Human fibronectin is synthesized as a pre-propolypeptide

Alejandro Gutman, Kenneth M. Yamada+ and Alberto Kornblihtt°

Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI-CONICET), Vuelta de Obligado 2490, 1428 Buenos Aires, Argentina and *+Laboratory of Molecular Biology, National Cancer Institute, NIH, Bethesda, MD 20892, USA*

Received 9 July 1986; revised version received 25 August 1986

Fibronectins (FNs) are extracellular glycoproteins consisting of dimers or multimers of similar but not identical subunits. The subunit differences result from variations in internal primary sequence due to alternative splicing in at least 2 regions of the pre-mRNA. The complete amino acid sequence of mature human cellular FN has been reported recently from cDNA cloning and sequencing. The same approach has now enabled us to deduce, for the first time, that FN has a 26 amino acid signal peptide and that it undergoes proteolytic processing at its N-terminus to eliminate a 5 amino acid pro-sequence (Ser-Lys-Ser-Lys-Arg). The signal sequence matches the consensus format, while this pro-sequence is a distinctive, very hydrophilic and basic peptide.

(Human) Fibronectin Signal peptide Pro-peptide cDNA

1. INTRODUCTION

Fibronectin (FN) represents a family of very closely related glycoproteins found in blood plasma and on cell surfaces. FN is involved in a wide variety of cell-cell and cell-substrate processes, including cell attachment and spreading, cell migration during embryogenesis, opsonization, oncogenic transformation and wound healing. This biological versatility is due to the co-existence, in the same protein, of binding sites for collagen, fibrin, heparin, bacteria, DNA and cell surfaces [1–3].

Two major forms of FN have been identified: a plasma form which is a soluble heterodimer and a cellular form which consists of dimers or highly insoluble multimers. The two forms contain disulfide-linked subunits taken from a pool of similar but not identical polypeptides. All possible subunits are produced by a single gene [4], and their primary structural differences are the results of a complex pattern of alternative splicing in at least two regions of the primary transcript termed ED and IIICS [5–12].

Recently, the complete primary structure of mature human FN has been determined by sequencing a series of overlapping cDNA clones [11]. Although the literature on partial peptide sequencing of FN is very extensive, no data are available on the N-terminal amino acid sequence of in vitro translated FN to show the existence of a prepro-FN. We report here the isolation and sequencing of cDNA clones that demonstrate the presence of signal and pro-peptides in human FN.

2. MATERIALS AND METHODS

2.1. *RNA preparation*

Human cell line Hs578T [13] was cultured in Dulbecco's modified Eagle medium containing 10% fetal calf serum. Total RNA was extracted from confluent cell monolayers by the guanidine-HCl method [14]. Poly(A)$^+$ RNA was obtained by chromatography on oligo-dT cellulose [15].

° To whom reprint requests should be addressed

Dedicated to Dr Luis F. Leloir on the occasion of his 80th birthday, 6 September 1986

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

2.2. *cDNA cloning*

Isolation of clone pFH6 by 'mRNA walking' has been described [11]. An oligonucleotide primer (3'-CGAGTCGTTTACCAAGTCG-5') complementary to the mRNA region close to the 5'-end of pFH6 was synthesized by the phosphotriester method [16]. The oligonucleotide was used to prime reverse transcription of poly(A)$^+$ RNA from Hs578T cells. Blunt-ended ds cDNA was prepared by the RNase H method [17] and cloned into pAT153/Pvu 11/8 [18]. Colonies were screened using as probe a 27 bp *Hin*dIII-*Dde*I restriction fragment from the 5'-end of pFH6 (fig.1), labelled by filling-in at the *Hin*dIII end. Whatman 541 paper replicas were obtained [19] and hybridized with the probe in 5 × SSC, 1% SDS, 1 × Denhardt's solution, 100 μg/ml herring sperm DNA, at 45°C overnight. After washing with 5 × SSC, 0.1% SDS at 25°C, two independent clones, pFH 2 and pFH 3, were obtained.

