# Human Cellular Fibronectin: Comparison of the Carboxyl-Terminal Portion with Rat Identifies Primary Structural Domains Separated by Hypervariable Regions†

Michael P. Bernard,‡ Margret Kolbe,§ Dominique Weil,‡,∥ and Mon-Li Chu*,§

*Departments of Biochemistry and Obstetrics/Gynecology, University of Medicine and Dentistry of New Jersey-Rutgers Medical School, Piscataway, New Jersey 08854*

*Received July 24, 1984*

ABSTRACT: We report the isolation and characterization of four overlapping cDNA clones coding for human cellular fibronectin which continuously cover more than 3 kilobases in length. The nucleotide sequence of these cDNAs has been determined, thus elucidating the amino acid sequence of the C-terminal 794 residues of human fibronectin, which cover the edge of cellular-, heparin-, and fibrin-binding domains of this protein. Comparisons of the nucleotide sequences and the deduced amino acid sequences with those of rat [Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R., & Hynes, R. O. (1983) *Cell (Cambridge, Mass.) 35*, 421] indicate a high degree of conservation at both nucleotide and amino acid levels. Comparison with previously published data on amino acid sequences of bovine fibronectin made it possible to identify structurally important features of the protein during the evolution of human, calf, and rat. The deduced human amino acid sequences contain five type III and three type I repeats of internal homologies. The interspecies conservation in amino acids is more pronounced in regions containing the internal repeats and within each functional domain. The implications of these interspecies conservation and divergence are discussed.

The function of cellular fibronectin, a large glycoprotein found on the surface of many different cells, is to contribute to the adhesive interactions both between cell and cell and between cell and substratum. A similar but not identical protein is present in large amounts in blood plasma and is believed to play a central role in thrombosis. As a result of these numerous interactions, it is known that these glycoproteins exhibit several binding domains specific for cells as well as for a number of macromolecules such as collagen, heparin, fibrin, actin, and DNA. [For recent reviews see Hynes & Yamada (1982), Yamada (1983), and Furcht (1983).] It has been shown that each attachment domain exhibits at least one of three different types of repeating units (I–III), varying in size from 50 to 90 amino acids. A great deal of work has been done to identify, characterize, and compare these repetitive portions (Skorstengaard et al., 1982; Petersen et al., 1983).

The large percentage of amino acid data available (Pande & Shively, 1982; Garcia-Pardo et al., 1983, 1984; Pierschbacher et al., 1982) has recently been supplemented with those derived from cDNA clones for chicken, calf, human, and rat (Fagan et al., 1979; Kornblihtt et al., 1983, 1984a; Oldberg et al., 1983; Schwarzbauer et al., 1983). Analysis of human and rat cDNAs revealed the existence of multiple fibronectin mRNAs which may be explained by alternative splicing of a single gene (Schwarzbauer et al., 1983; Kornblihtt et al., 1984a). These splicing regions have been shown to occur at different sites along the message. In man, this region is further 5′ and, unlike the rat, makes up an entire type III repeat.

We describe here the isolation and nucleotide sequence of four overlapping fibronectin cDNA clones prepared from normal human fibroblasts and encode for the C-terminal one-third of human cellular fibronectin. Our data were compared with the recently published data from bovine and rat; degrees of homology and divergence between species are analyzed and the implications discussed.

## Materials and Methods

*Materials.* Avian myeloblastoma virus reverse transcriptase was supplied by Dr. Joseph Beard (Life Sciences Inc., St. Petersberg, FL). *Escherichia coli* DNA polymerase I (Klenow fragment) and proteinase K were purchased from Boehringer Mannheim. Terminal deoxynucleotidyltransferase and deoxynucleotide triphosphates were obtained from P-L Biochemicals, Inc. Restriction enzymes and T4 polynucleotide kinase were purchased from New England Biolabs or Bethesda Research Laboratories and were used according to the manufacturer's recommendations. Labeled nucleotides were purchased from Amersham Corp.

*RNA Purification.* Normal human fibroblasts (CRL 1106, American Type Culture Collection, Rockville, MD) were grown in Dulbecco's modified Eagle's medium containing 10% fetal calf serum. Total poly(A) RNA was prepared from $10^8$ confluent cells by proteinase K digestion and oligo(deoxythymidylate)–cellulose chromatography and centrifuged through a 5–20% sucrose gradient as previously described (Chu et al., 1982). High molecular weight RNA (>28 S) were pooled and translated in a cell-free protein-synthesizing system prepared from rabbit reticulocyte (Pelham & Jackson, 1976). The degree of enrichment for fibronectin mRNA was estimated by polyacrylamide gel electrophoresis of the translation products.

*Construction of cDNA Clones.* cDNA was prepared by a modification of the procedure of Wickens et al. (1981). Poly(A) RNA that had been enriched for fibronectin sequences was used to synthesize a cDNA by using reverse transcriptase and oligo(dT) as primer. After separation of mRNA template by boiling, double-stranded DNA was synthesized with the Klenow fragment of DNA polymerase I, made blunt ended

† This work was supported by grants from National Institutes of Health (AM 32380 and AM 16516) and Institut National de la Sante et Recherche Medicale (France).

‡ Department of Obstetrics/Gynecology.

§ Department of Biochemistry.

∥ On leave of absence from Unite 12 INSERM-Hopital des Enfants Malades, Paris, France.

0006-2960/85/0424-2698$01.50/0 

with S1 nuclease, and fractionated on a sucrose gradient. Poly(G) was then added to the 3′ ends of the double-stranded DNA by use of terminal deoxynucleotidyltransferase (Roychoudhury et al., 1976).

