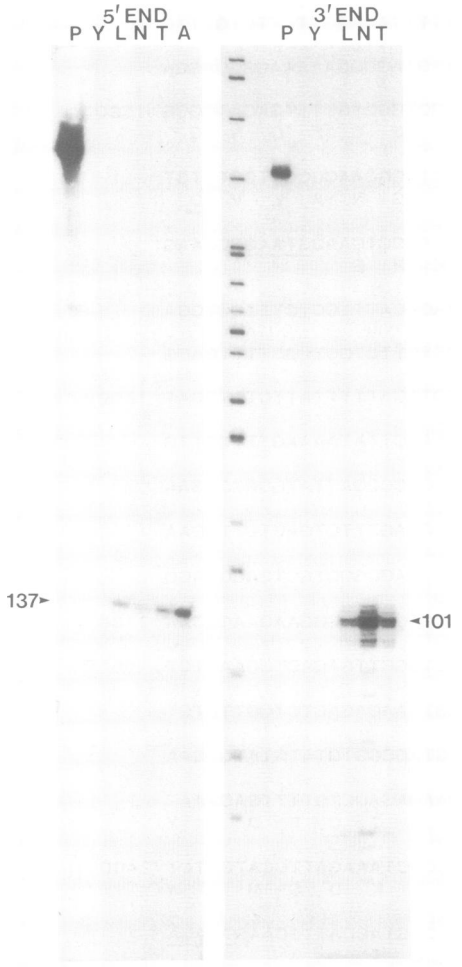


Fig. 6. RNase protection experiments locate unique 5′ and 3′ ends of rat fibronectin transcript. Complementary RNA transcripts were prepared using genomic fragments subcloned in pGEM2, hybridized with total cellular RNA and digested with ribonucleases. Protected fragments were analyzed on sequencing gels. The 5′ end was analyzed using an *Ava*I−*Pst*I fragment (Figure 5) and the 3′ end with an *Eco*RI−*Hin*dIII fragment (Figure 4). In each case, a single fragment or two fragments differing by a single base were protected by RNA preparations from a variety of cell types (L = liver, N = rat 1 normal fibroblast line, T = RSV-transformed rat 1, A = astrocytes, Y = yeast control, P = undigested probe).

and the last 210 bases are 90−95% identical. The reason for this strong conservation of 3′ untranslated segments is unknown.

Discussion

The data reported here establish the exon structure for more than 50% of the type I and II repeats of rat fibronectin. Each of these repeats is encoded by a single exon and, with a single exception, introns fall precisely between repeats. Although we have not sequenced the relevant exons, we presume that the other type I and II repeats are also encoded by single exons. A recent paper (Owens and Baralle, 1986) describes the exon structure for both type II and two type I (I-6 and I-7) repeats of human fibronectin. The results are in complete agreement with those presented here. It is clear that two basic exonic units encoding respectively type I and II repeats must have undergone duplication during evolution of the fibronectin gene. This endoduplication must have occurred long before the divergence of mammals since, while the rat sequences shown in Figures 2 and 3 are ~95% identical with the sequences of human and bovine fibronectins (Kornblihtt *et al.*, 1985; Skorstengaard *et al.*, 1986), individual repeats within a species are significantly less homologous with each other. The two type II repeats remain adjacent but the type I repeats have become separated into three blocks of six, three and three. Representative exons from each of these blocks are included in our sample (Table I, Figures 2−5).

The single exceptional type I repeat is the last one, which differs from the others in several respects. It has six cysteine residues as compared with four in all other repeats. All the cysteines are disulfide-bonded and the disulfide-bonding pattern has been determined to be 1−3 and 2−4 for several repeats (Skorstengaard *et al.*, 1982, 1984). The disulfide-bonding pattern for type I-12 has not been determined but has been assumed to be 1−4, 3−5 and 2−6 since cysteine residues 1, 3, 4 and 5 are homologous with those in other repeats. As shown here, cysteines 5 and 6 fall in a separate exon from cysteines 1−4 (Figures 2 and 4). Thus, the homologous block of sequence encoded by single exons in all other cases is split between two exons here. It would be of interest to know whether or not this change in exon structure correlates with a change in pattern of disulfide-bonding.

