

Human plasma fibronectin

Demonstration of structural differences between the A- and B-chains in the III CS region

Tim TRESSEL,* James B. McCARTHY,† Jimmy CALAYCAY,* Terry D. LEE,* Kassu LEGESSE,* John E. SHIVELY* and Hema PANDE*‡

*Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, CA 91010, U.S.A., and

†Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, U.S.A.

Fibronectins are a class of cell-adhesion proteins produced from a single gene. The soluble plasma form is synthesized by hepatocytes and the insoluble cellular form by fibroblasts and other cell types. The proteins possess multiple binding domains for macromolecules including collagen, fibrin and heparin along with at least one cell-binding domain. Cellular as well as plasma fibronectins are dimers of similar but not identical polypeptides. Their differences are the result of internal amino acid sequence variability due to alternative RNA splicing in at least three regions (ED-A, ED-B and III CS). We have been studying this polymorphism at the protein level in plasma fibronectin (pFn). Cathepsin D-digested pFn applied to a heparin-agarose column and eluted with an NaCl stepwise gradient (0.1 M, 0.25 M and 0.5 M) released two polypeptides (75 kDa and 65 kDa) in the 0.5 M-NaCl peak. Immunoblots with monoclonal antibodies IST-2 (specific for the C-terminal heparin-binding domain) and AHB-3 (specific for the III CS domain) suggest that both peptides contain the C-terminal heparin-binding (Hep-2) domain, but that only the larger fragment possesses the III CS region. These two polypeptides (75 kDa and 65 kDa) were digested with trypsin, and the resulting peptides were analysed by fast-atom-bombardment mass spectrometry and compared with the known cDNA-derived peptide sequence. Peptides that were unique to the III CS region were further characterized by micro sequence analysis. The 75 kDa fragment is derived from the A-chain and contains the III CS region (89 amino acid residues) along with the C-terminal heparin-binding (Hep-2) domain and the fibrin-binding (Fib-2) domain. A single galactosamine-based carbohydrate group was detected at Thr-73/74 of the III CS region present in the 75 kDa fragment. The 65 kDa fragment is derived from the B-chain and lacks the entire III CS region but does contain the Hep-2 and Fib-2 domains.

INTRODUCTION

Fibronectins (FNs) constitute a class of high-molecular-mass adhesive glycoproteins that are present in virtually all body fluids, on the surfaces of most cell types and in extracellular matrices. They are involved in a variety of cell functions, including cell attachment and spreading, embryogenesis and development, haemostasis, thrombosis, opsonization and wound healing. These biological functions of fibronectins are based on their affinities to cell surfaces and a number of macromolecules, including collagen, fibrin, heparin, and DNA, and certain bacteria (Hynes, 1985; Akiyama & Yamada, 1987).

There are several distinct forms of fibronectin, which appear to be produced from a single gene. Plasma fibronectin (pFN) is a soluble protein and is composed of two similar but non-identical polypeptide chains. Cellular FN is insoluble and anchors cells to the substrata. Alternative mRNA splicing of a common transcript is believed to give rise to structural polymorphism between the various forms of FN as well as between different subunits within each type. Three distinct regions of alternative splicing (ED-A, ED-B and III CS) have been identified so far, and these may contribute to the generation of up to 20 FN isoforms (Bernard *et al.*, 1985; Kornblihtt *et al.*, 1985; Paul *et al.*, 1986; Gutman & Kornblihtt, 1987). Interestingly, two of these three regions displaying alternative splicing (ED-A and III CS) are located close to the C-terminal heparin-binding domain, suggesting that this section of the FN molecule is highly variable.

In the present paper we describe our results on human pFN. We have examined the C-terminal heparin-binding domain of

pFN that was produced by digesting the protein with cathepsin D. In earlier reports, digestion with thermolysin resulted in the cleavage and loss of the III CS region (Pande *et al.*, 1987). Digestion of pFN with cathepsin D allows the isolation of a C-terminal fragment that contains the intact III CS region (Sekiguchi *et al.*, 1985). This region of the molecule is composed of two disulphide-linked non-identical polypeptides, each containing a heparin-binding (Hep-2) domain and a fibrin-binding (Fib-2) domain. Analysis of these fragments by immunoblotting, protein sequencing and fast-atom bombardment mass spectrometry (f.a.b.-m.s.) revealed that unique structural differences exist between the A and B subunits of pFN near the C-terminal heparin-binding domain.

MATERIALS AND METHODS

Materials

Cathepsin D, pepstatin A, phenylmethanesulphonyl fluoride, trypsin and heparin-agarose were purchased from Sigma Chemical Co. The Zorbax RP-1 h.p.l.c. column was purchased from DuPont. Ultrapure urea was purchased from Schwartz-Mann Laboratories. Immobilon [poly(vinylidene difluoride)]membranes were obtained from Millipore. Trifluoroacetic acid was purchased from Aldrich Chemical Co. Acetonitrile was obtained from Fisher Scientific Co. The monoclonal antibodies (mAbs) IST-2, directed to the Hep-2 domain (Castellani *et al.*, 1986), and IST-9, directed to the ED-A domain (Carnemolla *et al.*, 1987), were gifts from Dr. Luciano Zardi (Laboratoria di Biologia Cellulare, Istituto Scientifico Tumori, Genoa, Italy). The mAb AHB-3,

Abbreviations used: FN, fibronectin; pFN, plasma fibronectin; mAb, monoclonal antibody; f.a.b.-m.s., fast-atom-bombardment mass spectrometry.

‡ To whom correspondence should be addressed.

directed to the III CS domain, was prepared by Dr. James B. McCarthy and is described below. Mini gels were run on Mini-Slab (Idea Scientific, Corvallis, OR, U.S.A.).

