ABSTRACT

Clathrin-mediated endocytosis (CME) engages over 30 proteins to secure efficient cargo and membrane uptake. While the function of most core CME components is well established, auxiliary mechanisms crucial for fine-tuning and adaptation remain largely elusive. In this study, we identify ArhGEF37, a currently uncharacterized protein, as a constituent of CME. Structure prediction together with quantitative cellular and biochemical studies present a unique BAR domain and PI(4,5)P2-dependent protein–membrane interactions. Functional characterization yields accumulation of ArhGEF37 at dynamin 2-rich late endocytic sites and increased endocytosis rates in the presence of ArhGEF37. Together, these results introduce ArhGEF37 as a regulatory protein involved in endocytosis.

KEY WORDS: BAR domain, Endocytosis, Clathrin-mediated endocytosis, CME, Dynamin 2, ArhGEF37

INTRODUCTION

Endocytosis is essential to internalize membranes and extracellular material (Hopkins et al., 1985; Rosenbluth and Wissig, 1964). As such, it contributes to a variety of cellular functions during development and in the adult organism, including trans-cellular signaling (Ferguson and De Camilli, 2012; Traub, 2009), nutrient uptake (McMahon and Boucrot, 2011), maintenance of membrane tension (Boulant et al., 2011; Ferguson et al., 2009) and pathogen entry (Saheki and De Camilli, 2012). In mammalian cells, uptake occurs among others through CME, the clathrin-independent pathway or caveolae (Sorkin, 2000), with CME being the rate-limiting step (Saheki and De Camilli, 2012). In mammalian cells, uptake occurs among others through CME, the clathrin-independent pathway or caveolae (Sorkin, 2000), with CME being the rate-limiting step (Saheki and De Camilli, 2012).

CME is initiated by invagination of the plasma membrane (PM) through clustering of curvature-inducing proteins (Dauemke et al., 2014; Frost et al., 2009; McMahon et al., 2010). Upon indentation, the PM is therefore further formed by polygonal clathrin scaffolds (i.e. clathrin-coated pits, CCPs), ultimately creating a spherical invagination (Ferguson and De Camilli, 2012; McMahon and Boucrot, 2011). In a final step, assembly and contraction of dynamin at the neck of CCP, supported by actin polymerization (Almeida-Souza et al., 2018), triggers scission of the newly formed endocytotic vesicle from the membrane (Ferguson and De Camilli, 2012; Morlot et al., 2012; Taylor et al., 2011).

The whole process, which takes on average 35–50 s (Ehrlich et al., 2004; Shamir et al., 2016), is characterized by finely orchestrated recruitment and dissociation of various endocytic proteins (Posor et al., 2013; Taylor et al., 2011). Intriguingly, roughly one-third of the currently identified endocytic proteins contain curvature-sensitive domains (Daumke et al., 2014; Qualmann et al., 2011). As endocytosis is associated with dynamic changes in membrane geometry (e.g. low/high curvature; spherical/cylindrical/saddle shape), it has been put forward that the combination of unique curvature-sensing properties (Gallo et al., 2006; Henne et al., 2007; Shimada et al., 2007) and additional functional domains present in these proteins (Qualmann et al., 2011; Taylor et al., 2011) are crucial for the correct spatio-temporal assembly of the endocytotic machinery.

Considering the prevalence of Bin–Amphiphysin–Rvs (BAR) domain-containing proteins during endocytosis (Taylor et al., 2011), and curvature-dependent properties associated with many of its family members (Peter et al., 2004), we set out to identify uncharacterized proteins containing a BAR domain and probe their role in endocytosis.

RESULTS

To identify potential candidates, we performed a database search for uncharacterized BAR proteins (data not shown). This focused bioinformatics screen led us to ArhGEF37, a 676 amino acid-long protein of unknown function. Published expression profiles indicate differential expression across tissues (https://www.proteinatlas.org, http://biogps.org) and age (http://www.brainspan.org/). According to basic local alignment search tools, ArhGEF37 consists of an N-terminal actin-regulatory Rho guanine nucleotide exchange factor (RhoGEF) domain, followed by a BAR domain and two Src homology 3 (SH3) domains associated with protein–protein interaction (Fig. 1A). To gain further insight into protein organization, we examined the secondary structure of individual domains by using PSIPRED (Jones, 1999). We observed α-helices and coil structures in the GEF and BAR domain, whereas β-strands and coils were predicted for both SH3 domains (Table S1). A subsequent search using pGenTHREADER yielded predominantly CME-associated proteins (Table S2). Specifically, when searching separately for homologs with known structure for each domain, the
Fig. 1. See next page for legend.
Fig. 1. Structure prediction and recruitment dynamics of ArhGEF37. (A) Domain organization of ArhGEF37. (B) Sequence alignment of the BAR domains of ArhGEF37 and DNMBP showed conserved hydrophobic (blue) and non-hydrophobic (gray) residues. BAR domain alignment shows 24.9% identical (50/201) and 28.4% similar (57/201) amino acids between ArhGEF37 and DNMBP. (C) Antiparallel BAR-domain homodimer of ArhGEF37 as shown in the first and third model (monomers in green and orange), surface colored model (second and fourth model) show the electrostatic potential (red, $−10 \text{ KTe}^−1$; blue, $+10 \text{ KTe}^−1$). (D) Isosurface of BAR-domain dimers of ArhGEF37, amphiphysin (N-BAR, PDB ID: 1URU), FCHO2 (F-BAR, PDB code: 2V0O) and IRSp53 (I-BAR, PDB ID: 1Y2O); red and blue show isosurfaces of $−0.75 \text{ KTe}^−1$ and 0.75 $\text{ KTe}^−1$, respectively. (E) ArhGEF37 binds to planar model membranes. Flow-chart depicting ArhGEF37 purification and membrane-binding assay. Top box: Immobilized metal ion chromatography elution profile showing three peaks (P1, P2, P3), the imidazole elution gradient is indicated in green. Below, Coomassie Blue staining of the P1 fraction (left) and western blot analysis (right) using anti-ArhGEF37 and anti-HIS-tag antibodies as indicated. Middle box: Size-exclusion chromatography (top) followed by Coomassie Blue staining (middle) and mass spectrometry (bottom) establish ArhGEF37 as the main component of the P2 fraction. Bottom box: Membrane binding of purified ArhGEF37 probed using QCM. ArhGEF37 binding to various lipid compositions (left) and at different NaCl concentrations (right). Bar graphs depict mean frequency change ($\pm$4 experiments/condition). (F) ArhGEF37 forms puncta at the basal membrane. Main image shows magnification of boxed area of inset (top left) image. (F’) ArhGEF37 puncta are transient. (Top) Time series (main image) showing magnification of boxed area of inset (top left) image. Protein dynamics are depicted in false colors. (Bottom) Kymograph of line scan shows ArhGEF37 kinetics (arrowheads); boxed area in bottom image is shown magnified above. Scale bars: 5 $\mu$m (F, F’ top panel), 2 $\mu$m (F’ bottom panel). See also Fig. S1 and Movie 1.

BAR domain presented the N-BAR domain proteins amphiphysin1 and 2, (PDB IDs: 2FIC, 1URU, 4ATM) as the three homologs with the highest certainty scores, whereas both SH3 domains yielded DNMBP (PDB ID: 1UHC) as the best match (Tables S2 and S3). Likewise, the closest homolog of ArhGEF37 with respect to domain structure (i.e. DNMBP) showed high levels of conserved hydrophobic (blue) and non-hydrophobic residues (gray) for the BAR domain (Fig. 1B and Fig. S1A).

