EVH1/WH1 domains of VASP and WASP proteins belong to a large family including Ran-binding domains of the RanBP1 family

Isabelle Callebaut[a], Pascale Cossart[b], Pierre Dehoux[b,*]

[a] *Systèmes moléculaires et Biologie structurale, LMCP, CNRS UMRC 7590, U.P.6/U.P.7, Paris, France*
[b] *Unité des Interactions Bactéries Cellules, Institut Pasteur, 28 rue du Dr Roux, 75015 Paris, France*

Received 30 September 1998; received in revised form 13 November 1998

Abstract The two cytoskeletal proteins VASP and WASP and the protein Homer share a conserved domain, currently designated the WH1 domain (WASP homology domain 1) or EVH1 domain (ENA/VASP homology domain 1), which could play an important role in various cellular events such as transport, folding of proteins, and signal transduction. We report here additional occurrences of this domain in Ran-binding proteins of the RanBP1 family and various others proteins, or putative proteins of eukaryotic organisms, suggesting that the EVH1/WH1 domain may be more widely used than originally thought.

Key words: Sequence analysis; Polyproline; WH1 domain; EVH1 domain

1. Introduction

The ENA/VASP family proteins initially included the vasodilatator-stimulated phosphoprotein (VASP) [1,2], the Enabled protein of *Drosophila* (Ena) [3] and murine Mena and Evl proteins [3]. ENA/VASP proteins share two homology domains, one of which is the EVH1 domain (ENA/VASP homology domain 1), an N-terminal 115 amino acid domain present in the six currently known members of the VASP family. The EVH1 domain is responsible for binding of VASP and Mena to cytoskeleton-interacting proteins, including ActA of the motile intracellular bacterial pathogen *Listeria monocytogenes*, and the eukaryotic proteins vinculin and zyxin [3,4]. The biochemical data obtained with EVH1 domains of VASP and Mena indicate that it mediates interactions with polyproline-rich domains of bacterial actin/cytoskeleton-related protein ActA [4]. The N-terminal region of Mena, which contains the EVH1 domain, can mediate interactions with proteins containing an Acta-like polyproline motif such as zyxin and vinculin [3]. Previous analyses showed that EVH1 shares similarity with the N-terminus of another cytoskeleton-related protein, the Wiskott-Aldrich syndrome protein (WASP) [5]. This domain, also termed WH1 (for WASP homology domain 1), is present in the seven currently known members of the WASP family and is responsible for binding of WASP with a polyproline-containing domain of an actin-related protein WIP [6]. WH1 also mediates binding of N-WASP with phosphatidylinositol 4,5-bisphosphate (PIP2), and it has therefore been postulated that WH1 could be a pleckstrin homology (PH) domain [7]. Significant similarities were also found between these domains and the N-terminus of the recently identified Homer protein [8]. This 189 amino acid protein consists of an amino-terminal WH1 domain, followed by a putative PEST degradation motif [9]. The Homer WH1 domain, which has been postulated to be a divergent PDZ domain, binds selectively the C-terminus of phosphoinositide-linked metabotropic glutamate receptors [10].

Here, we have characterized in detail the EVH1/WH1 domains and identified new members of the EVH1/WH1/Homer family possessing one or more copies of the domain.

2. Materials and methods

Sequence analyses were performed on the mainframe computer run by the Institut Pasteur Computer Services. Profile analysis [11] was performed using the Genetics Computer Group software package version 9.1 [12] with Blosum matrix series. Iterative database searches were performed using the new Blast family programs gapped-Blast2 and PSI-Blast [13] with the non-redundant NCBI database (release March 1998). Compositional analysis was performed with the SEG program [14].

Multiple alignments were performed using a combination of automatic procedures such as ClustalW [15], MACAW [16] and the bidimensional hydrophobic cluster analysis (HCA) [17,18]. HCA combines sequence comparison with secondary structure prediction and is particularly efficient in comparing sequences sharing low levels of sequence identity (e.g. see the identification of the BRCT module [19]). The HCA-based secondary structure predictions were strengthened by those performed by the profile neural network prediction PHD program [20].