MAb

The mAb AHB-3 was prepared against a 33 kDa tryptic/catheptic-digest heparin-binding fragment originating from the A-chain of human FN (McCarthy *et al.*, 1988). Balb c-mouse immune splenocytes were fused with mouse P3-NS-1-AGY-1 myelomas. The fused cells were dispensed into 24-well tissue-culture plates in Dulbecco's modified Eagle's medium (GIBCO Laboratories, Grand Island, NY, U.S.A.) containing 20% (v/v) horse serum and supplemented with 22 mM-hypoxanthine, 9 μ M-aminopterin and 8 mM-thymidine (HAT). Actively secreting hybridomas were detected by an e.l.i.s.a on plates coated with 33 kDa heparin-binding fragment. Cells were cloned by limiting dilution at one cell/two wells. Antibody-producing clones were injected into the peritoneal cavities of Balb/c mice that had been injected 1 month previously with 0.5 ml of pristane. The ascites were harvested and rescreened for the presence of antibody. The antibody was characterized as IgG1 by using Ouchterlony plates and commercially available typing sera (Cooper Biomedical/Cappel Laboratories, Malvern, PA, U.S.A.). This antibody does not react with electrophoretic transfers of B-chain-derived heparin-binding fragments (McCarthy *et al.*, 1988), which lack the III CS sequence.

Purification of pFN and generation of cathepsin D-digest fragments

Human pFN was purified by chromatography on gelatin-Sepharose as previously described (Engvall & Ruoslahti, 1977; Zardi *et al.*, 1980). Cathepsin D digestion and purification of fragments were performed by the method of Richter *et al.* (1981), with some modifications.

FN was eluted from the gelatin-Sepharose column with 4 M-urea in 50 mM-Tris/HCl buffer, pH 7.6. Fractions were pooled and pH was adjusted to 3.5 with 1.0 M-sodium acetate to achieve a final concentration of 0.1 M-sodium acetate. The final concentration of urea for cathepsin D digestion was approx. 3.6 M. FN was digested with cathepsin D (1:300, w/w, enzyme/substrate ratio) at room temperature with gentle rocking. After 4 h the reaction was stopped with the addition of pepstatin A (1:5, w/w, pepstatin A/cathepsin D ratio) and pH was adjusted to 8.0 with 1 M-Tris/HCl buffer, pH 8.4. Digested FN was dialysed overnight against 50 mM-Tris/HCl buffer, pH 7.6, containing phenylmethanesulphonyl fluoride (0.2 mM). The dialysed pFN was mixed batchwise with heparin-agarose resin and shaken overnight at 4 °C. The FN/heparin-agarose mixture was packed into a column and eluted with a stepwise gradient (0.1 M, 0.25 M and 0.5 M) of NaCl. Fractions were assayed at 280 nm and pooled. After the initial purification conditions were established with the heparin-agarose column, pFN was routinely loaded on the column in 0.3 M-NaCl so that only the Hep-2 domain bound to the heparin-agarose column. The two other heparin-binding fragments could then be removed in the wash through. The fragments containing Hep-2 domain were subsequently eluted with 0.5 M-NaCl. Samples were further purified by h.p.l.c. on a Zorbax RP-1 column.

SDS/PAGE and electrotransfer

SDS/polyacrylamide gels (9%) were prepared according to the procedure of Laemmli (1970). Samples were reduced with 2% (v/v) 2-mercaptoethanol. Proteins on the gels were either stained directly with Coomassie Blue or electroblotted to either nitrocellulose membranes (Towbin *et al.*, 1979) for Western blots

or Immobilon membranes (Xu & Shively, 1988) for amino acid composition or sequence analysis.

After the transfer of proteins, the nitrocellulose membranes were blocked with 1.5% (w/v) BSA for 1 h at room temperature and immunostained with either mAb IST-2 (reactive to C-terminal heparin-binding domain) or mAb AHB-3 (reactive to III CS) overnight. Proteins were detected by using appropriate second antibody and substrate. The apparent molecular mass was estimated with the use of prestained protein standards from BRL. The standards were as follows: myosin (200 kDa), phosphorylase *b* (97 kDa), BSA (68 kDa), ovalbumin (43 kDa) and chymotrypsin (25 kDa).

Samples were electrotransferred to Immobilon membranes before amino acid composition analysis by the method of Xu & Shively (1988) with minor modifications. Two layers of Immobilon paper were used. Before electrotransfer, the Immobilon membranes were wetted with methanol and equilibrated for 15 min with transfer buffer [25 mM-Tris/HCl buffer, pH 8.0, containing 10 mM-glycine, 5 mM-dithiothreitol and 10% (v/v) methanol]. The gel was also equilibrated in transfer buffer for 10 min. For 9% minigels (8 cm \times 10 cm), electrotransfer was performed for a total of 80 min (15 min at 100 mA, 15 min at 200 mA and 50 min at 250 mA).

Following electrotransfer, the Immobilon membranes were rinsed with water for 30 s, stained for 10–15 s with 0.2% Coomassie Blue R-250 (methanol/water/acetic acid, 5:4:1, by vol.), destained for 15 min with methanol/water/acetic acid (45:48:7, by vol.), rinsed with water for 2 min and air-dried. The bands were cut out and submitted for amino acid analysis.

Amino acid analysis

Amino acid compositions of the 65 kDa and 75 kDa peptides were obtained on a Beckman System 6300 analyser. The 65 kDa and 75 kDa polypeptides were separated by SDS/PAGE (9% gels) followed by transfer to Immobilon membranes. The proteins were stained with Coomassie Blue and the bands were cut out. Samples were hydrolysed in evacuated tubes for 24 h at 110 °C in 6 M-HCl containing 0.2% 2-mercaptoethanol. The tubes were dried down and amino acids were extracted with 300 μ l of 5% (v/v) acetic acid. The samples were concentrated to dryness and loading buffer was added. The samples were then run as previously described (Del Valle & Shively, 1979).

