

cDNA cloning and sequencing of the protein-tyrosine kinase substrate, ezrin, reveals homology to band 4.1

Kathleen L. Gould^{1,4}, Anthony Bretscher²,
Fred S. Esch^{3,5} and Tony Hunter¹

¹Molecular Biology and Virology Laboratory, The Salk Institute, San Diego, CA 92138, ²Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 and ³Laboratories for Neuroendocrinology, The Salk Institute, San Diego, CA 92138, USA

⁴Present address: ICRF Cell Cycle Control Laboratory, Microbiology Unit, South Parks Road, Oxford OX1 3QU, UK

⁵Present address: Athena Neurosciences, 800F Gateway Boulevard, South San Francisco, CA 94080, USA

Communicated by P. Nurse

Ezrin is a component of the microvilli of intestinal epithelial cells and serves as a major cytoplasmic substrate for certain protein-tyrosine kinases. We have cloned and sequenced a human ezrin cDNA and report here the entire protein sequence derived from the nucleotide sequence of the cDNA as well as from partial direct protein sequencing. The deduced protein sequence indicates that ezrin is a highly charged protein with an overall pI of 6.1 and a calculated molecular mass of 69 000. The cDNA clone was used to survey the distribution of the ezrin transcript, and the 3.2 kb ezrin mRNA was found to be expressed in the same tissues that are known to express the protein and at the same relative levels. Highest expression was found in intestine, kidney and lung. The cDNA clone hybridized to DNAs from widely divergent organisms indicating that its sequence is highly conserved throughout evolution. The amino acid sequence of ezrin revealed a high degree of similarity within its N-terminal domain to the erythrocyte cytoskeletal protein, band 4.1 and secondary structure predictions indicate that a second region of ezrin contains a long α -helix, a feature also common to band 4.1. The structural similarity of ezrin to band 4.1 suggests a mechanism for the observed localization to the membrane, and a role for ezrin in modulating the association of the cortical cytoskeleton with the plasma membrane.

Key words: band 4.1/cDNA/ezrin/sequencing

Introduction

Much work has been directed towards identifying and characterizing substrates of protein-tyrosine kinases with the expectation that some, if not all, phosphotyrosine-containing proteins are involved in controlling important cellular processes such as growth, transformation and differentiation. The number of identified substrates for tyrosine phosphorylation is still small. It includes the human homolog of the yeast cell cycle control gene, *cdc2* (Draetta *et al.*, 1988), *c-raf* (Morrison *et al.*, 1988), MAP-2 kinase (Ray and Sturgill, 1988), phosphatidylinositol kinase (Courtneidge and

Heber, 1987; Kaplan *et al.*, 1987), lactate dehydrogenase, phosphoglycerate mutase and enolase (Cooper *et al.*, 1983a), vinculin (Sefton *et al.*, 1981) talin (Pasquale *et al.*, 1986), the fibronectin receptor (Hirst *et al.*, 1986), calpactins I (Radke and Martin, 1979) and II (Fava and Cohen, 1984), ezrin (Gould *et al.*, 1986), a 21 kd subunit of the T cell receptor (Samelson *et al.*, 1986), and several proteins known only by their apparent molecular masses, p50, p42, p41, p120 (for review see Hunter and Cooper, 1985, 1986). In order to evaluate the importance of phosphorylation of these proteins on tyrosine to cellular phenotype, it is necessary to understand their structure and function and ultimately the effect of phosphorylation on their function.

We have shown previously that an 81 kd substrate for tyrosine phosphorylation in human cells is identical or highly related to an 80 kd component of chicken microvilli termed ezrin (Gould *et al.*, 1986). Ezrin is localized to microvilli and other plasma membrane structures in a wide variety of cell types and it partitions in biochemical fractionation experiments as if it is associated with or is a part of the cytoskeleton (Bretscher, 1983; Gould *et al.*, 1986). In the human epidermoid carcinoma cell line, A431, ezrin is redistributed rapidly into microvilli and membrane ruffles after epidermal growth factor (EGF) treatment, and its redistribution is coincident with an increase in tyrosine and serine phosphorylation (Bretscher, 1980). These observations raise the possibility that ezrin phosphorylation might be involved in regulating cell surface topography.

To gain more information about the structure, function and possible phosphorylation sites of ezrin, we have isolated and sequenced a human ezrin cDNA encoding the entire protein. In keeping with the proposed role of ezrin in cytoskeletal function, the sequence reveals a continuous stretch of ~180 amino acids which is strongly predicted to form an α -helix and is related to α -helical domains in other cytoskeletal proteins. Furthermore, the N-terminal 250 amino acids of ezrin are homologous to the N terminus of the erythrocyte cytoskeletal protein, band 4.1.

Results

An oligonucleotide probe based on the amino acid sequence EMYGINYF derived by Edman degradation of both chicken and human ezrin tryptic peptides identified two identical clones from an Okayama–Berg cDNA library of RNA from HeLa cells (Hanks, 1987). The DNA sequence of one clone predicted several of the sequenced ezrin tryptic peptides verifying that this cDNA was derived from ezrin mRNA. The partial cDNA was used to screen 300 000 additional colonies of the HeLa cDNA library. Of 35 clones which hybridized to both the partial cDNA and a second oligonucleotide probe based on the amino acid sequence EVYWFG the longest cDNA insert (from clone F6) was selected for sequence analysis.

AGCGAGGGCGGGCGGGCGCTCTAAGGGTTCTGCTCTGACTCCAGGTTGGGACAGCGTCTTCGCTGCTGCTGGATAGTCGTGTTTTCGGGGATCGAGGATACTCACCAGAAACCGAAAATG 120

CGGAAACCAATCAATGTCGAGTTACCACCATGGATGCGAGCTGGAGTTTGCATCCAGCCAAATACAACCTGGAAAACAGCTTTTTGATCAGTGGTAAAGACTATCGGCCTCCGGGAA 240
 P K P I N V R V T T M D A E L E F A I Q P N T T G K Q L F D Q V V K T I G L R E 40
 P K P I N V R V V T V D A W L E F A I Q P N T T G K Q L L F D Q V V K T I

GTGTGGTACTTTGGCCTCCACTATGTGGATAATAAAGGATTTCTACCTGGCTGAAGCTGGATAAGAAGGTGTCTGCCAGGAGGTGAGGAAGGAGAATCCCTCCAGTTCAAGTTCGGG 360
 V W Y F G L H Y V D N K G F P T W L K L D K K V S A Q E V R K E N P L Q F K F R 80

GCGAAGTTCTACCTCGAAGTGTGGCTGAGGAGCTATCCAGGACATCACCAGAAACTTTTCTTCCCAAGTGAAGGAAGGAATCCTTAGCGATGAGATCTACTGCCCCCTGAGACT 480
 A K F Y P E D V A E E L I Q D I T Q K L F F L Q V K E G I L S D E I Y C P P E T 120

GCCGTGCTCTTGGGGTCTACGCTGTGAGGCAAGTTTGGGGACTACAACAAGAAGTGCACAAGTCTGGGTACCTCAGCTCTGAGCGGCTGATCCCTCAAAGTGTGACCAGC 600
 A V L L G S Y A V Q A K F G D Y N K E V H K S G Y L S S E R L I P Q R V M D Q H 160

AAACTTACCAGGGACCGTGGGAGGACCGGATCCAGGTGTGGCATGCGGAACCCGTGGGATGCTCAAAGATAATGCTATGTTGGAATACCTGAAGATTGCTCAGGACCTGAAAATGTAT 720
 K L T R D Q W E D R I Q V W H A E H R G M L K D N A M L E Y L K I A Q D L E M Y 200
 E N A M L E Y L K I A Q D L E M Y

GGAATCAACTATTTTCGAGATAAAAAACAAGAAAGGAACAGACCTTTGGCTTGGAGTTGATGCCCTTGGACTGAATATTTATGAGAAAGATGATAAGTTAACCCCAAAGATTGGCTTTCTCT 840
 G I N Y F E I K N K K G T D L W L G V D A L G L N I Y E K D D K L T P K I G F P 240
 G I N Y F E I K