Plasmid pBR322 DNA was cleaved with *Eco*RI, made blunt ended with reverse transcriptase, and tailed with an average of 15 dCMP residues by use of terminal deoxynucleotidyltransferase (Rougeon & Mach, 1977). The double-stranded cDNA and plasmid DNA were annealed, and the resulting chimeric plasmids were used to transform *E. coli* strain RR1 as described previously (Peacock et al., 1981). This procedure allows the inserts to be excised by *Eco*RI cleavage. Transformants were selected for ampicillin resistance. Recombinant DNA was handled in accordance with the National Institutes of Health guidelines.

*Screening of Recombinant Plasmids.* Transformants were screened by the colony hybridization procedure of Hanahan & Meselson (1980) using a [$^{32}$P]cDNA synthesized from the enriched poly(A) RNA. Positive clones were grown overnight at 37 °C in 3 mL of Luria broth containing ampicillin at 50 μg/mL and then stored at 4 °C. Pools of these clones were made by inoculation of 0.5 mL of each of five saturate cultures into 100 mL of Luria broth containing ampicillin and grown at 37 °C overnight. Plasmid DNA was isolated from the cultures and purified by CsCl/ethidium bromide gradient centrifugation (Kupersztoch & Helinski, 1973). DNA from each plasmid pool was labeled with [$^{32}$P]dNTPs by nick translation (Rigby et al., 1977) and hybridized with filter-bound RNA. RNA was subjected to electrophoresis on a 0.8% agarose/6% formaldehyde gel (Rave et al., 1979) and transferred to nitrocellulose paper as described by Thomas (1980).

*DNA-Sequencing Determination.* DNA sequencing was carried out essentially as described by Maxam & Gilbert (1980). The 5′ ends of restriction fragments were labeled with [$\gamma$-$^{32}$P]ATP and $T_4$ polynucleotide kinase. Either the labeled fragments were restricted with a second restriction endonuclease or the strands were separated by electrophoresis. The polyacrylamide gels for sequencing ranged from 0.1 to 0.45 mm thickness and from 40 to 80 cm in length depending on the size of the fragment and quality of the label. The effects due to salt were minimized either by employing a 20% gel or by merely diluting the sample before addition to the gel.

## Results

*Enrichment of Fibronectin mRNA.* Cultured human fibroblasts synthesize and secrete large amounts (up to 1–3% of total protein) of fibronectin (Hynes, 1981). Total poly(A) RNA isolated from fibroblasts was further enriched for fibronectin sequences by size fractionation on a sucrose gradient. RNA sedimenting greater than 28 S was pooled. This fraction represented about 10% of the total poly(A) RNA applied to the sucrose gradients; thus, there was a 10-fold enrichment of the fibronectin mRNA. [$^{32}$P]cDNA synthesized from the enriched RNA hybridized predominantly to three major poly(A) RNA species, 7.8, 5.8, and 4.8 kilobases (kb) in length (Figure 1). We have previously shown that the 5.8- and 4.8-kb mRNA encoded for type I collagen (Chu et al., 1982). The intensity and the size of the 7.8-kb band suggests it may encode for fibronectin. The hybridization results showed that [$^{32}$P]cDNA synthesized from the enriched mRNA may be used as a probe to screen fibronectin clones.

*Construction and Screening of Recombinant Clones.* Double-stranded cDNA synthesized from enriched poly(A) RNA was inserted into the *Eco*RI site of plasmid pBR322 as described above. Two thousand independent transformants were obtained from 50 ng of size-fractionated cDNA. The

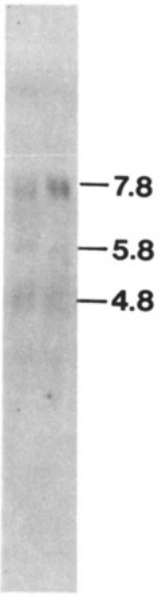



FIGURE 1: Hybridization of human fibroblast poly(A) RNA with $^{32}$P-labeled cDNA synthesized from RNA enriched for fibronectin sequences. Two identical lanes of poly(A) RNA (about 1 μg) were separated on agarose gel, transferred to nitrocellular paper, and hybridized with $^{32}$P-labeled ($5 \times 10^7$ cpm/μg) cDNA synthesized from size-fractionated poly(A) RNA (see Materials and Methods). RNA markers were run on a parallel lane and visualized by ethidium bromide staining.

