Discovery of a new evolutionarily conserved short linear F-actin binding motif

Themistoklis Paraschiakos[1*], Biao Yuan[2,5, 6, 3*], Kostiantyn Sopelniak[4], Michael Hecht-Bucher[1], Lisa Simon[1], Ksenija Zonjic[1], Dominic Eggers[1], Franziska Selle[1], Jing Li[1], Stefan Linder[4], Thomas C. Marlovits[2, 5, 6], and Sabine Windhorst[1]

[1]Department of Biochemistry and Signal Transduction, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, D-20246 Hamburg

[2] Institute of Microbial and Molecular Sciences, University Medical Center Hamburg-Eppendorf, Notkestrasse 85, D-22607 Hamburg

[4] Institute for Microbiology, Virology and Hygiene, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, D-20246 Hamburg[4]

[5] CSSB-Center for Structural Systems Biology, Notkestrasse 85, D-22607 Hamburg

[6] Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, D-22607 Hamburg

[3] current address: Molecular Structural Biology, Helmholtz Center for Infection Research, Inhoffenstraße 7, 38124, Braunschweig

*These authors contributed equally to this work

Address correspondence to:

Sabine Windhorst

Department of Biochemistry and Signal Transduction,

University Medical Center Hamburg-Eppendorf,

Martinistrasse 52,

D-20246 Hamburg

Germany

Tel. +49-40-7410-56013

E-Mail: s.windhorst@uke.de

Abstract

Regulation of the actin cytoskeleton by actin binding proteins (ABPs) is essential for cellular homeostasis, and the mode of actin binding determines the activity of ABPs. Here, we discovered a novel “Short linear F-actin binding motif (SFM)” on the basis of the cryo-EM structure of the ITPKA-F-actin complex. We developed the computational pipeline SLiMFold, which identified 103 human SFM containing-proteins exhibiting diverse cellular functions. The SFM probably developed *ex nihilo* and remained conserved in eukaryotes, with a binding affinity to F-actin ranging from 13 to 89 µM. Furthermore, we uncovered the essential amino acids of this SFM for F-actin binding and affinity modulation. Together, the SFM seems to serve as a low affinity anchor to target proteins to F-actin, in order to connect the regulation of actin dynamics with broad cellular functions. These findings will shed new light on the role of a wide variety of proteins.

Introduction

Actin filaments (F-actin) form a dynamic cytoskeletal framework in eukaryotic cells, driving essential processes such as cell motility, division, and morphogenesis. These processes are orchestrated by different interacting proteins, referred to as actin-binding proteins (ABPs), along with signalling and scaffolding proteins.[1] These proteins assist in polymerisation, nucleation, elongation or capping, severing, crosslinking, or bundling of actin filaments, and are required to coordinate spatially restricted actin turnover. Their impact on actin dynamics depends on the affinity of the actin-binding domain (ABD) and, in most cases, on cellular stimulation[2]. Many ABPs are inhibited in resting cells, and a lot of them are activated by small GTPases, RhoA, Cdc42, or Rac[3]. Failures in the regulation or expression of ABPs as well as mutations can lead to severe diseases, such as neuronal and immunological disorders and cancer[4–8].

Among the ABPs, a lot of conserved domains and motifs have been identified, upon which ABP-families with similar functions were defined. Supplementary Table 1 illustrates how the remarkable variety of motifs and domains matches the diversity of cellular actin functions. Despite this diversity, a lot of ABPs bind to a hydrophobic cleft between actin subdomain 1 and 3, such as cofilin-1, T-plastin, myosin and gelsolin[9]. Here, the binding modus and/or the affinity is different, resulting in competition of the ABPs to each other in response to cellular stimulation or under certain pathological conditions[10,11].

Alongside globular domains with properly folded three-dimensional structure, such as that of cofilin-1, T-plastin, myosin and gelsolin, unstructured regions have been shown to play an essential role in many dynamic cellular processes[12–21]. Among these, the most common functional modules are short amino acid stretches known as short linear motifs (SLiMs) [22–25]. The abundant G-actin binding WH2 (Wiskott-Aldrich homology 2) motif is a prominent example of a SLIM, and structure elucidation revealed that it binds to G-actin at the hydrophobic cleft between actin subdomains 1 and 3, thereby modulating polymerisation and filament dynamics[26]. Missense mutations located inside the WH2 motif are associated with diseases,

i.e. the human Leiomodin-3 (LMOD3), where mutations R543L[27] or L550F[28] are associated with Nemaline myopathy 10 (NEM10). Moreover, WH2 motifs can also be part of bacterial effector proteins, often leading to restructuring of the host cell actin cytoskeleton and promoting either efficient colonisation (VopL/VopF) or bacterial motility during infection (Sca2 and RickA)[29–32].

In this study, we identify a novel short linear F-actin binding motif (SFM) that mediates F-actin binding in diverse human proteins with low affinity. Comparative sequence analyses indicate that the SFM is under purifying selection and acts as a low-affinity anchor, targeting various cytosolic proteins to the actin cytoskeleton. Notably, our findings not only unify previously observed yet unclassified sequences under this defined motif, but also enable the prediction of new SFM-containing proteins. These observations point to a broader role for the SFM in cytoskeletal organization under both physiological and pathophysiological conditions and may uncover new cellular actin networks and mechanisms.

Results

Structural insight into ITPKA in F-actin binding

A high number of actin-binding proteins (ABPs) contain uncharacterized domains responsible for F-actin binding, including the human Inositol-trisphosphate 3-kinase A (ITPKA)[10,33–36]. This $InsP_3$Kinase isoform constitutively binds to F-actin, and due to homodimer formation it exhibits F-actin bundling activity[36]. Under physiological conditions, ITPKA contributes to control the postsynaptic actin architecture of hippocampal neurons. However, ITPKA is also expressed in different malignant tumor cells, and its actin bundling activity induces the formation of cellular protrusions essential for invasion of cancer cells[34,37].

To investigate the molecular basis of ITPKA for F-actin binding in detail, we solved the cryo-EM structure of the ITPKA-F-Actin complex at 2.97 Å resolution. In this structure, the F-actin binding domain (ABD) of ITPKA is clearly visible, whereas the catalytic domain and the linker

connecting it to the ABD were not resolved (Figure 1a, and Extended Data Figure 1), confirming that the C-terminal $InsP_3$Kinase domain is not directly involved in F-actin binding. The amino acids Arg28 to Ala49 bind to the hydrophobic cleft between actin subdomains 1 and 3, and also contacts the DNase-binding loop of the adjacent actin subunit (A_{-2}) (Figure 1a, and Extended Data Figure 1). In parallel to us, the Belyy group resolved the cryo-EM structure of F-actin bound by F-tractin, which represents the core actin binding domain of ITPKA and is used as F-actin marker for live cell imaging[11,38]. This structure matches our findings of the F-actin binding mode[11].

Collectively, we found that the core actin binding domain of ITPKA only comprises 21 amino acids folded into an alpha helix. Interestingly, this peptide exhibits the characteristics of a short linear motif (SLiM), and its structure resembles that of the F-actin probe Lifeact[11,39,40]. Based on this consideration, we performed a structural alignment of ITPKA$^{Arg28\text{-}Ala49}$ with Lifeact (PDB: 7BTE and 7AD9) and found highly similar helix positioning, including the aligned positions Val (P1), Leu (P4), Phe (P8), and Glu (P9) (Figure 1b,c, d).

This discovery let us to propose the F-actin binding SLiM VxxLxxxFE, and to determine whether other human proteins include this motif, we developed a computational pipeline, called SLiMFold (Figure 1e). First, we derived a Position Specific Scoring Matrix (PSSM) from the aligned sequences of ITPKA$^{Arg28\text{-}Ala49}$–and Lifeact (see Figure 1e and Extended Data Figure 2), and used it to screen against the NCBI human protein database. Next to this PSSM cutoff, we applied PSIPRED[41], ANCHOR[42], and IUPred[43] cutoffs to highlight candidates containing this SLiM (Figure 1e upper panel). For each of these 539 hits a multiple sequence alignment (MSA) was performed via jackhmmer[44] against UniRef90 database, with optimized parameters to improve both speed and coverage.

Once the MSAs were compiled, we ran a custom ColabFold notebook to predict the structure of various peptide–actin complexes, leveraging the user-provided MSAs instead of recalculating new ones. Each peptide–actin model was scored by pLDDT, pTM, and ipTM, and

any candidate with mean ipTM below 0.6 (averaged over the top three ranked models) was excluded (Figure 1e lower panel, and Figure 1f).

To assess structural convergence with the reference ITPKA$^{Arg28\text{-}Ala49}$ motif, each candidate underwent a detailed comparison against the ITPKA$^{Arg28\text{-}Ala49}$–actin complex. We measured RMSD for α-carbon atoms (positions P1–P9), helix orientation (Δθ, Δφ) using a polar coordinate system, and helix polarity (Figure 1f, g, h). The resulting models were then clustered using HDBSCAN[45], which effectively segregates core clusters from outliers (Figure 1i). Cluster 3—enriched in 59 gene sequences—displayed minimal RMSD and orientation shifts relative to ITPKA, suggesting a highly similar helix arrangement for actin binding (Figure 1i). A sequence logo of these 59 clustered hits confirmed strong positional conservation at P1, P4, P8, and P9, with P1 and P8 showing the highest degree of conservation (Figure 1j). Since this actin binding motif belongs to the SLiMs, it was designated as Short linear F-actin binding Motif (SFM).

