Isolation and characterization of a vinculin cDNA from chick-embryo fibroblasts

Glyn J. PRICE, Peter JONES, Matthew D. DAVISON, Bipin PATEL, Ian C. EPERON and David R. CRITCHLEY*

Department of Biochemistry, University of Leicester, Leicester LE1 7RH, U.K.

A chick-embryo fibroblast λ gt11 cDNA library was screened with affinity-purified antibodies to chick gizzard vinculin. One recombinant was purified to homogeneity and the fusion protein expressed in *Escherichia coli* strain C600. The fusion protein was unstable, but polypeptides that reacted with vinculin antibodies, but not non-immune immunoglobulin, were detected by Western blotting. The recombinant contained a single *Eco*RI fragment of 2891 bp with a single open reading frame. The deduced protein sequence could be aligned with that of six CNBr-cleavage peptides and two tryptic peptides derived from chicken gizzard vinculin. AUG-247 has tentatively been identified as the initiation codon, as it is contained within the consensus sequence for initiation sites of higher eukaryotes. The cDNA lacks 3' sequence and encodes 74% of the vinculin sequence, presuming the molecular mass of vinculin to be 130 000 Da. Analysis of the deduced sequence showed no homologies with other protein sequences, but it does display a triple internal repeat of 112 amino acid residues covering residues 259–589. The sequences surrounding the seven tyrosine residues in the available sequence were aligned with the tyrosine autophosphorylation consensus sequence found in protein tyrosine kinases. Tyr-822 showed a good match to this consensus, and may represent one of the two major sites of tyrosine phosphorylation by pp60^{v-src}. Northern blots showed that the 2.89 kb vinculin cDNA hybridized to one size of mRNA (approx. 7 kb) in chick-embryo fibroblasts, chick smooth muscle and chick skeletal muscle. Southern blots revealed multiple hybridizing bands in genomic DNA.

INTRODUCTION

In vitro, the interaction between cells and adhesive glycoproteins such as fibronectin elicits a transmembrane signal leading to reorganization of intracellular actin into microfilament bundles (Hynes *et al.*, 1982). In many cases these bundles terminate at specialized contacts between the ventral surface of the cell and the substrate that are termed adhesion plaques. *In vivo*, similar structures, collectively known as adherens junctions, are found at sites of intercellular contact (Geiger *et al.*, 1985; Drenckhahn & Franz, 1986) and between cells and the extracellular matrix (Geiger *et al.*, 1985; Drenckhahn & Wagner, 1986). The cytoskeletal reorganization is thought to be triggered by interaction of the cell-binding domain of fibronectin (Pierschbacher & Ruoslahti, 1984) with a transmembrane cell-surface receptor. Of a number of candidate receptors (Yamada *et al.*, 1985), a family of 140 kDa glycoproteins has clearly been found to show a close spatial distribution with fibronectin in adhesion plaques (Damsky *et al.*, 1985) and to bind to fibronectin (Pytela *et al.*, 1985), and the primary sequence of one of these glycoproteins has been determined (Tamkun *et al.*, 1986).

Immunocytochemical studies have shown that the proteins talin (215 kDa) (Burrage & Connell, 1983), vinculin (130 kDa) (Geiger *et al.*, 1980), and α -actinin (100 kDa) (Lazarides & Burrage, 1975) are wholly or in

part localized on the cytoplasmic face of the adhesion plaque (Geiger *et al.*, 1985) and may serve to complete the link between the transmembrane fibronectin receptor and microfilamentous actin (Heath, 1986). Indeed, experiments have shown that the 140 kDa glycoprotein can bind directly to talin, but not to vinculin or α -actinin (Horwitz *et al.*, 1986). Western-blot experiments have in turn provided evidence that vinculin binds to talin (Burrage & Mangeat, 1984). Although α -actinin is known to bind to the ends of F-actin filaments (Bennett *et al.*, 1984), the precise nature of the link between this complex and the other proteins of the adhesion plaque remains uncertain.

Much of the current interest in the molecular details of these interactions stems from the observations that adhesion plaques are rapidly disrupted following tumour-virus transformation (Rohrschneider *et al.*, 1982), or exposure of cells to tumour promoters (Meigs & Wang, 1986) or platelet-derived growth factor (Herman & Pledger, 1985). Vinculin is a substrate for both protein kinase C (Werth & Pastan, 1984) and the tyrosine-specific protein kinases encoded by the transforming genes of a number of retroviruses, including pp60^{v-src} (Sefton *et al.*, 1981; Rohrschneider *et al.*, 1982; Antler *et al.*, 1985; Kellie *et al.*, 1986b). It was therefore tempting to speculate that adhesion-plaque integrity could be modulated through phosphorylation. However, we (Kellie *et al.*, 1986a,b) and others (Rohrschneider *et al.*,

Abbreviations used: PAGE, polyacrylamide-gel electrophoresis; SSC, standard saline citrate (0.15 M-NaCl/0.15 M-sodium citrate buffer, pH 7.0); poly(A)⁺ RNA, polyadenylated RNA.

* To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00312.

1982; Antler *et al.*, 1985) have found no simple relationship between the tyrosine-specific phosphorylation of vinculin by pp60^{v-src} and the disruption of adhesion plaques. It remains possible that the reported phosphorylation of talin (Pasquale *et al.*, 1986) and the fibronectin receptor (Hirst *et al.*, 1986) by pp60^{v-src} will be more closely correlated with such events.

Because of the obvious importance of vinculin and α -actinin in cell adhesion, and the paucity of information available on the structure-function relationships within these proteins, we have recently initiated a programme to determine their primary sequence by cDNA cloning. The fascinating observation that the v-fgr (Naharro *et al.*, 1984) and trk (Martin-Zanca *et al.*, 1986) oncogenes contain actin-related and tropomyosin-related sequences respectively also highlights the importance of determining the sequence of other cytoskeletal proteins. We now report the isolation and characterization of a vinculin cDNA that encodes 74% of the coding sequence of the protein. Our studies on α -actinin cDNAs are reported elsewhere (Baron *et al.*, 1987).