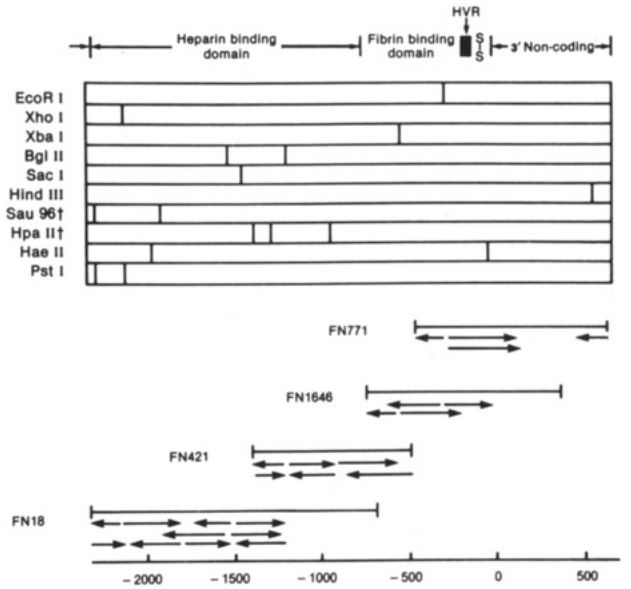
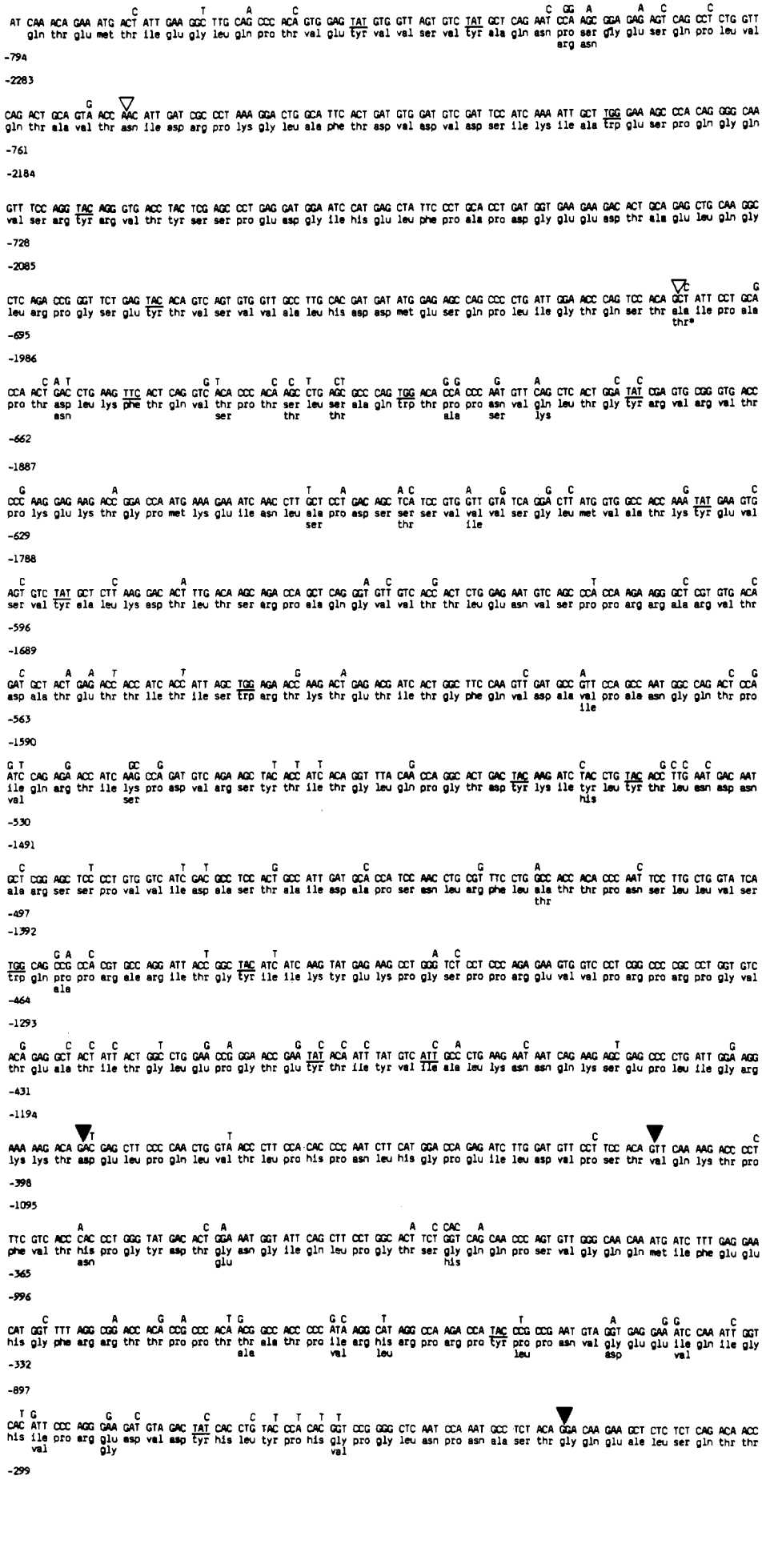



FIGURE 2: Restriction map and sequencing strategy of four cloned cDNAs for human cellular fibronectin (†) partial restriction map. Nucleotides (scale at bottom) are numbered with the first nucleotide of the stop codon designated zero. (HVR) Hypervariable region; (S–S) cysteine binding region for dimer formation.

clones were screened by in situ hybridization using [$^{32}$P]cDNA synthesized from the enriched mRNA. About 80 positive clones were obtained, and plasmid DNA was collected into 16 pools of five clones. Plasmid pools were nick translated and hybridized to Northern blotted fibroblast poly(A) RNA. Ten pools were found to hybridize to the 7.8-kb fibroblast RNA. The five plasmid strains from one of these positive pools were grown separately, and the plasmid DNAs were assayed for their ability to hybridize to the 7.8-kb RNA. The resulting positive clone Hf771 contained an insert of 1300 base pairs (bp) and was further characterized by restriction endonuclease mapping and nucleotide sequence analysis. The restriction


