Recognition specificity of individual EH domains of mammals and yeast

**Serena Paoluzi, Luisa Castagnoli,
Ilde Lauro, Anna Elisabetta Salcini[1],
Laura Coda[1], Silvia Fre'[1],
Stefano Confalonieri[1],
Pier Giuseppe Pelicci[1,2],
Pier Paolo Di Fiore[1,3] and Gianni Cesareni[4]**

Department of Biology, Enrico Calef, University of Rome Tor Vergata,
Rome 00133, [1]Department of Experimental Oncology,
European Institute of Oncology, Milan 20140, [2]Istituto di Patologia
Speciale Medica, University of Parma, Parma 43100 and
[3]Istituto di Microbiologia, University of Bari, Bari 70100, Italy

[4]Corresponding author
e-mail: cesareni@uniroma2.it

S.Paoluzi and L.Castagnoli contributed equally to this work

The Eps homology (EH) domain is a recently described protein binding module that is found, in multiple or single copies, in several proteins in species as diverse as human and yeast. In this work, we have investigated the molecular details of recognition specificity mediated by this domain family by characterizing the peptide-binding preference of 11 different EH domains from mammal and yeast proteins. Ten of the eleven EH domains could bind at least some peptides containing an Asn-Pro-Phe (NPF) motif. By contrast, the first EH domain of End3p preferentially binds peptides containing an His-Thr/Ser-Phe (HT/SF) motif. Domains that have a low affinity for the majority of NPF peptides reveal some affinity for a third class of peptides that contains two consecutive amino acids with aromatic side chains (FW or WW). This is the case for the third EH domain of Eps15 and for the two N-terminal domains of YBL47c. The consensus sequences derived from the peptides selected from phage-displayed peptide libraries allows for grouping of EH domains into families that are characterized by different NPF-context preference. Finally, comparison of the primary sequence of EH domains with similar or divergent specificity identifies a residue at position +3 following a conserved tryptophan, whose chemical characteristics modulate binding preference.
Keywords: End3/Pan1/phage display/protein binding modules/YBL47c

Introduction

The dynamic assembly of macromolecular complexes inside the cell is often mediated by relatively small protein recognition modules that are widespread in several proteins (Pawson and Scott, 1997). Each module family binds relatively short peptides with different chemical or structural characteristics. For instance, SH2 and PTB bind phosphotyrosine peptides, while SH3 domains require the target peptide to be folded into a proline type II helix. Within families, the recognition specificity of each member rests on subtle chemical variations, on a common structural theme, of the recognition domains that are reflected by complementary variations on the target peptide (Songyang *et al.*, 1993, 1994, 1997; Rickles *et al.*, 1994; Sparks *et al.*, 1994; Dente *et al.*, 1997).

EH is a recently described protein recognition domain that was first identified as a 100 amino acid module repeated three times in the N-terminus of the epidermal growth factor (EGF) receptor substrate Eps15, and of the related protein Eps15R (Fazioli *et al.*, 1993; Wong *et al.*, 1995; Di Fiore *et al.*, 1997). Searching the protein database for similar peptide sequences revealed that EH domains are found in several proteins in species as diverse as yeast and humans. When functional information is available, EH-containing proteins are often associated with regulation of protein transport/sorting and membrane trafficking. Eps15 and Eps15R are both components of clathrin-coated pits, and co-localize with AP2 (Tebar *et al.*, 1996; van Delft *et al.*, 1997; Coda *et al.*, 1998) and with synaptojanin 1 in coated endocytic intermediates in nerve terminals (Haffner *et al.*, 1997). *In vitro*, their C-terminal DPF-rich domain has been shown to be essential for binding to the so-called 'ear' of α-adaptin (Benmerah *et al.*, 1996; Iannolo *et al.*, 1997). It was recently demonstrated that Eps15 is an essential component of the endocytic machinery, since endocytosis of EGF and transferrin can be blocked by anti-Eps15 antibodies or by over-expression of protein fragments encompassing either the N-terminal EH domain or the C-terminal DPF domain (Carbone *et al.*, 1997; Benmerah *et al.*, 1998). Furthermore, intersectin, or its *Drosophila* homolog Dap160, which contain two EH domains, bind to dynamin, another molecule whose participation in coated pits-mediated endocytosis is clearly established (Roos and Kelly, 1998; Yamabhai *et al.*, 1998). Yeast strains mutated in two of the three genes encoding EH-containing proteins, PAN1 and END3, are defective in endocytosis (Benedetti *et al.*, 1994; Munn *et al.*, 1995; Tang *et al.*, 1997; Wendland and Emr, 1998).

Actin cytoskeleton regulation is another function linked to EH-containing proteins, as shown by the involvement of Pan1p and End3p in the organization of the actin cytoskeleton (Tang and Cai, 1996; Tang *et al.*, 1997). Recently, an EH-containing protein was found that binds RalBP1, a GTPase activating protein for CDC42 and Rac GTPases (Yamaguchi *et al.*, 1997; Ikeda *et al.*, 1998).

It is assumed that the peptide recognition specificity of EH domains plays an important role in determining the biological properties of EH-containing proteins, possibly by modulating the formation of macromolecular complexes. By screening a nonapeptide repertoire displayed by fusion to the major coat protein of filamentous phage

(Felici *et al.*, 1991), we have found that the N-termini of Eps15 and Eps15R bind peptides that contain an NPF (Asn-Pro-Phe) motif (Salcini *et al.*, 1997). Similarly, the EH domains of intersectin bind linear or constrained NPF-containing peptides (Yamabhai *et al.*, 1998). Indeed Eps15-binding partners such as Rab, Numb, RabR, NumbR and synaptojanin contain single or multiple NPF motifs that, at least in the case of Numb, mediate *in vivo* recognition. Although only the NPF tripeptide was found to be essential for binding, alanine scanning mutagenesis of peptide targets suggested that positions +1, –1 and –2 with respect to NPF also contribute to modulate binding affinity (Salcini *et al.*, 1997). The following studies were undertaken to investigate the molecular basis of EH recognition specificity.

Results

The six EH domains of Eps15 and Eps15R display diverse recognition preferences

The six EH domains of Eps15 and Eps15R were expressed by fusing their coding sequences to the glutathione-*S*-transferase (GST) gene in pGEX expression vectors. The purified hybrid proteins were used to screen a multivalent nonapeptide phage display library as previously described (Felici *et al.*, 1991; Dente *et al.*, 1997). In each panning experiment, after two panning cycles ~20 single clones were tested by phage ELISA, and those confirmed to be positive for binding were sequenced. The selected peptides (Figure 1) indicate that each individual EH domain from Eps15 and Eps15R can bind to NPF-containing peptides. However, from the comparison of the peptide sequences in each series, it is possible to recognize a certain degree of recognition specificity. Both the first and third EH domains of Eps15R (EH1R and EH3R, respectively), with a single exception, selected peptides that display an R after the conserved NPF motif. Furthermore, EH3R displays a significant preference for Q at position +2. These results are consistent with the observation that Eps15R, but not Eps15, binds preferentially to NPFR peptides (Salcini *et al.*, 1997). In contrast, the first two EHs of Eps15 (EH1 and EH2) and the second of Eps15R (EH2R) are less selective and tolerate several residues at position +1. Small hydrophilic residues such as T, N or S are found preferentially at position –1 in most of these domains.

The behaviour of the third domain of Eps15 (EH3) is strikingly different. Although two of the peptides selected by this domain contain the NPF motif, the majority is characterized by the presence of FW (Phe-Trp). Since one peptide contains the sequence NPFW, it is possible that the FW dipeptide is part of an extended NPFW consensus.

