Biochem. J. (1987) 241, 923–928 (Printed in Great Britain)

Primary structure of human plasma fibronectin

Characterization of a 38 kDa domain containing the *C*-terminal heparin-binding site (Hep III site) and a region of molecular heterogeneity

Angeles GARCIA-PARDO,*‡ Agueda ROSTAGNO* and Blas FRANGIONE*†

*Department of Pathology and †Kaplan Cancer Center, New York University Medical Center, 550 First Avenue, Rm. H-427, New York, NY 10016, U.S.A.

The primary structure of a 38 kDa heparin-binding domain from human plasma fibronectin has been determined. This domain contains 380 residues arranged in three type-III homology regions of approx. 90 residues each, and a 67-amino-acid *C*-terminal segment. This segment has been shown to be encoded by certain mRNA species only, due to alternative splicing [Kornblihtt, Vibe-Pedersen & Baralle (1984) Nucleic Acids Research **12**, 5853–5868], and therefore represents a region of heterogeneity in fibronectin. Our data indicate that at least one of the constituent polypeptide chains contains this region.

INTRODUCTION

The glycoprotein fibronectin is present in plasma and other body fluids, cell surfaces and extracellular matrices. Major functions of this protein include a role in cellular adhesion and in the functioning of the reticuloendothelial system. Fibronectin interacts with other macromolecules by means of specific binding sites located within structural domains of the molecule (reviewed by Pearlstein *et al.*, 1980; Furcht, 1983; Yamada, 1983; Mosher, 1984).

The interaction of fibronectin with heparin appears to take place through multiple sites which differ from each other in affinity (Gold *et al.*, 1983) and bivalent-cation-dependency (Hayashi & Yamada, 1982). So far three of these sites have been described and isolated as part of proteolytic fragments obtained after enzymic digestions of fibronectin (Pande & Shively, 1982; Smith & Furcht, 1982; Hayashi & Yamada, 1983; Garcia-Pardo *et al.*, 1983; Gold *et al.*, 1983; Calaycay *et al.*, 1985). The complete primary structure of two of these sites (Hep I and Hep II) has been recently determined (Pande & Shively, 1982; Garcia-Pardo *et al.*, 1983; Gold *et al.*, 1983; Pande *et al.*, 1985; Calaycay *et al.*, 1985). Only a partial amino acid sequence (68 residues) has been reported for the third heparin-binding site (Hep III) (Gold *et al.*, 1983). This paper is concerned with this site.

The Hep III site has been mapped by several investigators towards the *C*-terminal region of fibronectin (Smith & Furcht, 1982; Hayashi & Yamada, 1983; Sekiguchi & Hakomori, 1983). Studies performed with cDNA clones isolated from human (Kornblihtt *et al.*, 1984, 1985) and rat (Schwarzbauer *et al.*, 1983) cells have demonstrated the existence of several mRNAs for fibronectin. These data, together with previous reports on enzymic cleavage of this molecule (Hayashi & Yamada, 1983; Sekiguchi & Hakomori, 1983) have indicated the presence of heterogeneity between the two subunits of fibronectin. In plasma fibronectin this variability appears to be restricted to the *C*-terminal region, in particular to the segment located between the Hep III site and the second fibrin-binding domain (Kornblihtt *et al.*, 1984).

To determine how mRNA heterogeneity is translated into the mature protein, studies were carried out on the primary structure of a fragment containing the Hep III site. We and others have previously reported that, after fragmentation of fibronectin with trypsin, five major fragments are liberated of M_r 200000, 180000, 31000, 29000 and 6000 (Smith & Furcht, 1982; Garcia-Pardo *et al.*, 1983; Sekiguchi & Hakomori, 1983). The complete amino acid sequence of the 31, 29 and 6 kDa fragments has already been determined (Garcia-Pardo *et al.*, 1983, 1984, 1985). In the present paper the 200/180 kDa fragment mixture was further digested with trypsin to produce a 38 kDa fragment which contains the Hep III site. On the basis of the cDNA sequence (Kornblihtt *et al.*, 1984), this fragment immediately precedes the 31 kDa domain and extends into the region of heterogeneity. We present here its nearly entire amino acid sequence (20 residues were not sequenced) and show that plasma fibronectin contains the amino acid segment encoded by one of the mRNAs. Whether this segment is present in both polypeptide chains remains to be elucidated.

