Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1876–1880, April 1987
Biochemistry

Cloning and analysis of the promoter region of the human fibronectin gene

(DNA sequencing/RNA mapping/signal peptide/propeptide)

DOUGLAS C. DEAN, CHRISTOPHER L. BOWLUS, AND SUZANNE BOURGEOIS*

Regulatory Biology Laboratory, The Salk Institute for Biological Studies, P.O. Box 85800, San Diego, CA 92138-9216

Communicated by Klaus Hofmann, December 8, 1986 (received for review, November 5, 1986)

ABSTRACT **Human fibronectin (FN) genomic clones were isolated by screening a human genomic library with a 75-base oligonucleotide. The sequence of the oligonucleotide corresponds to a region near the 5′ end of the human FN cDNA clone pFH6 that contains the amino-terminal coding sequences but does not extend to the 5′ end of the mRNA [Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* 4, 1755–1759]. The 5′ end of the FN gene is found on a 3.7-kilobase-pair *Eco*RI fragment that contains about 2.7 kilobase pairs of flanking sequence. The first exon is 414 base pairs long, with a 5′ untranslated region of 267 base pairs. As deduced on the basis of the position of the initiation codon, FN is synthesized with a 31-residue amino acid extension on the amino terminus that is not present in the mature polypeptide. This amino-terminal extension appears to contain both a signal peptide and a propeptide. The first 200 base pairs of 5′-flanking sequence is very G+C rich. Upstream of this the sequence becomes relatively A+T rich. The sequence ATATAA is found at −25 and the sequence CAAT is present at −150. The sequence GGGGCGGGGC at −102 exhibits homology to the binding site for the transcription factor SP1, and the sequence TGACGTCA at −173 exhibits homology to 5′-flanking sequences important for induction by cAMP.**

Fibronectin (FN) is a large glycoprotein found both in a soluble form in plasma and in an insoluble form that binds to the cell surface through a FN receptor (1–4). The two types of FN differ slightly in amino acid sequence, the result of alternative mRNA splicing (5, 6).

The FN protein can be visualized as a series of globular domains that can bind independently to a number of different molecules such as heparin, DNA, collagen, actin, and fibrin as well as to the cell surface. The elongated shape, dimeric structure, and diversified binding sites permit several simultaneous interactions and probably account for the role of FN in a variety of processes, including cell adhesion, wound healing, blood coagulation, cell differentiation and migration, maintenance of the cellular cytoskeleton, and tumor metastasis (reviewed in ref. 1).

FN expression is greatly inhibited upon neoplastic transformation, which may account at least partially for alterations in both morphology and adhesion of transformed cells (7, 8). In addition, FN synthesis is stimulated by glucocorticoid hormones, which may play a key role in the regulation of this multifunctional protein *in vivo* (9–13). It is therefore of interest to determine the structure of the FN gene promoter as a first step in the study of its regulation. Recently chicken, rat, bovine, and human FN cDNA and genomic clones have been isolated (14–22). Here we describe the isolation and characterization of genomic clones containing the 5′ end and flanking sequence of the human FN gene.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Cell Culture. The human fibrosarcoma cell line HT1080 and human lung fibroblasts (WI38) were grown as described (12). Dexamethasone was added at 10^{-7} M where indicated and cells were incubated for 48 hr before isolation of RNA.

RNA Extraction and Mapping. Total RNA was extracted from confluent cell monolayers by the guanidinium thiocyanate method (23). Primer extensions and nuclease S1 mapping were done as described previously (24). Hybridizations for primer extensions were done in 70% (vol/vol) formamide/0.3 M NaCl at 42°C for 16 hr, using 5′-^{32}P-labeled primers and 50 or 75 μg of total RNA. Hybridizations for nuclease S1 mapping were at 52°C for 16 hr in 70% formamide/0.3 M NaCl with 50 or 75 μg of total RNA. Probes for nuclease S1 analysis were labeled with ^{32}P on the 5′ end by using polynucleotide kinase or on the 3′ end by using T4 polymerase (23).