To obtain direct evidence that different EH domains have distinct preferences when challenged with different peptides we selected 10 EH-binding phage clones and measured their ability to bind to the EH domains by ELISA type assay (see Materials and methods). The results shown in Figure 2 support the notion that different domains preferentially recognize NPF peptides within specific contexts. EH2R is the less selective among the tested domains and binds to most NPF peptides with comparable efficiency. EH1 and EH2 display a similar pattern with somewhat reduced affinity. In agreement with the consensus sequences derived in Figure 1, EH1R and
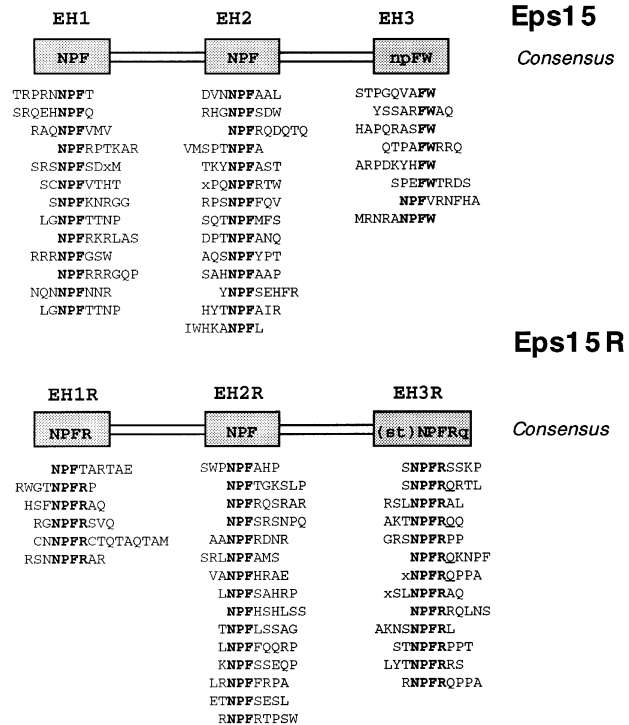

Fig. 1. Selection of peptides that bind to the EH domain of Eps15 and Eps15R. The six EH domains of Eps15 and Eps15R were utilized as GST-fusion proteins to pan a random nonapeptide phage displayed library as described in Materials and methods. The NPF motif, found in most of the peptides, is utilized to align the selected sequences. Inside the boxes, representing schematically the EH domain organization of Eps15 and Eps15R, we have reported the consensus sequences that were derived considering only the residues that are conserved in >50% of the peptides. 'x' in the peptide sequences refer to residues that could not be identified unambiguously from the DNA sequence.

EH3R are highly selective for peptides with an R after the NPF consensus while EH3 hardly binds to NPF peptides, with the possible exception of NPFL and NPFW, and is the only domain that binds to FW-containing peptides efficiently.

Qualitatively, the binding properties of the Eps15 and Eps15R fragments containing the three EH domains can be looked at as the sum of the binding properties of their constituent domains, suggesting that in the native protein, each domain is available for binding to its favourite target.

The diverse recognition specificity of the EH domains of Eps15 were further confirmed in experiments in which synthetic peptides were presented to the domains outside the phage capsid context (Figure 3).

Binding specificity of yeast EH domains

EH domains are found in proteins of many species, from lower eukaryotes to mammals. We have used the PSI_BLAST program to search the entire database for proteins containing single or multiple EH domains (Figure 4). As the entire nucleotide sequence of *Saccharomyces cerevisiae* is now available, it is possible to identify the entire 'EH repertoire' of a simple eukaryotic organism. Two reiterations of the PSI_BLAST program with an E-value of 10^{-4}, and utilizing as a bait the EH1 domain of Eps15, identified five yeast proteins with an E-value above the threshold. Starting the search with a different
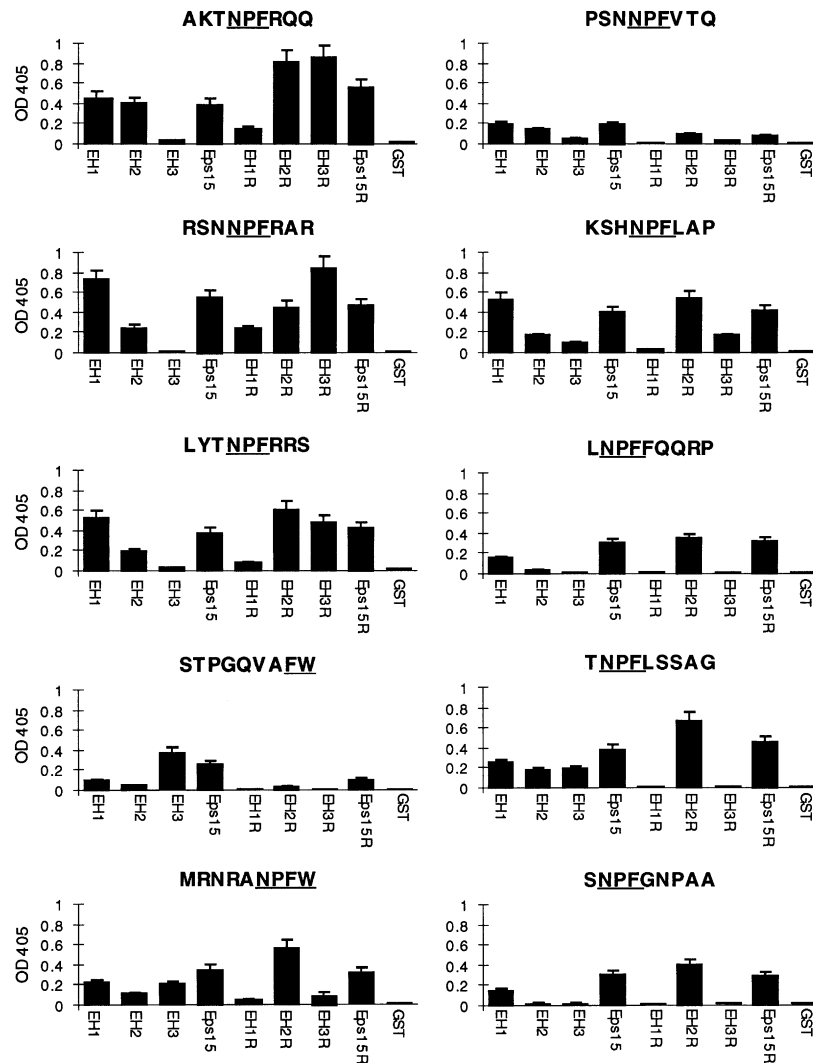
Fig. 2. Binding specificity of EH domains. Ten phage clones displaying different peptides were adsorbed to a plastic microtiter plate and challenged with equimolar amounts of different EH domains fused to GST. Bound domains were identified with an anti-GST antibody and a secondary antibody linked to alkaline phosphatase. The values reported in the histogram are an average of at least two independent experiments whose results differ by no more than 15%. The background due to non-specific binding of GST (~0.1 OD in these conditions) was subtracted from the values obtained in the presence of GST–EH fusion proteins.

EH domain did not alter the results substantially. Two of the yeast EH-containing proteins, YBL047cp and Pan1p, contain multiple EH domains, three and two, respectively. The remaining three proteins, End3p, YKR019cp and YJL083wp, display a single EH domain. A second divergent EH domain could be identified in End3p at a lower stringency and was therefore included in our analysis (E-value of 0.065 after two iterations with an E-value of 0.001).

The nine EH-domain coding sequences were cloned into pGEX expression vectors and the corresponding fusion proteins were purified by affinity chromatography. In the case of YKR019cp and YJL083wp, we have not been able to purify sufficient hybrid protein, possibly due to protease degradation. Thus, these two domains were not investigated further. The seven EH domains of the remaining three proteins were utilized in panning experiments, and with five of them we were able to identify target peptides after two or three selection cycles (Figure 5).

