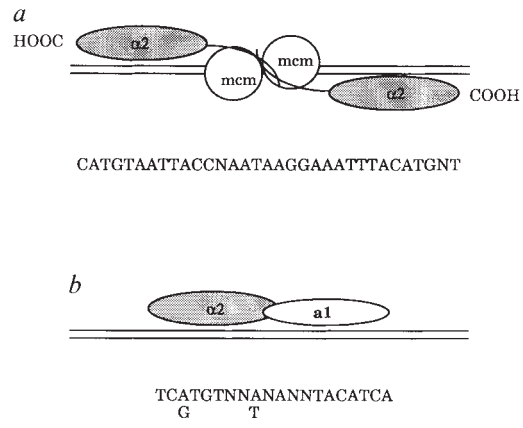



FIG. 4 Two modes by which α2 repressor regulates two different sets of genes. *a*, α2 bound to the operator of an **a**-specific gene. A dimer of α2 binds two half-sites situated 25 base pairs apart, on opposite faces of the helix[10]. A dimer of MCM1 occupies the centre of the operator[18–20]. A consensus α2/MCM1 sequence is shown, which was derived by Jarvis *et al.* from the **a**-specific genes *STE2, STE6, BAR1, MFa1* and *MFAa2* (ref. 29). *b*, α2 bound with **a**1 to the operator of a haploid-specific gene. The two proteins are shown binding as a heterodimer predominantly to a single face of the helix, although they wrap part of the way around the helix. A consensus sequence of **a**1/α2 operator DNA is shown, which was derived by Miller *et al.* from the haploid-specific genes *MATα1, STE5* (two repeated **a**1/α2 elements) and *HO* (10 elements)[5]. The centre of this sequence, unlike that of the α2/MCM1 operator, is not conserved. This supports the arguments that **a**1 and α2 are sufficient to bind the haploid-specific operator, and that **a**1 does not simply substitute for MCM1. The protein contacts seem to wrap part of the way around the helix (shown in Fig. 3*c*), perhaps in a way similar to that of λ repressor, which uses a flexible region of the protein to wrap around the DNA (for review see ref. 30). Alternatively, the DNA may wrap around the protein complex. The **a**1/α2 operator used here, like most haploid-specific operators[5], contains a short run of (dA) · (dT). Such short runs of A or T longer than three base pairs have been proposed to facilitate the wrapping of DNA around protein (for review see ref. 31).

in the second dimension. The upper spot has the mobility of α2; the lower spot, of **a**1. This demonstrates that **a**1 is actually present in the DNA–protein complex. Densitometric quantitation of the spots in duplicate gels revealed a ratio of **a**1 to α2 of 0.8 ± 0.1 (s.e.m.). One can conclude that each DNA–protein complex contains one molecule of α2 and one of **a**1.

Methylation interference was used to investigate which DNA bases make contact with the **a**1/α2 complex. These data are summarized in Fig. 3*c*, which shows the DNA–protein contacts positioned on the double helix. This pattern is quite different from the one produced by α2 at its operator at the **a**-specific genes[10]. At the **a**-specific genes, two α2 molecules bind cooperatively with a second, non-cell-type-specific protein, MCM1[18–20]. A dimer of MCM1 occupies the centre of the α2 operator[18,19], whereas the α2 dimer binds to the ends of the operator, on opposite faces of the double helix, spanning MCM1[10] (Fig. 4*a*).

The stoichiometry of binding demonstrated here excludes a model in which **a**1 substitutes for MCM1 at the haploid-specific genes. Figure 4*b* illustrates a model of α2 binding with **a**1 which is derived from this work. A heterodimer of **a**1 and α2 is shown binding cooperatively, predominantly to a single face of the helix.

