

A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution

WILLIAM T. WONG*, CHRISTOPH SCHUMACHER†, ANNA ELISABETTA SALCINI*‡, ALFREDO ROMANO‡§, PAOLA CASTAGNINO*, PIER GIUSEPPE PELICCI‡, AND PIER PAOLO DI FIORE*‡¶

*Laboratory of Cellular and Molecular Biology, and §Laboratory of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; †The Rockefeller University, New York, NY 10021-6399; and ‡IEO, European Institute of Oncology, 20141 Milan, Italy

Communicated by Stanley Cohen, Vanderbilt University, Nashville, TN, June 22, 1995

ABSTRACT In this report we structurally and functionally define a binding domain that is involved in protein association and that we have designated EH (for Eps15 homology domain). This domain was identified in the tyrosine kinase substrate Eps15 on the basis of regional conservation with several heterogeneous proteins of yeast and nematode. The EH domain spans about 70 amino acids and shows ≈60% overall amino acid conservation. We demonstrated the ability of the EH domain to specifically bind cytosolic proteins in normal and malignant cells of mesenchymal, epithelial, and hematopoietic origin. These observations prompted our search for additional EH-containing proteins in mammalian cells. Using an EH domain-specific probe derived from the *eps15* cDNA, we cloned and characterized a cDNA encoding an EH-containing protein with overall similarity to Eps15; we designated this protein Eps15r (for Eps15-related). Structural comparison of Eps15 and Eps15r defines a family of signal transducers possessing extensive networking abilities including EH-mediated binding and association with Src homology 3-containing proteins.

Eps15 is a recently identified substrate for the epidermal growth factor receptor (EGFR) and other receptor tyrosine kinases (1, 2). The predicted amino acid sequence of Eps15 identifies a modular protein with three domains (refs. 1 and 2 and Fig. 1A). Domain I is ≈300 aa long and composed of three nonidentical repeats of ≈100 aa each. Two tyrosines, at positions 19 and 132, within this domain possess features of candidate sites for phosphorylation (3–5). In addition, domain I contains EF-hand-type calcium-binding domains (6). Domain II spans the central region of the protein and presents the characteristic heptad repeats of coiled-coil proteins (7). Domain III, located at the carboxyl terminus of Eps15, displays a proline-rich region and a repeated aspartic acid-proline-phenylalanine (DPF) motif. The proline-rich region has been shown to bind to the Src homology 3 (SH3) domain of Crk (24). A motif containing DPF is conserved in several methyltransferases possessing different specificity (8) and has been proposed to represent the recognition sequence for *S*-adenosylmethionine, the methyl donor in the methyltransferase reaction (8).

The product of the *eps15* gene is a 140- to 150-kDa protein with predominant cytosolic localization (1). Eps15 can be directly phosphorylated *in vitro* by EGFR and is tyrosine phosphorylated at high stoichiometry *in vivo* following activation of the EGFR (1). There is evidence that perturbation of Eps15 function affects cell proliferation. Overexpression of *eps15* can transform NIH 3T3 cells, albeit with low efficiency (1). Furthermore, the human *eps15* gene, which maps to chromosome 1 band p32 (2), is rearranged with the HRX/ALL-1/MLL gene (9, 10) in the t(11;11)(p32;q23)