Reduction and alkylation of cystine disulphides

The reduction/alkylation was performed as described by Rugg & Rudinger (1977) with minor modifications. The sample was dissolved in buffer (6 M-guanidinium chloride in 0.5 M-Tris/HCl buffer, pH 8.5, and 1 mM-EDTA)/propan-1-ol (3:1, v/v). Tributylphosphine and vinylpyridine (1:4 molar ratio, sample to reagent) were then added and the reaction was allowed to proceed for 2 h at room temperature under nitrogen. The sample was purified by reverse-phase h.p.l.c., monitoring at 254 nm and 214 nm. Samples were run over a Zorbax RP-1 column (6.2 mm \times 8 cm) or desalted quickly with an Aquapore butyl column (30 mm \times 2.1 mm) using a 0–100% gradient. Solvent A was aq. 0.1% (v/v) trifluoroacetic acid and solvent B was 90% (v/v) acetonitrile in aq. 0.1% (v/v) trifluoroacetic acid.

Tryptic peptide maps of 65 kDa/75 kDa fragment

Trypsin digestion was performed as previously described (Pande *et al.*, 1987). Approx. 500 pmol of 65 kDa/75 kDa fragment was digested with 1% trypsin for 96 h at 37 °C in 0.2 M-ammonium bicarbonate buffer, pH 8.0. The peptides were fractionated on a Vydac C4 column (2.1 mm \times 250 mm) at 40 °C. A linear gradient was used from 98% solvent A to 100% solvent B, over 60 min at a flow rate of 0.15 ml/min.

Microsequence analysis

Samples were sequenced on a gas-phase microsequencer equipped with a continuous-flow reactor (Shively *et al.*, 1987) built at City of Hope (Hawke *et al.*, 1985). The amino acid phenylthiohydantoin derivatives were analysed by on-line reverse-phase h.p.l.c. and the data were collected and analysed with a Perkin-Elmer LIM system (Shively *et al.*, 1987).

F.a.b.-m.s. analysis

Samples were concentrated to dryness in polypropylene microcentrifuge tubes using a vacuum centrifuge, redissolved in 2 μ l of 5% (v/v) acetic acid, and added to 2 μ l of a liquid matrix on a 1.5 mm \times 6 mm stainless-steel sample stage. The liquid matrix consisted of dithiothreitol/dithioerythritol (5:1, w/w) containing 6 mM-camporsulphonic acid. Positive-ion f.a.b. mass spectra were obtained with a JEOL HX100HF high-resolution, double-focusing magnetic-sector mass spectrometer operating at 5 kV accelerating potential and a nominal resolution of 3000. Sample ionization was accomplished using a 6 keV Xe atom beam. Spectral data were collected for the mass range 400–4500 by using a JEOL DA5000 data system. Mass assignments were accurate to within 0.5 Da, and values for the molecular ion clusters are reported to the nearest integer values of the monoisotopic mass, unless otherwise stated.

RESULTS

Purification and initial characterization of 65 kDa and 75 kDa fragments.

Cathepsin D-digested pFN yielded an interchain disulphide-linked peptide, corresponding to the C-terminal region of the molecule. This peptide was eluted in the 0.5 M-NaCl fraction from a heparin-agarose column (results not shown) and appeared as two bands (65 kDa and 75 kDa) after reduction and analysis by SDS/PAGE (Fig. 1). Immunoblot analysis was carried out to characterize this heterodimer further (Fig. 1). Two bands (65 kDa and 75 kDa) were observed when stained with mAb IST-2 (specific for Hep-2 domain), and one band (75 kDa) was observed when stained with mAb AHB-3 (specific for III CS region). This implies that only the 75 kDa band contains the III CS region. Immunoblot analysis with mAb IST-9 (specific for the ED-A region) did not stain either band (results not shown).

The amino acid compositions of the 65 kDa band and 75 kDa fragments were determined after their electrophoretic separation by SDS/PAGE followed by transfer to Immobilon membranes. The agreement between the observed amino acid composition and that predicted from the cDNA sequence is good, with the exception of methionine and alanine (Table 1). Important to note are the amino acid differences (threonine, glutamic acid/glutamine and proline) found between the two samples. Threonine, glutamic acid/glutamine and proline make up over 42% of the amino acid residues of the III CS region.

In initial work, separation of the two peptides was achieved after reduction/alkylation followed by reverse-phase h.p.l.c. on a Zorbax RP-1 column (results not shown). N-Terminal sequence analyses on 100 pmol of h.p.l.c.-purified 65 kDa and 75 kDa pFN fragments gave identical sequences through ten amino acid residues (Table 2). However, for preparative-scale analysis, including tryptic mapping, it was necessary to analyse 75 kDa/65 kDa peptides together because the separation on the Zorbax column deteriorated and could not be reproduced on new columns.

Peptide mapping, f.a.b.-m.s. and microsequence analysis

To investigate the primary structure of 65 kDa and 75 kDa

pFN domains further, we digested these fragments with trypsin after reduction and alkylation with vinylpyridine. Trypsin digestion of the two fragments resulted in the isolation of 28 different fractions after separation by h.p.l.c. (Fig. 2). Several of these peaks were rechromatographed to separate the tryptic peptides further (peptides indicated with a letter designation). Rechromatography was conducted on the same column but with a less-steep gradient (results not shown). F.a.b.-m.s. analysis was performed on each of the peptides to determine their molecular masses accurately. The m/z values for the protonated molecular

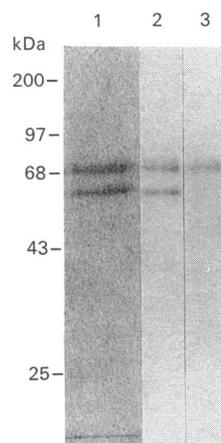


Fig. 1. Analysis of cathepsin D-digested 65 kDa/75 kDa fragments of human pFN by SDS/PAGE and immunoblotting

The 65 kDa/75 kDa fragments of pFN were eluted from the heparin-agarose column in 0.5 M-NaCl (see the text), then analysed by SDS/PAGE and stained with Coomassie Blue (lane 1), or electrotransferred to nitrocellulose membrane and probed with mAbs IST-2 (lane 2) and AHB-3 (lane 3).

Table 1. Amino acid composition of 65 kDa and 75 kDa fragments of human pFN

Approx. 1 μ g portions of 65 kDa and 75 kDa fragments were subjected to amino acid analysis (see the Materials and methods section). Values in parentheses represent the calculated numbers. Abbreviation: N.D., not determined.