Screening of the Library and DNA Characterization. Human spleen partial *Mbo* I Charon 28 library 10165 was kindly provided by Glen Evans (Salk Institute). Oligonucleotides were synthesized by Callie Mack (Salk Institute). Screening was done on nylon membranes (Amersham) with a 5′-^{32}P-labeled 75-base oligonucleotide corresponding to a region near the 5′ end of the human FN cDNA clone pFH6 (6) (see Fig. 5 for the sequence). The DNA sequence of pFH6 was kindly provided by Alberto Kornblihtt (INGEBI, Buenos Aires). Prehybridization was at 37°C for 24 hr in a solution containing 50% (vol/vol) formamide, 5× SSPE (1× SSPE is 0.18 M NaCl/1 mM EDTA/10 mM NaH_2PO_4, pH 7.4), 8× Denhardt's solution (1× Denhardt's solution is 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), 0.1% sodium deodecyl sulfate, and denatured herring sperm DNA at 0.4 mg/ml. Hybridization was done under similar conditions with probe at 10^6 cpm/ml for 24 hr. Filters were washed at 52°C in 0.5× SSPE for 1 hr.

DNA fragments were subcloned in the plasmid pGEM-4 (Promega Biotec, Madison, WI) for restriction enzyme mapping and DNA sequencing. DNA sequencing was done by using both the chemical (25) and dideoxy (26) methods. For dideoxy sequencing, subclones in the vector pGEM-4 were denatured with alkali and the sequence was obtained by using primers from the T7 and SP6 promoters in pGEM-4 or from chemically synthesized oligonucleotides (27).

RESULTS AND DISCUSSION

Human FN cDNA clones covering the entire coding region have been isolated (6). The subclone pFH6 contains the 5′-most position of the FN coding sequence but does not extend to the 5′ end of the mRNA (6). The deduced amino acid sequence from this clone was recently published, and its nucleotide sequence is available through the European Molecular Biology Laboratory data bank (6). An oligonucleotide

Abbreviations: FN, fibronectin; IVS, intervening sequence.

*To whom reprint requests should be addressed.

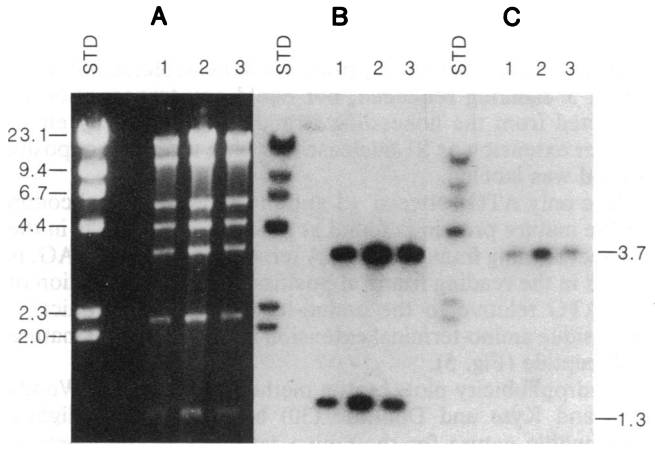

FIG. 1. Southern analysis of genomic clones. (*A*) *Eco*RI digests of human genomic clones λgHF23-3 (lane 1), λgHF26 (lane 2), and λgHF41 (lane 3) separated on a 1% agarose gel in the presence of ethidium bromide. (*B*) The gel was dried and the ^{32}P-labeled 75-mer used for screening was hybridized as described previously (24). (*C*) The gel was washed free of probe and hybridized to a ^{32}P-labeled 32-mer that contains only the 5′ portion of the 75-mer (Fig. 5). The size standard (STD) is *Hin*dIII-digested phage λ DNA, and fragment sizes are given in kb.