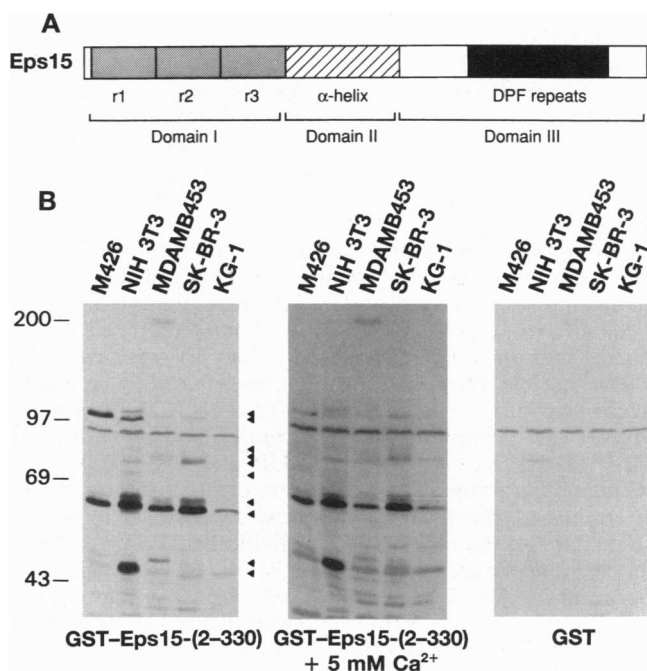


FIG. 1. Binding of GST fusion proteins containing various domains of Eps15. (A) Schematic of the structure of Eps15, highlighting the structural features of domains I–III. The three nonidentical repeats of domain I are designated r1–r3. The portions of Eps15 present in the various GST fusion proteins are indicated with brackets. (B) Far-Western assay of total cellular proteins (100 μg) from the indicated cell lines. Proteins were detected with GST-Eps15-(2–330) (Left), GST-Eps15-(2–330) in the presence of 5 mM CaCl₂ (Center) or control GST (Right) followed by affinity-purified anti-GST antibody coupled to ¹²⁵I-protein A. Molecular size markers are indicated on the left in kilodaltons. Bands specifically recognized by GST-Eps15-(2–330) are indicated by arrowheads.

translocation in acute myeloid leukemias (11). The der(11) of this translocation, which is most likely the biologically active fusion gene (12), is predicted to encode an HRX-Eps15 chimeric protein containing the amino-terminal portion of HRX and the entire Eps15 except for the first 11 aa (11).

There is no immediate indication of the physiological function of Eps15, nor of its role in neoplastic transformation. No enzymatic activity has been identified for the protein, nor does it contain any established motif of signal transducers, such as SH2 or SH3 domains (reviewed in ref. 13). The present studies were undertaken to gain insight into the molecular interactions of Eps15 and its function.

MATERIALS AND METHODS

Cell Lines. NIH 3T3 mouse fibroblasts were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (vol/vol) calf serum. Human tumor cell lines of epithelial and mesodermal derivation were maintained in DMEM supplemented with 10% fetal bovine serum. Human tumor cell lines of hematopoietic derivation were propagated in RPMI 1640 plus 10% fetal bovine serum.

Production of Recombinant Proteins and *in Vitro* Binding Studies. Glutathione S-transferase (GST)-Eps15 fusion proteins were obtained by recombinant PCR of the appropriate fragments from the Eps15 cDNA followed by cloning in the pGEX-2T expression vector (Pharmacia), in frame with the GST moiety. Purification of the fusion proteins on glutathione agarose has been described (1).

For the far-Western experiments, total cellular proteins (100 μ g) prepared as described (1) were fractionated by SDS/PAGE and transferred onto poly(vinylidene difluoride) membranes (Immobilon; Millipore). Blots were blocked with 2% (wt/vol) bovine serum albumin (BSA) in TTBS (20 mM Tris-HCl, pH 7.5/150 mM NaCl/0.05% Tween 20) for at least 2 hr at room temperature and then with 3 μ M reduced glutathione (Sigma) in TTBS with 0.5% BSA for 1 hr at room temperature. Blots were then incubated for 1 hr at room temperature with the appropriate GST fusion proteins (10 nM) in TTBS containing 3 μ M reduced glutathione and 0.5% BSA. Fusion proteins were detected with affinity-purified anti-GST antibody coupled to 125 I-protein A, as described (1).

For subcellular fractionation, NIH 3T3 cells were either harvested during logarithmic growth in the presence of serum (growing) or serum-starved for 48 hr (resting). Subcellular fractions (1) were assayed for the presence of lactate dehydrogenase (cytosolic marker) and alkaline phosphodiesterase (membrane marker) to assess purity (1).