Statistical significance of the multiple alignment was evaluated using BLAST- and PSI-BLAST-associated *E* values, estimation of *P* values by MACAW [16] and calculation of *Z* scores as described in [18].

3. Results and discussion

We searched (see Section 2) the non-redundant NCBI database for proteins which contain sequences similar to EVH1/WH1/Homer domains, hereafter designated for simplicity WH1 domains. Using an iterative strategy, we found interesting similarities between the WH1 domains of the VASP family and several sequences not described so far as members of the WH1 family, such as the Ran-binding domain found in RanBP1 and proteins of the nuclear pore complex. For example, using the ENA/VASP-like protein RNB6 as a seed (amino acids 1–150), PSI-BLAST converged after five iterations (threshold *E* values of 0.001; 0.001; 0.01; 0.01 and 0.001 for the successive steps) giving an unambiguous connection to the RanBP1 family (e.g. RANG_MOUSE; *E* value 10^{-42}). Other interesting matches were also found with rat Homer and *Drosophila* AE33 proteins (PSI-BLAST *E* values (following convergence after two iterations, threshold *E* value 0.01) of 2×10^{-4} and 2×10^{-9}, respectively), as well as with

*Corresponding author. Fax: (33) (1) 45 68 87 06.
E-mail: pdehoux@pasteur.fr

PII: S0014-5793(98)01541-5

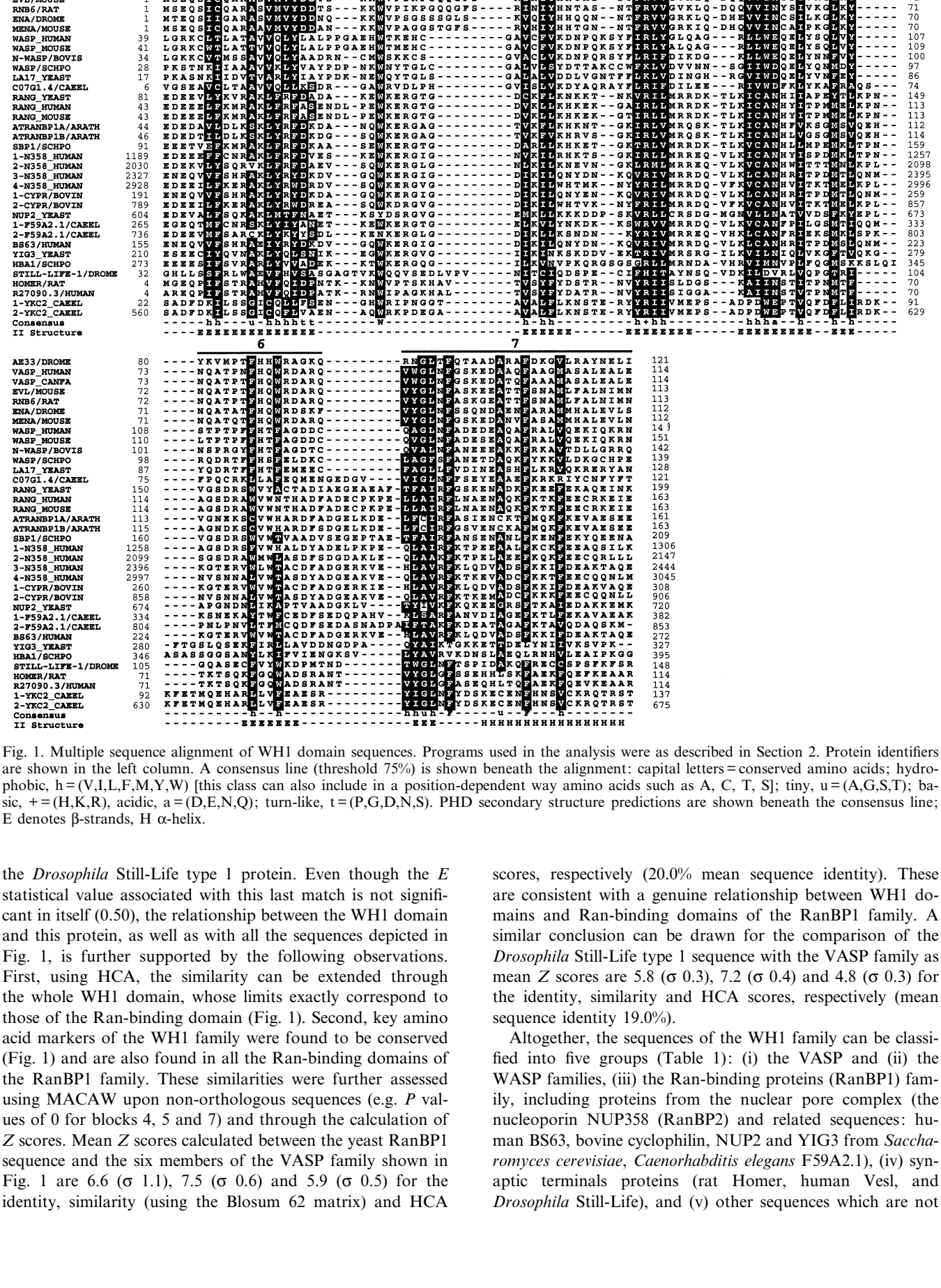

Fig. 1. Multiple sequence alignment of WH1 domain sequences. Programs used in the analysis were as described in Section 2. Protein identifiers are shown in the left column. A consensus line (threshold 75%) is shown beneath the alignment: capital letters = conserved amino acids; hydrophobic, h = (V,I,L,F,M,Y,W) [this class can also include in a position-dependent way amino acids such as A, C, T, S]; tiny, u = (A,G,S,T); basic, + = (H,K,R), acidic, a = (D,E,N,Q); turn-like, t = (P,G,D,N,S). PHD secondary structure predictions are shown beneath the consensus line; E denotes β-strands, H α-helix.

