

## Structural diversity of band 4.1 superfamily members

Kosei Takeuchi<sup>1,2</sup>, Akiyoshi Kawashima<sup>1,3</sup>, Akira Nagafuchi<sup>1</sup> and Shoichiro Tsukita<sup>1,2,\*</sup>

<sup>1</sup>Laboratory of Cell Biology, Department of Information Physiology, National Institute for Physiological Sciences, Myodaiji-cho, Okazaki, Aichi 444, Japan

<sup>2</sup>Department of Physiological Sciences, School of Life Sciences, The Graduate University of Advanced Studies, Myodaiji-cho, Okazaki, Aichi 444, Japan

<sup>3</sup>Department of Geriatric Medicine, Faculty of Medicine, Kyoto University, Yoshida-Konoe-cho, Kyoto 606-01, Japan

\*Author for correspondence at address 1

### SUMMARY

Several proteins contain the domain homologous to the N-terminal half of band 4.1 protein, indicating the existence of a superfamily. The members of this 'band 4.1' superfamily are thought to play crucial roles in the regulation of cytoskeleton-plasma membrane interaction just beneath plasma membranes. We examined the structural diversity of this superfamily by means of the polymerase chain reaction using synthesized mixed primers. We thus identified many members of the band 4.1 superfamily that were expressed in mouse teratocarcinoma F9 cells and mouse brain tissue. In total, 15 cDNA clones were obtained; 8 were identical to the corresponding parts of cDNAs for the known members, while 7 appeared to encode novel proteins (NBL1-7: novel band 4.1-like proteins). Sequence analyses of these clones revealed that the band 4.1 superfamily can

be subdivided into 5 gene families; band 4.1 protein, ERM (ezrin/radixin/moesin/merlin/NBL6/NBL7), talin, PTPH1 (PTPH1/PTPMEG/NBL1-3), and NBL4 (NBL4/NBL5) families. The NBL4 family was first identified here, and the full-length cDNA encoding NBL4 was cloned. The deduced amino acid sequence revealed a myristoylation site, as well as phosphorylation sites for A-kinase and tyrosine kinases in its N-terminal half, suggesting its involvement in the phosphorylation-dependent regulation of cellular events just beneath the plasma membrane. In this study, we describe the initial characterization of these new members and discuss the evolution of the band 4.1 superfamily.

Key words: ERM family, merlin, band 4.1 superfamily, PCR

### INTRODUCTION

The band 4.1 superfamily includes several distinct proteins that contain the conserved domain homologous to the N-terminal half of band 4.1 protein (reviewed by Tsukita et al., 1993). Band 4.1 protein is a major constituent of the undercoat of erythrocyte membranes (Bennett, 1989). This protein forms a complex with spectrin molecules and short actin filaments (Fowler and Taylor, 1980; Ohanian et al., 1984; Ungewickell et al., 1979) and binds to an integral membrane protein called glycophorin C (glycoconnectin) (Anderson and Lovrien, 1984; Anderson and Marchesi, 1985; Bennett, 1989; Leto et al., 1986; Mueller and Morrison, 1981). The N-terminal half of band 4.1 protein is responsible for glycophorin C-binding (Leto et al., 1986).

The first protein to be identified with homology to band 4.1 protein was ezrin. Ezrin (also called cytovillin or p81) was originally identified as a protein localizing just beneath the plasma membrane of intestinal microvilli (Bretscher, 1983; Pakkanen et al., 1987), and is a good substrate for some tyrosine kinases *in vivo* (Bretscher, 1989; Gould et al., 1986; Hunter and Cooper, 1981, 1983). Sequence analysis revealed that the N-terminal half of ezrin has homology with that of band 4.1 protein (~30% identity)(Conboy et al., 1986; Gould

et al., 1989; Turunen et al., 1989). Recently, radixin was identified as a barbed-end capping actin-modulating protein from isolated cell-to-cell adherens junctions from the liver (Tsukita et al., 1989a). Its cDNA sequencing revealed that radixin is highly homologous to ezrin (~75% identity overall; ~85% identity for the N-terminal half)(Funayama et al., 1991). Furthermore, moesin is also very similar to ezrin (~75% identity overall; ~85% identity for the N-terminal half) (Lankes and Farthmayr, 1991; Sato et al., 1992). Moesin was originally identified as an extracellular heparin-binding protein (Lankes et al., 1988) and was later found to be an intracellular protein (Lankes and Farthmayr, 1991). This high homology between these proteins indicates a gene family (ERM family) included in the band 4.1 superfamily (Sato et al., 1992). Recent detailed immunolocalization of these ERM family members revealed that they are highly concentrated at the specialized domain of plasma membranes where actin filaments are densely associated with plasma membranes, suggesting that these ERM family members are directly involved in the actin filament/plasma membrane association (Sato et al., 1992; Tsukita et al., 1992; Berryman et al., 1993; Franck et al., 1993). The tumor suppressor gene of neurofibromatosis 2 has been identified (Trofatter et al., 1993; Rouleau et al., 1993). Surprisingly, its gene product was similar in amino acid sequence

to those of the ERM family (~49% identity overall; ~62% for the N-terminal half). This product was named merlin (moesin-ezrin-radixin-like protein).

In addition to band 4.1 protein and ERM family, talin also belongs to the band 4.1 superfamily (Rees et al., 1990). Talin is a protein with a molecular mass of ~200 kDa, which is concentrated at the undercoat of cell-to-substrate adherens junctions (focal contacts) (Burrige and Connell, 1983; Burrige and Mangeat, 1984). This protein is thought to play an important role in linking integrins to actin-based cytoskeletons. Recent cDNA cloning and sequence analysis has revealed that there is a domain homologous to the N-terminal half of band 4.1 protein in the N-terminal end of talin (Rees et al., 1990). Furthermore, two distinct protein tyrosine phosphatases called PTPH1 and PTPMEG are reportedly members of band 4.1 superfamily, containing the band 4.1-like domain and the phosphatase domain at their N- and C-terminal regions, respectively (Gu et al., 1991; Yang and Tonks, 1991).

Considering that the N-terminal half of band 4.1 protein is responsible for its binding to glycoporphin C and that this domain is homologous to the N-terminal regions of the band 4.1 superfamily, all the members should be located just beneath the membrane, through the specific binding of their N-terminal ends with integral membrane proteins. Furthermore, the fact that this superfamily includes protein tyrosine phosphatases and a tumor suppressor such as merlin as well as proteins responsible for the cytoskeleton/membrane interaction, suggests that the band 4.1 superfamily members are directly involved in the regulation of cell growth by affecting cytoskeleton/membrane interactions.

For a better understanding of the physiological functions of the band 4.1 superfamily, it is necessary to know how many and what types of proteins are included. Therefore, to identify as many mouse band 4.1 superfamily members as possible, we used the polymerase chain reaction (PCR) and found several new members. In this study, we initially characterize these new members and discuss the evolution of the band 4.1 gene superfamily.

