Protein Science (1999), *8*:435–438. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Identification of a novel domain shared by putative components of the endocytic and cytoskeletal machinery

BRIAN K. KAY,[1] MONTAROP YAMABHAI,[1] BEVERLY WENDLAND,[2] AND SCOTT D. EMR[3]

[1]Department of Pharmacology, University of Wisconsin-Madison, Madison, Wisconsin 53706-1532
[2]Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218
[3]Division of Cellular and Molecular Medicine and Howard Hughes Medical Institute, University of California, San Diego, School of Medicine, La Jolla, California 92093-0668

(RECEIVED August 18, 1998; ACCEPTED October 22, 1998)

Abstract: We have identified a ~140 amino acid domain that is shared by a variety of proteins in budding and fission yeast, nematode, rat, mouse, frog, oat, and man. Typically, this domain is located within 20 residues of the N-terminus of the various proteins. The percent identity among the domains in the 12 proteins ranges from 42 to 93%, with 16 absolutely conserved residues: **N**-$x_{11–13}$-**V**-x_2-**A**-**T**-$x_{34–36}$-**R**-$x_{7–8}$-**W**-**R**-x_3-**K**-x_{12}-**G**-x-**E**-x_{15}-**L**-$x_{11–12}$-**D**-x-**G**-**R**-x_{11}-**D**-x_7-**R**. Even though these proteins share little beyond their segment of homology, data are emerging that several of the proteins are involved in endocytosis and or regulation of cytoskeletal organization. We have named this protein segment the ENTH domain, for Epsin N-terminal Homology domain, and hypothesize that it is a candidate for binding specific ligands and/or enzymatic activity in the cell.

Keywords: Af10; clathrin; cytoskeleton; Dap160; DPF; DPW; EH domains; endocytosis; Eps15; Epsin; homology; intersectin; MP90; NPF; Pan1; protein–protein interaction

The ENTH domain is a highly conserved amino terminal domain of ~140 amino acids that is found in a variety of proteins from numerous species. ENTH domain-containing proteins were first identified in several laboratories because their divergent carboxy-termini interact with Eps15 Homology (EH) domains (Chen et al., 1998; Yamabhai et al., 1998; Wendland & Emr, 1998). The EH domain has recently been described as a protein interaction module contained in proteins involved in endocytosis (Di Fiore et al., 1997) and regulation of the actin cytoskeleton (Wendland et al., 1998). To understand its function in yeast, EH domain-binding proteins have been isolated by two-hybrid screening with the EH domains from the yeast protein, Pan1p, an essential protein for normal organization of the actin cytoskeleton (Tang & Cai, 1996) and endocytosis (Wendland et al., 1996). The screen yielded the yeast homolog of AP180, a clathrin-assembly protein, and three uncharacterized open reading frames (ORFs) as candidate interacting proteins (Wendland & Emr, 1998). Not only did the yeast AP180 and ORF YDL161w proteins carry multiple copies of the putative EH ligand motif, asparagine-proline-phenylalanine (NPF), in their C-terminal regions (Salcini et al., 1997), but their N-termini were similar. A computer search of the yeast genome revealed two other proteins, ORFs YLR206w and YJR125c, with N-terminal sequences, which strongly resembled those carried by the YDL161w protein.

Simultaneous to the work in yeast, two other proteins bearing similar N-terminal sequences have been discovered in mammals. These proteins were isolated from rat and mouse through affinity purification (Chen et al., 1998) and cDNA expression screening with different EH domains (Yamabhai et al., 1998), respectively. The rat protein, termed Epsin, binds to the EH domains of Eps15, a substrate for the epidermal growth factor tyrosine kinase with three EH domains. Eps15 is involved in clathrin-mediated endocytosis (Fazioli et al., 1993; Tebar et al., 1996; Carbone et al., 1997; Benmerah et al., 1998), and has been observed localized in nerve terminals where clathrin-mediated endocytosis of synaptic vesicles occurs (Chen et al., 1998). The mouse protein, Intersectin binding protein 2 (Ibp2), binds to Intersectin, a novel protein that was identified in *Xenopus laevis* that contains two EH domains and five Src Homology (SH) domains in its N- and C-termini, respectively (Yamabhai et al., 1998). cDNAs encoding a human homolog of Intersectin (Guipponi et al., 1998) and a related protein in *Drosophila melanogaster* (Roos & Kelly, 1998), termed Dynamin associated protein of 160,000 molecular mass (Dap160), have been cloned recently.