The formation of multiple types of heterodimers is important in the regulation by several of the proteins that contain leucine zippers and the helix–loop–helix family of proteins (see ref. 21 for review). The **a**1 and α2 proteins are members of the class of DNA-binding proteins that contains a homeobox[22]. The homeobox-containing proteins have critical roles in development in other fungi[23] as well as in *Drosophila*, frogs and mammals (see ref. 24 for review). The regulatory activities of many 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.

1. Herskowitz, I. *Nature* **342,** 749–757 (1989).
2. Johnson, A. D. & Herskowitz, I. *Cell* **42,** 237–247 (1985).
3. Strathern, J., Hicks, J. & Herskowitz, I. *J. molec. Biol.* **147,** 357–372 (1981).
4. Goutte, C. & Johnson, A. D. *Cell* **52,** 875–882 (1988).
5. Miller, A. M., Mackay, V. L. & Nasmyth, K. A. *Nature* **314,** 598–603 (1985).
6. Hope, I. A. & Struhl, K. *Cell* **43,** 177–188 (1985).
7. Saiki, R. K. *et al. Science* **239,** 487–491 (1988).
8. Miller, A. M. *EMBO J.* **3,** 1061–1065 (1984).
9. Ner, S. S. & Smith, M. *Molec. cell. Biol.* **9,** 4613–4620 (1989).
10. Sauer, R. T., Smith, D. L. & Johnson, A. D. *Genes Dev.* **2,** 807–816 (1988).
11. Astell, C. R. *et al. Cell* **27,** 15–23 (1981).
12. Tatchell, K., Nasmyth, K. A., Hall, B. D., Astell, C. & Smith, M. *Cell* **27,** 25–35 (1981).
13. Siliciano, P. G. & Tatchell, K. *Proc. natn. Acad. Sci. U.S.A.* **83,** 2320–2324 (1986).
14. Hall, M. & Johnson, A. D. *Science* **237,** 1007–1012 (1987).
15. Smith, D. R., Jackson, I. J. & Brown, D. D. *Cell* **37,** 645–652 (1984).
16. Hope, I. A. & Struhl, K. *EMBO J.* **6,** 2781–2784 (1987).
17. Muller, M. *et al. EMBO J.* **7,** 4299–4304 (1988).
18. Keleher, C. A., Goutte, C. & Johnson, A. D. *Cell* **53,** 927–936 (1988).
19. Passmore, S., Elble, R. & Tye, B.-K. *Genes Dev.* **3,** 921–935 (1989).
20. Keleher, C. A., Passmore, S. & Johnson, A. D. *Molec. cell. Biol.* **9,** 5228–5230 (1989).
21. Jones, N. *Cell* **61,** 9–11 (1990).
22. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., DeRobertis, E. M. & Gehring, W. J. *Nature* **310,** 70–71 (1984).
23. Schulz, B. *et al. Cell* **60,** 295–306 (1990).
24. Gehring, W. J. & Hiromi, Y. *A. Rev. Genet.* **20,** 147–173 (1986).
25. Dranginis, A. M. *Molec. cell. Biol.* **9,** 3992–3998 (1989).
26. Melton, D. A. *et al. Nucleic Acids Res.* **12,** 7035–7056 (1984).
27. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
28. Ausubel, F. M. *et al.* (eds) *Current Protocols in Molecular Biology,* 12.3.1–12.3.6 (Wiley, New York, 1989).
29. Jarvis, E. E., Hagen, D. C. & Sprague, G. F. Jr *Molec. cell. Biol.* **8,** 309–320 (1988).
30. Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* **53,** 293–321 (1984).
31. Travers, A. *A. Rev. Biochem.* **58,** 427–452 (1989).