## MATERIALS AND METHODS

### Polymerase chain reaction (PCR)

Oligonucleotide primers were synthesized based on 6 conserved amino acid sequences described in detail in Fig. 1. The degeneracy of primers A,B,C,D,E and F were 2<sup>7</sup>-, 2<sup>8</sup>-, 2<sup>9</sup>-, 2<sup>10</sup>-, and 2<sup>9</sup>-fold, respectively. Phage DNA of a mouse embryonic teratocarcinoma F9 cell cDNA library (Nagafuchi et al., 1987), isolated by plate lysis (Sambrook et al., 1989), was used as the template. Alternatively, mRNA was extracted from F9 cells or mouse brain tissue by means of a guanidium isothiocyanate, and it was applied to reverse transcriptase-PCR (RT-PCR). The PCR conditions were essentially the same as those described by Saiki (1989). The PCR cycle, repeated 30 times, consisted of denaturation at 94°C for 1 minute, annealing at 47°C for 1 minute, and extension at 72°C for 2 minutes. The PCR products were analyzed on Nusieve GTG agarose (FMC Bioproducts, CO., Rockland, ME) and DNA fragments of appropriate sizes were excised, eluted from the gel, subcloned into the *EcoRV* site of pBlue-script SK(-) plasmid (Stratagene), and sequenced.

### Screening of the F9 cell cDNA library and DNA sequencing

Two  $\lambda$ gt11 libraries made of mouse F9 poly(A) RNAs were prepared

using either a random mixture of hexanucleotides or oligo (dT) as the primer for the first-strand synthesis (Nagafuchi et al., 1987). The PCR product of NBL4 was labeled with DIG according to the procedure developed by Boehringer Mannheim Biochemicals (Indianapolis, IN). Using this DIG-labeled fragment, ~5×10<sup>6</sup> plaques from the cDNA library was screened at high stringency, and one positive phage recombinant ( $\alpha$ 1) was cloned (see Fig. 3). Using two distinct fragments (Probes I and II in Fig. 3) of  $\alpha$ 1 as probes, two more positive phage recombinants ( $\beta$ 5,  $\gamma$ 3) were cloned. All clones were subcloned into pBluescript SK(-) and sequenced with the 7-deaza Sequenase Version 2.0 kit (US Biochemical Corp., Cleveland, OH), or with the Taq DyeDeoxy Terminator Cycle Sequencing Kit on an Applied Biosystems 373A Sequencer (Applied Biosystems, Inc. Foster City, CA).

### Northern blotting

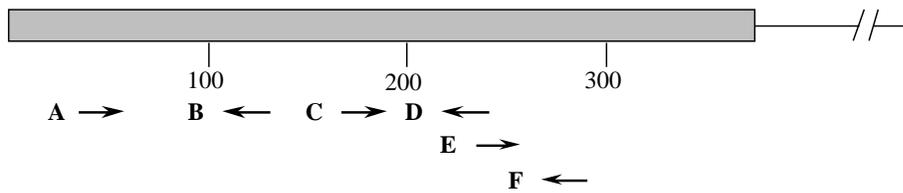
Total RNA was extracted from various mouse tissues and poly(A)<sup>+</sup> RNA was purified using oligo-dT-conjugated latex beads (Oligotex-dT30; Takara Shuzo Co., Ltd, Kyoto, Japan). Samples (10  $\mu$ g) of poly(A)<sup>+</sup> RNA from each tissue applied to a formaldehyde/agarose gel, were separated electrophoretically and transferred to a nitrocellulose membrane. An RNA ladder (Bethesda research laboratories, Bethesda, MD) was used as a size marker. The probes were labeled with [ $\alpha$ -<sup>32</sup>P]dCTP using the Random Primer Labeling Kit (Takara Shuzo Co., Ltd, Kyoto, Japan). Hybridization proceeded under conditions of high stringency (50% formamide/5× Denhardt's solution without BSA/5× SSC/0.5% SDS/100  $\mu$ g/ml denatured salmon sperm DNA).

## RESULTS

### Isolation of cDNAs encoding members of the band 4.1 superfamily

The cDNAs that encode members of band 4.1 superfamily were isolated by PCR. We compared the amino acid sequences of the conserved N-terminal domains of all the band 4.1 superfamily members so far identified. We selected six highly conserved amino acid sequences consisting of 6-8 amino acids, and synthesized the corresponding degenerate oligonucleotides (Fig. 1). PCR and reverse transcriptase-PCR (RT-PCR) were performed with the mixed oligonucleotides as primers in every pair of the six conserved sequences. A mouse teratocarcinoma cell line (F9) cDNA and mouse brain mRNA were used as templates, since band 4.1 superfamily should be abundant in these sources. Products of a size appropriate for the combination of primers were isolated and subcloned into the Bluescript SK(-) vector. In total, 217 clones with appropriate inserts were isolated (Table 1). Sequence analysis of these clones revealed that 169 corresponded to ezrin, radixin, moesin, band 4.1 protein, PTPH1, PTPMEG, talin, or merlin, while the other 48 appeared to encode 7 new members of the band 4.1 superfamily. They were designated NBL1-7 (novel band 4.1-like proteins). Judging from their deduced amino acid sequences, these new proteins were subclassified into three types: Three (NBL1-3) and two (NBL6,7) of them were highly similar to the corresponding sequences of PTPH1/PTPMEG and ezrin/radixin/moesin/merlin, respectively (Fig. 2A,C). The other two (NBL4,5) showed significant homology to the band 4.1 superfamily, but were less closely related to any of the currently known members than are NBL1,2,3,6 and 7 (Fig. 2B). We thus isolated the full-length cDNA encoding NBL4 from a  $\lambda$ gt11 library made from mouse F9 cells.

## Band 4.1 homology domain (370 a.a.)



<b>A:</b> EK(E/D/W)YFGL	38-44aa	<b>D:</b> (K/N)ISF(K/N)R	232-237aa
<b>B:</b> LAS(Y/F)AVQ	119-125aa	<b>E:</b> RK(K/Q/S)F(F/V)I	237-242aa
<b>C:</b> AR(T/D)L(D/F)(F/L/M)YG	181-188aa	<b>F:</b> CVEHHTFF or CMGNHELY	271-277aa

**Fig. 1.** Highly-conserved amino acid sequences among the known band 4.1 superfamily members. For the PCR analysis, 6 sequences (A-F) were chosen in the conserved domain (~370 a.a.) among band 4.1 superfamily members. Corresponding degenerate oligonucleotides were each synthesized in one direction (indicated by arrows).