TGGAGTGAATCAGGAACATCTCTTCAATGACAAAAAGTTTGTCTTAAACCCATCGACAAGAAGGCACCTGACTTTGTGTTTTATGCCCCAGCTCTGAGAATCAACAAGCGGATCCTG 960
 W S E I R N I S F N D K K F V I K P I D K K A P D F V F Y A P R L R I N K R I L 280

CAGCTCTGCATGGGCAACCATGAGTTGTATATGCGCCGAGGAAGCCTGACACCATCGAGGTGCAGCAGATGAAGGCCAGGCCGGGAGGAGAAGCATCAGAAGCAGCTGGAGCGGCAA 1080
 Q L C M G N H E L Y M R R R K P D T I E V Q Q M K A Q A R E E K H Q K Q L E R Q 320
 Q

CAGCTGAAAACAGAGAAGAAAAGGAGAGAAAACCGTGGAGAGAGAGAAAAGAGCAGATGATGCGCGAGAAGGAGGAGTTGATGCTGCGGCTGCAGGACTATGAGGAGAAGACAAAAGAAAGGCA 1200
 Q L E T E K K R R E T V E R E K E Q M M R E K E E L M L R L Q D Y E E K T K K A 360
 Q L E N E K E K E Q M L R L Q E Y E V K

GAGAGAGACTCTCGGACAGATTGAGAGGGCCCTGCAGCTGGAGGAGGAGGAAAGCGGGCACAGGAGGCGGAGCCGAGCGCTAGAGGCTGACCGTATGGCTGCACTGCGGGTAAAGGAG 1320
 E R E L S E Q I Q R A L Q L E E E R K A Q E E A E R L E A D R M A A L R A K E 400

GAGCTGGAGAGACAGCGGTGGATCAGATAAAGAGCCAGGAGCAGCTGGCTGCGGAGCTGCAGAAATACACAGCCAAAGATTGCCCTCTGGAAGAGGCGGAGGCGCAAGGAGGATGAA 1440
 E L E R Q A V D Q I K S Q E Q L A A E L A E Y T A K I A L L E E A R R R K E D E 440
 S Q E Q L A A E L A E Y T A K I A L L E E A R E S E

GTTGAAGAGTGGCAGCAGCGCCAAAGAACCCAGGATGACCTGGTGAAGACCAAGGAGGAGCTGCACCTGGTGTGACAGCAGCCCGCCCGCCACCACCCCGGTGTACGAGCCGGTG 1560
 V E E W Q H R A K E A Q D D L V K T K E E L H L V M T A P P P P P P P V Y E P V 480
 V E E W Q I R

AGCTACCATGTCCAGGAGAGCTTGCAGGATGAGGGCGCAGAGCCACGGGCTACAGCGGAGCTGTCTAGTGAGGGCATCCGGGATGACCGCAATGAGGAGAAGCGCATCACTGAGGCA 1680
 S Y H V Q E S L Q D E G A E P T G Y S A E L S S E G I R D D R N E E K R I T E A 520

GAGAAGAACGAGCGTGTGACGCGGAGCTCGTACGCTGAGCAGCGAGCTGCCAGGCCGAGATGAGAATAAGAGGACCCACAATGACATCATCCACAACGAGAATGAGGCAAGGC 1800
 E K N E R V Q R Q L V T L S S E L S Q A R D E N K R T H N D I I H N E N M R Q G 560
 Q G

CGGACAAGTACAAGACGCTGCGGCAGATCCGGCAGGGCAACACCAAGCAGCGCATCGACGAGTTCGAGGCCCTGTAAACAGCCAGGCCAGGCAAGGGCAGAGGGTGTCTATAGCGGG 1920
 R D K Y K T L R Q I R Q G N T K Q R I D E F E A L * 585
 R D K Y Q R I D X F K

CGCTGCCAGCCCCGCCAGCTGTCTTTAGTGTCTCAAGTCTAGGAACTCCCTCAGATCCCAGTCTCTTTAGAAAGCAGTTACCCAACAGAAACATTCTGGGCTGGGAACAGGGAGGGCG 2040

CCCTGGTTGTTTTCCCCAGTTGTAATAGTCCAAAGCAGGCTGATTCTCGGATTATTCTCGAATCACCTCCTGTGTGTGCTGGGAGCAGGACTGATTGAATTACGGAAAATGCCTGT 2160

AAAGTCTGAGTAAGAACTTCACTGCTGGCCTGTGTATACAAGAGTCAAGTATCAATTAAGGAAACGTGGCAGGACTTCCATCTGTGCCATACTGTTCTGTATTGAAATGAGCTCAAAT 2280

TGATTTTTTAATTTTATGAAAGATCCATCTTTGTATATTTACATGCTTAGAGGGGTGAAAATATTTTTGAAAATTGAGTCTGAAGCACTCTCGCACACAGTATTCCCTCCTCCCG 2400

TCACTCCACGAGCTGGCAGAGAGCAGTATCACCAGCGTGAAGTGGTGGAGGAGGACTTGGATATTTTTTAGTCTTTTTTTTTGGCTTAAAGTATTTAGAATACATTGACTT 2520

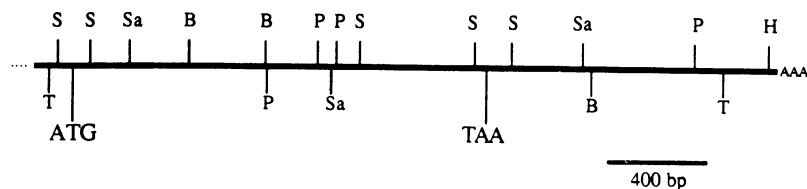
ATACACCTTATTAATGATCAGCTATATACTATTTATATACAAGTGATAATACAGATTTGTAACATTAGTTTTAAAAAGGGAAGTTTTGTTCTGTATATTTGTTACCTTTTACAGAATA 2640

AAAGAATTACATATGAAAAACCCCTCTAAACCATGGCAGTGTGATGTGATGGCAGGAGGAGTGGTGGAGCTGGACCTGCCTGCTGCAGCTGCAGTCAAGTAAACAGGATTATTATT 2760

AGTGTATGATGATGAAAGTATGACACTTTTAAATTTTGTGACAGTTACACATGCCACTATGAGCTTTTCAAGTCCAGCTGTGAAGAGACTCTGTCTGCTGTGTTGTTGCGAGT 2880

CTCTCTGCCATGGCCTTGGCAGGCTGTGGAAGGCAGCTTGTGGAGGCCGTTGGTCCGCCACTCATTCTCTCGTGCAGTCTTCTCTTCCACAGCTAAGATGCCATGTGCAGG 3000

TGGATCCATGCCGAGACATGAAATAAAGCTTTGCAAAGGC[A]_n 3044



The 3044 nucleotide long sequence of ezrin clone F6 and the predicted protein sequence of ezrin are shown in Figure 1. The 3'-untranslated region is 1166 bp long and contains a canonical poly(A) addition signal (AAUAAA) (Proudfoot and Brownlee, 1976) ending 15 nucleotides from the start of the poly(A) tract. Two pieces of evidence suggested that the protein initiates at the AUG codon at positions 118–120. Firstly, Edman degradation of intact chicken ezrin performed by P.Matsudaira (MIT, Cambridge, MA) yielded the sequence beginning PKPIN predicting that the protein was initiated at a methionine preceding PKPIN and that the methionine was removed by a post-translational processing event (see Figure 1). Secondly, it is the first AUG in the cDNA after an inframe stop codon at nucleotides 22–24.

The amino acid sequence deduced from the single open reading frame included all of the sequenced human and related chicken ezrin tryptic peptides and coded for a protein with a calculated molecular mass of 69 290. Since the purified protein migrated on 12.5% SDS–polyacrylamide gels as if it were considerably larger (81 kd), sequences containing the open reading frame were subcloned into an expression vector to ensure that they could direct expression of a full length protein. RNA was synthesized from ezrin cDNA in pGEM-4Z using SP6 polymerase and translated in a mRNA-dependent rabbit reticulocyte lysate in the presence of [³⁵S]methionine. The protein synthesized *in vitro* co-migrated on an SDS–polyacrylamide gel with ³⁵S-labeled ezrin immunoprecipitated from HeLa cells (Figure 2) establishing that clone F6 contained all ezrin protein coding sequences and that the initiator codon had been identified correctly.

