The EMBO Journal vol.3 no.1 pp.221 – 226, 1984

Human fibronectin: molecular cloning evidence for two mRNA species differing by an internal segment coding for a structural domain

A.R. Kornblihtt*, K. Vibe-Pedersen and F.E. Baralle

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

*To whom reprint requests should be sent

Communicated by F.E.Baralle

Two different fibronectin (FN) mRNA species were detected in the human cell line Hs578T. One species, mRNA I, contains an additional 270 nucleotide long insert (ED) that encodes exactly one of the internally repeated structural domains of the protein. The 90 amino acid extra domain belongs to the so-called type III homology and it is located in the carboxy-terminal half of FN, in between the cell attachment and the heparin binding sites of the protein. The evidence of two mRNAs is provided by the isolation and characterisation of four independent cDNA clones from a library prepared with a synthetic oligonucleotide primer, and it was confirmed by S1 nuclease analysis of cDNA/mRNA hybrids. This kind of analysis also showed that in the human cell line, mRNA I is present at a lower level than mRNA II (the mRNA species without the ED), whilst in human liver, mRNA I is virtually undetectable. Since liver tissue has recently been reported to be the source of plasma FN, our results indicate that the presence of the ED insert could be a particular feature of cellular FN.

Key words: fibronectins/mRNA 'walking'/cDNA sequencing/S1 mapping/structural domains

Introduction

Fibronectins (FNs) are high mol. wt. adhesive glycoproteins found in plasma and in extracellular matrices. They are involved in various contact processes such as attachment of cells onto substrata, cell migration, maintenance of normal cell morphology, opsonization and wound healing. Several binding activities have been located in different domains of the FN molecule. Indeed, FN has affinities for collagen, heparin and other glycosaminoglycans, fibrin, cell surfaces, bacteria and itself (for reviews, see Mosesson and Amrani, 1980; Hynes and Yamada, 1982; Yamada, 1983).

Two forms of FN have been identified. Plasma FN is a soluble protein and is a dimer of polypeptide chains of mol. wt. ~220 000 (Engvall *et al.*, 1978). Cellular FN can be either dimeric or multimeric, and is secreted by various cell types (Hynes and Destree, 1977). Although plasma and cellular FNs are very similar in structure, they differ in solubility, electrophoretic behaviour, certain biological activities (Yamada and Kennedy, 1979) and immunogenicity (Atherton and Hynes, 1981).

It has long been speculated that the differences between cellular and plasma fibronectins and also between the polypeptide chains of both forms could be due to multiple genes, multiple mRNAs, or differential post-translational modifications. However, little is known of the differences in primary structure both between and amongst the chains of cellular and plasma FNs. The recent isolation of a human FN cDNA clone (Kornblihtt *et al.*, 1983) has enabled us to test directly for the first time some of these hypotheses. Here we report the existence of two mRNA species for FN in the human cell line Hs578T. One of them, the less abundant, contains an additional 270 nucleotide insert that encodes exactly one of the internally repeated structural domains of the protein (Petersen *et al.*, 1983). We also found that liver tissue, reported to be a source of plasma FN (Tamkun and Hynes, 1983), does not contain significant amounts of the insert-bearing mRNA species.

Results

Isolation of human cDNA clones

We have previously isolated and characterized a cDNA clone for human cellular FN (Kornblihtt *et al.*, 1983). This clone (pFH1) extended 2.1 kb from the poly(A) tail, representing approximately one quarter of the estimated size of human FN mRNA (7900 nucleotides). To further our knowledge of FN mRNA(s), we undertook the construction of longer cDNA clones by 'mRNA walking'. For this purpose, a synthetic oligonucleotide complementary to the sense strand of a region near the 5′ end of clone pFH1 was prepared. The oligonucleotide sequence: 5′ GTGCATCAATGGCAG 3′ is located 30 bp downstream of the *Sac*I site of clone pFH1 (see Figure 1). This oligonucleotide was used to specifically prime reverse transcription of total RNA from the human cell line Hs578T. Double-stranded cDNA was cloned into the vector pAT153/*Pvu*II/8 and cDNA clones were screened by probing with an end-labelled restriction fragment of pFH1 located upstream of the synthetic primer sequence. A large number of FN cDNA clones (~180) with varying insert lengths were obtained.