Next, using the homologs listed in Table S2, we built homology models by using MODELLER9.19 (Eswar et al., 2007, see Materials and Methods) for each domain present in ArhGEF37 (Fig. S1B). To model the BAR domain homo-dimer of ArhGEF37, we symmetrically superimposed two modeled BAR domains analogous to the published crystal structure (PDB ID: 1URU) (Peter et al., 2004). As in the templates used to build the model, we observed an elongated banana-shaped dimer with positive patches (blue) on the convex (i.e. positively curved) surface (Fig. 1C and PDB file). Strikingly, however, isosurface comparisons between ArhGEF37 and prototypic BAR domain members (i.e. N-BAR, F-BAR and I-BAR) yielded stark differences in shape and surface-charge distribution (Fig. 1D).

Individual BAR-domain subtypes substantially differ in their membrane interactions (Peter et al., 2004). To explore possible membrane-binding properties of ArhGEF37, we expressed C-terminal HIS$_6$-tagged ArhGEF37 in E. coli and purified the protein by using immobilized metal ion affinity chromatography (Fig. 1, top). Coomassie Blue staining followed by immunostaining of ArhGEF37 and HIS$_6$ confirmed purification of intact ArhGEF37 (Fig. 1E, top and Fig. S1C). To eliminate unspecific proteins from the sample, we performed size-exclusion chromatography, and probed for protein identity using mass spectroscopy. With 41 unique sequence hits, covering 77% of the whole protein sequence, we confirmed ArhGEF37 to be the principal component of P2 (Fig. 1E, middle). By using purified ArhGEF37, we finally analyzed its binding ability to planar membranes using quartz crystal microbalance (Sauerbrey, 1959). Strikingly, we observed binding of ArhGEF37 to model membranes formed from brain lipids (Fig. 1E, bottom left; Fig. S1D). We further observed binding of ArhGEF37 to model membranes composed of POPC/ DOPC/cholesterol/PI(4,5)P$_2$ but not to membranes devoid of charged PI(4,5)P$_2$, establishing charge-dependent protein-membrane interactions. Consistently, increases in the concentration of NaCl, which reduces electrostatic attraction by shielding charged lipid head groups and amino acids, prevented membrane binding of ArhGEF37 (Fig. 1E, bottom right and Fig. S1E).

Next, we aimed to elucidate the function of ArhGEF37 in a cellular context. For this, cells were transfected with ArhGEF37, and the basal membrane was imaged by using spinning disk confocal microscopy (Fig. 1F). Intriguingly, we observed a punctate pattern of ArhGEF37 that rapidly appeared and disappeared (Fig. 1F, Movie 1).

To further characterize the spatio-temporal kinetics of these puncta, we coexpressed fluorescently labeled full-length ArhGEF37 with several proteins present at different stages of CME (Fig. 2A), with all data first being subjected to A trous wavelet filtering (Olivo-Marin, 2002) (Fig. S2A-C and Movie 2). When comparing the overlap percentage of ArhGEF37 with the other proteins (Materials and Methods), we found the highest ArhGEF37 colocalization with dynamin 2 (DYN2), followed by colocalization with clathrin light chain (CLTA) or amphiphysin 1 (Amph1). No apparent colocalization of ArhGEF37 was observed with FCHO2, FBPI7 or APPL1 (Fig. 2A, left). To further probe these findings, images were examined by using spatial cross-correlation (Matis et al., 2012). Consistent with results from the overlap analysis, strongest cross-correlation scores were observed for ArhGEF37 and DYN2, followed by CLTA and Amph1. Again, no cross-correlation was apparent for FCHO2, FBPI7 or APPL1 (Fig. 2A, middle), suggesting recruitment of ArhGEF37 to coincide predominantly with DYN2 during late stages of CME (Fig. 2B).

To gain further insight into the recruitment dynamics of ArhGEF37, we performed a temporal cross-correlation analysis (Fig. S2D). To establish a baseline, we first took advantage of published results on the recruitment kinetics of key endocytotic components (i.e. CLTA, DYN2 and Snx9) (Fig. S2E, left). When using DYN2 as reference, the maximal peak for CLTA shifted towards the left (i.e. CLTA precedes DYN2), whereas Snx9 moved to the right (i.e. Snx9 follows DYN2) (Fig. S2E, right), establishing accurate temporal analysis. We then probed the recruitment kinetics of ArhGEF37 vs DYN2. Using cells co-transfected with full-length ArhGEF37 and DYN2, we observed a maximal correlation score for ArhGEF37 at $+1$ s (Fig. 2C), suggesting DYN2 enrichment slightly before ArhGEF37.

To further elucidate the link between ArhGEF37 and DYN2, we probed protein dynamics in the presence of endocytotic inhibitors. Consistent with published work, addition of Dynasore (Macia et al., 2006) and Dyn-4A (Harper et al., 2011) both led to an increase in DYN2 puncta at the PM (Fig. 3A), whereas no changes were observed for a cytosolic reference protein (Fig. S3A). Similarly, dynamin inhibition increased ArhGEF37 puncta density at the PM (Fig. 3B and Fig. S3B). In addition, we find a significant increase in overlap percentage of DYN2 and ArhGEF37 puncta density at the PM in the presence of Dynasore and Dyn-4A (Fig. 3C), as well as colocalization of both proteins at tubular structures that formed at the PM in the presence of Dynasore-4A (Fig. 3D). We further observed an increase in ArhGEF37 puncta upon expression of dominant-negative DYN2 (K44A) (Damke et al., 1994) (Fig. S3C,D), as well as upon addition of the clathrin-specific endocytotic inhibitor Pitstop-2 (von Kleist et al., 2011) (Fig. 3E). Collectively, all perturbation results are...
consistent with kinetic data showing ArhGEF37 enrichment at late endocytotic sites. Importantly, as changes in temperature yield stark differences in inhibitor efficacy (Fig. S3E), all experiments were performed at 37°C.

To further characterize the interplay between DYN2 and ArhGEF37, we aimed at acutely inducing protein recruitment to the PM. Published work established that hyperosmotic shock decreases membrane tension, thereby reducing turnover of endocytotic proteins.
In agreement with previous studies (Morlot et al., 2012), hyperosmolarity led to an increase in DYN2 puncta at the PM (Fig. 4A, top and Movie 2). Similarly, a hyperosmotic shock augmented the number of ArhGEF37 puncta at the PM (Fig. 4A, bottom and Movie 3). To rule out that protein accumulation was due to the perturbation protocol (e.g. cell shape changes or focal shift), we expressed cytosolic CFP. We find no changes in CFP intensity, pattern or puncta number upon hyperosmotic shock (Fig. S4A and Movie 4). Finally, we measured recruitment kinetics. As in the cross-correlation experiments, we observed DYN2 recruitment preceding ArhGEF37 enrichment at the PM (Fig. 4B).

To determine the origin of these differences in recruitment kinetics, we tested ArhGEF37 deletion mutants composed of the sole GEF domain (GEF; aa 1-263), the BAR domain (BAR; aa 215-509) or the SH3 domains (SH3; aa 452-676). When expressed in cells, the GEF domain yielded predominantly cytoplasmic localization, whereas BAR and SH3 showed punctate expression pattern at the basal membrane (Fig. 4C, left). As above, we tested for colocalization with DYN2. No significant overlap was found for the GEF or the BAR domain, whereas the SH3 domain showed strong colocalization (overlap percentage 35±16%, n=13 cells) (Fig. 4C, right). Likewise, the SH3 domain showed elevated colocalization when coexpressed with CTLA (Fig. S4B).