A hydrophathy plot of ezrin showed that there are no long hydrophobic stretches which could encode a signal peptide or a transmembrane domain (Figure 3A). The amino acid composition indicated that ezrin has a large percentage of charged amino acids (38.5%), and this may explain the anomalously slow migration of ezrin on SDS–polyacrylamide gels. However, the average pI of the protein is 6.15, in excellent agreement with its previously reported migration on isoelectric focusing gels (Cooper *et al.*, 1983b; Hunter and Cooper, 1981, 1983; Bretscher, 1983; Gould *et al.*, 1986). Secondary structure analysis of ezrin predicted that residues 290–470 form an α -helix (Figure 3A).

A plot of the predicted ezrin sequence against itself indicates that there is a region within the protein containing short (7–10 amino acids), repeated similarities (Figure 3B). This region corresponds exactly to the domain predicted to form an α -helix. Three larger repeat units can also be identified in this stretch of amino acids; the optimal alignment of these three units is presented in Figure 3C. There is 13% identity and an additional 13% conservation between the three units aligned in this manner. If only repeats 1 and 2 are considered, then 29% of the residues (16 out of 55) are identical and an additional eight residues are similar in structure and charge.

Comparison of the deduced protein sequence for ezrin with protein sequences entered into the EMBL and SWISSPROT databases revealed that two regions of the protein possess similarity to other proteins in the databases. The region predicted to form an α -helix (amino acids ~299–366) has limited sequence similarity (15–26%) to cytoskeletal proteins containing α -helices including myosin, tropomyosin, keratin, lamin, desmin and troponins. The second region of ezrin that shares sequence similarity to a previously sequenced protein is the N-terminal ~260 amino acids. Amino acids 3–257 are 34% identical to the N-terminal ~240 amino acids of the erythrocyte cytoskeletal protein, band 4.1. In addition a large number of other residues show conservative substitutions in this region. An alignment of these sequences is shown in Figure 4.

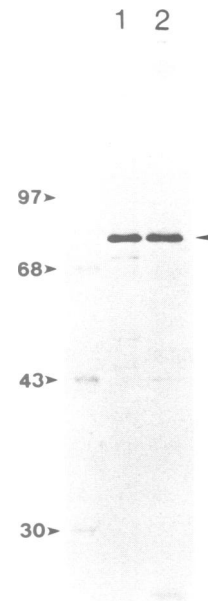


Fig. 2. Expression of ezrin from the cDNA clone. The putative coding region of ezrin clone F6 was inserted in the expression vector pGEM-4Z (Promega) as described in Materials and methods in an orientation that allowed synthesis of sense RNA from the SP6 promoter. Capped RNA was translated in a nuclease-treated rabbit reticulocyte lysate containing [³⁵S]methionine. The total reaction products in the *in vitro* translation reaction were analyzed on a 12.5% polyacrylamide gel (lane 2) next to an immunoprecipitate of ezrin from [³⁵S]methionine-labeled HeLa cells (lane 1) generated as reported previously (Gould *et al.*, 1986). Following fluorography, the gel was exposed to pre-sensitized Kodak XAR film for 3 days at -70°C . The position of ezrin is indicated with an arrowhead on the right. Mol. wt markers are phosphorylase b (97 kd), BSA (68 kd), ovalbumin (43 kd) and carbonic anhydrase (30 kd).

Fig. 1. Nucleotide sequence of the ezrin clone F6 and deduced primary structure of the encoded protein. The DNA sequence was determined as described in Materials and methods and used to predict the protein sequence. Human ezrin tryptic peptides sequenced by Edman degradation are underlined. Chicken ezrin tryptic peptide sequences were aligned with the human sequence and written underneath. The N-terminal sequence of chicken ezrin was obtained by Edman degradation of intact ezrin by P.Matsudaira (MIT). The nucleotide and amino acid which could not be identified unambiguously are denoted by Xs. The nucleotide sequence is numbered starting with the first base in clone F6. The amino acid sequence is numbered starting with the second amino acid as the mature protein apparently lacks the initiating methionine (P.Matsudaira, personal communication). The polyadenylation consensus sequence found near the end of the cDNA is underlined. A restriction map of the ezrin cDNA is shown below the sequence. The putative initiator ATG and terminator TAA codons are shown as are the sites of various restriction enzymes. B = *Bam*HI, P = *Pst*I, S = *Sau*3A, Sa = *Sac*I, T = *Th*1111.

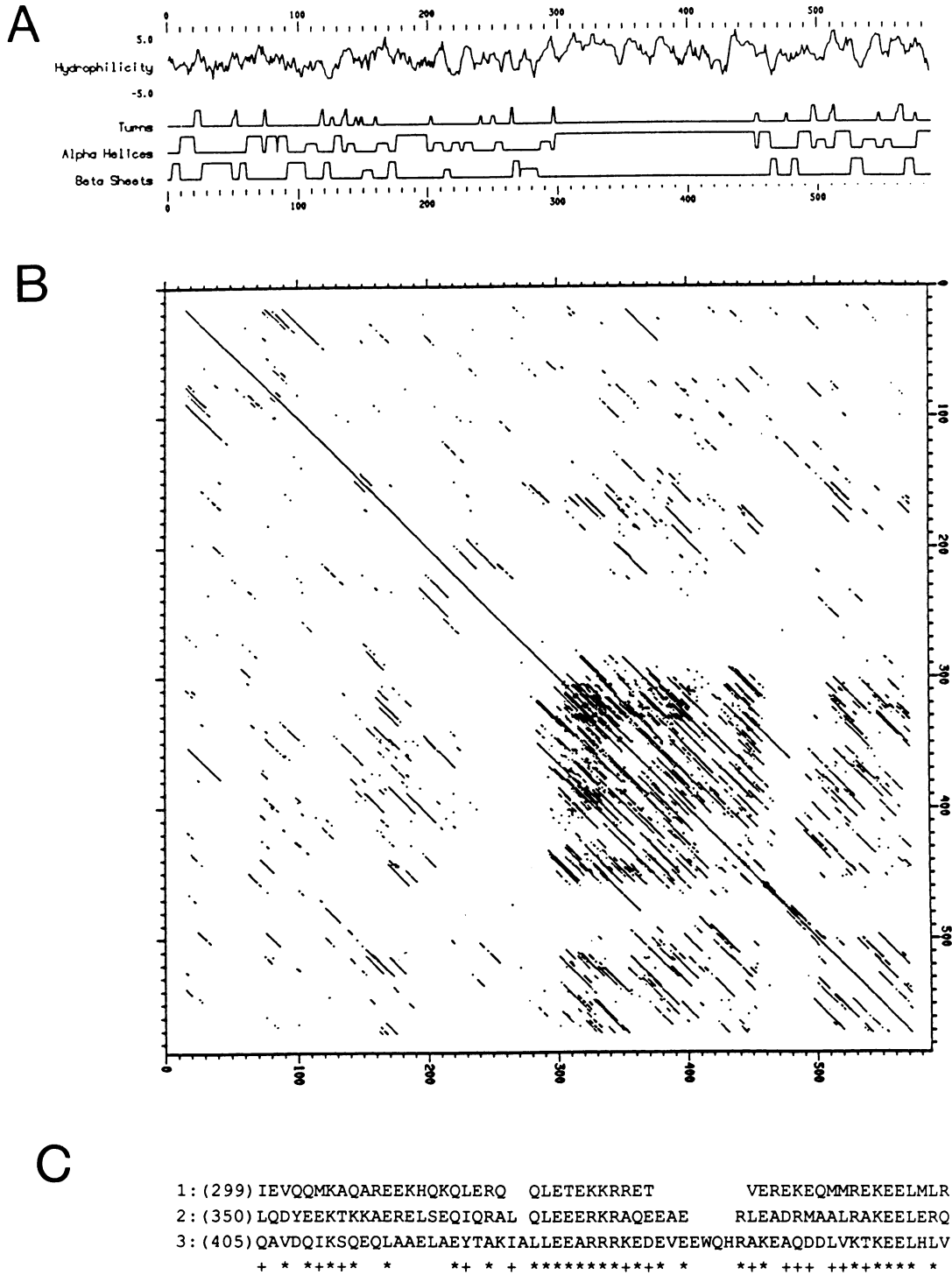


Fig. 3. Secondary structure predictions of ezrin and internal similarities. **(A)** A hydropathy plot was obtained from the predicted sequence of ezrin by the PEPTIDESTRUCTURE and PEPLOT programs of UWGCG according to the paradigms of Kyte and Doolittle (1982). Secondary structure predictions for ezrin were also obtained by the UWGCG programs based on the algorithms of Chou and Fasman (1978). The axes are labeled in residue numbers. **(B)** The predicted amino acid sequence was analyzed for internal similarities by the UWGCG program COMPARE with a window length of 30 and a stringency of 12. The output of this comparison was plotted by DOTPLOT (UWGCG). The lines parallel to the central diagonal represent internal similarities. The axes are labeled in residue numbers. **(C)** The predicted α -helical region of ezrin which also appeared to have the greatest density of internal similarities was aligned by BESTFIT (UWGCG) and then by eye into three repeat units. Two gaps were introduced to optimize the alignment. The number of the first residue in each repeat is shown in parentheses. Residues identical in two or three of the repeats are denoted with asterisks and conserved residues in two or three of the repeats are indicated with + signs.