**Table 1.** Number of independent cDNA subclones

No. of clones sequenced	No. with identical DNA sequence														
	Band4.1	Ezrin	Radixin	Moesin	Merlin	Talin	PTPH1	PTPMEG	NBL 1	NBL 2	NBL 3	NBL 4	NBL 5	NBL 6	NBL 7
20	---	2	4	3	---	---	4	2	2	---	1	2	---	---	---
112	5	17	12	21	---	---	16	15	6	4	2	10	4	---	---
20	---	4	5	6	---	---	---	2	---	---	---	3	---	---	---
41	3	8	6	8	1	---	2	2	3	---	2	---	---	5	2
9	---	2	---	2	---	---	3	1	1	---	---	---	---	---	---
14	2	2	1	1	---	3	4	---	---	---	1	---	---	---	---

### Isolation and sequencing of cDNA encoding NBL4

To isolate the rest of the NBL4 gene, we screened  $\sim 5 \times 10^6$  plaques from a  $\lambda$ gt11 cDNA library made from mouse F9 cells by DNA hybridization with the DIG-labeled PCR product at high stringency, and cloned one positive phage recombinant ( $\alpha 1$ ). Using two DIG-labeled fragments from  $\alpha 1$  (Probes I and II in Fig. 3) as probes, two more positive phage recombinants ( $\beta 5$ ,  $\gamma 3$ ) were cloned. Fig. 3 shows the overall relationship of the PCR products,  $\alpha 1$ ,  $\beta 5$ , and  $\gamma 3$ , and the restriction sites relevant for subcloning and sequencing. The complete nucleotide sequence and deduced amino acid sequence of the cloned molecule are shown in Fig. 4. The composite cDNA is 2,503 nucleotides long, consisting of 51 bp of the 5'-untranslated region, an ORF of 1,662 bp encoding 554 amino acids, and a 790 bp 3'-untranslated region with a poly(A)<sup>+</sup> tail. The calculated molecular mass of the protein encoded by this cloned cDNA is 61.0 kDa.

This cDNA had no long hydrophobic stretches that could

encode a signal peptide or a transmembrane domain, indicating that NBL4 is neither a secretory nor a transmembrane protein. At the amino acid residues 2-7, there is a typical myristoylation consensus (Towler and Gordon, 1988). Furthermore, there are several consensus sequences for A-kinase and tyrosine kinase phosphorylation sites (Fig. 4).

The N-terminal half (amino acid residue 1-327) shows a similarity to the N-terminal region of band 4.1 protein (39.9% identity), radixin (29.2% identity), merlin (29.7% identity), PTPH1 (36.9% identity), and talin (26.6% identity) (Fig. 5). Considering that another PCR product, NBL5, shows 95% identity to NBL4, this molecule together with NBL4, appears to form a new family in the band 4.1 superfamily. A search of the data bases identified no proteins with significant homology to the C-terminal half of NBL4.

The mRNA encoding NBL4 was detected by northern blotting with poly(A)<sup>+</sup> RNA from various types of mouse tissues. As shown in Fig. 6, a 2.5 kb band was detected in the

## (A)

```

* * *
Band4.1 VDLHKAKDLEGVDIILGVCSSGLLVYKDKLRINRF-PWPKVL-KISYKRSSFFIKIRPGEQ-EQYESTIGFKLPSYRAAKLWKV
PTPH1 VELHGGRDLDNLDLMIGIASAGIAVCRKYICTSFY-PWVNIL-KISFKRKKFFIHQRQKQQAESREHIVAFNMLNYSCKNLWKS
PTPMEG VEFHYARDQSNNEILLIGVMSGGILLIYKNR-VRMNTFLWLKIV-KISFKCKQFFIQLR-KELHESRETLLGFNMVNYRACKTLWKA
NBL1 VELHKGRDLDQLHLMIGVASGGILLVCRKR-VRTNRFWPVNILVKISFKRKKFFIQQRQKEAEE-RESTVGFNMLNYSCKNLWKS
NBL2 VELHKARDLDQLHGENKSAYGGILTPRKRVRVTRNKKPWGKIF-WPRIKRVHFFITQFEKEVLEK-ESTETFNFLNARASCTLWKS
NBL3 VEFHKARDLSQNEIENGSMGSLGLTPRKRVRMNNKPKLIK-WKISFKRKKFITQRQKEVLSR-ESTETFNFLNYSCKKWKKS

```

## (B)

```

*****
NBL4 VDLHPVYGENKSEYFLGLTPSGVVVYKNNKQVGKYFWPRITKVIHFKETQFELRVLGKDCNETSFFFEARSKTACKHLWKC
NBL5 VDLHPVYGENKSEYFLGVTSPGVVVYKRRKQVEKYFWPRITKVIHFKETQFELRVLGKDCNETSFFFEARSKTACKHLWKC

```

## (C)

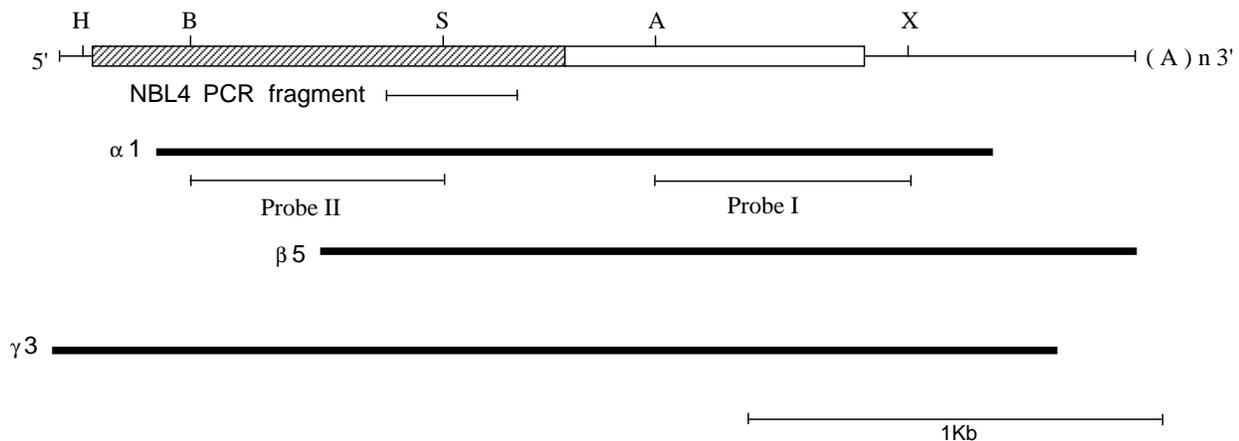
```

**** * ***** ** *
Radixin VNYFEIKNKKGTELWLGVDALGLNIYEHDDKLTTPKIGFPWSEIR
Merlin VNYFAIRNKKGTELELLGVDALGLHIYDPENRLTPKISFPWNEIR
NBL6 VNYFAIRNKKGTELWLGVDALGLNIYDQKDRLTPKISFPWNEIR
NBL7 VNYFAIRNKKGTELWLGVDDEMGLHIYDPENRLTPKISFPWNEIR

```

**Fig. 2.** Alignment of PTPH1 family member sequences (A), NBL4 family member sequences (B), and ERM family member sequences (C). (A) A comparison of the deduced amino acid sequence of the band 4.1 PCR product, with that of the PTPH1 family members. Conserved amino acids shared by more than 5 proteins are shadowed. The positions where all sequences have an

identical amino acid are indicated by asterisks. Gaps, introduced to maximize alignment, are indicated by dashes. (B,C) Amino acid sequences of PCR products of NBL4 family members (B) and ERM family members (C). Identities are indicated by asterisks. These are not PCR artifacts, since at the nucleotide sequence level the identity for NBL4/NBL5, merlin/NBL6, NBL6/NBL7 are 72%, 86.4% and 86.4%, respectively.



