Primary sequence and domain structure of chicken vinculin

Glyn J. PRICE,* Peter JONES,* Matthew D. DAVISON,* Bipin PATEL,* Ronit BENDORI,†
Benjamin GEIGER† and David R. CRITCHLEY*‡
*Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, U.K.,
and †Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel

We have determined the complete sequence of chick vinculin from two overlapping cDNA clones. The vinculin mRNA consists of 262 bp of 5' untranslated sequence, an open reading frame of 3195 bp (excluding the initiation codon) and a long 3' untranslated sequence (> 2 kb). Chick vinculin contains 1066 amino acid residues, and has a deduced molecular mass of 116933 Da. Analysis of the domain structure of vinculin shows that the molecule can be cleaved by V8 proteinase into a 90 kDa globular head and a 32 kDa tail region, the latter of which could further be cleaved into a 27 kDa polypeptide. The 90 kDa globular head contains the N-terminus of vinculin, three 112-residue repeats (residues 259–589), and extends to approximately residue 850. Gel overlay experiments show that it also contains a binding site for the cytoskeletal protein talin. The talin-binding domain was further localized to the N-terminal 398 amino acid residues of the protein by expression in vitro of this region from a vinculin cDNA cloned into the Bluescript SK+ vector. The head and tail domains are apparently separated by a proline-rich region that contains V8-proteinase-cleavage sites and a candidate tyrosine (822)-phosphorylation site. Secondary-structure prediction suggests that the head and tail domains contain α-helical regions separated by short stretches of turn/coil. Comparison of the chick with a partial human sequence reveals that vinculin is a highly conserved protein. In chickens Southern-blot analysis is consistent with a single vinculin gene, and it is therefore likely that vinculin, and its higher-molecular-mass isoform termed metavinculin, arise through alternative splicing.

INTRODUCTION

Vinculin is a cytoskeletal protein associated with a family of specialized cell–cell and cell–substrate contacts called adherens junctions (Geiger, 1979; Geiger et al., 1980, 1985). In both types of contacts vinculin is thought to form a complex with a number of other components that serve to anchor microfilamentous actin to the membrane. The details of these molecular interactions are best understood in adhesion plaques (or focal contacts), the areas of closest apposition between the ventral surface of cultured cells and the substrate (for a review see Burridge, 1986). Studies on the molecular substructure of adhesion plaques suggest that they contain several domains. These include integral transmembrane components of the integrin family (Tamkun et al., 1986; Buck & Horwitz, 1987), a submembrane plaque containing vinculin (Geiger et al., 1980) and talin (Burridge & Connell, 1983), and a cytoskeletal domain enriched in α-actinin (Lazarides & Burridge, 1975), actin and several associated proteins (reviewed in Burridge, 1986).

Recent biochemical studies suggest a sequence of molecular interactions that might be involved in the transmembrane assembly of adhesion plaques. It has been shown that the cytoplasmic C-terminus of the β-subunit of integrin can bind talin (Horwitz et al., 1986), and that talin can bind to vinculin (Otto, 1983; Burridge & Mangeat, 1984). Furthermore, binding studies indicate

that vinculin may also interact with α-actinin via a low-affinity binding site (Belkin & Koteliansky, 1987; Wachsstock et al., 1987), and possibly with itself (Geiger et al., 1984). α-Actinin in turn is known to bind to and cross-link actin filaments (Bennett et al., 1984). This series of molecular interactions might account for the apparent transmembrane linkage of the extracellular matrix to the cytoplasmic microfilament system. Studies on the molecular diversity of adherens-type junctions have, however, suggested that the associations of vinculin with the junctional plaque may occur via two or more alternative routes. Thus it has been shown that the intercellular adherens-type junctions do not contain integrin and talin (Geiger et al., 1985), but are associated with A-CAM (Volk & Geiger, 1986), L-CAM (Boller et al., 1985) and plakoglobin (Cowin et al., 1986).

Further understanding of the detailed molecular interactions between the various components of adhesion plaques has been largely restricted by the lack of information on the structure of the proteins involved and their functional domains. We have previously determined the complete sequence of a chick α-actinin (Baron et al., 1987a,b), and have published the partial sequence of a chick vinculin deduced from a 2.89 kb cDNA clone encoding approx. 96 kDa of the N-terminal part of the protein (Price et al., 1987). In the present paper we report the completion of the elucidation of the chick vinculin sequence by use of a vinculin cDNA clone (cVin5)

---

isolated by Bendori et al. (1987). The deduced sequence of vinculin is analysed vis-à-vis the known biochemical and structural properties of the protein.

MATERIALS AND METHODS

Isolation of vinculin cDNAs

The two chick vinculin cDNAs described in this study were isolated from a chick-embryo fibroblast λgt11 cDNA library (large insert) (Tamkun et al., 1986) by antibody screening. One of the cDNAs contained 2891 bp (2.89 kb cDNA) (Price et al., 1987); the other was approx. 5 kb in length (cVin 5) (Bendori et al., 1987). A partial human vinculin cDNA was isolated from a human fibroblast λgt10 cDNA library kindly provided by Dr. A. R. Macleod (Ludwig Institute, Cambridge, U.K.). The library was screened with the 2.89 kb chick vinculin cDNA purified by agarose-gel electrophoresis and labelled with [α-^{32}P]dCTP by the random priming method of Feinberg & Vogelstein (1984). Plaque screening by DNA–DNA hybridization was by standard procedures (Maniatis et al., 1982) with a final wash in 2 × SSC containing 0.1 % SDS at 65 °C.