MATERIALS AND METHODS

Isolation and limited tryptic digestion of fibronectin were performed essentially as described (Garcia-Pardo *et al.*, 1983, 1985).

Purification of the 38 kDa fragment

Tryptic digests of fibronectin were applied to a gelatin–Sepharose column equilibrated in 25 mM-Tris/HCl/50 mM-NaCl/0.5 mM-Na_2EDTA buffer, pH 7.6 (25 mM-Tris buffer), at a ratio of 2 mg of protein to 1 ml of swollen gel; bound fragments, namely those of M_r

Abbreviations used: PAGE, polyacrylamide-gel electrophoresis; SV.8, *Staphylococcus aureus* V8 proteinase; PMSF, phenylmethanesulphonyl fluoride.

‡ Present address and address for correspondence: The Lindsley F. Kimball Research Institute of the New York Blood Center, New York, NY 10021, U.S.A.

200000 and 180000, were eluted with the same buffer containing 4 M-urea and applied to a Sephacryl S-200 column in 25 mM-Tris for further purification. Material eluted in the first peak of this column was digested with trypsin [1:100 (w/w) at 37 °C for 45 min] and, after termination of the digestion with 1 mM-PMSF, the resulting fragments were applied to a heparin–Sepharose affinity column (4–5 mg of protein/ml of swollen gel). The column was washed with 25 mM-Tris buffer/1 mM-PMSF to remove unbound fragments, followed by 25 mM-Tris/0.1 M-NaCl buffer. Fragments that bound strongly to heparin were eluted with 25 mM-Tris/0.5 M-NaCl buffer, dialysed against 10 mM-Tris, pH 7.0, and stored at −20 °C.

The heparin-binding fraction was further resolved by DEAE-cellulose chromatography; if necessary, solid urea to a final concentration of 4 M was added to the protein solution in order to dissolve any precipitate formed during the dialysis. This fraction was applied to a DEAE-cellulose column (Whatman DE-52) equilibrated in 10 mM-Tris/4 M-urea, pH 7.0 (10 mg of protein/ml of matrix). The unbound material containing the 38 kDa fragment was recovered in the 10 mM-Tris/4 M-urea buffer, dialysed against water and freeze-dried. When required, further purification of the 38 kDa fragment was achieved by reverse-phase h.p.l.c. (Waters Associates instrument) on a μBondapak C_{18} column. The elution system consisted of two solvents: A, aq. 0.05% trifluoroacetic acid; B, trifluoroacetic acid/acetonitrile/water (1:1320:679, by vol.). Chromatography was performed in a stepwise manner with increasing concentrations of solvent B.

Citraconylation of the 38 kDa fragment

Lysine residues of the 38 kDa domain were blocked with citraconic anhydride (Pierce Chemical Co.) exactly as described by Garcia-Pardo *et al.* (1978).

Enzymic digestions

Trypsin digestion of the 38 kDa fragment was performed in 0.2 M-NH_4HCO_3, pH 8.5, at an enzyme/protein ratio of 1:100 (w/w) and 37 °C for 2 h; the digestion was stopped by freezing followed by freeze-drying. Citraconylated 38 kDa fragment was digested with trypsin in the same buffer, but only for 1 h. After this time, the pH was adjusted to 2.0 with drops of HCl and the digestion mixture was left at 4 °C overnight. It was subsequently frozen and freeze-dried. Peptides T1 and CT1 were digested with SV.8 (Miles Laboratories) in 0.1 M-NH_4HCO_3/2 mM Na_2EDTA, pH 7.8, at a 1:50 enzyme/peptide ratio and 37 °C for 2 and 24 h respectively and freeze-dried.

Purification of tryptic and SV.8-derived peptides

Peptides from intact and citraconylated 38 kDa domain were separated by h.p.l.c. on a μBondapak C_{18} column and with the A and B solvents described above. After the column had been washed with solvent A for 10 min, a linear gradient from 0 to 100% solvent B was applied over a 1 h period. Peptides were detected by absorbance at 214 nm.