Based on restriction enzyme analysis, two families of cDNA clones were found. Restriction maps of four independent clones representative of the two families are aligned in Figure 1 to show overlapping areas. Family I clones (pFH23 and pFH111) contain a 270-bp internal fragment (named ED, extra domain, for convenience) characterized by the presence of the additional sites for the restriction enzymes *Xho*I and *Pst*I, while family II clones (pFH37 and pFH154) lack this fragment. Clones pFH1 and pFH154 together cover 4.6 kb which represents ~60% of the estimated length of human FN mRNA (7.9 kb).

S1 nuclease analysis

To determine whether the two types of cDNA were faithful copies of mRNA species present in the cell, S1 nuclease mapping of cDNA/RNA hybrids was performed. Two probes were prepared: probe I was the 446 nucleotide 3′ end-labelled anti-sense strand *Sau*961-*Hae*II of clone pFH23 and probe II was the 176 nucleotide 3′ end-labelled anti-sense strand

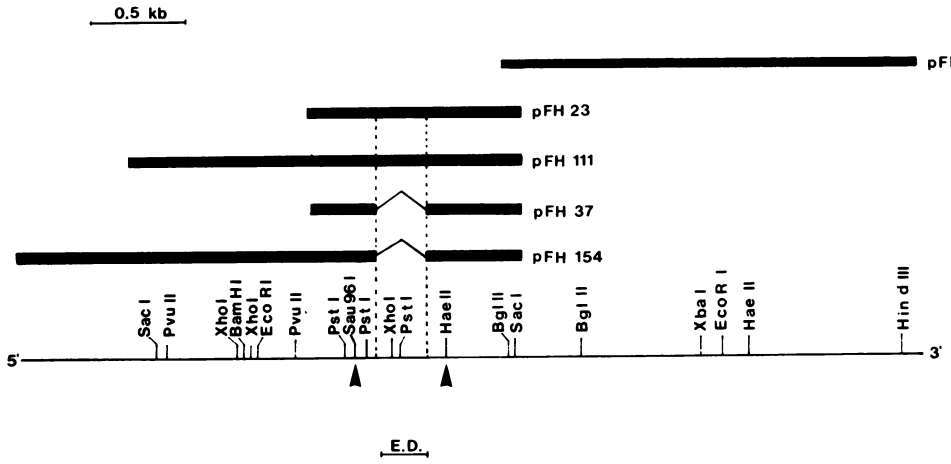

Fig. 1. Restriction enzyme map of different human fibronectin clones. The isolation of clone pFH1 was previously reported (Kornblihtt *et al.*, 1983). Isolation of clones pFH23, pFH111, pFH37 and pFH154 is outlined in the text and in Materials and methods. The 3′ end of the map is the 3′ end of the FN mRNA while the 5′ end of the map corresponds to approximately the middle of the FN mRNA molecule. Clone inserts (thick lines) are aligned to show overlapping areas. Arrows indicate the sites used in making the probes for S1 nuclease analysis. ⟋⟍ indicates contiguity. ED, extra domain.

*Sau*961-*Hae*II of clone pFH37 (see arrows in Figure 1). When probe I was hybridized to total RNA from Hs578T cells, two S1 nuclease-resistant DNA fragments were detected (Figure 2, products A and B). Product A (446 nucleotides) represents the fully protected probe and indicates the existence of mRNA molecules (mRNA I) containing the 270 nucleotides ED fragment that give rise to family I clones. Product B (119 nucleotides) exactly covers the distance between the *Sau*961 site and the base in which families I and II clones diverge (shown below by DNA sequencing).

When probe II was used, again two S1 nuclease-resistant products were observed (Figure 2, products C and D). Band C (176 nucleotides) is the fully protected probe, this time indicating the existence of mRNA molecules (mRNA II) lacking the ED fragment that give rise to family II clones; while band D, co-migrating with band B, again represents the distance between the labelled end of the probe and the point of divergence between the two types of clones. The combined data from both S1 nuclease mapping experiments has allowed us to conclude that Hs578T cells have at least two different species of FN mRNA. Densitometer tracings (Figure 3) indicated that mRNA I (bands A and D) is present at ~25% the level of mRNA II (bands B and C). This observation is in agreement with the fact that most of the positive clones analysed fell into family II.