MATERIALS AND METHODS

Isolation of vinculin cDNAs

A chick-embryo fibroblast cDNA library (large insert) constructed in the expression vector λ gt11 (Tamkun *et al.*, 1986), generously given by Dr. R. O. Hynes (Massachusetts Institute of Technology, Cambridge, MA, U.S.A.), was screened for recombinants of interest by using minor modifications to published methods (Huynh *et al.*, 1985). Rabbit antibodies against SDS/PAGE-purified chicken gizzard vinculin were affinity-purified as previously described (Kellie *et al.*, 1986b), and used at a dilution sufficient to detect 1 ng of antigen. Bound antibody was detected by using pig anti-rabbit IgG followed by peroxidase-rabbit anti-peroxidase complex (Dakopatts, Copenhagen, Denmark). Positive plaques were purified to homogeneity and re-screened with rabbit antibodies to chicken gizzard vinculin kindly made available to us by Dr. Keith Burridge (University of North Carolina, Chapel Hill, NC, U.S.A.). Large-scale preparations of recombinant λ bacteriophage were prepared by CsCl-gradient centrifugation, and the sizes of the inserts were determined by agarose-gel electrophoresis of *Eco*RI-digested λ DNA (Maniatis *et al.*, 1982). Agarose-gel-purified inserts were labelled with [α -³²P]dCTP by a random priming method (Feinberg & Vogelstein, 1984). Plaque screening by DNA-DNA hybridization was by standard procedures (Maniatis *et al.*, 1982), with a final wash in 0.1 $\times$ SSC containing 0.1% SDS at 65 °C.

Western-blot analysis of fusion proteins

Fusion proteins were obtained by growing the recombinant λ phage in *Escherichia coli* strain C600, which lacks the *supF* mutation required by λ gt11 for lysis. Cells were grown to an A_{450} of 0.64, and 1 ml of the culture (approx. 2×10^8 cells) was pelleted by centrifugation at 12000 *g* for 2 min in a Microfuge. The bacterial pellets were resuspended in 360 μ l of L-broth (Maniatis *et al.*, 1982) containing 10 mM-MgSO₄, and 40 μ l of the λ gt11 recombinant was added at a multiplicity of 5–10. Cells were incubated at 37 °C for 3 h with occasional shaking, pelleted by centrifugation, and dissolved in

100 μ l of electrophoresis sample buffer. Half the sample was analysed by SDS/PAGE (Laemmli, 1970), and proteins were transferred to nitrocellulose sheets (Towbin *et al.*, 1979). Proteins were detected by staining the filters with undiluted Ponceau S (BDH Chemicals, Poole, Dorset, U.K.), the positions of the fusion proteins marked with ink and the filters blocked with 3% dried milk/0.1% Tween 20 in 10 mM-Tris/0.9% NaCl, pH 7.5. Filters were incubated with rabbit anti-vinculin antibodies, and the bound antibodies were detected as for plaque screening.

DNA sequencing

DNA fragments were subcloned into M13mp18 (Norrandar *et al.*, 1983) with the *E. coli* strain JM103 as host, and sequenced by the dideoxy chain termination method of Sanger *et al.* (1977). Partial sequence was obtained by using standard shotgun cloning techniques with the enzymes *Sau*III, *Alu*I and *Hae*III. Sequences derived from the ends of non-overlapping contigs were used to design synthetic 17-mers. The oligonucleotides were then used as primers on the full cDNA to extend the sequence from each contig into neighbouring contigs (Brenner & Shaw, 1985), and also to confirm the sequence in any regions of ambiguity. The cDNA was

Fig. 1. Analysis of fusion proteins expressed by putative vinculin recombinants by using SDS/PAGE and Western blotting

A λ gt11 recombinant that reacted strongly with rabbit anti-(chicken gizzard vinculin) immunoglobulin in a plaque assay was expressed in *E. coli* strain C600, and the proteins were separated by SDS/PAGE (7.5% gel). Either the gel was stained with Coomassie Blue (tracks a–c) or the proteins were transferred to nitrocellulose for immunostaining (tracks d and e) as described in the Materials and methods section. Track a, C600-cell proteins; track b, C600 cells infected with wild-type λ gt11; track c, C600 cells infected with recombinant λ gt11. Tracks d and e were stained with affinity-purified rabbit anti-(chicken gizzard vinculin) immunoglobulin (2 μ g/ml). Track d, C600 cells infected with wild-type λ gt11; track e, C600 cells infected with recombinant λ gt11. Molecular-mass markers are shown on the left.