Homologous type I and II repeats occur in several other proteins (Figures 2 and 3; see Introduction). The homologies are extensive, strongly suggesting that these structural units share common evolutionary origins. Given the exon structure of the fibronectin gene it seems highly likely that exon units encoding type I and II homologies have become distributed among the various genes during evolution. The only other gene for a protein containing one of these repeats whose structure is known is that for tissue plasminogen activator (t-PA; Ny *et al.*, 1984). The single type I repeat in t-PA is encoded precisely by an exon exactly as are the type I repeats in fibronectin. These data provide strong support for the idea of exon shuffling suggested by Gilbert (1978). We predict that the type II modules in factor XII and PDC-109 will also prove to be encoded by single exons. As reported elsewhere (Vibe-Pedersen *et al.*, 1984; Hynes, 1985; Odermatt *et al.*, 1985; Oldberg and Ruoslahti, 1986; Schwarzbauer *et al.*, 1987b) the type III modules of fibronectin are encoded by pairs of exons in the gene but have not yet been detected in other proteins.

The sequence of the 5′ end of the rat fibronectin gene (Figure 5) also reveals the sequence of the primary translation product. There is a methionine residue, encoded by a sequence fitting the consensus for initiation codons (Kozak, 1984) which precedes the known N-terminal sequence of the mature protein (QAQQ) by 32 amino acids. However, analysis of the putative signal sequence using the values suggested by von Heijne (1982, 1986) does not predict removal of these 32 amino acids as a signal propeptide. Instead, this analysis suggests cleavage of a signal sequence of 19 or 21 amino acids. This putative signal sequence is followed by 11 or 13 amino acids, 5 of which are charged, ending at a lysine−arginine doublet before the N-terminal glutamine. While there is no proof of this supposition, the polar nature of this stretch suggests that it may be a prosequence removed by a subsequent cleavage after the dibasic pair rather than being part of a typical signal. Gutman *et al.* (1987) have recently reported a similar pre-pro sequence for human fibronectin. The purpose of such a pro sequence on fibronectin is obscure.

We also present sequence data on the 5′ and 3′ flanking regions of the rat fibronectin gene: these data lead to several conclusions. The nuclease protection experiments (Figure 6) suggest that the transcription starts at a single initiation site and ends at a single poly(A) addition site. These experiments do not absolutely rule out the presence of exons upstream or downstream of the regions

Table I. Intron–exon boundaries of rat fibronectin gene

N-TERM						.....GTC	AGT	CAG	AGC	AAG	C	gtgagtaccgac
						V	S	Q	S	K		
I-1	ctttcccctcaattcttcacag	CT	GGC	TGT	TTT	GAC.....GAG	AGC	AAG	CCT	GAA	C	gtaagtggaagg
		P	G	C	F	D	E	S	K	P	E	
I-2	ctttttttttttccttttccacag	CT	GAA	GAG	ACC	TGT.....AGC	TGT	ACC	AAT	GCA	A	gtaagaaagg
		P	E	E	T	C	S	C	T	I	A	
I-3	tccctttcccctaatgtttatttaccttgatctgtaacag	AT	CGC	TGC	CAT	GAA.....ACC	TGC	AAG	CCA	ATA	G	gtaagtgg
		N	R	C	H	E	T	C	K	P	I	
I-5	cttctttccccacacag	AC	AGA	TGC	AAT	GAT.....CAG	AGT	GCT	TCA	GCT	G	gtgag
		N	R	C	N	D	Q	S	A	S	A	
II-2	ccttcctgtctgtgtgttgtcttgggcag	TT	TTG	GTT	CAG	ACT.....TTC	TGC	CCA	ATG	GCT	G	gtaagaggaag
		V	L	V	Q	T	F	C	P	M	A	
I-7	tttctcctccatttccaaacag	CC	CAT	GAG	GAG	ATC.....TAC	TCC	CAG	CTC	CGA	G	gtactggg
		A	H	E	E	I	Y	S	Q	L	R	
I-10	tcttttccttttttttccccttttctcctccttctcatag	TC	AAC	GAA	GGC	CTG.....AGA	TGC	GAT	TCA	TCT	A	gtgagtag
		V	N	E	G	L	R	C	D	S	S	
I-11	tccccactttttccag	AA	TGG	TGC	CAT	GAC.....TTC	AAA	TGC	GAT	CCC	C	gtacg
		K	W	C	H	D	F	K	C	D	P	
I-12	tttctgctctgtgcccacgtgcag	AT	GAA	GCA	ACG	TGT.....TTC	GGG	GGC	CAG	CGG		gtaag
		H	E	A	T	C	F	G	G	Q	R	
HINGE	ttatgatatcctctcctctcag	GGC	TGG	CGC	TGT	GAC.....CAG	AGA	ACG	AAC	ACT		gtaagtg
		G	W	R	C	D	Q	R	T	N	T	
C-TERM	tttttcttctattttgacag	AAT	GTA	AAT	TGC	CCA.....						
		N	V	N	C	P						