DNA sequencing

DNA sequencing was carried out by the dideoxy chain-termination method of Sanger et al. (1977) after cloning into bacteriophage, M13 mp18 (Norrander et al., 1983). Partial sequence was obtained by shotgun cloning, and elucidation of the complete sequence was achieved by using synthetic 17-mer oligonucleotides as primers on the full-length cDNA (Brenner & Shaw, 1985).

Primer extension analysis

Primer extension analysis was carried out by using 2 μg of polyadenylated RNA, isolated from cultured chick-embryo fibroblasts, and a 201 bp single-stranded DNA primer (nucleotide residues 28–229) universally labelled with [α-^{32}P]dCTP, as described by Williams & Mason (1985).

Southern blots

DNA was prepared from chick brain by phenol/ chloroform extraction and RNAase A digestion by standard procedures (Maniatis et al., 1982). Restriction-endonuclease digests of the DNA were fractionated in agarose gels (0.8 %), the DNA was transferred to Hybond N nylon membranes (Amersham International, Amersham, Bucks., U.K.) by passive diffusion, and hybridization was performed at 65 °C for 15–20 h (Dalgleish et al., 1986). Filters were washed to a final stringency of 0.5 × SSC containing 0.1 % SDS at 65 °C.

Purification and sequencing of V8-proteinase-cleavage fragments of vinculin

Vinculin was purified by the method of Evans et al. (1984) from adult chicken gizzard, and was more than 95 % pure as judged by SDS/PAGE (Laemmli, 1970). Vinculin (1 mg/ml) was incubated with V8 proteinase (10 μg/ml) in 20 mM-NaCl / 0.1 % (v/v) 2-mercapto-ethanol/0.1 mM-EDTA/20 mM-Tris/HCl buffer, pH 7.6, for 1 h at 37 °C. The 90 kDa, 32 kDA and 27 kDa fragments from 3 mg of total digest were resolved by SDS/PAGE, and the fragments were recovered by the electroelution method of Baron et al. (1987a) and sequenced on an Applied Biosystems 470A gas-phase

sequencer. V8-proteinase-cleavage fragments resolved by SDS/PAGE were also electroblotted to poly(vinylidene difluoride) membranes, and individual bands were excised for sequencing, as described by Matsudaira (1987).

Binding of talin to V8-proteinase-cleavage fragments of vinculin

Vinculin (150 μg) was incubated with V8 proteinase (3 μg) in 600 μl of buffer for various times, and the resulting fragments were separated by SDS/PAGE and transferred to nitrocellulose filters (Towbin et al., 1979). Filters were incubated in Ponceau S (BDH Chemicals, Poole, Dorset, U.K.), and the positions of the separated polypeptides and molecular-mass standards were marked in ink. Filters were then incubated with 3 % (w/v) dried milk dissolved in 0.9 % NaCl/20 mM-Tris/HCl buffer, pH 7.4, followed by 100 μg/ml of a 190 kDa talin-derived fragment (O'Halloran & Burridge, 1986) in the same buffer, each for 1 h at room temperature. The filters were washed, and the bound 190 kDa talin fragment was located by using a rabbit antibody to talin (serum number F16/36, kindly provided by Dr. K. Burridge, University of North Carolina, Chapel Hill, NC, U.S.A.) diluted 1:1000, and an indirect immunoperoxidase technique (Price et al., 1987).

Generation of transcripts in vitro

The 2.89 kb vinculin cDNA was cloned into the EcoR1 site of the Bluescript vector SK + (Stratagene, La Jolla, CA, U.S.A.), and RNA transcripts encoding the N-terminal 398 amino acid residues of vinculin were generated in vitro from plasmid DNA linearized by BamH1 cleavage within the cDNA insert. The reaction mix (10 μl) contained 1 μg of plasmid DNA, 15 units of T7 RNA polymerase (Boehringer, Mannheim, Germany), 1 mM-nucleotide triphosphates, 5 mM-dithio-threitol, 14 units of RNasin (Amersham International), 40 mM-Tris/HCl buffer, pH 8, 8 mM-MgCl$_2$, 2 mM-spermidine and 50 mM-NaCl. The nucleic acids were extracted and precipitated by standard procedures (Maniatis et al., 1982), and the amount and integrity of the RNA synthesized were determined by agarose-gel electrophoresis. The RNA was redissolved in sterile water to a concentration of 1 μg/μl.

Capping of RNA transcript and translation in vitro

RNA transcripts were capped with guanylyltransferase (GIBCO–BRL, Paisley, Strathclyde, U.K.) as described in the manufacturer's instructions, with the addition of 1 mM-S-adenosylmethionine. The capped transcript was extracted and precipitated as above, and translated in vitro by using the rabbit reticulocyte lysate system (Amersham International) in the presence of [^{35}S]-methionine (Amersham International) as per the manufacturer's instructions.

Binding of translation products to talin

Micro-titre wells were incubated overnight with 100 μg of talin or bovine serum albumin/ml in NET buffer (150 mM-NaCl/5 mM-EDTA/50 mM-Tris/HCl buffer, pH 7) containing 1 mM-phenylmethanesulphonyl fluoride, unbound protein was removed by washing, and excess protein-binding sites were blocked with 1 % (w/v) bovine serum albumin (1 h at 37 °C). Total 'in vitro' translation reaction mixtures diluted in NET buffer

containing 1 mM-phenylmethanesulphonyl fluoride were added to the micro-titre wells, and incubated for 1–2 h at 37 °C. Unbound material was removed by rinsing the wells with three changes of NET buffer containing 0.05 % Nonidet P-40. Bound radiolabelled polypeptides were extracted in SDS sample buffer (2 h at 37 °C), and were analysed by SDS/PAGE and autoradiography.