The first EH domain of YBL047cp binds preferentially to peptides displaying a (F/L)WR consensus, reminiscent of the FW consensus recognized by the EH3 domain of Eps15. However, one peptide was found that contains an NPF motif. NPF is also one of the consensus sequences derived from the peptides selected by the second EH domain, the other being WWxxad. Finally, the third YBL047cp EH domain has preferences for NPFR-containing peptides.

NPF peptides could also be selected by panning with the second domain of Pan1p. However, with this domain, four panning cycles were necessary to enrich for binding peptides, and the eight phages that were characterized were found to display only three different peptides. As confirmed later by phage ELISA (not shown), this suggests that the selected phage binds poorly to the Pan1p–EH2 domain. By contrast, the first EH domain of Pan1 did not select any phage in panning experiments. Furthermore, a GST-fusion protein containing both the first and second domain of Pan1p bound to phages that were indistinguishable from those selected by EH2 alone (not shown).
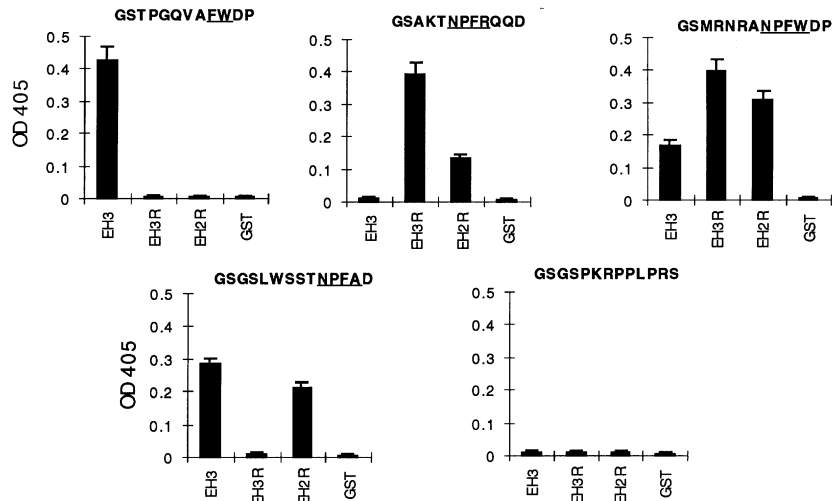
Fig. 3. Binding of EH domains to synthetic peptides. Four peptides, representative of different EH-domain binding preferences, were biotinylated in their N-terminus and bound to microtiter wells coated with 1 µg of streptavidin. Each well was incubated with 0.25 µg of GST–EH domain hybrid proteins and the bound domain revealed with an anti GST antibody. GSGSPKRPPLPRS is a peptide normally recognized by a class of SH3 domains (unpublished) and here used as a negative control.
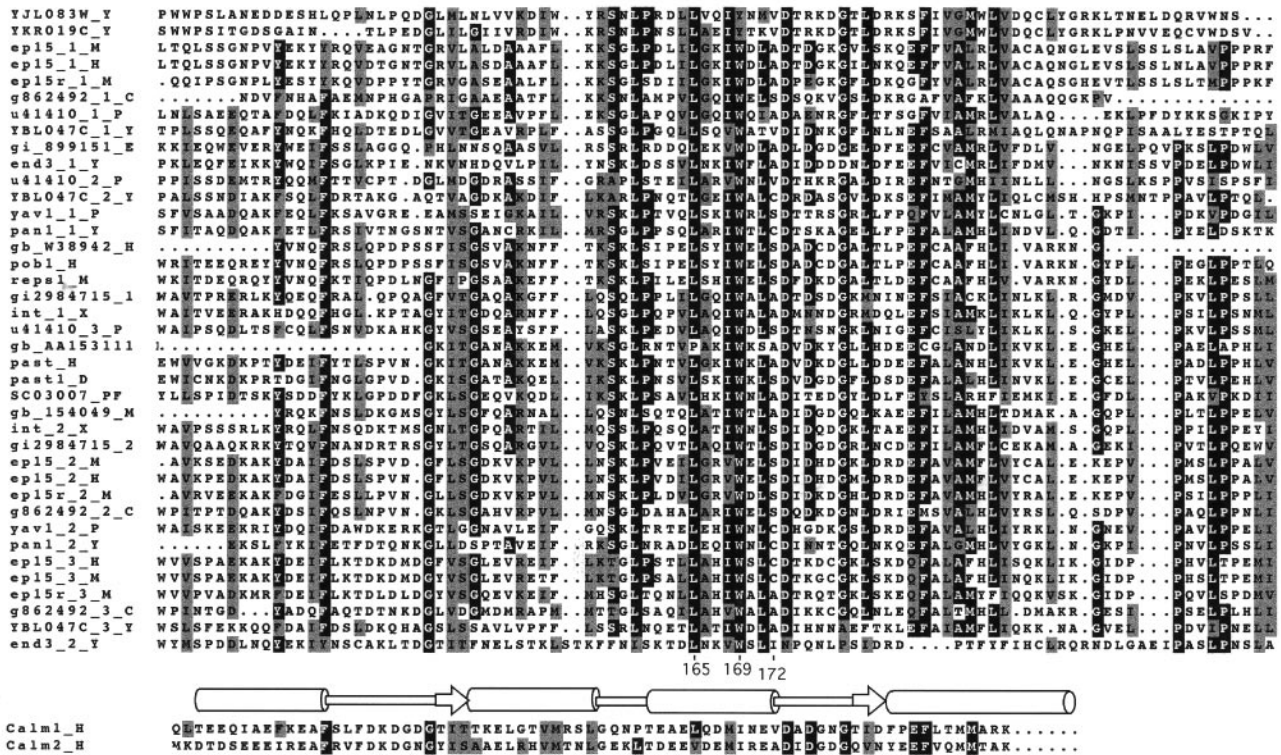

Fig. 4. Multi-alignment of EH domains. EH domains were searched on May 25th 1998 at http://www.ncbi.nlm.nih.gov/BLAST/ by the PSI-BLAST algorithm (Altschul *et al.*, 1997), using the EH1 domain of Eps15 as a starting query sequence. After two iterations with an E value of 10^{-4}, the sequences with an E value of 10^{-10} or better were aligned by the Pileup program of the UWGCG package. The SWISS-PROT Database accession numbers of the sequences that are identified in the Figure by their common names are ep15_M, sp|P42567|; ep15_H, sp|P42566|; ep15R_M, gi|968973; end3_Y; int_X, gi|2642625; reps1_M, gi|2677843; past_H, gi|2529707; past1_D, gi|1572719; yav1_P, sp|Q10172|; pan1_Y, sp|P32521|. The figures below the alignment refer to the corresponding positions in the primary sequence of Eps15–EH2 and identify the residues that have been mutated in this work. The lower part of the figure represents the two halves of the calmodulin sequence aligned with the EH domains. The secondary structure of Eps15–EH2 domain, as determined by de Beer *et al.* (1998), is illustrated with cylinders and arrows representing α-helices and β-strands, respectively.