**Fig. 3.** Restriction map and cDNA fragments of NBL4. The box indicates the coding region, and the band 4.1 homology region is shadowed.  $\alpha 1$  was first obtained from a cDNA library made from mouse F9 cells using the PCR product (NBL4 PCR fragment) as a probe.  $\beta 5$  and  $\gamma 3$  were obtained using Probes I and II, respectively. A, *Ava*I; B, *Bam*HI; H, *Hind*III; S, *Sac*I; X, *Xba*I.

brain, heart, lung, liver, and spleen, but was undetectable in thymus and kidney.

### The mouse merlin homologue

The deduced amino acid sequence of one of our PCR products was identical to that of the corresponding region of mouse merlin (Claudio et al., 1994). Interestingly, we found two other PCR products (NBL6 and 7) with a high similarity to mouse merlin (86.4 and 90.9% identity, respectively), suggesting that these proteins together with ezrin, radixin, moesin and merlin, form a gene family within the band 4.1 superfamily (see Fig. 2C).

The expression profile of these merlin-like proteins in various types of tissues was examined by northern blotting.

NBL6 was mainly expressed in brain and thymus, while the expression of NBL7 was restricted to brain. Both northern blots revealed three hybridizing species of 2.5, 4.0 and 5.5 kb (data not shown).

### DISCUSSION

Using the PCR method with mixed primers, we attempted to provide a more detailed picture of the band 4.1 superfamily. So far, among band 4.1 superfamily members, 8 proteins have been identified; band 4.1 protein, ezrin, radixin, moesin, merlin, talin, PTPH1, and PTPMEG (Tsukita et al., 1992). In this study, cDNA fragments were obtained for all of these,

```

1  GCGGGGCCCGGTCGACGCTCTCGCCTTGCTGGAGAAGCTTTACGCCGACATGGGCTGTTTCTGCGCGTTCGCGAAGAATTTTACTGCGAAGTTTGTCTCTGGACGAGTCCAAGCTC
      M G C F C A V P E E F Y C E V L L L D E S K L
121 ACCCTCACCCAGCAGCAGGATCATCAAGAAGTCAACAAAAGGCTCTGTGTCTTGACCAGTTCCTCGTACATAAACCTCGTGGAGATAGATTATTTGGGCTCGGCTACTGTGAC
      T L T T Q Q Q G I K K S T K G S V V L D H V F R H I N L V E I D Y F G L R Y C D
      -----
241 AGAAGCCATCAGACATATTGGCTGGATCCCGCAAAACCTTGACAGAACACAGGAGCTGATCAACACTGGACCTCCATATACTTTGTATTTGGTATTAAATCTATGCTGAAGATCCA
      R S H Q T Y W L D P A K T L A E H K E L I N T G P P Y T L Y F G I K F Y A E D P
361 TGTAAGCTCAAGAAGAAATAACAGATATCAGTTTTTCTTACAGGTGAAGCAAGATGCCCTTCAGGGCCGGTGCCTTGCCCTGTCAATATTGTCTCAGATGGGGCATATGCCATC
      C K L K E E I T R Y Q F F L Q V K Q D A L Q G R L P C P V N I A A Q M G A Y A I
      -----
481 CAGGCCGAACCTGGAGACCCAGCACCATACAAGCACACTCGGGGTATGTGTCGGAGTACCGGTTTGTCTCTGATCAGAAGGAAGAGCTGGAAGAACCCATAGAAAGGATTATAAACT
      Q A E L G D H D P Y K H T A G Y V S E Y R F V P D Q K E E L E E A I E R I H K T
601 CTAATGGTCCAGGCTCTCCGAAGCTGAGCTGAATTACTTGAGGACTGCCAAATCCCTGGAGATGTATGGTGTGACCTCCATCCTGTCTATGGAGAAAATAAGTCCGAGTACTTCTTA
      L M G Q A P S E A E L N Y L R T A K S L E M Y G V D L H P V Y G E N K S E Y F L
721 GGCTTGACTCCATCGGGAGTTGTGCTCTATAAGAATAAAAAGCAAGTGGGAAGTATTTCTGCGCTCGGATTACAAGGTGCACCTCAAGAAAACCCAGTTTGAGCTCAGAGCTTTGGGA
      G L T P S G V V V Y K N K K Q V G K Y F W P R I T K V H F K E T Q F E L R V L G
841 AAAGACTGTAATGAACTCATCTTTTGAAGCTCGAAGCAAACTGCTTGCAAGCACCTCGGAAATGCAGCGTAGAGCACCATACGTTTTTCAGAATGCCGGACACAGAATCAAT
      K D C N E T S F F F E A R S K T A C K H L W K C S V E H H T F F R M P D T E S N
961 TCATTATCAAGAAAACCTCAGCAAGTTTGGTCCATAAGTTATAAACATCGGTACAGGACAGCTTTGCAAAATGAGCCGAGATCTTTCTATCCAACCTCCCGGCCCAATCAGAACGTGGTG
      S L S R K L S K F G S I S Y K H R Y R T A L Q M S R D L S I Q L P R P N Q N V V
      -----
1081 AGAAGTCGAAGCAAGACTTACCCCAAGGGTAGCAGACTCAGCCCACTGGATCAAACAACATCAATCGGATAACTGCAAACACGGAAGGAGAAATGAAGAACCCACCAAGATC
      R S R S K T Y P K R V A Q T Q P T G S N N I N R I T A N T E N G E N E G T T K I
1201 ATTGCGCTTCCCCAGTGAAGTTTAAAGAAAGCAAGAAATGAAACACGCCCTGACCCCAAGAAAGCAAAATCTCATGCACCTTGGGAAGAAAACGGACCGCAGAGTGGACTATACAAC
      I A P S P V K S F K K A K N E N S P D P Q R S K S H A P W E E N G P Q S G L Y N
1321 TCCTCCAGTGACCCGACTAAGTACCAAAGTTTCCCTGTGCCCGGCAGCGCAACCTCTCCTGCGGCAGCGACAATGACTCCTCACAACCTGATGAGGCGAAGGAAAGCTCACAACAGTGGC
      S S S D R T K S P K F P C A R Q R N L S C G S D N D S S Q L M R R R K A H N S G
1441 GAAGACTCAGATCTCAGCAGAGGAGGAGTACGCTCAGCTGTAACACAAGCAGTGGTAGTGAGTACAGAAAATCTAACAGAGAACCCGGAAGAAAAGAGAACAGAACCCGCGCAGGAG
      E D S D L K Q R R R S R S R C N T S S G S E S E N S N R E H R K K R N R T R Q E
      -----
1561 AATGACATGGTTGACTCGGGCCCTCAGTGGGAAGCCGTGTTAAGGAGACAGAAGAAAAGAACCAAGCGGACCTAACACCGGAGATCCAGACACAGATCTCGCTCAAGAAAGTCTCTGAC
      N D M V D S G P Q W E A V L R R Q K E K N Q A D P N N R R S R H R S R S P D
1681 ATCCAAGCAAAAGAAGAACTGTGAAGACACATCTAGAAGAGCTTGTGGACCCATCTGGACTGTCAAGAGAGCTGAAGGAGATCCCGTACACTAAAGTAGAGACTCAGGGTACCCCG
      I Q A K E E L W K H I *
1801 GTCCGCATCCGGCATTCTCATTCCCGCGGAGTTACCGCAGTACCGAAGATCCAGTGTTCAGACGGAGAACGATCCGTTCTCTCGGAAGTGAAGTCAAGACAGATCTGTACCACCA
1921 CTTCCGGTCAACCGTCTCTCAGATGCTCAGGGGCTGGAGGCTCGACAGTTCATCAGAGGAGAATGGGCTTAAAGACAGCTTGATTGAAGAAAAGTCTCAGCTATCTACAATCAACCCG
2041 GCTGGAAACCTACAGCAAAACCAATAAAACCATCCAGGCTGCTCGCCTCAAGGCAGAGACTTGACGGGCTGGTGCGGGTAGGAAGGAAATTCGTGCCGTGCCACTGGTACTCTG
2161 AAAGTGAATTAGTTCATTATCCATGCACCCCTGGACATGTGCTTTAGCATCGTCTAAGGATTTATGTACAGCATTCTGTGCTCTGGGCTTTGGGAGGACAACTTATAAACTCACT
2281 GGATACCACATCTATTTCTGAAATTCGACCATCCAACCTCCACCCAGTCAAGACAGCAACTCCGATGAGCCCTGGACTCTGGGCTGCATCAGACTGGTACTTTTATCCCA
2401 CCAAAGTGTTCATAATTGTATCCCGTGGCACTAAGTCCCAACCGTATCTCAAATAATGTAAAAGCTGTTAAGTGTGATTTGAAAAAATAAAAAAAAAA