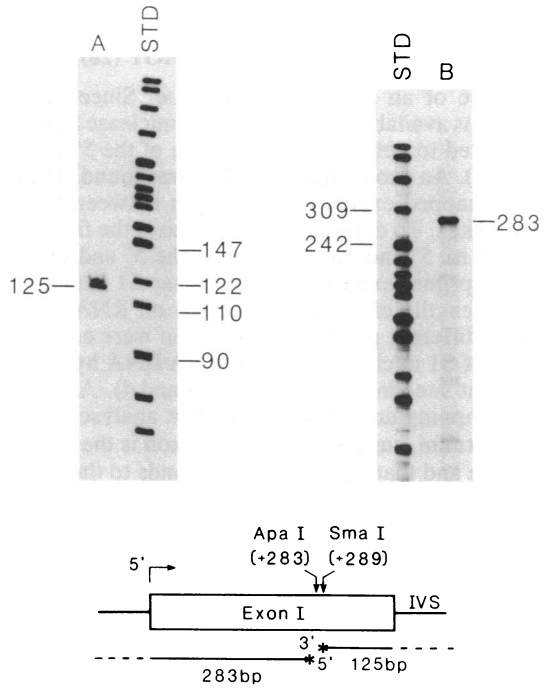

FIG. 3. Nuclease S1 mapping of the 5′ and 3′ ends of the first exon of the FN gene. For mapping of the 5′ end of the exon (*B*), the noncoding strand was labeled on the 5′ end at the *Apa* I site. The 3′ end of the noncoding strand of the *Sma* I site was labeled for analysis of the 3′ end of the exon (*A*). Total RNA (50 μg) from dexamethasone-induced HT1080 cells was used for mapping. The size standard (STD) is pBR322 digested with *Msp* I, and sizes are given in base pairs (bp).

of 75 bases (75-mer) was synthesized that corresponds to sequences near the 5′ end of pFH6 (see Fig. 5). The 75-mer was then used to screen a human genomic λ phage library. The library, which originally contained 2×10^6 members, was amplified prior to screening. Approximately 1×10^6 phage plaques were screened and 6 positive clones were identified. All of the clones appeared to be identical by restriction enzyme mapping.

An *Eco*RI digest of three of the genomic clones is shown in Fig. 1*A*. The clones contain about 20 kilobases (kb) of human DNA. When the gel was probed with the 75-mer used in the library screening, two bands, 3.7 and 1.3 kb, hybridized, suggesting that in the genomic clones the 75-mer is interrupted by an intervening sequence (IVS) (Fig. 1*B*). The same gel was then probed with a 32-base oligonucleotide (32-mer) corresponding to only the 5′ portion of the 75-mer (Fig. 5). This resulted in only the 3.7-kb band hybridizing (Fig. 1*C*), which indicated that the 3.7-kb *Eco*RI band was 5′ to the 1.3-kb *Eco*RI band. The 3.7-kb *Eco*RI fragment was then subcloned (clone pgHF3.7) and its restriction enzyme map is presented in Fig. 2.

A portion of the 3.7-kb *Eco*RI fragment from clone λgHF41 containing the transcriptional start site was sequenced. The sequencing strategy is shown in Fig. 2. Initially two complementary 32-mers corresponding to the 5′ end of pFH6 were used to sequence in both directions. The sequence in the 3′ direction was identical to that found in pFH6 up to the point of the intervening sequence junction, indicating that the genomic clones contain human FN DNA sequences (Figs. 2 and 5). In addition to this, DNA sequences in the genomic clones are transcribed and the level of RNA is regulated by dexamethasone in a fashion similar to that found for FN (Figs. 3 and 4) and the sizes of the transcripts are also similar to those of FN (data not shown).

It was then possible to define the 3′ end of the exon by the divergence of cDNA and genomic sequences. The sequence

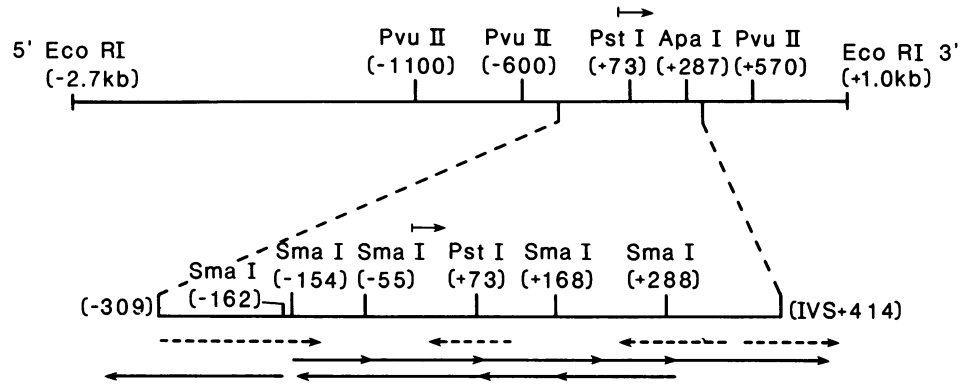

FIG. 2. Restriction enzyme map of the insert subclone pgHF3.7, which contains the 5′ end of the FN gene. The top portion of the figure shows a restriction enzyme map of the 3.7-kb *Eco*RI fragment from Fig. 1. The lower part of the figure shows the portion of the fragment that was sequenced. The restriction enzymes that are shown were used for subcloning. The DNA sequencing strategy is shown below. Solid lines represent DNA sequence obtained from subclones, while the broken lines indicate sequence obtained by using synthetic oligonucleotide primers. The arrows indicate the start site of transcription, and the nucleotide numbers in parentheses are based on this site.