Cleavage of the 38 kDa Fragment with CNBr

The 38 kDa domain was dissolved in 70% (v/v) formic acid and treated with a 5-fold excess of CNBr for 24 h at room temperature. The sample was diluted with distilled water before being freeze-dried. CNBr-produced fragments were separated by h.p.l.c. under the conditions described above.

Other methods

SDS/polyacrylamide-gel electrophoresis, amino acid analysis and automated amino acid sequence determination were performed essentially as described (Garcia-Pardo *et al.*, 1983, 1985); microsequence analyses were performed on an Applied Biosystems Model 470A gas-phase sequencer equipped with a model 120A phenylthiohydantoin analyser.

RESULTS AND DISCUSSION

Isolation of the 38 kDa domain from tryptic digests of fibronectin

Trypsin digestion of the 200/180 kDa-fragment mixture resulted in four major fragments with heparin-binding properties (Fig. 1, lane 3): two large fragments of 145 and 120 kDa, a 58 kDa and a 38 kDa fragment. These four fragments therefore contained a high-affinity heparin-binding site (Hep III site); their relationship will not be discussed further here. Highly purified 38 kDa domain could be obtained after DEAE-cellulose chromatography and h.p.l.c. of this fraction (Fig. 1, lanes 4 and 5).

Primary structure of the 38 kDa domain

Initial characterization of this fragment included its amino acid composition and *N*-terminal amino acid sequence determination. As shown in Table 1, the 38 kDa fragment contains no half-cystine or free thiol groups and is characterized by an increased number of proline and threonine residues (11 and 13% respectively). Automated sequence analysis on 23 nmol of this fragment identified the first 28 residues (Fig. 2). The remaining sequence was obtained by a combination of enzymic digestions with trypsin (before and after citraconylation) and SV.8 as well as chemical cleavage with CNBr.

Trypsin digestion of 84 nmol of 38 kDa fragment resulted in 27 peptides, which were separated by h.p.l.c. Their amino acid composition and amino acid sequence were determined and are shown in Table 1 and Fig. 2. Only the first 20 residues of peptide T1 (positions 1–51) were identified by this approach; further digestion of T1 with SV.8 rendered peptide SV.8-1, which corresponded to residues 20–32 (Fig. 2). Characterization of other peaks from this digest or from SV.8-digested CT1 (positions 1–75, see below) did not identify residues 33–51; since we have previously reported the sequence of residues 40–73 as part of a subtilisin fragment of fibronectin (SH_{fr} in Fig. 2) (Gold *et al.*, 1983), the only segment of T1 not yet sequenced extends from positions 33 to 39. The presence of these seven amino acids, however, is inferred from the composition of peptide T1 (Table 1). A correction has been made for positions 66, 108 and 124, which were originally reported as tyrosine, glycine and valine respectively due to h.p.l.c. artefacts.

Peptide T11 and T19 with identical composition (Ala-Arg) as well as the free arginine residue at position 141 (T10) were recovered in the second peak of the chromatogram. The amino acid composition of this peak was $Ala_{1.1}Arg_{1.7}$; the sequence and location of the 38 kDa fragment was deduced from trypsin specificity