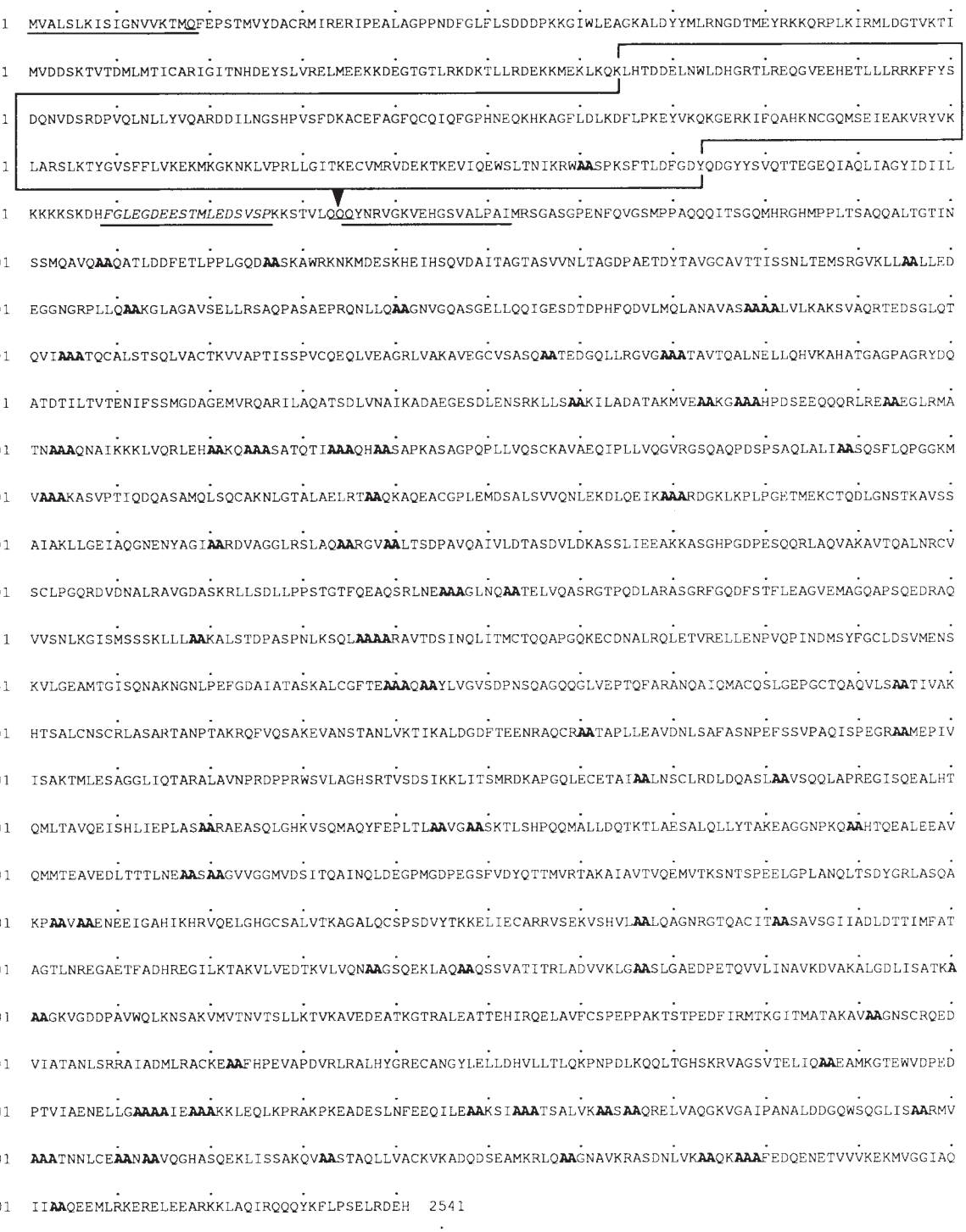
ACKNOWLEDGEMENTS. I thank R. T. Simpson, in whose laboratory this work was performed, for his advice and support, A. Hinnebusch, R. T. Simpson, S. Y. Roth, R. H. Morse and A. Wolffe for critical reading of the manuscript, and C. Szent-Gyorgyi for oligonucleotide synthesis.