Next, we applied a hyperosmotic shock to the truncated versions of ArhGEF37. Curiously, we observed for all constructs elevated signal levels at the PM after hyperosmotic shock (Fig. 4D, and Movie 5). To further characterize these unexpected findings, we examined their recruitment kinetics. For the GEF domain, we find an increase in transient puncta, whereas the SH3 domain of ArhGEF37 yielded long-lasting puncta with almost identical kinetics to those of DYN2 (Fig. 4E). Intriguingly, and unlike the SH3 domain, recruitment of the BAR domain strongly resembled full-length ArhGEF37.

Collectively, these studies are consistent with BAR-dependent recruitment of ArhGEF37 to late endocytotic sites, followed by protein–protein interactions through the SH3 domain. As DNMBP, one of the predicted ArhGEF37 homologs (Table S2), has been described to bind to dynamin through its SH3 domain (Salazar et al., 2003), we next performed pull-down assays. Specifically, we used...
Fig. 4. See next page for legend.
endocytosis. Consistently, full-length ArhGEF37 and the isolated BAR domain show comparable recruitment kinetics (Fig. 4E). Once recruited, our data suggests that ArhGEF37 increases endocytosis rates (Fig. 4F,G) by augmenting the dynamin-dependent vesicle scission rate. Intriguingly, ArhGEF37 carries a RhoGEF domain, raising the possibility that ArhGEF37 can, analogous to other endocytotic proteins (Almeida-Souza et al., 2018; Brinas et al., 2013; Salazar et al., 2003), after actin-dynamics to augment uptake rates (Fig. 4E and Fig. S4I-L).

However, although recruitment of DYN2 and the SH3 domain of ArhGEF37 coincide spatio-temporally (Fig. 4E), pull-down assays did not identify direct binding between these proteins (Fig. S4C). Hence, our studies do not define any interactions with the canonical endocytic machinery. Finally, when considering the role of DYN2 in clathrin-independent types of endocytosis (Cao et al., 2007; Henley et al., 1998; Sauvonnnet et al., 2005), ArhGEF37 might play additional roles beyond CME. Future work will unveil the full scope of functions employed by ArhGEF37 to modulate endocytosis.

MATERIALS AND METHODS

Cell culture

HeLa cells (Leibniz Institute DSMZ, ACC-57) and NIH 3T3 embryonic fibroblasts (Leibniz Institute DSMZ, ACC-59) were cultured in Dulbecco’s modified Eagle’s medium (DMEM) containing 4.5 g/l D-glucose, GlutaMax-I and Pyruvate (31966-021, Gibco), supplemented with 10% fetal bovine serum (L11-004, BioChrom AG) and 1% penicillin-streptomycin 10 µg/ml (15140-122, Thermo Fisher Scientific). Cells were incubated at 37°C in 5% CO₂ and passaged two to three times a week.

Structure prediction and modeling

Prediction of the secondary structure of the individual domains (GEF, BAR, SH3-1 and SH3-2) was performed by PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred). For building structural models of the individual domains, we used homology modeling, a method that uses existing structures of homologous proteins (templates) to build structural models for a given sequence (query).

We first performed a search for templates using the fold prediction of pGenTHREADER (http://bioinf.cs.ucl.ac.uk/psipred). The best-fitting structures were then sorted according to the prediction certainty. Top scorers (and the alignments generated by pGenTHREADER) were then chosen for homology modeling (Table S2). By using MODELLER v9.19, 1200 homology models were generated for each individual domain, using a ‘slow’ optimization protocol followed by a ‘slow’ molecular dynamics-based refinement protocol. Next, the generated models of individual domains were ranked based on a normalized DOPE score (similar to a Z-score), with the top scoring model being the one shown in the paper. To generate a model of the BAR dimer structure of ArhGEF37, the modeled individual BAR domains were superimposed on the symmetrical units that form the dimer in the amphiphysin crystal structure (PDB ID: 1URU). The chosen homology models and the modeled BAR dimer were then solvated in a truncated octahedron water box with TIP3P water molecules, in a way that a layer of water of at least 12 Å surrounded the solute in any direction. Na⁺ ions were added to counteract the charges, as well as 150 mM KCl. The solvated systems were then subjected to a 50,000 steps energy minimization, with gradually relaxing constraints, combining steepest descent and conjugated gradient steps. All minimizations were done in explicit solvent, using the AMBER ff14SB force field model and the AMBER 18 software (ambermd.org/). Finally, by using the Adaptive Poisson Boltzmann Software (APBS; http://www.poissonboltzmann.org/), the non-linear Poisson-Boltzmann equation was solved (at a salt concentration of 0.15 M) to obtain the electrostatic potentials of the minimized structures.

Sequence alignment

Alignments of the BAR domain of ArhGEF37 and DNM1P (Fig. 1B) and the BAR domain of ArhGEF37 with its top scoring homologs (Fig. S1B) were generated using CLUSTALX (http://www.clustal.org/omega/).
**Protein purification**

C-terminal His$_{6}$-tagged ArhGEF37 (aa 1-1676) was cloned into pET21a vector and transformed into E.coli BL21-DE3 cells (70235-3, Millipore). Cells were grown in YT medium containing 16 g/l tryptone/peptone (8952.3, Carl Roth), 10 g/l yeast extract (2363.3, Carl Roth), and 5 g/l NaCl pH 7 at 37°C until an OD$_{600}$ of 0.6–0.7 reached. Protein expression was induced with 0.5 mM IPTG (R0392, Thermo Fisher Scientific) for 24 h at 18°C. Cells were then spun down at 10,000 g for 30 min at 4°C and the pellet was resuspended in lysis buffer containing 30 mM Tris-HCl (4855.2, Carl Roth), 0.5% Triton X-100 (T9284, Sigma), 45 mM imidazole (3899.2, Carl Roth), 5% glycerol (G6279, Sigma), 1:100 protease inhibitor mix (39106.01, Serva), 1% lysosyme (L4919, Sigma), 0.1% DNAse (D5025, Sigma) pH 7.5 for 15 min on ice followed by sonication. Cell lysates were precipitated by centrifugation at 30,000 g for 30 min at 4°C. The target protein was purified using immobilized metal-affinity chromatography (IMAC) by 3 Hi-Trap-chelating columns (17-0409-03, GE Healthcare) incubated with Ni$^{2+}$ using the Äkta Prime Plus system (11001313, GE Healthcare). Specifically, the protein sample was injected into the system at a flow rate of 5 ml/min and washed with IMAC running buffer containing 20 mM HEPES pH 8, 500 mM NaCl, 45 mM imidazole, 5% glycerol. The target protein was then eluted with 20 mM HEPES pH 8, 500 mM NaCl, 500 mM imidazole, 5% glycerol. The elution fractions were concentrated using Amicon Ultra-4 centrifugal filters 30K (UFC803024, Millipore) and loaded onto a HiLoad 16/600 Superdex 200 pg column (28989335, GE Healthcare) for further purification by using size exclusion chromatography. Finally, all fractions were collected in 20 mM HEPES pH 8, 500 mM NaCl, 20 mM HEPES pH 8, 500 mM NaCl, 45 mM imidazole, 5% glycerol and purity was confirmed by SDS/PAGE and western blot. The purified protein was concentrated and stored at −80°C for further studies.

**Mass spectrometry**

Sample preparation. Proteins were precipitated by adding three volumes of ice-cold acetone overnight at −20°C. After centrifugation (20,000 g for 20 min at 4°C) samples were dried and solubilized in 8 M Urea at room temperature (RT). By adding 10 mM (NH$_4$)$_2$HCO$_3$ buffer (pH 7.8) the concentration was reduced to 2 M Urea. Samples were subjected to cysteine reduction and carboxamidomethylation using 10 mM Tris (2-carboxyethyl) phosphine at 37°C for 30 min and 30 mM iodoacetamide at RT for 30 min.