-2382
AT CAA ACA GAA ATG ACT ATT GAA GGC TTG CAG CCC ACA GTG GAG TAT GTG GTT AGT GTC TAT GCT CAG AAT CCA AGC GGA GAG AGT CAG CCT CTG GTT
gln thr glu met thr ile glu gly leu gln pro thr val glu tyr val val ser val tyr ala gln asn pro ser gly glu ser gln pro leu val
arg asn
-794
-2283
CAG ACT GCA GTA ACC AAC ATT GAT CGC CCT AAA GGA CTG GCA TTC ACT GAT GTG GAT GTC GAT TCC ATC AAA ATT GCT TGG GAA AGC CCA CAG GGG CAA
gln thr ala val thr asn ile asp arg pro lys gly leu ala phe thr asp val asp val asp ser ile lys ile ala trp glu ser pro gln gly gln
-761
-2184
GTT TCC AGG TAC AGG GTG ACC TAC TCG AGC CCT GAG GAT GGA ATC CAT GAG CTA TTC CCT GCA CCT GAT GGT GAA GAA GAC ACT GCA GAG CTG CAA GGC
val ser arg tyr arg val thr tyr ser ser pro glu asp gly ile his glu leu phe pro ala pro asp gly glu glu asp thr ala glu leu gln gly
-728
-2085
CTC AGA CCG GGT TCT GAG TAC ACA GTC AGT GTG GTT GCC TTG CAC GAT GAT ATG GAG AGC CAG CCC CTG ATT GGA ACC CAG TCC ACA GCT ATT CCT GCA
leu arg pro gly ser glu tyr thr val ser val val ala leu his asp asp met glu ser gln pro leu ile gly thr gln ser thr ala ile pro ala
thr*
-695
-1986
CCA ACT GAC CTG AAG TTC ACT CAG GTC ACA CCC ACA AGC CTG AGC GCC CAG TGG ACA CCA CCC AAT GTT CAG CTC ACT GGA TAT CGA GTG CGG GTG ACC
pro thr asp leu lys phe thr gln val thr pro thr ser leu ser ala gln trp thr pro pro asn val gln leu thr gly tyr arg val arg val thr
asn ser thr thr ala ser lys
-662
-1887
CCC AAG GAG AAG ACC GGA CCA ATG AAA GAA ATC AAC CTT GCT CCT GAC AGC TCA TCC GTG GTT GTA TCA GGA CTT ATG GTG GCC ACC AAA TAT GAA GTG
pro lys glu lys thr gly pro met lys glu ile asn leu ala pro asp ser ser ser val val val ser gly leu met val ala thr lys tyr glu val
ser thr ile
-629
-1788
AGT GTC TAT GCT CTT AAG GAC ACT TTG ACA AGC AGA CCA GCT CAG GGT GTT GTC ACC ACT CTG GAG AAT GTC AGC CCA CCA AGA AGG GCT CGT GTG ACA
ser val tyr ala leu lys asp thr leu thr ser arg pro ala gln gly val val thr thr leu glu asn val ser pro pro arg arg ala arg val thr
-596
-1689
GAT GCT ACT GAG ACC ACC ATC ACC ATT AGC TGG AGA ACC AAG ACT GAG ACG ATC ACT GGC TTC CAA GTT GAT GCC GTT CCA GCC AAT GGC CAG ACT CCA
asp ala thr glu thr thr ile thr ile ser trp arg thr lys thr glu thr ile thr gly phe gln val asp ala val pro ala asn gly gln thr pro
ile
-563
-1590
ATC CAG AGA ACC ATC AAG CCA GAT GTC AGA AGC TAC ACC ATC ACA GGT TTA CAA CCA GGC ACT GAC TAC AAG ATC TAC CTG TAC ACC TTG AAT GAC AAT
ile gln arg thr ile lys pro asp val arg ser tyr thr ile thr gly leu gln pro gly thr asp tyr lys ile tyr leu tyr thr leu asn asp asn
val ser his
-530
-1491
GCT CGG AGC TCC CCT GTG GTC ATC GAC GCC TCC ACT GCC ATT GAT GCA CCA TCC AAC CTG CGT TTC CTG GCC ACC ACA CCC AAT TCC TTG CTG GTA TCA
ala arg ser ser pro val val ile asp ala ser thr ala ile asp ala pro ser asn leu arg phe leu ala thr thr pro asn ser leu leu val ser
thr
-497
-1392
TGG CAG CCG CCA CGT GCC AGG ATT ACC GGC TAC ATC ATC AAG TAT GAG AAG CCT GGG TCT CCT CCC AGA GAA GTG GTC CCT CGG CCC CGC CCT GGT GTC
trp gln pro pro arg ala arg ile thr gly tyr ile ile lys tyr glu lys pro gly ser pro pro arg glu val val pro arg pro arg pro gly val
ala
-464
-1293
ACA GAG GCT ACT ATT ACT GGC CTG GAA CCG GGA ACC GAA TAT ACA ATT TAT GTC ATT GCC CTG AAG AAT AAT CAG AAG AGC GAG CCC CTG ATT GGA AGG
thr glu ala thr ile thr gly leu glu pro gly thr glu tyr thr ile tyr val Ile ala leu lys asn asn gln lys ser glu pro leu ile gly arg
-431
-1194
AAA AAG ACA GAC GAG CTT CCC CAA CTG GTA ACC CTT CCA CAC CCC AAT CTT CAT GGA CCA GAG ATC TTG GAT GTT CCT TCC ACA GTT CAA AAG ACC CCT
lys lys thr asp glu leu pro gln leu val thr leu pro his pro asn leu his gly pro glu ile leu asp val pro ser thr val gln lys thr pro
-398
-1095
TTC GTC ACC CAC CCT GGG TAT GAC ACT GGA AAT GGT ATT CAG CTT CCT GGC ACT TCT GGT CAG CAA CCC AGT GTT GGG CAA CAA ATG ATC TTT GAG GAA
phe val thr his pro gly tyr asp thr gly asn gly ile gln leu pro gly thr ser gly gln gln pro ser val gly gln gln met ile phe glu glu
asn glu his
-365
-996
CAT GGT TTT AGG CGG ACC ACA CCG CCC ACA ACG GCC ACC CCC ATA AGG CAT AGG CCA AGA CCA TAC CCG CCG AAT GTA GGT GAG GAA ATC CAA ATT GGT
his gly phe arg arg thr thr pro pro thr thr ala thr pro Ile arg his arg pro arg pro tyr pro pro asn val gly glu glu ile gln ile gly
ala val leu leu asp val
-332
-897
CAC ATT CCC AGG GAA GAT GTA GAC TAT CAC CTG TAC CCA CAC GGT CCG GGG CTC AAT CCA AAT GCC TCT ACA GGA CAA GAA GCT CTC TCT CAG ACA ACC
his ile pro arg glu asp val asp tyr his leu tyr pro his gly pro gly leu asn pro asn ala ser thr gly gln glu ale leu ser gln thr thr
val gly val
-299