1	CGCCGCCGGGGGAAGGCGAGAGGTCACGAGTTCGAGTCCCGCCGTCGCCGGAACTGCC	1501	GCAGAAATGTGTGAGGAGCCTAAAGAAAGAGATGATATCCTTCCTCTGGGGGAAATC	438
61	GGTCCGACCCGCGCAGCCGCTGAGGGGTGAGGAGTTCGGGCGGCGTCTTCTCTAT		A E L C E E P K E R D D I L R S L G E I	
121	TTCTTCTCGTTTTTCTCTCTCGGCGGAGGCTCGGCTTCTCCGTCGGGCGCCAAACCG	1561	TCTGCTCTACAGCTAAGCTGTGAGATCTTGCACGATGGGAAAGGCGACTCTCTGAG	458
181	CGGGGACAGAGTTTTCCAGGAGTTGTGCGGACTTCGGGCTCGGCGCCCTGCCCGCT		S A L Q L S C Q I L R R H G K G D S P E	
241	GCCGCCATGCCCGTCTTCCACACGCGACCATCGAGAGCATCTTGGAGCCGCTGGCTCAG	1621	GCCCGTCATTTGGCCAAAGTAAGTACATCACTTTCAGAACTTACAGCTCAAAACAAAC	478
	M P V F H T R T I E S I L E P V A Q		A R A L A K Q I A T S L Q N L Q S K T N	
301	CAGATCTCCACCTGGTCATCATGACAGGAGGGGGAGGTAGACGGCAAGGCCATCCCG	1681	AGAGCTGTAGCAATACTAGACAGTAAAGCTGTCTTCTTGGAGGGCAAGATTGAG	498
	Q I S H L V I M H E E G E V D G K A I P		R A V A N T R P V K A A V H L E G K I E	
361	GACCTCACGCCCCGCTGTCGGCGTGCAGGCGCTGTGACGAACTGGTGGCGGTTGGA	1741	CAAGCTCAGAGGTGGATAGATAATCTACAGTGTATGATGGGGAGTAGGCCAGGCAGCA	518
	D L T A P V S A V Q A A V S N L V R V G		Q A Q R W I D N P T V D D R G V G Q A A	
421	AAAGAACTGTGCAGACACAGAGACAGATCTTGAAGGAGATAGCCACGACATTC	1801	ATTCGGGTTTGGTTCAGAGGTCGTCTGCTAGCCAAATGTCATGATGGGACCTTATCGT	538
	K E T V Q T T E D Q I L K R D M P P A F		I R G L V A E G R R L A N V M M G P Y R	
481	ATCAAAGTAGAGATGCTGCGACCAAGCTCGTTCGAGCAGCCAGATGCTGCAAGCAGAT	1861	CAGGACCTGCTTCCAAATGTGACCGTGTAGACAGCTGGCTGTGCTGCTGACCTT	558
	I K V E N A S T K L V R A A Q M L Q A D		Q D L L A K C D R V D Q L A A Q L A D L	
541	CTTATTCAGTACAGCTCGTACCTAATGATGGATCAAGAGGCAATCTTCTGGA	1921	GCAGCAAGGGGGAAGGAGTCTCTCAGGCTAGGGCAATGTCTGCTCAGCTCAGGAC	578
	P Y S V P A R D Y L I D G S R G I L S G		T A R G E G E S P Q A R A I A A Q L Q D	
601	ACATCAGACTTACTTGTACATTTGATGAAGCGAGGTCCTGAAATCATCGGTGCTGC	1981	TCCTGAAGGATCTCAAAGCAGGATGCAAGAGCAATGACCCAGGAGGTGTCTGATGT	598
	T S D L L L T F D E A E V R K I I R V C		S L K D L K A R M Q E A H T Q E V S D V	
661	AAAGGAATTTGGAATATCTGACTGTGCGAGAAGTAGTAGACTATGGAGATTGGTG	2041	TTTAGTGACACTCAACTCTTATTAAGTTGTAGCAGTAGCAGCCACTGCTCTCTGAT	618
	K G I L E Y L T V A E V V E T M E D L V		F S D T T T P I K L L A V A A T A P S D	
721	ACATACAAAGAATCTAGGCGCAGGAATGACAAAGATGGCAAAATGATCGATGAGA	2101	ACTCCAATAGAGAAGAGGTGTTGAAGAAGAGCAGCAAAATTTGAAACCATGCTGT	638
	T Y T K N L G P T G M K A K M I D E R		T P N R E E V F E E R A A N F E N H A A	
781	CAACAGGAATTAACATCAGGAACATAGGTTATGCTGGTGAATCATGAATACTGTG	2161	AGACTGGAGCAACAGCAGAAAAAGCAGCTGCAAGTGGAACTGCTAATAAACTACTGTG	658
	Q Q E L T H Q E H R V M L V N S M N T V		R L G A T A E K A A A V G T A A N K T T V	
841	AAGGAGCTATTGCTGTACTTATTTCCAGCTATGAAGATCTTTGTAACCAAAAAACACT	2221	GAAGGCATTGAGCAACAGTCAAACTGCAAGGGAGCTTACACCAAGGTAGTATCGCT	678
	K E L L P V L I S A M K I F V T T K N T		E G I Q A T V K S A R E L T P Q V V S A	
901	AAAAGCAGGGAATAGAAGGCTTGAAGATCGCAATTTACAGTAGAAGAAATGAGT	2281	GCTCGAATCTCTGAGAAATCTGGAATCAAGCTGCTTATGAGCATTTTGAGACAAATG	698
	K S Q G I E A C T K N R N F T V E K M S		A R I L L R N P G N Q A A Y E H F E T M	
961	GCTGAGATAATGAATAATCCGTGATTACAACCTACTTCTGGGATGAAGATGCTGG	2341	AAAAACCACTGGATTGATAATGTAGAAAAGTACAGGGCTGGTGGATGAGGCCATCGAT	718
	A E I N E I I R V L Q L T S W D E D A M		K M H M I D N V E K M T G L V D E A I D	
1021	CCAGCAAGGACACTGAAGCCATGAAGAGCTTTAGCCCTAATGATTAAGATGAAC	2401	ACCAAGTCTCTGTTGGATGATCAGAAAGGCTATTGAAGGATCTTGTAAATGTAA	738
	A S K D T E A M K R A L A L I D S K M N		T K S L L D A S E E A I K R D L D K C K	
1081	CAGGCAAAAGGCTGGCTGAGAGATCCAAAGCAGCTCCAGGGGATGCTGGTGAGCAAGCA	2461	GTGCAATGGCCAAATATGCAACCTCAGATGCTGGTAGCTGGAGCCAGCATGCTAGA	758
	Q A K G M L R D P N A P P G D A G E Q A		V A M A N M Q P Q H L V A G A T S I A R	
1141	ATAAGGCAGATCTTGTATGAAGCTGGAAGCAGGAGATTTGTGTCAGGCAAGAACGC	2521	CGAGCAACCGAATCTGCTTGTGGCAAAAGGAGGTTGAAATTCAGAGACCTTAA	778
	I R Q I L D E A G K A G E L C A G K E R		R A N R I L L V A K R E V E N S E D P K	
1201	AGGGAGTCTGGGAACATCTGGAAGCTTTGGCCAAATGACTGATCACTTGTCTGATCTC	2581	TTGAGGAGGCTGTTAAAGCAGCTTCTGATGAGCTGAGCAAAACCATATCCAGGTGTA	798
	R E I L G T C K T L G Q H T D Q L A D L		F R E A V K A A S D E L S K T I S P M V	
1261	CGAGCTAGAGGACAGGGTGTACACGATGGCAAGGAGGAGGACAGGAGTGTACAA	2641	ATGGATGCTAAAGCTGTAGCAGGAATATCTCTGATCTGGTTTGCAGAGAGTTCTTG	818
	A R G Q G A M Q K A Q Q V S Q		M D A K A V A G N I S D P G L Q K S F L	
1321	GGCTTGATTTGCTCACTGCAAAAGTGGAATGCAGCCCGGAAATGGAGGCGATGACA	2701	GATTCGGATACAGGATCTGGGAGCTGTGGCCAAAGTCAGAGAGCGCTTTCAGCCTCAG	838
	S L D L L T A K V E N A A R K L E A M T		D S G Y R I L G A V A K V R E A F O P Q	
1381	AACTCTAAGCAGGCTATTGCAAGAAGATGATGCTGCTCAGAACTGGCTTGGCGATCCC	2761	GAGCCAGACTTTCCTCTCTCTCTGACCTGAGCATCTCCATCTGACTGATGAGCTT	858
	N S K Q A I A K K I D A A Q N M L A D P		E P D F P P P P P P D L E H L H L T D E L	
1441	AACGGTGAAGTGAAGGAGAAGAACACATTCGAGGAATATGCTGAGGCAAGGAAAGTT	2821	GCTCTCCAAAACCAACCTCCAGAGGTGAGGTTCCCAACCCAGACCCACCCCT	878
	N G G S E G E E H I R G I M S E A R K V		A P P K P P L P E G E V P P P P P P P P	
		2881	GAAGAAAGG	881
			E K K	