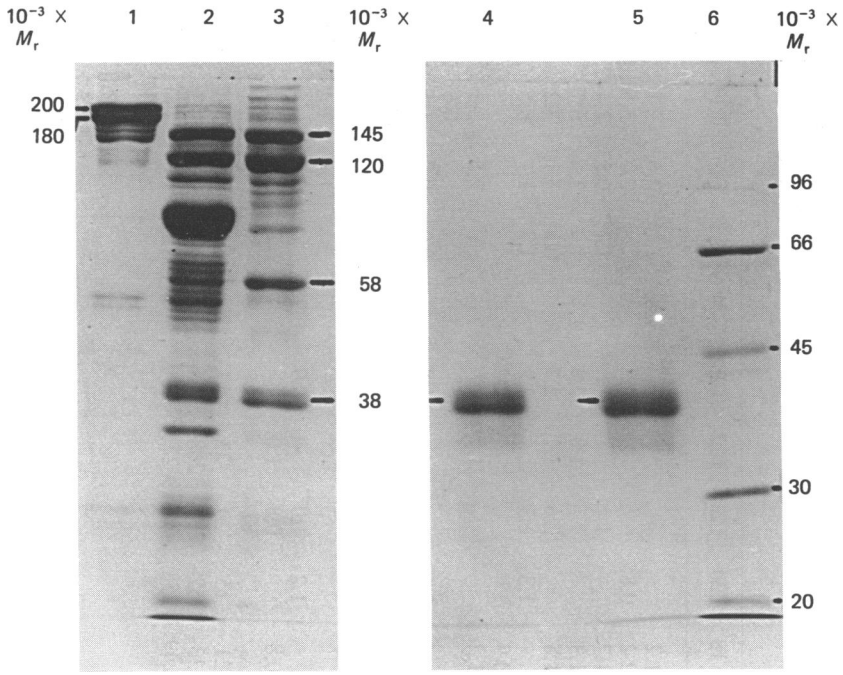

Fig. 1. SDS/10%-(w/v)-polyacrylamide-gel electrophoresis of the 200/180 kDa fragment mixture and its tryptic products

Lane 1, starting 200/180 kDa mixture; lane 2, whole tryptic digest (enzyme/substrate, 1:100; 45 min, 37 °C); lane 3, fragments that bound to heparin–Sepharose and were eluted with 0.5 M-NaCl; lanes 4 and 5, purified 38 kDa fragment before (lane 4) and after (lane 5) reduction; lane 6, M_r markers consisting of phosphorylase *b* (96000), bovine serum albumin (66000), ovalbumin (45000), carbonic anhydrase (30000), soybean trypsin inhibitor (20000) and α-lactalbumin (14000).

and comparison with the sequence deduced from cDNA clones (Kornblihtt *et al.*, 1984).

Peptides T26 (residues 311–341) and T27 (residues 342–380) were recovered in low yield and could not be completely sequenced. Since these peptides belong to a region that could be variable in length (or not present at all), owing to alternative splicing mechanisms (Schwarzbauer *et al.*, 1983; Kornblihtt *et al.*, 1984), the citraconylated peptide CT10 covering residue 311–380 was produced. CT10 had the following amino acid composition: $Asp_{4.7}Thr_{6.5}Ser_{3.2}Glu_{10.4}Pro_{8.3}Gly_{7.9}Val_{4.8}Met_{0.9}Ile_{2.4}Leu_{6.2}Tyr_{1.1}Phe_{2.5}His_{3.7}Lys_{2.8}Arg_{1.2}$. Sequence determination by microsequencing methods gave the first 31 residues of this fragment. Attempts to digest CT10 with trypsin (after deblocking the lysine residues), chymotrypsin and SV.8 were unsuccessful in producing a peptide comprising the *C*-terminal region of CT10; however, the presence of residues 360–372 (shown in parentheses in Fig. 2) was inferred from the composition of peptides T27 and CT10. Peptide CT10 could be cleaved with CNBr to release the *C*-terminal peptide CB-2 (residues 373–380), whose amino acid sequence was determined. The ordering of all peptides in the 38 kDa domain was obtained by comparison with the sequence deduced from human (Kornblihtt *et al.*, 1984, 1985) and rat (Schwarzbauer *et al.*, 1983) cDNA clones and by overlapping relevant citraconylated peptides (results not shown).

The amino acid sequence and composition of peptide CT10 showed that, following the first three residues KKT (one-letter amino acid notation), which belong to the third type-III homology region (see below), the segment contains 67 amino acids until the last residue at position 380. However, the cDNA clone that contains the nucleotide stretch (known as the 'III CS region') coding for peptide CT10 (minus the first three residues) consists of 267 base-pairs, which would translate into 89 amino acids (Kornblihtt *et al.*, 1984). We have recently shown (Garcia-Pardo *et al.*, 1985) that the last ten amino acids of this region constitute the first ten residues of the 31 kDa fragment. The remaining 12 amino acids are probably lost as a result of a secondary cleavage during the initial trypsin digestion of the 200/180 kDa fragments. To prove this, a CNBr cleavage of 200 nmol of 38 kDa fragment was carried out and peptide CB-2 was obtained in good yield. Sequence determination of this peptide confirmed that it corresponded to residues 373–380. Therefore the 38 kDa domain isolated in this study ends at this position.