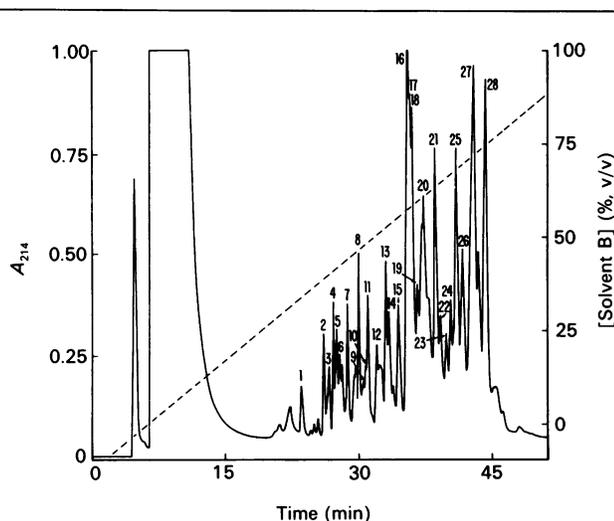
Amino acid	Amino acid composition (residues/molecule)			
	65 kDa fragment		75 kDa fragment	
Asx	57	(51)	60	(57)
Thr	61	(62)	73	(73)
Ser	41	(41)	50	(44)
Glx*	66	(66)	77	(77)
Pro	46	(42)	50	(58)
Gly*	43	(44)	51	(52)
Ala	34	(29)	40	(30)
Val	40	(40)	45	(46)
Met	3	(6)	4	(7)
Ile	26	(26)	30	(30)
Leu	36	(31)	43	(37)
Tyr	20	(22)	24	(24)
Phe	14	(12)	15	(15)
His	8	(8)	10	(13)
Lys	24	(24)	25	(25)
Arg	28	(33)	36	(38)
Cys	N.D.	(17)	N.D.	(17)
Trp	N.D.	(9)	N.D.	(9)

* Values for Glx and Gly have been corrected for a small background peak.

Table 2. *N*-Terminal sequence analysis of 65 kDa and 75 kDa fragments from human pFN

Approx. 100 pmol portions of h.p.l.c.-purified 65 kDa and 75 kDa fragments were sequenced. Values in parentheses are in pmol.

Cycle	Residue	
	65 kDa fragment	75 kDa fragment
1	Ala (27)	Ala (30)
2	Gln (16)	Gln (22)
3	Asn (20)	Asn (22)
4	Pro (31)	Pro (25)
5	Ser (20)	Ser (15)
6	Gly (21)	Gly (15)
7	Glu (20)	Glu (20)
8	Ser (10)	Ser (10)
9	Glu (15)	Glu (10)
10	Pro (10)	Pro (10)

**Fig. 2.** Tryptic peptide map of 65 kDa/75 kDa fragment

Approx. 500 pmol portions of 65 kDa/75 kDa fragment were digested with 1% trypsin for 96 h at 37 °C in 0.2 M-ammonium bicarbonate buffer, pH 8.0. The peptides were fractionated on a Vydac C4 column (2.1 mm × 250 mm) at 40 °C. A linear gradient was used from 98% solvent A to 100% solvent B over 60 min at a flow rate of 0.15 ml/min. T-28 contained undigested peptide.

ions of each peptide were compared with the cDNA-derived peptide sequences (Table 3). Peptides that were unique to either 65 kDa or 75 kDa fragment were also analysed by micro-sequencing. The alignment of tryptic peptides made by comparison with cDNA-derived FN sequence established the primary structure of 65 kDa/75 kDa fragments as shown in Fig. 3. The sequence shown here includes 510 residues, corresponding to Ala-1 through to Arg-510. The sequence beyond Arg-510, not shown in Fig. 3, is identical in pFN A- and B-chains.

Five unique peptides (T-5, T-21c, T-25a, T-25b and T-27a) were obtained. Four of these (T-5, T-21c, T-25b and T-27a) correspond to the III CS region found in the A-chain. The peptide T-27a corresponds to Thr-288–Lys-316 present in the A-chain of pFN. The *N*-terminal threonine residue of peptide T-27a corresponds to the last amino acid residue of the penultimate type III repeat of pFN and is located adjacent to the splice junction. The remaining 28 amino acid residues of peptide T-27a are derived from the III CS region. The sequence was determined for the first 13 cycles. A molecular ion was not obtained for this

Table 3. F.a.b.-m.s. analysis of tryptic peptides generated from 65 kDa and 75 kDa pFN fragments

Mass measurements were made as described in the text. All measurements reported reflect the monoisotopic mass, except those in parentheses which are average masses. Abbreviation: N.D., not determined.

Tryptic peptide	Sequence position	Calculated (<i>M</i> + <i>H</i>)	Experimental (<i>M</i> + <i>H</i>)
T-1a	490–496	785	785
T-1b	85–93	1072	1072
T-2	234–242	1030	N.D.
T-3	279–285	771	771
T-4	157–163	828	828
T-5	356–367	1812†	1812†
T-6	356–367	1812†	1812†
T-7a	53–56	444	N.D.
T-7b	530–532	476	N.D.
T-7c	100–115	1665	1665
T-8	567–576	1276	1276
T-9	533–549	1863*	1863*
T-10	94–99	692	692
T-11	516–525	1341*	1341*
T-12	418–430	1266	1266
T-13	164–177	1544	1544
T-14a	227–233	807	807
T-14b	416–430	1462	1462
T-15	133–156	(2574)	(2574)
T-16a	178–188	1356	1356
T-16b	497–510	1643*	1643*
		1748**	1748**
T-17	189–207	1913	1913
T-18	119–132	1594	1594
T-19	1–26	2663	2664
T-20	431–442	1292	1292
T-21a	431–446	1819	1819
T-21b	27–50	(2694)	(2694)
T-21c	317–355	(4218)	(4218)
T-22	64–84	2117	2117
T-24	208–224	1927	1927
T-25a‡	288–415	4382	N.D.
T-25b	368–415	5494	N.D.
T-26	248–274	(2891)	(2891)
T-27a	288–316	3188	N.D.
T-27a-2	286–316	(3446)	(3446)§
T-27b	451–489	4358	N.D.