# Sequence and domain structure of talin

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

* Center for Cancer Research and † Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
‡ Department of Biology, B-022, University of California San Diego, La Jolla, California 92093, USA

**TALIN is a high-molecular-weight cytoskeletal protein concentrated at regions of cell–substratum contact[1] and, in lymphocytes, at cell–cell contacts[2,3]. Integrin receptors are involved in the attachment of adherent cells to extracellular matrices[4,5] and of lymphocytes to other cells[6]. In these situations, talin codistributes with concentrations of integrins in the cell surface membrane[3,7–9]. Furthermore, *in vitro* binding studies suggest that integrins bind to talin, although with low affinity[10]. Talin also binds with high affinity to vinculin[11], another cytoskeletal protein concentrated at points of cell adhesion[12]. Finally, talin is a substrate for the $Ca^{2+}$-activated protease, calpain II[13,14], which is also concentrated at points of cell–substratum contact[14]. 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 submembranous 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[15]. 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. 2*a* 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 sequence commencing 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 protease cleavages[33,34]. 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 bias[35]). 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 λgt11 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 λgt11 (Clontech) and one prepared with cDNA from Babl/c 3T3 RNA using the Librarian system from Invitrogen and cloned in λZAPII 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 polymerase[36], or modified T7 DNA polymerase[37]. Sequence data were assembled and analysed using the programs of Staden[35] 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 gizzards[1] and cleaved with calpain II purified from bovine heart[38]. The 47K and 190K fragments were separated on 5% SDS-polyacrylamide gels and transferred to Immobilon membranes[39]. Bands were cut out and sequenced in an Applied Biosytems Protein Sequencer.

do not reflect any readily discernible repeating structures. Secondary structure predictions suggest a high content of $\alpha$ helix consistent with the high $\alpha$-helix content of talin[16], but the sequence is not consistent with a coiled-coil structure. These structural predictions are consistent with electron microscope images of talin[16], 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 ($M_r$ 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 $M_r$ of the 433-residue N-terminal domain is 49,981 consistent with the apparent $M_r$ of 47,000. The predicted $M_r$ of the C-terminal fragment is 219,873, which is larger than its apparent $M_r$ on SDS-PAGE (190K). The same is true for intact talin; actual $M_r$ is 269,854K, apparent $M_r$ 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 database[17] 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 $\alpha$ helices) rather than to evolutionary relatedness. In particular, there are no significant relationships with other large cytoskeletal proteins with extended coiled $\alpha$ helices (myosin, tropomyosin) or repeating segments of $\alpha$ helix (spectrin, $\alpha$-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. 2*b* and *c*) and displayed in Fig. 3. These proteins are band 4.1, originally identified in erythrocytes but also present in many other cells[18,19], and a protein, known variously as ezrin[20], cytovillin[21,22] or p81[23,24], originally described in