Isolation and Characterization of the *eps15r* cDNA. A mouse keratinocyte cDNA library was screened with a 987-bp fragment of the *eps15* cDNA that was obtained by PCR and corresponded to the region encoding aa 2-330 (domain I of Eps15). Hybridization was performed under conditions of reduced stringency, in 5 \times standard saline citrate (SSC)/35% formamide at 42°C and was followed by washing in 0.4 \times SSC at 50°C. Under these conditions a reduction of 10.5°C in the apparent melting temperature of the hybrids is achieved with respect to standard high-stringency conditions (hybridization in 50% formamide, wash in 0.1 \times SSC, leaving the other conditions unaltered) (14). Positive plaques were further screened in duplicate under reduced- and high-stringency conditions. Only those plaques exhibiting signals in reduced-stringency conditions but not in high-stringency conditions were further characterized.

DNA sequence was determined by the dideoxy termination method on both strands of the *eps15r* cDNA with a Sequenase kit (United States Biochemical).^{||} The GenBank and EMBL databases were screened with the BLAST program (15). Signatures of the predicted protein sequence were identified by the PROSITE program (3). Multiple sequence alignments were obtained with the PILEUP program (16).

RESULTS

Binding Properties of Eps15 Domains. To investigate the interactions of Eps15 with other proteins, we engineered GST-based fusion proteins of the three Eps15 domains (Fig. 1A): GST-Eps15-(2-330), containing domain I; GST-Eps15-(321-520), encompassing domain II; and GST-Eps15-(501-874), representing domain III. Total cellular proteins from cell

lines of different embryological derivation (fibroblastic, epithelial, and hematopoietic) were fractionated by SDS/PAGE and analyzed in a far-Western assay using the GST-Eps15 fusions as probes and detecting the interactions with an affinity-purified anti-GST antibody. The GST fusions of domains II and III did not specifically recognize cellular proteins, even when the assay was performed after renaturation of the membrane-bound cellular proteins (data not shown). GST-Eps15-(2-330), however, specifically interacted with several proteins contained in NIH 3T3 and M426 fibroblasts, MDAMB453 and SK-BR-3 human mammary carcinoma cells, and KG-1 acute myeloblastic leukemia cell lines (Fig. 1B Left). In addition, agarose-immobilized GST-Eps15-(2-330) specifically recognized, in 35 S-labeled lysates, proteins similar in size to the proteins detected in the far-Western assay (data not shown). A control GST protein did not detect these proteins (Fig. 1B Right). Domain I of Eps15 contains two calcium-binding motifs of the EF-hand type (1, 2). The possibility was explored that binding to domain I might be influenced by calcium. However, binding was not significantly affected by addition of 5 mM CaCl₂ to the reaction mixture (Fig. 1B Center).

We next analyzed the subcellular distribution of proteins binding to domain I of Eps15. For this purpose, NIH 3T3 cell lysates were fractionated into cytosolic, nuclear, and membrane fractions. Subcellular fractions were obtained from resting, serum-starved cells or from cells in logarithmic phase of growth (Fig. 2). Each fraction was shown to be >99% pure, by assaying lactate dehydrogenase (cytosolic marker) and alkaline phosphodiesterase (membrane marker) enzymatic activities (data not shown). Aliquots of each fraction, representative of the same number of cells, were then analyzed in far-Western assay with the GST-Eps15-(2-330), in comparison with an aliquot of total cellular proteins. Most of the domain I-interacting proteins were in the cytosolic fraction (Fig. 2), consistent with the predominant cytoplasmic localization of Eps15 (1). In addition, no difference in the pattern of domain I-binding proteins was detectable between resting and growing cells (Fig. 2).

Identification of a Binding Domain Conserved in Evolution.