When similar S1 analysis was carried out with human liver poly(A)$^+$ RNA, only bands B and C were detected. Bands A and D cannot be distinguished above the background, suggesting that liver tissue produces mainly the form of FN mRNA without the ED (mRNA II). Figure 3 shows a comparison of densitometer tracings between S1 mapping experiments performed with Hs578T cell line and liver RNAs. It is evident that in liver mRNA I is <5% the level of mRNA II.

DNA sequence analysis

The inserts of clones pFH23 and pFH37 were sequenced by the chemical degradation method of Maxam and Gilbert. DNA sequence analysis (Figure 4A) confirmed that the portions of clone pFH23 outside the 270-bp ED fragments are identical to their clone pFH37 counterparts with the exception of residue 798 where a T→A transversion is observed. The ED sequence presents an open reading frame which is in phase with the rest of the sequence of pFH23. More strikingly, if the ED segment was taken out of the pFH23 sequence and the remaining ends joined together, the deduced amino acid sequence obtained would be identical to the one deduced from pFH37 (Figure 4B). It should be noted that although amino acid 181 (alanine) is outside the ED, the first base of its codon belongs to the ED sequence, so that in the pFH37 clone that amino acid is changed to a threonine (arrow in Figure 4B, line II).

Accumulative evidence (Petersen *et al.*, 1983; Hirano *et al.*, 1983; Skorstengaard *et al.*, 1982) suggests that FN arose from gene duplications of at least three units corresponding to the three types of internal homologies (I, II and III) found in bovine plasma fibronectin. We found that the polypeptide coded by the 270-bp ED fragment is exactly one 90 amino acid long domain of type III homology.

Discussion

We have shown that at least two different species of FN mRNA occur in the human cell line Hs578T. This observation is based on the isolation and characterization of four independent cDNA clones and it was confirmed by S1 nuclease mapping. The minor species (mRNA I) contains a 270 nucleotides insert while the major species (mRNA II) lacks it. It seems important to point out that the presence of the ED segment may not be the only length difference between the two mRNAs detected.

It could be possible that mRNA I represents a partially processed precursor of a mature mRNA (mRNA II). This seems unlikely because the ED is an open reading frame, in phase with the rest of the sequence, and the polypeptide coded by it has marked homologies with one type of protein domain sequence known to exist in bovine plasma FN (homology type III).

The 90 amino acids long polypeptide coded by the ED is located in the COOH-terminal half of fibronectin in between the cell attachment and the heparin binding domains of the protein. In fact, the deduced amino acid sequence of the regions that flank the ED (amino acids 1–90 and 181–270 in Figure 4A) represent another two domains of type III homology. The domain flanking the NH_2 terminus of the ED