H(S/T)F is the somewhat unorthodox consensus identified by peptides selected by the EH1 domain of End3p. Conversely, we could not enrich for any specific binding peptide after four panning cycles with the second EH domain of this protein. Also in this case, panning with an End3p protein fragment containing both domains resulted in selection of peptides very similar to those selected by EH1 alone (not shown). When the EH1 domain of End3p was used to pan a phage library displaying pentadecapeptides a different consensus could be deduced sWGxxxw.

```
                200    400    600    800   1000   1200   1400
YBL047C
              EH1 EH2  EH3                NPF
          yxxx(F/L)WRp               WWxxad              rxxNPFR

     IYAASFWRP  2        RQGNPFSKW              RTRSTNPFQ
     YPPSFWLPR  5        DATNPFGSA              SRSNPFRSL
     IYEANFWRP           HVMNPFNTR              SKNNPFRPL
     IYAASFWRP           QKPNPFHPT              RSNPFRAPS
     KYAPDYWRR           QYSNPFYAP              QVRRANPFR
     PSQPALWRV                                  RSRNPFRSS
     RHPYRLWRR           SGWWDAPGQ              RSNNPFRAR
     RPWNPQFWR           SATEWWHAAD  2          RRQNPFRPT
                         LWWHSADST              RNPFRQPP
     RNPFIWWHQ           EWWSIADQ               RSLNPFRAQ
                                                ARNPFRGAP

PAN1
         Q rich    EH1        EH2
              No
            enrichment     nNPFxD

                         YKKSNNPFSD  5
                         KSRYNNPFWD
                         TRRSRNPFWD  2

END3
              EH1 EH2
           H(s/t)F          No
           sWGxxxw       enrichment
                                          ● Ca binding
     WHSFRIAHQ   WSWGGVVVYKKDLLR
     YHTFRTRPP   FCWGTSCWLPPGLGFM
     YHTFQVRLE   FCWGETCWGASARTFS
     SHTFTVGPL   FSWGTVSWSPSSGFPW
     NHTFALYSS   FSWGTVSWSPRRGSRG
     YHTFQILAP   FSWGGVKFSVLNERAQ  4
     SHTFPLLES   FSWGVSEVMSVKQSEL
     GHTFMLLTE
     AHTFRISSS
     SHSFAIASQ
     LHTFNVGHG
     RHTFAINAQ
```
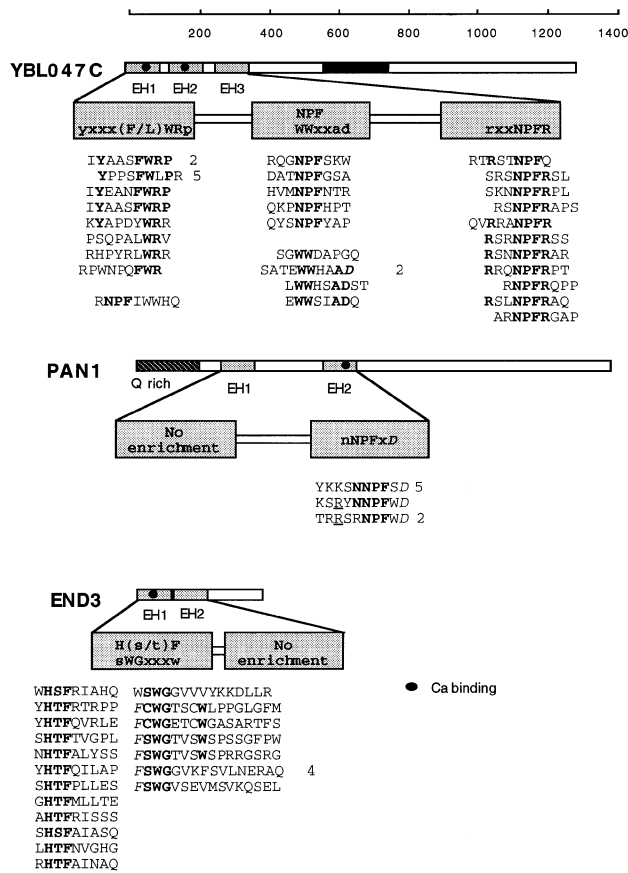
Fig. 5. Peptides selected by yeast EH domains. Panning experiments were carried out using different GST fusion proteins as baits to select nonapeptides from a random library. After two panning cycles, ~20 selected phage clones were tested by phage ELISA and the amino acid sequence of the peptides displayed by the positive phages deduced from the DNA sequence of the hybrid pVIII genes. The amino acid sequence of the selected peptides is reported below the corresponding gene structure. The residues that are found >50% of the time in the collection of peptides selected with a specific domain are shown in bold. Figures refer to the number of times that the corresponding peptide has been found independently among the positive clones. The sequence of the consensus peptides are boxed. Amino acids found in every sequence are upper case, while the residues found in >50% of the peptides are lower case. Black circles indicate putative Ca^{2+} binding sites. We have represented the residues of the pVIII capsid protein that flank the peptide inserts and could participate in domain binding in italic.

The three peptide classes, selected by the second domain of Pan2p and by the first of End3p, can be aligned without shifting their frame, suggesting that one or more residues in the coat protein context may be important in modulating binding affinity. The coat residues that might contribute to binding are indicated in italic in Figure 5.

The C-terminal boundary of the EH domain
By searching the database of proteins whose three-dimensional structure has been determined (Protein Data Bank), with a protein profile derived from an alignment of EH domains, a significant homology was detected between a portion of the EH domain and proteins that display an EF-hand fold, such as calmodulin, recoverin and sarcoplasmic binding protein. The structural significance of this homology has been confirmed recently by the determination of the solution structure of the second EH


```
EH  DOMAIN  COOH  TERMINUS                                              % BINDING

ep15_1_M   QEFFVALRLVACAQNGLEVSLSSLSLAVPPPRFHDS                         100
Δep15_1_M  QEFFVALRLV                                                     0
ep15_2_M   DEFAVAMFLVYCALEKEPV...PMSLPPALVPPSKRK                         100
Δep15_2_M  DEFAVAMFLV                                                      0
ep15_3_M   DQFALAFHLINQKLIKGIDP..PHSLTPEMIPPSDRSS                        100
Δep15_3_M  DQFALAFHLI                                                      0
end3_1_Y_S EEFVICMRLIFDMVNKNISSVPDE                                        0
end3_1_Y   EEFVICMRLIFDMVNKNISSVPDELPDWLIPGSK                               0
end3_1_Y_L EEFVICMRLIFDMVNKNISSVPDELPDWLIPGSKVNLIKERKKRKQIENADLPPKKEIKVD  100

                                Helix D
```
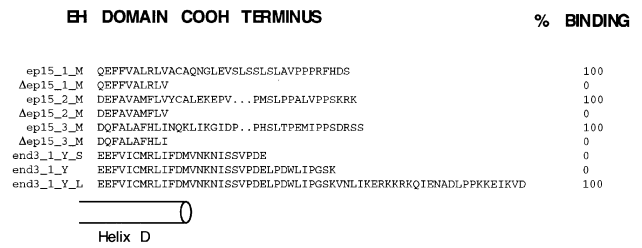
Fig. 6. Mapping the C-terminal-end of the EH domain. The C-terminal residues of the EH domains that were expressed by fusion to the GST protein are reported in the figure. The End3–EH1 domains were tested, by phage ELISA, against a panel of peptides containing the HTF or the HSF motif, while the EH domains of Eps15 were tested against peptides containing the NPF consensus. '100' indicates that the corresponding domains displayed full binding activity, while '0' indicates that the phage ELISA with the corresponding domain gave a result indistinguishable from background. The cylinder in the lower part of the figure represents the C-terminal part of the fourth helix (de Beer et al., 1998). The deletions in the Eps15 domains extend into the last turn of helix D because they were designed on an EH model that we assembled, by homology modelling on the calmodulin structure, before the NMR structure had been reported.

domain of Eps15 (de Beer et al., 1998). The N-terminal 70–80 residues of EH domain fold into two helix–loop–helix motifs connected by a short antiparallel β-sheet. The C-terminal 25 residues that are included in our standard constructions display a conserved pattern of repeated hydrophobic and proline residues that, in the three-dimensional structure, zigzag over the third and fourth helix (Figure 4). However, the C-terminal boundaries of a functional EH domain cannot be unequivocally identified by sequence alignment due to variable amino acid conservation, which is lower than in the N-terminal portion.

In order to examine whether the C-terminal regions of EH are necessary for peptide binding, we expressed derivatives of the three domains of Eps15 and the first domain of End3p, and tested their ability to bind to a set of target peptides by phage ELISA. We have observed that a deletion of ~25 residues, including the hydrophobic-Pro motif, in any of the three Eps15 domains completely abolishes binding (Figure 6). Thus, we conclude that the core EH domain containing only the two helix–loop–helix motifs is not sufficient for NPF peptide binding.