```

**Fig. 4.** Nucleotide and deduced amino acid sequences of NBL4. The ORF encodes a 554 amino acid polypeptide with a predicted molecular mass of 61.0 kDa. The band 4.1 homology domain is shadowed. The consensus sequences for myristoylation ( $\equiv$ ), A-kinase phosphorylation (---) and tyrosine-kinase phosphorylation (—) are underlined. These sequence data are available from EMBL/GenBank/DBJ under accession number D28818.

indicating that our screening system with the PCR method was powerful enough to identify novel band 4.1 superfamily members.

Using this system, in addition to cDNA clones for the known members, we identified 7 cDNA clones that appear to encode novel band 4.1 superfamily proteins (NBL1-7). Sequence analyses of these clones revealed that three (NBL1-3) and two (NBL6,7) of them had high similarity to the corresponding sequence of PTPH1/PTPMEG and ezrin/radixin/moesin/merlin, respectively, and that the other two members (NBL4,5) were less closely related to any of the currently known band 4.1 superfamily members than are NBL1,2,3,6 and 7. These data suggested that the band 4.1 superfamily can be subclassified into at least 5 gene families; band 4.1 (band 4.1 protein), ERM (ezrin, radixin, moesin, merlin, NBL6, NBL7), talin (talin), PTPH1 (PTPH1, PTPMEG, NBL1, NBL2, NBL3), and NBL4 (NBL4, NBL5). In the band 4.1 and talin families, only

one member has been identified so far for each family. Therefore, in a strict sense, they should not be called 'family' at present. Of course, since we screened only mouse F9 cells and brain, other families may be identified in different types of cells in future.

As shown in Fig. 2A, PCR products of NBL1-3 show strong similarity to the N-terminal half of PTPH1 and PTPMEG, indicating the existence of a gene family, which we tentatively called PTPH1 family. Recently, various protein tyrosine phosphatases have been identified, among which these PTPH1 and PTPMEG were characterized by the presence of a band 4.1 homology domain in their N-terminal half (Gu et al., 1991; Yang and Tonks, 1991). These proteins contain a protein tyrosine phosphatase domain in their C-terminal half. At present, it is not clear whether or not NBL1-3 also have this domain. However, these novel members may also be involved in the signal transduction just beneath the plasma membrane