Careful comparison of the primary structures of the proteins that interact with the EH domains of Pan1p, Eps15, and Intersectin, revealed that they share a ~140 amino acid conserved segment. This segment, first noted in a plant (Oat) protein (Jones & Hooley, 1997), occurs in 12 proteins currently in GenBank, in such diverse genomes as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Rattus norvegicus*, *Mus musculus*, *Xenopus laevis*, *Avena fatua*, and *Homo sapiens*. The alignment of this

Reprint requests to: Brian K. Kay, University of Wisconsin-Madison, 1300 University Avenue, Madison, Wisconsin; e-mail: bkkay@facstaff.wisc.edu.

novel domain is shown in Figure 1. The primary structural identity among the sequences ranges between 42 and 93%, with 16 absolutely conserved residues, **N**-x_{11-13}-**V**-x_2-**A**-**T**-x_{34-36}-**R**-x_{7-8}-**W**-**R**-x_3-**K**-x_{12}-**G**-x-**E**-x_{15}-**L**-x_{11-12}-**D**-x-**G**-**R**-x_{11}-**D**-x_7-**R**. Interestingly, this conserved segment is present at or near the N-terminus of most proteins, even though their C-terminal sequences vary. Computer algorithms suggest that the domain is mostly α-helical in secondary structure and lacks any transmembrane spanning regions. We have named this conserved protein segment the "ENTH" domain, for Epsin N-Terminal Homology domain, after Epsin, 1 of the 12 proteins.

In GenBank searches, the only characterized protein with a region that weakly matches (22% similarity) the ENTH domain is the clathrin assembly protein, AP180 (also known as AP-3, NP185, pp155, F1-20). Interestingly, 9 of the 16 residues absolutely conserved in the ENTH domain are also conserved between the mouse,