the *Drosophila* Still-Life type 1 protein. Even though the *E* statistical value associated with this last match is not significant in itself (0.50), the relationship between the WH1 domain and this protein, as well as with all the sequences depicted in Fig. 1, is further supported by the following observations. First, using HCA, the similarity can be extended through the whole WH1 domain, whose limits exactly correspond to those of the Ran-binding domain (Fig. 1). Second, key amino acid markers of the WH1 family were found to be conserved (Fig. 1) and are also found in all the Ran-binding domains of the RanBP1 family. These similarities were further assessed using MACAW upon non-orthologous sequences (e.g. *P* values of 0 for blocks 4, 5 and 7) and through the calculation of *Z* scores. Mean *Z* scores calculated between the yeast RanBP1 sequence and the six members of the VASP family shown in Fig. 1 are 6.6 (σ 1.1), 7.5 (σ 0.6) and 5.9 (σ 0.5) for the identity, similarity (using the Blosum 62 matrix) and HCA scores, respectively (20.0% mean sequence identity). These are consistent with a genuine relationship between WH1 domains and Ran-binding domains of the RanBP1 family. A similar conclusion can be drawn for the comparison of the *Drosophila* Still-Life type 1 sequence with the VASP family as mean *Z* scores are 5.8 (σ 0.3), 7.2 (σ 0.4) and 4.8 (σ 0.3) for the identity, similarity and HCA scores, respectively (mean sequence identity 19.0%).

Altogether, the sequences of the WH1 family can be classified into five groups (Table 1): (i) the VASP and (ii) the WASP families, (iii) the Ran-binding proteins (RanBP1) family, including proteins from the nuclear pore complex (the nucleoporin NUP358 (RanBP2) and related sequences: human BS63, bovine cyclophilin, NUP2 and YIG3 from *Saccharomyces cerevisiae*, *Caenorhabditis elegans* F59A2.1), (iv) synaptic terminals proteins (rat Homer, human Vesl, and *Drosophila* Still-Life), and (v) other sequences which are not