2.3. *DNA sequencing*

It was performed by the chemical degradation procedure [20].

2.4. *S_1 mapping*

S_1 nuclease mapping was performed as follows: a 501 bp *Taq*I fragment from pFH3 was labelled with [γ-^{32}P]ATP and polynucleotide kinase after dephosphorylation with calf intestinal phosphatase. The probe was hybridized to 25 μg of Hs578T total RNA in formamide at 54°C. Hybridization and subsequent steps were as described [5].

3. RESULTS

Fig.1 shows the organization of clones pFH2 and pFH3. These cDNA clones were constructed with a synthetic primer complementary to the 5'-end of a previously isolated clone, pFH6, which reached the N-terminus of mature FN. Primer extension experiments (not shown) using the same primer showed that the sequences of hFN mRNA upstream of those coding for the N-terminal glutamine of FN [21,22] should be no longer than 380 bases. The insert of clone pFH2 corresponds to the first 119 bases of the 5'-sequences. Clone pFH3 has a 258 bp insert beginning, as did the pFH2 insert, with the primer sequence. However, the furthest upstream 139 bp of the pFH3 insert do

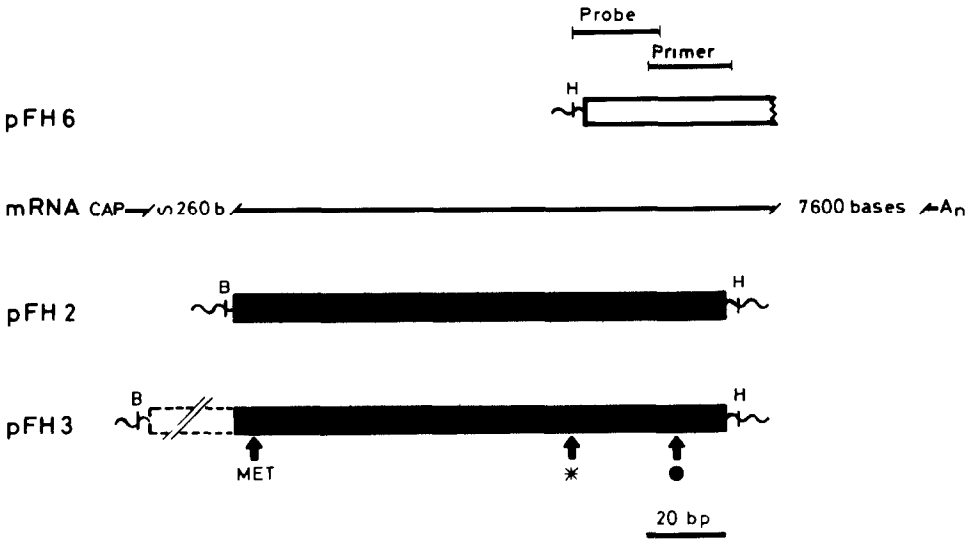

Fig.1. Localization of the inserts of clones pFH6, pFH2 and pFH3 along FN mRNA. Wavy lines are vector sequences. B = *Bam*HI; H = *Hin*dIII. (*) Region encoding the predicted cleavage site for signal peptidase. (●) Region encoding the cleavage site between pro- and mature FN. pFH3 contains 139 bp of FN-unrelated DNA due to a cloning artifact (dotted empty box).

```
        1                                   10
        M  L  R  G  P  G  P  G  L  L  L  L  A  V  Q  C  L  G  T
...AACATGCTTAGGGGTCCGGGGCCCGGGCTGCTGCTGCTGGCCGTCCAGTGCCTGGGGACA

20                        *          30    ●
/A  V  P  S  T  G  A  S /K  S  K  R  Q  A  Q  Q  M  V  Q
GCGGTGCCCTCCACGGGAGCCTCGAAGAGCAAGAGGCAGGCTCAGCAAATGGTTCAGC...
                          TaqI
```

Fig.2. Nucleotide sequence and deduced amino acid sequence of human FN cDNA clone pFH2. DNA: sequence complementary to the primer and *Taq*I site used in S_1 mapping (fig.3) are underlined. Protein: the processing 'window' is shown by slashes. (*) Predicted cleavage site for signal peptidase. (●) Cleavage site between pro- and mature FN.