Validation of SFM-peptides *in cellulo* and *in vitro*

To validate binding of the predicted SFM-peptides to F-actin, we randomly selected candidates, and split them into two groups. Group 1 included the RhoGTPase-guanine exchange factors (GEFs) FGD1, FGD4, DENNDC1, and ARHGEF11, as well as DIXDC1, a protein that is involved in regulation of wnt-signalling (Figure 2a). Group 2 comprised the actin bundling proteins SHROOM 3, ESPNL, and PPP1R9A, the ubiquitin specific protease USP54, the z-disc protein CEFIP and the tight junction protein CGNL1 (Figure 3a). Group 1 candidates were validated by F-actin-colocalization with EGFP-peptides or full-length proteins in primary human macrophages as well as in H1299 lung cancer cells, and Group 2 candidates were validated by F-actin co-sedimentation assays with recombinant peptides.

The Group 1 SFM peptides, including the positive controls ITPKA$^{Arg28\text{-}Ala49}$ and Lifeact, were transiently expressed as EGFP-fusion peptides in primary human macrophages, allowing a

precise quantification of co-localization with F-actin subsets[46]. We additionally tested site-specific mutants (V→A and F→A) at positions 1 and 8. As shown in Figure 2 (and Extended Data Figure 4), Lifeact, ITPKA$^{Arg28\text{-}Ala49}$, FGD4 and ARHGEF11 showed prominent co-localization with podosome cores. However, DENND1C mainly localized to the podosome cap, and also FGD1 showed a broader distribution, encompassing parts of the podosome core. On the other hand, all P1/P8 mutants were diffusely distributed, confirming that positions P1 and P8 inside the SFM are essential for F-actin binding. Similar results were obtained also in H1299 lung cancer cells (Extended Data Figure 5). Together, these data indicate that the SFM binds to F-actin with different efficiency, and preference to F-actin subpopulations.

To further confirm these findings, we tested full-length versions of selected proteins in H1299 cells and introduced the mutations at P1/P8 position. Colocalization with F-actin was observed for the wild-type full-length proteins but completely disappeared upon mutation of the residues at P1/P8 position (Extended Data Figure 6). This result highlights the critical role of the SFM in mediating F-actin binding, not only in isolated peptides but also in full-length proteins.

For Group 2 candidates affinities of peptides to F-actin were tested. For this purpose, the peptides were recombinantly expressed as EGFP-fusion peptides in bacteria, and after purification its binding to F-actin was assessed by a co-sedimentation assay. After employing different F-actin concentration, dissociation constants (K_D) were calculated, showing that the SFMs of CEFIP and USP54, of ITPKA and PPP1R9A, and of CGNL1 and ESPNL bound with very similar affinities to F-actin (13 µM, 30 µM, and 50 µM) while the affinity of SHROOM 3 was lowest (90 µM) (Figure 3). These data reveal a relative low F-actin binding affinity of the SFM-peptide with high variations among the different peptides.

Amino acids in between the conserved Valine (P1) and Phenylalanine (P8) determine affinity of the SFMs

Our data depicted in Figure 3b reveal that although all SFM-peptides tested included the conserved amino acids at P1/P8 position, their binding affinities to F-actin were different. This raises the question as to whether amino acids (aa) flanking the motif modulate the affinity.

To address this, we aimed to identify the aa inside the SFM mediating the highest affinity to F-actin. For this purpose, the jackhammer generated MSAs of the peptide sequences from ITPKA, CEFIP, PPP1R9A, USP54, FGD4, ARHGEF11, DIXDC1, and Lifeact were used to generate a position-specific frequency matrix (PSFM) (Figure 3c).

Aligning homologous sequences from these motifs served as a “contrast enhancer”, revealing positions that appear intolerant to certain physiochemically similar amino acid groups (white squares in Figure 4c) or potentially essential for F-actin binding (red squares in Figure 4c). On the basis of this matrix, three hotspots were defined, an N-terminal, an internal, and a C-terminal motif. To test whether substitution of low frequency amino acid with high frequency amino acid would enhance the binding affinity to F-actin, targeted mutations in SHROOM3 and USP54 were introduced according to these hotspots. In SHROOM3, point substitutions Y1I/C6S (M2), K11L (M3), and Q17A/E19C/Q23A (M4) were produced, along with a triple-hotspot variant (M1). Similarly, USP54 mutants F1I/L7S (M2), A12I/Q16E (M3), M18I/M23A

Indeed, after assessing binding of the wt and mutant peptides to F-actin we found that mutation of the amino acids between the conserved V(P1)/F(P8) positions (internal motif) increased binding to F-actin four-fold in SHROOM 3 (mutant S2) and three-fold in USP54 (Mutant U2) compared to wt protein. However, mutations N- or C-terminal to the V(P1)/F(P8) positions (Mutants S1, S3, U1, U3) had no significant effects (Figure 3e). Moreover, determination of USP54 M1 (triple-hotspot variant) affinity to F-actin revealed a K_D of 3.7 µM (Figure 3f), confirming increased binding of the mutant to F-actin.

In summary, our data show that the conserved amino acids at P1 and P8 position are required for binding of the SFM-motif to F-actin, and reveal that those residues in between modulate affinity.

The amino acids inside the SFM are conserved over evolution

To examine the evolutionary conservation of the SFM motif, we generated sequence logos for 12 validated SFM-containing proteins across *Crocodylia*, *Aves*, *Testudines*, *Lepidosauria*, *Mammalia*, *Amphibia*, *Coelacanthiformes*, and *Actinopteri*. These seven classes were chosen because, with the exception of Lifeact (which is found only in *Saccharomyces* and thus not represented in these classes), they cover all known occurrences of the SFM in our studied dataset. The resulting logos consistently highlighted Val (P1), Leu (P4), Phe (P8), and Glu (P9) as predominantly invariant across the examined taxa. Although a small number of substitutions were observed at these positions in certain species, these core residues remained largely unchanged, indicating strong evolutionary constraints (Figure 4a). This observation aligns with our initial comparison of ITPKA and Lifeact (Figure 1), where a similar set of core residues were conserved. To accurately quantify the selective forces acting on these conserved positions, we calculated the non-synonymous to synonymous substitution ratio (dN/dS) within each SFM-containing gene across the taxa in which the motif was confidently detected. We then averaged these values around the SFM's core (Val (P1) to Glu (P9)) for all proteins. The resulting low dN/dS ratio (relative to nearby flanking regions) strongly suggests that the SFM is under purifying (negative) selection, consistent with its implied functional importance (Figure 4b). Consistent with AlphaMissense pathogenicity scores[47] (Extended Data Figure 8), this highlights the critical necessity of these amino acids for cell survival.

To get further insight into evolutionary development of the motif, phylogenetic trees were constructed and traced each validated SFM-containing protein back to its most recent common ancestor (MRCA). This analysis revealed a relative early appearance of SHROOM3 in *Bilateria*, followed by PPP1R9A, DIXDC1 and CGNL1 in *Vertebrata*. CEFIP and ITPKA emerged in *Euteleostomi*, and by the time of *Gnathostomata* USP54, FGD1, FGD4, ESPN, DENND1C were present (Figure 4c). Despite this broad taxonomic distribution, we found no

evidence of exon duplication or shuffling that might have repurposed an older motif. These observations support a convergent evolution and a spontaneous origin (*ex nihilo*) of the SFM.

To conclude, phylogenetic analysis confirms the conserved nature of the SFM motif, and indicates an *ex nihilo* development.

Final iteration of the SLiMFold pipeline delivers 103 new SFM-candidates with different biological roles

Finally, we refined the PSSM of the SLiMFold pipeline by incorporating our newly identified SFM peptides, then performed two additional iterations to compile a definitive list of putative human SFM proteins (Figure 5a and Extended Data Figure 9). In total, 124 final SFM sequences emerged in 103 genes (summarized in Table 1). Of these, 26 proteins—including both our experimental validations and those reported in the literature[48–68]—have been tested using truncation-based approaches for the SFM. Additionally, 16 other proteins are known to bind F-actin through more indirect methods (e.g., colocalization or co-sedimentation of the full-length protein)[69–85]. Further testing and more detailed truncation studies are still needed to validate the remaining predicted proteins.

In order to get information about the biological role of the new SFM-proteins, Gene Ontology analysis and manual database research, including Uniprot, NCBI, and PubMed, revealed that they are involved in the regulation of diverse cellular functions (Figure 5b, Extended Data Figure 10 and Table 1). The largest group encompassed the proteins that directly bind to and modulate actin. These include well-described actin modulators, such as Cortactin (CTTN), Caldesmon (CALD1) and Supervillin (SVIL), and many other less investigated proteins. This group is followed by actin regulators, mostly belonging to the RhoGTPase activators (GEFs) or silencers (GAPs). Also, we found microtubule associated proteins, signaling proteins, and proteins involved in the regulation of chromatin, DNA and RNA. The SFM was also identified in proteins, involved in adhesion, junction, focal adhesion, and cilia-proteins.

In conclusion, the SFM is included in highly diverse proteins which is in accordance with our finding that the peptides have arisen spontaneously in evolution.

Discussion

Actin binding proteins (ABPs) are essential for the control of cellular actin dynamics, and mutation, mis-expression or deletion of ABPs can induce severe diseases[4–8]. The biological role of ABPs is defined by their actin binding mode, and its targeting to actin can alter cytoskeletal dynamics, enzyme activity, or cellular localization (reviewed in [86,87]). In this study, we present the discovery of a new short linear F-actin binding motif (SFM), and provide first evidence for its biological role.

Detailed mutation analysis revealed that inside the SFM VxxLxxxFE, the valin (P1) and phenylalanine (P8) are essential to bind to F-actin while the amino acids in between modulate affinity, thus are important. This finding was confirmed by evolution analysis. Also, a low dN/dS ratio at these positions evidenced purifying selection and a strong evolutionary drive to retain functional F-actin-binding capacity.