Protein hydrolysis was carried out with trypsin (Promega) at a ratio of 1:100 (trypsin:protein) by incubating the samples overnight at 37°C. Tryptic digestion was stopped by adding formic acid to decrease the pH to below 1:100 (trypsin:protein) by incubating the samples overnight at 37°C. Tryptic phosphine at 37°C for 30 min and 30 mM iodoacetamide at RT for 30 min.

**Quality control**

Proteolytic digests were checked for complete hydrolyzation after desalting by using monolithic column separation (PepSwift monolithic PS-DVB PL-CA200-200-PM, Dionex) on an inert Ultimate 3000 HPLC (Dionex, Germering, Germany) by injection of 1 µg sample. A binary gradient (solvent A: 0.1% TFA, solvent B: 0.08% TFA, 84% ACN) in the range of 5-12% B in 5 min followed by 12-50% B in 15 min at a flow rate of 2.2 µl/min and at 60°C, was applied. UV traces were acquired at 214 nm.

Nano-LC-MS/MS analysis. All samples were analyzed by using an Ultimate 3000 RSLC nano system (Dionex) coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). In total, 1 µg of each sample was injected. Peptides were pre-concentrated on a 75 µm×2 cm C18 trapping column for 10 min using 0.1% TFA (v/v) at a flow rate of 10 µl/min, followed by separation on a 75 µm×50 cm C18 main column (Acclaim Pepmap, Thermo Fisher Scientific) with a 110 min LC gradient in the range of 3-45% B (84% ACN in 0.1% FA) at a flow rate of 250 nl/min. MS survey scans were acquired on the Q Exactive mass spectrometer with a 300–1500 mass-to-charge (m/z) ratio at a resolution of 60,000 using the polysiloxane ion at 371.1012 m/z as lock mass. The 15 most-intense ions were subjected to high-energy collision-induced dissociation (HCD), taking into account a dynamic exclusion of 15 s. HCD spectra were acquired with normalized collision energy of 27%. AGC target values were set to 1e6 for MS1 and 5e4 for MS2 scans and maximum injection times were set to 120 ms for MS1 and 50 ms for MS2 scans, respectively. The isolation window was set to 1.6 m/z.

Data analysis. The data analysis was performed by using proteome discoverer 1.4, using Mascot 2.6.1 as search and identification engine. The search parameters were set as follows. Protease was selected to be trypsin with full specificity and a maximum of missed cleavage sites of two. The precursor m/z tolerance was set to 10 ppm, whereas the fragment m/z tolerance was set to 0.02 Da. The searches were performed against a targeted/decoy human UniProt database (downloaded October 26th 2017) containing 40,340 sequences. Carboxamidomethylation of cysteine was set as fixed, oxidation of methionine as variable modification. The false discovery rate was set to 1%.

**Quartz crystal microbalance**

To create model membranes, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), Cholesterol (Chol), 1,2-dioleoyl-sn-glycerol-3-phosphoinositol-4,5-bisphosphate/triammonium salt (PI(4,5)P$_2$), and lipid brain extract were purchased from Avanti Polar Lipids Inc. (Alabaster, AL). The citrate buffer contained 10 mM trisodiumcitrate (Merk, Darmstadt, Germany) and 150 mM NaCl pH 4.6 (at RT). The HBS buffer consisted of 10 mM HEPES and different NaCl concentration at pH 8.0 (at RT). Solvents (HPLC grade) were purchased either from Merck, Carl Roth or AppliChem (Darmstadt, Germany). For small unilamellar vesicle (SUV) preparation, the respective lipids and cholesterol were dissolved in chloroform/methanol (1:1 v/v) except for PI(4,5)P$_2$ that was dissolved in chloroform/methanol/water (20:9:1 v/v/v). The stock solutions were mixed and the organic solvents were removed under a stream of nitrogen above the lipid gel-fluid phase transition temperature. Residual traces of the organic solvent were then evaporated in high vacuum for 4 h at the same temperature and lipid films were stored at 4°C until use. Lipid films (i.e. model membranes) of the following composition (given in molar ratios) were used: POPC/DOPC/Chol/PI(4,5)P$_2$ (60:20:10:10). POPC/DOPC/Chol (70:20:10) or the lipid brain extract. Dry lipid films were suspended in the citrate buffer at 60°C for 30 min with subsequent steps of vortex mixing every 5 min. Multilamellar vesicles (MLVs) were extruded 31 times using a polycarbonate membrane with a pore size of 50 nm (Avestin Liposofast, Ottawa, Canada) to obtain SUVs.

**Preparation of solid supported bilayers**

Silicon-coated quartz sensor (QSX 303, 50 nm SiO$_2$, 4.95 MHz) were cleaned in 2% (w/v) SDS (SERVA, Heidelberg, Germany) and hydrophobilized during a 10-min O$_2$-plasma treatment (Harrick Plasma, Ithaca, NY). Subsequently, surfaces were rinsed with ultra-pure water and dried in a stream of nitrogen. To prepare the supported bilayer membrane on the quartz sensor by vesicle rupture, SUVs (0.5 mg/ml, 50 nm) were applied during quartz crystal microbalance with dissipation (QCM-D) measurements. Following membrane formation, the citrate buffer was exchanged to a HBS-Puffer pH 8.0 containing different salt (NaCl) concentrations between 125 mM and 1 M.

**QCM-D measurements**

Quartz crystal microbalance with dissipation (QCM-D) measurements were performed on a Q-Sense E4 QCM-D Analyzer (Q-Sense, Gothenburg, Sweden) equipped with four temperature-controlled flow cells in a parallel configuration. Flow cells were connected to a peristaltic pump (Ismatec IPC, Glattbrugg, Switzerland) employing a flow rate of 80.4 µl/min. Binding analysis was performed at 20°C in HEPES-buffered saline supplemented with NaCl at different concentrations (125 mM to 1 M). Frequency and dissipation shifts of the third overtone resonance frequency of the quartz sensor (QSX 303, 50 nm SiO$_2$, 4.95 MHz) were monitored and considered for data evaluation. Calculations were carried out using OriginPro v. 9.1 (OriginLab Corp., Northampton, MA).

**SDS/PAGE and western blots**

Brain tissue lysates (PK-AB718-1403-0/7/14, Proma Kine) and other protein samples were diluted 1:1 with 2×Laemmli buffer containing 125 mM Tris-HCl, 0.005% Brilliant Blue (19598.1, Carl Roth), 20% glycerol, 4% SDS (20765.3, Serva) and 5% β-mercaptoethanol (M-7154, Sigma), incubated for 2 min at 92°C and loaded onto 12% polyacrylamide gel.
DNA plasmids

Full-length ArhGEF37 (GEF) (aa 1-676), as well as deletion mutants ArhGEF37 (GEP) (aa 1-263), ArhGEF37 (BAR) (aa 215-509) and ArhGEF37 (SH3) (aa 452-676) were PC-amplified from mouse cDNA (ORF clone MG222459, Origene), and subsequently cloned into Vivid Color-pcDNA 6.2-C-YFP-DEST vector (V357-20, Life Technologies). All plasmids were verified by DNA sequencing before use. The following commercially available mammalian expression vectors were used: FCH02-pmCherryC1 (Addgene plasmid #27686), CLTA-mCherry (Addgene plasmid #27680), DYNO2-pmCherryN1 (Addgene plasmid #27689), FBPl7-pmCherryC1 (Addgene plasmid #27688), APPL1-pmCherryC1 (Addgene plasmid #27683) all described by Taylor et al., 2011; DYNN2 (K44A)-GFP (Addgene plasmid #22301) from Ochoa et al., 2000, pmCherry Paxillin- (Addgene plasmid #50526) from Kenneth Yamada, Amphiphysin1 (our own, Galic et al., 2012), and cytosol-CFP (our own, Galic et al., 2014). For bacterial expression, C-terminal His-tagged ArhGEF37 (Genebank NM_001001669) was cloned into pET21a vector (Promab Biotechnologies, Inc., Richmond, CA).