Expression of ezrin mRNA

The expression and size of ezrin mRNA was examined in a variety of human and two murine cell lines by Northern

blotting. RNA expression was highest in A431 epidermoid carcinoma and CCRF-CEM T lymphoma cells (Figure 5, lanes 2 and 7), and intermediate in HeLa, K562 and ANN-1

Ezrin	1	PKPINVRVTMDAE LE FAIQPNTT GKQ LFDQVVKTIGLRE VWYFGL HVDNKGF PTWL KL	60
Band 4.1	1	MHCKVSLDDTVY ECV VEKHAK GQD LLKRVCEHLN LEEDYFGL AIWDNATSK TWLD S	58
Match	 D . . . E G . . L . . . V L . E . . YFGL . . . DN TWL . .	
Ezrin	61	DKK VSAQ EV RKEN PLQ FKF RA KFY PE DVAE ELIQDIT QKLF FL . . Q VKE GIL SDEI Y CPP	118
Band 4.1	59	A KEIK Q . VR . G VPWN FT FN V KFY PP DPAQ . L T E D I T R . . YYL CL Q L RQ D I V AGRL PC SF	110
Match		. K Q . VR . . . P . . F KFY P . D . A . . L . . DIT L . Q I C . .	
Ezrin	119	ET AVLL GS YAV Q AKF GD YNKE V HK SG YL SS ERL IP QR VMD QHK L TRD Q WED RI QV W HAEH	178
Band 4.1	111	AT LALL GS YTI Q SEL GD YD PE LHG V DYV S DFK L AP NQ T KE . L EE KV M EL H KSY	165
Match		. T . . LLGSY . . Q . . GDY . . E . H S . . L . P T E H . . .	
Ezrin	179	RG MLK DN AM LE YLK IA QD LE MYG IN YFE IK NK KG TD L WLG VD AL GL NI Y EK DD KL TP KIG	238
Band 4.1	166	R SM TP AQ AD LE F LEN AK KL S MYG VD LHK AK DLE GV DI LG VC SS G LL VY . KD . KLR . INR	222
Match		R . M A . LE . L . . A . L . MYG K . . . G . D . . LGV . . . GL . . Y . KD . KL	
Ezrin	239	FP WSE IR NIS F ND KK FV IK P ID KK AP DF V F . Y AP RL R IN KR IL Q LC MGN HE LY M RRR K P DT I	299
Band 4.1	223	FP W PK V L KIS Y KR SS FF IK I R P GE Q E Q Y EST IG F KL P SY RA AK LW K VC VE HHT FF RL T STD	285
Match		FPW IS F . IK L R . . . T	

Fig. 4. Homology between ezrin and band 4.1. Amino acid residues 1–299 of ezrin are aligned with residues 1–285 of human band 4.1 (Conboy *et al.*, 1986). Identities are shown as bolded residues in each sequence, and also in the match line below. Gaps in the two sequences introduced to maximize the alignment are indicated (.).

cells (Figure 5, lanes 3, 8 and 11). In rat tissue RNAs, the highest level of ezrin mRNA expression was found in skin, small intestine, kidney and lung (Figure 6, lanes 1, 6, 7 and 9). Ezrin mRNA was also detected at lower levels in RNAs from ovary, heart, brain, spleen, thymus, cerebellum and submaxillary gland and was not detected in RNA samples from testes, skeletal muscle or liver (Figure 6, lanes 2, 3, 5, 8 and 11).

The size of the major human ezrin mRNA was estimated based on its migration relative to 18S and 28S rRNAs to be 3.2 kb in length. As clone F6 is 3044 nucleotides long without its poly(A) tail, clone F6 is probably close to full length although primer extension experiments would have to be performed to determine its precise 5' end. The nature of the three other RNA species from human cells is unknown. However, since their presence correlates exactly with the presence of the major transcript, they are likely to be related by incomplete or alternative splicing. In mouse and rat cells, there were two predominant ezrin mRNAs of similar size to the major human ezrin mRNA (see Figure 6). The nature of the difference between them is not known. As in human cells, a larger RNA species was present which probably represents unspliced RNA.

Ezrin genomic sequences

A Southern blot of human A431 cell DNA digested with a variety of restriction enzymes was probed with four single-stranded segments corresponding to the entire ezrin clone F6. At high stringency the sizes of bands detected in most digests added up to only 5–6 kb (Figure 7). As the cDNA is 3.0 kb in length, there is most likely a single gene encoding ezrin.

The similarity of tryptic peptide sequences between human and chicken ezrin indicated that the protein was highly conserved throughout evolution (Gould *et al.*, 1986). To determine if human ezrin genomic sequences could be detected in other species, we probed a Southern blot containing DNAs from a variety of different organisms

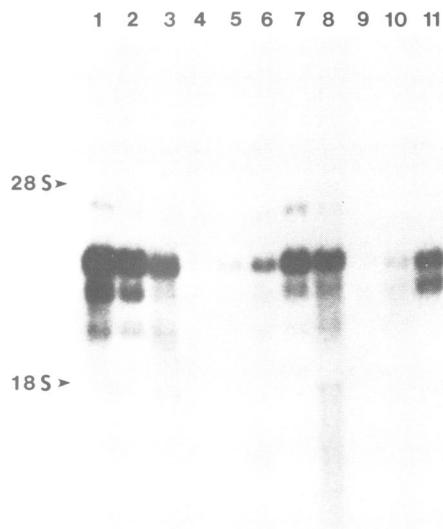


Fig. 5. Ezrin mRNA size and cell type expression. Ten μ g of total RNA (lanes 2–9) or 2 μ g of poly(A)-selected RNA (lanes 1, 10 and 11) were separated on a 1% agarose–formaldehyde gel, transferred to nitrocellulose and hybridized with a 32 P-labeled single-stranded fragment of the ezrin clone as described in Materials and methods. RNA samples were from human HeLa cervical carcinoma cells (lanes 1 and 3), A431 epidermoid carcinoma cells (lane 2), MG-63 osteosarcoma cells (lane 4), HepG2 hepatoma cells (lane 5), 1321N1 astrocytoma cells (lane 6), CCRF-CEM T lymphoma cells (lane 7), K562 chronic myelogenous leukemia cells (lane 8), SK-N-SH neuroblastoma cells (lane 9), murine AtT20 pituitary tumor cells (lane 10) and ANN-1 Abelson murine leukemia virus-transformed NIH3T3 cells (lane 11). The stringency of the final wash was $0.1 \times$ SSPE, 0.1% SDS at 50°C. 28S and 18S rRNAs were used as molecular size markers. Exposure time was 3 days with pre-sensitized Kodak XAR film at -70°C with an intensifying screen.