Moreover, phylogenetic analysis indicates that the SFM emerged spontaneously (*ex nihilo*) rather than via exon shuffling or domain duplication. Most likely the motif randomly developed from point mutations, and since it provided a new functionality, it was negatively selected. This *ex nihilo* development seems to be typical for SLiM evolution, among which classical examples are nuclear import and export motifs as well as the calmodulin binding site[88]. Similar to the SFM, they target highly diverse proteins to cellular destinations or mediate binding of protein interaction partners (reviewed in [89]).

Accordingly, the SFM-containing proteins can be categorized into eight main groups: (1) regulation of microtubule, (2) cilia, (3) cell junctions, (4,5) actin, (6) signal transduction (7) nuclear structure and RNA processing, (8) chromatin regulation and DNA repair. Among these, the SFM-containing proteins involved in actin regulation encompass the highest number of

proteins. These include actin binding proteins known to bind to actin via canonical actin binding motifs such as CTTN[52,90], FHOD3, SVIL, MYO18A, MYO15B and CALD1 as well as RhoGEFs and RhoGAPs whose actin binding activity is partly described.

Since the SFM only mediates weak F-actin binding, we assume that it has a synergistic or additive effect on proteins targeted to actin by canonical actin binding domains. Indeed, for CTTN (cortactin) it has been shown that in addition to the cortactin repeats an adjacent linker, which includes the SFM, is required to mediate high affinity binding to F-actin[52]. Since cortactin stabilizes Arp2/3 F-actin branches, this activity may be important for the maintenance of cellular protrusions. Also, for CALD1 (Caldesmon) the SFM has been identified, and it has been shown that its binding to actin could be displaced by calmodulin, indicating that SFM-mediated actin binding is weakened after cellular stimulation[50]. A synergistic or additive effect of the SFM on actin binding may also work in actin binding proteins containing SFM repeats, including XIRP1, XIRP2, PPP1R9B, and SHROOM3. Also, the actin-regulating Rho-GTPase PLEKHG3 include SFM repeats, and its deletion abrogates cell polarity of 3T3 fibroblasts[61]. Moreover, depletion of the SFM-based actin-binding site in the RhoGEF ARHGEF11 disrupts the balance between Rac and Rho signaling, impairing the formation of leading-edge protrusions and trailing-edge retractions in cancer A431 cells[91]. Likewise, SFM repeats are present in the signaling proteins, SORBS1, SORBS2, and SSH2.

However, ITPKA only includes one SFM, but its deletion had strong effects on the actin architecture as well as on invasion of lung cancer cells[37], showing that SFM repeats are not essentially necessary for actin targeting. Also, ITPKA is a good example for actin binding proteins involved in the regulation of signal transduction. Deletion of the ITPKA-SFM significantly reduced the $InsP_3$Kinase activity of ITPKA[36], and thereby alters calcium signaling in hippocampal neurons and cancer cells[10,34]. Likewise, the cancer-related protein kinase PEAK is targeted to actin, and phosphorylates the focal adhesion protein paxillin, indicating that its actin targeting coordinates the architecture of focal adhesions after integrin activation[60]. In lymphocytes, HCLS1 is a substrate of tyrosine kinases, and activates the Arp2/3 complex

after SFM-controlled actin binding[54]. Thus, it seems that SFM-mediated actin targeting represents a signaling hub integrating actin dynamics and cellular signal transduction.

Moreover, for ZO-1 (TJP1) it could be shown that deletion of its SFM alters the barrier function in kidney MDCK cells, indicating that the SFM exhibits an important function in connecting the actin cytoskeleton with cellular junctions[67]. However, currently there are no studies analyzing actin targeting of SFM-proteins involved in the regulation of cilia, microtubule, chromatin modification, DNA repair, nuclear structure and RNA processing. These groups encompass proteins with broad cellular function such as the kinesins KIF21B and KIF23, as well as the DNA demethylating TET3, and it will be very exciting to elucidate their role in actin targeting in future studies. Finally, two main questions should be addressed in future studies: (1) Does the SFM provide a broad cellular network, or (2) does it define local molecular circuits refined during evolution to adapt to environmental changes? Since the SFM may have developed *ex nihilo,* the second hypothesis is more likely. However, future studies are necessary to prove this assumption.

In conclusion, SFM-mediated F-actin targeting regulates the properties of diverse cellular proteins, and we strongly assume that future studies will uncover yet not identified biological roles. Moreover, expanding our motif search to include non-human databases could reveal the conservation and functional diversity of SFMs across different species, providing deeper evolutionary insights and potentially identifying novel SFM-containing proteins. Looking forward, we envision that the elucidation of the biological role of SFM-proteins will uncover yet unknown cellular regulation mechanisms involved in physiological and pathological settings.

Methods

Pipeline Overview

The identification of F-actin-binding SLiMs was performed using the SLiMFold pipeline, which integrates multiple bioinformatic tools to systematically identify, filter, and validate candidate motifs. The pipeline was executed in three sequential iterations, refining results at each step to improve specificity and accuracy.

1. **Prerun**

The process begins with **Prerun.ipynb**, where hypothesized SLiMs (in this work, the aligned sequences of ITPKA and Lifeact ABDs from P1 to P9) were used along with the BLOSUM62 substitution matrix[92] to create a position-specific scoring matrix (PSSM). This PSSM was used to scan the human proteome (NCBI Taxonomy ID: 9606) for putative motif hits. To further filter out false positives, we incorporated the following criteria:

- **PSSM Score**: The matrix was used to score sequences in the database, identifying motifs similar to the hypothesized SLiMs[93].
- **IUPRED**: Predicts intrinsically disordered regions, highlighting regions suitable for motif embedding. The mean IUPRED score was calculated for the motif and its 60-residue flanking regions (N- and C-terminal)[43].
- **ANCHOR**: Identifies regions in intrinsically disordered regions likely to bind structured partners. The mean ANCHOR value was calculated within the motif[42].
- **PSIPRED**: Predicts the probability of secondary structures (helix, beta strand, and random coil)[41].

All cut-offs for these criteria were progressively relaxed across iterations (see below). This approach allowed the identification of additional hits while iteratively strengthening the PSSM, reducing false positives. Each hit was extended by ±20 flanking residues for subsequent homology searches, and redundant sequences were removed. Each hit was then paired with the bait sequence (in this study, the human actin sequence) and saved as a separate FASTA file, formatted for direct compatibility with ColabFold.

For each hit, multiple sequence alignments (MSAs) were generated using **jackhmmer**[44] with the UniRef90 database. The following jackhmmer parameters were applied: number of iterations: 5, e-value: 1e-5, no f1 or f2 filter (Original AF2 code: number of iterations: 1, e-value: 1e-5, f1: 0.0005. f2: 0.00005). The modified jackhmmer parameters improved the number of sequences retrieved, likely enhancing detection of more distant homologs. Additionally, parallelization was implemented to process multiple hits simultaneously,

significantly speeding up the alignment step. The sto-file for the bait sequence was constructed in the same manner. Output alignments were reformatted to A3M format using the **HH-suite's reformat.pl**[94] script to ensure compatibility with downstream structure prediction tools. Alignments were sorted and combined to meet AlphaFold2 Multimer input criteria.

2. ColabFold looped

The **ColabFold_looped.ipynb** notebook, originally developed by the Steinegger lab[95], was modified to enable batch processing and the integration of custom MSAs, making it more suitable for the SLiMFold workflow. These modifications were designed to streamline the process and enhance flexibility when handling multiple input sequences. In the modified version, the notebook was adapted to loop through all FASTA-files in a specified folder, such as one located in Google Drive. This automation allowed for seamless processing of multiple sequences without the need for manual input, significantly improving the workflow's efficiency. Additionally, the modified notebook incorporated the ability to retrieve precomputed A3M files from another designated folder. These alignments, created during the Prerun.ipynb step, were matched to their corresponding FASTA-files based on their names. This ensured accurate integration of the custom MSAs into the ColabFold predictions. To manage outputs effectively, the notebook allowed users to specify a folder for storing results. This organizational setup made it easier to handle and analyze the results of large-scale computations.

Compared to the original ColabFold batch notebook, these modifications introduced enhanced flexibility by supporting custom MSAs and allowing adjustments to the number of seeds used in the AlphaFold2 predictions.

3. Postanalysis

The post-analysis step was performed using the **Postanalysis.ipynb** script, which systematically processed and analyzed the outputs from the SLiMFold pipeline. This step integrated sequence and structural data to evaluate the conformational and functional relevance of candidate F-actin-binding SLiMs. The pLDDT, pTM, and ipTM scores were extracted to assess prediction confidence and interaction reliability of each predicted structure. Structures with ipTM scores below 0.6 were excluded from further analysis, ensuring a focus on high-confidence predictions. Each remaining structure was aligned to a reference PDB file (in this study, the ITPKA-actin complex) to evaluate conformational similarity.

The **root-mean-square deviation (RMSD)** between the alpha carbon (Cα) atoms of the predicted structure and the reference structure at the motif positions (P1 to P9) was calculated to quantify structural similarity. RMSD was computed as:

$$\mathrm{RMSD} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - x_{ref}\right)^2$$

where N is the number of aligned residues, x_i represents the atomic coordinates of a residue in the predicted structure, and x_{ref} represents the corresponding coordinates in the reference structure.