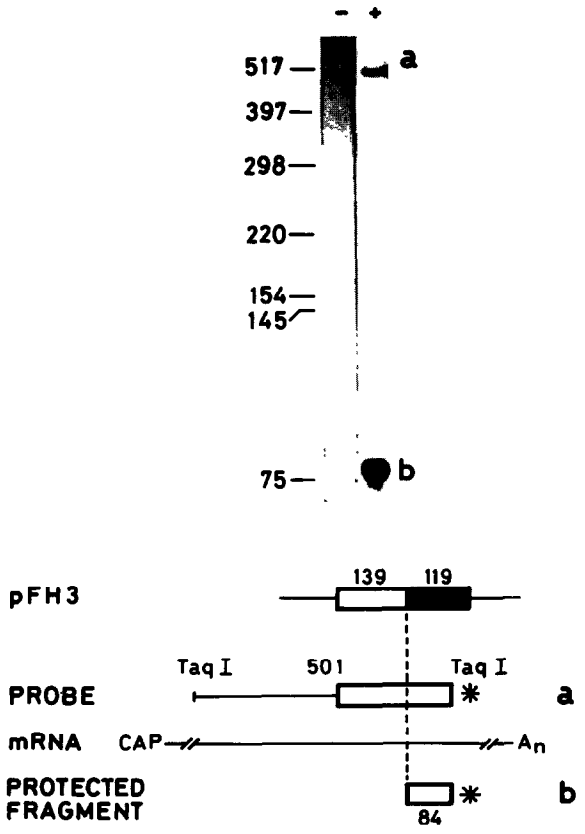

Fig.3. S_1 nuclease analysis of Hs578T cell total RNA/cDNA hybrids. The probe was the 501 bp *Taq*I fragment of pFH3 carrying 222 bp of insert sequences upstream of the labelled site (▭) plus 279 bp of vector sequences (——). + and − indicate presence and absence of Hs578T cell RNA in the hybridization mixture, respectively.

not seem to correspond to FN mRNA, as shown by the S_1 mapping experiment in fig.3. Only the downstream 84 bp of the probe remained S_1 resistant after hybridization with the same RNA used for the construction of the clones. Furthermore, the upstream 139 bp of clone pFH3 failed to hybridize to FN mRNA in 'Northern blots' (not shown). We then concluded that, as an artifact of cloning, a piece of DNA unrelated to FN co-ligated with FN cDNA in pFH3.

Fig.2 shows the amino acid sequence deduced from the nucleotide sequence of clone pFH2. A stretch of 31 amino acids preceding the N-terminal glutamine of hFN could be recognized as leader and pro-sequences. Although no protein data exist for comparison, these sequences must belong to FN based on the following evidence. First, there is a 35 bp overlap between pFH2 and pFH6. Second, the leader sequence present in the *Taq*I probe is totally protected in S_1 mapping (fig.3) and third, a nucleotide sequence identical to the one shown in fig.2 was read when sequencing directly single-stranded cDNA bands obtained by reverse transcription using the 19-mer oligonucleotide (section 2.2) as primer (not shown).

4. DISCUSSION

Fibronectin is an extracellular glycoprotein, and would thus be expected to have a signal peptide. Our data confirm this prediction but also show the existence of a pro-peptide. Amino acid 31 in fig.2 should be the carboxy end of a pro-peptide but not of a signal peptide: first, with the sole exception of apolipoprotein A1 [23], all reported pro-peptides