Immune precipitation

Portions of the 'in vitro' translation reaction mixtures were diluted to 1 ml with NET buffer containing 1 mM-phenylmethanesulphonyl fluoride and incubated with 10 μl of a rabbit antiserum against chicken gizzard vinculin (Kellie et al., 1986) or pre-immune serum from the same animal. Immune complexes were precipitated by use of 40 μl of a 10 % (w/v) suspension of Protein A–Sepharose in NET buffer containing 1 % bovine serum albumin. Precipitated radiolabelled proteins were analysed by SDS/PAGE and autoradiography.

Secondary-structure prediction

Secondary-structure predictions were carried out on the complete chick and partial human vinculin sequences by using the joint prediction method of Eliopoulos et al. (1982). In its present form this combines the results of eight separate predictive methods, which were integrated into a single joint prediction by using rules based on the total number of methods that predict α-helix, β-strand and turn coil for each residue within a sequence. The following thresholds were used, based on tests of the joint prediction method applied to predominantly α-helical proteins of known crystal structure: a residue was assigned as helical when five or more of the eight methods predicted helix, or when four or more methods predicted helix when two or more methods failed to give a prediction. Two methods predicting turn/coil took precedence over up to three methods predicting helix, and three methods predicting turn/coil took precedence over up to four methods predicting helix. When the number of methods predicting β-strand equalled or was one less than the number predicting helix, the β-strand prediction took precedence. At the few positions where the number of methods predicting turn/coil was equal to those predicting β-strand the residue was scored as possibly either structure.

RESULTS

We have previously published the partial sequence of chick vinculin deduced from a 2.89 kb cDNA clone that was isolated from a chick-embryo fibroblast λgt11 library (Price et al., 1987). The clone contained 246 bp of 5' untranslated sequence, followed by an open reading frame through to the 3' end of the clone. We have subsequently authenticated AUG 247 as the initiation codon by direct protein sequencing. The clone encoded a polypeptide of 96162 Da, but lacked 3' coding sequence corresponding to the C-terminus of vinculin. Comparison of the restriction map of this clone with that of clone cVin5 isolated from the same library by Bendori et al. (1987) indicated that there was extensive overlap between the two clones, with the 2.89 kb cDNA containing additional 5' sequence, and cVin5 containing additional 3' sequence (Fig. 1). In the region of overlap, the restriction maps of the two clones were identical except for an Aval site that was uniquely located in the 2.89 kb
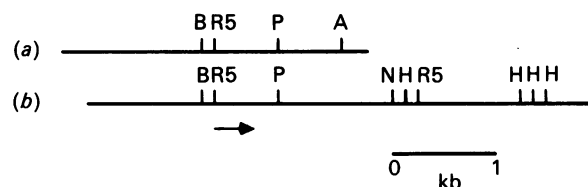

Fig. 1. Restriction maps of two chick vinculin cDNAs

(a) The 2.89 kb cDNA isolated by Price et al. (1987); (b) the 5 kb cVin5 cDNA isolated by Bendori et al. (1987). The maps of the two clones within the region of overlap were identical with the exception of the Aval site, present exclusively at the 3' end of the 2.89 kb cDNA. Bendori et al. (1987) have described a HindIII site unique to the 5' end of cVin5. This was later found to be mislocated. The nucleotide sequence of cVin5 was determined from the 5' EcoRV site (arrow). Key to restriction-endonuclease sites: A, Aval; B, BamH1; H, HindIII; N, Ncol; P, Pst1; R5, EcoRV.

cDNA. DNA sequencing of cVin5 from the 5' EcoRV site in the 3' direction showed that the two clones were essentially identical in this region of overlap. There were five single-base substitutions, one of which contributed the unique Aval site in the 2.89 kb cDNA. All of the differences were in the third base of the codon, and none affected the deduced amino acid sequence of the protein.

DNA sequencing of the unique 3' region of cVin5 has allowed the complete sequence of chick vinculin to be determined (Fig. 2). Alignment of the sequence of the two cDNAs showed that initiation codon AUG 247 is followed by a single open reading frame of 3195 bp ending at a termination codon TAA (3445). This stop codon is followed by a long 3' untranslated sequence (1760 bp) rich in A + T residues, and containing several in-frame stop codons, but lacking a polyadenylation signal and poly(A) tail. Primer extension studies carried out with the 2.89 kb cDNA showed that the corresponding vinculin mRNA contains 262 ± 1 bp of 5' untranslated sequence. Northern-blot analysis has in turn shown that both cDNAs hybridized to a single vinculin mRNA band of the order of 6–7 kb (Price et al., 1987; Bendori et al., 1987). It is therefore clear that the 3' untranslated region of the vinculin mRNA will be in excess of 2.5 kb in length. Inspection of the deduced sequence reveals that vinculin is a protein of 1066 amino acid residues with a molecular mass of 116933 Da, without taking into account any post-translational modifications.

In smooth and cardiac muscle a higher-molecular-mass isoform of vinculin has been described and termed metavinculin (Feramisco et al., 1982; Siliciano & Craig, 1982; Geiger, 1982). Such vinculin variants either could be the products of different genes or might arise through alternative splicing of the primary transcript of a single gene. Southern blots of chick genomic DNA, digested with a variety of restriction enzymes and probed with the 2.89 kb vinculin cDNA (Price et al., 1987), or with a variety of cDNA clones including cVin5 (Bendori et al., 1987), produced a complex pattern of hybridization. When this experiment was repeated with a 400 bp EcoRl–Bgl1 restriction enzyme fragment from the extreme 5' end of the 2.89 kb cDNA, only a single band was detected in each digest (Fig. 3). This result strongly suggests that there is a single chicken vinculin gene.