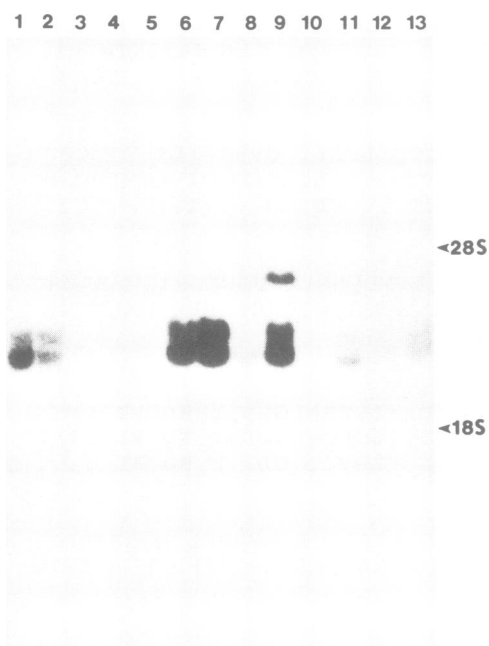


Fig. 6. Ezrin mRNA tissue distribution. Ten μg of total RNAs were separated on a 1% agarose-formaldehyde gel, transferred to nitrocellulose and hybridized with a ^{32}P -labeled single-stranded fragment of the ezrin clone as described in Materials and methods. RNA samples were from rat skin (lane 1), ovary (lane 2), heart (lane 3), skeletal muscle (lane 4), brain (lane 5), small intestine (lane 6), kidney (lane 7), spleen (lane 8), lung (lane 9), testes (lane 10), thymus (lane 11), cerebellum (lane 12) and submaxillary gland (lane 13). The stringency of the final wash was $0.1 \times \text{SSPE}$, 0.1% SDS at 50°C . 28S rRNA and 18S rRNA were used as molecular size markers. Exposure time was 3 days with pre-sensitized Kodak XAR film at -70°C with an intensifying screen.

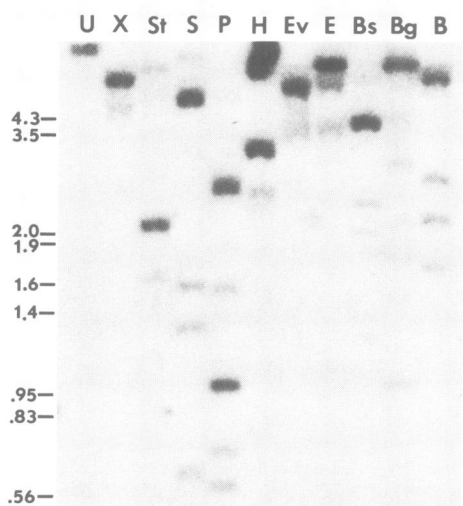


Fig. 7. Southern analysis of human ezrin genomic sequences. Southern analysis of A431 DNA was carried out as described in Materials and methods. Ten μg of A431 DNA was either left untreated (U) or digested with the following restriction enzymes: X = *Xba*I, St = *Stu*I, S = *Sph*I, P = *Pst*I, H = *Hind*III, Ev = *Eco*RV, E = *Eco*RI, Bs = *Bst*EII, Bg = *Bgl*II, B = *Bam*HI, resolved on a 1% agarose gel, and transferred to a Gene Screen Plus filter. The filter was hybridized to four ^{32}P -labeled single-stranded probes representing the entire cDNA fragment in 50% formamide, $5 \times \text{SSPE}$, $5 \times \text{Denhardt's}$, 1% SDS and $100 \mu\text{g}$ hydrolyzed yeast RNA/ml at 42°C . The stringency of the final wash was $0.1 \times \text{SSPE}$, 0.1% SDS at 65°C . The sizes of marker DNAs are shown on the left. The exposure time at -70°C with an intensifying screen was 3 days with pre-sensitized Kodak XAR film.

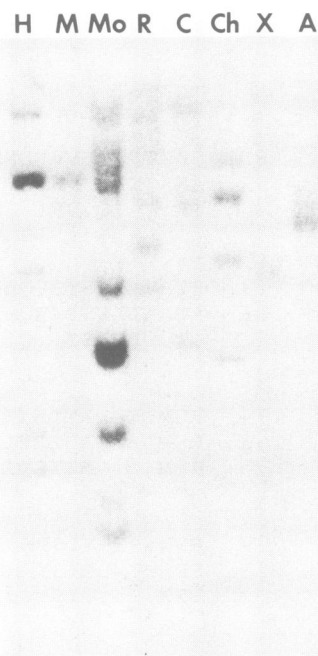


Fig. 8. Southern analysis of ezrin-related sequences in other organisms. Ten μg of DNA from the genomes of human (H), monkey (M), mouse (Mo), rat (R), Chinese hamster (C), chicken (Ch), frog (X) and *Amphioxus* (A), was digested with the restriction enzyme *Pst*I, resolved on a 1% agarose gel and transferred to a Gene Screen Plus filter. The filter was hybridized to a mixture of ^{32}P -labeled single-stranded segments spanning the entire cDNA fragment in 30% formamide, $5 \times \text{SSPE}$, $5 \times \text{Denhardt's}$, 1% SDS and $100 \mu\text{g}$ hydrolyzed yeast RNA/ml at 42°C . The stringency of the final wash was $0.1 \times \text{SSPE}$, 0.1% SDS at 52°C . The sizes of marker DNAs are shown on the left. The exposure time at -70°C with an intensifying screen was 1 day with presensitized Kodak XAR film.

digested with *Pst*I. Human ezrin-related sequences were detected in DNAs from monkey, mouse, rat, Chinese hamster, chicken, frog and *Amphioxus* (Figure 8). Bands hybridizing to the ezrin probe were also detected in DNA from sea squirt and *Drosophila melanogaster* (data not shown).

Discussion

We have isolated and sequenced a cDNA corresponding to the mRNA for human ezrin, a protein which we have previously referred to as p81 (Gould *et al.*, 1986). The single open reading frame of 585 amino acids predicted by the DNA sequence includes all of the directly sequenced human and related chicken tryptic peptides (Figure 1). Furthermore, the predicted N-terminal sequence matches the sequence derived from Edman degradation of the intact protein with the exception of the initiating methionine residue (P.Matsudaira, personal communication). It appears that this methionine residue is removed co- or post-translationally revealing an unblocked N-terminal proline residue.

The predicted molecular mass of ezrin is 69 000, significantly smaller than that expected based on its mobility in 12.5% SDS-polyacrylamide gels. This discrepancy appears to be caused by anomalous migration in such gels perhaps due to the highly charged character of the protein, since the *in vitro* translation product specified by the cDNA co-migrated in an SDS-polyacrylamide gel with the mature protein isolated from [^{35}S]methionine-labeled HeLa cells

(Figure 2). Consistent with the predicted molecular mass, the protein sediments in a sucrose gradient coincidentally with a 68 kd marker (Gould *et al.*, 1986). Ezrin mRNA was efficiently translated in a rabbit reticulocyte lysate, although the sequence surrounding the initiator codon [AAACCGAAAUGC] does not conform well to the consensus motif for initiator codons of higher eukaryotes [(GCC)GCCA/GCCAUGG] (Kozak, 1986). However, ezrin mRNA has an A at position -3, and this is the dominant requirement for a good initiation site (Kozak, 1986).

Several features of the predicted sequence are consistent with known properties of ezrin. Although it contains a high percentage of charged amino acids, the predicted pI of ezrin is 6.15. This agrees well with its previously reported pI on two-dimensional gels (Hunter and Cooper, 1981, 1983; Bretscher, 1983; Cooper *et al.*, 1983b; Gould *et al.*, 1986). There are no long stretches of hydrophobic amino acids that would provide a signal peptide for secretion or a membrane-spanning region, and indeed ezrin is not a secreted protein nor does it behave in biochemical fractionation experiments as an integral membrane protein (Bretscher, 1983; Gould *et al.*, 1986). Structural predictions based on the protein sequence suggest that 25–30% of the protein forms a long α -helix beginning at residue 290 near the middle of the sequence (Figure 3). If this structure indeed forms, it could account for the protein's sedimentation characteristics which suggest that ezrin has a somewhat elongated shape (Bretscher, 1983). This α -helical 150–180 amino acid stretch, which has a high content of charged amino acids, was found to be 15–25% similar to sequences in myosin, tropomyosin, troponins, keratin, desmin and lamins. However, this region of ezrin is not similar to all proteins containing long α -helices, and it is worth noting that the proposed α -helix is not predicted to be amphipathic.