Fig. 2. Sequence of a chick-embryo fibroblast vinculin cDNA and the deduced amino acid sequence

Numbers on the left refer to the nucleotide sequence, those on the right to the amino acid sequence. The putative initiation codon is at base 247. The position of the sequences determined on CNBr-cleavage and tryptic peptides derived from chicken gizzard vinculin are shown by the continuous lines, with dashes indicating those amino acids in the peptide sequences to which an absolute identity could not be assigned.

sequenced on both strands, with at least two gel readings covering any particular point in the sequence. Sequence alignments were made by using the DB system of Staden (1982b). Both DNA and deduced protein sequences were analysed by using the sequence-analysis software package of the University of Wisconsin Genetic Computer Group.

Southern and Northern blotting

Total RNA was extracted from tissues by using the guanidine isothiocyanate procedure (Maniatis *et al.*, 1982), and poly(A)⁺ RNA was isolated by chromatography on oligo(dT)-cellulose (Collaborative Research, Levington, MA, U.S.A.). DNA was prepared from chick brain by phenol/chloroform extraction and RNAase A digestion by standard procedures (Maniatis *et al.*, 1982). Both RNA and DNA fractionated in agarose gels was transferred to Hybond N nylon membranes (Amersham International, Amersham, Bucks., U.K.) by passive diffusion. Hybridization was

performed at 65 °C for 15–20 h as described by Dalglish *et al.* (1986). Filters were washed to a final stringency of 0.5 × SSC containing 0.1% SDS at 65 °C.

Vinculin purification and sequencing

Vinculin was purified from adult chicken gizzard by the method of Evans *et al.* (1984) and appeared to be > 95% pure by SDS/PAGE. The protein (2 mg) was reduced and pyridylethylated by a modification of the method of Friedman *et al.* (1970). The purified protein was solubilized in 0.3 M-Tris/HCl buffer, pH 7.5, containing 6 M-guanidinium chloride, dithiothreitol was added to a final concentration of 5 mM, and the solution incubated in the dark for 1 h at room temperature. 4-Vinylpyridine was then added to a final concentration of 50 mM, and the incubation was continued for a further 2 h. The solution was then exhaustively dialysed against distilled water and freeze-dried. Cleavage of the protein with CNBr was performed as described by Pappin & Findlay (1984), except that free tryptophan was added (a

The first AUG codon is in position 247 and is contained within the consensus sequence CCGCCAUG(G) for initiation sites of higher eukaryotes (Kozak, 1984). Several other lines of evidence suggest that this is indeed the correct initiation codon. There is a marked difference in codon usage on either side of AUG-247. For example, the preferred codons for alanine in the sequence 5' to the putative initiation codon were GCC and GCG. In the coding sequence the preferred codons were GCT and GCA. The sequence 5' to AUG-247 has a high G + C content (approx. 70%), and this may explain the lack of termination codons. Finally, primer-extension studies have shown that the 5' end of the cDNA is close to the end of the mRNA (results not shown). This putative initiation codon is followed by an open reading frame extending to the 3' end of the cDNA.

Assuming that we have correctly identified the initiation codon, the cDNA encodes a polypeptide chain of 96 162 Da. The published molecular mass of vinculin is 130 000 Da (Evans *et al.*, 1984), and the cDNA would therefore contain 74% of the coding sequence. However, in SDS/PAGE vinculin co-migrates with *E. coli* β -galactosidase, which has a molecular mass of 116 349 Da calculated from the DNA sequence (Fowler & Zabin, 1978). It is therefore possible that we are only lacking a small portion of 3' coding sequence.