Table 1
Protein families containing WH1 domains

Protein	Organism	Description	Swissprot/Trembl accession number
VASP family			
VASP_HUMAN	human	Vasodilatator phosphoprotein	P50552
VASP_CANFA	*Canis*	Vasodilatator phosphoprotein	P50551
EVL/MOUSE	mouse	ENA-VASP-like protein (Evl)	P70429
RNB6/RAT	rat	RNB6, a member of the ENA/VASP family	Ø08719
MENA/MOUSE	mouse	Mena, a relative of VASP and Enabled	P70430
ENA/DROME	*Drosophila*	Abl substrate ena (enabled)	Q24035
WASP family			
WASP_HUMAN	human	Wiskott-Aldrich syndrome protein (WASP)	P42768
WASP_MOUSE	mouse	WASP homolog	P70315
N-WASP/BOVIS	bovine	Actin-depolymerizing protein, N-WASP	Q95107
LA17_YEAST	*S. cerevisiae*	Proline-rich protein LAS17 or YOR181W	Q12446
WASP/SCHPO	*S. pombe*	SPAC4F10.15c similar to LAS17	Ø36027
C07G1.4/CAEEL	*C. elegans*	Similarity to WASP proteins	Q17795
RANBP1 family/nucleoporins			
RANG_YEAST	*S. cerevisiae*	RAN-binding protein 1 homolog (RANBP1)	P41920
RANG_HUMAN	human	RAN-specific binding protein 1 (RANBP1)	P43487
RANG_MOUSE	mouse	RAN-specific binding protein 1 (RANBP1)	P34023
RANBP1A/ARATH	*A. thaliana*	RAN-binding protein 1 homolog (RANBP1)	Ø04149
RANBP1B/ARATH	*A. thaliana*	RAN-binding protein 1 homolog (RANBP1)	Ø04150
SBP1/SCHPO	*S. pombe*	RAN/SPI1-binding protein	Q09717
N358_HUMAN	human	Nuclear pore complex protein NUP358	P49792
BS63/HUMAN	human	BS-63	Q99666
NUP2_YEAST	*S. cerevisiae*	Nuclear pore protein nup2 or YLR335W	P32499
F59A2.1/CAEEL	*C. elegans*	F59A2.1 homolog to nucleoporin	Q21021
CYPR/BOVIN	bovine	Cyclophilin (fragment)	Q28091
YIG3_YEAST	*S. cerevisiae*	Hypothetical 36.1 kDa protein	P40517
Synaptic terminal proteins			
HOMER/RAT	rat	GLGF domain protein Homer	Ø08567
R27090.3/HUMAN	human	Hypothetical 41.3 kDa human protein	Ø14580
STILL-LIFE-1/DROME	*Drosophila*	Still-Life type 1	P91621
Others			
YKC2_CAEEL	*C. elegans*	Hypothetical 115.2 kDa protein	P41993
AE33/DROME	*Drosophila*	Proline-rich protein	JC5909[a]
HBA1/SCHPO	*S. pombe*	Brefeldin A resistance protein	Q09146

[a]Swissprot/Trembl accession number is not yet available for the AE33 sequence. The GenBank accession number is given instead.

classified including a *C. elegans* hypothetical protein (YKC2) and *Drosophila* AE33.