siRNA knockdown

All siRNAs used in this study were 19-mers including 5′-dTdT sense-5′. Scrambled control siRNA used in all experiments was 5′-GGAGGAAG-GAGATAGTAAATATTT-3′. Transfection

For transient overexpression of ArhGEF37, deletion mutants and other endocytic proteins, HeLa and NIH 3T3 fibroblast cells were transfected with plasmids 24 h prior to analysis using FuGENE HD transfection reagent (E2311, Promega) according to described protocol. To achieve optimal knockdown efficiency, HeLa cells were transfected with all three siRNAs together by using Lipofectamine 2000 (11668-019, Invitrogen) according to the manufacturer’s protocol and analyzed after 48 h.

Immunocytochemistry

Cells were fixed in 4% paraformaldehyde (18505, Ted Pella) containing 4% sucrose (S7903, Sigma) in 1×PBS (phosphate-buffered saline, 10010-02, Sigma) for 10 min at RT. Then the cells were kept on ice for 10 min, followed by washing twice with ice-cold acidic buffer (0.1 M glycine, 150 mM NaCl pH 3) and twice with ice-cold 1×PBS. Cells were dissociated with 6 mM EGTA and centrifuged using a table-top centrifuge at 1000 rpm (100 g) for 3 min at 4°C. Cell pellets were re-suspended in 6 µl ice-cold dilution buffer and washed thrice by using the DynaMag-2 magnet (12321D, Thermo Fisher Scientific). For pull down, previously equilibrated beads were incubated with 450 µl of cell lysates at 4°C for 1 h under rotating conditions. Following the beads, were then magnetically separated using DynaMag-2 and washed thrice with ice-cold dilution buffer. At last, beads were suspended in 2× sample buffer for further analysis by SDS/PAGE and western blotting.

Fluorescence microscopy

Images were captured using an EMCCD camera (IXON Ultra, DU-888U3-CSO-BV, Andor), 1024×1024 pixel, 13 µm×13 µm pixel size, mounted on the side port of an inverted microscope (Nikon Eclipse Ti). The setup was equipped with a Yokogawa CSU-X1 spinning-disc scanning unit; 60× or 63× objective were used. Lasers for excitation were used at 445 nm for Alexa-Fluor®488 serum albumin, A9085.25G, Sigma) for 15 min at RT. Cells were then washed with ice-cold 1×PBS containing 2.5% BSA (bovine serum albumin, A9085.25G, Sigma) for 15 min at RT. Then the cells were kept on ice for 10 min, followed by incubation with 30 µg/ml Alexa-Fluor 647 (T23366, Molecular Probes) for 15 min at 4°C. The cells were stopped with 2× sample buffer for further analysis by SDS/PAGE and western blotting.

Endocytosis inhibition

Cells were initially cultured in DMEM and then transfected into live-cell imaging solution (LCIS) at equal osmolarity. To inhibit endocytosis, cells were incubated either with 80 µM Dynasore (D7693, Sigma), 30 µM Dyngo-4A (ab120689, Abcam) or 30 µM Pitstop-2 (ab120687, Abcam) for 30 min at 37°C. 0.04% dimethylsulfoxide (DMSO), D2650, Sigma) was used as a vehicle control. For quantification, individual cells were imaged before (pre) and after (post) incubation with drugs. Cells expressing DYNN2 (K44A) were fixed 24 h after transfection.

RNA extraction and qPCR

Total ribonucleic acid (RNA) from cells was extracted using Qiagen RNeasy Mini kit (74104, Qiagen) according to the manufacturer’s protocol. First-strand cDNA was synthesized from 1 µg RNA with Thermo Fisher Scientific’s SuperScript III Reverse Transcriptase System (180080093) in 20 µl final volume reaction mix containing 30 ng random primers (48190-011, Thermo Fisher Scientific), 1 µl 10 mM dNTPs (18427-013, Thermo Fisher Scientific), 1 µl 0.1 M DTT; 4 µl 5× first-strand buffer, 1 µl RNase OUT (10777-019, Thermo Fisher Scientific) and 1 µl SuperScript III (200 U/µl). Quantitative real-time polymerase chain reaction (qPCR) was performed using a BioRad Myq single color real-time PCR detection.
system with KAPA SYBR FAST Bio-Rad iCycler (KK4606, Sigma) for the real-time amplification detection. Specific primers were designed and tested for qPCR analysis (Table S4).

Quantification and statistical analysis

Image analysis

All images (unless otherwise mentioned) were processed using Arous wavelet filtering (Oliveo-Marín, 2002), and a custom-built MATLAB script (www.mathworks.com) (Fig. S2A,B). For calculation of the overlap percentage, wavelet transformed channels were first binarized and subsequently processed using logical ‘OR’ and ‘AND’ operators, thus resulting in ‘sum’ and ‘overlap’ images, respectively. Finally, the overlap percentage was calculated. For temporal analysis, wavelet transformed stacks were binarized, followed by ‘OR’ operation, therefore generating superimposed images of both channels. Next, an ‘AND’ operation was performed and a mask was generated, representing the ROIs. Finally, an object detection algorithm was used to index island like structures (i.e. ROIs) in the mask. Corresponding to each ROI, mean intensity values were acquired from the original stacks and temporal cross-correlation was calculated (Fig. S2D,E). Software used for image analysis is available on our homepage (https://www.medizin.uni-muenster.de/en/imphd/das-institut/nanoscale-forces-in-cells/software/).

qPCR analysis

For relative quantification of gene expression, expression of two reference genes (i.e. OAZ1 and RPS13) were measured and used to normalize all qPCR data. Gene expression was calculated with REST software (Pfaffl et al., 2004) according to the following equation:

\[
\text{Ratio} = \frac{(E_{\text{target}} \times CP_{\text{target}} \times 2^{-\Delta\text{ACT}})}{(E_{\text{ref}} \times CP_{\text{ref}} \times 2^{-\Delta\text{ACT}})}
\]

Where E is the efficiency, and the target is ArhGEF37; ref is OAZ1 and RPS13, and CP is the respective crossing point.

Statistics

Experiments are composed of at least three biological repeats, unless stated otherwise in the figure legends. Statistical calculations were accomplished using GraphPad Prism, version 5.03. To test for normal distributions, the Kolmogorov–Smirnov test was applied. To test for equal variance, the F-test was performed for statistics. Following a significant F-test, the Student’s t-test was used. Box-plots represent median, interquartile range and percentage, wavelet transformed channels were first binarized and

Supplementary information

Supplementary information available online at http://jcs.biologists.org/lookup/doi/10.1242/jcs.226530/supplemental

References


d


**Fig. S1. ArhGEF37 structure and membrane binding analysis.**

**(A)** Sequence alignment of BAR domains from ArhGEF37 and different homologs (indicated). Colour-code indicates conserved hydrophobic (blue) and non-hydrophobic (grey) residues.  