Domain I of Eps15 is entirely composed of three nonidentical repeats of \approx 100 aa each. The binding properties of domain I must, therefore, reside in the individual repeats. Using the BLAST program (15), we searched for homologies to the

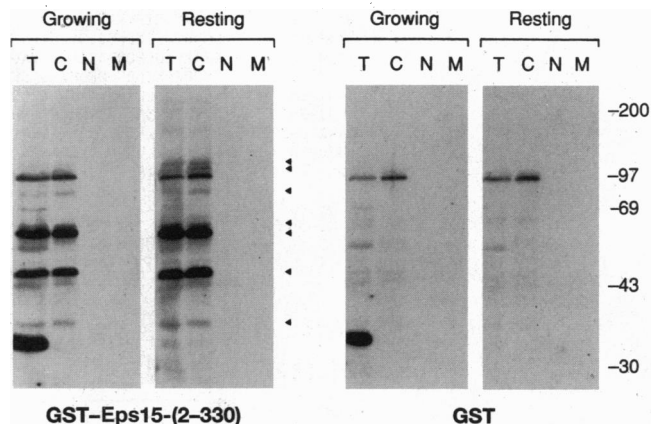


FIG. 2. Subcellular fractionation of proteins binding to domain I of Eps15. NIH 3T3 cells were either serum-starved for 48 hr (Resting) or grown in the continuous presence of 10% calf serum (Growing). Aliquots of each subcellular fractions (cytosolic, C; nuclear, N; and membrane, M) or of the total cellular lysate (T), representative of the same number of cells (5×10^5), were then analyzed by far-Western blotting with either GST-Eps15-(2-330) or control GST. Molecular size markers are indicated in kilodaltons. Bands specifically recognized by GST-Eps15-(2-330) are indicated by arrowheads.

^{||}The nucleotide sequence of *eps15r* has been deposited in the GenBank database (accession no. U29156).

individual repeats of domain I of Eps15 in database sequences. Five sequences were detected, including PAN1, a poly(A)-binding protein (PAB)-dependent poly(A) ribonuclease identified in yeast (17); PAN-1b, a PAN1-related sequence of yeast (Protein Identification Resource database, accession no. S48440); END3, a yeast protein required for internalization (18); YBL0520, a yeast protein of unknown function (19); and YNJ6, a hypothetical protein of *Caenorhabditis elegans* (20). The homology between Eps15 and END3 has already been reported (18). Homology between these proteins (see Fig. 3A) and the repeats of Eps15 identified a structural module shared by mammalian, nematode, and yeast proteins, which we named EH (for Eps15 homology).

Multiple sequence alignment of EH domains was obtained by the PILEUP program (16) and optimized by visual inspection (Fig. 3A). EH domains from a mammalian *eps15*-related gene (*eps15r*, described in the next paragraph) were also incorporated. Conversely, EH domains from PAN1-b were excluded, in order not to bias the derived consensus sequence of EH, since this protein is >98% identical to PAN1. The alignment of the sequences and the EH consensus are shown in Fig. 3A. The domain is ≈ 70 aa long and $\approx 60\%$ of the positions are conserved when a plurality of >50% is used to calculate overall homology (see legend to Fig. 3A for details). Four residues—Leu³⁴, Leu³⁹, Trp⁴³, and Phe⁵⁹—are invariant. Calcium-binding domains of the EF-hand type (6) were found in 5 of 13 aligned EH domains (Fig. 3A), indicating that their presence is not obligatory. EH domains are repeated in some of the proteins: PAN1 and PAN-1b display two EH domains (Fig. 3B), and YBL0520, three (Fig. 3B), while Eps15 contains three EH domains.

Isolation of the *eps15r* cDNA and Structure of the *eps15r* Gene Product. With the exception of Eps15, all other EH-containing sequences originated from yeast or *C. elegans*. Because of the heterogeneity of the EH-containing proteins, we reasoned that also in mammalian cells other EH-containing proteins must exist. We therefore screened a mouse keratinocyte cDNA library with a probe derived from the region encoding domain I of Eps15 (EH-containing region), under

reduced stringency conditions (see *Materials and Methods*). Positive phage plaques were further subjected to hybridization under high- and reduced-stringency conditions. Phages hybridizing at high stringency were eliminated, as they probably represented *eps15* cDNAs. Among the cDNAs hybridizing only under reduced stringency, the longest one (pCEV-*eps15r*) was sequenced and found to contain the entire open reading frame of an *eps15*-related cDNA (Fig. 4). The 3' nucleotide sequence of this clone was identical to a partial cDNA clone isolated from a mouse embryo expression library screened for Crk-SH3-binding proteins which showed homology to *eps15* and was designated *eps15r*, for *eps15*-related (24). We concluded that we had isolated a cDNA containing the entire open reading frame of *eps15r*.