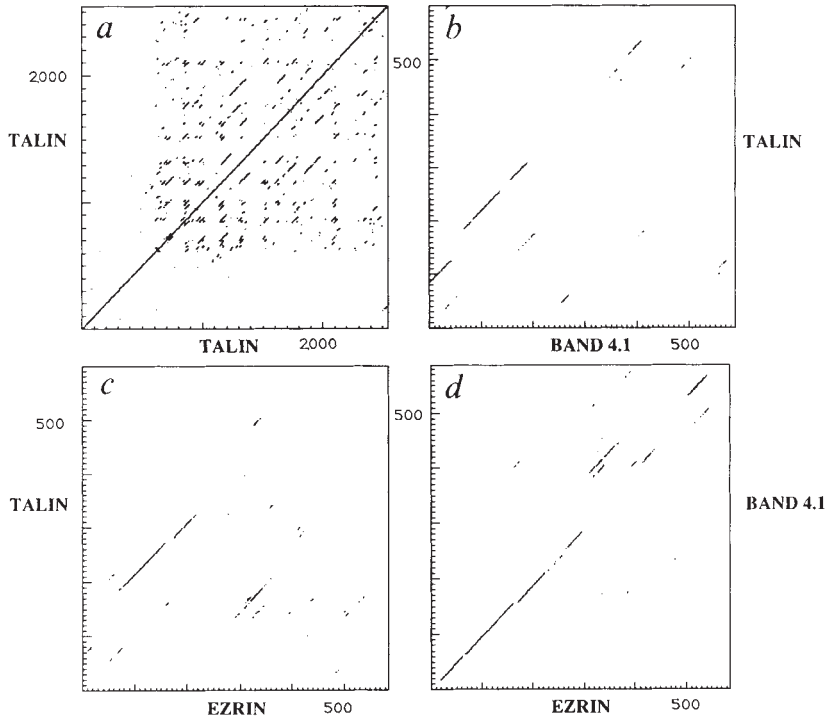



FIG. 2 Comparison of mouse talin amino-acid sequence with itself and with the sequences of human band 4.1 and ezrin/cytovillin/p81. *a*, Self comparison shows that the sequence can be considered in two parts. The first 600 residues are unrelated to the rest of the sequence. The last 1,900 residues show 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.

```
Human Band 4.1  47  WDNATSKTWLDSAKEIKKQV..RGVPWNFTFNVKFYPPDPA.QLTEDITRYYLCLQLRQDIVAGRLPCSFATLAL
Mouse Talin    165  LHTDDELNWLDHGRTLREQG..VEEHETLLLRRKFFYSDQNVDSRDPVQLNLLYVQARDDILNGSHPVSFDKACE
Human Ezrin     49  VDNKGFPTWLKLDKKVSAQEVRKENPLQFKFRAKFYPEDVAEELIQDITQKLFFLQVKEGILSDEIYCPPETAVL
X.laevis 4.1   239  WESPTCKVWLDPLKDIRKQV..HGGPCEFTSNVKFYPPDPA.QLSEDITRYYLCLQLRKDIFSGRLPCSFATLAL

Human Band 4.1 119  LGSYTIQSELGDYDPELHGVDYVSDFKLAPNQTKE........LEEKVMELHKSYRSMTPAQADLEFLENAKKLS
Mouse Talin    238  FAGFQCQIQFGPHNEQKHKAGFLDLKDFLPKEYVKQKG......ERKIFQAHKNCGQMSEIEAKVRYVKLARSLK
Human Ezrin    124  LGSYAVQAKFGDYNKEVHKSGYLSSERLIPQRVMDQHKLTRDQWEDRIQVWHAEHRGMLKDNAMLEYLKIAQDLE
X.laevis 4.1   311  LGSYTVQSEVGDYEEDLHGVDYVSEFKLSPNQTKD........LEEKVGELHKSYRSMTPAQADLEFLENAKKLT

Human Band 4.1 186  MYGVDLHKAKDLEG.....VDIILGVCSSGLLVYKDKLR.INRFPWPKVLKISYKR..SSFFIKIRPG    245
Mouse Talin    307  TYGVSFFLVKEKMKGKNKLVPRLLGITKECVMRVDEKTK.EVIQEWSLTNIKRWAASPKSFTLDFGDY    373
Human Ezrin    199  MYGINYFEIKNKKGTDLWLGVDALGLN...IYEKDDKLTPKIGFPWSEIRNISFND..KKFVIKPIDK    261
X.laevis 4.1   378  MYGVDIHQAKDLEG.....VDIKLGVCSGGLMVFKDNLR.INRFPWPKVLKISYKR..SSFFIKIRPG    437
```

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[24], human band 4.1 (refs 33 and 40) and *Xenopus* band 4.1 (ref. 41). Residues 19–46 in the human erythroid band 4.1 sequence can be alternatively spliced[42]; the homology between talin and the other proteins follows this alternatively spliced segment. Identities are marked by lines and conservative substitutions (D, E; K, R; 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 (9.6 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).

intestinal brush borders[25] but also present in many other cell types[26]. 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 glycophorin[27,28] and lower affinity for the anion transporter, band III (ref. 29). The binding site for glycophorin has been mapped to a 30K fragment of band 4.1 that contains the region of homology with talin[30]. 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[21]. Ezrin also colocalizes with a variety of cytoskeletal elements[31,32].