Fig. 2. Complete cDNA nucleotide sequence and deduced amino acid sequence of chick vinculin

The complete sequence of vinculin was compiled from (1) a 2.89 kb cDNA that contained 246 bp of 5' untranslated sequence, the initiation codon (AUG 247), and encoded 881 amino acid residues, but lacked 3' coding and non-coding sequence (Price et al., 1987), and (2) a second vinculin cDNA (cVin5) isolated by Bendori et al. (1987), which contained the additional 3' sequence missing from the 2.89 kb cDNA. The nucleotide sequence of cVin5 determined from the 5' EcoRV site (Fig. 1, arrow) was identical with that of the 2.89 kb cDNA in the region of overlap, except for five third-base differences at positions 1701 (G in cVin5, A in the 2.89 kb cDNA), 1746 (C to T), 2043 (C to T), 2058 (C to T) and 2793 (A to C). Numbers on the left refer to nucleotide sequence, those on the right to the deduced amino acid sequence.

In recent years biochemical and electron-microscopic studies have shed some light on the domain substructure of vinculin. Rotary-shadowing studies have shown chicken gizzard vinculin to be composed of a globular head and an extended tail region (Milam, 1985). The globular head region can be liberated by V8-proteinase cleavage as a 90 kDa fragment (Milam, 1985). Analysis of the time course of the proteolysis indicates that vinculin
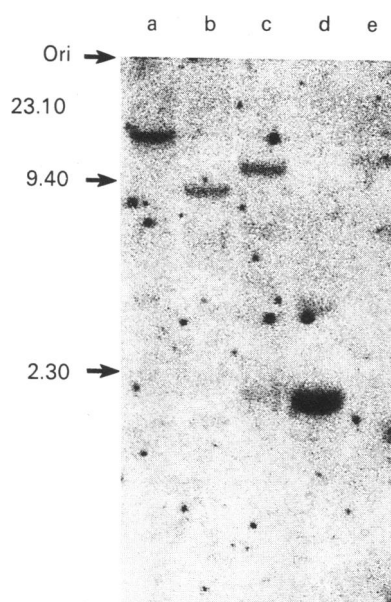
Fig. 3. Southern-blot analysis of chick genomic DNA digested with a variety of restriction endonucleases

Chick brain DNA (4 μg) was digested with *Bam*H1 (lane a), *Eco*R1 (lane b), *Hind*III (lane c), *Pst*1 (lane d) and *Sst*1 (lane e) and fractionated on a 0.8 % agarose gel. The DNA was transferred to a nylon filter and hybridized with a ^{32}P-labelled 400 bp *Eco*R1–*Bgl*1 restriction fragment of the 2.89 kb cDNA as described in the Materials and methods section. Molecular-size markers (kb) are shown on the left; Ori is the gel origin.

Fig. 4. Time course of V8-proteinase cleavage of vinculin

Vinculin was incubated with V8 proteinase as described in the Materials and methods section, and samples were taken after times indicated in the Figure, subjected to SDS/PAGE (5–20 %-acrylamide gradient gel) and stained with Coomassie Blue. The positions of molecular-mass marker proteins (kDa), are shown on the left, and of intact vinculin (V) and vinculin fragments to the right.

is cleaved first to 90 kDa and 32 kDa fragments (Fig. 4, lane b), the latter presumably representing the tail of the molecule. The 90 kDa fragment is subsequently cleaved to a 60 kDa fragment, and the 32 kDa fragment to a 27 kDa fragment (Fig. 4, lanes c and d). To localize the head and tail domains within the vinculin molecule, the above V8-proteinase-cleavage fragments were isolated and partially sequenced. Sequencing of the 90 kDa fragment showed that it contained the same *N*-terminus as intact vinculin. Conversion of the 90 kDa fragment into the 60 kDa fragment occurred by cleavage between residues 243 and 244. The *N*-terminal residues of the 32 kDa and 27 kDa fragments were identified as histidine-851 and leucine-858 respectively. Because of the apparent 5 kDa size difference between these two fragments, it is likely that conversion of the 32 kDa into the 27 kDa fragment involves cleavage at both *N*- and *C*-terminal sites. Interestingly, the V8-proteinase-cleavage sites that liberate the 90 kDa, 32 kDa and 27 kDa fragments are located within a proline-rich region (Fig. 5). We therefore conclude that the 90 kDa globular head contains the *N*-terminus of vinculin, while the 32 kDa tail polypeptide contains the *C*-terminus of the protein.

Vinculin has been shown to bind to another adhesion-plaque protein talin, through a binding domain thought to be located within the globular head region of the vinculin molecule (Burridge & Mangeat, 1984; Milam, 1985). To investigate this point more thoroughly, vinculin was cleaved with V8 proteinase, and the 90 kDa and 32 kDa fragments were resolved by SDS/PAGE and

transferred to nitrocellulose. The filter was then overlaid with the 190 kDa talin fragment and washed, and the bound talin fragment was detected by using an indirect immunoperoxidase technique. Talin binding was limited to the 90 kDa globular head region of vinculin, and there was no binding to the 60 kDa or 32 kDa fragments (Fig. 6).