Angular analysis was performed using vector-based calculations, where the spatial coordinates of the alpha carbon (Cα) atoms of key residues in the SLiM (P1 to P9) were used to define vectors representing the orientation of the motif. Two angles, **phi** (***φ***, azimuthal angle in the x-y plane) and **theta** (***θ***, polar angle relative to the z-axis), were calculated to quantify the motif's orientation relative to the reference vector:

$$\varphi = arctan\left(\frac{y}{x}\right) \qquad \theta = arccos\left(\frac{z}{|v|}\right)$$

where x and y are the coordinates of the vector projection in the x-y plane and where z is the z-coordinate of the vector and $|v|$ is the magnitude of the vector. Helix polarity was also calculated to capture the directionality of the predicted motif's structural alignment. The **delta phi (Δ_φ)** and **delta theta (Δ_θ)** values, representing the differences between the angles of the predicted and reference structures, were calculated for each candidate SLiM. These metrics, along with RMSD and polarity, provided a comprehensive structural profile for clustering.

Clustering was performed using the HDBSCAN[45] algorithm, leveraging RMSD, Δ_φ, Δ_θ, and polarity as input features. The clustering parameters, including minimum cluster size and minimum samples, were optimized using the following metrics:

- **Silhouette Score**: Evaluates cluster separation and cohesion.
- **Davies-Bouldin Index**: Measures intra-cluster similarity relative to inter-cluster separation.
- **Calinski-Harabasz Index**: Assesses the ratio of between-cluster dispersion to within-cluster dispersion.

Optimal clustering parameters were selected based on these indices, ensuring robust and meaningful grouping of structurally similar SLiMs. For further analysis, all structures belonging to a single cluster were exported as a PyMOL session file (.pml), enabling manual inspection and visualization. For each cluster, the corresponding sequences were compiled into a FASTA file. These sequences were used to generate sequence logos[96], highlighting conserved

residues and identifying potential functional motifs. Additionally, each sequence was mapped to its corresponding gene using the NCBI database. This mapping facilitated gene ontology (GO) enrichment analysis, allowing functional insights into the biological roles of the identified SLiMs.

Cloning strategies

All constructs were generated via PCR-based cloning using high-fidelity polymerases (Q5 or Phusion; M0491S or M0530S, New England Biolabs Inc., Massachusetts, USA) following the manufacturer's instructions. For inserting longer DNA fragments such as full-length protein sequences, we employed T4 SLIC[97] (Sequence and Ligation Independent Cloning) (T4 DNA Polymerase; M0203S, New England Biolabs Inc., Massachusetts, USA), which relies on ~30 nucleotide overlaps between vector and insert for seamless assembly. Short peptides or specific mutations were introduced using a "QuickChange-like" site-directed mutagenesis protocol, wherein primer pairs incorporated short (8–12 nt) overlaps at their 5′ ends; the PCR products were subsequently treated with DpnI (R0176S; New England Biolabs Inc., Massachusetts, USA) to remove the template and then transformed into *E. coli* XL1-Blue (200249, Agilent Technologies, Inc., California, USA). Alternatively, the PCR products were processed using the KLD (Kinase–Ligase–DpnI) enzyme mix (M0554S, New England Biolabs Inc., Massachusetts, USA) to remove the template and ligate the new amplicon. For bacterial expression, PCR products were cloned into pSF421-based expression vectors (e.g., pSF421_10xHis_GFP_TEV), while for eukaryotic expression in mammalian cells (e.g., H1299), target genes were inserted into an mEGFP-N1 backbone. All final plasmids were verified by Sanger sequencing (Microsynth Seqlab GmbH, Göttingen, Germany). Verified constructs were subsequently transformed into *E. coli* Rosetta™ (DE3) (Novagen, Merck KGaA, Darmstadt, Germany) for bacterial protein expression or transfected into mammalian cells. Complete plasmid lists, vector maps, sequencing results, and primer sequences are provided in Supplementary Tables 2 and 3 and Source Data.

Purification of Actin

Actin was prepared from Gallus gallus (chicken) skeletal muscle as described in reference[98]. The final purification step was performed using a HiLoad 16/600 Superdex 200 column equilibrated with G-actin buffer (5 mM Tris-HCl, pH 7.5, 0.2 mM $CaCl_2$, 0.5 mM DTT and 0.2 mM ATP).

Expression and Purification of ITPKA and GFP-Peptides

All constructs were expressed in *E. coli* Rosetta™ (DE3) competent cells (Novagen, Merck KGaA, Darmstadt, Germany). Cells were grown in Terrific Broth (TB) medium at 37°C to an optical density (OD_{600}) between 0.6 and 1.0. Protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18°C for 16 hours for ITPKA and with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37°C for 4 hours for the GFP-Peptides. After induction, cells were harvested by centrifugation at 3,000 × g for 20 minutes at 4°C. The cell pellets were resuspended in cold phosphate-buffered saline (PBS) buffer, centrifuged at 2,000 × g for 30 minutes at 4°C, flash-frozen in liquid nitrogen, and stored at −80°C.

For protein purification, the cell pellets were resuspended in ice-cold Buffer A (50 mM Tris-HCl, pH 7.4, 400 mM NaCl, 3 mM $MgCl_2$ 1 mM β-mercaptoethanol for ITPKA and 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 1 mM β-mercaptoethanol for GFP-Peptides) and homogenized using an IKA® ULTRA-TURRAX® disperser (IKA-Werke GmbH & Co. KG, Staufen, Germany). DNase I was added, and the cells were lysed with a Constant Cell Disruption System (Constant Systems Limited, Daventry, UK) at 1.8 kbar. Following lysis, phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 1 mM and imidazole to a final concentration of 25 mM. Cell debris was removed by centrifugation at 43,000 × g for 30 minutes at 4°C, and the supernatant was incubated with Ni-NTA agarose resin (SERVA Electrophoresis GmbH, Heidelberg, Germany) for 30 minutes at 4°C. The bound proteins were eluted using a gradient of Buffer B (Buffer A with 500 mM imidazole). Fractions were analyzed by SDS-PAGE, and those with >90% purity were pooled and then dialyzed overnight at 4°C in Buffer A.

For the ITPKA protein, an additional cleavage step was performed to remove His tag. The pooled protein fractions were dialyzed overnight at 4°C in the presence of TEV protease, at a ratio of 1 mg TEV protease per 40 mg protein substrate, in Buffer A. Following dialysis, the cleaved protein was separated from the tags and TEV protease by reverse immobilized metal affinity chromatography (IMAC) using a Ni^{2+}-NTA column (Bio-Rad Laboratories, California, USA) on an NGC™ Liquid Chromatography System (Bio-Rad Laboratories, California, USA).

Subsequently, the pooled reversed IMAC fractions (for ITPKA) or the dialyzed proteins (for GFP Peptides) were concentrated to 0.5–2 mL using Amicon® Ultra Centrifugal Filters (Merck-Millipore, Massachusetts, USA) and further purified by size-exclusion chromatography using a HiLoad Superdex 16/600 200pg Gel Filtration Column (Cytiva, Massachusetts, USA) on an NGC™ Liquid Chromatography System (Bio-Rad Laboratories, California, USA), equilibrated with Buffer D (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 3 mM $MgCl_2$ 1 mM β-mercaptoethanol for ITPKA and 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM β-mercaptoethanol for GFP-Peptides) . The peak fractions were concentrated, aliquoted, then flash-frozen in liquid nitrogen, and stored at −80°C.

Cryo-EM data acquisition and processing

We initiated all sample preparations using frozen aliquots of G-actin. To convert G-actin into F-actin, we first thawed the samples and performed centrifugation at 100,000 x g for 30 minutes to eliminate potential aggregates. The polymerization process involved adding F buffer to achieve a final concentration of 20 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM ATP, 1 mM EGTA, and 1 mM DTT, followed by incubation for one hour at room temperature. Subsequently, we collected the filaments by centrifuging at 100,000 x g for one hour at 4°C and resuspended them in F buffer supplemented with Phalloidin at twice the molar concentration of F-actin.

For cryo-EM grid preparation, we used Quantifoil 200-mesh 2.0/1.0 holey carbon grids. Specifically, for Phalloidin-stabilized F-actin, we applied 3.5 µl of 1 µM solution onto glow-discharged grids. For the Phalloidin-stabilized F-actin-ITPKA complex, we initially applied 2 µl of 1 µM Phalloidin-stabilized F-actin to the grids. Subsequently, we added 2 µl of 10 µM ITPKA and mixed it directly on the grids immediately before plunge-freezing in a liquid ethane/propane mixture using a Vitrobot Mark IV (FEI). Datasets were collected on an FEI Titan Krios (300 KV; Gatan K3 camera; pixel size: 0.826 Å; dose per frame:1.4; defocus range: 0.8 to 3.0 µm).

Single-particle helical reconstruction was performed using Relion 4[99]. Movies were firstly motion-corrected using MotionCor2[100], and then the contrast transfer function (CTF) estimation was done using CTFFIND4[101]. Particle segments picking model was trained in cryolo 1.7[102]. The helical segments were extracted into 360 × 360 boxes and the junk segments were excluded after 2D classification. The initial helical parameters of a helical rise of 27.3 Å and a helical twist of −166.5° were applied for 3D classification and autorefinement in Relion 4.0. Overall, gold-standard resolution (Fourier shell correlation = 0.143) was calculated in Relion 4.0. The statistics for data collection and processing are listed in Supplementary Table 4.

Model building and refinement

Previously published models of phalloidin bound F-actin (PDB: 6T20[103], 7BTI[40]), and the AlphaFold2 structure of ITPKA were used as initial models. Models were initially built in ChimeraX 1.7[104], and further refined against the cryo-EM maps using ISOLDE[105], and real space refinement in Phenix[106]. The detailed model information and validation statistics for final models are described in Supplementary Table 4.