Comparison of the deduced sequence for vinculin with protein sequences entered in the data base revealed no significant homologies with other proteins. When the deduced protein sequence of vinculin was compared with itself by using the program DIAGON (Staden, 1982a), a domain of approx. 110 residues was found to be repeated three times within the molecule, extending from about residue 260 to 590 (Fig. 3). A more detailed sequence alignment obtained by using the program BESTFIT (Smith *et al.*, 1983) showed that residues 259–589 consisted of a unit of 112 residues repeated three times within the vinculin sequence (Fig. 4). Of these residues 18% were conserved in all three repeats, and a further 17% of the residues consisted of conservative substitu-

The consensus sequence for autophosphorylation is shown at the top of the Figure, with variations on the consensus shown underneath the sequence. The sequence surrounding the seven tyrosine residues in vinculin were optimally aligned to this sequence by using pad-character positions (*) employed by Hunter & Cooper (1985) to align the different tyrosine kinase autophosphorylation sites. Residues showing absolute homology with the consensus are boxed, and those showing conservative changes according to the scheme used in Fig. 4 are denoted by a line above and below the residue. Two possible alignments are shown for Tyr-537. The first represents the best alignment to the consensus for the tyrosine-autophosphorylation site; the second alignment positions a basic residue seven residues *N*-terminal to Tyr-537. The determined tyrosine-phosphorylation sites for substrates of pp60^{v-src}, e.g. lactate dehydrogenase (LDH), enolase (Cooper *et al.*, 1984) and p36 (calpactin) (Saris *et al.*, 1986), have also been aligned to the consensus for tyrosine autophosphorylation. The enolase sequence was derived from sequencing a tryptic peptide, and is shown in parentheses, since the whole of the primary sequence remains to be determined. The one determined site of phosphorylation in p36 (Tyr-23) and the additional candidate site (Tyr-187) are also shown. Also shown above the autophosphorylation sequence are those residues (+) that are thought to be important in determining viable tyrosine-phosphorylation sites in substrates of tyrosine kinases.

tutions. Residues 678–791 showed partial homology with this repeat structure, and may represent a fourth repeat unit. The structure–function implications of this observation remain to be elucidated.

Vinculin is a substrate for a number of retroviral tyrosine-specific protein kinases (Sefton *et al.*, 1981; Rohrschneider *et al.*, 1982; Antler *et al.*, 1985; Kellie *et al.*, 1986b), and the sequence was therefore analysed for candidate tyrosine-phosphorylation sites. In Fig. 5, the sequences surrounding the tyrosine residues in vinculin have been aligned with part of the consensus sequence for autophosphorylation sites in tyrosine-specific protein kinases (Hunter & Cooper, 1985) and with sequences surrounding phosphorylated tyrosines in substrates for tyrosine kinases. Residues thought to be important in tyrosine-phosphorylation sites are a basic amino acid seven residues *N*-terminal to the tyrosine, and one or more acidic residues in the first five residues *N*-terminal to the tyrosine (Hunter & Cooper, 1985). Of the seven tyrosine residues in the available vinculin sequence, Tyr-822 fits the above consensus sequence. It is also noteworthy that residues 9–14 *N*-terminal to Tyr-822 show significant homology with the consensus sequence of known tyrosine kinases, although there is no homology between vinculin and tyrosine kinases outside this region. The only other candidate is Tyr-537, which has the required basic amino acid but lacks the intervening acidic residue.

Northern-blot analysis of poly(A)⁺ RNA isolated from chick-embryo fibroblasts with the ³²P-labelled

Fig. 6. Northern-blot analysis of vinculin mRNA

Poly(A)⁺ RNA (2 µg) extracted from either chick-embryo fibroblasts (tracks a and b) or chicken gizzard smooth muscle (track c) or skeletal muscle (track d) was separated on a 1% agarose gel containing 2.2 M-formaldehyde. RNA was transferred to a nylon filter, hybridized with the ³²P-labelled 2.89 kb vinculin cDNA at 65 °C for 20 h and washed as described in the Materials and methods section; 18S rRNA and 28S rRNA were used as molecular-size markers. Ori is the gel origin. Tracks b–d were exposed to X-ray film for the same length of time. Track a is a shorter exposure of track b.

Fig. 7. Hybridization of chicken genomic DNA with a vinculin cDNA probe

Chick brain DNA (4 µg) was digested with (track a) *Bam*HI, (track b) *Eco*RI, (track c) *Hind*III, (track d) *Pst*I or (track e) *Sst*I and fractionated on a 0.8% agarose gel. The DNA was transferred to a nylon filter, hybridized with the ³²P-labelled 2.89 kb vinculin cDNA at 65 °C for 20 h and washed as described in the Materials and methods section. Molecular-size markers at the left are in kb. Ori is the gel origin.

vinculin cDNA is shown in Fig. 6 (lanes a and b). Under conditions of high stringency, the probe hybridized to a single discrete mRNA species of about 7 kb. The labelled cDNA also hybridized to a single mRNA species of similar size in adult chicken skeletal-muscle and smooth-muscle poly(A)⁺ RNA, although the degree of hybridization to this mRNA was much lower in these tissues (Fig. 6, lanes c and d). Whether this is a reflection of the lower abundance of vinculin mRNA in muscle compared with non-muscle tissues or to the expression in muscle of mRNAs only partially homologous with the fibroblast cDNA probe is unclear.

A Southern blot of chick genomic DNA digested with a variety of restriction endonucleases and probed with the ³²P-labelled vinculin cDNA under conditions of high stringency is shown in Fig. 7. The pattern of hybridization with the DNA fragments produced by all five restriction enzymes was complex, considering that the cDNA probe contains single recognition sites for two of the enzymes (*Bam*HI and *Pst*I) and lacks sites for the others (*Eco*RI, *Hind*III and *Sst*I). These results are most readily explained either by the presence of multiple introns in the vinculin gene or by cross-hybridization of the probe to one or more closely related genes.