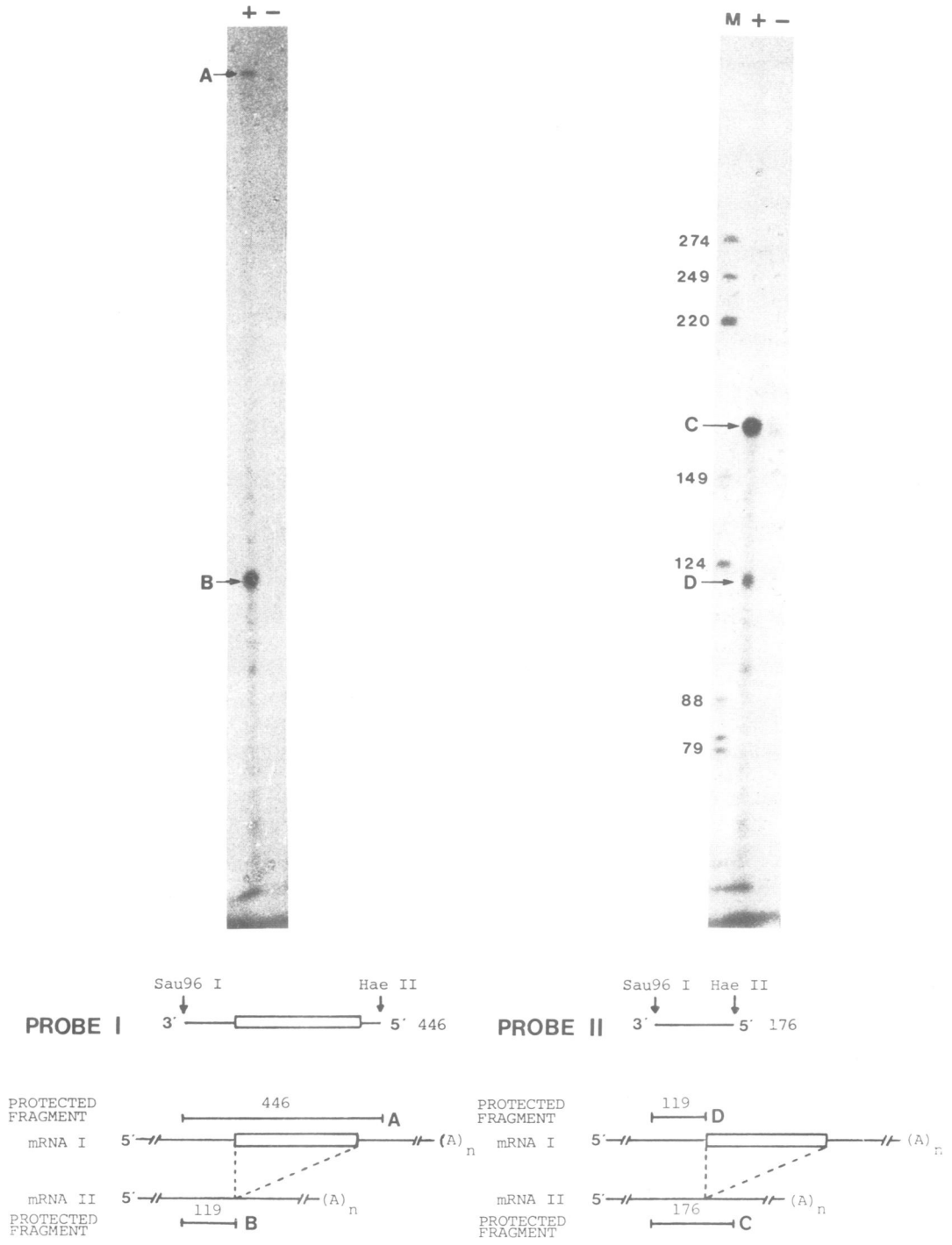

Fig. 2. S1 nuclease analysis of Hs578T cells total RNA/cDNA hybrids. Probe I was the *Sau*961-*Hae*II fragment of pFH23 insert and probe II was the *Sau*961- *Hae*II fragment of pFH37 insert. Probe fragments, S1 nuclease-resistant DNA products and mRNA species are shown in the diagram below the figure. + and − indicate the presence and absence of Hs578T cells RNA in the hybridization mixture, respectively. Correspondence between the diagrams and the gels are indicated by letters. Marker (M) lengths are given in nucleotides.

polypeptide overlaps a sequence reported to be contained in the cell-binding fragment of human plasma FN (Pierschbacher *et al.*, 1982), while the domain flanking the carboxy terminus of the ED is homologous to a fragment of bovine plasma FN carrying the heparin-binding activity (Petersen *et al.*, 1983). A detailed analysis of sequence homologies of the clones depicted in Figure 1 will be published elsewhere.

Two explanations could account for the origin of the two species of mRNA. Either both mRNAs are transcribed from the same gene or each one comes from a different gene. Genomic blot experiments are consistent with the existence of a single fibronectin gene in the human haploid genome (Kornblihtt *et al.*, 1983 and unpublished results). The observation that, except for the ED region and for residue number 798, the sequenced portions of clones pFH23 and pFH37 are identical for an extension of ~540 residues (1−271, 542−797, 799−810) may indicate that they are transcribed from the same DNA sequence. Although the 798 T→A transversion could reflect the presence of two genes, it can also be explained by allelic variations. The two mRNA species may