Cell culture

NCI-H1299 (H1299) cells were kindly provided by Cagatay Günes (Hamburg, Germany). For detailed cellular characteristics, refer to the American Type Culture Collection (ATCC, Rockville, USA). The cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% (v/v) fetal calf serum (FCS), 4 mM L-glutamine, 100 µg/ml streptomycin, and 100 U/ml penicillin. The culturing conditions were rigorously maintained to ensure optimal growth and viability. Additionally, the cells were tested regularly for mycoplasma contamination to ensure experiments were conducted with mycoplasma-free cells.

Primary human monocytes were isolated from buffy coats (kindly provided by Frank Bentzien, Transfusion Medicine, UKE, Hamburg, Germany). 20 ml blood was coated on 15 ml Lymphocyte Separation Medium 1077 (PromoCell, Heidelberg, Germany) and centrifuged for 30 min at 4°C and 460×g. Buffy coats were transferred to a new 50 ml Falcon tube and filled up to 50 ml with cold RPMI (Gibco, Paisley, UK). Leukocyte fractions were washed twice in RPMI and centrifuged for 10 min, as described above. Enriched leukocytes were resuspended in 400 µl monocyte buffer (5 mM EDTA and 0.5% human serum albumin in Dulbecco's PBS [DPBS], pH 7.4), mixed with 100 µl of magnetic beads suspension coupled to anti-CD14 antibodies (Miltenyi Biotec, Bergisch Gladbach, Germany) and incubated for 15 min on ice. The mixture was subsequently loaded onto Separation columns LS (Miltenyi Biotec, Bergisch Gladbach, Germany) that were previously placed in a magnetic holder and equilibrated with 500 µl cold monocyte buffer. Trapped CD14+ monocytes were washed on column with 500 µl monocyte buffer and then eluted with 1ml monocyte buffer into 15 ml cold RPMI after removal from the magnets. After centrifugation for 10 min at 4°C and 460×g, the supernatant was removed and cells were resuspended in 40 ml RPMI and seeded on a 6-well plate (Sarstedt, Nuembrech, Germany) at a density of 2*106 cells per well. After adhesion of monocytes for 1h, RPMI medium was replaced by 2 ml monocyte culture medium (RPMI substituted with 20% human serum and 1% penicillin/streptomycin (Sigma-Aldrich, Missouri, USA). Monocytes were cultivated in an incubator at 37°C, 5% CO2, and 90% humidity. Isolated monocytes were differentiated for at least 6 days.

Transfection of Cells

For H1299 cells, a total of 2.5 × 10^4 cells were seeded into 8 well chamber slides (Ibidi, Gräfelfing, Germany). After 16 h, the cells were transfected with 0.5 µg with any mEGFP-N1-Peptide or mEGFP-N1-FL-Proteins using the K2® Transfection Reagent (T060-0.75, Biontex Laboratories GmbH, München, Germany) according to the manufacturer's instructions. After

24 h, the cells were fixed with 4% paraformaldehyde/4% sucrose, stained with rhodamine-conjugated phalloidin (ab235138, Abcam Limited., Cambridge, UK), and analyzed by fluorescence microscopy using the Olympus IXplore Live microscope imaging system and the FV3000 confocal laser scanning microscope (Olympus Corporation, Tokyo, Japan).

Macrophages were detached by incubation of Accutase (Invitrogen, Massachusetts, USA) for at least 30 min in culturing conditions. Cells were collected with monocyte culture medium, washed in PBS pH 7.3 and resuspended in R-Buffer (at a concentration of 106 cells per 100 µl buffer, 10 µg DNA), provided by the Neon Transfection System (Invitrogen, Massachusetts, USA). Macrophages were transiently transfected with plasmid DNA with the following settings: voltage 1,000 V; width 40 ms, and 2 pulses. Transfected cells were resuspended in RPMI and seeded on 12-mm glass coverslips (105 cells per coverslip). The cells were left to adhere for 1 h at culturing conditions. Thereafter, 1 ml of monocyte culture medium was added, and cells were incubated overnight.

Immunofluorescence and microscopy

Cells were fixed for 10 min in 3.7% formaldehyde in PBS, and permeabilized for 10 min in PBS containing 0.5% TritonX-100. Thereafter cells were incubated for 60 min in blocking solution (2% BSA in PBS) with 1:400 phalloidin-568. Cells were washed three times in PBS and mounted on glass slides with FluoromountG (Invitrogen Massachusetts, USA) containing DAPI (Sigma-Aldrich, Missouri, USA). Images of fixed samples were acquired with a confocal laser-scanning microscope Olympus FV3000 equipped with an 60x UPlanApo HR Oil objective and Olympus FV-3000 software.

Poji Macro Analysis

Semi-automated Poji analysis was performed as previously described in reference[107]. Profile analysis was performed on ROI of transfected cells with podosomes defined by phalloidin-568 staining with a circle size of single isolated podosomes of 25 pixels. For more detailed information, see https://github.com/roherzog/Poji.

Cosedimentation assay and K_D Determination

The dissociation constant (K_D) and binding strength of the interaction between the actin-binding protein (ABP) and filamentous actin (F-actin) were determined using an F-actin cosedimentation assay.

For this, preparations were made using frozen aliquots of G-actin. To convert G-actin into F-actin, the samples were thawed and centrifuged at 100,000 × g for 30 minutes to eliminate potential aggregates. The polymerization process involved adding F-buffer to achieve a final concentration of 20 mM Tris (pH 7.5), 100 mM KCl, 2 mM $MgCl_2$, 0.5 mM ATP, 1 mM EGTA, and 1 mM DTT. The samples were then incubated for one hour at room temperature. The resulting F-actin was mixed with an equivalent amount of ABP and incubated for one hour at room temperature before cosedimentation at 100,000 × g for 30 minutes. Pellet and supernatant fractions were resuspended in sample buffer, heated at 95 °C for 15 min, and analyzed by SDS-PAGE and Western blot or Coomassie staining as detailed below.

For ABPs exhibiting relatively **low affinity** ($K_D > [ABP]$, such that free actin remains largely undepleted), we measured only the pellet fraction by Western blot and plotted densitometry (D) as a function of the total actin concentration $[Actin]_{total}$. These data were fit in GraphPad Prism using the "One site – Specific binding" hyperbolic model:

$$D = \frac{B_{max} \times [Actin]_{total}}{K_D + ([Actin]_{total})}$$

Here, B_{max} corresponds to the maximal densitometry signal at saturation.

For **higher-affinity** ABPs ($K_D < [ABP]$), the free actin can be substantially depleted, so a simple hyperbolic fit may be inaccurate. In these cases, we quantified both pellet and supernatant fractions by SDS-PAGE to determine the fraction of total ABP bound, Y. We then fit Y against $[Actin]_{total}$ using the „Quadratic" (or "Tight Binding") model[108]:

$$Y = \frac{[ABP\ Actin]}{[ABP]_{total}}$$

$$Y = \frac{[Actin]_{total} + [ABP] + K_D - \sqrt{([Actin]_{total} + [ABP] + K_D)^2 - 4 \cdot [ABP] \cdot [Actin]_{total}}}{2 \cdot [ABP]}$$

All experiments were performed in triplicate to ensure reproducibility. Varying concentrations of actin (0–120 µM) were incubated with a fixed concentration of ABP (4 µM). In each replicate, a negative control (0 µM actin) verified background signal. For low-affinity peptides, blot exposures were selected to remain in the linear range of detection, and densitometry signals (D) were directly used in the hyperbolic fit. For high-affinity assays, we confirmed mass balance by checking that pellet + supernatant matched the total ABP added. Nonlinear regression in GraphPad Prism was used to extract K_D and B_{max}.

For the binding strength determination of USP54 and SHROOM3 mutants (M1, M2, M3, and M4), the proteins were employed at a final concentration of 4 µM with F-actin at 12 µM. Densitometric analysis was performed on the Western blot results, quantifying the signal

intensity of the bound ABP in the pellet fraction to assess differences in binding strength among the mutants. The densitometry values were corrected by subtraction of the negative control (SFM with 0 µM Actin).

SDS-PAGE and Western Blotting

Equal volumes of samples were loaded onto 12% polyacrylamide gels for SDS-PAGE. Electrophoresis was performed at 120 V until the dye front reached the bottom of the gel. SDS-PAGE was transferred onto nitrocellulose membranes using a wet transfer system at 60 V for 90 minutes in 1× blot buffer. Membranes were blocked with 5% (w/v) non-fat dry milk or 5% BSA in Tris-buffered saline with Tween 20 (TBST) for 1 hour at room temperature to block nonspecific binding sites. The membranes were incubated overnight at 4°C with the primary antibody specific to GFP (mouse anti-GFP, 1:1,500, Article No. 11814460001, Roche Applied Science, Penzberg, Germany) or actin (rabbit anti-actin, 1:10,000, Article No. A2066, Sigma Aldrich, St. Louis, MO, USA) diluted in blocking buffer. After three washes with TBST, the membranes were incubated for 1 hour at room temperature with the appropriate HRP-conjugated secondary antibody. Goat anti-mouse secondary antibody was used for GFP detection (1:10,000 dilution in TBST), while goat anti-rabbit secondary antibody was used for actin detection (1:10,000). Protein bands were detected using a chemiluminescence reagent (Cytiva Amersham™ ECL™ Prime Western Blot Detection Reagent) and imaged using the INTAS ECL CHEMOCAM (INTAS Science Imaging Instruments GmbH, Göttingen, Germany). Band intensities were analyzed using ImageJ[109].