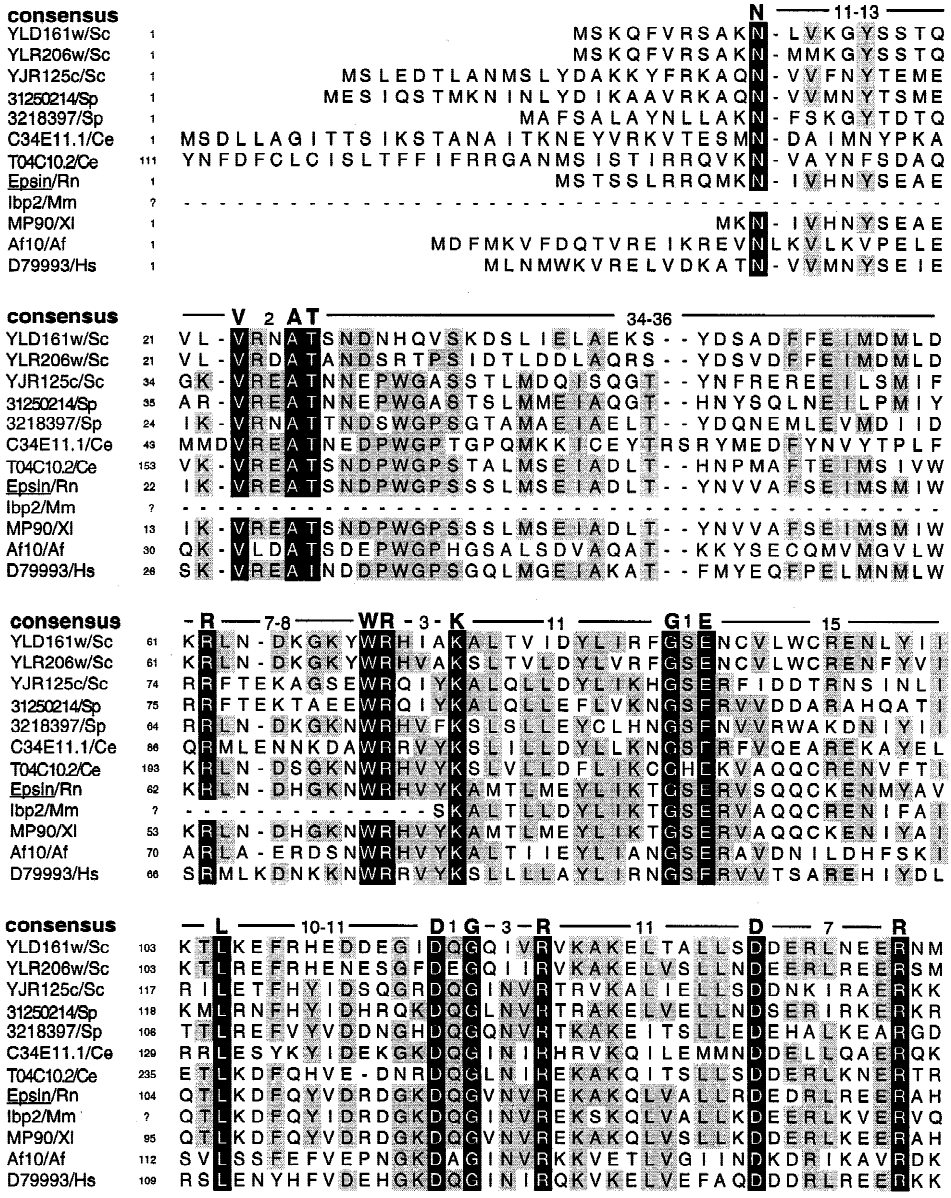

Fig. 1. Alignment of the primary structures of the ENTH domains in 11 proteins. The domains were aligned with ClustalW (http://www2.ebi.ac.uk/clustalw/) and the figure was generated with SeqVu (v1.1). The amino acid positions are numbered on the left. Gaps have been introduced in some of the sequence to maximize their alignment. Identical residues shared by ≥50% of the 12 entries have been highlighted; residues shared by all the entries are highlighted in black. The absolutely conserved residues, along with the number of intervening amino acids, are shown in the consensus. All sequences are full-length, except for Ibp2 which is missing its N-terminus. Sc, Sp, Ce, Rn, Mm, Xl, Af, and Hs abbreviations correspond to *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Rattus norvegicus*, *Mus musculus*, *Xenopus laevis*, *Avena fatua*, and *Homo sapiens*, respectively. The GenBank accession numbers are listed on the left column, except for Epsin, Ibp2, MP90, and Af10 which are 3249559, 3063649, 2072301, and 1724114, respectively. Further descriptions regarding some of the proteins, YLD161w (Wendland et al., 1998), MP90 (Stukenberg et al., 1997), Epsin (Chen et al., 1998), Ibp2 (Yamabhai et al., 1998), and Af10 (Jones & Hooley, 1997), can be found elsewhere.

human, and yeast AP180 proteins (Wendland & Emr, 1998); more specifically, these are N-$x_{11–13}$-V-x_2-A-T-$x_{34–36}$-R-$x_{7–8}$-W-R-x_3-K-x_{12}-G-x-E-x_{15}-L-$x_{11–12}$-D-x-G-R-x_{11}-D-x_7-R (matching residues are underlined). This region of similarity is contained within a larger segment (N-terminal 33 kDa) of AP180 previously shown to bind polyphosphoinositides and inhibit clathrin assembly (Ye & Lafer, 1995; Hao et al., 1997). This segment includes a motif implicated in the binding of phosphatidylinositol 3,4,5 trisphosphate by centaurin-α (Hammonds-Odie et al., 1996). While this motif falls within the region of AP180 similar to the ENTH domain, the consensus residues for lipid binding are not conserved in the ENTH domain.