Similarly, the first domain of End3p requires a C-terminal region that extends beyond the EF-hand homology boundary since an End3p that terminates at residue K96 does not show any detectable binding to phages displaying various HTF or HSF peptides (Figure 6).

Residues that are involved in peptide binding
The residues corresponding to L165 and W169 in the Eps15–EH2 sequence are among the most conserved in the EH domain family. We therefore tested whether these residues are involved in ligand binding. We constructed four mutant Eps15–EH2 domains by changing L165 into A, and W169 into Y, F or A. All the mutant domains seem to fold properly, judging by the yield of soluble GST–EH hybrid protein obtained in overproducer strains. However, both L165A and W169A have completely lost their ability to bind NPF-containing peptides (Figure 7). In contrast, when W169 is changed into either F or Y, most of the binding activity is retained.
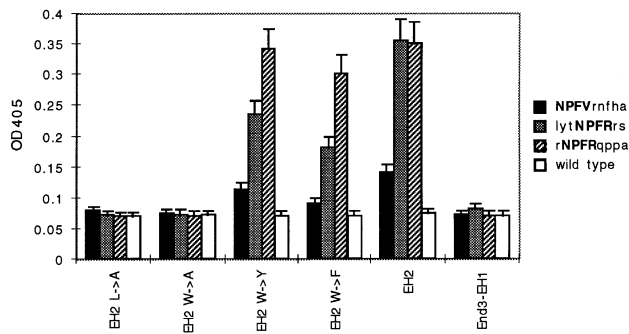
Fig. 7. Binding of Eps15–EH2 mutants to NPF peptides. Four Eps15–EH2 mutants were obtained by site-directed mutagenesis and tested by phage ELISA for binding to three different NPF-containing peptides. The End3–EH1 domain was used as a negative control.

Two classes of peptides

The EH-domain ligands, as deduced from our panning experiments, can be grouped into two different classes. Most EH domains select peptides that contain a variation of the typical NPF motif. In contrast, some domains like Eps15–EH3, the first and second domain of YBL047cp, or the first of End3p, select peptides that are characterized by consensi containing aromatic and hydrophobic residues (FW, WW, SWG, etc.) (Figures 1 and 5). We term the latter class II peptides. Notably, some of the domains with a marked preference for NPF peptides, such as Eps15–EH1, also displayed some affinity for class II peptides (not shown).

The third EH domain of Eps15 selects both NPF-containing (class I) and FW-containing (class II) peptides, thus representing a good model to test whether the two different classes of peptides bind to the same or different sites (Figure 1). Using competition experiments, we tested the ability of NPF and FW peptides to bind to the EH3 domain of Eps15. As shown in Figure 8, the synthetic GSTPGQVAFWDP peptide is equally efficient (IC50 ≈ 50 µM) in competing the binding of EH3 to either the cognate FW peptide (GSTPGQVAFWDP) or the NPFA (GSGSLWSSTNPFAD) and NPFW (GSMRNRANPF-WDP) peptides. The same FW peptide, however, cannot compete with an NPFR peptide (GSAKTNPFRQQD) for binding to the EH3 domain of Eps15R, consistent with its inability to bind to this domain. Similarly, the NPFW peptide competes with the same efficiency (IC$_{50}$ ≈ 200 µM) with binding of EH3 to both peptide classes This result strongly suggests that the binding sites for the two peptide classes are either coincident or very close.

A residue that modulates binding specificity

By comparing the primary sequences of domains with similar or divergent recognition specificity we have identified a residue, at position 172 (+3 with respect to the conserved tryptophan), whose chemical characteristics correlate with recognition specificity (see Figure 4 and Discussion). Domains that prefer NPF have an Ala or Ser at this position (Ala being preferred by those requiring an Arg after the NPF motif), while the domains that bind to class II peptides have slightly larger side chains, either Cys or Val.

To test whether this residue could modulate recognition specificity, we have constructed four derivatives of EH3
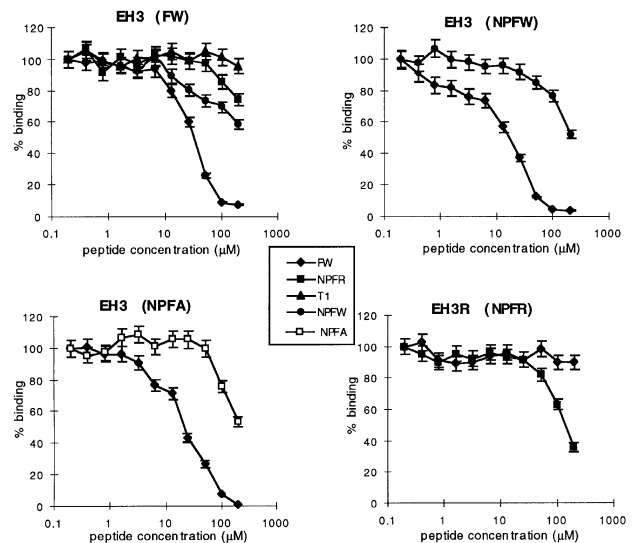

Fig. 8. Inhibition of EH-domain binding by different peptides. Aliquots (0.1 µg) of each GST–EH domain, as indicated at the top of each graph, were incubated in microtiter wells coated with 0.3 µg of streptavidin linked to biotinylated peptides (in parentheses). Binding was measured by revealing, with an anti-GST antibody, the amount of retained GST–EH domain fusion protein. Similar reactions were carried out in the presence of increasing concentrations of competing peptides. The sequence of the peptides were as follows: NPFA (GSGSLWSSTNPFAD), FW (GSTPGQVAFWDP), NPFR (GSAKTNPFRQQD), NPFW (GSMRNRANPFWDP), T1 control peptide (HDGYLQGLSGGG). Each point is obtained as an average of at least two independent experiments.

and EH3R by replacing the Cys that is present in EH3 at this position with either an Ala, or a Ser and the Ala of EH3R into either Cys or Val. We have then assayed, by phage ELISA, the ability of these mutants to bind to a sample of NFP or FW peptides.

The results reported in Figure 9 support our prediction, since EH3 domains that have Cys replaced by either Ala or Ser bind more efficiently to NPF-containing peptides, while EH3 domains with Cys and Val acquire the ability to bind to FW and NPFW peptides.

The observed change in recognition patterns, however, cannot be simply described as a shift from NPF to FW specificity, but rather as a relaxation of recognition specificity. EH3 C→A, for instance, acquires the ability to bind to NPFR peptides but retains the ability to bind FW peptides. Conversely, EH3R A→C (or A→V) maintain their ability to bind NPFR peptides, albeit with somewhat reduced affinity.

Discussion

Previous work (Salcini *et al.*, 1997) had shown that the N-termini of Eps15 and Eps15R, containing three 100-amino-acid repeats (EH domains), bind peptides that share a common NPF motif. Furthermore, it was concluded that binding of Eps15 to the protein NUMB *in vivo* is mediated by binding of the EH domain to an NPF tripeptide located near the C-terminus (Salcini *et al.*, 1997). Recently, we have found that the EH domains of intersectin also bind to NPF peptides (Yamabhai *et al.*, 1998).