The average length of the WH1 domain, which is globular as assessed by compositional analysis and HCA structural predictions, is 115 amino acids. Pairwise alignments between non-orthologous sequences revealed an identity ranging from 18% to 31%. All aligned sequences share a convincingly similar distribution of charged, hydrophobic and turn-promoting residues (Fig. 1). In particular, in addition to hydrophobic residues, most of which would participate in the structural core of the domain, three aromatic residues are strikingly conserved (tryptophan in block 2; two phenylalanine in block 7), as well as a basic residue (Arg, Lys or His in block 4). A position containing asparagine, aspartic acid or glutamic acid is also particularly well conserved in block 5. Secondary structure predictions suggest that the WH1 domain consists of seven or eight β-strands (the eighth one which would be included in block 2 containing the conserved tryptophan) associated in one or more β-sheet(s) and a single amphiphilic helix. The C-terminus block (block 7) includes a β-strand containing a typical motif tightly followed by an amphiphilic α-helix. This typical motif is constituted by turn-promoting residues associated with one hydrophobic residue and one invariant phenylalanine and has been previously described in the Homer sequence as a divergent PDZ 'GLGF motif' [10]. However, as already noticed by Ponting [8], this motif has been improperly assigned to a PDZ domain which in fact actually differs from WH1 domains. In particular, the GLGF motif of the PDZ domain is located in a loop between the two first β-strands of a β-sandwich, with the phenylalanine participating in the second β-strand, and not, as in the WH1 domain, at the end of the last strand of a predicted β7α (or β8α) structure. In this context, it is worth noting that the predicted secondary structures of the WH1 domain are strikingly similar to those of the PH domain, comprising an antiparallel β-sandwich, with two roughly perpendicular β-sheets, followed by an amphipathic α-helix. This observation raises the possibility that, despite the lack of sequence similarity, WH1 and PH domains could be structurally related. This hypothesis is strengthened by the fact that this fold is apparently widespread in signal transduction modules without any obvious sequence similarities, as illustrated by the structural similarity of the PH/PTB domains [21].

Aligned sequences do not share conserved phosphorylation motifs. The two invariant hydrophobic aromatic residues may be exposed at the surface and may be functional as in the case of SH3 [22] and profilin [23].

Alteration of the WH1 domain was observed in Wiskott-Aldrich syndrome patients. In most cases, mutations were localized within the WH1 domain indicating that this domain is critical for WASP function [24,25]. Moreover, in vitro analysis of a point mutation within the WH1 domain of N-WASP protein [26] and EVH1 domain of ENA [27] confirm these clinical observations.

Sequence similarities uncovered during this work confirm the pleiotropy of the WH1 domain, since proteins which are not directly related to actin/cytoskeleton components also contain WH1 domains. Beside rat Homer, which would act as a dominant negative regulator for receptor and/or component of the adhesion/cytoskeleton apparatus [10,28], two other synaptic terminal proteins, the human Vesl [28] and *Drosophila* Still-Life type 1 [29] comprise a WH1 domain. Human Vesl is a protein which shares a striking similarity with Homer in the amino-terminal WH1 domain and also presents a PEST motif [9]. In addition, it contains two coiled-coil domains that could be responsible for subsequent oligomerization. However, no function has yet been postulated for this protein. *Drosophila* Still-Life type 1 is a protein with multiple interactive domains which presents a WH1 domain at the N-terminus followed by a region showing similarities with the mouse TIAM-1 guanine nucleotide exchange factor (GEF) protein [29]. No ligand for the WH1 domain of this protein has yet been characterized, but deletion of the N-terminus including this domain induces membrane ruffling with altered actin localization [29].

As revealed by this analysis, another group of proteins, composed of Ran-binding proteins of the RanBP1 family which contains proteins from the nuclear pore complex (NPC), have one or more copies of the WH1 domain. RanBP1 is a predominantly cytoplasmic protein which forms a trimeric complex with GTP and the small GTPase Ran. This abundant GTPase of the Ras superfamily plays an essential role in nuclear transport, as well as in the maintenance of the integrity of nuclei and cell cycle control [30,31]. Ran appears to be bound to GTP in the nucleus, while in the cytoplasm it is mainly bound to GDP, as a result of the asymmetric distribution of factors that promote guanine nucleotide exchange (in the nucleus), and GTP hydrolysis (in the cytoplasm). Imbalances between these two forms are thought to specify transport direction. RanBP1 triggers Ran to hydrolyze GTP by acting as a cofactor for the Ran GTPase-activating protein, RanGAP1. RanBP1 also stabilizes the interaction between Ran and importin β/kariopherin β [32] and, in the absence of bound nucleotide, can form a stable heterotrimeric complex with Ran and the Ran-specific GEF RCC1 [33]. RanBP1s are 200–220 amino acid proteins presenting an extremely conserved domain of approximately 150 amino acids including the WH1 domain identified here which is also the minimal Ran-binding domain [34]. The exact ligand-binding site within the Ran sequence has not yet been determined, but the acidic C-terminal domain of Ran is necessary for efficient binding, suggesting that the C-terminal may also be involved in the interaction [35]. Some nuclear pore complex proteins are Ran-binding proteins and they present one or more WH1 domains. The human NPC protein NUP358, also termed RanBP2, is a multiple domain protein located at the cytoplasmic fibrils of the NPC. It contains four WH1 domains and numerous FXFG motifs typical for a subgroup of nucleoporins, as well as a cyclophilin A homologous domain which could function as a peptidyl-prolyl *cis-trans* isomerase that accelerates protein folding [36]. Bovine cyclophilin and human BS-63 share very high similarity with NUP358, suggesting that they are also NPC proteins. Another protein from the NPC, the yeast NUP2, also presents one WH1 domain [37]. Finally, two genome project protein sequences, *C. elegans* F59A2.1 and *S. cerevisiae* YIG3, contain two and one copies of the WH1 domain, respectively. The YIG3 sequence contains features that are typical of proteins of the nucleopore complex whereas F59A2.1 is a putative microbody protein by virtue of the presence of a microbody C-terminal targeting signal [38]. Interestingly, a nuclear phosphoprotein, Hba1p, conferring brefeldin A resistance in the yeast *Schizosaccharomyces pombe*, also includes a WH1 domain [39].