The *eps15r* open reading frame (Fig. 4A) predicts a 907-aa protein with a calculated molecular mass of ≈ 99 kDa and displaying 47% identity with the predicted Eps15 protein (Fig. 4A). Homology between Eps15 and Eps15r is colinear, allowing identification in Eps15r of three domains analogous to those of Eps15 (Fig. 4B). Domain I (aa 15–338) displays 70% identity to Eps15 and contains three EH domains (Figs. 3A and 4). While in Eps15 the three EH domains are contiguous, a 49-aa insertion separates EH2 and EH3 in Eps15r (Fig. 4). In domain II, identity between Eps15r and Eps15 drops to $\approx 45\%$ (Fig. 4A). However, heptad motifs biased for hydrophobic amino acid residues at positions 1 and 4 are present (Fig. 4A), predicting formation of an α -helix (Fig. 4B). In domain III, there is little conservation between Eps15 and Eps15r (Fig. 4A), with two notable exceptions. First, multiple DPF motifs are present in both proteins (Fig. 4A). Second, a proline-rich region (positions 775–790 and 770–786, for Eps15r and Eps15, respectively) is well conserved. This region has been demonstrated to mediate the association of Eps15 and Eps15r with the Crk SH3 domain (24). Thus, Eps15 and Eps15r identify a family of substrates endowed with SH3-binding properties and possessing a previously unknown binding domain, EH.

DISCUSSION

We have identified a protein-binding domain of ≈ 70 aa in the amino-terminal portion of Eps15. This domain, which we have

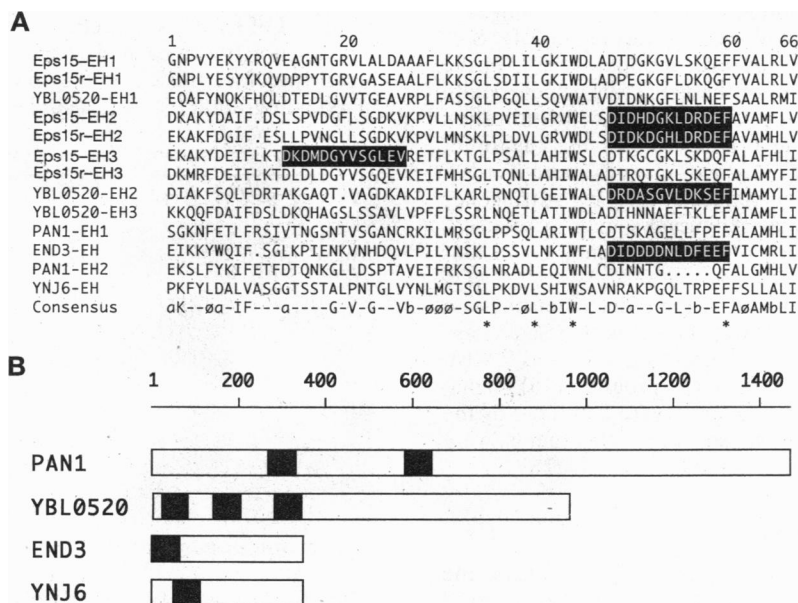


FIG. 3. Identification and structure of the EH domain. (A) Sequence alignment and EH domain consensus. Proteins displaying homology to either of the three repeats of domain I of Eps15 were initially identified by a BLASTP search of protein databases. Sequences displaying scores with $P < 0.01$ were selected for further analysis and proteins whose homology was mainly in the EF-hand type calcium-binding domain were discarded. Multiple sequence alignment was obtained with the PILEUP program. The consensus was determined by a plurality of >50% (7 conserved positions out of 13 sequences) in the alignment. Accepted conservations were as follows: a, E/D; b, K/R/H; ø, F/Y/L/I/V. Asterisks indicate invariant amino acids. Calcium-binding domains of the EF-hand type are shown on a black background. (B) Position of EH domains (solid boxes) in various proteins. Amino acid positions are indicated by the scale.