Figure 2 diagrams the relative locations of the ENTH domain and other short motifs within the 12 proteins. Based on the biochemical and genetic characterization of only a few of the proteins, we postulate that these proteins are likely to function in endocytosis and/or regulation of the actin cytoskeleton. Ten of the 12 proteins contain sequences related to the clathrin-binding motifs described in arrestin (Krupnick et al., 1997), adaptor protein AP-3 (Dell'Angelica et al., 1998), and amphiphysin II (Ramjaun & McPherson, 1998). The proteins share either the motif L(L,I,V)(D,E,G,N)(L,F)(D,E,Q) or the related sequence, LIDL-COO^-. Interestingly, the rat Epsin protein, which carries the motif LVDLD, has recently been reported to bind clathrin (Chen et al., 1998). Many of the 12 proteins also carry the motif asparagine-proline-phenylalanine (NPF) in their C-terminal regions. This motif appears to be the ligand for the EH domains of Eps15 (Salcini et al., 1997), Pan1p (Wendland & Emr, 1998), and Intersectin (Yamabhai et al., 1998). Finally, many of the proteins also carry multiple copies of the tripeptide sequences DPF or DPW, even within the ENTH domain itself (i.e., C34E11.1, T04C10.2, MP90, epsin, D79993). At present, it is unclear what function(s) these tripeptides serve.

In conclusion, we propose that the ENTH domain is an important domain within proteins involved in endocytosis and the cytoskeleton. The amino acid conservation among the ENTH domains in proteins of fungal, plant, and animal origin strongly suggests that this domain is functionally important. Future experiments will determine whether or not the ENTH domain has ligand binding or enzymatic activity in the cell.

References

Benmerah A, Lamaze C, Begue B, Schmid SL, Dautry-Varsat A, Cerf-Bensussan N. 1998. AP-2/Eps15 interaction is required for receptor-mediated endocytosis. *J Cell Biol 140*:1055–1062.

Carbone R, Fre S, Iannolo G, Belleudi F, Mancini P, Pelicci PG, Torrisi MR, Di Fiore PP. 1997. eps15 and eps15R are essential components of the endocytic pathway. *Cancer Res 57*:5498–5504.