DISCUSSION

In the present study we have isolated and sequenced a 2891 bp vinculin cDNA from a chick-embryo fibroblast cDNA library constructed in the expression vector λ gt11 (Tamkun *et al.*, 1986). The Western-blot analysis of the fusion protein encoded by the recombinant was initially confusing. The recombinant clearly encoded a protein of higher molecular mass than β -galactosidase, but the protein, and what we presume to be high-molecular-mass degradation products, reacted only weakly, though specifically, with affinity-purified anti-vinculin immunoglobulin. The material that reacted most strongly with the antibodies was of much lower molecular mass, and did not obviously correspond to protein bands unique to *E. coli* strain C600 infected with the recombinant. The subsequent sequencing of the 2891 bp insert and the alignment of the deduced protein sequence with that of sequence determined in CNBr-cleavage peptides purified from chick smooth-muscle vinculin have confirmed the authenticity of the recombinant. The insert encodes a protein of at least 96 162 Da, and the predicted size of the fusion protein would therefore have been in excess of 200 000 Da, but no such protein was observed. We therefore conclude that the fusion protein was unstable in *E. coli* C600 cells, and was quantitatively degraded by proteolytic cleavage within the vinculin portion of the molecule to liberate polypeptides of 40 000–75 000 Da that were highly cross reactive with anti-vinculin antibodies.

The high-molecular-mass fusion proteins, which were readily detected by protein staining, are presumed to contain smaller and variable amounts of residual vinculin polypeptide sequence that reacted to a much lesser extent with the antibody. Some of the additional protein sequence of the higher-molecular-mass fusion proteins may have been encoded by cDNA sequences 5' to the normal initiation codon for vinculin, but in frame with the β -galactosidase gene. Such amino acid sequences would not normally be found in vinculin and would not be recognized by the anti-vinculin antibody. *E. coli* strain C600, which was used to express the fusion proteins encoded by the recombinant, lacks the *lon* mutation, which inactivates one of the major protein-degradation systems of *E. coli*. This may account in part for the instability of the fusion proteins expressed in these cells. The advantage of the system is that it is rapid, and avoids the need to isolate lysogens of the recombinant λ bacteriophage in suitable *E. coli* host strains.

Analysis of the 2891 bp cDNA sequence shows that the most 5' AUG codon (247) lies within a sequence matching the extensively documented consensus for higher-eukaryotic initiation codons (Kozak, 1984), and we therefore tentatively identified this as defining the *N*-terminus of the protein. Although the upstream sequence lacks termination codons and maintains the same open reading frame to the 5' end of the cDNA, the codon usage 5' of AUG 247 changes abruptly, consistent with the idea that this sequence is non-coding. The region is also G+C-rich, a feature found 5' to the putative initiation codons for protein kinase C (Parker *et al.*, 1986) and integrin (Tamkun *et al.*, 1986), cDNAs for which also display open reading frames to the 5' end. Finally, we have recently determined the *N*-terminal sequence of chick smooth-muscle vinculin and shown it to match the deduced sequence exactly, confirming

AUG-247 as encoding the initiating methionine residue (M. D. Davison, unpublished work). Initial inspection of the deduced protein sequence gives little clear information on the structure-function relationships within the protein. Vinculin has been reported to bind directly to membranes rich in acidic phospholipids (Ito *et al.*, 1983), and experiments with photoactivatable lipid probes suggest that a portion of the molecule actually penetrates the lipid bilayer (Niggli *et al.*, 1986). However, there is no obvious region of hydrophobic or non-charged amino acids that might account for these observations. It remains possible that the observed post-translational acylation of vinculin is important in this respect (Kellie & Wigglesworth, 1987; Burn & Burger, 1987). The sequence lacks an *N*-terminal glycine residue, the signal for myristoylation (Schultz *et al.*, 1985), although it is possible that vinculin has a cysteine residue close to the C-terminus that is palmitoylated, as in the case of p21 ras (Chen *et al.*, 1985). Since our sequence lacks the C-terminus of the protein, we are unable to comment further on this point. A family of proteins has been described that interact with acidic phospholipids in a Ca^{2+} -dependent manner (Geisow & Walker, 1986), including p36, a substrate for pp60^{v-src} (Saris *et al.*, 1986). Such proteins contain one or more copies of a 17-amino-acid-residue consensus sequence (Geisow & Walker, 1986). Although it is not clear whether vinculin binding to acidic phospholipids is Ca^{2+} -dependent, it was decided to inspect the vinculin sequence for the above consensus sequence, but no significant homologies were found. Interestingly, protein kinase C also interacts with acidic phospholipids in a Ca^{2+} -dependent manner (Nishizuka, 1984), but lacks the above consensus sequence (Parker *et al.*, 1986). There are, however, no significant homologies between vinculin and protein kinase C.

Vinculin is also known to bind to the cytoskeletal protein talin (Burrige & Mangeat, 1984), which may account for the localization of a proportion of the total cellular vinculin in adhesion plaques. Rotary-shadowing studies show that vinculin has a globular head and an extended tail that may be important in the ability of vinculin to self-aggregate (Milam, 1985). The globular head can be released by V8-proteinase cleavage. It has a molecular mass of 100 kDa and retains the ability of the intact molecule to bind to talin (Milam, 1985). As yet this fragment has not been isolated for sequencing so that its position within the deduced protein sequence can be determined. A clearer view of the functionally important domains in the vinculin sequence may emerge from a comparison of the conserved residues between the chick sequence and those of other species when they have been established.