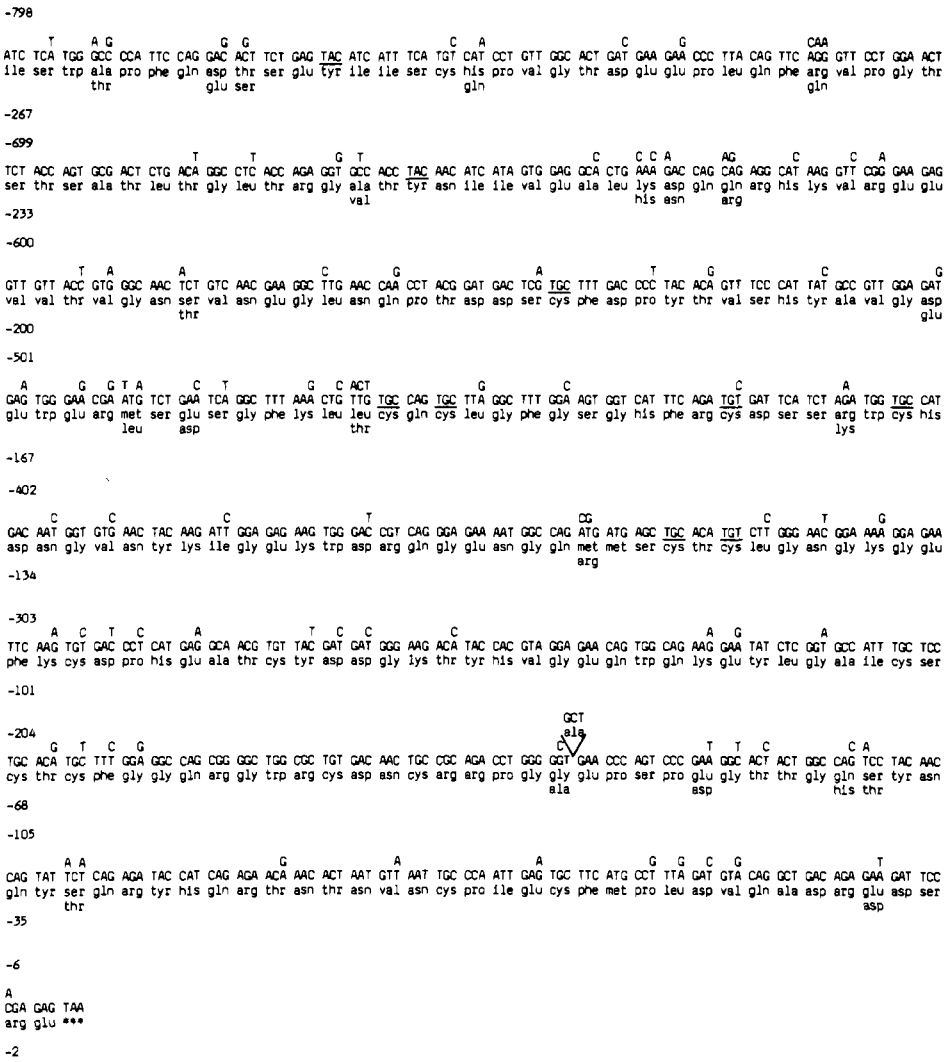


FIGURE 3: Comparison of human and rat cDNAs for cellular fibronectin. First line: nucleotide sequence in the rat cDNAs where this differs from that of the human. Second line: nucleotide sequence obtained from the human cDNAs for cellular fibronectin. Third line: amino acids encoded by the human cDNAs. Fourth line: amino acid sequence for the rat cellular fibronectin where this differs from that of the human. White triangles indicate region of alternate splicing in man (Kornblihtt et al., 1984a) (region not compared to rat). Black triangles indicate region of alternate splicing in rat (Schwarzbauer et al., 1983). Underlined amino acids refer to those conserved amino acids that best align the consensus sequences of type I and type III homologies. (Asterisk) Amino acid difference caused by possible codon recomposition in rat. Numbering system signifies amino acids, below the line, and nucleotides, above the line. The first nucleotide of the stop codon designates nucleotide 0; the stop codon designates codon 0.

map of this clone is identical with that of the 3′-end portions of a recently published human fibronectin cDNA clone (pFH1) (Kornblihtt et al., 1983). Nucleotide sequence of selected areas matched with those derived from pFH1.

Clones from positive cDNA pools were screened again by using the nick-translated *Eco*RI insert of Hf771 as a probe, and nine positive clones were obtained. Restriction endonuclease mapping of plasmids isolated from these colonies showed that they contained cDNA inserts of 700–2000 nucleotides. The 5′ end of the inserts from five independent clones were almost identical with that of Hf771, suggesting that this region of mRNA may contain secondary structure which causes the reverse transcriptase to preferentially stop at this point. One clone (FN 1646) was found to extend 300 bp further to the 5′ end. Positive pools were screened again with an *Eco*RI fragment derived from the 5′ end of FN 1646. By successive screening of the clones, we isolated two additional overlapping cDNA clones, FN 421 and FN 18 (Figure 2).

*Nucleotide Sequence Analysis of Fibronectin cDNAs.* Nucleotide sequences were obtained from four overlapping cDNA clones (Figure 3). Together the four cDNAs endoded 794 amino acid (aa) residues of the C-terminus and approximately 700 bp of the 3′-noncoding region. A restriction endonuclease map was generated and used to develop a strategy for nucleotide sequencing (Figure 2). With the exception of the 3′-noncoding portion of the gene, all sequences were confirmed by sequencing both strands of the DNA. Restriction maps of five smaller cDNAs obtained during the same cloning experiments were consistent with the data presented here. In the overlapping regions, there is 100% homology among all four clones in areas sequenced. There was no evidence of any nucleotide rearrangement during the preparation of the cloned cDNAs as has been previously encountered (Lehrach et al., 1978, 1979; Fuller & Boedtker, 1981).