In an attempt to define further the talin-binding domain in the globular head region of vinculin, we expressed the *N*-terminal 398 amino acid residues of vinculin by means of a transcription/translation system *in vitro*. Transcripts encoding this region were synthesized from the 2.89 kb vinculin cDNA cloned into the Blue-script vector, and linearized within the insert by using *Bam*H1. Translation of the capped RNA by the rabbit reticulocyte system resulted in the synthesis of a poly-peptide with a molecular mass of 45 kDa, in close agreement with the predicted size based on a knowledge of the amino acid sequence (Fig. 7, lanes a and b). The polypeptide was specifically immunoprecipitated from the total translation products with an antibody to chick vinculin (Fig. 7, lanes c and d), thereby confirming its identity. The 45 kDa polypeptide bound to micro-titre

90 kDa head containing talin-binding site
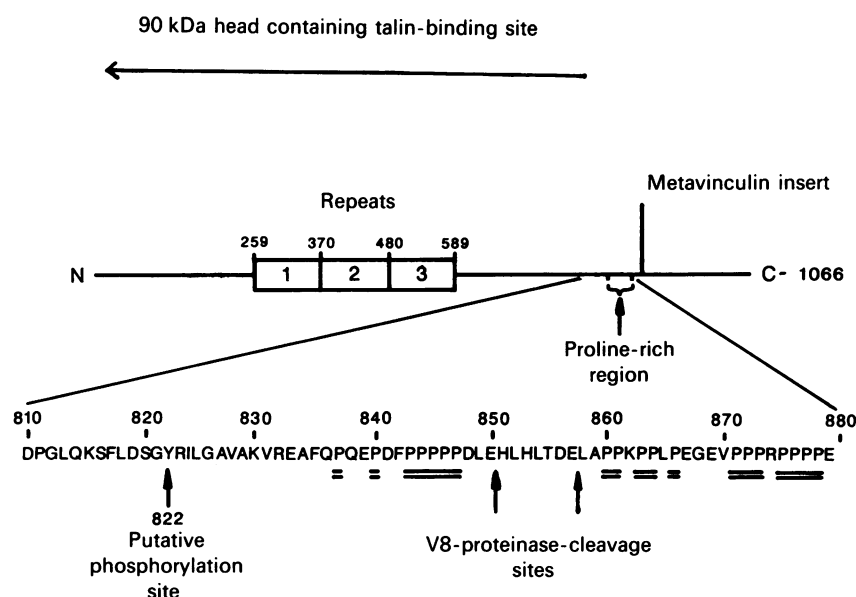
Fig. 5. Location of domains within the vinculin primary sequence

The upper portion of the Figure shows the position of the three 112-amino-acid-residue repeats, the proline-rich region and the probable location of a sequence found to be unique to metavinculin (pig) by Gimona *et al.* (1988). The location of the 90 kDa V8-proteinase-cleavage fragment, which contains the globular head of vinculin and the talin-binding site, is also indicated. The amino acid sequence of the proline-rich region is shown in the lower part of the Figure (proline residues are underlined), along with the V8-proteinase-cleavage sites that give rise to the 32 kDa and 27 kDa *C*-terminal fragments. The position of a candidate tyrosine-phosphorylation site is indicated.

wells coated with talin, but not bovine serum albumin (Fig. 7, lanes e–g), and the binding was clearly specific in that none of the polypeptides synthesized from the endogenous reticulocyte mRNA bound to talin. Binding was also progressively inhibited by the addition of increasing amounts of unlabelled vinculin to the assay (results not shown). These results indicate that a talin-binding site is located within the *N*-terminal 398 amino acid residues of vinculin.

We have previously shown that amino acid residues 259–589 of the chick vinculin sequence deduced from the 2.89 kb cDNA consist of three repeats of 112 amino acid residues (Price *et al.*, 1987). There is no significant sequence similarity to the vinculin repeats before residue 259, although residues 678–724 do show similarity to the first part of the repeat consensus. Further examination indicated that there is no apparent similarity to the repeat consensus sequence *C*-terminal to this region. The repeats are therefore contained within the globular head of the protein. We have also sequenced a 1.5 kb human vinculin cDNA that encodes amino acid residues equivalent to 253–771 in the deduced chick vinculin sequence, and therefore includes the sequence of all three repeats (results not shown). The nucleotide sequence of the human and chick cDNAs showed an 80% level of identity. The deduced amino acid sequence showed a 90% level of identity, increasing to 94% similarity if conservative substitutions are included. Clearly vinculin is a highly conserved protein.

Secondary-structure prediction analysis of the complete chick vinculin sequence indicates a high α-helical content (62.5%) with 21% turn/coil and 16.5% β-strand. The two short stretches of sequence rich in proline residues (residues 837–848 and 860–878) are strongly predicted to be regions of turn/coil structure.
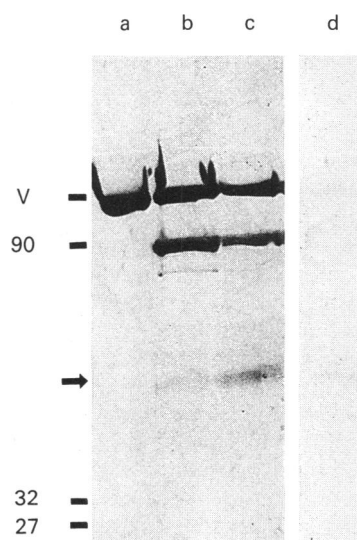
Fig. 6. Talin binding to V8-proteinase-cleavage fragments of vinculin

Vinculin was incubated with V8 proteinase at 37 °C for 0 min (lane a), 15 min (lane b) and 30 min (lanes c and d). Vinculin fragments were resolved by SDS/PAGE and transferred to nitrocellulose, and binding of the 190 kDa talin-derived polypeptide to these fragments was determined by using a rabbit anti-talin antibody and an indirect immunoperoxidase procedure, all as described in the Materials and methods section. The specificity of the detection system was determined by omitting the 190 kDa talin fragment from the assay (lane d). The position of intact vinculin (V) and the 90 kDa, 32 kDa and 27 kDa vinculin fragments are shown to the left of the Figure. The arrow indicates a region where talin binding was not reproducible.