The DNA sequences at the 5′ and 3′ boundaries of the exons encoding the N-terminal segment, eight type I repeats, one type II repeat, the hinge region and the C-terminal segment are shown. Intron sequences are in lower case and exon sequences are in upper case and arranged in codons above the corresponding amino acids which are given in single letter code. Note that all exons except the last three begin and end between the first and second bases of a codon.

whose sequences are given in Figures 4 and 5 although the sequence preceding the putative initiation site does not look like a 5′ splice site and no data suggest the existence of other exons. Therefore, it appears that transcription starts at a C residue 207 bases upstream of the ATG initiator codon, which is therefore the first ATG in the transcript. There is a TATAA box at −29 relative to the putative initiator site and CCAAT boxes at −40 and −283. Two GC boxes fitting the sequence for SP1 recognition sites (GGGCGGG, Dynan and Tjian, 1985) lie at −50 and −104 relative to the putative initiation site.

There are also several other potential transcriptional control elements in the 5′ flanking region and in the first intron of the gene (Figure 5). The fibronectin gene is subject to complex control mechanisms, being regulated by cell density and growth rate, oncogenic transformation, differentiation state and glucocorticoids (Adams *et al.*, 1982; Senger *et al.*, 1983; Oliver *et al.*, 1983; Tyagi *et al.*, 1983, 1985; Allebach *et al.*, 1985; Jochemsen *et al.*, 1986; Leibovitch *et al.*, 1986). Three octanucleotides similar to a suggested consensus sequence for glucocorticoid receptor binding sites (AGAA/T CAG A/T, Payvar *et al.*, 1983) lie at −21, −203 and −474. Three further reasonable matches to this octanucleotide sequence lie in the first intron (Figure 5); Moore *et al.* (1985) have reported glucocorticoid regulatory elements in the first intron of the human growth hormone gene. There are no obvious matches to other consensus sequences which have been suggested for glucocorticoid regulatory elements (Karin *et al.*, 1984; von der Ahe *et al.*, 1986). At position −161, the palindromic sequence CC..GTGACGTCAC..GG appears related to the consensus sequence suggested for genes regulated by cAMP in bacteria (Ebright *et al.*, 1984) and the central octanucleotide, TGACGTCA, is found in several eukaryotic genes which are regulated by cAMP (Montminy *et al.*, 1986). At position +110 is a 30-base sequence which is a good match to a consensus sequence found in the 5′ regions of genes regulated by interferons (Friedman and Stark, 1985). Fibronectin mRNA levels in human fibroblasts are increased by interferon (M.Zern, personal communication). A 38-base sequence near the start of the first intron is homologous with a consensus sequence noted near the 5′ end of several mRNAs for acute phase proteins (Dente *et al.*, 1985). Fibronectin is known to be an acute phase protein in rats (Pick-Kober *et al.*, 1986). Finally, there is a good match to the consensus sequence (C--GAA--TTC--G, Pelham, 1986) for a heat shock regulatory element at −765. It remains to be tested whether any or all of these potential control elements actually function in transcriptional control of the fibronectin gene.

In conclusion, the information on fibronectin gene structure presented here and elsewhere demonstrates the existence of a single gene encoding this family of glycoproteins which arose by endoduplication of three primordial minigenes, one for each of the three types of repeating units. Type I and II repeats are each encoded by one exon per repeat and these units also occur in other genes and proteins presumably via exon shuffling during evolution. Type III repeats are generally encoded by two exons per repeat and have not yet been detected in other genes or proteins. Certain of the type III repeats are alternatively spliced and the exon structure of these repeats is altered. Based on the unique start and stop sites for transcription of the fibronectin gene which we report here, this alternative splicing must be regulated by elements internal to the transcript and probably proximal to the regions of alternative splicing.

Materials and methods

Isolation of genomic clones

Genomic clones were isolated from a rat genomic library in EMBL3B (Tamkun

et al., 1984) using segments from the 5′ end of λrFN2 and subsequently the 5′ ends of successive clones. The segments were subcloned in pGEM vectors (Promega Biotec), checked for repetitive sequences and used to screen lambda plaques by standard methods (Maniatis *et al.*, 1982). Clones were analyzed by restriction enzyme mapping and Southern blotting and suitable fragments were subcloned into pGEM vectors for further analysis.