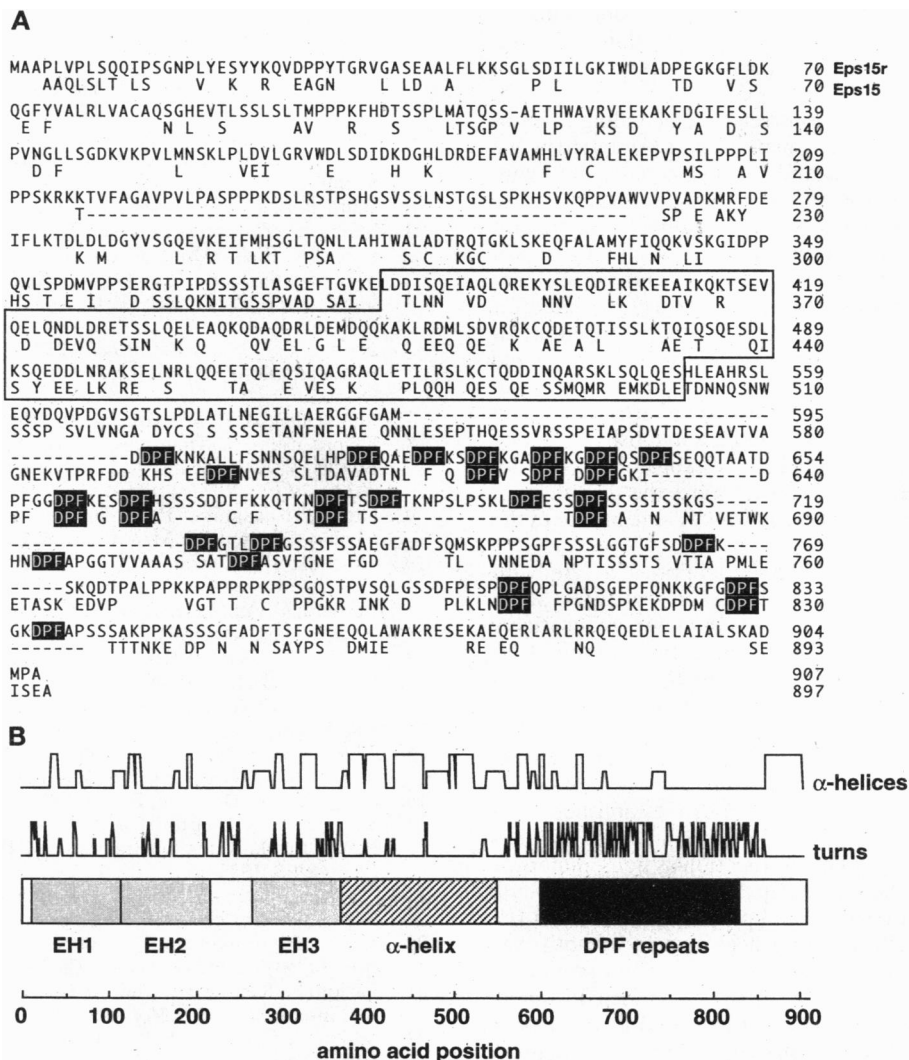


FIG. 4. (A) Comparison of the predicted amino acid sequences of Eps15 and Eps15r. In the Eps15 sequence, only nonidentical amino acids are reported, except for the DPF motifs. Dashes indicate gaps introduced to maximize the alignment. Domain II, containing the heptad repeats is boxed. In domain III, the DPF motifs are highlighted on a black background. (B) Predicted structure of Eps15r. The secondary structure prediction (Chou-Fasman) is aligned with a schematic of Eps15r which shows the three EH domains, the central α -helical region, and the DPF-containing carboxyl-terminal domain.