A second CNBr fragment isolated from this digestion was CB-1, which, by amino acid sequence, was shown to correspond to positions 65–87 (Fig. 2). The production of CB-1 was due to a cleavage by CNBr at tryptophan residue 64 present in T2; cleavage at tryptophan residues by CNBr has been previously reported by Ozols & Craig (1977).

Examination of the primary structure of the 38 kDa domain (Fig. 2) shows that it is composed of three type-III homology regions (approx. 90 amino acids each) (Petersen *et al.*, 1983) followed by the 67-residue *C*-terminal segment. Fragments containing the Hep II site are also located in type-III homology regions and are positively charged (Pande *et al.*, 1985; Calaycay *et al.*, 1985). The net charge of the 38 kDa fragment was +2, and, when the three homology regions were considered individually, their charges were +3, +2 and +6

Table 1. Amino acid composition of the 38 kDa fragment (38 kDa F) and its tryptic peptides T1–T27

Amino acid	Composition (residues/molecule)																					
	38 kDa F	T1	T2	T4	T6	T7	T8	T9	T12	T13	T14	T15	T16	T17	T18	T20	T21	T22	T23	T24	T25	T26	T27
Asp	23.4	3.0	1.1			2.2		1.9	1.3	2.2	1.3	0.8	3.2	3.0	1.2					1.8		2.6	2.1
Thr	41.7	7.6	5.2	0.9	1.0	1.3		3.7	4.5	5.0	1.1	2.7	1.3	1.2	2.2	0.8			4.9			3.0	3.5
Ser	23.7	3.2	1.8			2.4	1.2	2.4	1.4			1.0		3.5	1.7		1.1				1.1	0.8	2.4
Glu	32.4	7.5	3.1			1.2	1.3	2.4	1.3	4.1		1.0			1.3		1.1	1.0	2.9	1.1	1.3	3.6	6.7
Pro	40.3	3.9	2.7	0.9	0.7	0.9		2.5		1.6	1.0	0.9		2.1	1.8		2.6	1.2	2.7		0.8	5.1	3.2
Gly	26.2	3.0	1.2		0.8	1.3		1.1		2.4		1.8				1.3	1.3		3.2		1.3	1.4	6.5
Ala	21.5	3.4	1.0			1.7	1.0	1.0	1.1	2.1			1.0	2.8	0.9				2.2				
Val	32.6	6.4	1.8	0.9		3.6	1.8	1.9		2.1	0.8			1.7	1.2			1.7				2.5	2.3
Met	3.2	0.5			0.9	0.7																	0.6
Ile	18.6	1.6				0.9			1.6	2.0	1.1	0.6	1.0	1.9		2.6			1.8		0.9	0.9	1.5
Leu	24.5	2.7	2.0			2.1	1.0	2.3				1.0	1.8	1.1	2.7				1.8		1.1	5.0	1.2
Tyr	12.4	1.3	0.9				1.8					1.7	1.8			0.9	0.8		1.6				1.1
Phe	4.7		0.9							0.8					0.8								1.5
His	3.5																					1.8	1.4
Lys	14.4	0.9		1.0	1.1	1.0	1.2			1.2	0.9	1.2				1.2	1.0		0.9	0.9		2.8	
Arg	16.2		0.9					1.6	0.8	0.9	0.8		1.0	0.8	0.8		0.8	1.0			0.8		0.9
Trp*	N.D.		+												+								
Yield†...		35	12	23	9	13	20	12	9	19	22	29	15	19	21	15	20	23	17	20	21	6	6
Position...	1–380	1–51	52–75	78–81	84–88	89–109	110–118	119–140	144–157	158–181	182–188	189–202	208–213	214–232	233–249	252–258	259–267	268–272	273–299	300–303	304–310	311–341	342–380

* Determined by amino acid sequencing.
† Expressed in nmol.

