of these proteins are known to be influenced by the presence of other homeobox-containing proteins. Heterodimerization may be one means by which these combinatorial interactions occur. A few DNA-binding proteins could thus generate many different regulatory specificities.

Received 21 June; accepted 3 August 1990.


ACKNOWLEDGEMENTS: I thank R. T. Simpson, in whose laboratory this work was performed, for his advice and support; A. Hennetibo, R. T. Simpson, S. Y. Roth, R. H. Morali and A. Wolfels for critical reading of the manuscript, and C. Senti-Gyorgy for oligonucleotide synthesis.

Sequence and domain structure of talin

D. J. G. Rees‡, Sarah E. Ades†, S. J. Singer‡ & Richard O. Hyne‡

‡ Center for Cancer Research and † Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Talin is a high-molecular-weight cytoskeletal protein concentrated at regions of cell-substratum contact and in lymphocytes, at cell-cell contacts. Integrin receptors are involved in the attachment of adherent cells to extracellular matrices and of lymphocytes to other cells. In these situations, talin codistributes with concentrations of integrins in the cell surface membrane. Furthermore, in citrate binding studies suggest that integrins bind to talin, although with low affinity. Talin also binds with high affinity to vinculin, another cytoskeletal protein concentrated at points of cell adhesion. Finally, talin is a substrate for the Ca²⁺-activated protease, calpain II, which is also concentrated at points of cell-substratum contact. To learn more about the structure of talin and its involvement in transmembrane connections between extracellular adhesions and the cytoskeleton, we have cloned and sequenced murine talin. We describe a model for the structure of talin based on this sequence and other data. Homologies between talin and other proteins define a novel family of submembraneous cytoskeleton-associated proteins all apparently involved in connections to the plasma membrane.
The initial talin complementary DNA clone was isolated by antibody screening and validated by western blotting of fusion proteins and of talin and by immunofluorescence studies (data not shown). Antisera raised against the fusion proteins also reacted with talin in similar assays (data not shown, see Fig. 1). Further cDNA clones were then isolated by DNA hybridization to cover the complete sequence and selected clones were sequenced (Fig. 1). The sequence defines an open reading frame encoding 2,541 amino acids (Fig. 1). The position of the N-terminal methionine was confirmed by amino-acid sequencing (see Fig. 1 and below) and is encoded by a nucleotide sequence (CCACCATGG) conforming to the consensus typical of eukaryotic initiation sites. A typical polyadenylation signal, AATAAA, precedes a poly(A) tail by 11 nucleotides. Restriction enzyme analyses and partial sequence data on a large number of overlapping cDNA clones revealed no evidence for variant forms (alternative splicing etc.) other than the polymorphic residues described in Fig. 1.

The sequence is markedly nonhomogenous (Fig. 1) and can be considered in two blocks. The first 600 residues are highly polar (28% charged residues), whereas the last 1,900 residues are highly enriched in alanine (18%) and poor in aromatic residues (2%). The C-terminal 60 residues are highly charged (>30%). The division of the talin sequence into blocks is readily seen in Fig. 2a which shows a comparison of the mouse talin sequence with itself; whereas the first 600 residues show no internal homologies, the last 1,900 show apparent homologies. These homologies are due to the high content of alanine and...
FIG. 1. Amino-acid sequence (single-letter notation) of mouse talin deduced from cDNA sequence. The N-terminal sequence was confirmed by amino-acid sequencing of the 43K fragment of chicken gizzard talin which agreed with the deduced mouse sequence at 16 out of 17 positions commencing with residue 2 (underlined). The sequencing commences at residue 434 (underlined) is closely similar to the sequence obtained from the 190K fragment produced by calpain digestion of chicken gizzard talin (identities at 11 out of 18 positions). This sequence defines the point of calpain cleavage (arrow), which is preceded by a segment of protein (italics) that matches so-called PEST sequences, which have been hypothesized to be involved in protein degradation22-24. The segment homologous with band 4.1 and ezrin (see Figs 2 and 3) is boxed and the alanine-rich nature of the tail segment is highlighted by bolding of runs of alanine residues. The coding region is preceded by 159 nucleotides of 5' untranslated sequence and followed by a 440 nucleotide long 3'-untranslated region (all segments confirmed by analysis of the positional base bias25). The complete nucleotide sequence determined is 8,226 bases long without poly(A) tail, corresponding well with the size of the ~8.5-kilobase (kb) messenger RNA detected on northern blots (data not shown). The total cDNA sequence has been submitted to GenBank/EMBL/DDBJ (accession number: X56123).