Position specific frequency matrix (PSFM)

We developed a pipeline for generating position-specific frequency matrices (PSFMs) that quantify amino acid conservation at each position across a set of sequences, focusing specifically on peptides with high affinity or strong colocalization. Multiple sequence alignments (MSAs) derived from SLiMFold's jackhmmer output were manually curated to remove alignments with names indicating different proteins, ensuring retention of only true homologs. These curated MSAs served as input for generating individual frequency tables for each peptide, which were then averaged to account for variations in the number of homologous sequences. A 23×N matrix (where N is the sequence length) was generated, with each element representing the proportion of a specific amino acid at a given position. To reduce skewness in the data and emphasize subtle conservation trends, a square root transformation was applied to the matrix. Heatmaps of the PSFMs were visualized using matplotlib, with conserved

positions annotated and residue indices labeled for interpretation. Detailed code is provided in the reference[110].

Phylogenetic and Motif Analysis

To investigate the evolutionary history of F-actin binding motif-containing proteins and their isoforms, we conducted a comprehensive phylogenetic analysis. Orthologous sequences were retrieved from OrthoDB[111] and aligned using DECIPHER[112] (100 iterations, 200 refinements). Maximum likelihood phylogenetic trees were constructed with IQ-TREE 2[113] (WAG+G model, 1000 ultrafast bootstraps). Isoform sequences were identified and filtered based on header annotations, and their phylogenetic clustering was visualized. Taxonomic information was incorporated to determine the MRCA of isoform clusters. F-actin binding motifs were extracted using regular expressions and filtered based on predicted disorder (IUPRED[43]), anchor regions (ANCHOR[42]), and random coil propensity (PSIPRED[41]). Position-Specific Scoring Matrices (PSSMs) were generated and iteratively used to refine motif identification. The MRCA of motif-containing sequences was determined using isoform-specific trees. Motif distribution was validated by mapping motifs to a taxon-based tree. Motif development across isoforms and taxonomic classes was analyzed using frequency matrices. Sequence logos[96] were generated from PSSMs. For nucleotide-level analyses, corresponding DNA sequences were retrieved from NCBI, filtered for ambiguous bases and stop codons, and codon-aligned using PRANK[114]. Phylogenetic trees were inferred with IQ-TREE 2[113] (GTR+G model), and dN/dS analysis was performed using HyPhy[115]. Detailed code and parameter settings are provided in the Source Data.

GO Enrichment analysis

To elucidate the biological functions and cellular components associated with Short Linear F-actin Binding Motif (SFM)-containing proteins, we performed Gene Ontology (GO) enrichment analysis using ShinyGO[116]. The analysis aimed to identify overrepresented GO terms within three primary categories: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). We compiled a comprehensive list of SFM-containing proteins identified in this study as the query set. Utilizing ShinyGO with its standard parameters, we input the query set to perform the enrichment analysis. The analysis employed the default settings for multiple testing correction, including the False Discovery Rate (FDR), ensuring that only GO terms with an adjusted p-value below 0.05 were considered significantly enriched.

AlphaMissense

For each of the 12 proteins examined in this study, the AlphaMissense pathogenicity score[47] was retrieved using the canonical UniProt ID. The relevant amino acid positions were then manually aligned based on the SFM motif coordinates. Residues flanking the motif (positions −30 to −20 and +30 to +40) served as control values for comparative analysis. After extracting the AlphaMissense scores for all 12 proteins, the mean score at each aligned position was calculated and the standard error of the mean (SEM) determined. Statistical evaluation of differences between motif residues and the flanking control positions was performed via one-way ANOVA with multiple comparisons.

Data availability

Coordinates and cryo-EM maps for the F-actin structure have been deposited in the Electron Microscopy Data Bank (EMDB) under accession code EMD-18866, with corresponding Protein Data Bank (PDB) entry 8R3H. The F-actin–ITPKA complex has been deposited in the EMDB under accession code EMD-18868 and in the PDB under accession code 8R3J. Additionally, PDB entries 6T20, 7BTI, and 7AD9 were used for structural comparisons. The SLiMFold pipeline code is available at https://github.com/thp42/SLiMFold, and the PSFM code is provided at https://github.com/thp42/PSFM. Source data underlying this study are provided with the paper.

Acknowledgements

We thank Christine Blechner and Andrea Mordhorst for excellent technical assistance. We also acknowledge Christian Dahlstroem (AG Windhorst, Universitätsklinikum Hamburg-Eppendorf) Nicholas Kley and Joel Fauser (AG Itzen, Universitätsklinikum Hamburg-Eppendorf) for continuous support. We are grateful to Frank Bentzien (Transfusion Medicine, Universitätsklinikum Hamburg-Eppendorf) for providing buffy coats, and to Robert Herzog and Pasquale Cervero for help in validating podosome detection. We thank Malik Alawi, Philipp Dirksen and Susanne Dobler for critical evaluation of our phylogenetic analysis. We further appreciate the support of the UKE Microscopy Imaging Facility (UMIF, DFG code RI_00489 and INST 152/933-1) for advanced microscopy access and image analysis, as well as Martin Aepfelbacher and Aymelt Itzen for continuous support. This project was funded by Erich und Gertrud Roggenbuck-Stiftung (to S.W. and T.P.). It is part of the doctoral thesis of K.S. and was additionally supported by the Deutsche Forschungsgemeinschaft (RTG2771/P4, to S.L.).

References

1. Pollard, T. D. & Borisy, G. G. Cellular Motility Driven by Assembly and Disassembly of Actin Filaments. *Cell* **112**, 453–465 (2003).
2. Heasman, S. J. & Ridley, A. J. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat Rev Mol Cell Biol* **9**, 690–701 (2008).
3. Hall, A. Rho GTPases and the Actin Cytoskeleton. *Science* **279**, 509–514 (1998).
4. Fox, J. W. *et al.* Mutations in filamin 1 Prevent Migration of Cerebral Cortical Neurons in Human Periventricular Heterotopia. *Neuron* **21**, 1315–1325 (1998).
5. Motley, W. W., Züchner, S. & Scherer, S. S. Isoform-specific loss of dystonin causes hereditary motor and sensory neuropathy. *Neurol Genet* **6**, e496 (2020).
6. Standing, A. S. I. *et al.* Autoinflammatory periodic fever, immunodeficiency, and thrombocytopenia (PFIT) caused by mutation in actin-regulatory gene *WDR1*. *Journal of Experimental Medicine* **214**, 59–71 (2017).
7. Kahr, W. H. A. *et al.* Loss of the Arp2/3 complex component ARPC1B causes platelet abnormalities and predisposes to inflammatory disease. *Nat Commun* **8**, 14816 (2017).
8. Haffner, M. C. *et al.* AIM1 is an actin-binding protein that suppresses cell migration and micrometastatic dissemination. *Nat Commun* **8**, 142 (2017).
9. Dominguez, R. Actin-binding proteins – a unifying hypothesis. *Trends in Biochemical Sciences* **29**, 572–578 (2004).
10. Köster, J.-D. *et al.* Inositol-1,4,5-trisphosphate-3-kinase-A controls morphology of hippocampal dendritic spines. *Cellular Signalling* **28**, 83–90 (2016).
11. Shatskiy, D., Sivan, A., Wedlich-Söldner, R. & Belyy, A. Structure of the F-tractin–F-actin complex. *Journal of Cell Biology* **224**, e202409192 (2025).
12. Dunker, A. K. *et al.* Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* **19**, 26–59 (2001).
13. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293**, 321–331 (1999).

14. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are ?natively unfolded? proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).
15. Tompa, P. Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **27**, 527–533 (2002).
16. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208 (2005).
17. Xie, H. *et al.* Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J. Proteome Res.* **6**, 1882–1898 (2007).
18. Galea, C. A., Wang, Y., Sivakolundu, S. G. & Kriwacki, R. W. Regulation of Cell Division by Intrinsically Unstructured Proteins: Intrinsic Flexibility, Modularity, and Signaling Conduits. *Biochemistry* **47**, 7598–7609 (2008).
19. Gsponer, J. & Madan Babu, M. The rules of disorder or why disorder rules. *Progress in Biophysics and Molecular Biology* **99**, 94–103 (2009).
20. Tompa, P. Unstructural biology coming of age. *Current Opinion in Structural Biology* **21**, 419–425 (2011).
21. Babu, M. M., van der Lee, R., de Groot, N. S. & Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology* **21**, 432–440 (2011).
22. Dinkel, H. *et al.* The eukaryotic linear motif resource ELM: 10 years and counting. *Nucl. Acids Res.* **42**, D259–D266 (2014).
23. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. BioSyst.* **8**, 268–281 (2012).
24. Mohan, A. *et al.* Analysis of Molecular Recognition Features (MoRFs). *Journal of Molecular Biology* **362**, 1043–1059 (2006).
25. Mi, T. *et al.* Minimotif Miner 3.0: database expansion and significantly improved reduction of false-positive predictions from consensus sequences. *Nucleic Acids Research* **40**, D252–D260 (2012).