**(B)** Modelled domains of individual ArhGEF37 domains. Corresponding structure models depicting electrostatic surface potential (red, $-10kT/e$, blue, $+10kT/e$) are shown next to it.  

**(C)** Complete Coomassie and Western blots (corresponding to Fig. 1E).  

**(D)** From left to right, measurements of ArhGEF37 binding to POPS/DOPC/Chol/Pl(4,5)P$_2$, POPC/DOPC/Chol and Brain lipid extract (corresponding to Fig. 1E).  

**(E)** From left to right, measurements of ArhGEF37 binding to POPS/DOPC/Chol/Pl(4,5)P$_2$ at 1000 mM, 250 mM and 125 mM, respectively (corresponding to Fig. 1E).
**Fig. S2. Image analysis tools to study ArhGEF37 function.** (A) Scheme depicting the principle of à trous wavelet filtering. In brief, the filter divides the input image into wavelet planes based on different frequencies of signals in the image. The 1st order wavelet plane consists of noise (highest frequency), followed by wavelet planes with decreasing frequencies associated with increasing structure size in the spatial domain. By combining the 2nd and 3rd wavelet planes, followed by a subsequently thresholding, the final output image is generated. (B) Quality controls for à trous wavelet filtering analysis. To the left, raw images of cells transfected with fluorescently labelled Dynamin2 (DYN2), transferrin and ArhGEF37, respectively. In the middle, result of à trous wavelet filtering for respective images (red). To the right, overlay of raw data (grey) and filter results (red). (C) Line-scan (left) and overlay of 100 individual puncta (right) suggest diffraction-limited ArhGEF37 aggregates. (D) Scheme showing workflow of custom made script for temporal cross-correlation analysis. In brief, individual puncta were isolated, and intensity profiles measured along the z-axis (i.e. time). To determine the temporal order of appearance, intensity profiles within individual puncta (green and red) are then cross-correlated. (E) In silico testing of temporal cross-correlation analysis. Published data on recruitment kinetics of CTLA, DYN2 and Snx9 was used to generate intensity traces. Note that the software accurately predicts that CTLA precedes DYN2 (blue) while Snx9 follows DYN2 (red). Scale bar, (B) 5 µm.
**Fig. S3. Perturbation controls for ArhGEF37 recruitment.** (A) No pattern formation in cytosol after Dynasore addition. To the left, cell expressing cytosolic CFP before (pre) and after (post) exposure to Dynasore. To the right, analysis for change in dots/µm² upon addition of DMSO (0.00±0.03 dots/µm², black, n = 12 cells, mean±SD) and Dynasore (0.00±0.01 dots/µm², red, n = 13 cells, mean ± SD). (B) DMSO does not trigger increased co-localization of DYN2 and ArhGEF37. Cells were transfected with ArhGEF37 (green) and DYN2 (magenta). Note no significant change in total overlap percentage upon DMSO addition (pre: 21 ± 10%, black; post: 27 ± 13%, red; n = 11 cells, mean ± SD). (C) Specificity test for ArhGEF37 antibody. Cells were transfected with ArhGEF37 for 24 hours, fixed and stained with antibody directed against ArhGEF37. Following wavelet-transformation, overlap percentage was determined (62±19%, n = 19). Note: since antibody directed against ArhGEF37 yields multiple bands of various molecular weights on Western blot (Supplemental Fig. 4G), unspecific binding partners need to be considered. (D) Expression of dominant-negative DYN2(K44A) yields an increase in ArhGEF37 antibody signal. Cells transfected with DYN2(K44A) for 24 hours were fixed and stained with antibody directed against ArhGEF37. Following wavelet-transformation, puncta density was determined for transfected (0.08±0.02 dots/µm²; n = 35 cells) and non-transfected (0.04±0.02 dots/µm², n = 18 cells) cells. (E) Temperature shift changes Dyngo4A-dependent DYN2 recruitment. Cells were transfected with DYN2 for 24 hours and incubated for 30 minutes with Dyngo-4A at room temperature. Note differences in enrichment of DYN2 to the membrane compared to 37 °C (shown in Fig. 3A). Scale bars, (A, E) 10 µm; (B) 5 µm; (C, D) 20 µm.
Fig. S4. Localization controls and ArhGEF37 knockdown validation.

(A) Hyperosmotic shock does not change cytosolic signal. Cells expressing cytosolic CFP before (pre) and after (post) hyperosmotic shock. Below, quantification of particles (dots/µm²) is shown (n = 11 cells). (B) Co-localization for different ArhGEF37 domains and CLTA. Left, cells co-expressing indicated ArhGEF37 constructs (green) and CLTA (magenta). Middle panel depict spatial cross-correlation analysis of truncated versions of ArhGEF37 vs. CLTA. To the right, overlap percentage for CTLA and the GEF (5 ± 4%, n = 21 cells), BAR (7 ± 3%, n = 16 cells) and SH3 (20 ± 9%, n = 18 cells) domains (Mean ± SD). Blue boxes in the graphs are added as guidance to the eye. (C) Pulldown assay yields no apparent interaction between ArhGEF37 and Dynamin2. From left to right, cell lysate (left), pulldown (middle) and supernatant (right) of cells transfected with full length ArhGEF37 (lane 1), the isolated SH3 domain of ArhGEF37 (lane 2), and the isolated BAR domain of ArhGEF37 (lane 3). Upon pulldown, samples were loaded on gel and stained with antibodies directed against GFP (top gel) and Dynamin2 (bottom gel), respectively. (D) Representative image of cells co-transfected with siRNA directed against ArhGEF37 (siRNA_{ArhGEF37}) and fluorescence marker (green). Graph depicts transfection efficiency (65 ± 3% n = 4 technical repeats, mean ± SD). (E) Analysis of ArhGEF37 knockdown efficiency via qPCR. 48 hours post transfection, cells transfected with siRNA_{ArhGEF37} yield 31± 9% (S1), 53± 23% (S2) and 26± 40% (S3) reduction in mRNA levels (3 biological repeats, each with n = 3 technical repeats, median ± SD). Note that transfection efficiency of 65 ± 3% needs to be considered. (F) Knockdown of ArhGEF37 yields reduced ArhGEF37 immunofluorescence signal. To the left cells transfected with siRNA\textsubscript{control} (black, top) or siRNA\textsubscript{ArhGEF37} (red, bottom) are shown. Difference in immunofluorescence in control (21 ± 11 au, black, n = 17...
cells) and ArhGEF37 knockdown (12 ± 3 au, red, n = 19 cells) are shown to the right.
(Mean ± SD; Mann Whitney test; *p ≤ 0.05). (G) Western blot analysis of ArhGEF37
knockdown efficiency. Cells were transfected for 48 hours with siRNAcontrol or
siRNAArhGEF37. Upon isolation, protein samples were loaded and stained with antibody
directed against ArhGEF37 (top) and tubulin (bottom). Two separate biological
repeats yield signal reduction of 38% (left) and 22% (right), respectively. Again, note
that transfection efficiency of 65±3% needs to be considered. (H) FACS analysis
(10’000 cells/condition) depicting signal separation for GFP-transfected cells (black)
vs. non-transfected cells (blue). Threshold used in Fig. 4G for gating (dashed red
line) was set at an arbitrary grey value of 670, yielding <1% false positives. (I)
ArhGEF37 retains DYN2 at endocytotic sites. Cells co-transfected with siRNAcontrol
(black) or siRNA ArhGEF37 (red) and fluorescently tagged DYN2 before (top) and after
(bottom) hyperosmotic shock. To the right, kinetics and statistical analysis (blue) are
shown. As above, bold lines depict the median, thin lines individual experiments (n =
14 cells for siRNAcontrol and siRNAArhGEF37, respectively). (J) Knockdown of ArhGEF37
reduces transferrin uptake and area of adherent cells. Cells transfected with
fluorescence marker and control siRNA or siRNA directed against ArhGEF37 after
incubation with transferrin-Alexa647 for 10 minutes at 37 °C. To the right, scatter plot
depicting cell area vs. transferrin uptake for siRNAcontrol (black, n = 22 cells) and
siRNAArhGEF37 (red, cells, n = 29 cells), respectively. Below, quantification of
transferrin uptake for siRNAcontrol (0.14±0.02 particles/µm², black) and siRNAArhGEF37
(0.10±0.02 particles/µm², red), respectively, show a slight but significant reduction in
transferrin uptake. Likewise, quantification of cell area upon transfection with
siRNAcontrol (2216±490 µm², black) and siRNAArhGEF37 (1585±446 µm², red) show a
significant difference in cell area. (K) Knockdown of ArhGEF37 does not change cell volume in FACS analysis. Cells co-transfected with fluorescence marker and siRNAcontrol (blue) or siRNAArhGEF37 (black). Forward scatter does not yield apparent changes in cell volume. (L) The SH3 domain of ArhGEF37 localizes to paxillin-positive focal adhesions. Cells were co-transfected with the isolated SH3 domain of ArhGEF37 (green) and the focal adhesion protein paxillin (magenta). Note slight enrichment of the SH3 domain of ArhGEF37 at focal adhesions. Statistics: * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001; Mann Whitney test; Error bars represent SEM. Scale bars, (A, I) 10 µm; (B, L) 20 µm, 5 µm; (D) 50 µm; (F) 20 µm.
Movie 1. ArhGEF37 recruitment analysed via à trous wavelet filtering. HeLa cell transfected with ArhGEF37. To the left, raw data of basal PM. In the middle, same movie subjected to à trous filtering. To the right, overlay of raw (grey) and filtered (red) data. Frames captured at 1 Hz. Scale bar, 10 µm.
Movie 2. DYN2 recruitment to the PM increases upon hyperosmotic shock. HeLa cell transfected with DYN2 and subjected to hyperosmotic shock (star). Individual frames captured at 1 Hz. Scale bar, 10 µm.
Movie 3. ArhGEF37 recruitment to the PM increases upon hyperosmotic shock.
HeLa cell transfected with ArhGEF37, followed by hyperosmotic shock (star). Frames were taken at 1 Hz. Scale bar, 10 µm.
Movie 4. No change in cytosolic perturbation after hyperosmotic shock.