*Comparison of Nucleotide Sequences with Those Derived from a Human Tumor Cell Line.* Other investigators have published sequences derived from a human tumor cell line (Kornblihtt et al., 1983, 1984a) which coincide with regions covered by our clones. Briefly, these investigators had shown that, in a human carcinoma cell line, there are mRNAs present in two forms: a major and a minor species. The minor species, as in our clones, contains an insert of 270 bp. A comparison (not shown) of this insert region and regions further down-

stream (aa −100 to aa −1, Figure 3) with our data, which covers a single type I repeat of the fibrin binding domain as well as the cysteine dimer region, exhibits complete identity.

*Comparison of Amino Acid and Nucleotide Sequences between Rat and Man.* Our human nucleotide sequence overlapped those published for rat (Schwarzbauer et al., 1983). The region compared includes the 360 bp sequence from rat, which is differentially spliced and leads to as many as three RNA species. A high degree of conservation was noted at both the nucleotide and amino acid level. There are a total of 243 nucleotide substitutions resulting in 54 of 704 amino acids differing between the two species, in the regions compared. Most of the base differences either do not change the amino acid or are conservative because they do not change the charge distribution along the chain. The largest conserved stretch at the base level is 60 bp between base pair −794 and base pair −853. Long stretches of conserved amino acids are more common, the longest being 99 amino acids long (residue −363 to −461).

Common amino acid substitutions include serine to threonine exchanges and isoleucine to valine exchanges. In only one case do these amino acids substitute for any amino acid other than the ones listed. Together, they make up nearly a third of the total amino acid changes. Other differences include two extra methionines present in the human sequence located at amino acid positions −163 and −114, where in rat these positions are occupied by leucine and arginine, respectively.

Absolute conservation of cysteines was observed throughout, including the single cysteine residue located in the last type III homology of the heparin binding region. These cysteines are connected by vertical lines in Figure 4 with the exception of the single cysteine residue. As with the cysteines, the aromatic residues that align the consensus sequence of the type III homologies are conserved and are also connected by vertical lines.

*Comparison of Amino Acid Sequences of Human with Amino Acid Sequences from Rat and Calf.* The amino acid sequences deduced from human nucleotide sequence cover the edge of the cellular-binding domain, the heparin-binding domain, the fibrin-binding domain. They have been arranged into five type III and three type I homologies in Figure 4. Select amino acid sequences for bovine fibronectin have been determined by protein analysis (Skorstengaard et al., 1982, 1984; Petersen et al., 1983). Therefore, it was possible to compare the primary structure of the bovine FN with that derived from human and rat cDNAs.

There is an absolute conservation of 27 amino acids at the C-terminal end when the human sequences are compared to rat. This degree of homology expands to 38 amino acids when the same region is compared to bovine (Figure 4). This region contains two half-cystine residues which form interchain disulfide bonds in the dimer of plasma fibronectin. The derived amino acid sequence is identical with the recently published sequence of human plasma fibronectin (Garcia-Pardo et al., 1984). This high degree of conservation is also noted throughout the type I and type III homologies.

A hypervariable region between the end of the type I homologies and the C-terminal fragment noted previously by Kornblihtt et al. (1983) was immediately apparent. Within a 12 amino acid stretch (aa −49 to −38 in Figures 3 and 4), six amino acids differ from bovine, where five differ from rat, along with the insertion of an alanine in the rat sequence. Only two of the differences these species have with man lie on the same amino acid, leaving only four amino acids unaltered in the comparison. A similar divergence is shown in the area from amino acid −637 to amino acid −664 where the excision of the human insert would define the boundary of the heparin and cellular binding domains. Here the rat sequence is much more divergent with eight amino acid changes and bovine only with four. This latter stretch which includes these changes, however, is larger, such that the percentage difference when compared to the C-terminal hypervariable region is smaller. Other regions which are divergent to a lesser degree might be found near amino acid −163 and mark the boundary between the type I and type III repeats. As yet, there is not enough data to substantiate this.

*Comparison of Differential Splicing Regions of Rat and Man.* The differential splicing region reported for man (Kornblihtt et al., 1984a) is not found in rat, so comparison was not possible. The sequences surrounding this region, however, show an interesting divergence which may be best envisioned in Figure 3. In man, the 3′-junction amino acid changes from an alanine to a threonine because of the position of the splicing site within the first base of the codon.

The alternate splicing region in rat (Schwarzbauer et al., 1983) when compared to those sequences derived from man showed a striking similarity at the base level as well as the amino acid level. The longest homology, which is made up of 60 bp, is located in the region surrounding the 3′-splice junction. In the region between both 5′-splice junctions, a conserved stretch of 47 bp was found. If four mismatches are allowed, this stretch increases to 116 bp. At the amino acid level, the 5′ junction is contained within the largest conserved stretch, 99 amino acids, when the two species were compared.

## Discussion

The data presented here provide the first extensive information about the primary structure of more than a third of cellular fibronectin from human fibroblasts. Comparison of sequences for fibronectin from different species has shown a dramatic similarity at the base level and even greater similarity at the amino acid level. This conservation is more pronounced in regions containing repeats of internal homology (e.g., type I and type III homology), substantiating the data previously developed by amino acid and nucleotide sequencing. For example, it has been reported that, in the cell- and collagen-binding domain, the interspecies homology ranges between 88 and 96% (Petersen et al., 1983; Pierschbacher et al., 1982; Skorstengaard et al., 1984). Assuming that structural features that are conserved through evolution are important for biological functions, most of the structure in each FN domain has a critical biological role. The observation of conserved functional domains is indeed consistent with the multifunction feature of fibronectin.