MERLIN	M-SFSSLKR	KQ-----	PKFTT----	VRIVTMDAEM	-EFNCEMKWK	GKDLFDLVC	TLGLRWF	GLOY-	58
RADIXIN	M-----	-----	PKPIN----	VRVVTMDAEL	-EFAIQPNTT	GKQLFDQVVK	TVGLREVWF	GLOYV	50
BAND41	MH-----	-----	-----	CKVSLDDTV	YECVVEKHAK	GODLLKRVCE	HLNLEEDMF	GLAIW	47
PTPH1	MTSRLRALGG	RINNIRTSSEL	PKEKTRSEVI	CSIHFLDGTV	QTFKVKQDT	QVLLDMVHN	HLGVTEKEYF	GLQHD	75
NBL 4	MGC-----	-----	FCAV	PEEFYCEVLL	LDESKLTLTT	QQGGIKKSTK	GSVVLDHVFR	HINLVEIDMF	62
Consensus	M-----	-----	PK.....	.....LD...	.EF...K..K	G...LLD.V...	HLGL.E...YE	GLOY.	75
MERLIN	TIK-DTVVNL	KMDKKVL-DH	DV-SKEEPT	FFFLAKFYPE	NAEEELVQEI	TDHLFFLQVK	KQILDEKIYC	PPEAS	130
RADIXIN	DSK-GYSTWL	KLNKKVT-QQ	DV-KKENPLQ	FKFRAKFFPE	DVSEELIQEI	TORLFFLQVK	EAILNDEIYC	PPETA	122
BAND41	DNA-TSKTWL	DSAKEI--KK	QV-RG-VPWN	PTFNVMFYPP	DPA-CLTEDI	TRYYLQLQLR	QDIVAGRLFC	SFATL	116
PTPH1	DDSDVSPFWL	EASKPI--RK	QL-KGGFECT	LHFRVFFIP	DPN-TLQEQE	TRHLYFLQVK	MDICEGRLLC	PLNSA	146
NBL 4	DRS-HQTYWL	DPAKTLAEHK	ELINTGEPNT	LYFELIKYAE	DPC-KLKEEI	TRYQFFLQVK	QDALQGRLEC	PVNIA	135
Consensus	D.....WL	...K...-K	.V-...E.T	F.E...KEYPE	DP-..L.QEI	TR.LFFLQVK	.DIL.GRL.C	P...A	150
MERLIN	VLLASYAVQA	KYGDYDPSVH	KRGFLAQEEL	IFKRVINLYQ	MTPEMWEERI	TAWYAEHRGR	ARDEAEMEYL	KIADD	205
RADIXIN	VLLASYAVQA	KYGDYDNKEIH	KPGYLANDRL	IFQRVLEQHK	LTKEQWEERI	QNWHEEHRGM	LREDSMMEYL	KIADD	197
BAND41	ALLGSYTIOS	ELGDYDPEIH	GVDYVSDFKL	IFNQTKLEE	-----KV	MELHKSYSRM	TPAQADLEFL	ENAKK	183
PTPH1	VVLSASYAVQS	HFGDYNSSIH	HPGYLSDSHF	IFDQNEFLT	-----KV	ESLHEQHSGL	KQSEAESCYI	NIART	213
NBL 4	AQMGAAYAQA	ELGDHDPYKH	TAGYVSEYRF	VEDQKEELE	-----AI	ERIHKTLMGQ	APSEAEINYL	RTAKS	202
Consensus	VLLASYAVQA	.GDYDP..H	..GYLS...L	E.Q...EL..	-----I	...H..HRG.	...EAE.EYL	.IAD.	225
MERLIN	LEMVGVNYFA	IRNKKGTELL	IGVDAIGLHI	IFPENRLTPK	ISFPWNEIRN	ISYSD-KEFT	IKPLDKKIDV	PKFNS	279
RADIXIN	LEMVGVNYFE	IKNKKGTELW	IGVDAIGLNI	YEHDDKLTPK	IGFPWSEIRN	ISFND-KKFV	IKPIDKKAPD	FVFFA	271
BAND41	LSMYGVDLHK	AKDLEGVDII	IGVCSGSLV	YKDKLRINRF	PWPVKLKISY	KRSSFFKIR	PGEQEQYEST	IGFNL	258
PTPH1	LDFYGVLEHS	GRDLNLDLM	IGIASAGVAV	YRKYICTSFY	PWVNILKISF	KRKKFFIHQR	QKQAESREHI	VAFNM	288
NBL 4	LEMVGVDLHP	VYGENKSEYF	IGLTPSGVVV	YKNNKQVGKY	FWPRIKTVHF	KETQF--ELR	VLGKDCNETS	FFFEA	275
Consensus	LEMVGV.LH.	.....G.EL.	IGV...GL.V	M.....	.W...KI..	K...F...R	.K..D..E..	F.FE..	300
MERLIN	SKLRVRLIL	QLCIGNHDLF	M-----	-----	-----	-----	-----	-----	300
RADIXIN	PRLRINKRIL	ALCMGNHELY	M-----	-----	-----	-----	-----	-----	300
BAND41	PSYRAKKLW	KVCVEHHTFF	RLTSTDITPK	SKFLALGSKF	RY	-----	-----	-----	300
PTPH1	LNYSCKNLW	KS-----	-----	-----	-----	-----	-----	-----	300
NBL 4	RSKTACKHLW	KCSVEHHTFF	RMPDT-----	-----	-----	-----	-----	-----	300
Consensus	...R..K.LW	K.C...H..F	-----	-----	-----	-----	-----	-----	342

Fig. 5. Alignment of the deduced amino acid sequences of the N-terminal half of five band 4.1 superfamily members whose full-length cDNAs were obtained. Conserved amino acids are indicated in 'consensus'. The positions where all sequences share an identical amino acid are boxed. Gaps, introduced to maximize alignment, are indicated by dashes.

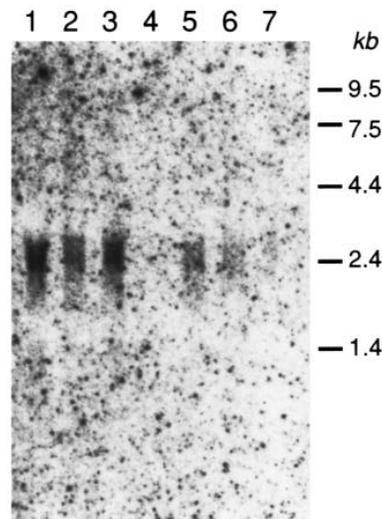


Fig. 6. Detection of NBL4 mRNA in various tissues by northern hybridization. The poly(A)<sup>+</sup> RNA samples were separated by electrophoresis in an agarose gel and transferred to a nitrocellulose membrane. The HindIII-AvaI fragment of NBL4 (see Fig. 3) cDNA was used to probe the RNA blot. The mobility of the RNA standards (kb) is shown at the right. Lane 1, brain; lane 2, heart; lane 3, lung; lane 4, thymus; lane 5, liver;

lane 6, spleen; lane 7, kidney. The expression level of NBL4 is rather high in the brain, heart and lung.

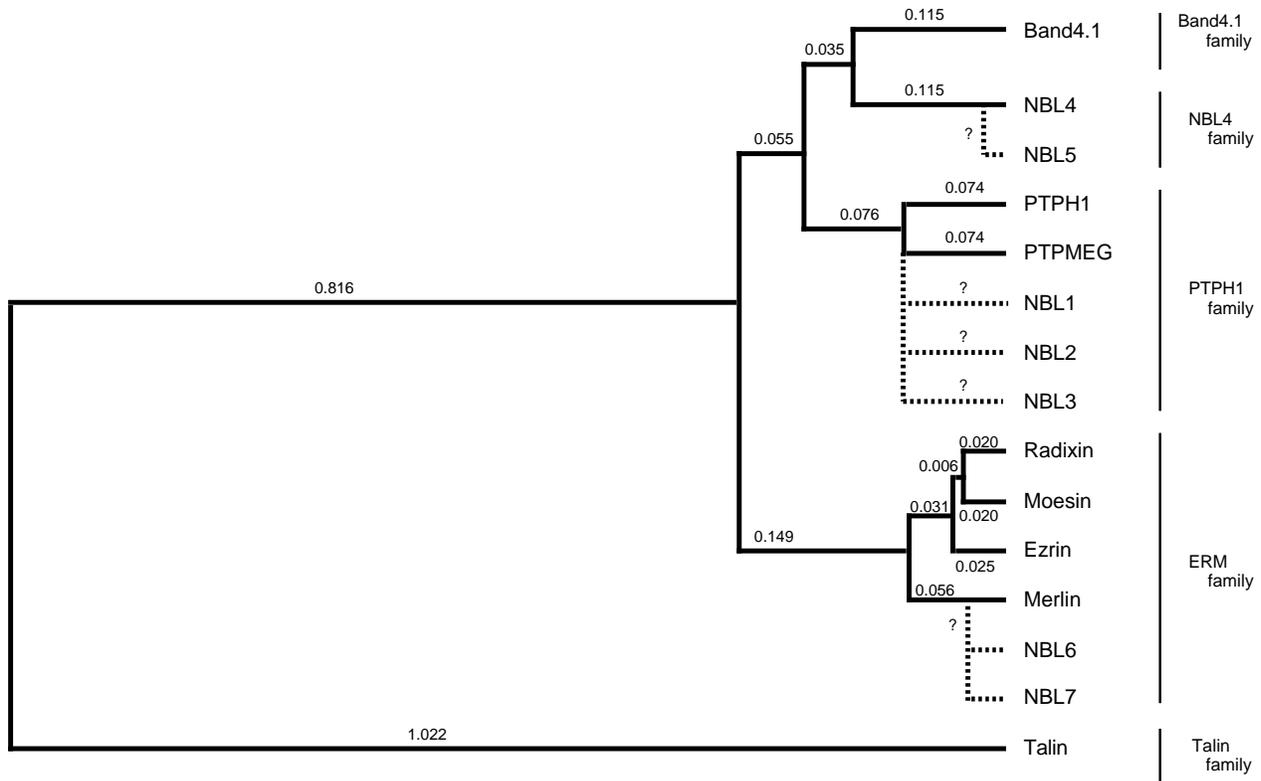
through some enzymic activities such as phosphatase or kinase activities.