Preparation of cDNA clones from genomic fragments

Genomic fragments were subcloned into a murine retroviral vector, pLJ, which is a derivative of DOL (Korman *et al.*, 1987) from which the 5′ splice site has been deleted. As described elsewhere (Schwarzbauer *et al.*, 1987a) genomic fragments subcloned into such retroviral vectors are accurately spliced during generation of recombinant retrovirus. Cells derived by infection with these viruses therefore contain cDNAs derived from the genomic clones and these cDNAs can be recovered by fusion rescue. cDNA clones covering the 5′ 2 kb of rat fibronectin encoding nine type I repeats and two type II repeats were prepared in this way.

DNA sequencing

Parts of λrFN-1 were sequenced by the method of Maxam and Gilbert (1980) but most sequencing was by the dideoxy method of Sanger *et al.* (1977). Subclones in pGEM vectors were sequenced using Klenow fragment of DNA polymerase or reverse transcriptase and SP6 and T7 primers and the conditions suggested by Promega Biotec.

Nuclease protection analysis

RNAs were prepared by the guanidine thiocyanate procedure (Chirgwin *et al.*, 1979). End-labelled DNA probes were prepared and used according to standard methods (Maniatis *et al.*, 1982). RNA probes were prepared using subclones in pGEM vectors transcribed with SP6 or T7 polymerase (Promega Biotec) and purified on acrylamide gels. Total RNA (3 μg) plus 5 μg yeast tRNA were incubated with $1-3 \times 10^6$ c.p.m. of probe in 30 μl 80% formamide, 40 mM Pipes pH 6.4, 400 mM NaCl, 1 mM EDTA overnight at 30°C. The nucleases were removed by digestion with proteinase K and extraction with phenol–chloroform. Protected fragments were analyzed on sequencing gels. The exact sizes of these fragments were determined by comparison with adjacent sequencing reactions.

Acknowledgements

This work was supported by a grant from the USPHS National Cancer Institute (POI CA26712) and by grants to the M.I.T. Cancer Center from the National Cancer Institute, Bristol Meyers, Inc., and Ajinomoto, Inc. J.E.S. was a fellow of the Charles A.King Trust. E.O. was an EMBO fellow.