There has been considerable interest in the observation that vinculin is a substrate for both pp60^{v-src} (Sefton *et al.*, 1981; Ito *et al.*, 1983) and protein kinase C (Werth & Pastan, 1984), although it is now apparent that there is no clear relationship between tyrosine-specific phosphorylation of vinculin by pp60^{v-src} and a variety of transformation parameters, including disruption of adhesion-plaque integrity (Sefton *et al.*, 1981; Rohrschneider *et al.*, 1982; Antler *et al.*, 1985; Kellie *et al.*, 1986b). We have found that one out of the seven tyrosine residues in the available vinculin sequence fits the proposed consensus for a tyrosine-phosphorylation site (Hunter & Cooper, 1985). Thus Tyr-822 has both a basic

amino acid seven residues *N*-terminal to the tyrosine, and an acidic residue within five residues of the tyrosine. Also, residues 9–14 *N*-terminal to Tyr-822 show significant homology with the tyrosine kinase auto-phosphorylation consensus sequence, making it a strong candidate for the phosphorylation site. A second residue, Tyr-537, has a basic amino acid seven residues from the tyrosine, but lacks the acidic residue. It should be noted, however, that the sequences surrounding tyrosine-phosphorylation sites in substrates for protein tyrosine kinases do not show a perfect match with the consensus sequence for tyrosine phosphorylation. Thus, although the determined phosphorylated tyrosine residue in lactate dehydrogenase (Cooper *et al.*, 1984) fits the tyrosine phosphorylation consensus sequence, that in p36 has the basic but not the acidic residue (Saris *et al.*, 1986), and that in enolase lacks both basic and acidic residues (Cooper *et al.*, 1984) (see Fig. 5). The conclusion that there is one good and one less good tyrosine-phosphorylation site in vinculin is consistent with the results of peptide-mapping studies on ³²P-labelled vinculin isolated from Rous-sarcoma-virus-transformed chick cells (Sefton *et al.*, 1981). Only two phosphotyrosine-containing peptides were detected, one containing 80% of the label. In contrast, there were multiple sites of serine and threonine phosphorylation.

The deduced protein sequence of chick-embryo fibroblast vinculin is identical with the partial sequence determined from CNBr-cleavage peptides derived from chick gizzard smooth-muscle vinculin. It would therefore appear that fibroblast and smooth-muscle vinculins are closely related or even identical. Although there are a number of isoelectric variants of vinculin (Geiger, 1982), it is unclear whether they represent slightly different polypeptides or arise through post-translational modification. Smooth muscle also expresses a 152000 Da protein that is both immunologically and structurally very closely related to vinculin, and has been called metavinculin (Siliciano & Craig, 1982). Translation experiments performed *in vitro* suggest that there are unique mRNAs for vinculin and metavinculin in smooth muscle, whereas fibroblasts only contain vinculin mRNA (Feramisco *et al.*, 1982).

Northern blots of fibroblast, smooth-muscle and skeletal-muscle extracts showed a single discrete mRNA in all three tissues, although we cannot exclude the possibility that our cDNA probe hybridizes to a number of closely related mRNA species of similar size. Interestingly, the size of the vinculin mRNA (approx. 7 kb) is substantially larger than that needed to code for a protein of 130000 Da. Preliminary experiments suggest that the 5' untranslated region of the vinculin mRNA is approx. 260 bp long, and we therefore infer that there must be a large 3' untranslated sequence. The mRNA for the oestrogen receptor (Green *et al.*, 1986) and tropomyosin (MacLeod *et al.*, 1986) have been shown to have 2–4 kb 3' untranslated sequences. Whether the mRNAs for vinculin and metavinculin arise through differential splicing of a primary transcript or are the products of separate genes remains to be elucidated.

We are grateful to Jim Turner and John Keyte for synthesis of oligonucleotides, to Dr. Mike Waterfield, Geoff Scrase and Nick Totty (Imperial Cancer Research Fund Laboratories, London, U.K.) for sequencing of vinculin tryptic peptides, and to Professor William J. Brammar for his continued encourage-

ment and advice throughout this project. The work was support by the Cancer Research Campaign and Medical Research Council (U.K.).