HeLa cell transfected with cytosolic marker and subjected to hyperosmotic shock (star). Individual frames were taken at 1 Hz. Scale bar, 10 µm.
Movie 5. PM recruitment of BAR and SH3 domain of ArhGEF37 increase after hyper-osmotic shock. HeLa cells transfected with truncated versions of ArhGEF37 (GEF; BAR, SH3), and subjected to hyperosmotic shock (star). Individual frames were taken at 1 Hz. Scale bar, 10 µm.

Movie 6. 3D model depicting potential sites for ArhGEF37 enrichment during late phase of CME. Clathrin (blue), filamentous actin (green), Dynamin (yellow), and ArhGEF37 (red) are shown.
Table S1. Protein secondary structure prediction. Graphical representation of secondary structures predicted via PSIPRED for the following ArhGEF37 domains: GEF domain (A), BAR domain (B) and the two SH3 domains (C and D). For every amino acid, blue bars indicate confidence of the prediction. Black lines (letter code C), yellow arrows (letter code E) or purple cylinders (letter code H) indicate a predicted coil region, β-strand or α-helix, respectively.
<table>
<thead>
<tr>
<th>Domain</th>
<th>PDB ID</th>
<th>Homologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEF</td>
<td>3EO2</td>
<td>RhoGEF domain of h-neuroepithelial cell transforming gene-1</td>
</tr>
<tr>
<td></td>
<td>3GF9</td>
<td>Crystal structure of Human intersectin 2 RhoGEF domain</td>
</tr>
<tr>
<td></td>
<td>1F5X</td>
<td>NMR structure of the Y174 autoinhibited DBL homology domain</td>
</tr>
<tr>
<td>BAR</td>
<td>4AVM</td>
<td>Crystal structure of N-BAR domain of h-bridging integrator 2</td>
</tr>
<tr>
<td></td>
<td>2FIC</td>
<td>Crystal structure of BAR domain from human Bin1/Amphiphysin II</td>
</tr>
<tr>
<td></td>
<td>1URU</td>
<td>Amphiphysin BAR domain from <em>Drosophila</em></td>
</tr>
<tr>
<td></td>
<td>4ATM</td>
<td>Crystal structure of BAR domain of human Amphiphysin, isoform 1</td>
</tr>
<tr>
<td></td>
<td>3CAZ</td>
<td>Crystal structure of a BAR protein from Galdieria sulphuraria</td>
</tr>
<tr>
<td>SH3-1</td>
<td>1UHC</td>
<td>Solution structure of RSG1 RUH-002, SH3 domain of human DNMBP/ K1AA1010 protein</td>
</tr>
<tr>
<td></td>
<td>1UG1</td>
<td>SH3 domain of hypothetical protein BAA76854.1</td>
</tr>
<tr>
<td></td>
<td>1SEM</td>
<td>Structural determinants of peptide-binding orientation &amp; of sequence specificity in SH3 domains.</td>
</tr>
<tr>
<td>SH3-2</td>
<td>1UHC</td>
<td>Solution structure of RSG1 RUH-002, SH3 domain of human DNMBP/ K1AA1010 protein</td>
</tr>
<tr>
<td></td>
<td>2KBT</td>
<td>Attachment of an NMR-invisible solubility enhancement tag</td>
</tr>
<tr>
<td></td>
<td>1OOT</td>
<td>SH3 domain from a <em>S. cerevisiae</em> hypothetical 40.4 kDa protein</td>
</tr>
<tr>
<td></td>
<td>2K9G</td>
<td>Solution structure of 3rd SH3 domain of the Cin85 adapter protein</td>
</tr>
</tbody>
</table>