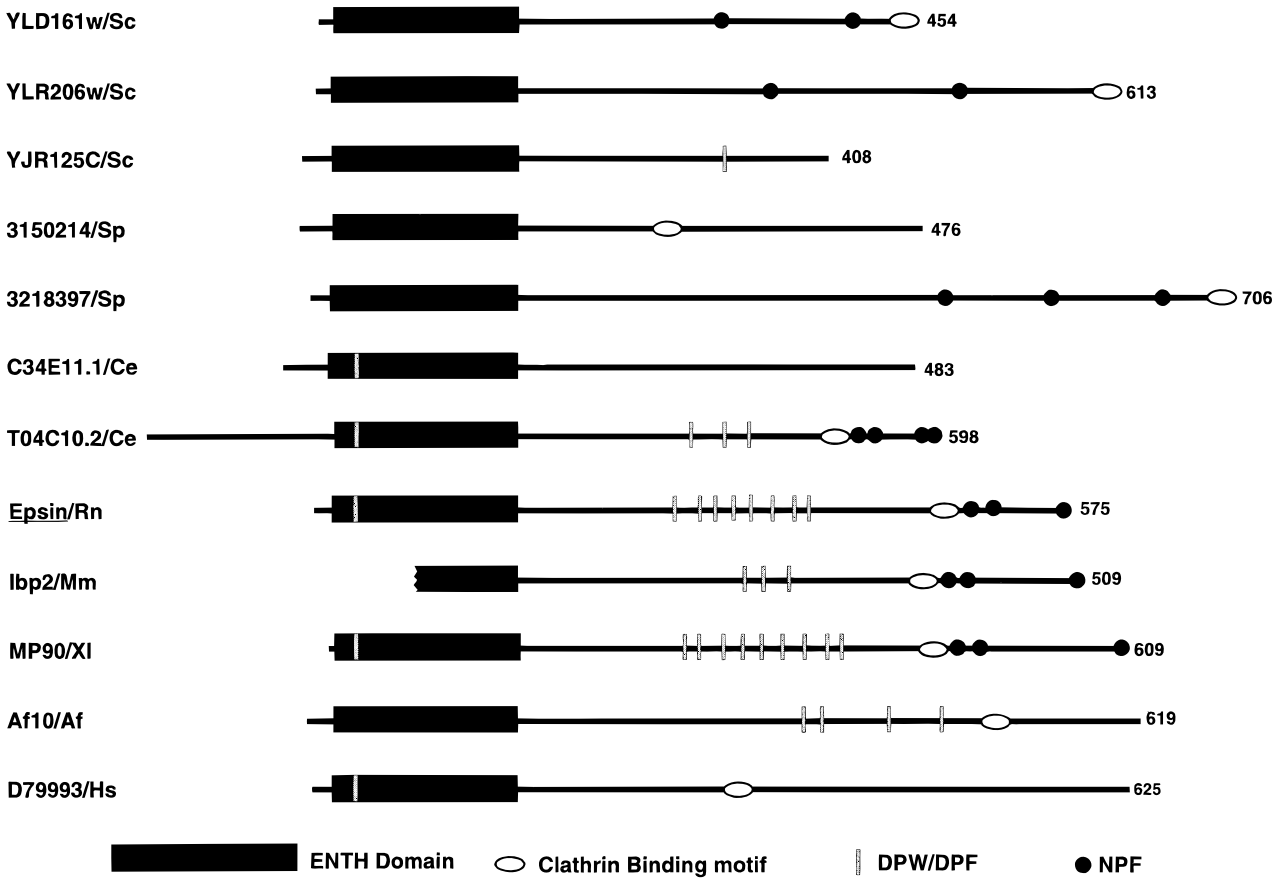

Fig. 2. Diagram of the proteins described in Figure 1. The ENTH domain is shown as a solid box. The proteins are referred to by their GenBank accession name, source, and length in amino acids. Sc, Sp, Ce, Rn, Mm, Xl, Af, and Hs abbreviations correspond to *S. cerevisiae*, *S. pombe*, *C. elegans*, *R. norvegicus*, *M. musculus*, *X. laevis*, *A. fatua*, and *H. sapiens*, respectively. The DPW/DPF, clathrin-binding, and NPF motifs are denoted by shaded bars, hollow ovals, and filled circles, respectively.

Chen H, Fre S, Slepnev V, Capua M, Takei K, Butler M, Di Fiore P, De Camilli P. 1998. Epsin, an EH domain binding protein implicated in clathrin-mediated endocytosis. *Nature (London) 394*:793–797.

Dell'Angelica E, Klumperman J, Stoorvogel W, Bonifacino J. 1998. Association of the AP-3 adaptor complex with clathrin. *Science 280*:431–434.

Di Fiore PP, Pelicci PG, Sorkin A. 1997. EH: A novel protein-protein interaction domain potentially involved in intracellular sorting. *Trends Biochem Sci 22*:411–413.

Fazioli F, Minichiello L, Matoskova B, Wong WT, Di Fiore PP. 1993. eps15, a novel tyrosine kinase substrate, exhibits transforming activity. *Mol Cell Biol 13*:5814–5828.