REFERENCES

- Antler, A. M., Greenberg, M. E., Edelman, G. M. & Hanafusa, H. (1985) *Mol. Cell. Biol.* **5**, 263–267
- Baron, M. D., Davison, M. D., Jones, P., Patel, B. & Critchley, D. R. (1987) *J. Biol. Chem.* **262**, 2558–2561
- Bennett, J. P., Zamer, K. S. & Stossel, T. P. (1984) *Biochemistry* **23**, 5081–5086
- Brenner, D. G. & Shaw, W. V. (1985) *EMBO J.* **4**, 561–568
- Burn, P. & Burger, M. M. (1987) *Science* **235**, 426–479
- Burridge, K. & Connell, L. (1983) *J. Cell Biol.* **97**, 359–367
- Burridge, K. & Mangeat, P. (1984) *Nature (London)* **308**, 744–746
- Chen, Z. Q., Ulsh, L. S., DuBois, G. & Shih, T. Y. (1985) *J. Virol.* **56**, 607–612
- Cooper, J. A., Esch, F. S., Taylor, S. S. & Hunter, T. (1984) *J. Biol. Chem.* **259**, 7835–7841
- Dalgleish, R., Williams, G. & Hawkins, J. R. (1986) *Hum. Genet.* **73**, 91–92
- Damsky, C. M., Knudsen, K. A., Bradley, D., Buck, C. A. & Horwitz, A. F. (1985) *J. Cell Biol.* **100**, 1528–1539
- Drenckhahn, D. & Franz, H. (1986) *J. Cell Biol.* **102**, 1843–1852
- Drenckhahn, D. & Wagner, J. (1986) *J. Cell Biol.* **102**, 1738–1747
- Evans, R. R., Robson, R. M. & Stromer, M. H. (1984) *J. Biol. Chem.* **259**, 3916–3924
- Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267
- Feramisco, J. R., Smart, J. E., Burridge, K., Helfman, D. M. & Thomas, G. P. (1982) *J. Biol. Chem.* **257**, 11024–11031
- Fowler, A. V. & Zabin, I. (1978) *J. Biol. Chem.* **253**, 5521–5525
- Friedman, M., Krull, L. H. & Cavins, J. F. (1970) *J. Biol. Chem.* **245**, 3868–3871
- Geiger, B. (1982) *J. Mol. Biol.* **159**, 685–701
- Geiger, B., Tokuyasu, K. T., Dutton, A. H. & Singer, S. J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4127–4131
- Geiger, B., Avnur, Z., Volberg, T. & Volk, T. (1985) in *The Cell in Contact: Adhesions and Junctions as Morphogenetic Determinants* (Edelman, G. M. & Thiery, J. P., eds.), pp. 461–489, John Wiley and Sons, New York
- Geisow, M. J. & Walker, J. H. (1986) *Trends Biochem. Sci.* **11**, 420–423
- Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J. M., Argos, P. & Chambon, P. (1986) *Nature (London)* **320**, 134–139
- Heath, J. P. (1986) *Nature (London)* **320**, 484–485
- Herman, B. & Pledger, W. J. (1985) *J. Cell Biol.* **100**, 1031–1040
- Hirst, R., Horwitz, A. F., Buck, C. A. & Rohrschneider, L. R. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6470–6474
- Horwitz, A. F., Duggan, K., Buck, C. A., Beckerle, M. C. & Burridge, K. (1986) *Nature (London)* **320**, 531–533
- Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54**, 897–930
- Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning* (Glover, D. M., ed.), vol. 1, pp. 49–78, IRL Press, Oxford
- Hynes, R. O., Destree, A. T. & Wagner, D. D. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **46**, 659–670
- Ito, S., Werth, D. K., Richert, N. D. & Pastan, I. (1983) *J. Biol. Chem.* **258**, 14626–14631
- Kellie, S. & Wigglesworth, N. M. (1987) *FEBS. Lett.* **231**, 428–432
- Kellie, S., Patel, B., Mitchell, A., Critchley, D. R., Wigglesworth, N. M. & Wyke, J. A. (1986a) *J. Cell Sci.* **82**, 129–142
- Kellie, S., Patel, B., Wigglesworth, N. M., Critchley, D. R. & Wyke, J. A. (1986b) *Exp. Cell Res.* **165**, 216–228
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872

- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lazarides, E. & Burridge, K. (1975) *Cell (Cambridge, Mass.)* **6**, 289–298
- MacLeod, A. R., Houlker, C., Reinach, F. C. & Talbot, K. (1986) *Nucleic Acids Res.* **14**, 8413–8426
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Martin-Zanca, D., Hughes, S. H. & Barbacid, M. (1986) *Nature (London)* **319**, 743–748
- Meigs, J. B. & Wang, Y. L. (1986) *J. Cell Biol.* **102**, 1430–1438
- Milam, L. M. (1985) *J. Mol. Biol.* **184**, 543–545
- Naharro, G., Rubbins, K. C. & Reddy, E. P. (1984) *Science* **223**, 63–66
- Niggli, V., Dimitrov, D. P., Brunner, J. & Burger, M. (1986) *J. Biol. Chem.* **261**, 6912–6918
- Nishizuka, Y. (1984) *Nature (London)* **308**, 693–698
- Norrandner, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106
- Pappin, D. J. C. & Findlay, J. B. C. (1984) *Biochem. J.* **217**, 605–613
- Parker, P. J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., Stabel, S., Waterfield, M. D. & Ullrich, A. (1986) *Science* **233**, 853–859
- Pasquale, E. B., Maher, P. A. & Singer, S. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5507–5511
- Pierschbacher, M. D. & Ruoslahti, E. (1984) *Nature (London)* **309**, 30–33
- Pytela, R., Pierschbacher, M. D. & Rouslahti, E. (1985) *Cell (Cambridge, Mass.)* **40**, 191–198
- Rohrschneider, L., Rosok, M. & Shriver, K. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **46**, 953–965
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5466
- Saris, C. J., Tack, B. F., Kristensen, T., Glenney, J. R. & Hunter, T. (1986) *Cell (Cambridge, Mass.)* **46**, 201–212
- Schultz, A. M., Henderson, L. E., Oroszlan, S., Garber, E. A. & Hanafusa, H. (1985) *Science* **227**, 427–429
- Sefton, B. M., Hunter, T., Ball, E. H. & Singer, S. J. (1981) *Cell (Cambridge, Mass.)* **24**, 165–174
- Siliciano, J. D. & Craig, S. W. (1982) *Nature (London)* **300**, 533–535
- Smith, T. F., Waterman, M. S. & Sadler, J. R. (1983) *Nucleic Acids Res.* **11**, 2205–2220
- Staden, R. (1982a) *Nucleic Acids Res.* **10**, 2951–2961
- Staden, R. (1982b) *Nucleic Acids Res.* **10**, 4731–4751
- Tamkun, J. W., De Simone, D. W., Fonda, D., Patel, R. S., Buck, C. A., Horwitz, A. F. & Hynes, R. O. (1986) *Cell (Cambridge, Mass.)* **46**, 271–282
- Towbin, H., Staehlin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- Werth, D. K. & Pastan, I. (1984) *J. Biol. Chem.* **259**, 5264–5270
- Yamada, K. M., Humphries, M. J., Hasegawa, T., Hasegawa, E., Olden, K. & Chien, W. T. (1985) in *The Cell in Contact: Adhesions and Junctions as Morphogenetic Determinants* (Edelman, G. M. & Thiery, J. P., eds.), pp. 303–332, John Wiley and Sons, New York

Received 12 January 1987/1 May 1987; accepted 7 May 1987