*, ** Peptides with one or two cysteines respectively modified with vinylpyridine (see the Materials and methods section).

† Peptide containing carbohydrate.

‡ Peptide with III CS region missing, contains Thr-288 and continues with Gly-378 through to Arg-415.

§ Mass for fragment T-27a-2 (286–316) was obtained from a second tryptic map.

peptide. However, we did obtain an *m/z* of 3446 (average mass) for peptide T-27a-2 from a previous map that was digested with trypsin for a shorter period of time (48 h instead of 96 h). This resulted in lack of cleavage at some of the tryptic-cleavage sites and hence isolation of some larger fragments. The *m/z* value of 3446 for peptide T-27a-2 agrees with a peptide sequence Lys-286–Lys-Thr-Asp...Thr-Val-Gln-Lys-316 found in the pFN A chain. A sequence for ten cycles was obtained for peptide T-27a-2 and was in agreement with the cDNA-derived peptide sequence. These data indicate that pFN A-chain contains the first 25 residues corresponding to the first splice junction located within the III CS region. Peptide T-21c corresponds to a fragment of the middle section of the III CS region (III CS residues 29–67). This fragment contains 39 amino acid residues and has an experi-

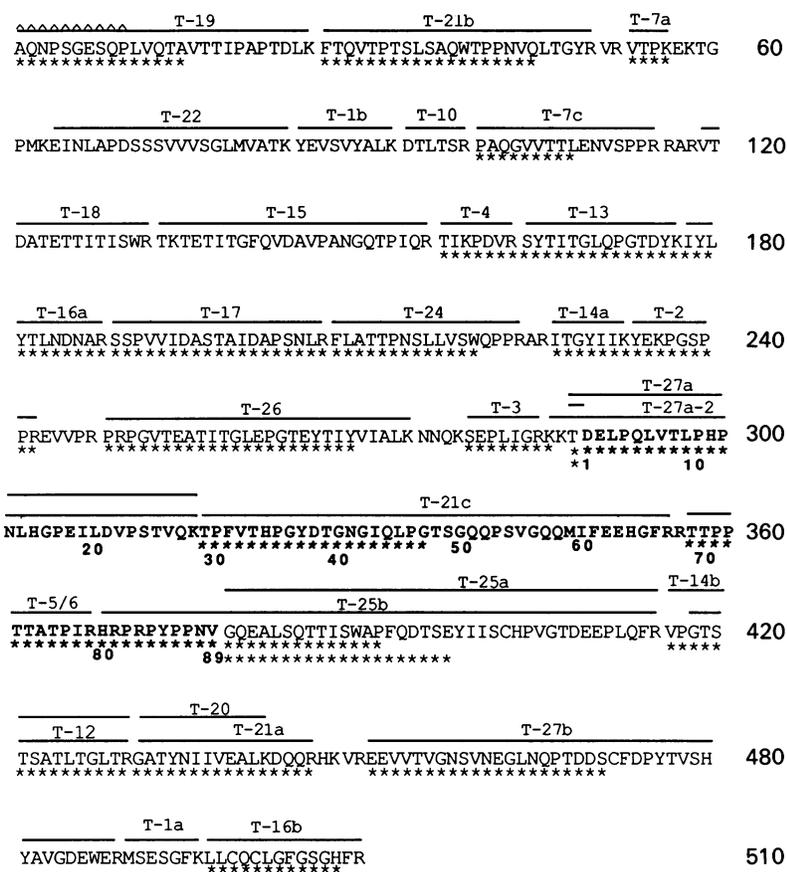


Fig. 3. Primary structure of cathepsin D-digested 65 kDa and 75 kDa fragments of pFN

The sequence of tryptic peptides was determined by using f.a.b.-m.s. (*m/z* values shown in Table 3) and/or microsequencing and comparing these results with cDNA-predicted peptide sequence. There are 60 amino acid residues per row, as indicated by the large numbers on the right. Tryptic peptides are indicated by T-*n* and correspond to fractions from the tryptic map found in Fig. 2. The results of *N*-terminal amino acid sequence analysis on intact 65 kDa and 75 kDa fragments is shown by ^. The residues identified by amino acid sequence analysis on the tryptic fraction are shown by *. The amino acid residues in bold type represent the 89-amino acid-residue III CS region. The III CS region is also numbered 1–89, starting with Asp-289 as III CS residue 1. The numbering of the III CS region is indicated below the amino acid sequence of this region, using numbers in bold type.

mental average molecular mass of 4218 Da, which agrees with the cDNA-derived peptide sequence shown in Fig. 3. The first 19 amino acid residues were confirmed by using *N*-terminal sequence analysis. Peptide T-5 corresponds to residues 69–79 of the III CS region. The complete sequence of this peptide was determined by *N*-terminal sequence analysis. It gave lower yields of threonine phenylthiohydantoin derivative at cycles 5 and 6 compared with cycles 1, 2 and 8, which also contain threonine. This peptide gave a molecular ion at 1812 instead of 1156 due to the presence of a galactosamine-based oligosaccharide group attached to Thr-73/74 of the III CS region. Peptide T-6 also gave a molecular ion at 1812 and contained the same sequence as peptide T-5. Peak T-25 contains two important peptides present in approximately equal molar yields. Peptide T-25b contains the last ten amino acid residues of the III CS region (III CS residues 80–89) plus 38 amino acid residues of the last type III repeat, which are common to both the A- and B-chains. The first 24 amino acid residues of peptide T-25b were confirmed by *N*-terminal sequence analysis. A molecular ion was not obtained for this fragment because of the amount of material needed at this high mass. Peptide T-25a was lacking the entire III CS region. The fragment starts with Thr-288, which is one amino acid residue before the start of the III CS domain, and continues with Gly-378, which is the first amino acid residue of the last type III repeat in pFN, while lacking the entire III CS region. The peptide contained 39 amino

acid residues, with the first 21 confirmed by *N*-terminal protein sequence analysis. A molecular ion was not obtained for this fragment because of its large size. Since we were unable to obtain molecular ions for peptides T-25a and T-25b, and the sequence of both these peptides ended prematurely, the indicated *C*-termini of these peptides as Arg-415 (Fig. 3) are based on the predicted cleavage site for trypsin.