References

Adams,S.L., Boettiger,D., Focht,R.J., Holtzer,H. and Pacifici,M. (1982) *Cell*, **30**, 373–384.
Allebach,E.S., Boettiger,D., Pacifici,M. and Adams,S.L. (1985) *Mol. Cell. Biol.*, **5**, 1002–1008.
Baker,M.E. (1985) *Biochem. Biophys. Res. Commun.*, **130**, 1010–1014.
Banyai,L., Varadi,A. and Patthy,L. (1983) *FEBS Lett.*, **163**, 37–41.
Caput,D., Beutler,B., Hartog,K., Thayer,R., Brown-Shimer,S. and Cerami,A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1670–1674,
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5194–5299.
Dente,L., Ciliberto,G. and Cortese,R. (1985) *Nucleic Acids Res.*, **13**, 3941–3952.
Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774–778.
Ebright,R.H., Cossart,P., Gicquel-Sanzey,B. and Beckwith,J. (1984) *Nature*, **311**, 232–235.
Esch,F.S., Ling,N.C., Bohlen,P., Ying,S.Y. and Guillemin,R. (1983) *Biochem. Biophys Res. Commun.*, **113**, 861–867.
Friedman,R.L. and Stark,G.R. (1985) *Nature*, **314**, 637–639.
Gilbert,W. (1978) *Nature*, **271**, 501.
Gutman,A., Yamada,K.M. and Kornblihtt,A. (1987) *FEBS Lett.*, **207**, 145–148.
Hirano,H., Yamada,Y., Sullivan,M., deCrombugghe,B., Pastan,I. and Yamada, K.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 46–50.
Hynes,R.O. (1985) *Annu. Rev. Cell Biol.*, **1**, 67–90.
Hynes,R.O. (1986) *Sci. Am.*, **254**, 42–51.
Hynes,R.O. and Yamada,K.M. (1982) *J. Cell Biol.*, **95**, 369–377.
Jochemsen,A.G., Bernards,R., van Kranen,H.J., Houweling,A., Bos,J.L. and van der Eb,A.J. (1986) *J. Virology*, **59**, 684–691.
Karin,M., Haslinger,A., Holtgreve,H., Richards,R.I., Krauter,P., Westphal,H.M. and Beato,M. (1984) *Nature*, **308**, 513–519.
Korman,A.J., Frantz,J.D., Strominger,J.L. and Mulligan,R.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2150–2154.
Kornblihtt,A.R., Vibe-Pedersen,K. and Baralle,F.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3218–3222.
Kornblihtt,A.R., Umezawa,K., Vibe-Pedersen,K. and Baralle,F.E. (1985) *EMBO J.*, **4**, 1755–1759.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Leibovitch,S.A., Hillion,J., Leibovitch,M.P., Guillier,M., Schmitz,A. and Harel,J. (1986) *Exp. Cell Res.*, **166**, 526–534.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–580.
McMullen,B.A. and Fujikawa,K. (1985) *J. Biol. Chem.*, **260**, 5328–5341.
Montminy,M.R., Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6682–6686.
Moore,D.D., Marks,A.R., Buckley,D.I., Kapler,G., Payvar,F. and Goodman,H.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 699–702.
Ny,T., Elgh,F. and Lund,B. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5355–5359.
Odermatt,E., Tamkun,J.W. and Hynes,R.O. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6571–6575.
Oldberg,A. and Ruoslahti,E. (1986) *J. Biol. Chem.*, **261**, 2113–2116.
Oliver,N., Newby,R.F., Furcht,L.T. and Bourgeois,S. (1983) *Cell*, **33**, 287–296.
Owens,R.J. and Baralle,F.E. (1986) *FEBS Lett.*, **204**, 318–322.
Payvar,F., DeFranco,D., Firestone,G.L., Edgar,B., Wrange,O., Okret,S., Gustafsson,J.-A. and Yamamoto,K.R. (1983) *Cell*, **35**, 381–392.
Pelham,H. (1986) *Trends Genet.*, **1**, 31–35.
Pick-Kober,K.M., Munker,D. and Gressner,A.M. (1986) *J. Clin. Chem. Clin. Biochem.*, **24**, 521–528.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5466.
Schwarzbauer,J.E., Tamkun,J.W., Lemischka,I.R. and Hynes,R.O. (1983) *Cell*, **35**, 421–431.
Schwarzbauer,J.E., Muligan,R.C. and Hynes,R.O. (1987a) *Proc. Natl. Acad. Sci. USA*, **84**, 754–758.
Schwarzbauer,J.E., Patel,R.S., Fonda,D. and Hynes,R.O. (1987b) *EMBO J.*, **6**, 2573–2580.
Senger,D.R., Destree,A.T. and Hynes,R.O. (1983) *Am. J. Physiol.*, **245**, C144–150.
Skorstengaard,K., Thøgersen,H.C., Vibe-Pedersen,K., Petersen,T.E. and Magnusson,S. (1982) *Eur. J. Biochem.*, **128**, 605–623.
Skorstengaard,K., Thøgersen,H.C. and Petersen,T.E. (1984) *Eur. J. Biochem.*, **140**, 235–243.
Skorstengaard,K., Jensen,M.S., Sahl,P., Petersen,T.E. and Magnusson,S. (1986) *Eur. J. Biochem.*, **161**, 441–453.
Tamkun,J.W., Schwarzbauer,J.E. and Hynes,R.O. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5140–5144.
Tyagi,J.S,. Hirano,H., Merlino,G.T. and Pastan,I. (1983) *J. Biol. Chem.*, **258**, 5787–5793.
Tyagi,J.S., Hirano,H. and Pastan,I. (1985) *Nucleic Acids Res.*, **22**, 8275–8284.
Vibe-Pedersen,K., Kornblihtt,A.R. and Baralle,F.E. (1984) *EMBO J.*, **3**, 2511–2516.
von der Ahe,D., Renoir,J.M., Buchou,T., Baulieu,E.E. and Beato,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2817–2821.
von Heijne,G. (1982) *Eur. J. Biochem.*, **133**, 17–21.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
Yamada,K.M. (1983) *Annu. Rev. Biochem.*, **52**, 761–799.

Received on April 7, 1987; revised on June 5, 1987

Notes added in proof

The cDNA and parts of the genomic sequences have been submitted to the GENBANK and EMBL databases.

During processing of this manuscript, Dean *et al.*, *Proc. Natl. Acad. Sci. USA*, **84**, 1876–1880 (1987) reported a sequence of the human fibronectin promoter which is homologous with the rat sequence we report here.