However, the molecular basis of recognition of specific peptides by individual EH domains remained to be elucidated. In this study we demonstrate that a large fraction of
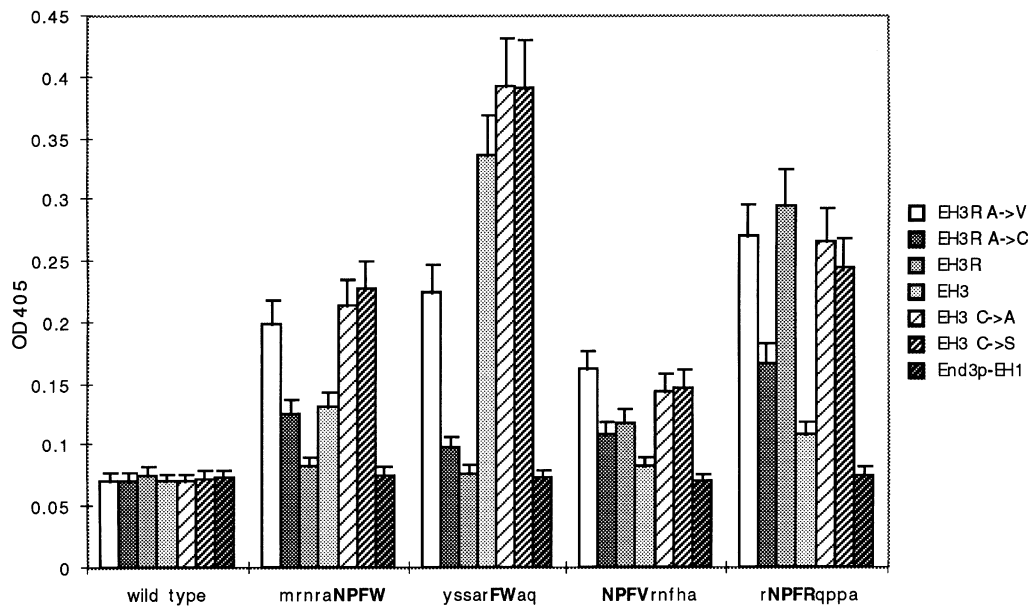
Fig. 9. Recognition specificity of EH3 and EH3R derivatives with substitutions at position 172. GST–EH3 and GST–EH3R hybrid proteins (0.25 μg of either wild-type or mutant proteins), were adsorbed to microtiter wells and incubated with 10^9 phage particles displaying the peptides indicated in the legend. Bound phage was revealed with HRP conjugated anti-filamentous phage antibodies (Pharmacia). The End3–EH1 domain was used as a negative control.

EH domains recognises NPF peptides albeit with a different sequence context preference.

Panning a phage-displayed nonapeptide library with 13 different EH domains revealed that at least 10 can bind peptides containing the NPF motif. Among the remaining three, the first EH domain of the yeast protein Pan1p and the second of End3p could not find any specific target structure in our peptide library.

The first EH domain of End3p, in contrast, binds to peptides that share the HTF/HSF tripeptide consensus, thus defining a new EH domain recognition specificity. We considered the possibility that NPF and HTF/HSF peptides could share a common conformation, despite the sequence difference. Thus, we compared the conformation of the NPF or H(T/S)F peptides that are present in the proteins of known structure in the PDB database. We observed, however, that while a large fraction of NPF share a common turn-like conformation, stabilized by a conserved pattern of hydrogen bonds, HTF peptides have a much wider distribution of conformations (not shown).

When all the selected NPF peptides are considered together, no strong preference can be identified in the position immediately preceding or following the NPF motif. However, the preference for Asn and Thr at position –1, Arg at position +1 and Gln at +2, is statistically significant, with an occurrence that is approximately four times higher than predicted by the respective codon frequency. More striking is the strong negative bias for some residues that are never found at specific positions. For instance, negatively charged residues are never found at positions –1 and +1. Ile, Val and Trp are also not tolerated at –1, while Ile and Tyr are heavily under-represented at any position following NPF.

However, by grouping NPF binding domains into broad families, further regularities can be identified. There are domains belonging to the first family, such as the first and second domain of Eps15R or the third domain of the yeast protein YBL047c, which display a strong preference for R at position +1. The peptides selected by domains of this first group also display a statistically significant bias for positive residues at position –2 and –3, and for glutamine and alanine at position +2.

A second family of EH domains, including the first and second domain of Eps15 and the second domain of Eps15R, is less selective, showing only a slight preference (~3-fold) for Thr at position –1 and Ala at +1.

Finally, the last family possesses domains, such as the first and second domain of YBL047cp, the third domain of Eps15, or the second domain of Pan1p, which generally have a lower affinity for NPF peptides and show a marked preference for Asn at –1 and for Trp at +1. An additional characteristic of this last domain family is the ability to select another class of peptides (class II) that does not contain NPF and is characterized by two aromatic residues: FW in the case of YBL047cp–EH1 and Eps15–EH3, and WW for YBL047cp–EH2. In addition, the End3p–EH1 selects, aside from HTF peptides, a second peptide family containing a Trp (SWG peptides).

The functional significance of the NPF consensus, identified by phage display, has been demonstrated by the isolation of several EH-binding proteins containing one or more NPF motifs (Haffner *et al.*, 1997; Salcini *et al.*, 1997; Yamabhai *et al.*, 1998), and by the observation that the C-terminal motif of the protein NUMB is essential for NUMB-Eps15 co-immunoprecipitation (Salcini *et al.*, 1997). It is not clear at present whether class II peptides, displaying the consensus with two aromatic residues, would also represent a second functionally relevant binding mode of EH domain or if they should be considered as mimotopes.

The consensus sequences that we have determined can be used as bait to search protein databases with a pattern search algorithm for putative protein targets. In *S.cerevisiae,* where the entire genome sequence has been deter-

Table I. Details of expression vector constructions

Domain	Limits	Oligonucleotides		Vector
EPS15 EH1	L7-S106	GCGGCAGGGATCCTCTCCCTGACACAG	(R370)	pGEX5×-3
		ACGGACCCCGGGAGTCATGAAATCTTGG	(R371)	
EPS15 EH2	A122-K216	TGGGCTGTAAAGTCTGAA		pGEX-2T
		TTTTCTCTTAGAAGGTGG		
EPS15 EH3	W218-S315	TGGGTTGTATCCCCTGCA		pGEX-2T
		ACTGGATCTGTCTGATGG		
EPS15R EH1	L7-T105	CCCCTCTCGCAGCAGATT		pGEX-2T
		ATCATGAAATTTTGGTGG		
EPS15R EH2	A121-K205	TGGGCTGTGCGGGTGGAG		pGEX-2T
		CTTCCTCTTAGAGGGTGG		
EPS15R EH3	W266-T364	TGGGTTGTGCCTGTGGCA		pGEX-2T
		AGTGCCTCTCTCAGAGGG		
PAN1 EH1-EH2	P259-T669	GATATCGGATCCCCTGCCATTAGATTGTCGTT	(R294)	pGEX-2T
		ATTTCCGGATCCGAAAAATCTTTGTTTTATA	(R291)	
PAN1 EH1	P259-F360	GATATCGGATCCCCTGCCATTAGATTGTCGTT	(R294)	pGEX-2T
		AATGGCGAATTCCAAAACTTGAAACTTCGTT	(R295)	
PAN1 EH2	E600-T687	TGTCTAGAATTCCTGTGCTGGAGGGAATTAA	(R290)	pGEX-2T
		ATTTCCGGATCCGAAAAATCTTTGTTTTATA	(R291)	
END3 EH1-EH2	M1-K236	GATCTCGGATCCATGCCCAAGTTGGAACAATTT	(R282)	pGEX-2T
		CTTCTTGAATTCCCTTTGTAGGAGGTTGAGAGGA	(R283)	
END3 EH1-S	M1-E86	GATCTCGGATCCATGCCCAAGTTGGAACAATTT	(R282)	pGEX-2T
		GAGACAGAATTCCATCTACTTTGATTTCCT	(R292)	
END3 EH1	M1-D123	GATCTCGGATCCATGCCCAAGTTGGAACAATTT	(R282)	pGEX-2T
		ATCAGGGAATTCATCTGGAACAGA	(R380)	
END3 EH1-L	M1-K96	GATCTCGGATCCATGCCCAAGTTGGAACAATTT	(R282)	pGEX-2T
		GATTAGAATTCCCTTACTCCCTGGAAT	(R416)	
END3 EH2	W124-C219	TCAAAGGATCCTGGTACATGTCTCCAGATG	(R293)	pGEX-2T
		TTCTTGAATTCCCTTTGTAGGAGGTTGAGAGGA	(R283)	
YBL047c EH1	A2-L102	GCATCTATTACTTT		pGEX-2T
		CAGTTGAGTAGGTGT		
YBL047c EH2-S	N124-L220	AACAACACTGATAT		pGEX-2T
		CAGCTGAGTAGGCAACA		
YBL047c EH2	N124-P228	AACCGGGTCGACAAACAACACTGATATT	(R381)	PYex
		GATTAAGAATTCCAGGTTCTAAACGAAT	(R382)	
YBL047c EH3	A263-L365	GCATTCAGTAATGCT		pGEX-2T
		TAAAGCGGGGGACTGTAATAATT		
YKR019c	470–570	AGCAACAGACACCGTCATTT		pGEX-2T
		TCTATCTACGCTAT		
YJL083w	453-557	TCAAACAGATACTC		pGEX-2T
		GCCGTCAACGCTATT		