A few small peptides including six dipeptides corresponding to residues 51–52, 57–58, 117–118, 225–226, 447–448 and 449–450, one tetrapeptide corresponding to residues 275–278 and two pentapeptides corresponding to residues 59–63 and 243–247 could not be identified. The molecular ions produced from small quantities of low-molecular-mass peptides are often obscured by the sample matrix ions in mass spectra. However, all the continuous peptides, including the ones corresponding to the III CS region, were isolated and characterized.

DISCUSSION

Analysis of pFN fragments released by limited proteolysis and chemical cleavage have suggested that pFN subunit polymorphism arises from variation at the *C*-terminal region. Use of specific monoclonal or polyclonal antibodies have further supported this conclusion. At the molecular level, the A versus B subunit polymorphism in pFN has been attributed to alternative

splicing of FN pre-mRNA at the III CS region, which is located between sequences coding the Hep-2 domain and Fib-2 domain. Furthermore it has been demonstrated that in human liver, which is the major source of pFN, at least three mRNAs exist. Two of these mRNAs, one containing 192 bases coding for 64 amino acid residues of the III CS region and the second completely lacking the III CS coding sequences, are abundantly transcribed relative to the third (Sekiguchi *et al.*, 1986). This third mRNA codes for the 89-amino acid-residue segment of the III CS domain.

To analyse precisely this subunit polymorphism in pFN, we have attempted protein structural analysis. In earlier work (Pande *et al.*, 1987), we have analysed the structure of 29 kDa/38 kDa Hep-2 domains corresponding to the A- and B-chains of pFN that were produced by thermolysin digestion. These studies revealed the region of variability between the two chains to be located in the III CS region. It was demonstrated that whereas the B-chain of pFN completely lacks the III CS domain, the A chain contains at least a portion of it. These findings were confirmed by other investigators, who have subsequently reported sequences of pFN fragments containing portions of the III CS region (Garcia-Pardo *et al.*, 1987; McCarthy *et al.*, 1988). We also noticed that inclusion of III CS sequences introduces into the pFN molecule a site highly susceptible to thermolysin digestion. Therefore the resulting 29 kDa peptide described previously contained only the first five amino acid residues of this region (Pande *et al.*, 1987). In order to obtain FN fragments containing the intact III CS region, cathepsin D was employed for pFN digestion. Through the use of cathepsin D, we have isolated and characterized two larger fragments (65 kDa and 75 kDa), one of which contains the intact III CS region.

In the present work we have shown that both 65 kDa and 75 kDa fragments have identical N-terminal sequences. Both fragments start with alanine, which corresponds to residue 1583 of intact pFN. The cleavage site giving rise to these two fragments was thus identified as Tyr-1582-Ala-1583. Cathepsin D was also found to cleave placental fibronectin in this exact same position (Tressel *et al.*, 1988). This peptide cleavage pattern is consistent with the preference of cathepsin D for peptide bonds involving aromatic residues.

In order to detect the domain changes in the A- and B-chains of pFN, we employed an immunoblotting experiment using mAbs that were reactive to specific pFN domains. One mAb, IST-2, recognizes determinants on the A-chain and B-chain Hep-2 regions, whereas the second mAb, AHB-3, recognizes determinants present only in the A-chain. Both 65 kDa and 75 kDa fragments reacted with mAb IST-2, indicating that these contain the Hep-2 domain, especially the type III repeats 12-14 of pFN. However, only the 75 kDa fragment reacted with mAb AHB-3, indicating that only the A-chain of pFN contains the III CS region. Since the precise location of the epitope recognized by mAb AHB-3 within the III CS region has not been determined, these immunoblotting results did not provide information regarding the exact size of the III CS domain present in pFN A-chain.

The exact structure of this splice junction region in pFN was therefore determined by analysing the tryptic peptides produced from the 65 kDa and 75 kDa fragments by using f.a.b.-m.s. and microsequence analyses. The *m/z* values for the protonated molecular ions of the tryptic peptides matched well with the cDNA-derived peptide sequence. The tryptic peptides corresponding to the splice junction region were determined by microsequence analysis, and these data confirmed the f.a.b.-m.s. results. Peptides T-27A, T-21C, T-5 and T-25b contained sequences corresponding to the III CS region (89 residues, Fig. 3). These results demonstrate that the A-chain of pFN contains

Table 4. Schematic representation of different FN isoforms that could arise by alternative splicing in the III CS region and a list of tryptic peptides that could be generated from these different isoforms

FN isoforms	Tryptic peptides	(M + H)	Identification criteria
1. 	1. TGQEALSQTTSWAPFQDTSEYIHSCHPVGTDDEEPLQFR* 1. TDELPLQLVTLPHPNLHGPEILDVPSVQK 2. TPFVTHPGYDTGNGIQLPGTSGQQPSVGGQQMIFEEHGF 3. TTPPTTATPIR 4. HRPRYPNVGEEIQIGHIPR* 5. EDVDYHLPHGPGLNPNASTQEALSQTTSWAPFQDTSEYIHSCHPVGTDDEEPLQFR*	4382 3187 4218 1812 2462 N.O. N.O.	Sequence analysis Mass and sequence analysis Mass and sequence analysis Mass and sequence analysis N.O. N.O.
2. 	1. TDELPLQLVTLPHPNLHGPEILDVPSVQK 2. TPFVTHPGYDTGNGIQLPGTSGQQPSVGGQQMIFEEHGF 3. TTPPTTATPIR 4. HRPRYPNVGQEALSQTTSWAPFQDTSEYIHSCHPVGTDDEEPLQFR*	3187 4218 1812 5494	Mass and sequence analysis Mass and sequence analysis Mass and sequence analysis Sequence analysis
3. 	1. TVQK* 2. TPFVTHPGYDTGNGIQLPGTSGQQPSVGGQQMIFEEHGF 3. TTPPTTATPIR 4. HRPRYPNVGEEIQIGHIPR* 5. EDVDYHLPHGPGLNPNASTQEALSQTTSWAPFQDTSEYIHSCHPVGTDDEEPLQFR*	475 4218 1812 2462 N.O.	N.O. Mass and sequence analysis Mass and sequence analysis N.O.
4. 	1. TVQK* 2. TPFVTHPGYDTGNGIQLPGTSGQQPSVGGQQMIFEEHGF 3. TTPPTTATPIR 4. HRPRYPNVGQEALSQTTSWAPFQDTSEYIHSCHPVGTDDEEPLQFR*	475 4218 1812 5494	N.O. Mass and sequence analysis Mass and sequence analysis Sequence analysis