Translation   Immune pptn.   Talin binding
in vitro

45 —

a   b   c   d   e   f   g
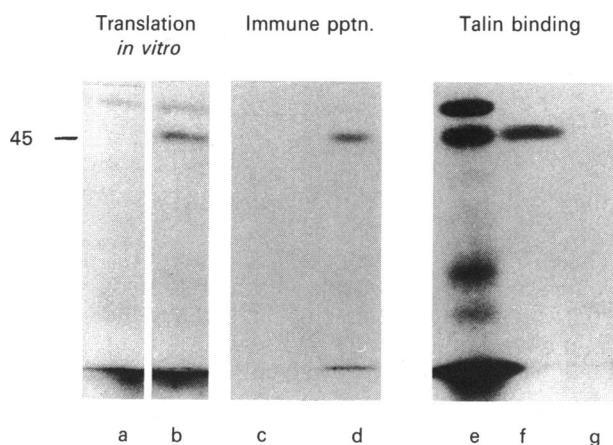
Fig. 7. Talin binding to the N-terminal 45 kDa vinculin fragment synthesized by using a transcription/translation system in vitro

RNA transcripts encoding the N-terminal 398 amino acid residues of vinculin were synthesized from the 2.89 kb vinculin cDNA clones into the Bluescript SK + vector, as described in the Materials and methods section. Capped transcript was translated in vitro by using a rabbit reticulocyte system in the presence of [^{35}S]methionine, and the proteins synthesized from endogenous mRNA (lane a), or in response to added transcript (lane b), were analysed by SDS/PAGE. Lanes c and d, immune precipitation of b either with pre-immune rabbit serum (lane c) or with rabbit anti-vinculin antibody (lane d). Lanes e–g, binding of proteins synthesized by the translation system in vitro in response to added transcript (lane e), to micro-titre wells coated with talin (lane f) or to bovine serum albumin (lane g). The position and molecular mass (kDa) of the expressed vinculin polypeptide is shown to the left of the Figure.

The consensus prediction for the aligned chick and human 112-amino-acid-residue repeats within the globular head domain is shown in Fig. 8, which also shows the number and position of proline residues in the alignment. Below the histogram is a diagram delineating the helical, turn/coil and β-strand regions of the consensus prediction for the repeats. The repeats are predicted to consist of three helices joined by turn/coil regions, with no consistent prediction of β-strand. Helix 1 (34 residues) is longer than helices 2 and 3 (20 and 24 residues respectively), and there is some indication for discontinuity in helix 3 (alignment positions 68 and 69).

DISCUSSION

In this study, we have completed the elucidation of the amino acid sequence of chicken vinculin, and aligned the major structural domains of the molecule along it. Two overlapping cDNA clones have been used to determine the complete vinculin sequence. One clone, referred to as the 2.89 kb cDNA, had been sequenced previously (Price et al., 1987). The other clone (cVin5), isolated by Bendori et al. (1987), showed extensive overlap with the 2.89 kb cDNA, but contained the 3' coding and non-coding sequence missing from the 2.89 kb clone. Comparison of the sequences of the two clones in the region of overlap 3' to the common EcoRV restriction enzyme site pointed to nearly identical sequences, with the exception of five single-base substitutions, all of which were 'silent' and probably reflect allelic variations. The complete sequence deduced from these two overlapping cDNAs showed that chick vinculin is a protein of 1066 amino acid residues with a deduced molecular mass of 116933 Da. This value is in excellent agreement with that estimated
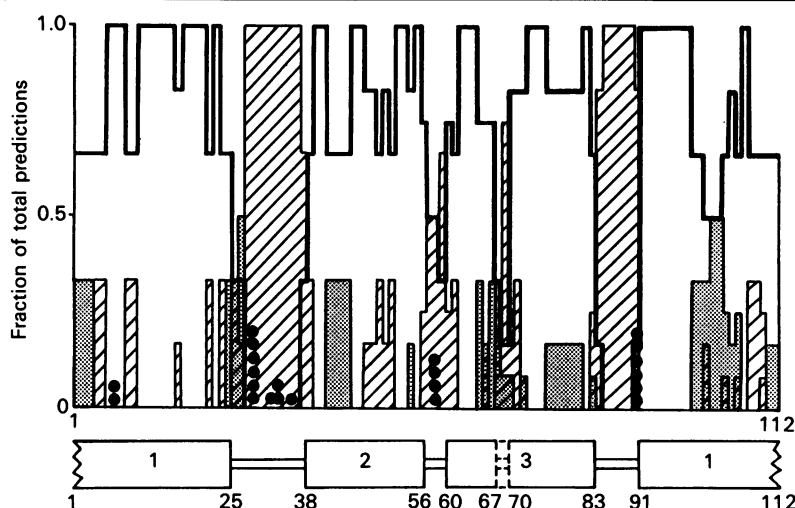
Fig. 8. Consensus secondary structure for the chick and human repeats in vinculin