Two sequences presenting features that are typical of proteins related to the actin/cytoskeleton network were also detected in this study. The AE33 gene was identified in *Drosophila* [40] and may play a role during photoreceptor cells development of *Drosophila*. The deduced amino acid sequence includes a N-terminal WH1 domain associated with two polyproline-containing domains, a cysteine-rich domain and a potential transmembrane domain. The *C. elegans* genome project sequence YKC2 consists of two WH1 domains and a carboxy-terminal region similar to WASP's C-terminus.

Many proteins containing WH1 domains have only recently been described, hence little is known relating to their functions. Otherwise numerous WH1 domains belong to proteins implicated in a diverse range of signalling, nuclear transport and cytoskeletal events. Six proteins are known to bind to WH1 domains: the actin cytoskeleton-related proteins ActA, vinculin, zyxin (VASP) and the WIP protein (WASP); the small GTPase Ran (RanBP1); the cytoplasmic domain of metabotropic glutamate receptors (Homer). The polyproline region of ActA and the C-terminal 10 amino acids of metabotropic glutamate receptors have been demonstrated to be the binding partner for the WH1 domain, but in most cases ligand-interactive domains have not yet been characterized. Otherwise the WH1 domain of N-WASP is responsible for binding the lipid PIP2 leading to the proposition that, as for some PH domains, the WH1 domain could be a multifaceted binding domain responsible for a number of different interactions including lipids as well as proteins. The increasing recognition of the WH1 domain in intracellular proteins, presenting one or multiple copies and frequently associated with other interactive domains, suggests that it is an important member of the growing family of intracellular modules.

4. Note added during revision

After completion of this study, a preliminary entry of the PROSITE profile database (PS50229) was found to report independently on the WH1/RanBP1 similarity.

Acknowledgements: We wish to thank Gordon Langsley, Jean-Antoine Girault and Jean-Paul Mornon for their helpful suggestions regarding the manuscript. This work received finacial support from ARC (CTG223).

References

[1] Reinhard, M., Halbrugge, M., Scheer, U., Wiegand, C., Jockusch, B.M. and Walter, U. (1992) EMBO J. 11, 2063–2070.
[2] Haffner, C., Jarchau, T., Reinhard, M., Hoppe, J., Lohmann, S.M. and Walter, U. (1995) EMBO J. 14, 19–27.
[3] Gertler, F.B., Niebuhr, K., Reinhard, M., Wehland, J. and Soriano, P. (1996) Cell 87, 227–239.
[4] Niebuhr, K. et al. (1997) EMBO J. 16, 5433–5444.
[5] Symons, M., Derry, J.M., Karlak, B., Jiang, S., Lemahieu, V., Mccormick, F., Francke, U. and Abo, A. (1996) Cell 84, 723–734.