Unique peptides from each isoform are given in bold type and marked with an asterisk. Abbreviation: N.O., not observed.

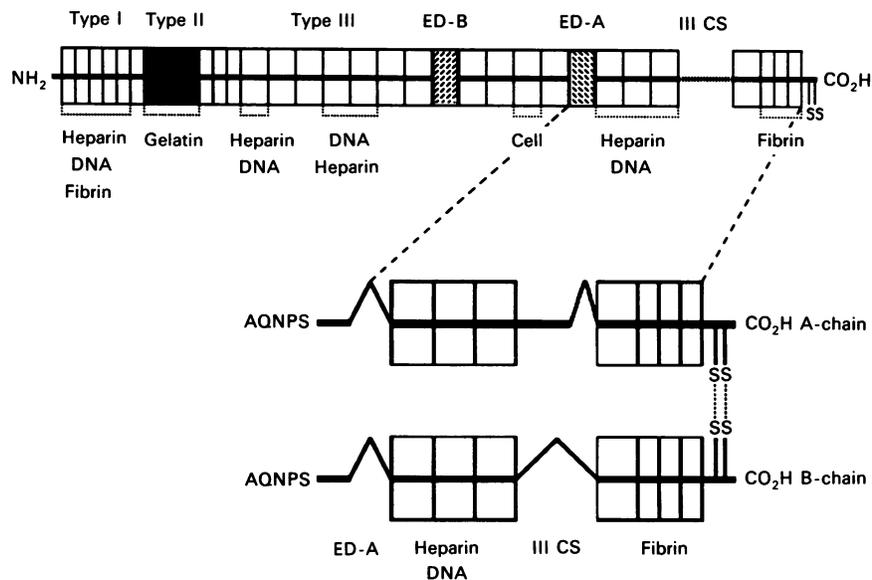


Fig. 4. Domain model of pFN and schematic representation of cathepsin D-digested pFN fragments possessing structural variations between the A- and B-chains

A-chain of pFN is shown to have the ED-A spliced out and to contain 89 amino acid residues of the III CS domain. The B-chain is shown to have both the ED-A and III CS spliced out.

only the first 89 residues of the III CS region (120 amino acid residues maximum size). The B-chain, however, lacks the entire III CS region, as evidenced by the sequence of peptide T-25a, which starts at Thr-288 and then continues with Gly-378.

Regarding the post-translational modification, Skorstengaard *et al.* (1986) have reported that bovine pFN contains a galactosamine-based carbohydrate linked to Thr-1943/1944 and a glucosamine-based carbohydrate linked to Asn-1987, both located in the III CS region. The III CS region of oncofetal FN, however, appears to be highly glycosylated, having as many as six *O*-linked carbohydrate moieties (Matsuura *et al.*, 1988). Our tryptic peptide T-5 corresponds to residues 1939–1949 of human pFN. The peptide contains the exact same sequence as found in bovine pFN and oncofetal FN. F.a.b.-m.s. analysis of this peptide gave a molecular ion at 1812 instead of 1156, predicted from amino acid sequence. This difference of 656 is due to the presence of a carbohydrate moiety and is in agreement with the proposed trisaccharide (NeuAc-Hex-HexNAc) structure (Matsuura *et al.*, 1988). The peptide T-21c corresponds to residues 29–67 of the III CS region. This region in oncofetal FN contains three carbohydrate moieties, including the one that has been reported to be essential for maintaining the oncofetal epitope specificity (Matsuura *et al.*, 1988). For peptide T-21c, we obtained a molecular ion at 4218 (Table 3). This observed average mass value agrees well with the predicted mass of the peptide indicating that, unlike oncofetal FN, this region in pFN is not glycosylated. The peptide containing the glucosamine linked to asparagine, which is derived from the C-terminal 31-amino acid residue section of the III CS region, was not found in pFN.

The results presented here demonstrate that the 65 kDa and 75 kDa fragments generated from the two chains of pFN are identical, with the exception of the III CS region. Previous studies based mainly on cDNA data have suggested that up to five different FN isoforms may be generated by sequence variations in the III CS region. To identify which of these five isoforms are present in pFN, we have compared our tryptic peptide data with the various peptides that may be generated

from these different isoforms. This information is presented in Table 4, where the unique peptides from each possible isoform are given in **bold type**. The first possibility is that the complete III CS region is lacking, as shown for isoform 1 (Table 4). This form will generate a single unique peptide containing sequences from type III repeats 14 and 15 while missing the III CS region completely. As discussed above, the peptide T-25a contains this exact sequence, indicating that this isoform is present in pFN and corresponds to the B-chain of pFN. The second possibility is that the full-length III CS of 120 amino acid residues is present. This isoform would produce five tryptic peptides of which two, no. 4 and no. 5, are unique (Table 4). The fact that we did not obtain these unique peptides indicates that this form is not present in pFN. The third possibility is that the III CS region of 89 amino acid residues, which is lacking the C-terminal 31 residues, is present. This FN form would generate four peptides, of which one peptide, no. 4, is unique. Since we obtained all the four peptides, including the unique peptide, this FN isoform must be present and corresponds to the A-chain of pFN. The fourth possibility is that the III CS region is lacking the N-terminal 25 residues. This form would produce five tryptic peptides, three of these being unique. None of these unique peptides was detected in our map, again suggesting that this isoform is not present in pFN. The last possibility is that the size of the III CS region is only 64 amino acid residues and it is lacking the N-terminal 25 as well as C-terminal 31 residues. This isoform would generate four tryptic peptides, of which no. 1 (Thr-Val-Gln-Lys) is unique (Table 4). We did not detect this unique peptide in our map. Although it is often difficult to obtain a mass for a short peptide like this, we would have identified it by sequence analysis had it been present. Furthermore, the molar yields of peptides corresponding to isoforms 1 and 3 (Table 4) are approximately equal, indicating that pFN contains only two isoforms (Fig. 4).