METHODS. The initial cDNA clone was isolated from an oligo(dT)-primed Agt11 cDNA library prepared from poly(A) RNA isolated from mouse Swiss 3T3 cells, using a polyclonal anti-talin antibody affinity-purified on 210-240K proteins from Swiss 3T3 cells. This clone was confirmed by immunoblotting of the fusion protein, which was then used to affinity-purify antibodies that reacted with chicken talin on immunoblots of total chicken embryo fibroblast protein and stained focal contacts in 3T3 cells and chick embryo fibroblasts in immunofluorescence experiments. Further clones were then isolated from two other libraries, one of cDNA made from Balb/c 3T3 RNA, cloned in Agt11 (Clontech) and one prepared with cDNA from Balb/c 3T3 RNA using the Libranych system from Invitrogen and cloned in xZAPII according to the manufacturer's instructions. Selected clones were subcloned into M13mp18 or M13mp19 as sonicated fragments and sequenced using either the Klenow fragment of DNA polymerase26, or modified T7 DNA polymerase27. Sequence data were assembled and analyzed using the programs of Staden28 and UWGCG. Two residues were found to be variant in this analysis; residue 3,473 was either T or C, coding for either leucine or proline at amino-acid 1,105, and residue 6,698 was either A or T, coding for either methionine or lysine at amino acid 2,180. Talin was purified from chicken gizzards1 and cleaved with calpain II purified from bovine heart28. The 47K and 190K fragments were separated on 5% SDS-polyacrylamide gels and transferred to Immobilon membranes29. Bands were cut out and sequenced in an Applied Biosystems Protein Sequencer.

FIG. 2. Comparison of mouse talin amino-acid sequence with itself and with the sequences of human band 4.1 and ezrin/cytovillin/p81. a. Self comparison shows that the sequence can be considered in two parts. The first 600 residues are unrelated to the rest of the sequence. The last 1,900 residues show some degree of self-relatedness, but on examination this is largely attributable to the high content of alanine and does not reflect repeated structural units. b, c. Comparisons with band 4.1 and ezrin show in each case that a segment of talin (roughly residues 150-350) is homologous with the N-terminal 200 residues of each of these proteins. d. Comparison of band 4.1 with ezrin shows that these two proteins are homologous over the first 300 residues.

Methods. Comparisons were made using the DIAGON program of Staden with a window of 51 residues and a threshold of 540.

do not reflect any readily discernible repeating structures. Secondary structure predictions suggest a high content of α helix consistent with the high α-helix content of talin16, but the sequence is not consistent with a coiled-coil structure. These structural predictions are consistent with electron microscopy images of talin16, which show an elongated molecule (60 nm) with a globular head and a flexible tail.

Also consistent with the suggestion of two structural domains in talin is its cleavage by calpain II into two fragments of apparent relative molecular masses 47,000 and 190,000 (M, 47K and 190K) (ref. 14). We isolated these fragments from a calpain II digest of chicken gizzard talin and determined their N-terminal amino-acid sequences. Identities with the mouse talin sequence unambiguously identified the 47K fragment as the N terminus of talin and the 190K fragment as the C-terminal domain (Fig. 1). The calpain II cleavage site is before residue 434. The predicted Mr of the 433-residue N-terminal domain is 49,981 consistent with the apparent Mr of 47,000. The predicted Mr of the C-terminal fragment is 219,873, which is larger than its apparent Mr on SDS-PAGE (190K). The same is true for intact talin; actual Mr is 269,854K, apparent Mr is 225-235K. This discrepancy is likely to be a consequence of the unusual amino-acid composition and/or secondary structure of the C-terminal domain of talin.