designated EH, is conserved in a heterogeneous group of proteins from yeast to mammals. The diversity of EH-containing proteins has two implications. First, it suggests that this domain is not likely to be involved in determining intrinsic enzymatic activities of EH-containing proteins. However, it might be involved in establishing protein-protein interactions relevant to the regulation of these proteins or to their signaling abilities. Consistent with this hypothesis, we established a specific protein-binding function associated with the EH domains of Eps15. Second, it implies that a family of EH-containing proteins exists in mammalian cells. Indeed, we isolated a cDNA, *eps15r*, which predicts an EH-containing protein related to Eps15. In addition, relaxed-stringency Southern hybridization with an EH probe indicated the existence of other related genes in human and mouse genomes (data not shown). Thus our results demonstrate that EH is a binding domain conserved during evolution.

EH domains exist in multiple copies in some proteins, suggesting diversified binding abilities and intracellular networking. Structure-function analysis of the PAN1 protein appears to support this contention. PAN1 in yeast is responsible for both mRNA poly(A) shortening and translation initiation; consequently, the gene is essential for cell viability

(17). A PAN1 mutant bearing an internal deletion of aa 95-391, encompassing the first EH domain, was still able to support spore viability (17). In contrast, deletion of the region 391-662, containing the second EH domain, was lethal (17). By analogy, structure-function analysis of the END3 protein supports the idea of an essential role of its EH domain in determining protein function (18). The END3 protein of yeast is required for the internalization step of endocytosis and for actin cytoskeleton organization (18). Therefore *end3* mutants display temperature-sensitive growth defects and inability to internalize the α -factor (18). While the wild-type END3 protein can complement these defects, mutants bearing small deletions in the EH domain lack this property (18).

The involvement of END3 in the endocytotic pathway provides a direct working hypothesis to understand some of the functions of Eps15, including its possible involvement in receptor internalization. In this regard, it is interesting that Eps15 is efficiently phosphorylated by EGFR, but not by the related receptor tyrosine kinase ErbB-2 (1). While EGF is efficiently internalized, an EGFR-ErbB-2 chimera that transduces ErbB-2-specific signals upon EGF stimulation (21) is internalized very slowly upon EGF triggering (22). Thus, the ability to recruit Eps15 as a substrate correlates with the

internalization ability of some receptor tyrosine kinases. Therefore, the hypothesis that Eps15 is involved in endocytosis, possibly through its EH domains, warrants further investigation.

How EH-mediated interactions relate to signaling remains to be established. One intriguing hypothesis is presented by the finding that PAN1 copurifies and is possibly associated with the regulatory subunit of the cAMP-dependent protein kinase (17). Involvement of EH domains in the binding of this subunit might provide an explanation for their presence in a group of proteins with a range of different activities. Our far-Western experiments detected proteins in the size range expected for the regulatory subunit(s) of cAMP-dependent protein kinase—i.e., 50–55 kDa—but we have not yet characterized them further. An expression cloning approach has been used to clone several cDNAs encoding proteins that interact with the EH domain (L. Salcini and P.P.D.F., unpublished results). Their characterization should help us establish how EH domains bind to their targets, what determines the specificity of binding for various EH domains, and which is the nature of the EH-targeted proteins.

The presence of the EH domain identifies a family of molecules with heterogeneous functions and structures. Some characteristics are shared, however. EH domains are invariably present in the amino-terminal portions. In addition, EH-containing proteins display high propensity for α -helical structure in their carboxyl-terminal portions (data not shown). This feature could be directly correlated with the presence of heptad repeats biased for hydrophobic amino acid residues at positions 1 and 4, in most cases (refs. 1 and 17–19; this study). This type of α -helical structure has been established as a dimerization interface (e.g., ref. 23) and its presence suggests that EH-containing proteins might function as homodimers or heterodimers. Within the family, Eps15 and Eps15r identify a distinct subclass with the additional properties of domain III. In this region a PALPPK amino acid sequence motif was shown to bind to Crk (24). In addition, the conservation of DPF motifs implies relevance for function. Thus, Eps15-family molecules show extensive networking potential through their EH domains, SH3-binding regions, and α -helical coiled-coil structures and are likely to act at multiple points in the signaling cascade.