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 glycophorin-binding domain of 4.1 (see Fig. 3) are involved in membrane attachment. It has been suggested that talin binds to integrins[10]. There are no obvious homologies between the cytoplasmic domains of glycophorin and integrin subunits. It is also worth noting that the affinity of band 4.1 for glycophorin is enhanced by phosphorylated derivatives of phosphatidylinositol[28]; 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

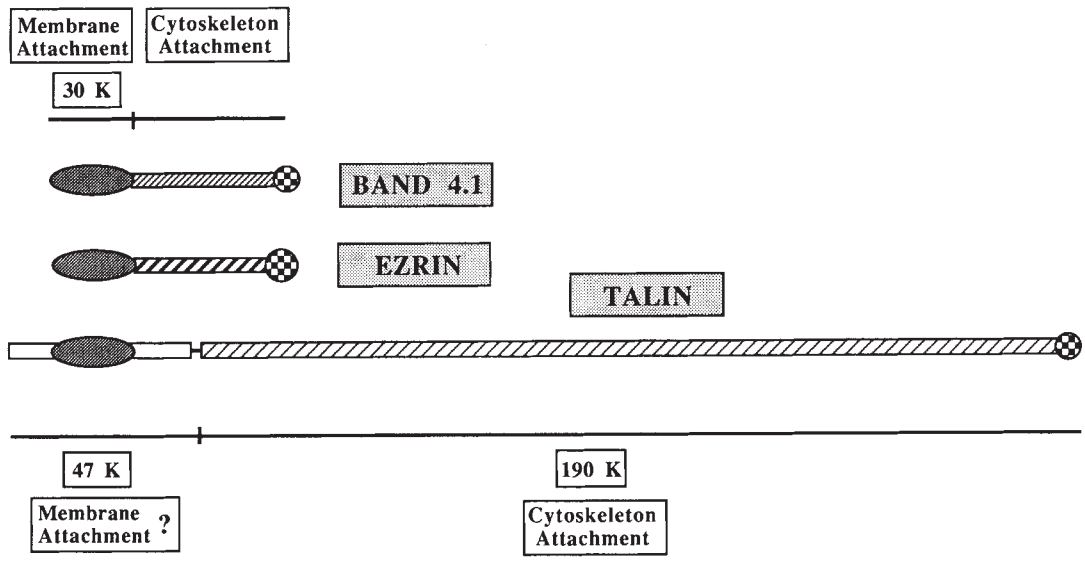



FIG. 4 Models of the structures of talin, band 4.1 and ezrin. Each protein has a homologous N-terminal domain (shaded oval) followed by a domain rich in $\alpha$ helix (hatched rectangles; different shading indicates that these domains are not homologous with each other). The C-terminal segments (chequered 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 $\alpha$-helical[33]. The 190K fragment of talin binds to vinculin[11] and is also predicted to contain significant $\alpha$-helical structure. No binding data are available for ezrin; however, it also has a segment predicted to be $\alpha$-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 $\alpha$ helix. These $\alpha$-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 that 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 family 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. □

Received 29 May; accepted 15 August 1990.