The mouse talin sequence was compared with the Owl protein sequence database1 using the FASTA program. The C-terminal tail segment showed only weak relationships with various structural proteins. We attribute these apparent relationships to similarities in secondary structure (such as α helices) rather than to evolutionary relatedness. In particular, there are no significant relationships with other large cytoskeletal proteins with extended coiled α helices (myosin, tropomyosin) or repeating segments of α helix (spectrin, α-actinin, dystrophin). A segment from the N-terminal region of talin showed homology with two other proteins, however, as shown on the DIAGON plots (Fig. 2b and c) and displayed in Fig. 3. These proteins are band 4.1, originally identified in erythrocytes but also present in many other cells30,31, and a protein, known variously as ezrin32, cytovillin32,33 or p8123,24, originally described in
intestinal brush borders but also present in many other cell types. A segment of 200–220 amino acids is homologous in talin, ezrin and band 4.1 (Figs 2 and 3).

The existence of a homologous N-terminal domain in these three proteins seems likely to have a bearing on the interaction of each with protein components of the plasma membrane. Band 4.1 has been analysed most extensively; it links the spectrin–actin cytoskeleton of erythrocytes to the plasma membrane. Band 4.1 is reported to bind two integral membrane proteins, with highest affinity for glycoporin and lower affinity for the anion transporter, band III (ref. 29). The binding site for glycoporin has been mapped to a 30K fragment of band 4.1 that contains the region of homology with talin. Although the interactions of ezrin with other proteins have not been defined, it is clearly localized in a submembranous region in the brush border of intestinal epithelial cells and more generally in other cells. Ezrin also colocalizes with a variety of cytoskeletal elements.

Thus, both these proteins, like talin, are found at interfaces between the cytoskeleton and the plasma membrane and it seems likely that the domains homologous with the glycoporin-binding domain of 4.1 (see Fig. 3) are involved in membrane attachment. It has been suggested that talin binds to integrins. There are no obvious homologies between the cytoplasmic domains of glycoporin and integrin subunits. It is also worth noting that the affinity of band 4.1 for glycoporin is enhanced by phosphorylated derivatives of phosphatidylinositol; it will be of interest to determine whether the same is true for the binding of talin to integrins. It is not known whether or not this binding occurs through the 47K N-terminal domain, although the homology shown in Fig. 3 might suggest such a model. The availability of cDNA clones encoding this domain should facilitate further analyses of these questions.

Turning next to the C-terminal domains of the three homologous proteins, the available data suggest that they are spliced segment. Identities are marked by lines and conservative substitutions (D, E, K, L, I, V, M, C, Y, F, W, S, T) by colons. The region displayed exhibits 20% identity between talin and human band 4.1 (8.6 s.d. above random), 23% identity between talin and ezrin (9.8 s.d. above random) and 35% identity between ezrin and band 4.1 (16.5 s.d. above random).

FIG. 3 Comparison of the homologous N-terminal domains of talin, band 4.1 and ezrin. Alignment of the sequences of mouse talin, human ezrin and X. laevis 4.1 (ref. 41). Residues 19–46 in the human erythroid band 4.1 sequence can be alternatively spliced25; the homology between talin and the other proteins follows this alternatively spliced segment. Identities are marked by lines and conservative substitutions (D, E, K, L, I, V, M, C, Y, F, W, S, T) by colons. The region displayed exhibits 20% identity between talin and human band 4.1 (8.6 s.d. above random), 23% identity between talin and ezrin (9.8 s.d. above random) and 35% identity between ezrin and band 4.1 (16.5 s.d. above random).