[6] Ramesh, N., Anton, I.M., Hartwig, J.H. and Geha, R.S. (1997) Proc. Natl. Acad. Sci. USA 94, 14671–14676.
[7] Miki, H., Miura, K. and Takenawa, T. (1996) EMBO J. 15, 5326–5335.
[8] Ponting, C.P. and Phillips, C. (1997) J. Mol. Med. 75, 769–771.
[9] Rechsteiner, M. and Rogers, S.W. (1996) Trends Biochem. Sci. 21, 267–271.
[10] Brakeman, P.R., Lanahan, A.A., O'Brien, R., Roche, K., Barnes, C.A., Huganir, R.L. and Worley, P.F. (1997) Nature 386, 284–288.
[11] Gribskov, M. and Veretnik, S. (1996) Methods Enzymol. 266, 198–212.
[12] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[13] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[14] Wootton, J.C. (1994) Comput. Chem. 18, 269–285.
[15] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[16] Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) Proteins 9, 180–190.
[17] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149–155.
[18] Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J.P. (1997) Cell. Mol. Life Sci. 53, 621–645.
[19] Callebaut, I. and Mornon, J.P. (1997) FEBS Lett. 400, 25–30.
[20] Rost, B. (1996) Methods Enzymol. 266, 525–539.
[21] Lin, D. and Pawson, T. (1997) Trends Cell Biol. 7 (8), centerfold.
[22] Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. and Saraste, M. (1992) Nature 359, 851–855.
[23] Mahoney, N.M., Janmey, P.A. and Almo, S.C. (1997) Nature Struct. Biol. 4, 953–960.
[24] Schindelhauer, D., Weiss, M., Hellebrand, H., Golla, A., Hergersberg, M., Seger, R., Belohradsky, B.H. and Meindl, A. (1996) Hum. Genet. 98, 68–76.
[25] Schwarz, K. (1996) Immunol. Today 17, 496–502.
[26] Miki, H. and Takenawa, T. (1998) Biochem. Biophys. Res. Commun. 243, 73–78.
[27] Ahern-Djamali, S.M., Comer, A.R., Bachmann, C., Kastenmeier, A.S., Reddy, S.K., Beckerle, M.C., Walter, U. and Hoffmann, F.M. (1998) Mol. Biol. Cell 9, 2157–2171.
[28] Kato, A., Ozawa, F., Saitoh, Y., Hirai, K. and Inokuchi, K. (1997) FEBS Lett. 412, 183–189.
[29] Sone, M. et al. (1997) Science 275, 543–547.
[30] Ohno, M., Fornerod, M. and Mattaj, I.W. (1998) Cell 92, 327–336.
[31] Dahlberg, J.E. and Lund, E. (1998) Curr. Opin. Cell Biol. 10, 400–408.
[32] Lounsbury, K.M. and Macara, I.G. (1997) J. Biol. Chem. 272, 551–555.
[33] Bischoff, F.R., Krebber, H., Smirnova, E., Dong, W. and Ponstingl, H. (1995) EMBO J. 14, 705–715.
[34] Beddow, A.L., Richards, S.A., Orem, N.R. and Macara, I.G. (1995) Proc. Natl. Acad. Sci. USA 92, 3328–3332.
[35] Richards, S.A., Lounsbury, K.M. and Macara, I.G. (1995) J. Biol. Chem. 270, 14405–14411.
[36] Wu, J., Matunis, M.J., Kraemer, D., Blobel, G. and Coutavas, E. (1995) J. Biol. Chem. 270, 14209–14213.
[37] Loeb, J.D., Davis, L.I. and Fink, G.R. (1993) Mol. Biol. Cell 4, 209–222.
[38] Bairoch, A., Bucher, P. and Hofmann, K. (1997) Nucleic Acids Res. 25, 217–221.
[39] Turi, T.G., Mueller, U.W., Sazer, S. and Rose, J.K. (1996) J. Biol. Chem. 271, 9166–9171.
[40] DeMille, M.M., Kimmel, B.E. and Rubin, G.M. (1996) Gene 183, 103–108.