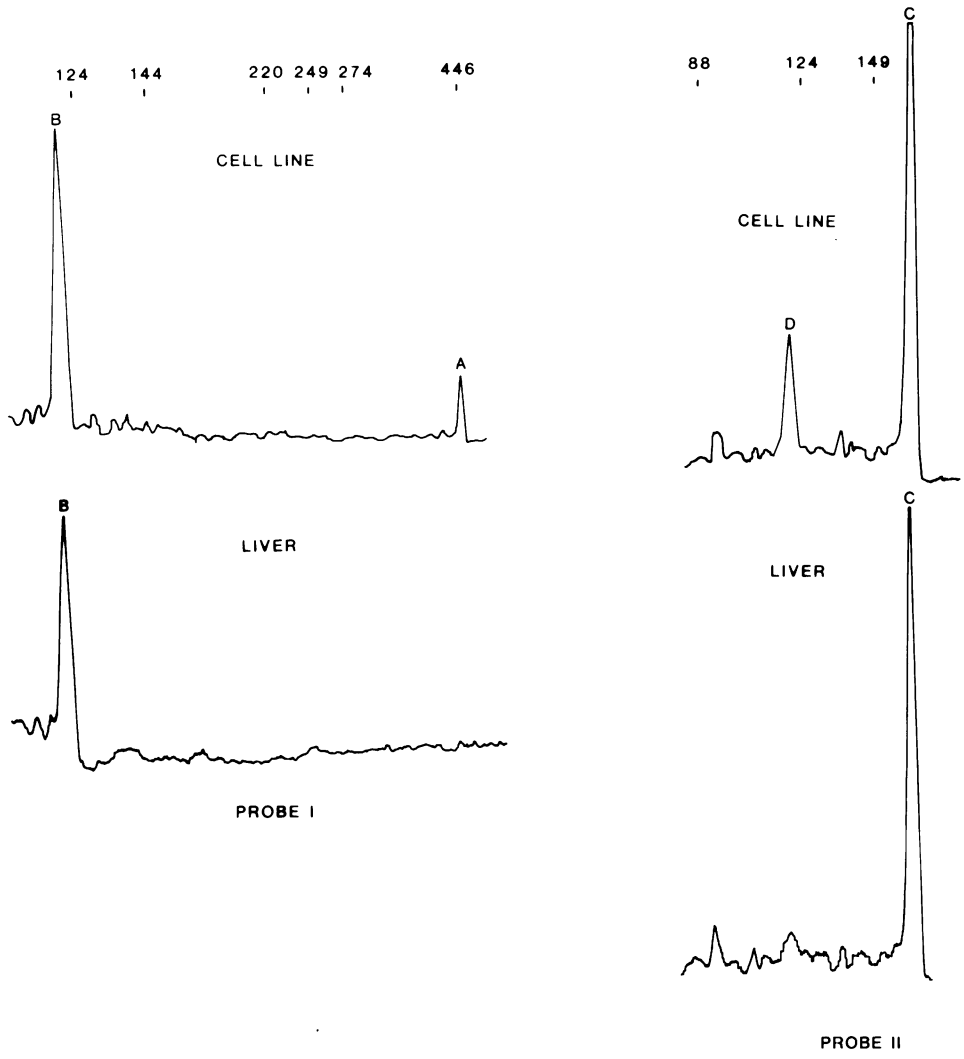

Fig. 3. Densitometer tracings of the autoradiograms of S1 nuclease mapping experiments. **Left**: S1-resistant DNA products obtained when probe I was hybridized to Hs578T cells total RNA (**top**) or human liver poly(A)$^+$ RNA (**bottom**). **Right**: S1-resistant DNA products obtained when probe II was hybridized to Hs578T cells total RNA (**top**) or human liver poly(A)$^+$ RNA (**bottom**). The letters refer to the protected fragments shown in Figure 2. Marker lengths are given in nucleotides.

then arise by a process of alternative splicing (Hagenbüchle *et al.*, 1981; Amara *et al.*, 1982; Crabtree and Kant, 1982) of a common precursor similar to the one described for the α A-crystallin gene (King and Piatigorsky, 1983). Mapping and sequence analyses of human FN genomic clones containing the ED exon are currently under way. In fact, preliminary sequence data show the presence of introns in the ED boundaries.

Finally, the possible biological significance of two FN mRNAs deserves consideration. It has been widely accepted that plasma FN consists of two non-identical subunit chains with mol. wts. varying from 210 000 to 230 000 daltons depending on the species, whereas cellular FN would be composed of identical subunit chains of ~220 000 daltons (Hynes and Yamada, 1982). One interesting possibility is that the two mRNAs reflect two different polypeptide subunits that form cellular FN dimers, in which case the concept of identical subunit chains for cellular FN would have to be revised. It would be conceivable that cellular FN is made of two protein subunits that, although different in primary structure, co-migrate in SDS gel electrophoresis, due to differential post-translational modifications like glycosylation, phosphorylation and sulfation.