mined, this search can be exhaustive (Chervitz *et al.*, 1997). In a typical search one may retrieve from a few to a few hundred putative targets, depending on the stringency of the pattern utilized in the search (not shown). Some judicious selection, for instance disregarding the putative target proteins that are known to be in a different cellular compartment to the bait, allows us to focus on a sufficiently low number of candidates to be approached experimentally. Recently, Wendland and Emr (1998) reported that the region encompassing the two EH domains of Pan1p binds yAP180Ap, one of the two yeast homologues of a class of clathrin assembly proteins (AP180). The same authors identified a genetic interaction between PAN1 and SJL1, one of the three synaptojanin-like genes in yeast. Interestingly, yAP180Ap contains five NPF motifs and SJL1p has one NPFXD motif that exactly matches the consensus that we have identified by phage display. Furthermore, NPFXD was recently identified as a new class of endocytosis signal in *S.cerevisiae* (Tan *et al.*, 1996).

The solution structure of the second domain of Eps15 was recently reported by de Beer *et al.* (1998). The core of the EH domain (~75 residues), that shares a significant homology with EF-hand proteins, folds into two helix–loop–helix motifs connected by a short β-sheet. The

C-terminal 20 residues, which we have shown to contain information necessary for binding to NPF peptides, have a less regular structure. The residues that are involved in the formation of the 'NPF' binding pocket were identified by looking at the progressive changes in the ^{1}H, ^{13}C and ^{15}N resonances upon titration with the peptide (PTGSSST-NPFL) corresponding to the C-terminus of RAB (de Beer *et al.*, 1998). NPF peptides bind to a hydrophobic pocket that is formed by Trp169, Leu155, Leu156 and Leu165 in Eps15–EH2. In this work we have shown that when either Trp169 or Leu165 are changed into Ala, the ability to bind NPF peptides is lost completely. Consistent with the results of de Beer *et al.* (1998), more conservative changes at position 169 (W→Y and W→F) have less dramatic effects.

In an attempt to identify residues that modulate peptide recognition, we have tried to relate sequence recognition specificity to EH-domain primary structure. By aligning and comparing EH-domain sequences with identical or divergent recognition specificity, we could not identify simple revealing regularities in the primary structures of domains that bind to similar peptides. However, we have spotted a single position (three residues after the conserved tryptophan) whose side chains display some correlation

with recognition specificity (Figure 4). We started by comparing the amino acid sequence of the third domain of Eps15 (EH3) with that of Eps15R (EH3R), which although very similar (63% identity), has distinct recognition specificity (Figure 2). EH3 has a negligible affinity for most NPF peptides and preferentially binds to FW-containing peptides. In contrast, EH3R prefers peptides where the NPF motif is followed by an arginine. The EH3 and EH3R peptide sequences differ in 33 positions. When these variable residues are confronted with the residues that are present in more distant EH domains, a significant correlation between specificity and side-chain characteristics is found only at position +3 following the conserved tryptophan (position 172 in the Eps15–EH2 sequence); this is an Ala in EH3R and a Cys in EH3. However, only residues with small side chains are tolerated in the remaining EH domains. Significantly, all the domains that prefer NPF have an Ala or a Ser at this position (Ala being preferred by those requiring an Arg after the NPF motif), while the domains that bind to class II peptides have slightly larger side chains with either Cys or Val. By site-directed mutagenesis we have shown that it is possible to modulate recognition specificity by directed substitutions at position 172, Ala and Ser favouring binding to NPF peptides, and Cys and Val promoting binding to FW peptides. However, our results suggest that it is not possible to shift recognition specificity from FW to NPF binding and vice versa by only introducing appropriate changes at position 172, and that other changes may be required.

Most EH proteins contain multiple repeats of the EH domain. It is not clear whether, in vivo, each EH domain binds to different protein targets or whether multiple EH domains cooperate in binding to the same protein target, thereby increasing the stability and rigidity of the complex. The relatively low affinity of a single EH–NPF peptide interaction and the observation that most EH-protein targets have multiple NPF sequences would favour the second model.

Approximately 50% of the NPF peptides that we have selected have either a Ser or a Thr at position –1 or –2 (Yamabhai et al., 1998; this study). Accordingly, an equivalent (or even larger) enrichment of Ser and Thr is observed in the corresponding positions in the binding peptides that are found in the EH-domain protein targets (Salcini et al., 1997; Yamabhai et al., 1998). Since we have shown that negative charges at –1 and –2 would negatively affect NPF peptide binding to EH domains, one might speculate that peptide binding could be modulated by a mechanism involving Ser/Thr phosphorylation. Whether this convenient regulatory opportunity has been exploited by natural selection remains to be established.

Materials and methods

Strains and enzymes

Escherichia coli strain DH5α F+ (endA1 hsdR17 (r⁻k m k+) recA1 endA1 gyrA96 thi-1 relA1 △upE44 f80lacZDM15) was utilized for expression of recombinant proteins and for growth of filamentous phage. Restriction enzymes were purchased from New England Biolabs, Taq polymerase from Perkin Helmer and T4 DNA ligase from Amersham. Oligonucleotides were purchased from Genset and oligopeptides from Genosys.

Plasmid constructions

Expression plasmids were constructed by standard recombinant DNA technology by inserting DNA fragments encoding the relevant EH domain into an expression vector of the pGEX series (Pharmacia). The oligonucleotides utilized to amplify the different EH-coding sequences and the recipient expression vectors are indicated in Table I. After ligation and electroporation, recombinant clones were identified by PCR, and the inserted DNA fragment was sequenced to exclude the possibility that the PCR procedure had introduced unwanted mutations. EH mutants were obtained by standard site-directed mutagenesis techniques.

Biopanning (affinity selection)

The construction of the nonapeptide display library and the panning conditions were as previously described (Felici et al., 1991; Dente et al., 1997). The bait protein is prepared as fusion with the GST by affinity purification on a glutathione–Sepharose resin (Pharmacia).

Every panning cycle was carried out with 4 μg of bait protein and 10^{10} transducing units (~2 pmol of random nonamers) in 100 μl phosphate-buffered saline (PBS), 3% bovine serum albumin (BSA) for 1 h at 4°C. The GST–EH domain fusion protein was immobilized by binding to 20 μl of gluthatione–Sepharose matrix. After two or three panning cycles, the selected phage clones were tested in ELISA against their bait proteins. Single-stranded DNA was prepared as described previously (Dente et al., 1983) and sequenced using an ABI 310 Perkin-Elmer instrument.