The consensus structure prediction histogram was constructed as described in the Materials and methods section. The helix prediction is indicated by the thicker line, the turn/coil prediction as hatched areas and the β-strand prediction as the stippled areas. The vertical axis represents the fraction of the total number of predictions that gave each type of the three structures, and the horizontal axis the position in the repeat alignment given by Price et al. (1987). The position and number of proline residues within the alignment are indicated by ●. By using the thresholds stated below, the helical and turn/coil region of the consensus repeat was delineated, and is shown at the foot of the diagram. There is no clearly significant β-strand predictions within the repeat. The thresholds used to delineate helical and turn/coil regions were: (a) when the fraction of the total number of helix predictions for a particular position in the alignment was > 0.5, then the position was assigned as helix; (b) when the fraction predicting turn/coil was > 0.5, then the position was assigned as turn/coil; (c) when the fraction predicting helix = fraction predicting turn/coil = 0.5, the position was assigned as turn/coil; (d) when the fraction predicting helix = fraction predicting turn/coil = 0.33, the position was assigned as turn/coil. Other rules were: all assignment of helix must result in more than four contiguous residues of helix, and assignment of turn/coil in more than three contiguous residues of turn/coil.

from SDS/PAGE, where vinculin co-migrates with *Escherichia coli* β-galactosidase (molecular mass 116349 Da calculated from the DNA sequence; Fowler & Zabin, 1978).

The sequence data reported here combined with biochemical analysis provides new insight into the nature of the major structural and functional domains of the vinculin molecule. On the basis of V8-proteinase cleavage and the corresponding sequence data, we estimate that the globular head domain of vinculin (Milam, 1985) consists of the first 850 amino acid residues from the *N*-terminus, and that the *C*-terminal tail largely consists of amino acid residues 851–1066. The cleavage between these two domains occurs in a proline-rich region, which probably plays an important role in the separation of the extended tail from the globular head domain.

The head/tail structure of vinculin would seem to define functionally important domains. Indirect binding studies of fluorophore-labelled vinculin to focal contacts (Avnur *et al.*, 1983), as well as direct electron-microscopic analysis (Milam, 1985), suggests that vinculin undergoes self-association into multimeric clusters through sites thought to be located in the tail domain of the molecule (Milam, 1985). In contrast, the gel overlay experiments reported here clearly show that binding of vinculin to talin occurs through the 90 kDa globular head of vinculin, as suggested by the preliminary data of others (Burridge & Mangeat, 1984; Milam, 1985). However, we were unable to find consistent evidence for talin binding to the 60 kDa fragment that is liberated on further digestion of the 90 kDa fragment with V8 proteinase. Neither have we observed any lower-molecular-mass fragment able to bind talin. One possible explanation for these observations is that the talin-binding site is contained within the 30 kDa fragment that must be liberated from the 90 kDa fragment on prolonged digestion with V8 proteinase, but that it is further degraded with corresponding loss of binding activity. The fact that the *N*-terminus of the 60 kDa fragment is isoleucine-244 indicates that the 30 kDa fragment must contain the *N*-terminus of the protein. The conclusion that the *N*-terminal region of vinculin contains a talin-binding domain is further supported by the transcription/translation experiments *in vitro* described here, which show that a vinculin polypeptide containing the first 398 amino acid residues retains talin-binding activity. In addition, when this same vinculin fragment was expressed in *Cos* monkey cells, it localized to focal contacts that contain talin (R. Bendori, D. Salomon & B. Geiger, unpublished work). Furthermore, site-directed mutagenesis points to amino acid residues 167–207 as important determinants of talin binding (G. J. Price, P. Jones, B. Patel & D. R. Critchley, unpublished work). Such studies suggest that the talin-binding site lies just *N*-terminal to the three 112-amino-acid-residue repeats that begin at residue 259.

The relationship between vinculin and the muscle-specific isoform of the protein termed metavinculin (Feramisco *et al.*, 1982; Siliciano & Craig, 1982; Geiger, 1982) has been partially clarified by the studies performed by Gimona *et al.* (1988). They used a combination of protein fragmentation and sequencing to demonstrate the presence of a unique 68-amino-acid-residue sequence in pig metavinculin. Interestingly, the sequence both started and terminated with a KWSSK motif. The availability of the complete chicken vinculin sequence

reported in the present paper allows the position of this insert to be located between residues 912 and 913, assuming co-linearity between the pig and chicken sequences. The results of the Southern-blot analysis are consistent with the existence of a single chicken vinculin gene, and it would therefore seem likely that the two protein isoforms arise through alternative splicing of a single primary transcript.

The elucidation of the complete amino acid sequence of chick vinculin, together with a knowledge of the partial sequence of human vinculin and peptide sequence from pig vinculin (Gimona *et al.*, 1988), indicate that the protein has been highly conserved throughout evolution from chick to man. It is noteworthy that, in contrast with this striking homology, most antibody reagents prepared against chick vinculin react poorly or not at all with the mammalian proteins, suggesting that immunopotent epitopes are scarce, and probably restricted to areas in which considerable species-specific variability exist.