More revealing about the possible function of ezrin is its significant homology to a domain of the erythrocyte cytoskeletal protein, band 4.1. Band 4.1 has a similar predicted molecular mass to that of ezrin (66 000) (Conboy *et al.*, 1986) and also migrates in polyacrylamide gels as if it is an ~80 kd protein (Leto and Marchesi, 1984). Band 4.1 binds to the cytoskeletal protein spectrin (Tyler *et al.*, 1979; Ungewickell *et al.*, 1979) and promotes the binding of spectrin to F-actin. Band 4.1 also binds to the transmembrane glycoproteins glycophorin with a high affinity (Anderson and Lovrien, 1984) in a polyphosphoinositide-dependent manner (Anderson and Marchesi, 1985), and the anion transport protein band 3 with a lower affinity (Pasternack *et al.*, 1985). These properties indicate that band 4.1 plays a key role in linking the actin network to transmembrane proteins and in regulating cell shape by modifying the association between the cortical cytoskeleton and the erythrocyte membrane.

Band 4.1 is organized into four domains defined by partial proteolytic mapping (Leto and Marchesi, 1984; Correas *et al.*, 1986). The N-terminal domain, which exhibits 37% identity to the N terminus of ezrin, is responsible for the polyphosphoinositide-dependent binding to glycophorin (Anderson and Lovrien, 1984; Leto *et al.*, 1986). This band 4.1 domain has a high content of hydrophobic residues (Conboy *et al.*, 1986) and the same is true for ezrin. This raises the possibility that ezrin also binds to one or more membrane proteins through its N terminus, perhaps even to a glycophorin-like molecule. Sequences downstream of the N-terminal domain in band 4.1 bear no similarity to those

in ezrin. However, secondary structure predictions indicate that both proteins have a long α -helical domain followed by a highly charged region at their C-termini (Conboy *et al.*, 1986; this report). In band 4.1, the latter portion of the α -helical domain is responsible for spectrin binding (Correas *et al.*, 1986). Ezrin has no sequence identity in this region to band 4.1, but if an overall structure–function organization is maintained, this region of ezrin might be expected to bind a second protein.

This new knowledge allows us to speculate upon the mechanism whereby ezrin is localized to microvillar core structures in intestinal epithelial cells and other cell types (Bretscher, 1983, 1989; Gould *et al.*, 1986). Purified ezrin does not bind with high affinity to F-actin, villin or fimbrin, which are the major components of the microvillar core (Bretscher, 1983). However, the similarity between ezrin and band 4.1 suggests that ezrin will bind to a cytoskeletal protein through its C-terminal domain. As discussed above, it is unlikely that ezrin binds to spectrin based on sequence considerations. In addition, ezrin is not generally co-distributed with spectrin in cells. For instance, ezrin is located in the microvillar cores of intestinal epithelial cells, whereas spectrin is found in the transverse terminal web underlying the microvillar cores (Bretscher, 1983). The similarity between band 4.1 and ezrin in the glycophorin-binding domain suggests that ezrin also binds to a membrane protein, and that this is the means by which ezrin is localized at the plasma membrane in cultured cells. A similar interaction might be predicted to occur in the microvillus, but a possible association with integral membrane proteins of the microvillar membrane has yet to be investigated.

There have been a number of reports of band 4.1-like molecules in non-erythroid cells (e.g. Granger and Lazarides, 1984; Baines and Bennett, 1985), as well as RNAs of size similar to that of erythroid band 4.1 mRNA detected by band 4.1 cDNA (Conboy *et al.*, 1986). In brain one of these proteins is synapsin I (Baines and Bennett, 1985). Since the size of ezrin mRNA (3.8 kb) is considerably smaller than that of band 4.1 mRNA (5.6 kb, Conboy *et al.*, 1986), and since the degree of similarity between ezrin and band 4.1 is not high enough to allow cross-hybridization between their coding sequences, it is unlikely that these non-erythroid band 4.1-like proteins correspond to ezrin.

One characterized protein that may be closely related to, or even identical to ezrin is cytovillin, a 75 kd protein, originally isolated from choriocarcinoma cells, which is localized to the cytoplasmic side of the plasma membrane in surface microvilli, and which is expressed in a wide variety of cell types (Suni *et al.*, 1984; Narvanen, 1985; Pakkanen *et al.*, 1987; Pakkanen, 1988). The size, isoelectric point, subcellular location and cell type distribution of cytovillin are all very similar to those of ezrin. Two immunologically distinct but related forms of ezrin have been purified from placenta; the smaller 77 kd form does not appear to be an *in vitro*-derived proteolytic fragment of the larger 81 kd form, and therefore may be encoded by a separate mRNA (Bretscher, 1989). If this is the case, the multiple species of ezrin mRNA we have detected in some cell lines and tissues might correspond to alternately spliced forms of mRNA, which could encode these two forms of ezrin. It should be noted that the cDNA sequence described here specifies the 81 kd form, since this is the only form expressed in HeLa cells and translation of mRNA derived from the ezrin cDNA clone gives rise to a protein of the

appropriate molecular size. Clearly the identity of cytovillin with both forms of ezrin needs to be tested.

The binding of ezrin to cytoskeletal or membrane components might be regulated by phosphorylation and dephosphorylation. Indeed, the phosphorylation of ezrin is coincident with its relocation and the rapid cell surface changes induced by EGF in A431 cells (Bretscher, 1989). Since ezrin contains 19 tyrosines, 24 threonines and 18 serines, there are a large number of possible phosphorylation sites. Several of these sites may be used, since our earlier studies showed that all three phosphoamino acids can be detected in ezrin (p81), and that ezrin gives rise to multiple phosphopeptides (Hunter and Cooper, 1981; Gould *et al.*, 1986; Bretscher, 1989). Based on the predicted protein sequence, we would speculate from previous observations that Thr298 is the major site of threonine phosphorylation (K.L.Gould and T.Hunter, unpublished data). Whether this phosphorylation event influences the structure or regulation of ezrin can now be tested by site-directed mutagenesis. Several protein-serine/threonine kinases, such as cAMP-dependent protein kinase (Kemp *et al.*, 1976) and the multifunctional calmodulin-dependent protein kinase (Pearson *et al.*, 1985) recognize serine or threonine residues with basic amino acids situated N-terminal to them as is the case for Thr 298. Although the sequence surrounding Thr 298 does not confirm precisely to any consensus phosphorylation site motif, there are three consensus phosphorylation sites for cAMP-dependent protein kinase at positions 65, 213 and 332, which might also be utilized.

The identification of phosphorylated tyrosine residues in ezrin is complicated by the observation that there are at least two major sites of tyrosine phosphorylation, one predominating after EGF treatment and the other predominating in cells lines transformed by certain protein-tyrosine kinase oncogenes (K.L.Gould and T.Hunter, unpublished data). A comparison of the sequences surrounding the 19 tyrosines in the protein to the identified sites of tyrosine phosphorylation in the EGF receptor (Downward *et al.*, 1984), calpactin I (Glenney and Tack, 1985), pp60^{c-src} (Cooper *et al.*, 1986), pp60^{v-src} (Patschinsky *et al.*, 1982) or enolase and lactate dehydrogenase (Cooper *et al.*, 1984) indicates that there are no similarities in flanking sequences. Thus, the identity of phosphorylated tyrosines in ezrin has yet to be established.

The pattern of ezrin mRNA expression determined by Northern analysis agrees well with our previous measurements of ezrin protein levels by immunoblotting (Gould *et al.*, 1986). Both the mRNA and protein are expressed at high levels in small intestine, kidney and lung. Intermediate levels of both mRNA and protein are found in almost all other tissues, but neither the mRNA nor the protein is detectable in liver or skeletal muscle. These data suggest that ezrin expression is regulated predominantly at the RNA level and in a tissue-specific manner.

Our analysis of human ezrin genomic sequences by Southern blotting revealed that there is most likely a single gene encoding ezrin. The gene appears to be conserved throughout evolution, as we were able to detect sequences related to the ezrin cDNA in genomes of many different organisms even as distantly related to humans as *Amphioxus*, sea squirt and *Drosophila*. Out of the 129 amino acids of chicken ezrin sequence obtained by Edman degradation of

tryptic peptides 91% are identical to human ezrin, which is another indication of its high degree of conservation. These data establishing conservation in sequence suggest that ezrin function may also have been conserved throughout evolution.