We thank Steven R. Tronick for helping with the computer analysis and Matthias Kraus for critically reviewing the manuscript. This

research was supported in part by a grant from the Associazione Italiana Ricerca sul Cancro.

1. Fazioli, F., Minichiello, L., Matoskova, B., Wong, W. T. & Di Fiore, P. P. (1993) *Mol. Cell. Biol.* **13**, 5814–5828.
2. Wong, W. T., Kraus, M. H., Carlomagno, F., Zelano, A., Druck, T., Croce, C. M., Huebner, K. & Di Fiore, P. P. (1994) *Oncogene* **9**, 1591–1597.
3. Bairoch, A. (1993) *Nucleic Acids Res.* **21**, 3097–3103.
4. Songyang, Z., Carraway, K. L., III, Eck, M. J., Harrison, S. C., Feldman, R. A., Mohammadi, M., Schlessinger, J., Hubbard, S. R., Smith, D. P., Eng, C., Lorenzo, M. J., Ponder, B. A. J., Mayer, B. J. & Cantley, L. C. (1995) *Nature (London)* **373**, 536–539.
5. Hunter, T. (1982) *J. Biol. Chem.* **257**, 4843–4848.
6. Heizmann, C. W. & Hunziker, W. (1991) *Trends Biochem. Sci.* **16**, 98–103.
7. McLachlan, A. D. & Karn, J. (1983) *J. Mol. Biol.* **164**, 605–626.
8. Klimasauskas, S., Timinskas, A., Menkevičius, S., Butkiene, D., Butkus, V. & Janulaitis, A. (1989) *Nucleic Acids Res.* **17**, 9823–9832.
9. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M. & Canaani, E. (1992) *Cell* **71**, 701–708.
10. Tkachuk, D. C., Kohler, S. & Cleary, M. L. (1992) *Cell* **71**, 691–700.
11. Bernard, O. A., Mauchauffe, M., Mecucci, C., Van den Berghe, H. & Berger, R. (1994) *Oncogene* **9**, 1039–1045.
12. Rowley, J. D. (1992) *Genes Chromosomes Cancer* **5**, 264–266.
13. Cohen, G. B., Ren, R. & Baltimore, D. (1995) *Cell* **80**, 237–248.
14. Kraus, M. H. & Aaronson, S. A. (1991) *Methods Enzymol.* **200**, 546–556.
15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
16. Feng, D. F. & Doolittle, R. F. (1987) *J. Mol. Evol.* **25**, 351–360.
17. Sachs, A. B. & Deardorff, J. A. (1992) *Cell* **70**, 961–973.
18. Benedetti, H., Raths, S., Crausaz, F. & Riezman, H. (1994) *Mol. Biol. Cell* **5**, 1023–1037.
19. De Wergifosse, P., Jacques, B., Jonniaux, J. L., Purnelle, B., Skala, J. & Goffeau, A. (1994) *Yeast* **10**, 1489–1496.
20. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., *et al.* (1994) *Nature (London)* **368**, 32–38.
21. Fazioli, F., Kim, U. H., Rhee, S. G., Molloy, C. J., Segatto, O. & Di Fiore, P. P. (1991) *Mol. Cell. Biol.* **11**, 2040–2048.
22. Sorkin, A., Di Fiore, P. P. & Carpenter, G. (1993) *Oncogene* **8**, 3021–3028.
23. Engel, M., Williams, R. W. & Erickson, B. W. (1991) *Biochemistry* **30**, 3161–3169.
24. Schumacher, C., Knudsen, B. S., Ohuchi, T., Di Fiore, P. P., Glassman, R. H. & Hanafusa, H. (1995) *J. Biol. Chem.* **270**, 15341–15347.