1. Burridge, K. & Connell, L. *J. Cell Biol.* **97,** 359–367 (1983).
2. Kupfer, A., Singer, S. J. & Dennert, G. *J. exp. Med.* **163,** 489–498 (1986).
3. Burn, P., Kupfer, A. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **85,** 497–501 (1988).
4. Hynes, R. O. *Cell* **48,** 549–554 (1987).
5. Ruoslahti, E. & Pierschbacher, M. D. *Science* **238,** 491–497 (1987).
6. Anderson, D. C. & Springer, T. A. *A. Rev. Med.* **38,** 175–194 (1987).
7. Chen, W. T., Hasegawa, T., Hasegawa, C., Weinstock, C. & Yamada, K. M. *J. Cell Biol.* **100,** 1103–1114 (1985).
8. Damsky, C. M., Knudsen, K. A., Bradley, D., Buck, C. A. & Horwitz, A. *J. Cell Biol.* **100,** 1528–1539 (1985).
9. Kupfer, A. & Singer, S. J. *J. exp. Med.* **170,** 1697–1713 (1989).
10. Horwitz, A., Duggan, E., Buck, C., Beckerle, M. C. & Burridge, K. *Nature* **320,** 531–533 (1986).
11. Burridge, K. & Mangeat, P. *Nature* **308,** 744–746 (1984).
12. Geiger, B. *Cell* **18,** 193–205 (1979).
13. Fox, J. E. B., Goll, D. E., Reynolds, C. C. & Phillips, D. R. *J. biol. Chem.* **260,** 1060–1066 (1985).
14. Beckerle, M. C., Burridge, K., DeMartino, G. N. & Croall, D. E. *Cell* **51,** 569–577 (1987).
15. Kozak, M. *Nucleic Acids Res.* **12,** 857–872 (1984).
16. Molony, L., McCaslin, D., Abernethy, J., Paschal, B. & Burridge, K. *J. biol. Chem.* **262,** 7790–7795 (1987).
17. Akrigg, D. *et al. Nature* **335,** 745–746 (1988).
18. Bennett, V. *A. Rev. Biochem.* **54,** 273–304 (1985).
19. Marchesi, V. T. *A. Rev. Cell Biol.* **1,** 531–561 (1985).
20. Bretscher, A. *Meth. Eznym.* **134,** 24–37 (1986).
21. Pakkanen, R., Hedman, K., Turunen, O., Wahlstrom, T. & Vaheri, A. *J. Histochem. Cytochem.* **35,** 809–816 (1987).
22. Turunen, O. *et al. J. biol. Chem.* **264,** 16727–16732 (1989).
23. Gould, K. L., Cooper, J. A., Bretscher, A. & Hunter, T. *J. Cell Biol.* **102,** 660–669 (1986).
24. Gould, K. L., Bretscher, A., Esch, F. S. & Hunter, T. *EMBO J.* **8,** 4133–4142 (1989).
25. Bretscher, A. *J. Cell Biol.* **97,** 425–432 (1983).
26. Pakkanen, R. *J. cell. Biochem.* **38,** 65–75 (1988).
27. Anderson, R. A. & Lovrien, R. E. *Nature* **307,** 655–658 (1984).
28. Anderson, R. A. & Marchesi, V. T. *Nature* **318,** 295–298 (1985).
29. Pasternack, G. R., Leto, T. L., Anderson, R. A. & Marchesi, V. T. *J. biol. Chem.* **260,** 3676–3683 (1985).
30. Leto, T. L., Correas, I., Tobe, T., Anderson, R. A. & Horne, W. C. in *Membrane Skeletons and Cytoskeletal Membrane Associations* (eds Bennett, V., Cohen, C. M., Lux, S. E. & Palek, J.) 201–209 (Liss, New York, 1986).
31. Birgbauer, E. & Solomon, F. *J. Cell Biol.* **109,** 1609–1620 (1989).
32. Goslin, K., Birgbauer, E., Banker, G. & Solomon, F. *J. Cell Biol.* **109,** 1621–1631 (1989).
33. Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. *Proc. natn. Acad. Sci. U.S.A.* **83,** 9512–9516 (1986).
33. Rogers, S., Wells, R. & Rechsteiner, M. *Science* **234,** 364–368 (1986).
34. Wang, K. K. W., Villalobo, A. & Roufogalis, B. D. *Biochem. J.* **262,** 693–706 (1989).
35. Staden, R. *Meth. Enzym.* **183,** 163–180 (1990).
36. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463–5467 (1977).
37. Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84,** 4767–4771 (1987).
38. DeMartino, G. N. & Croall, D. E. *Biochemistry* **22,** 6287–6291 (1983).
39. Matsudaira, P. *J. biol. Chem.* **262,** 10035–10038 (1987).
40. Tang, T. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85,** 3713–3717 (1988).
41. Giebelhaus, D. H., Eib, D. W. & Moon, R. T. *Cell* **53,** 601–615 (1988).
42. Tang, T. K., Qin, Z., Leto, T., Marchesi, V. T. & Benz, E. J. Jr *J. Cell Biol.* **110,** 617–624 (1990).