at this splice junction AAG/CGTGAGT is very similar to the consensus donor sequence $\overset{A}{C}AG/GT\overset{A}{G}AGT$ (28) except for the presence of an additional C residue. Since no cDNA sequence was available 5′ of the 32-mer, nuclease S1 mapping was employed to determine the position of the 5′ end of the exon (Fig. 3). An exon length of 414 bp was found. The entire exon was mapped except for the region between the *Apa* I and *Sma* I sites. To determine if this exon is the first one and therefore if its 5′ end corresponds to the 5′ end of the FN RNA, both primer extension and nuclease S1 mapping were used. The lengths of primers extended on RNA templates from three different positions in the exon were compared to the length of S1 nuclease-resistant RNA·DNA hybrids from two different sites in the exon (Figs. 3 and 4). All of the S1 nuclease mapping and primer extension analyses indicated similar 5′ termini, suggesting that this exon is the first one in the FN gene and that its 5′ end corresponds to the transcriptional start site. This site corresponds to an A (see Fig. 4*E*) and is indicated as +1 in the DNA sequence (Fig. 5). In all the RNA mapping shown, RNA from the human fibrosarcoma cell line HT1080 was used; however, an identical transcriptional start site was observed in human lung fibroblasts (WI38) (data not shown).

The 3.7-kb *Eco*RI fragment then contains about 2.7 kb of 5′-flanking sequence and about 1.0 kb of downstream sequence (Fig. 2). No transcription could be detected in the 2.7 kb of 5′-flanking sequence, nor could any transcription be detected from the noncoding strand in exon one by either primer extension or S1 nuclease mapping when the opposite strand was labeled.

The only ATG between +1 and the amino-terminal codon of the mature protein is found at position +268 and is in the correct reading frame (Fig. 5). A termination codon, TAG, is found in the reading frame at position +178. The position of the ATG relative to the amino-terminal codon predicts a 31-residue amino-terminal extension not found on the mature polypeptide (Fig. 5).

Hydrophobicity plots by the method of Hopp and Woods (29) and Kyte and Doolittle (30) both indicate a slightly hydrophilic nature for the amino terminus of this peptide, which extends into the Gly-Pro-rich region. From this point the sequence, which extends to the Ser-Thr residues at +334, becomes very hydrophobic. The carboxyl terminus of the peptide is very hydrophilic. The high degree of hydrophobicity and the high leucine content near the middle of the peptide seem to be indicative of a signal peptide; however, the presence of the sequence Gly-Pro-Gly-Pro-Gly preceding