Guipponi M, Scott HS, Chen H, Schebesta A, Rossier C, Antonarakis SE. 1998. Two isoforms of the human intersectin protein are produced by brain-alternative splicing in a stop codon. *Genomics 53*:369–376.

Hammonds-Odie LP, Jackson TR, Profit AA, Blader IJ, Turck CW, Prestwich GD, Theibert AB. 1996. Identification and cloning of centaurin-alpha. A novel phosphatidylinositol 3,4,5-trisphosphate-binding protein from rat brain. *J Biol Chem 271*:18859–18868.

Hao W, Tan Z, Prasad K, Reddy KK, Chen J, Prestwich GD, Falck JR, Shears SB, Lafer EM. 1997. Regulation of AP-3 function by inositides. Identification of phosphatidylinositol 3,4,5-trisphosphate as a potent ligand. *J Biol Chem 272*:6393–6398.

Jones H, Hooley R. 1997. A full-length cDNA, Af10 (accession No. U80041), isolated from wild oat (*Avena fatua*) aleurone encodes a protein with similarity to clathrin assembly proteins (PGR97-018). *Plant Physiol 113*:663.

Krupnick JG, Goodman OB Jr, Keen JH, Benovic JL. 1997. Arrestin/clathrin interaction. Localization of the clathrin binding domain of nonvisual arrestins to the carboxy terminus. *J Biol Chem 272*:15011–15016.

Ramjaun AR, McPherson PS. 1998. Multiple amphiphysin II splice variants display differential clathrin binding: Identification of two distinct clathrin-binding sites. *J Neurochem 70*:2369–2376.

Roos J, Kelly RB. 1998. Dap160, a neural-specific eps15 homology and multiple SH3 domain-containing protein that interacts with *Drosophila* dynamin. *J Biol Chem 273*:19108–19119.

Salcini AE, Confalonieri S, Doria M, Santolini E, Tassi E, Minenkova O, Cesareni G, Pelicci PG, Di Fiore PP. 1997. Binding specificity and in vivo targets of the EH domain, a novel protein-protein interaction module. *Genes Dev 11*: 2239–2249.

Stukenberg PT, Lustig KD, McGarry TJ, King RW, Kuang J, Kirschner MW. 1997. Systematic identification of mitotic phosphoproteins. *Curr Biol 7*:338-348.

Tang HY, Cai M. 1996. The EH-domain-containing protein Pan1 is required for normal organization of the actin cytoskeleton in *Saccharomyces cerevisiae*. *Mol Cell Biol 16*:4897–4914.

Tebar F, Sorkin T, Sorkin A, Ericsson M, Kirchausen T. 1996. Eps15 is a component of clathrin-coated pits and vesicles and is located at the rim of coated pits. *J Biol Chem 271*:28727–28730.

Wendland B, Emr SD. 1998. Pan1p, yeast eps15, functions as a multivalent adaptor that coordinates protein-protein interactions essential for endocytosis. *J Cell Biol 141*:71–84.

Wendland B, Emr S, Reizman H. 1998. Protein traffic in the yeast endocytic and vacuolar protein sorting pathways. *Curr Opin Cell Biol 10*:513–522.

Wendland B, McCaffery JM, Xiao Q, Emr SD. 1996. A novel fluorescence-activated cell sorter-based screen for yeast endocytosis mutants identifies a yeast homologue of mammalian eps15. *J Cell Biol 135*:1485–1500.

Yamabhai M, Hoffman NG, Hardison NL, McPherson PS, Castagnoli L, Cesareni G, Kay BK. 1998. Intersectin, a novel adaptor protein with two EH and five SH3 domains. *J Biol Chem 273*:31401–31407.

Ye W, Lafer EM. 1995. Clathrin binding and assembly activities of expressed domains of the synapse-specific clathrin assembly protein AP-3. *J Biol Chem 270*:10933–10939.