ACKNOWLEDGEMENTS. We thank M. Adams for the preparation of anti-talin antibodies, D. Croall for advice on the purification of calpain II, R. Cook and the MIT/HHMI Biopolymer Laboratory for protein sequencing and T. Hunter for generous support, encouragement and assistance. We also thank E. Pasquale, J. Pines, L. Connell, A. Tomecka and P. Johnson for advice. This work was supported by the US Public Health Service (R.O.H. and S.J.S.) and the Howard Hughes Medical Institute (R.O.H.). S.J.S. is an American Cancer Society Professor, and D.J.G.R. was the recipient of a NATO/SERC postdoctoral fellowship.

# X-ray structure of phospholipase $A_2$ 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 $A_2$ play a part in a number of physiologically important cellular processes such as inflammation, blood platelet aggregation and acute hypersensitivity[1,2]. These processes are all initiated by the release of arachidonic acid from cell membranes which is catalysed by intracellular phospholipases $A_2$ and followed by conversion of arachidonic acid to prostaglandins, leukotrienes or thromboxanes[3]. An imbalance in the production of these compounds can lead to chronic inflammatory diseases such as rheumatoid arthritis and asthma. Inhibitors of phospholipase $A_2$ might therefore act to reduce the effects of inflammation, so structural information about the binding of phospholipase $A_2$ to its substrates could be helpful in the design of therapeutic drugs. The three-dimensional structure is not known for any intracellular phospholipase $A_2$, but these enzymes share significant sequence homology[4–6] with secreted phospholipases, for which some of the structures have been determined[7–10]. Here we report the structure of a complex between an extracellular phospholipase $A_2$ 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-amino-1-hexanol-phosphoglycol (Fig. 1). This inhibitor binds by over three orders of magnitude more strongly than the substrate (R)-1,2-di-dodecanoyl-glycerol-3-phosphocholine[11] to phospholipase $A_2$. A porcine phospholipase $A_2$ mutant obtained by site-directed mutagenesis was used because of its better crystallization properties. Moreover, this mutant, in common with cellular phospholipases $A_2$, lacks residues 62–66 and has a high activity on monomeric and aggregated substrates[12]. To increase the affinity for monomeric substrates still further, we created a single Trp phospholipase $A_2$ (W3F,L31W; see Table 1).

The complex of this mutant phospholipase $A_2$ and the inhibitor was crystallized (space group $P2_12_12_1$, with two

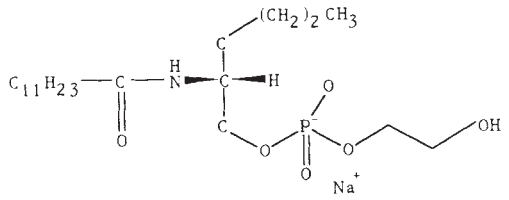



FIG. 1 The inhibitor (R)-2-dodecanoyl-amino-1-hexanol-phosphoglycol used in our investigation. The inhibitor was prepared from *nor*-leucine-methylester as described previously[22]. Inhibition measured at pH 8 by pH-stat titration using (R)-1,2-di-dodecanoyl-glycerol-3-phosphocholine (3 mM in 3 mM sodium taurodeoxycholate) as substrate[11] showed that the inhibitor binds 1,200 times more strongly than a natural substrate molecule.