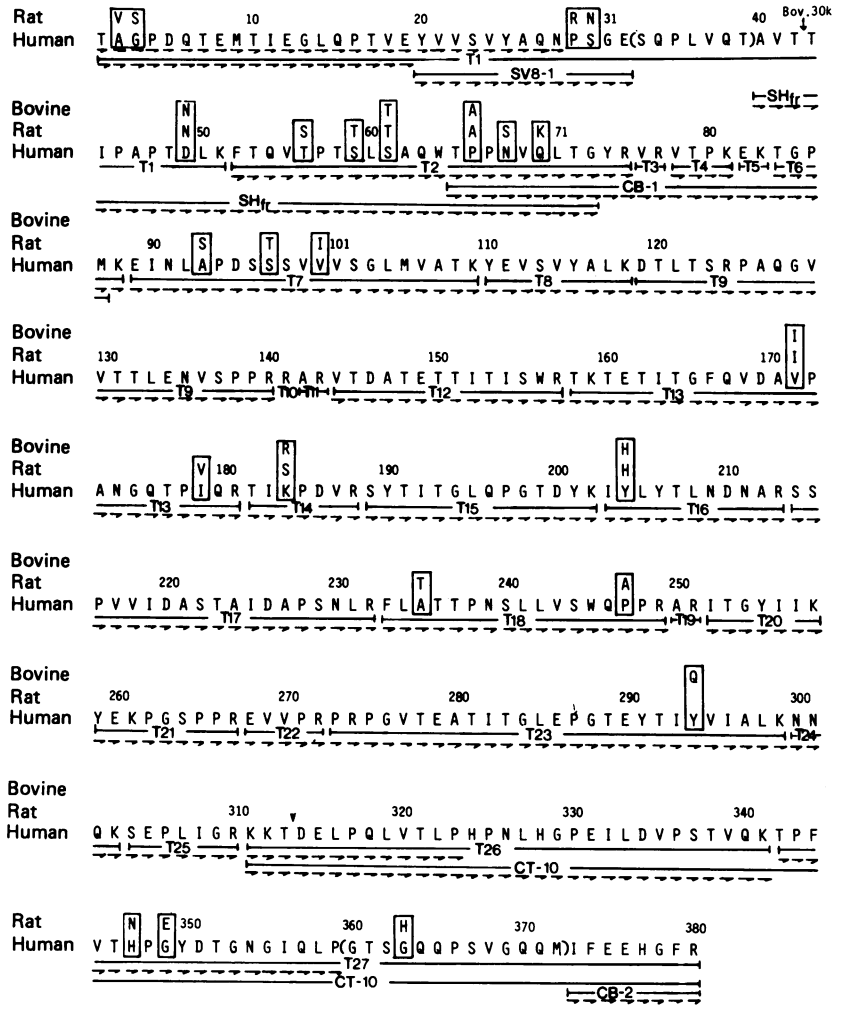

Fig. 2. Amino acid sequence of the 38 kDa heparin-binding domain of human plasma fibronectin (380 residues)

Abbreviations: T, tryptic; CT, citraconylated; CB, CNBr; ⌐ ¬ indicates automated determined sequence. Peptides were arranged by comparison with the sequence deduced from cDNA clones (Kornblihtt *et al.*, 1984). 'Bov. 30 k' marks the beginning of the chymotryptic fragment described by Skorstengaard *et al.* (1986) that extends as far as position 310. '▼' indicates the position of one of the sites where alternative splicing of pre-mRNAs can occur. The extra segment which might be present only in certain fibronectin subunits therefore comprises residues 314–380. Differences in the sequences between human, bovine (Skorstengaard *et al.*, 1986) and rat (Schwarzbauer *et al.*, 1983) fibronectin are shown in boxes.

respectively; it is therefore possible that all three regions contribute to the binding to heparin. The 67-residue *C*-terminal segment is acidic, rich in hydrophobic amino acids and contains the only four histidine residues present in the 38 kDa domain. The possible function of this segment is not known. It does not appear to be necessary for the binding to heparin, since a 30 kDa fragment from bovine fibronectin which ends at position 310 of the 38 kDa (Fig. 2) also binds to heparin (Skorstengaard *et al.*, 1986).