The availability of the ezrin sequence and the cDNA will allow genetic manipulations that should enable us to define the function of ezrin in cytoskeletal and membrane organization and also to identify the phosphorylation sites in ezrin and test their function. An obvious target for genetic manipulation is the region of similarity between band 4.1 and ezrin. If this region is involved in the association of ezrin with the membrane, deletion mutations in this sequence should preclude membrane localization. Finally, the relationship of ezrin to band 4.1 suggests that it will be worth determining whether ezrin interacts with glycophorin, and whether this might be regulated by polyphosphoinositides.

Materials and methods

Protein purification and amino acid sequencing

Ezrin was purified as described elsewhere from chicken intestine (Bretscher, 1983) and human placenta (Bretscher, 1986, 1989). The purified preparations (~45 µg from chicken intestine and ~190 µg of the 81 kd form from human placenta) were made 1 ml in volume with 50 mM ammonium bicarbonate, 0.05% β-mercaptoethanol (buffer 1), dialyzed extensively against buffer 1, and digested in a final volume of 2 ml for 4 h at 30°C with a 1:50 ratio (w/w) of trypsin (TPCK-treated, Worthington, Freehold, NJ) to ezrin. After tryptic digestion, 100% trifluoroacetic acid (TFA) was added to a final concentration of 0.1% and the samples were subjected to reverse phase high pressure liquid chromatography (HPLC) on a microbond C₁₈ column (30 cm × 3.9 mm, Aquapore, Brownlee Labs, Santa Clara, CA). Peptides were eluted from the column with a linear 0–50% gradient of acetonitrile in 0.1% TFA at a flow rate of 0.5 ml/min and collected manually. HPLC fractions containing apparently only one tryptic peptide (seven from chicken ezrin and 11 from human ezrin) were subjected to automated Edman degradation in an Applied Biosystems Inc. model 470 gas phase protein sequencer with on-line phenylthiohydantoin-amino acid analyzer from the same manufacturer.

Oligonucleotide probes

Based on the amino acid sequence of tryptic peptides from human ezrin, two regions were chosen for oligonucleotide probes on the basis of minimum redundancy in codon usage. Probes 1 and 2 contained sequences complementary to all the possible combinations of codons for the amino acids EMYGINYF and EVWYFG and were mixtures of 196 23 base long and 64 17 base long oligonucleotides respectively (the third base of the codons for the last amino acids were omitted).

cDNA cloning and nucleotide sequence determination

A HeLa cell cDNA library made in an Okayama and Berg vector (Hanks, 1987) was screened initially by colony hybridization on duplicate nitrocellulose filters with probe 1 labeled with ³²P by polynucleotide kinase (Maniatis *et al.*, 1982). Hybridization was carried out at 37°C for 15 h in 5 × SSPE (1 × SSPE = 0.15 M NaCl, 10 mM NaH₂PO₄·H₂O, 1 mM Na₂EDTA·2H₂O), 5 × Denhardt's (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% BSA), 0.05% Na₄P₂O₇·H₂O, 100 µg hydrolyzed yeast RNA/ml (Jay *et al.*, 1974). After hybridization, the filters were washed in 6 × SSC at 42°C and subjected to autoradiography. They were then washed in 3 M (CH₃)₄NCl, 2 mM EDTA, 0.1% SDS, 50 mM Tris-HCl, pH 8.0 according to Wood *et al.* (1985), at 58°C. From 60 000 colonies, two identical partial cDNA clones were obtained (clones E3 and E4). The DNA sequence of clone E3 encoded several ezrin tryptic peptides indicating that it was a partial ezrin clone. A 520 base long M13 single-stranded probe derived from clone E3 (nucleotides 440–960 of p81 clone F6) was labeled with ³²P as described below and used to screen the library a second time. The longest cDNA clone which hybridized to both the partial cDNA and probe 2 was analyzed by restriction mapping. Overlapping *Pst*I, *Bam*HI, *Sac*I and *Sau*3A restriction fragments were subcloned in M13mp18–mp19 and their sequences were determined by the dideoxy chain termination method (Sanger *et al.*, 1977) using Sequenase (U.S. Bichemical, Cleveland, OH). Five additional 18 base long oligonucleotide sequencing primers were needed to complete the entire sequence of both strands.

Computer analysis of the amino acid sequences

Both DNA and deduced protein sequences were analyzed by using the sequence analysis software packages of the University of Wisconsin Genetics Computer Group (UWGCG, Devereux *et al.*, 1984) and Intelligenetics. Secondary structure and hydropathicity were predicted by PEPSTRUCTURE (UWGCG), using the algorithms of Chou and Fasman (1978) and Kyte and Doolittle (1982), respectively, and represented graphically by PEPLOT (UWGCG). The amino acid sequence was searched for internal repeats by the method of Maizel and Lenk (1981) using COMPARE and DOT-PLOT of the UWGCG program. The PIR (National Biomedical Research Foundation) and SWISSPROT protein databases and the GenBank and EMBL nucleic acids databases were searched with the predicted amino acid sequence and the nucleotide sequence by the programs IFIND (Intelligenetics), WORDSEARCH (UWGCG), and FASTP based on the algorithms of Wilbur and Lipman (1985) and Pearson and Lipman (1988).

Expression of the cDNA-encoded protein

The predicted protein coding region of ezrin clone F6 was excised from the Okayama and Berg vector with the restriction enzyme *Tth*1111. The resultant 2802 bp long ezrin cDNA fragment was blunt ended with the Klenow fragment of DNA polymerase and ligated to *Sma*I-cut pGEM-4Z (Promega) according to standard procedures (Maniatis *et al.*, 1982). Capped RNA was made from ~0.5 µg of plasmid DNA by SP6 polymerase as per the manufacturer's instruction (Promega, Madison, WI). *In vitro* translations were performed in a mRNA-dependent reticulocyte lysate system as reported previously (Pelham and Jackson, 1975; Sefton *et al.*, 1978). [³⁵S]Methionine (>1000 Ci/mmol, Amersham, Arlington Heights, IL) was present at ~1 mCi/ml in a final reaction volume of 11 µl. HeLa cells were labeled with [³⁵S]methionine and ezrin was immunoprecipitated from RIPA lysates as described previously (Gould *et al.*, 1986).

Northern and Southern blot analysis

Cytoplasmic RNAs from the murine fibroblast line, ANN-1, and the pituitary tumor line, AtT20, were prepared as described previously (Gould *et al.*, 1984) and subjected to oligo(dT)-cellulose chromatography. Other RNAs were isolated from cultured human cells and from rat tissues according to Chomczynski and Sacchi (1987), and were generous gifts of S.K.Hanks, R.A.Lindberg and J.Meisenhelder (all at the Salk Institute, San Diego, CA). For Northern hybridizations, RNA (10 µg/well total or 2 µg/well poly(A)-selected) was size-fractionated by electrophoresis through 1% agarose-formaldehyde gels, transferred to nitrocellulose or Gene Screen Plus (NEN/Dupont, Boston, MA) membranes, and hybridized with a ³²P-labeled probe. The ezrin probe was synthesized from a M13 single-stranded template (Jeffreys *et al.*, 1985) and contained ~1000 nucleotides corresponding to the 5' 960 nucleotides of p81 clone F6 and ~40 nucleotides of G-C tail and vector sequences. Hybridizations were carried out for 24 h at 42°C in 50% formamide, 5 × SSPE, 5 × Denhardt's, 2% SDS, and 100 µg hydrolyzed yeast RNA/ml. Following hybridization the filters were washed in 0.1 × SSPE, 0.1% SDS at 50°C.

A431 cell DNA was prepared by standard procedures and was generously provided by J.Meisenhelder. All other DNAs were kind gifts from S.Sukumar (monkey and rat), R.A. Lindberg (mouse), S.Simon (Chinese hamster), P.Bull (chicken), M.Kindy (frog), M.Broome (Amphioxus) (all at the Salk Institute). For genomic Southern analysis, 10 µg of the DNA preparations were digested, fractionated on 1% agarose gels, and transferred to Gene Screen Plus membranes. Hybridizations were performed as described above for Northern blots except that four single-stranded probes covering the entire cDNA were used. Filters were washed in 0.1 × SSPE, 0.1% SDS up to a final stringency of 65°C.