end in a pair of basic amino acids (in our case K at 30 and R at 31). Second, signal peptides normally end in amino acids with small, uncharged side chains [24]. Our data do not prove that the ATG coding for M no.1 in fig.2 is necessarily the initiation codon. However, the presence of an A and a C three and one nucleotides upstream from the ATG, respectively, is consistent with the consensus sequence reported by Kozak [25] around functional initiator codons. Assuming that this ATG is the initiator triplet, we applied the criteria described by Von Heijne [26] to calculate the 'processing probability', and predicted the point of cleavage for signal peptidase between amino acids 26 and 27, within a 'processing window' spanning amino acids 20–27. The FN signal therefore fits the consensus format in its length (26 residues) and hydrophobicity, including a charged residue within the first five amino acids (R at 3), followed by a core of hydrophobic residues (amino acids 9–14, 17, 20 and 21) and ending with a small, uncharged residue (A at 26). The subsequent predicted propeptide SKSKR is very hydrophilic and basic; it was not present in any other protein in the NBRF protein databank. We have previously demonstrated that FN does not undergo proteolytic processing at the C-terminus [4]. The existence of a predicted pro-peptide indicates that FN undergoes more post-translational proteolysis at its N-terminus, than signal elimination; the function of this novel charged domain in FN processing remains to be determined.

ACKNOWLEDGEMENTS

We wish to thank Dr F.E. Baralle for his valuable help in the oligonucleotide synthesis, and Judith Frydman for determining optimal hybridization conditions for the 27 bp *Hin*dIII-*Dde*I probe.

REFERENCES

[1] Yamada, K.M. (1983) Annu. Rev. Biochem. 52, 761–799.
[2] Furcht, L. (1983) Mod. Cell Biol. 1, 53–117.
[3] Hynes, R.O. (1985) Annu. Rev. Cell Biol. 1, 67–90.
[4] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1983) Proc. Natl. Acad. Sci. USA 80, 3218–3222.
[5] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1984) EMBO J. 3, 221–226.
[6] Kornblihtt, A.R., Vibe-Pedersen, K. and Baralle, F.E. (1984) Nucleic Acids Res. 12, 5853–5868.
[7] Vibe-Pedersen, K., Kornblihtt, A.R. and Baralle, F.E. (1984) EMBO J. 3, 2511–2516.
[8] Umezawa, K., Kornblihtt, A.R. and Baralle, F.E. (1985) FEBS Lett. 186, 31–34.
[9] Schwarzbauer, J.E., Tamkun, J.W., Lemischka, I.R. and Hynes, R.O. (1983) Cell 35, 421–431.
[10] Tamkun, J.K., Schwarzbauer, J.E. and Hynes, R.O. (1984) Proc. Natl. Acad. Sci. USA 81, 5140–5144.
[11] Kornblihtt, A.R., Umezawa, K., Vibe-Pedersen, K. and Baralle, F.E. (1985) EMBO J. 4, 1755–1759.
[12] Odermatt, E., Tamkun, J.W. and Hynes, R.O. (1985) Proc. Natl. Acad. Sci. USA 82, 6571–6575.
[13] Hackett, A.J., Smith, H.S., Springer, W.L., Owen, R.B., Nelson Rees, W.A., Riggs, J.L. and Gardner, M.B. (1977) J. Natl. Cancer Inst. 58, 1795–1800.
[14] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[16] 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.
[17] Gubler, V. and Hoffman, B.J. (1983) Gene 25, 263–269.
[18] Anson, D.S., Choo, K.H., Rees, D.J.G., Giannelli, F., Gould, K., Huddleston, J.A. and Brownlee, G.G. (1984) EMBO J. 3, 1053–1060.
[19] Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979) Nucleic Acids Res. 7, 2115–2136.
[20] Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[21] Garcia-Pardo, A., Pearlstein, E. and Frangione, B. (1983) J. Biol. Chem. 20, 12670–12674.
[22] Petersen, T., 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.
[23] Shoulders, C.C., Kornblihtt, A.R., Munro, B.S. and Baralle, F.E. (1983) Nucleic Acids Res. 9, 2827–2837.
[24] Watson, M.E. (1984) Nucleic Acids Res. 12, 5145–5164.
[25] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[26] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.