There remains no information about the functions of NBL4 family members identified here. The amino acid sequence of

NBL4 deduced from its full-length cDNA revealed that its N-terminal half has similarity to band 4.1 superfamily members, but that its C-terminal half shows no homology with sequences in proteins so far identified. The only clue to the function of NBL4 was the occurrence of consensus sequences for myristoylation, A-kinase phosphorylation, and tyrosine kinase phosphorylation. Further analysis with this NBL4 cDNA will lead to a better understanding of the functions of this novel protein.

One of our PCR products was identical to the corresponding part of mouse merlin. Two other PCR products (NBL6 and 7) were highly similar to this mouse merlin (86.4% and 90.9% identity, respectively). Judging from the similarity between ezrin, radixin, moesin, merlin, NBL6, and NBL7, these 6 proteins are included in a gene family (ERM family). Northern blots revealed that both NBL6 and NBL7 are preferentially expressed in the brain. Considering that merlin plays a crucial role in the regulation of cell growth in the nervous tissue, it is likely that NBL6 and NBL7 are also involved in the same process. Further analyses on the interaction and relationship between merlin, NBL6 and NBL7 will be required to understand how merlin works as a tumor suppressor in neurofibromatosis 2.

To deduce the possible evolutionary relationships among the band 4.1 superfamily members, the amino acid sequences of the N-terminal homology domains of members were quantitatively compared, and the genetic distance for each pair was



**Fig. 7.** A potential phylogenetic tree of the band 4.1 superfamily. The tree was constructed by the unweighed pair group method (Fitch and Margoliash, 1967; Nei, 1987), using the calculated genetic distances (presented numerically) between pairs of members.

estimated according to the method described by Doolittle (1984). Based upon these distances, a phylogenetic tree of the band 4.1 superfamily members was constructed using the unweighed pair-group method (Fitch and Margoliash, 1967; Nei, 1987) and is shown in Fig. 7. The simplest interpretation of this tree is that talin and the primordial form of the other members were first generated and that the primordial form of the ERM family members differentiated from that of band 4.1/NBL4/PTPH1 families.

This study confirmed that various types of families are included in the band 4.1 superfamily, suggesting that it is involved in various cellular events. However, it is likely that the band 4.1 homology domain localized at the N-terminal domain of the members is responsible for their direct association with membrane proteins, suggesting that all the members work just beneath the plasma membrane. For a better understanding of the band 4.1 superfamily, we should next answer the question: in which cellular events just beneath the plasma membrane is each family (in the band 4.1 superfamily) involved?

We thank all the members of our laboratory (Laboratory of Cell Biology, National Institute for Physiological Sciences) for their helpful discussions. Our thanks are also due to Dr S. Yonemura for his critical reading of this manuscript, and to Dr M. Hayashi for his collaboration in the computer analysis on the phylogenetic tree. This work was supported by research grants from the Ministry of Education, Science and Culture of Japan and from the Ministry of Health and Welfare of Japan.

## REFERENCES

- Anderson, R. A. and Lovrien, R. E.** (1984). Glycophorin is linked by band 4.1 protein to the human erythrocyte membrane skeleton. *Nature* **307**, 655-658.
- Anderson, R. A. and Marchesi, V. T.** (1985). Regulation of the association of membrane skeletal protein 4.1 with glycophorin by a polyphosphoinositide. *Nature* **318**, 295-298.
- Bennett, V.** (1989). The spectrin-actin junction of erythrocyte membrane skeletons. *Biochim. Biophys. Acta* **988**, 107-121.
- Berryman, M., Franck, Z. and Bretscher, A.** (1993). Ezrin is concentrated in the apical microvilli of a wide variety of epithelial cells whereas moesin is found primarily in endothelial cells. *J. Cell Sci.* **105**, 1025-1043.
- Bretscher, A.** (1983). Purification of an 80,000-dalton protein that is a component of the isolated microvillus cytoskeleton, and its localization in nonmuscle cells. *J. Cell Biol.* **97**, 425-432.
- Bretscher, A.** (1989). Rapid phosphorylation and reorganization of ezrin and spectrin accompany morphological changes induced in A-431 cells by epidermal growth factor. *J. Cell Biol.* **108**, 921-930.
- Burridge, K. and Connell, L.** (1983). A new protein of adherens plaques and ruffling membranes. *J. Cell Biol.* **97**, 359-367.
- Burridge, K. and Mangeat, P.** (1984). An interaction between vinculin and talin. *Nature* **308**, 744-746.
- Claudio, J. O., Marineau, C. and Rouleau, G. A.** (1994). The mouse homologue of the neurofibromatosis type 2 gene is highly conserved. *Hum. Mol. Genet.* **3**, 185-190.
- Conboy, J., Kan, Y. W., Shoheit, S. B. and Mohandas, N.** (1986). Molecular cloning of protein 4.1, a major structural element of the human erythrocyte membrane skeleton. *Proc. Nat. Acad. Sci. USA* **83**, 9512-9516.
- Doolittle, R. F.** (1984). In *The Plasma Proteins*, vol. IV (ed. F. W. Putnam), pp. 317-360. Academic Press, Orlando, FL.
- Fitch, W. M. and Margoliash, E.** (1967). Construction of phylogenetic trees. *Science* **155**, 279-284.
- Fowler, V. and Taylor, D. L.** (1980). Spectrin plus band 4.1 cross-link actin. *J. Cell Biol.* **85**, 361-376.