Alternatively, the existence of two FN mRNAs reported in this paper may be related to the difference between plasma and cellular FNs. Differences in basic structure between plasma and cellular forms have been detected by monoclonal antibodies (Atherton and Hynes, 1981) and by peptide mapping (Hayashi and Yamada, 1981). These authors showed that, among other differences, heparing-binding fragments contain regions that are 10 000 daltons larger in cellular than in plasma chicken FNs. The size (10 000 daltons $\cong$ 90 amino acids) and location of this polypeptide are coincident with those of the ED. All the differences were mapped in internal parts of the molecules and cannot be attributed to proteolytic processing alone. We have recently determined by cDNA sequencing (Kornblihtt *et al.*, 1983) that a stop codon follows immediately after the C-terminal residue of mature FN, showing that no proteolytic process occurs at the C terminus of FN after it is synthesized in the ribosome. Recently, Tamkun and Hynes (1983) reported that the plasma form of FN is secreted by hepatocytes. The cellular form is also syn-

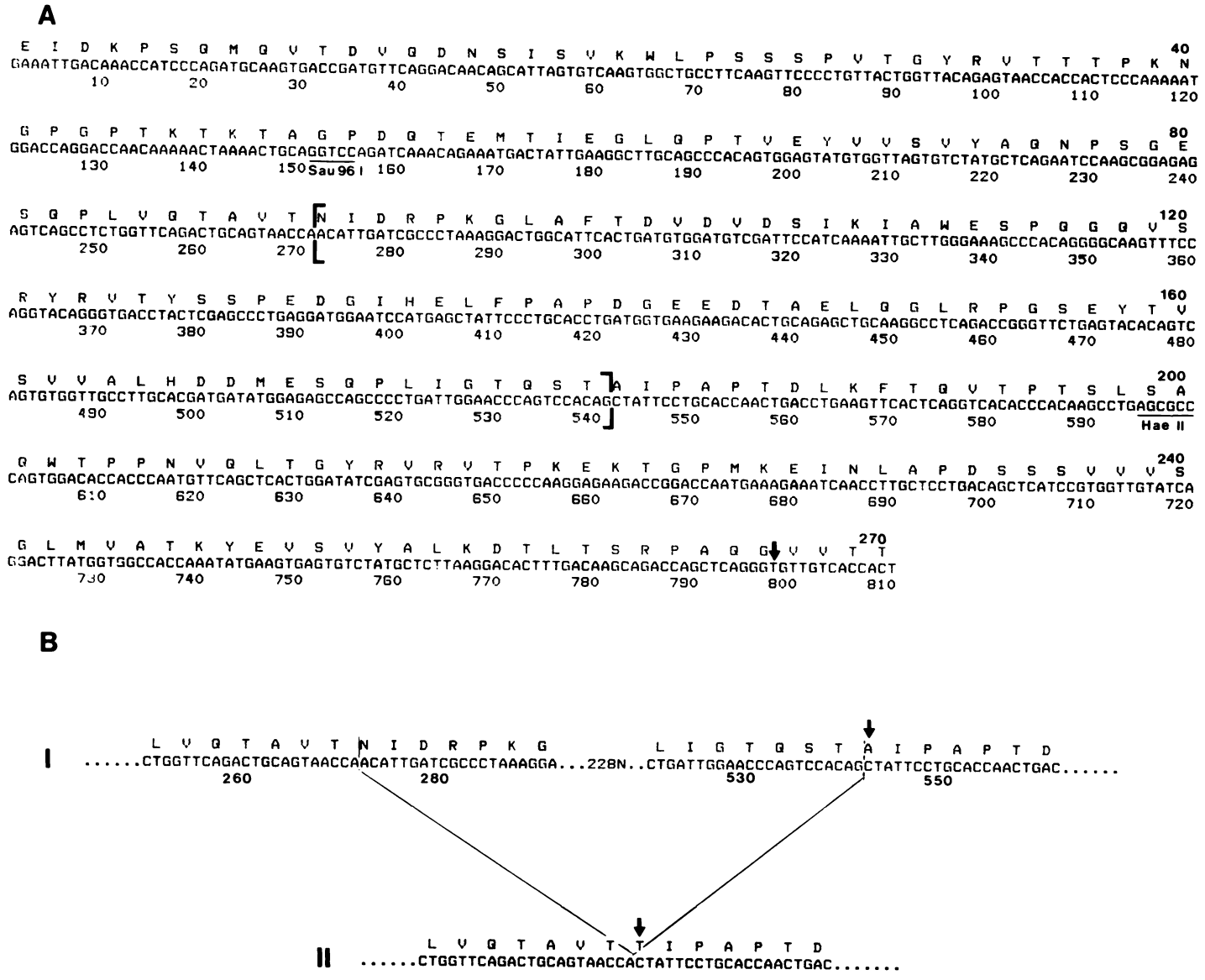

Fig. 4. Nucleotide sequence and predicted amino acid sequence (top row) of clone pFH23. Only the regions corresponding to three 90 amino acids long units of type III homology is shown. Brackets enclose the ED sequence. The nucleotide sequence of pFH37 can be obtained by reading continuously residues 1–271, 542–810. The arrow indicates nucleotide 798 which is a T in pFH23 but an A in pFH37. **(B)** Nucleotide sequence and deduced amino acid sequence (top row) of the ED boundaries in clone pFH23 (I) compared with the region of clone pFH37 around the point of divergence (II). Arrows indicate amino acids that change (Ala→Thr) as a consequence of a codon recomposition.

thesized by hepatocytes but the majority of the FN secreted is of the plasma type. Experiments depicted in Figure 3 indicate that liver produces mostly mRNA II, i.e., the species lacking the 270 bases ED insert. Furthermore, the predicted amino acid sequence of clone pFH37 (mRNA II) matches perfectly (except for a few interspecies changes) with the sequence of a peptide fragment of bovine plasma FN (segment S6 in Petersen *et al.*, 1983). These facts stress the possibility that the ED polypeptide is absent in plasma FN. Then the presence of FN mRNA containing the ED could be a particular feature of those cells that produce the 'cellular' form of FN, or a phenomenon restricted to certain cell lines. The presence of an extra repetitive unit (ED) located in between the cell-binding and heparin-binding domains of one of the FN forms may explain the differences in solubility and in the ability to control cell morphology found between cellular and plasma FNs.