Secondary-structure prediction analysis of the complete vinculin sequence indicates stretches of α-helix separated by regions of turn/coil distributed throughout the molecule, with little consistent prediction of β-strand. Although there are well-recognized uncertainties in empirical secondary-structure prediction methods, tests of the joint prediction method used here have shown that it correctly predicts α-helical and turn/coil regions in predominantly α-helical proteins with a 64–78% and 60–69% accuracy respectively (Eliopoulos *et al.*, 1982). Furthermore, the results of c.d. studies on vinculin (Evans *et al.*, 1984) are consistent with our prediction that the molecule contains a high α-helical content. The accuracy of the secondary-structure prediction method that we have used is likely to be considerably increased when a prediction is derived from a number of aligned sequences. Thus the consensus prediction for the aligned 112-amino-acid-residue repeats (three chick and three human) consisted of three helices joined by turn/coil regions. The occurrence of ten proline residues at alignment positions 29, 32, 33 and 35, four at positions 58 and six at positions 90 is consistent with their predicted turn/coil structure connecting the three helices. Only two proline residues at positions 7 exist within a region predicted to be helical. It may be noted that the end of a repeat and the beginning of the next is predicted to consist of one continuous helix (alignment positions 91–25; Fig. 8). The region leading to helix 1 at the start of repeat 1 and that leading out of helix 1 of repeat 3 are predicted to consist of helix, consistent with this observation. The function of the repeats is unknown, but repeat units of approximately similar length are found in other cytoskeletal proteins (Baron *et al.*, 1987*b*; Davison & Critchley, 1988).

Note added in proof (received 8 February 1989)

Further sequencing of the vinculin cDNA cVin5 showed that it lacked a segment of 123 bp 5′ to the *Bam*H1 site, and corresponding to amino acid residues 167–207 encoded by the 2.89 kb vinculin cDNA. The origin of this heterogeneity remains to be established.

REFERENCES

Avnur, Z., Small, J. V. & Geiger, B. (1983) J. Cell Biol. **96**, 1622–1630

Baron, M. D., Davison, M. D., Jones, P., Patel, B. & Critchley, D. R. (1987a) J. Biol. Chem. **262**, 2558–2561

Baron, M. D., Davison, M. D., Jones, P. & Critchley, D. R. (1987b) J. Biol. Chem. **262**, 17623–17629

Belkin, A. M. & Koteliansky, V. E. (1987) FEBS Lett. **220**, 291–294

Bendori, R., Salomon, D. & Geiger, B. (1987) EMBO J. **6**, 2897–2905

Bennett, J. P., Zamer, K. S. & Stossel, T. P. (1984) Biochemistry **23**, 5081–5086

Boller, K., Vestweber, D. & Kemler, R. (1985) J. Cell Biol. **100**, 327–332

Brenner, D. G. & Shaw, W. V. (1985) EMBO J. **4**, 561–568

Buck, C. A. & Horwitz, A. F. (1987) Annu. Rev. Cell Biol. **3**, 179–206

Burridge, K. (1986) Cancer Rev. **4**, 18–78

Burridge, K. & Connell, L. (1983) J. Cell Biol. **97**, 359–367

Burridge, K. & Mangeat, P. (1984) Nature (London) **308**, 744–746

Cowin, P., Kapprell, H. P., Franke, W. W., Tamkun, J. J. & Hynes, R. O. (1986) Cell **46**, 1063–1073

Dalgleish, R., Williams, G. & Hawkins, J. R. (1986) Hum. Genet. **73**, 91–92

Davison, M. D. & Critchley, D. R. (1988) Cell **52**, 159–160

Eliopoulos, E., Geddes, A. J., Brett, M., Pappin, D. J. C. & Findlay, J. B. C. (1982) Int. J. Biol. Macromol. **4**, 263–270

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., Helfmann, D. M. & Thomas, G. P. (1982) J. Biol. Chem. **257**, 11024–11031

Fowler, A. V. & Zabin, I. (1978) J. Biol. Chem. **253**, 5521–5525

Geiger, B. (1979) Cell **18**, 193–205

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., Kreis, T. E. & Schlessinger, J. (1984) Cell Muscle Motil. **5**, 195–234

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

Gimona, M., Small, J. V., Moermans, M., Van Damme, J., Puype, M. & Vandekerckhove, J. (1988) EMBO J. **7**, 2329–2334

Horwitz, A. F., Duggan, K., Buck, C. A., Beckerle, M. C. & Burridge, K. (1986) Nature (London) **320**, 531–533

Kellie, S., Patel, B., Wigglesworth, N. M., Critchley, D. R. & Wyke, J. A. (1986) Exp. Cell Res. **165**, 216–228

Laemmli, U. K. (1970) Nature (London) **227**, 680–685

Lazarides, E. & Burridge, K. (1975) Cell **6**, 289–298

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor

Matsudaira, P. (1987) J. Biol. Chem. **262**, 10035–10038

Milam, L. M. (1985) J. Mol. Biol. **184**, 543–545

Norrander, J., Kempe, T. & Messing, J. (1983) Gene **26**, 101–106

O'Halloran, T. & Burridge, K. (1986) Biochim. Biophys. Acta **869**, 337–349

Otto, J. J. (1983) J. Cell Biol. **97**, 1283–1287

Price, G. J., Jones, P., Davison, M. D., Patel, B., Eperon, I. C. & Critchley, D. R. (1987) Biochem. J. **245**, 595–603

Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5466

Siliciano, J. D. & Craig, S. W. (1982) Nature (London) **300**, 533–535

Tamkun, J. W., DeSimone, D. W., Fonda, D., Patel, R. S., Buck, C., Horwitz, A. F. & Hynes, R. O. (1986) Cell **46**, 271–282

Towbin, H., Staehelin, T. & Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4350–4354

Volk, T. & Geiger, B. (1986) J. Cell Biol. **103**, 1441–1450

Wachsstock, D. H., Wilkins, J. A. & Lin, S. (1987) Biochem. Biophys. Res. Commun. **146**, 554–560

Williams, J. G. & Mason, P. J. (1985) in Nucleic Acid Hybridization: A Practical Approach (Hames, B. D. & Higgins, S. J., eds.), pp. 139–160, IRL Press, Oxford