Phage ELISA

Each well was loaded with 10^{10} transducing units of the selected phage in 100 μl PBS, 3% BSA. The phage suspension was incubated for 8 h at 4°C, and after washing, 0.2 μg (5 pmol) of the appropriate GST-fusion protein was applied to each well in 50 ml of PBS, 5% dry milk, and incubated for 2 h at 37°C. After washing, a 1:1000 dilution of an anti-GST rabbit antiserum (a gift of F.Benfenati) was applied and incubated for 1 h at room temperature. Detection was performed using an alkaline phosphatase conjugated anti-rabbit goat antibody, diluted 1:1000 (Sigma) and pre-adsorbed for 2 h with 10^{11} particles of M13K07.

The chromogenic reaction was developed for 1 h at 37°C by adding 50 μl of 1mg/ml p-Nitrophenylphosphate disodium, hexahydrate (PNPP; Sigma 104) in 2 mM $MgCl_2$, 50 mM Na_2CO_3 pH 9.6 to each well. Reading was performed at 405 nm.

Acknowledgements

References

Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25, 3389–3402.

Benedetti,H., Raths,S., Crausaz,F. and Riezman,H. (1994) The END3 gene encodes a protein that is required for the internalization step of endocytosis and for actin cytoskeleton organization in yeast. Mol. Biol. Cell, 5, 1023–1037.

Benmerah,A., Begue,B., Dautry-Varsat,A. and Cerf-Bensussan,N. (1996) The ear of α-adaptin interacts with the COOH-terminal domain of the Eps15 protein. J. Biol. Chem., 20, 12111–12116.

Benmerah,A., Lamaze,C., Begue,B., Schmid,S.L., Dautry-Varsat,A. and 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,P.G. and Torrisi,M.R. (1997) eps15 and eps15R are essential components of the endocytic pathway. Cancer Res., 57, 5498–5504.

Chervitz,S.A., Cherry,J.M. and Botstein,D. (1997) Saccharomyces Genome Database On-line Pattern Matching Program, version 0.y. http://genome-www.stanford.edu/Sacch3D/patmatch.html.

Coda,L., Salcini,A.E., Confalonieri,S., Pelicci,G., Sorkina,T., Sorkin,A., Pelicci,P.G. and Di Fiore,P.P. (1998) Eps15R is a tyrosine kinase substrate with characteristics of a docking protein possibly involved in coated pits-mediated internalization. *J. Biol. Chem.*, **273**, 3003–3012.

de Beer,T., Carter,R.E., Lobel-Rice,K.E., Sorkin,A. and Overduin,M. (1998) Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science*, **281**, 1357–1360.

Dente,L., Cesareni,G. and Cortese,R. (1983) pEMBL: a new family of single stranded plasmids. *Nucleic Acids Res.*, **11**, 1645–1655.

Dente,L., Vetriani,V., Zucconi,A., Pelicci,G., Lanfrancone,L., Pelicci,P.G. and Cesareni,G. (1997) Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. *J. Mol. Biol.*, **269**, 694–703.

Di Fiore,P.P., Pelicci,P.G. and Sorkin,A. (1997) EH: a novel protein–protein interaction domain potentially in intracellular sorting. *Trends Biochem. Sci.*, **22**, 411–413.

Fazioli,F., Minichiello,L., Matoskova,B., Wong,W.T. and Di Fiore,P.P. (1993) Eps 15, a novel tyrosine kinase substrate, exhibits transforming activity. *Mol. Cell. Biol.*, **13**, 5814–5828.

Felici,F., Castagnoli,L., Musacchio,A., Jappelli,R. and Cesareni,G. (1991) Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. *J. Mol. Biol.*, **222**, 301–310.

Haffner,C., Takei,K., Chen,H., Ringstad,N., Hudson,A., Butler,M.H., Salcini,A.E., Di Fiore,P.P. and De Camilli,P. (1997) Synaptojanin 1: localization on coated endocytic intermediates in nerve terminals and interaction of its 170 kDa isoform with Eps15. *FEBS Lett.*, **419**, 175–180.

Iannolo,G., Salcini,A.E., Gaidarov,I., Goodman,O.B., Baulida,J.J., Carpenter,G., Pelicci,P.G., Di Fiore,P.P. and Keen,J.H. (1997) Mapping of the molecular determinants involved in the interaction between eps15 and AP-2. *Cancer Res.*, **57**, 240–245.

Ikura,M., Clore,G.M., Gronenborn,A.M., Zhu,G., Klee,C.B. and Bax,A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science*, **256**, 632–638.

Meador,W.E., Means,A.R. and Quiocho,F.A. (1992) Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin–peptide complex. *Science*, **257**, 1251–1255.

Munn,A.L., Stevenson,B.J., Geli,M.I. and Riezman,H. (1995) end5, end6 and end7: mutations that cause actin delocalization and block the internalization step of endocytosis in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, **6**, 1721–1742.

Pawson,T. and Scott,J.D. (1997) Signaling through scaffold, anchoring and adaptor proteins. *Science*, **278**, 2075–2080.

Rickles,R.J., Botfield,M.C., Weng,Z., Taylor,J.A., Green,O.M., Brugge,J.S. and Zoller,M.J. (1994) Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *EMBO J.*, **13**, 5598–5604.

Roos,J. and Kelly,R.B. (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,E., Confalonieri,S., Doria,M., Santolini,E., Tassi,E., Minenkova,O., Cesareni,G., Pelicci,P.G. and Di Fiore,P.P. (1997) Binding specificity and *in vivo* targets of the EH domain; a novel protein:protein interaction module. *Genes Dev.*, **11**, 2239–2249.

Songyang,Z. *et al.* (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell*, **72**, 767–778.

Songyang,Z. *et al.* (1994) Specific motifs recognised by the SH2 domains of Csk, 3BP2, fps/fes, Grb-2, HCP, SHC, Sykand Vav. *Mol. Cell. Biol.*, **14**, 2777–2785.

Songyang,Z. *et al.* (1997) Recognition of unique C-terminal motifs by distinct PDZ domains. *Science*, **275**, 73–77.

Sparks,A.B., Quilliam,L.A., Thorn,J.M., Der,C.J. and Kay,B.K. (1994) Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J. Biol. Chem.*, **269**, 23853–23856.

Tan,P.K., Howard,J.P. and Payne,G.S. (1996) The sequence NPFXD defines a new class of endocytosis signal in *Saccharomyces cerevisiae*. *J. Cell Biol.*, **135**, 1789–1800.

Tang,H.Y. and Cai,M. (1996) The EH-domain-containing protein Pan1 is required for normal organisation of the actin cytoskeleton in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**, 4897–4914.

Tang,H.Y., Munn,A. and Cai,M. (1997) EH domain proteins Pan1p and End3p are components of a complex that plays a dual role in organization of the cortical actin cytoskeleton and endocytosis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **17**, 4294–4304.

Tebar,F., Sorkina,T., Sorkin,A., Ericsson,M. and Kirchhausen,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.

van Delft,S., Schumacher,C., Hage,W., Verkleij,A.J. and van Bergen en Henegouwen,P.M. (1997) Association and colocalization of Eps15 with adaptor protein-2 and clathrin. *J. Cell Biol.*, **136**, 811–821.

Wendland,B. and Emr,S.D. (1998) Pan1p, Yeast eps15, functions as a multivalent adaptor that coordinates protein–protein interactions essential for endocytosis. *J. Cell Biol.*, **141**, 71–84.

Wong,W.T., Schumacher,C., Salcini,A.E., Romano,A., Castagnino,P., Pelicci,P.G. and Di Fiore,P.P. (1995) A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution. *Proc. Natl Acad. Sci. USA*, **92**, 9530–9534.

Yamabhai,M., Hoffman,N.G., Hardison,N.L., McPherson,P.S., Castagnoli,L., Cesareni,G. and Kay,B. (1998) Intersectin, a novel adaptor protein with two EH and five SH3 domains. *J. Biol. Chem.*, in press.

Yamaguchi,A., Urano,T., Goi,T. and Feig,L.A. (1997) An Eps homology (EH) domain that binds to the Ral-GTPase, RalBP1. *J. Biol. Chem.*, **272**, 31230–31234.