- Franck, Z., Gary, R. and Bretscher, A.** (1993). Moesin, like ezrin, colocalizes with actin in the cortical cytoskeleton in cultured cells, but its expression is more variable. *J. Cell Sci.* **105**, 219-231.
- Funayama, N., Nagafuchi, A., Sato, N., Tsukita, Sa. and Tsukita, Sh.** (1991). Radixin is a novel member of the band 4.1 family. *J. Cell Biol.* **115**, 1039-1048.
- Gould, K. L., Cooper, J. A., Bretscher, A. and Hunter, T.** (1986). The protein-tyrosine kinase substrate, p81, is homologous to a chicken microvillar core protein. *J. Cell Biol.* **102**, 660-669.
- Gould, K. L., Bretcher, A., Esch, F. S. and Hunter, T.** (1989). cDNA cloning and sequencing of the protein-tyrosine kinase substrate, ezrin, reveals homology to band 4.1. *EMBO J.* **8**, 4133-4142.
- Gu, M., York, J. D., Warshawsky, I. and Majerus, P. W.** (1991). Identification, cloning, and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase with sequence homology to cytoskeletal protein 4.1. *Proc. Nat. Acad. Sci. USA* **88**, 5867-5871.
- Hunter, T. and Cooper, J. A.** (1981). Epidermal growth factor induces rapid tyrosine phosphorylation of proteins in A431 human tumor cells. *Cell* **24**, 741-752.
- Hunter, T. and Cooper, J. A.** (1983). Role of tyrosine phosphorylation in malignant transformation by viruses and in cellular growth control. *Prog. Nucl. Acid Res. Mol. Biol.* **29**, 221-233.
- Lankes, W., Griesmacher, A., Grunwald, J., Schwartz-Albiez, R. and Keller, R.** (1988). A heparin-binding protein involved in inhibition of smooth-muscle cell proliferation. *Biochem. J.* **251**, 831-842.
- Lankes, W. and Farthmayr, H.** (1991). Moesin: A member of the protein 4.1-talin-ezrin family of proteins. *Proc. Nat. Acad. Sci. USA* **88**, 8297-8301.
- Leto, T. L., Correas, L., Tobe, T., Anderson, R. A. and Horne, W. C.** (1986). In *Membrane Skeleton and Cytoskeletal Membrane Associations* (ed. V. Bennet, C. M. Cohen, S. E. Lux, and J. Palek), pp. 210-209. Alan R. Liss, New York.
- Mueller, T. J. and Morrison, M.** (1981). In *Erythrocyte Membranes 2: Recent Clinical and Experimental Advances* (ed. Kruckeberg, Eaton and Brewer), pp. 95-112. Alan R. Liss, New York.
- Nagafuchi, A., Shirayoshi, Y., Okazaki, K., Yasuda, K. and Takeichi, M.** (1987). Transformation of cell adhesion properties by exogenously introduced E-cadherin cDNA. *Nature* **329**, 341-343.
- Nei, M.** (1987). *Molecular Evolutionary Genetics*. pp. 293-298. Columbia University Press, New York, NY.
- Ohanian, V., Wolfe, L. C., John, K. M., Pinder, J. C., Lux, S. E. and Gratzner, W. B.** (1984). Analysis of the ternary interaction of the red cell membrane skeletal proteins spectrin, actin, and 4.1. *Biochemistry* **23**, 4416-4420.
- Pakkanen, R., Hedman, K., Turunen, O., Wahlstrom, T. and Vaheri, A.** (1987). Microvillus-specific Mr 75,000 plasma membrane protein of human chorio carcinoma cells. *J. Histochem. Cytochem.* **135**, 809-816.
- Rees, D. J. G., Ades, S. E., Singer, S. J. and Hynes, R. O.** (1990). Sequence and domain structure of talin. *Nature* **347**, 685-689.
- Rouleau, G. A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B., Pulst, S. M., Le noir, G., Bijlsma, E., Fashold, R., Dumanski, J., de Jong, P., Parry, D., El drige, R., Aurias, A., Delattre, O. and Thomas, G.** (1993). Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2. *Nature* **363**, 515-521.
- Saiki, R. K.** (1989). *PCR Technology* (ed. H. A. Erlich), pp. 7-22. Stockton Press, NY.
- Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sato, N., Funayama, N., Nagafuchi, A., Yonemura, S., Tsukita, Sa. and Tsukita, Sh.** (1992). A gene family consisting of ezrin, radixin, and moesin. Its specific localization at actin filament/plasma membrane association sites. *J. Cell Sci.* **103**, 131-143.
- Towler, D. A. and Gordon, J. I.** (1988). The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem.* **57**, 69-99.
- Tsukita, Sa., Hieda, Y. and Tsukita, Sh.** (1989a). A new 82 kD-barbed end capping protein localized in the cell-to-cell adherens junction: Purification and characterization. *J. Cell Biol.* **108**, 2369-2382.
- Tsukita, Sh., Itoh, M., Nagafuchi, A., Yonemura, S. and Tsukita, Sa.** (1993). Submembranous junctional plaque proteins include potential tumor suppressor molecules. *J. Cell Biol.* **123**, 1049-1053.
- Tsukita, Sh., Tsukita, Sa., Nagafuchi, A. and Yonemura, S.** (1992). Molecular linkage between cadherins and actin filaments in cell-to-cell adherens junctions. *Curr. Opin. Cell Biol.* **4**, 834-839.
- Trofatter, J. A., MacCollin, M. M., Rutter, J. L., Murrell, J. R., Duyao, M. P., Parry, D. M., Eldridge, R. E., Kley, N., Menon, A. G., Pulaski, K., Haase, V. H., Ambrose, C. M., Munroe, D., Bove, C., Haines, J. L., Martuza, R. L., MacDonald, M. E., Seizinger, B. R., Short, M. P., Buckler, A. J. and Gusella, J. F.** (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* **72**, 791-800.
- Turunen, O., Winqvist, R., Pakkanen, R., Grzeschik, K. H., Wahlstrom, T. and Vaheri, A.** (1989). Cytovillin, a microvillar Mr 75,000 protein. cDNA sequence, prokaryotic expression, and chromosomal localization. *J. Biol. Chem.* **264**, 16727-16732.
- Ungewickell, E., Bennett, P. M., Calvert, R., Ohanian, V. and Gratzner, W.** (1979). In vitro formation of a complex between cytoskeletal proteins of the human erythrocyte. *Nature* **280**, 811-814.
- Yang, Q. and Tonks, N. K.** (1991). Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with homology to the cytoskeletal-associated proteins band 4.1, ezrin, and talin. *Proc. Nat. Acad. Sci. USA* **88**, 5949-5953.

(Received 20 December 1993 - Accepted 3 March 1994)