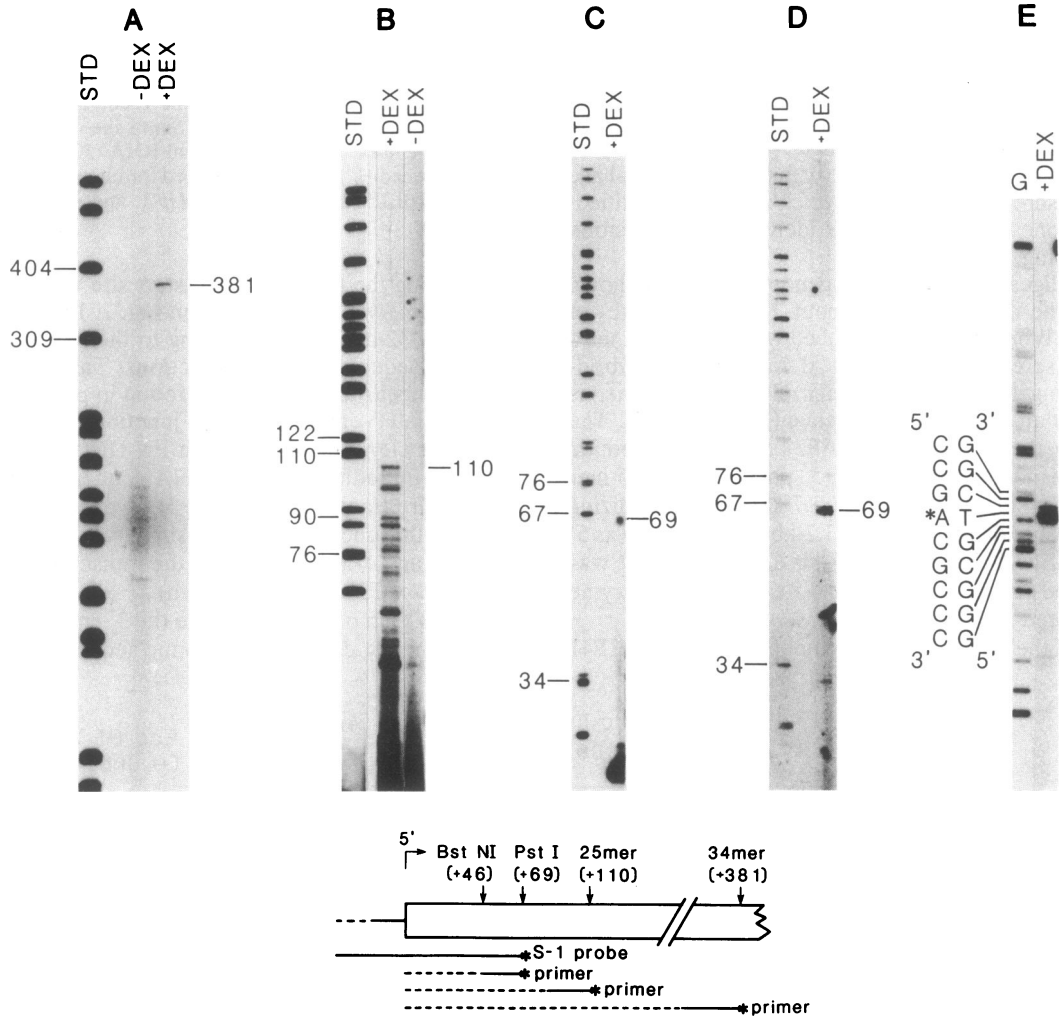

FIG. 4. Determination of the site of transcriptional initiation of the human FN gene. *A*, *B*, and *C* represent primer extension analyses using 5′-end-labeled primers from different points in the exon. The location of the primers is shown schematically at the bottom of the figure. The 34-mer is identical to the 32-mer (Fig. 5) except that GC was added to the 5′ end to assist in subcloning. *D* and *E* are nuclease S1 mapping experiments using the 5′-end-labeled probe indicated at the bottom of the figure. In lanes labeled +DEX, 75 μg of total RNA from HT1080 cells incubated for 48 hr in the presence of dexamethasone was used. Equal amounts of RNA were used in experiments without dexamethasone. In *E*, the lane labeled G is a chemical sequencing G reaction on the 5′-end-labeled nuclease S1 probe. The asterisk on the A residue indicates that it is the transcriptional start site. The size standard (STD) is *Msp* I-digested pBR322.