This analysis (Table 4) resolves the potential problem that may be encountered in analysing a mixture of peptides derived from both the A- and B-chains of pFN, namely that it would be difficult to decide from which chain a peptide arises. The problem

is resolved by assuming that all possible isoforms predicted by mRNA analysis may be present. Fortunately, each isoform is expected to produce at least one unique tryptic peptide. Therefore the protein data are compatible with only two isoforms, one of which must be ascribed to each chain. This conclusion is compatible with the mRNA data, which otherwise are flawed by the problem of analysing plasma and cellular forms together. Although it is also possible that cellular forms are present in the soluble pFN, they would be expected to be minor (< 5%) because of their lower solubility and distinct tissue distribution.

In summary, we have presented our results on immunoblotting and primary sequence on pFN subunits. These results clearly demonstrate that the B-chain of pFN generates the 65 kDa cathepsin D fragment, which lacks the III CS region, whereas the A-chain generates the 75 kDa fragment, which contains the first 89 amino acid residues of the III CS region. This is different from the A-chain of bovine pFN, which was shown to contain the 120-amino acid-residue III CS domain. Of particular interest is the fact that, whereas the mRNA encoding the B-chain of pFN is abundantly detected in liver, the A-chain appears to be encoded by a minor mRNA species (Sekiguchi *et al.*, 1986). Furthermore, our results clearly demonstrate that the third mRNA species, though detected abundantly in liver, does not encode a major form of pFN as has been previously implied (Sekiguchi *et al.*, 1986). It is therefore postulated that this mRNA encodes either a minor pFN form present at < 5% level or, more likely, a cellular form of FN, which is also known to be synthesized in the liver.

We are very grateful to Dr. Luciano Zardi for the monoclonal antibodies IST-2 and IST-9. We thank James Sligar for performing amino acid analysis. This research was supported by National Institutes of Health Grant GM 29007. The microsequencing part of this research was supported in part by the City of Hope Cancer Center Grant from the National Cancer Institute, CA 33572.

REFERENCES

- Akiyama, S. K. & Yamada, K. M. (1987) *Adv. Enzymol. Relat. Areas Mol. Biol.* **59**, 1–57
- Bernard, M. P., Kolbe, M., Weil, D. & Chu, M.-L. (1985) *Biochemistry* **24**, 2698–2704
- Carnemolla, B., Borsi, L., Zardi, L., Owens, R. J. & Baralle, F. E. (1987) *FEBS Lett.* **215**, 269–273
- Castellani, P., Siri, A., Rosellini, C., Infusini, E., Borsi, L. & Zardi, L. (1986) *J. Cell Biol.* **103**, 1671–1677
- Del Valle, U. & Shively, J. E. (1979) *Anal. Biochem.* **96**, 77–83
- Engvall, E. & Ruoslahti, E. (1977) *Int. J. Cancer* **20**, 1–5
- Garcia-Pardo, A., Rostagno, A. & Frangione, B. (1987) *Biochem. J.* **241**, 923–928
- Gutman, A. & Kornblihtt, A. R. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7179–7182
- Hawke, D. H., Harris, D. C. & Shively, J. E. (1985) *Anal. Biochem.* **147**, 315–330
- Hynes, R. (1985) *Annu. Rev. Cell Biol.* **1**, 67–90
- Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* **4**, 1755–1759
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Matsuura, H., Takio, K., Titani, K., Greene, T., Levery, S. B., Salyan, M. E. K. & Hakomori, S. (1988) *J. Biol. Chem.* **263**, 3314–3322
- McCarthy, J. B., Chelberg, M. K., Mickelson, D. J. & Furcht, L. T. (1988) *Biochemistry* **27**, 1380–1388
- Pande, H., Calaycay, J., Lee, T. D., Legesse, K., Shively, J. E., Siri, A., Borsi, L. & Zardi, L. (1987) *Eur. J. Biochem.* **162**, 403–411
- Paul, J. I., Schwarzbauer, J. E., Tamkun, J. W. & Hynes, R. O. (1986) *J. Biol. Chem.* **261**, 12258–12265
- Richter, H., Seidl, M. & Hormann, H. (1981) *Hoppe Seyler's Z. Physiol. Chem.* **362**, 399–408
- Ruegg, U. T. & Rudinger, J. (1977) *Methods Enzymol.* **47**, 111–126
- Sekiguchi, K., Siri, A., Zardi, L. & Hakomori, S. (1985) *J. Biol. Chem.* **260**, 5105–5114
- Sekiguchi, K., Klos, A. M., Kurachi, K., Yoshitake, S. & Hakomori, S. (1986) *Biochemistry* **25**, 4936–4941
- Shively, J. E., Miller, P. & Ronk, M. (1987) *Anal. Biochem.* **163**, 517–529
- Skorstengaard, K., Jensen, M. S., Sahl, P., Petersen, T. E. & Magnusson, S. (1986) *Eur. J. Biochem.* **161**, 441–453
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- Tressel, T., Shively, J. E. & Pande, H. (1988) *Arch. Biochem. Biophys.* **266**, 639–643
- Xu, Q.-Y. & Shively, J. E. (1988) *Anal. Biochem.* **170**, 19–30
- Zardi, L., Siri, A., Carnemolla, B., Cosulich, E., Viale, G. & Santi, L. (1980) *J. Immunol. Methods* **26**, 155–165

Received 23 July 1990/1 October 1990; accepted 10 October 1990