**Table S2. Structures used for homology modeling.** For each domain, both the PDB code and the name of the used homologs are indicated.
<table>
<thead>
<tr>
<th>Conf.</th>
<th>Net Score</th>
<th>P-value</th>
<th>Homology</th>
<th>Conf.</th>
<th>Net Score</th>
<th>P-value</th>
<th>Homology</th>
<th>Conf.</th>
<th>Net Score</th>
<th>P-value</th>
<th>Homology</th>
<th>Conf.</th>
<th>Net Score</th>
<th>P-value</th>
<th>Homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>CERT</td>
<td>93.553</td>
<td>2e-08</td>
<td>3e02</td>
<td>CERT</td>
<td>80.053</td>
<td>4e-07</td>
<td>4e-04</td>
<td>CERT</td>
<td>80.423</td>
<td>4e-07</td>
<td>2e02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>91.089</td>
<td>3e-08</td>
<td>3f00</td>
<td>CERT</td>
<td>77.736</td>
<td>7e-07</td>
<td>1e01</td>
<td>CERT</td>
<td>77.724</td>
<td>7e-07</td>
<td>4e-07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>90.790</td>
<td>3e-08</td>
<td>2e2k</td>
<td>CERT</td>
<td>66.051</td>
<td>1e-05</td>
<td>3e02</td>
<td>CERT</td>
<td>54.144</td>
<td>2e-04</td>
<td>2e12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>88.703</td>
<td>6e-08</td>
<td>15x</td>
<td>CERT</td>
<td>51.949</td>
<td>3e-04</td>
<td>12xw</td>
<td>CERT</td>
<td>49.056</td>
<td>5e-04</td>
<td>4e14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>84.763</td>
<td>1e-07</td>
<td>1x7</td>
<td>CERT</td>
<td>49.475</td>
<td>4e-04</td>
<td>14d</td>
<td>CERT</td>
<td>45.234</td>
<td>0.001</td>
<td>2rj</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>83.877</td>
<td>2e-07</td>
<td>15xoe</td>
<td>CERT</td>
<td>44.748</td>
<td>0.002</td>
<td>4sxe</td>
<td>CERT</td>
<td>41.585</td>
<td>0.003</td>
<td>3fhn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>82.078</td>
<td>2e-07</td>
<td>2x2r</td>
<td>CERT</td>
<td>39.811</td>
<td>0.005</td>
<td>2q12</td>
<td>CERT</td>
<td>43.403</td>
<td>0.002</td>
<td>2zsw</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>80.679</td>
<td>4e-07</td>
<td>5q94</td>
<td>CERT</td>
<td>39.102</td>
<td>0.006</td>
<td>2q13</td>
<td>CERT</td>
<td>41.055</td>
<td>0.003</td>
<td>3f8n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>79.170</td>
<td>5e-07</td>
<td>2p17</td>
<td>CERT</td>
<td>39.464</td>
<td>0.005</td>
<td>1y5y</td>
<td>CERT</td>
<td>39.176</td>
<td>0.006</td>
<td>3l2w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>78.270</td>
<td>6e-07</td>
<td>3r90</td>
<td>CERT</td>
<td>39.102</td>
<td>0.006</td>
<td>2q13</td>
<td>CERT</td>
<td>38.656</td>
<td>0.006</td>
<td>4h5s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>78.029</td>
<td>7e-07</td>
<td>1x01</td>
<td>CERT</td>
<td>38.172</td>
<td>0.007</td>
<td>3f9g</td>
<td>CERT</td>
<td>37.594</td>
<td>0.007</td>
<td>3p3y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>77.788</td>
<td>7e-07</td>
<td>1xq1g</td>
<td>CERT</td>
<td>39.046</td>
<td>0.005</td>
<td>1y5y</td>
<td>CERT</td>
<td>37.818</td>
<td>0.008</td>
<td>2e9f</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>75.961</td>
<td>1e-06</td>
<td>3y99</td>
<td>CERT</td>
<td>38.172</td>
<td>0.007</td>
<td>3f9g</td>
<td>CERT</td>
<td>37.373</td>
<td>0.009</td>
<td>3k2w</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>75.921</td>
<td>1e-06</td>
<td>1y1y</td>
<td>CERT</td>
<td>37.594</td>
<td>0.007</td>
<td>3p3y</td>
<td>CERT</td>
<td>35.000</td>
<td>0.000</td>
<td>3m3y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>74.619</td>
<td>1e-06</td>
<td>4g1o</td>
<td>CERT</td>
<td>37.818</td>
<td>0.008</td>
<td>2e9f</td>
<td>CERT</td>
<td>35.000</td>
<td>0.000</td>
<td>3m3y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>73.090</td>
<td>2e-06</td>
<td>1y1 1</td>
<td>CERT</td>
<td>38.172</td>
<td>0.007</td>
<td>3f9g</td>
<td>CERT</td>
<td>35.000</td>
<td>0.000</td>
<td>3m3y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>72.855</td>
<td>2e-06</td>
<td>4g2o</td>
<td>CERT</td>
<td>37.818</td>
<td>0.008</td>
<td>2e9f</td>
<td>CERT</td>
<td>36.058</td>
<td>0.005</td>
<td>2d1x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>58.196</td>
<td>7e-05</td>
<td>3mpx</td>
<td>CERT</td>
<td>37.373</td>
<td>0.009</td>
<td>3k2w</td>
<td>CERT</td>
<td>36.354</td>
<td>0.005</td>
<td>4xv5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>58.427</td>
<td>2e-04</td>
<td>1b1</td>
<td>CERT</td>
<td>37.000</td>
<td>0.000</td>
<td>3m3y</td>
<td>CERT</td>
<td>40.312</td>
<td>0.004</td>
<td>3m1z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>51.539</td>
<td>3e-04</td>
<td>3k5y</td>
<td>CERT</td>
<td>37.000</td>
<td>0.000</td>
<td>3m3y</td>
<td>CERT</td>
<td>40.058</td>
<td>0.005</td>
<td>2d1x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>52.702</td>
<td>0.025</td>
<td>39l7</td>
<td>CERT</td>
<td>36.010</td>
<td>0.012</td>
<td>4y6h</td>
<td>CERT</td>
<td>40.048</td>
<td>0.005</td>
<td>2v5f</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>52.144</td>
<td>0.025</td>
<td>4m17c</td>
<td>CERT</td>
<td>36.010</td>
<td>0.012</td>
<td>4f1h</td>
<td>CERT</td>
<td>39.937</td>
<td>0.005</td>
<td>2y3a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>51.665</td>
<td>0.025</td>
<td>3w9o</td>
<td>CERT</td>
<td>36.002</td>
<td>0.012</td>
<td>4e2z</td>
<td>CERT</td>
<td>39.883</td>
<td>0.005</td>
<td>2c2e</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>31.659</td>
<td>0.032</td>
<td>3w2v</td>
<td>CERT</td>
<td>35.573</td>
<td>0.013</td>
<td>2r17</td>
<td>CERT</td>
<td>36.697</td>
<td>0.005</td>
<td>2a2y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>31.364</td>
<td>0.035</td>
<td>3p9y</td>
<td>CERT</td>
<td>35.470</td>
<td>0.013</td>
<td>2f1j</td>
<td>CERT</td>
<td>35.393</td>
<td>0.005</td>
<td>1u1t</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CERT</td>
<td>31.364</td>
<td>0.035</td>
<td>3p9y</td>
<td>CERT</td>
<td>35.313</td>
<td>0.005</td>
<td>1n2g</td>
<td>CERT</td>
<td>35.473</td>
<td>0.005</td>
<td>2a8k</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conf.: The hit confidence category
Net Score: The GenTHREADER raw score

Table S3. Fold-based homolog search. A list of homologs found using the fold-based homolog searching server pGenTHREADER. For every individual domain, the 30 top scoring hits are detailed with indications of the level of confidence, the pGenTHREADER Net Score, the p-value and the PDB accession code.
<table>
<thead>
<tr>
<th>Gene</th>
<th>RefSeq access no.</th>
<th>Forward primer (5´-3´)</th>
<th>Reverse primer (5´-3´)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArhGEF37</td>
<td>NM_001001669</td>
<td>ATCCTCCAGGTCAGGGAGT</td>
<td>TGCAGGCAACTGCTGGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGAACCAGGTCATAGCCGC</td>
<td>CCAAGAAAGCCAGAAGGCACA</td>
</tr>
<tr>
<td>OAZ</td>
<td>NM_004152</td>
<td>TAACTGGCCAACAGTGCTGA</td>
<td>ATGAAGACATGGTGCCGCTCG</td>
</tr>
<tr>
<td>RPS13</td>
<td>NM_001017</td>
<td>GCTCTCCTTTCCGTCCCTGA</td>
<td>TAGGGTAAGCCGATGGGA</td>
</tr>
</tbody>
</table>

**Table S4. List of forward and reverse primers used for qPCR.** OAZ stands for Ornithine decarboxylase antizyme, and RPS13 for Ribosomal protein S13.