FIG. 4 Models of the structures of talin, band 4.1 and ezrin. Each protein has a homologous N-terminal domain (shaded oval) followed by a domain rich in α-helices (hatched rectangles); different shading indicates that these domains are not homologous with each other. The C-terminal segments (chequered circles) in each protein are highly charged. Evidence exists for membrane and cytoskeleton binding sites in the indicated domains of band 4.1 and for vinculin binding in the 190K domain of talin. Membrane attachment through the homologous N-terminal domains of each protein is hypothesized.
involved in binding to other cytoskeletal proteins. The spectrin-actin binding site of band 4.1 is located in the C-terminal half of the protein in a segment predicted to be predominantly α-helical. The 190K fragment of talin binds to vinculin and is also predicted to contain significant α-helical structure. No binding data are available for ezrin; however, it also has a segment predicted to be α-helical. A reasonable argument can thus be made that the three proteins—talin, band 4.1 and ezrin—have analogous structures (Fig. 4). Each molecule has a homologous N-terminal domain, known to be involved in membrane attachment in band 4.1 and hypothesized to be involved in interactions of talin with integrins and of ezrin with unknown membrane protein(s). In each molecule this homologous domain is followed by a segment postulated to be rich in α helix. These α-helical segments are not homologous but, in the case of band 4.1 and talin, there is evidence for functional as well as structural analogy in the way each binds to cytoskeletal proteins—spectrin-actin and vinculin, respectively. Finally, each of the three proteins ends in a short highly charged segment.

Thus, the sequence data we present here, together with data from the literature and the homologies we describe, allow us to propose that talin is a member of a subfamily of submembranous cytoskeletal proteins involved in connections of major cytoskeletal structures to the plasma membrane. The homologies and analogies among the proteins in this family suggest many lines for future investigation.

X-ray structure of phospholipase A₂ complexed with a substrate-derived inhibitor

Marjolein M. G. M. Thunnissen, Eiso AB, Kor H. Kalk, Jan Drenth, Bauke W. Dijkstra†, Oscar P. Kuipers‡, Ruud Dijkman†, Gerard H. de Haas† & Hubertus M. Verheij‡

Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands
† Department of Enzymology and Protein Engineering, University of Utrecht, Center for Biomembranes and Lipid Enzymology, PO Box 80054, 3508 TB Utrecht, The Netherlands
‡ To whom correspondence should be addressed

PHOSPHOLIPASES A₂ play a part in a number of physiologically important cellular processes such as inflammation, blood platelet aggregation and acute hypersensitivity. These processes are all initiated by the release of arachidonic acid from cell membranes which is catalysed by intracellular phospholipases A₂ and followed by conversion of arachidonic acid to prostaglandins, leukotrienes or thromboxanes. An imbalance in the production of these compounds can lead to chronic inflammatory diseases such as rheumatoid arthritis and asthma. Inhibitors of phospholipase A₂ might therefore act to reduce the effects of inflammation, so structural information about the binding of phospholipase A₂ to its substrates could be helpful in the design of therapeutic drugs.

The three-dimensional structure is not known for any intracellular phospholipase A₂, but these enzymes share significant sequence homology with secreted phospholipases, for which some of the structures have been determined. Here we report the structure of a complex between an extracellular phospholipase A₂ and a competitively inhibiting substrate analogue, which reveals considerable detail about the interaction and suggests a mechanism for catalysis by this enzyme.

The inhibitor used in these studies is (R)-2-dodecanoyl-1-hexanol-phosphoglycolyl (Fig. 1). This inhibitor binds by over three orders of magnitude more strongly than the substrate (R)-1,2-di-dodecanoyl-glycerol-3-phosphocholine to phospholipase A₂. A porcine phospholipase A₂ mutant obtained by site-directed mutagenesis was used because of its better crystallization properties. Moreover, this mutant, in common with cellular phospholipases A₂, lacks residues 62–66 and has a high activity on monomeric and aggregated substrates. To increase the affinity for monomeric substrates still further, we created a single Trp phospholipase A₂ (W3F, L31W; see Table 1).

The complex of this mutant phospholipase A₂ and the inhibitor was crystallized (space group P2₁2₁2₁, with two...