26. Chereau, D. *et al.* Actin-bound structures of Wiskott–Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16644–16649 (2005).
27. Marguet, F. *et al.* Association of fingerprint bodies with rods in a case with mutations in the LMOD3 gene. *Neuromuscular Disorders* **30**, 207–212 (2020).
28. Schatz, U. A. *et al.* Evidence of mild founder *LMOD3* mutations causing nemaline myopathy 10 in Germany and Austria. *Neurology* **91**, e1690–e1694 (2018).
29. Liverman, A. D. B. *et al.* Arp2/3-independent assembly of actin by *Vibrio* type III effector VopL. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17117–17122 (2007).
30. Tam, V. C. *et al.* Functional Analysis of VopF Activity Required for Colonization in Vibrio cholerae. *mBio* **1**, e00289-10 (2010).
31. Wood, H. & Artsob, H. Spotted Fever Group Rickettsiae: A Brief Review and a Canadian Perspective: Spotted Fever Group Rickettsiae. *Zoonoses and Public Health* **59**, 65–79 (2012).
32. Reed, S. C. O., Lamason, R. L., Risca, V. I., Abernathy, E. & Welch, M. D. Rickettsia Actin-Based Motility Occurs in Distinct Phases Mediated by Different Actin Nucleators. *Current Biology* **24**, 98–103 (2014).
33. Windhorst, S. *et al.* Inositol-1,4,5-trisphosphate 3-kinase A regulates dendritic morphology and shapes synaptic Ca2+ transients. *Cellular Signalling* **24**, 750–757 (2012).
34. Windhorst, S. *et al.* Inositol 1,4,5-Trisphosphate 3-Kinase-A Is a New Cell Motility-promoting Protein That Increases the Metastatic Potential of Tumor Cells by Two Functional Activities. *Journal of Biological Chemistry* **285**, 5541–5554 (2010).
35. Paraschiakos, T., Flat, W., Chen, Y., Kirchmair, J. & Windhorst, S. Mechanism of BIP-4 mediated inhibition of InsP3Kinase-A. *Bioscience Reports* **41**, BSR20211259 (2021).
36. Ashour, D. J. *et al.* The catalytic domain of inositol-1,4,5-trisphosphate 3-kinase-a contributes to ITPKA-induced modulation of F-actin: Effect of Full Length ITPKA on Actin Structure. *Cytoskeleton* **72**, 93–100 (2015).

37. Küster, L. *et al.* The actin bundling activity of ITPKA mainly accounts for its migration-promoting effect in lung cancer cells. *Bioscience Reports* **43**, BSR20222150 (2023).
38. Melak, M., Plessner, M. & Grosse, R. Actin visualization at a glance. *Journal of Cell Science* jcs.189068 (2017) doi:10.1242/jcs.189068.
39. Belyy, A., Merino, F., Sitsel, O. & Raunser, S. Structure of the Lifeact-F-actin complex. *PLoS Biol* **18**, e3000925 (2020).
40. Kumari, A., Kesarwani, S., Javoor, M. G., Vinothkumar, K. R. & Sirajuddin, M. Structural insights into actin filament recognition by commonly used cellular actin markers. *EMBO J* **39**, (2020).
41. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* **47**, W402–W407 (2019).
42. Dosztányi, Z., Mészáros, B. & Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746 (2009).
43. Dosztányi, Z. Prediction of protein disorder based on IUPred. *Protein Sci* **27**, 331–340 (2018).
44. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res* **43**, W30-38 (2015).
45. McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *JOSS* **2**, 205 (2017).
46. Linder, S., Cervero, P., Eddy, R. & Condeelis, J. Mechanisms and roles of podosomes and invadopodia. *Nat Rev Mol Cell Biol* (2022) doi:10.1038/s41580-022-00530-6.
47. Cheng, J. *et al.* Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).
48. Arai, A., Spencer, J. A. & Olson, E. N. STARS, a Striated Muscle Activator of Rho Signaling and Serum Response Factor-dependent Transcription. *Journal of Biological Chemistry* **277**, 24453–24459 (2002).

49. Banerjee, J., Fischer, C. C. & Wedegaertner, P. B. The Amino Acid Motif L/IIxxFE Defines a Novel Actin-Binding Sequence in PDZ-RhoGEF. *Biochemistry* **48**, 8032–8043 (2009).
50. Zhan, Q. Q., Wong, S. S. & Wang, C. L. A calmodulin-binding peptide of caldesmon. *Journal of Biological Chemistry* **266**, 21810–21814 (1991).
51. D'Atri, F. & Citi, S. Cingulin interacts with F-actin in vitro. *FEBS Letters* **507**, 21–24 (2001).
52. Scherer, A. N., Anand, N. S. & Koleske, A. J. Cortactin stabilization of actin requires actin-binding repeats and linker, is disrupted by specific substitutions, and is independent of nucleotide state. *Journal of Biological Chemistry* **293**, 13022–13032 (2018).
53. Marat, A. L., Ioannou, M. S. & McPherson, P. S. Connecdenn 3/DENND1C binds actin linking Rab35 activation to the actin cytoskeleton. *MBoC* **23**, 163–175 (2012).
54. Wang, X. *et al.* DIXDC1 isoform, l-DIXDC1, is a novel filamentous actin-binding protein. *Biochemical and Biophysical Research Communications* **347**, 22–30 (2006).
55. Hao, J.-J., Zhu, J., Zhou, K., Smith, N. & Zhan, X. The Coiled-coil Domain Is Required for HS1 to Bind to F-actin and Activate Arp2/3 Complex. *Journal of Biological Chemistry* **280**, 37988–37994 (2005).
56. Johnson, H. W. & Schell, M. J. Neuronal IP_3 3-Kinase is an F-actin–bundling Protein: Role in Dendritic Targeting and Regulation of Spine Morphology. *MBoC* **20**, 5166–5180 (2009).
57. Brehm, M. A., Schreiber, I., Bertsch, U., Wegner, A. & Mayr, G. W. Identification of the actin-binding domain of Ins(1,4,5) *P* 3 3-kinase isoform B (IP3K-B). *Biochemical Journal* **382**, 353–362 (2004).
58. Maul, R. S. *et al.* EPLIN regulates actin dynamics by cross-linking and stabilizing filaments. *The Journal of Cell Biology* **160**, 399–407 (2003).
59. Fauchereau, F. The RhoGAP activity of OPHN1, a new F-actin-binding protein, is negatively controlled by its amino-terminal domain. *Molecular and Cellular Neuroscience* **23**, 574–586 (2003).

60. Kelber, J. A. & Klemke, R. L. PEAK1, a Novel Kinase Target in the Fight Against Cancer. *Oncotarget* **1**, 219–223 (2010).
61. Nguyen, T. T. T. *et al.* PLEKHG3 enhances polarized cell migration by activating actin filaments at the cell front. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10091–10096 (2016).
62. Oliver, C. J. *et al.* Targeting Protein Phosphatase 1 (PP1) to the Actin Cytoskeleton: the Neurabin I/PP1 Complex Regulates Cell Morphology. *Molecular and Cellular Biology* **22**, 4690–4701 (2002).
63. Barnes, A. P., Smith, F. D., VanDongen, H. M., VanDongen, A. M. J. & Milgram, S. L. The identification of a second actin-binding region in spinophilin/neurabin II. *Molecular Brain Research* **124**, 105–113 (2004).
64. Smith, C. *et al.* Septin 9 Exhibits Polymorphic Binding to F-Actin and Inhibits Myosin and Cofilin Activity. *Journal of Molecular Biology* **427**, 3273–3284 (2015).
65. Hildebrand, J. D. & Soriano, P. Shroom, a PDZ Domain–Containing Actin-Binding Protein, Is Required for Neural Tube Morphogenesis in Mice. *Cell* **99**, 485–497 (1999).
66. Yeo, N. C. *et al. Shroom3* contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res.* **25**, 57–65 (2015).
67. Belardi, B. *et al.* A Weak Link with Actin Organizes Tight Junctions to Control Epithelial Permeability. *Developmental Cell* **54**, 792-804.e7 (2020).
68. Pacholsky, D. *et al.* Xin repeats define a novel actin-binding motif. *Journal of Cell Science* **117**, 5257–5268 (2004).
69. Schulze, J. O. *et al.* Structural and functional characterization of human Iba proteins. *The FEBS Journal* **275**, 4627–4640 (2008).
70. Boeynaems, S. *et al.* Aberrant phase separation is a common killing strategy of positively charged peptides in biology and human disease. Preprint at https://doi.org/10.1101/2023.03.09.531820 (2023).
71. Chrifi, I. *et al.* Cgnl1, an endothelial junction complex protein, regulates GTPase mediated angiogenesis. *Cardiovascular Research* **113**, 1776–1788 (2017).

72. Zhao, B.-W. *et al.* FBXO34 Regulates the G2/M Transition and Anaphase Entry in Meiotic Oocytes. *Front. Cell Dev. Biol.* **9**, 647103 (2021).
73. Estrada, L. Fgd1, the Cdc42 guanine nucleotide exchange factor responsible for faciogenital dysplasia, is localized to the subcortical actin cytoskeleton and Golgi membrane. *Human Molecular Genetics* **10**, 485–495 (2001).
74. Inokuchi, J. *et al.* Deregulated expression of KRAP, a novel gene encoding actin-interacting protein, in human colon cancer cells. *J Hum Genet* **49**, 46–52 (2004).
75. Rivera Alvarez, J. *et al.* The kinesin Kif21b regulates radial migration of cortical projection neurons through a non-canonical function on actin cytoskeleton. *Cell Reports* **42**, 112744 (2023).
76. Abé, T. *et al.* Ladinin-1 in actin arcs of oral squamous cell carcinoma is involved in cell migration and epithelial phenotype. *Sci Rep* **14**, 22778 (2024).
77. Wong, M. J., Malapitan, I. A., Sikorski, B. A. & Jongstra, J. A cell-free binding assay maps the LSP1 cytoskeletal binding site to the COOH-terminal 30 amino acids. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1642**, 17–24 (2003).
78. Cervero, P., Wiesner, C., Bouissou, A., Poincloux, R. & Linder, S. Lymphocyte-specific protein 1 regulates mechanosensory oscillation of podosomes and actin isoform-based actomyosin symmetry breaking. *Nat Commun* **9**, 515 (2018).
79. Sandeep, P. *et al.* Neuron navigators: A novel frontier with physiological and pathological implications. *Molecular and Cellular Neuroscience* **127**, 103905 (2023).
80. Maas, M. N. *et al.* Examining prestructured β-actin peptides as substrates of histidine methyltransferase SETD3. *Sci Rep* **14**, 26439 (2024).
81. Fredriksson-Lidman, K., Van Itallie, C. M., Tietgens, A. J. & Anderson, J. M. Sorbin and SH3 domain-containing protein 2 (SORBS2) is a component of the acto-myosin ring at the apical junctional complex in epithelial cells. *PLoS ONE* **12**, e0185448 (2017).
82. Mehta, V. *et al.* SPECC1L binds the myosin phosphatase complex MYPT1/PP1β and can regulate its distribution between microtubules and filamentous actin. *Journal of Biological Chemistry* **299**, 102893 (2023).