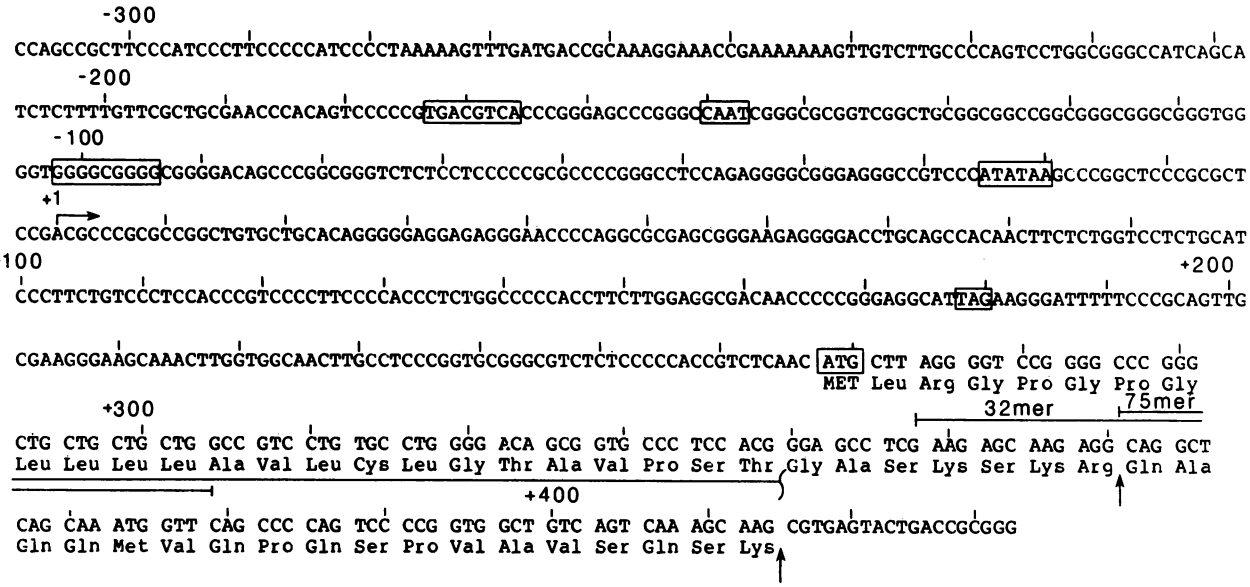

FIG. 5. DNA sequence of the first exon and 5′ flanking region of the human FN gene. The A designated as +1 is the start site of transcription. The arrow above residue +1 shows the direction of transcription. The Gln to the right of the first vertical arrow is the amino terminus of the mature polypeptide. The arrow after the Lys residue marks the 3′ end of the first exon. Sequence of the 75-mer that was used to probe the genomic library is marked with a bar. Sequence identity with the 75-mer diverges at the IVS junction. DNA sequence of the 32-mer is shown below the bar. The 5′ end of the 32-mer corresponds to the 5′ end of the human FN cDNA clone pFH6 (6). The genomic sequence is the same as the cDNA sequence of pFH6 to the point of the 3′ end of the exon. See text for discussion of sequences in boxes.

the hydrophobic region seems unusual (31). The presence of Lys-Arg immediately upstream of the amino-terminal Gln of the mature protein is inconsistent with a signal peptide cleavage site (31); however, such a cleavage site, like that cleaved by trypsin, is the processing point in a number of propeptides (32). It therefore seems likely that FN is synthesized as a precursor that contains both a propeptide and a signal peptide. Potential signal peptide cleavage sites can be found between 5 and 15 amino acids upstream of the mature amino-terminal Gln (+361).

The 5′-flanking region contains the sequence ATATAA at −25 and the sequence CAAT at −150. The sequence GGGGCGGGGC at −102 corresponds to sequences identified as binding sites for the transcription factor SP1 (33). The sequence TGACGTCA at −173 corresponds to 5′-flanking sequences necessary for cAMP induction in the phospho*enol*pyruvate carboxykinase gene (34), the somatostatin gene (35), and the α subunit of the chorionic gonadotropin gene (P. Mellon, personal communication). Recently we have found that synthesis of FN in HT1080 cells is stimulated by cAMP (unpublished result). Several regions of the sequence show at least a limited homology to a glucocorticoid regulatory element (−206, +218, and +302) (36).

The 5′-flanking sequence is very G+C rich up to about −200, where the sequence becomes relatively A+T rich. This G+C-rich nature as well as some sequence homology is shared by the 5′-flanking region of the collagen α_1 (I) gene, which is expressed in similar tissues and is also sensitive to cellular transformation (37).

Knowledge of the sequence surrounding the FN transcriptional start site should prove to be valuable in the analysis of how this gene is regulated.

Note Added in Proof. After our manuscript was submitted, we learned of a paper by Gutman *et al.* (38) showing cDNA sequence of the region coding for the human fibronectin prepropeptide. Our sequence is in agreement with theirs except for one discrepancy at position +311. Here we find an A and they find a T.

We thank Alberto Kornblihtt for the sequence of pFH6, Glen Evans for the human genomic library, Pamela Mellon for allowing us to cite unpublished work, Harvey Lodish for helpful discussions, Allen Koop and Ron Newby for constructive comments, and Gloria Laky Swart for preparing the manuscript. This work was supported by grants (to S.B.) from the National Science Foundation (DCB-8417395), the National Institutes of Health (GM35751), the Elsa U. Pardee Foundation, and the Gustavus and Louise Pfeiffer Foundation.