Alignment of type homologies in fibronectin showed the absolute conservation of cysteines as well as basic and aromatic residue groups recently hypothesized for heparin binding (Schwarzbauer et al., 1983). Conservation of cysteines is not surprising as interchain disulfide bridges are important in the formation of tertiary structure. The basic and aromatic residues, which are potential carbohydrate receptor sites, could very likely be attachment regions for heparin as are select regions further 5′ responsible for cellular binding. It is interesting to note that, recently, the cell binding activity has been attributed to a tetrapeptide within this domain (Pierschbacher & Ruoslahti, 1984). The conservation of other amino acids within the repeats are less apparent, because the degree of identity between each type I homology ranges from 18 to 60% and that between each type III homology is about 30% (Petersen et al., 1983). It is possible that this conserved

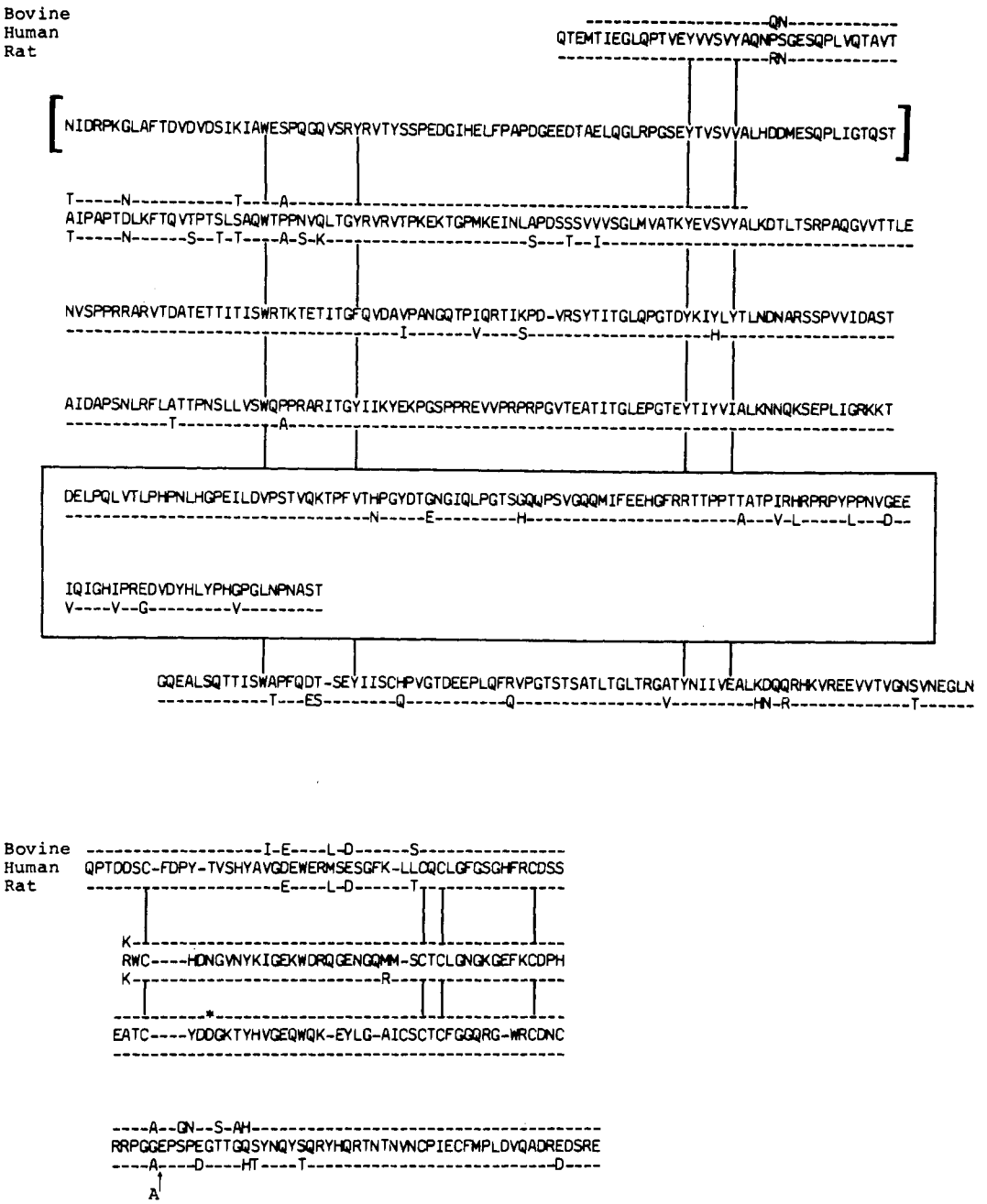



FIGURE 4: Comparison of amino acid sequences for cellular fibronectin among bovine, human, and rat arranged in type III and type I homologies. First line: amino acid sequences from bovine where this differs from that of human (Petersen et al., 1983; Skorstengaard et al., 1982). Second line: amino acid sequences encoded by cDNAs for human. Third line: amino acid sequences encoded by cDNAs for rat where this differs from that of human (Schwarzbauer et al., 1983). (Asterisk) Asparagine residue in bovine sequence has been replaced, as it has been shown to be an aspartate residue through sequencing of a cDNA for this region (Kornblihtt et al., 1983). Brackets indicate region of differential splicing found so far only in man. Boxed area refers to region of differential splicing in rat. Vertical lines refer to those amino acids that have shown to be highly conserved within the individual repeat types. Regions available for comparison are presented as dotted lines or replacement sites.

region serves some additional purpose such as interacting with other domains of the same chain or the correct alignment of the dimeric or multimeric structure. Conversely, there seems distinct regions of amino acid divergence located in areas that define boundaries of structural domains. This was first seen while comparing sequences from bovine and man at the junction between the last repeat of type I homology and the $M_r$ 3000 fragment of the C-terminus (Kornblihtt et al., 1983). With the introduction of the rat data, the divergence in the same region became amplified. In fact, the only amino acid insertion occurs in this region. Analysis of the region which hypothetically defines the boundary of the heparin and cellular binding regions again resulted in a divergent cluster between aa −637 and aa −658. Clusters of divergent amino acids between these binding domains show their function not to be as critical but may serve merely as a link.

Comparison of the amino acid sequence in the differential spliced region reported for rat showed this region also to be conserved. This difference region has been implicated to account for the apparent molecular weight difference of the two chains of plasma fibronectin. The observed interspecies conservation again suggests that this region is essential for normal function. If different fibronectin mRNAs are found to encode for different forms of fibronectin, the conservation of such an organized structure might be essential for the full expression of the biological activity of this complex protein.