While only limited deductions can be made from sequencing data alone about the biological relevance of the ED segment, the results described in this paper constitute the first evidence, at primary structure level, of differences between FN polypeptides.

Materials and methods

RNA preparation

The humam tumor Hs578T cells (Hacket *et al.*, 1977) were cultured in Dulbecco's modified Eagle's medium containing 10% foetal calf serum. Total RNA was extracted from confluent cell monolayers by the guanidine-HCl method (Chirwing *et al.*, 1979). Human liver poly(A)$^+$ RNA was prepared by chromatography of human liver total RNA through poly(U)-Sepharose (Pharmacia) column according to the supplier's instructions.

mRNA walking

The oligonucleotide synthesis was performed by the solid phase phosphotriester method developed by Gait *et al.* (1980). Double-stranded cDNA was

obtained essentially as described in Kornblihtt *et al.* (1983) using a synthetic oligonucleotide to prime reverse transcription of total RNA. Phosphorylation of the 5'-OH of the oligonucleotide primer was performed by incubating blunt ended double-stranded cDNA with ATP and polynucleotide kinase according to Maniatis *et al.* (1982). Double-stranded cDNA was blunt end ligated into the pAT153/*Pvu*II/8 vector without further treatment as described by Shoulders and Baralle (1982). The ligated molecules were used to transform *Escherichia coli* MC1061 (Casadaban and Cohen, 1980). Colonies were transferred to Whatman 541 papers and prepared for hybridization according to Gergen *et al.* (1979). For the screening of the colonies with an end-labelled restriction fragment of clone pFH1 (Kornblihtt *et al.*, 1983) paper replicas were prehybridized, hybridized and washed as previously described (Shoulders and Baralle, 1982). This fragment (78 bp long) extended from a *Hind*III site in the vector flanking the 5' end of clone pFH1 insert to the *Sac*I site in the insert.

Sequence determination

Restriction fragments were labelled at the 3' end by 'filling in' and then subjected to the chemical degradation procedure of Maxam and Gilbert (1977).