Acknowledgements

We are grateful to Steve Hanks for providing HeLa cell cDNA library, RNA samples and much valuable advice on the project. We also thank the following individuals for their generous gifts of RNA, DNA and/or useful suggestions and protocols: Jill Meisenhelder, Rick Lindberg, Detlev Jaehner, Jon Pines, Gerry Weinmaster, Clare Isacke, Suzanne Simon, Sara Sukumar, Martin Broome and Mark Kindy. We gratefully acknowledge Lisa Caballero for help with sequence analysis computer programs, and are indebted to Lois Tack for aligning the internal repeats in ezrin, Paul Matsudaira for communicating unpublished protein sequence information, Jim Woodgett for help with HPLC purification of tryptic peptides, and Jasper Rees for re-searching the protein databases for us. This work was supported by US Public Health Service Grant (CA-39785) and by a grant from the American Business Foundation for Cancer Research to T.H.

References

- Anderson.R.A. and Lovrien.R.E. (1984) *Nature*, **307**, 655–658.
 Anderson.R.A. and Marchesi.V.T. (1985) *Nature*, **318**, 295–298.
 Baines.A.J. and Bennett.V. (1985) *Nature*, **315**, 410–413.
 Bretscher.A. (1983) *J. Cell Biol.*, **97**, 425–432.
 Bretscher.A. (1986) *Methods Enzymol.*, **134**, 24–37.
 Bretscher.A. (1989) *J. Cell Biol.*, **108**, 921–930.
 Chomczynski.P. and Sacchi.N. (1987) *Anal. Biochem.*, **162**, 156–159.
 Chou.P.Y. and Fasman.G.D. (1978) *Annu. Rev. Biochem.*, **47**, 251–276.
 Conboy.J., Kan.Y.W., Shohet.S.B. and Mohandas.N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9512–9516.
 Cooper.J.A., Reiss.N.A., Schwartz.R.J. and Hunter.T. (1983a) *Nature*, **302**, 218–223.
 Cooper.J.A., Scolnick.E.M., Ozanne.B. and Hunter.T. (1983b) *J. Virol.*, **48**, 752–764.
 Coper.J.A., Esch.F.S., Taylor.S.S. and Hunter.T. (1984) *J. Biol. Chem.*, **259**, 7835–7841.
 Cooper.J.A., Gould.K.L., Cartwright.C.A. and Hunter.T. (1986) *Science*, **231**, 1431–1433.
 Courtneidge.S.A. and Heber.A. (1987) *Cell*, **50**, 1031–1038.
 Correas.I., Leto.R.L., Speicher.D.W. and Marchesi.V.T. (1986) *J. Biol. Chem.*, **261**, 3310–3315.
 Devereux.J., Haeberli.P. and Smithies.O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
 Downward.J., Parker.P. and Waterfield.M.D. (1984) *Nature*, **311**, 483–485.
 Draceta.G., Piwnica-Worms.H., Morrison.D., Druker.B., Roberts.T. and Beach.D. (1988) *Nature*, **336**, 738–744.
 Fava.R.A. and Cohen.S. (1984) *J. Biol. Chem.*, **259**, 2636–2645.
 Glenney.J.R.Jr. and Tack.B.F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7884–7888.
 Gould.K.L., Cooper.J.A. and Hunter.T. (1984) *J. Cell Biol.*, **98**, 487–498.
 Gould.K.L., Cooper.J.A., Bretscher.A. and Hunter.T. (1986) *J. Cell Biol.*, **102**, 660–669.
 Granger.B.L. and Lazarides.E. (1984) *Cell*, **37**, 595–607.
 Hanks.S.K. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 388–392.
 Hirst.R., Horwitz.A., Buck.C. and Rohrschneider.L.R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6470–6474.
 Hunter.T. and Cooper.J.A. (1981) *Cell*, **24**, 741–752.
 Hunter.T. and Cooper.J.A. (1983) *Proc. Nucleic Acid Res. Mol. Biol.*, **29**, 221–233.
 Hunter.T. and Cooper.J.A. (1985) *Annu. Rev. Biochem.*, **54**, 897–930.
 Hunter.T. and Cooper.J.A. (1986) *The Enzymes*, **17**, 191–246.
 Jay.E., Bambara.R., Padmanabhan.R. and Wu.R. (1974) *Nucleic Acids Res.*, **1**, 331–353.
 Jeffreys.A.J., Wilson.V. and Thein.S.L. (1985) *Nature*, **314**, 67–73.
 Kaplan.D.A., Whitman.M., Schaffhausen.B., Pallas.D.C., White.M., Cantley.L. and Roberts.T.M. (1987) *Cell*, **50**, 1021–1029.
 Kemp.B.E., Benjamini.E. and Krebs.E.G. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 1038–1042.
 Kozak.M. (1986) *Nucleic Acids Res.*, **15**, 8125–8130.
 Kyte.J. and Doolittle.R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
 Leto.R.L. and Marchesi.V.T. (1984) *J. Biol. Chem.*, **259**, 4603–4608.
 Leto.R.L., Correas.I., Tobe.R., Anderson.R.A., Horne.W.C. (1986) In Bennett.V., Cohen.C.M., Lux.S. and Palek.J. (eds), *Membrane Skeletons and Cytoskeletal Membrane Associations*. Liss, NY, pp. 137–147.
 Maizel.J.V. and Lenk.R.P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7665–7669.
 Maniatis.T., Fritsch.E.F. and Sambrook.J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 Morrison.D.K., Kaplan.D.R., Rapp.U. and Roberts.T.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8855–8859.
 Narvanen.A. (1985) *Biochem. J.*, **231**, 53–62.
 Pakkanen.R. (1988) *J. Cell. Biochem.*, **38**, 65–75.
 Pakkanen.R., Hedman.K., Turunen.O., Wahlstrom.T. and Vaheri.A. (1987) *J. Histochem. Cytochem.*, **35**, 809–817.
 Pasquale.E., Maher.P.A. and Singer.S.J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5507–5510.
 Pasternak.G.R., Anderson.R.A., Leto.T.L. and Marchesi.V.T. (1985) *J. Biol. Chem.*, **260**, 3676–3683.
 Patschinsky.T., Hunter.T., Esch.F.S., Cooper.J.A. and Sefton.B.M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 973–977.
 Pearson.W.R. and Lipman.D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.

- Pearson,R., Woodgett,J.R., Cohen,P. and Kemp,B.E. (1985) *J. Biol. Chem.*, **260**, 14471–14476.
- Pelham,H.R.B. and Jackson,R.J. (1975) *Eur. J. Biochem.*, **67**, 247–256.
- Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
- Radke,K. and Martin,G.S. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5212–5216.
- Ray,L.B. and Sturgill,T.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3753–3759.
- Samelson,L.E., Patel,M.D., Weissman,A.M., Harford,J.B. and Klausner,R.D. (1986) *Cell*, **46**, 1083–1090.
- Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Sefton,B.M., Beemon,K. and Hunter,T. (1978) *J. Virol.*, **28**, 957–971.
- Sefton,B.M., Hunter,T., Ball,E.H. and Singer,S.J. (1981) *Cell*, **24**, 821–830.
- Suni,J., Narvanen,A., Wahlstrom,T., Aho,M., Pakkanen,R., Vaheri,A., Copeland,T., Cohen,M. and Oroszlan,S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6197–6201.
- Tyler,J.M., Hargeaves,W.R. and Branton,D. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5192–5196.
- Ungewickell,E., Bennett,P.M., Calvert,R., Ohanian,V. and Gratzer,W.B. (1979) *Nature*, **280**, 811–814.
- Wilbur,W.J. and Lipman,D.J. (1985) *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.
- Wood,W.I., Gitschier,J., Lasky,L.A. and Lawn,R.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1585–1588.

Received on July 7, 1989

Note added in proof

The recently reported cDNA sequence for cytovillin (Turunen,O., Winqvist,R., Pakkanen,R., Grzeschik,K.-H., Wahlstrom,T. and Vaheri,A. (1989) *J. Biol. Chem.*, **264**, 16727–16732) proves that it is identical to ezrin. This has been confirmed by a direct comparison of the two proteins (Pakkanen,R. and Vaheri,A. (1989) *J. Cell. Biochem.*, **41**, 1–12).