Since submission of this paper, Kornblihtt et al. (1984b) published nucleotide sequences of 4383 bp of fibronectin cDNAs from a human tumor cell line. Our sequences in Figure 2 are identical with those reported for the human tumor fibronectin mRNA except for a region surrounding the alternate splicing region found in rat (Schwarzbauer et al., 1983). Our clone FN 421 contains a 360 bp insert between the last two repeats of type III homology. Thus, it is homologous to one of the three mRNA species reported in rat but different from the one reported in the human tumor in which a 267 bp (89 amino acids) insert was present. Recently, Tamkun et al. (1984) reported that the three fibronectin mRNAs in rat were generated by alternative splicing of a single exon. Close inspection of the 5′ junction of the different regions between fibroblasts and tumor cells reveals that the normal human sequence contains a possible donor splice site recognition sequence (AG/GTGAGG, −919 to −912 in Figure 2). This sequence is different from that seen in rat by a replacement of the second guanine by an adenine, hence destroying the concensus sequence. It is tempting to speculate that the mRNA species found in the tumor cell line may be unique to human and are generated by the excision of the latter half of this exon via an RNA splicing mechanism. Recently, we have performed several S1 nuclease mapping experiments using mRNAs prepared from both normal and SV40 transformed human fibroblasts and have also detected a mRNA species with a 267 bp insert in both cell lines (data not shown).

## Acknowledgments

We are grateful to Drs. Francesco Ramirez and Darwin J. Prockop for their stimulating discussions and support during the course of this study. We thank Dr. Richard O. Hynes for helpful discussions, especially in pointing out the extra donor splice site in the human sequences.

## References

Bernard, M. P , Myers, J. C., Chu, M.-L., Ramirez, F., Eikenberry, E. F., & Prockop, D. J. (1983) *Biochemistry 22*, 1139.

Chu, M.-L., Myers, J. C., Bernard, M. P., Ding, G., & Ramirez, F. (1982) *Nucleic Acids Res. 10*, 5925.

Fagan, J. B., Yamada, K. M., de Crombrugghe, B., & Pastan, I. (1979) *Nucleic Acids Res. 6*, 3471.

Fuller, F., & Boedtker, H. (1981) *Biochemistry 20*, 996.

Furcht, L. T. (1983) *Mod. Cell Biol. 1*, 53.

Garcia-Pardo, A., Pearlstein, E., & Frangione, B. (1983) *J. Biol. Chem. 258*, 12670.

Garcia-Pardo, A., Pearlstein, E., & Frangione, B. (1984) *Biochem. Biophys. Res. Commun. 120*, 1015.

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63.

Hynes, R. O. (1981) *Cell Biology of Extracellular Matrix* (Hay, E. D., Ed.) p 295, Plenum Press, New York.

Hynes, R. O., & Yamada, K. M. (1982) *J. Cell Biol. 95*, 369.

Kornblihtt, A. R., Vibe-Pedersen, K., & Barralle, F. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3218.

Kornblihtt, A. R., Vibe-Pedersen, K., & Barralle, F. E. (1984a) *EMBO J. 3*, 221.

Kornblihtt, A. R., Vibe-Pedersen, K., & Barralle, F. E. (1984b) *Nucleic Acids Res. 12*, 5853–5868.

Kupersztoch, Y. M., & Helinski, D. R. (1973) *Biochem. Biophys. Res. Commun. 54*, 1451.

Lehrach, H., Frischauf, A. M., Hanahan, D., Wozney, J., Fuller, F., Crkvenciakov, R., Boedtker, H., & Doty, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5417.

Lehrach, H., Frischauf, A. M., Hanahan, D., Wozney, J., Fuller, F., & Boedtker, H. (1979) *Biochemistry 18*, 3416.

Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184.

Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499.

Oldberg, A., Linney, E., & Ruoslahti, E. (1983) *J. Biol. Chem. 258*, 10193.

Pande, H., & Shively, J. E. (1982) *Arch. Biochem. Biophys. 213*, 258.

Peacock, S. L., McIver, C. M., & Monohan, J. J. (1981) *Biochim. Biophys. Acta 655*, 243.

Pelham, H. R., & Jackson, R. J. (1976) *Eur. J. Biochem. 67*, 247.

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*, 137.

Pierschbacher, M. D., & Ruoslahti, E. (1984) *Nature (London) 309*, 30.

Pierschbacher, M. D., Ruoslahti, E., Sundelin, J., Lind, P., & Peterson, P. A. (1982) *J. Biol. Chem. 257*, 9593.

Pierschbacher, M. D., Hayman, E. G., & Ruoslahti, E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1124.

Rave, N., Crkvenjakov, R., & Boedkter, H. (1979) *Nucleic Acids Res. 6*, 3559.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237.

Rougeon, F., & Mach, B. (1977) *J. Biol. Chem. 252*, 2209.

Roychoudhury, R., Jay, E., & Wu, R. (1976) *Nucleic Acids Res. 3*, 101.

Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R., & Hynes, R. O. (1983) *Cell (Cambridge, Mass.) 35*, 421.

Skorstengaard, K., Thogersen, H. C., Vibe-Pedersen, K., Petersen, T. G., & Magnusson, S. (1982) *Eur. J. Biochem. 128*, 605.

Skorstengaard, K., Thogersen, H. C., & Petersen, T. E. (1984) *Eur. J. Biochem. 140*, 235.

Tamkun, J. W., Schwarzbauer, J. E., & Hynes, R. O. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5140–5144.

Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201.

Wickens, M. P., Buell, G. N., & Schimke, R. T. (1978) *J. Biol. Chem. 253*, 2483.

Yamada, K. M. (1983) *Annu. Rev. Biochem. 52*, 761.