S1 nuclease mapping

S1 nuclease mapping was performed as described by Berk and Sharp (1977). Probes were labelled at their *Sau*961 sites by 'filling in' with the Klenow fragment of DNA polymerase I and [α-^{32}P]dGTP. Probes isolated by polyacrylamide gel electrophoresis were strand separated according to Maxam and Gilbert (1980). 10 μg of Hs578T cells total RNA or of human liver poly(A)$^+$ RNA were mixed with the appropriate probe in the presence of 20 μg of poly(A) and co-precipitated with ethanol. Pellets were resuspended in 12.5 μl of 80% formamide, 10 mM PIPES (pH 6.5), 1 mM EDTA, 0.4 M NaCl, heated at 73°C for 10 min under paraffin oil and hybridized overnight at 52°C (probe I) or 46°C (probe II). The hybrids were diluted with 150 μl of S1 buffer (250 mM NaCl, 30 mM NaOAc pH 4.4 1 mM $ZnSO_4$), digested with 3000 U of S1 nuclease (Boehringer) for 60 min at 30°C, mixed with 50 μl of S1 stop (15 mM EDTA, 600 μg/ml yeast tRNA), ethanol precipitated and analyzed on 6% polyacrylamide 7 M urea gels.

Acknowledgements

We thank Mr. K.K. Chada for the oligonucleotide synthesis and Drs. E. Whitelaw and N. Proudfoot for advice on S1 nuclease mapping. K.V.-P. received support from the Danish Natural Science Research Council.

References

Amara,S.G., Jonas,V., Rosenfeld,M.G., Ong,E.S. and Evans,R.M. (1982) *Nature*, **298**, 240-244.
Atherton,B.T. and Hynes,R.O. (1981) *Cell*, **25**, 133-141.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Casadaban,M.J. and Cohen,S.N. (1980) *J. Mol. Biol.*, **138**, 179-207.
Chirwing,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.
Crabtree,G.R. and Kant,J.A. (1982) *Cell*, **31**, 159-166.
Engvall,E., Ruoslahti,E. and Miller,E.J. (1978) *J. Exp. Med.*, **147**, 1584-1595.
Gait,M.J., Singh,M., Sheppard,R.C., Edge,M.D., Greene,A.R., Heathcliffe,G.R., Atkinson,T.C., Newton,C.R. and Markham,A.F. (1980) *Nucleic Acids Res.*, **8**, 1081-1096.
Gergen,J.P., Stern,R.H. and Wensink,P.C. (1979) *Nucleic Acids Res.*, **7**, 2115-2136.
Hackett,A.J., Smith,H.S., Springer,E.L., Owens,R.B., Nelson-Rees,W.A., Riggs,J.L. and Gardner,M.B. (1977) *J. Natl. Cancer Inst.*, **58**, 1795-1800.
Hagenbüchle,O., Tosi,M., Schibler,U., Bovey,R., Wellauer,P.K. and Young,R. (1981) *Nature*, **289**, 643-646.
Hayashi,M. and Yamada,K.M. (1981) *J. Biol. Chem.*, **256**, 11292-11300.
Hirano,H., Yamada,Y., Sullivan,M., de Crombrugghe,B., Pastan,I. and Yamada,K.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 46-50.
Hynes,R.O. and Destree,A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 2855-2859.
Hynes,R.O. and Yamada,K.M. (1982) *J. Cell Biol.*, **95**, 369-377.
King,C.R. and Piatigorsky,J. (1983) *Cell*, **32**, 707-712.
Kornblihtt,A.R., Vibe-Pedersen,K. and Baralle,F.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3218-3222.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Mosesson,M.W. and Amrani,D.L. (1980) *Blood*, **56**, 145-158.
Peterson,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.
Pierschbacher,M.D., Ruoslahti,E., Sundelin,J., Lina,P. and Petersen,P.A. (1982) *J. Biol. Chem.*, **257**, 9593-9597.
Shoulders,C.C. and Baralle,F.E. (1982) *Nucleic Acids Res.*, **10**, 4873-4882.
Skorstengaard,K., Thøgersen,H.C., Vibe-Pedersen,K., Petersen,T.E. and Magnusson,S. (1982) *Eur. J. Biochem.*, **128**, 605-623.
Tamkun,J.W. and Hynes,R.O. (1983) *J. Biol. Chem.*, **258**, 4641-4647.
Yamada,K.M. (1983) *Annu. Rev. Biochem.*, **52**, 761-799.
Yamada,K.M. and Kennedy,D.W. (1979) *J. Cell Biol.*, **80**, 492-498.

Received on 19 October 1983

Note added in proof

Since submission of this manuscript, we have also detected two species of FN mRNA differing by the ED segment in normal human fibroblasts.