1. Yamada, K. M. (1983) *Annu. Rev. Biochem.* **52,** 761–799.
2. Ruoslahti, E., Engvall, E. & Hayman, E. G. (1981) *Coll. Res.* **1,** 95–128.
3. Hynes, R. O. & Yamada, K. M. (1982) *J. Cell Biol.* **95,** 369–377.
4. Pytela, R., Pierschbacher, M. D. & Ruoslahti, E. (1985) *Cell* **40,** 191–198.
5. Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R. & Hynes, R. O. (1983) *Cell* **35,** 421–431.
6. Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* **4,** 1755–1759.
7. Yamada, K., Yamada, S. S. & Pastan, I. (1977) *J. Cell Biol.* **74,** 649–654.
8. Tyagi, J. S., Hirano, H., Merlino, G. T. & Pastan, I. (1983) *J. Biol. Chem.* **258,** 5787–5793.
9. Furcht, L. T., Mosher, D. F., Wendelschafer-Crabb, G., Woodbridge, P. A. & Foidart, J. M. (1979) *Nature (London)* **277,** 392–395.
10. Marceau, N., Goyette, R., Valet, J. P. & Deschenes, J. (1980) *Exp. Cell Res.* **125,** 497–502.
11. Baumann, H. & Eldredge, D. (1982) *J. Cell Biol.* **95,** 29–40.
12. Oliver, N., Newby, R., Furcht, L. T. & Bourgeois, S. (1983) *Cell* **33,** 287–296.
13. Lien, Y.-H., Wong, M. J., Golbus, M. S. & Stern, R. (1984) *J. Cell Physiol.* **120,** 103–107.
14. Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3218–3222.
15. Tamkun, J. W., Schwarzbauer, J. E. & Hynes, R. O. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5140–5144.
16. Fagan, J. B., Pastan, I. & de Crombrugghe, B. (1980) *Nucleic Acids Res.* **8,** 3055–3064.
17. Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1984) *EMBO J.* **3,** 221–226.
18. Vibe-Pedersen, K., Kornblihtt, A. R. & Baralle, F. E. (1984) *EMBO J.* **3,** 2511–2516.
19. Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1984) *Nucleic Acids Res.* **12,** 5853–5868.

20. Fagan, J. B., Sobel, M. E., Yamada, K. M., de Crombrugghe, B. & Pastan, I. (1981) *J. Biol. Chem.* **256,** 520–525.
21. Hirano, H., Yamada, Y., Sullivan, M., de Crombrugghe, B., Pastan, I. & Yamada, K. D. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 46–50.
22. Oldberg, A., Linney, E. & Ruoslahti, E. (1983) *J. Biol. Chem.* **258,** 10193–10196.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 197.
24. Knoll, B. J., Zarucki-Schultz, T., Dean, D. C. & O'Malley, B. W. (1983) *Nucleic Acids Res.* **11,** 6733–6754.
25. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
27. Promega Biotec (1986) *Technical Manual* (Promega Biotec, Madison, WI), pp. 7–10.
28. Cech, T. R. (1983) *Cell* **34,** 713–716.
29. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3824–3828.
30. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
31. Von Heijne, G. (1982) *Eur. J. Biochem.* **133,** 17–21.
32. Steiner, D. F., Kemmler, W., Tager, H. S. & Peterson, J. D. (1974) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **33,** 2105–2115.
33. Dynan, W. J. & Tjian, R. (1985) *Nature (London)* **316,** 775–778.
34. Short, J. M., Wynshaw-Boris, A., Short, H. P. & Hanson, R. W. (1986) *J. Biol. Chem.* **261,** 9721–9726.
35. Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6682–6686.
36. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. J., Krauter, P., Westphal, H. M. & Beato, M. (1984) *Nature (London)* **308,** 513–519.
37. Chu, M.-L., deWet, W., Bernard, M. & Ramirez, F. (1985) *J. Biol. Chem.* **260,** 2315–2320.
38. Gutman, A., Yamada, K. M. & Kornblihtt, A. (1986) *FEBS Lett.* **207,** 145–148.