Comparison of the 38 kDa sequence with this bovine fragment and with the sequence deduced from rat cDNA clones (Schwarzbauer *et al.*, 1983) shows a 97 and 94% homology respectively (Fig. 2). Except for the differences at positions 184, 294 and 363, the other substitutions involve only one base change.

The concept that the 38 kDa domain is involved in the heterogeneity observed in fibronectin is supported by two facts. First, the demonstration that, after proteolytic degradation of the individual chains of plasma fibronectin, a fragment with similar characteristics to our 38 kDa domain was produced from one of the chains only (Hayashi & Yamada, 1983; Click & Balian, 1985). Second, the existence of multiple mRNAs for fibronectin which arise by alternative splicing mechanisms occurring precisely at the region which encodes the 67-amino-acid *C*-terminal segment present in the 38 kDa domain (Schwarzenbauer *et al.*, 1983; Kornblihtt *et al.*, 1984).

It is therefore possible that these 67 residues constitute a structural difference between the two polypeptide chains of fibronectin. In this case it should be expected that the other polypeptide chain is shorter and that a fragment different from the 38 kDa is produced after trypsin digestion. Preliminary results from our laboratory indicate that the 58 kDa component (Fig. 1, lane 3) might be such a fragment.

We thank Ms. D. Tummola and Ms. N. Lagmay for their expert technical assistance and Ms. B. Coopersmith for typing the manuscript. This work was supported by U.S. Public

Health Service research grant no. AM 01431. A.G.P. is a recipient of a New Investigator Research Award (no. 1 R23 HL33860).

REFERENCES

Calaycay, J., Pande, H., Lee, T., Borsi, L., Siri, A., Shively, J. E. & Zardi, L. (1985) J. Biol. Chem. **260**, 12136–12141
Click, E. M. & Balian, G. (1985) Biochemistry **24**, 6685–6696
Furcht, L. T. (1983) Mod. Cell Biol. **1**, 53–117
Garcia-Pardo, A., Rosenwasser, E. S. & Frangione, B. (1978) J. Immunol. **121**, 1040–1044
Garcia-Pardo, A., Pearlstein, E. & Frangione, B. (1983) J. Biol. Chem. **258**, 12670–12674
Garcia-Pardo, A., Pearlstein, E. & Frangione, B. (1984) Biochem. Biophys. Res. Commun. **120**, 1015–1021
Garcia-Pardo, A., Pearlstein, E. & Frangione, B. (1985) J. Biol. Chem. **260**, 10320–10325
Gold, L. I., Frangione, B. & Pearlstein, E. (1983) Biochemistry **22**, 4113–4119
Hayashi, M. & Yamada, K. M. (1982) J. Biol. Chem. **251**, 5263–5267
Hayashi, M. & Yamada, K. M. (1983) J. Biol. Chem. **258**, 3332–3340
Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1984) Nucleic Acids Res. **12**, 5853–5868
Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) EMBO J. **4**, 1755–1759
Mosher, D. F. (1984) Annu. Rev. Med. **35**, 561–575
Ozols, J. & Craig, G. (1977) J. Biol. Chem. **252**, 5986–5989
Pande, H. & Shively, J. E. (1982) Arch. Biochem. Biophys. **213**, 258–265
Pande, H., Calaycay, J., Hawke, D., Ben-Avram, C. M. & Shively, J. E. (1985) J. Biol. Chem. **260**, 2301–2306
Pearlstein, E., Gold, L. I. & Garcia-Pardo, A. (1980) Mol. Cell. Biochem. **29**, 103–128
Petersen, T. E., Thogersen, H. C. Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. & Magnusson, S. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 139–141
Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R. & Hynes, R. O. (1983) Cell (Cambridge, Mass.) **35**, 425–431
Sekiguchi, K. & Hakomori, S. (1983) J. Biol. Chem. **258**, 3967–3973
Skorstengaard, K., Jensen, M. S., Petersen, T. E. & Magnusson, S. (1986) Eur. J. Biochem. **154**, 15–29
Smith, D. E. & Furcht, L. T. (1982) J. Biol. Chem. **257**, 6518–6523
Yamada, K. M. (1983) Annu. Rev. Biochem. **52**, 761–799

Received 11 August 1986/29 September 1986; accepted 16 October 1986