83. Chen, Y. *et al.* F-actin and Myosin II Binding Domains in Supervillin. *Journal of Biological Chemistry* **278**, 46094–46106 (2003).
84. Pestonjamasp, K. N., Pope, R. K., Wulfkuhle, J. D. & Luna, E. J. Supervillin (p205): A Novel Membrane-associated, F-Actin–binding Protein in the Villin/Gelsolin Superfamily. *The Journal of Cell Biology* **139**, 1255–1269 (1997).
85. Wulfkuhle, J. D. *et al.* Domain analysis of supervillin, an F-actin bundling plasma membrane protein with functional nuclear localization signals. *Journal of Cell Science* **112**, 2125–2136 (1999).
86. Kadzik, R. S., Homa, K. E. & Kovar, D. R. F-Actin Cytoskeleton Network Self-Organization Through Competition and Cooperation. *Annu. Rev. Cell Dev. Biol.* **36**, 35–60 (2020).
87. Pollard, T. D. Actin and Actin-Binding Proteins. *Cold Spring Harb Perspect Biol* **8**, a018226 (2016).
88. Davey, N. E., Cyert, M. S. & Moses, A. M. Short linear motifs – ex nihilo evolution of protein regulation. *Cell Commun Signal* **13**, 43 (2015).
89. Van Roey, K. *et al.* Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **114**, 6733–6778 (2014).
90. Weed, S. A. *et al.* Cortactin localization to sites of actin assembly in lamellipodia requires interactions with F-actin and the Arp2/3 complex. *J Cell Biol* **151**, 29–40 (2000).
91. Nanda, S. *et al.* Rho GTPase activity crosstalk mediated by Arhgef11 and Arhgef12 coordinates cell protrusion-retraction cycles. *Nat Commun* **14**, 8356 (2023).
92. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89**, 10915–10919 (1992).
93. Altschul, S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).
94. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).

95. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

96. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res.* **14**, 1188–1190 (2004).

97. Li, M. Z. & Elledge, S. J. SLIC: A Method for Sequence- and Ligation-Independent Cloning. in *Gene Synthesis* (ed. Peccoud, J.) vol. 852 51–59 (Humana Press, Totowa, NJ, 2012).

98. Pardee, J. D. & Aspudich, J. [18] Purification of muscle actin. in *Methods in Enzymology* vol. 85 164–181 (Elsevier, 1982).

99. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal* **478**, 4169–4185 (2021).

100. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).

101. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).

102. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218 (2019).

103. Pospich, S., Merino, F. & Raunser, S. Structural Effects and Functional Implications of Phalloidin and Jasplakinolide Binding to Actin Filaments. *Structure* **28**, 437-449.e5 (2020).

104. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14–25 (2018).

105. Croll, T. I. *ISOLDE* : a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).

106. Afonine, P. V. *et al.* Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).

107. Herzog, R., Van Den Dries, K., Cervero, P. & Linder, S. Poji: a Fiji-based tool for analysis of podosomes and associated proteins. *Journal of Cell Science* **133**, jcs238964 (2020).

108. Belyy, A. *et al.* Actin activates Pseudomonas aeruginosa ExoY nucleotidyl cyclase toxin and ExoY-like effector domains from MARTX toxins. *Nat Commun* **7**, 13582 (2016).

109. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).

110. Paraschiakos, T. Position Specific Frequency Matrix v1.0. Zenodo https://doi.org/10.5281/ZENODO.10702806 (2024).

111. Zdobnov, E. M. *et al.* OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Research* **49**, D389–D393 (2021).

112. Wright, E. S. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* **16**, 322 (2015).

113. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).

114. Löytynoja, A. Phylogeny-aware alignment with PRANK. in *Multiple Sequence Alignment Methods* (ed. Russell, D. J.) vol. 1079 155–170 (Humana Press, Totowa, NJ, 2014).

115. Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).

116. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).

Figure Legends

Figure 1: Structural alignment of ITPKA$^{Arg28\text{-}Ala49}$ and Lifeact points towards a novel F-actin binding motif, prompting proteome-wide discovery using the SLiMFold pipeline. (a) Structural superimposition of the ITPKA$^{Arg28\text{-}Ala49}$–actin complex (orange and red for the two actin subunits, cyan for ITPKA) with the Lifeact–actin structure (grey; PDB: 7AD9) shows near-identical actin conformations apart from a minor shift in the D-loop of the ITPKA-bound model. **(b)** Rotating the helices by 180° highlights the residues interacting with F-actin. Conserved side chains are represented as sticks and labeled P1, P4, P8, and P9. **(c)** Domain architecture of **S. cerevisiae** ABP140 and **H. sapiens** ITPKA. ABP140 comprises an N-terminal F-actin binding domain (ABD) and a C-terminal methyltransferase (MTase) domain. ITPKA features an N-terminal ABD, a central calmodulin-binding domain (CaMBD), and a C-terminal inositol 1,4,5-trisphosphate 3-kinase (IP3K) domain. Notably, the F-actin binding motifs, derived in this study, are located within intrinsically disordered regions, which is typical for short linear motifs (SLiMs). **(d)** Sequence comparison underscores structural and sequence conservation, conserved positions P1, P4, P8, and P9 are labeled in red. **(e)** SLiMFold pipeline overview. The SLiMFold pipeline started with the hypothesized SLiMs of Lifeact and ITPKA$^{Arg28\text{-}Ala49}$ Sequence, on which basis a position-specific scoring matrix (PSSM) was calculated to identify motif hits from NCBI databases which were filtered by PSSM scores, IUPRED, ANCHOR, and PSIPRED. This yielded 539 hits in the first iteration. Thereafter, multiple sequence alignments (MSAs) were calculated locally using UniRef90 with modified jackhmmer filters, followed by multimer predictions on ColabFold. The predicted structures were analyzed by extracting AlphaFold2 scores, computing RMSD and angle metrics (θ, φ). These data were clustered to identify 59 new sequences that carry the suspected motif. **(f-i)** Detailed illustration of calculated metadata from (e). **(f)** Scatterplot of predicted peptide–actin complexes plotting mean ipTM vs. RMSD relative to ITPKA$^{Arg28\text{-}Ala49}$. A red dotted line at ipTM=0.6 marks the threshold below which interactions are unreliable. **(g)** Helix orientation was assessed by comparing $\Delta\varphi$ and $\Delta\theta$ angles in a polar coordinate system after superimposing each complex onto the ITPKA$^{Arg28\text{-}Ala49}$–actin reference. **(h)** A 3D plot of $\Delta\varphi$, $\Delta\theta$, and helix polarity (1 = same direction, −1 = opposite) maps each predicted peptide, with data points colored by log(1+RMSD). Angles near zero indicate an orientation similar to that of ITPKA$^{Arg28\text{-}Ala49}$. **(i)** HDBSCAN clustering of these data points groups peptides based on orientation and RMSD. Outliers are assigned to Cluster –1, while Cluster 3 contains helices most closely matching ITPKA$^{Arg28\text{-}Ala49}$. **(j)** Sequence logo depicting positional conservation among the 59 identified peptides confirms robust conservation at positions P1, P4, P8, and P9, strongest at P1 and P8.

Figure 2: Validation and functional analysis of the novel short linear F-actin binding motif (SFM) by cellular F-actin colocalization. (a) Five peptides identified by the SLiMFold pipeline were aligned with the sequences of ITPKA and Lifeact; the conserved positions P1, P4, P8, P9 are highlighted in red. To probe their functional significance, the peptides were cloned into an pEGFP vector, and alanine mutations were introduced at P1 and P8. **(b)** Primary human macrophages were transfected with the vectors coding for the EGFP-tagged peptides and stained with phalloidin-568 to visualize F-actin (red). Representative cell images are shown including zoomed- into one podosome substructure, scale bars: 10 and 5 µm, respectively. Colocalization is evident when EGFP-tagged peptides (green) overlap with phalloidin-stained actin (red). Right panels: Imaris 3D reconstructions and Poji radial profiles confirm varying degrees of peptide–actin colocalization. Shown are Poji radial profiles of fluorescence intensities, at a z plane of highest F-actin intensity, with the mean values of F-actin and respective peptide ± standard deviation of at least 3 cells, and 240 podosomes per transfected peptide. Poji radial profiles were normalized to set the intensity values from 0% to 100%.

Figure 3: Determination of SFM-binding affinities to F-actin, and identification of affinity-modulating amino acids. (a) Six further peptides identified by the SLiMFold pipeline were aligned with the sequences of ITPKA and Lifeact, and the conserved positions P1, P4, P8, and P9 highlighting in red. The peptides were expressed in bacteria, and after purification F-actin pull-downs were performed. Binding of the peptides were analyzed by Western blotting, and to calculate relative dissociation constants (K_D values), the F-actin concentrations were plotted against normalized band intensities. The right panel provides the K_D-values for all peptides tested. **(c)** In order to identify the amino acids modulating the affinity of the peptides to F-actin, their frequency was analyzed by generating a position specific matrix (PSFM) using MSA from CEFIP, USP54, ITPKA, PPP1R9A, Lifeact, ARHGEF11, FGD4, DIXDC1. Y-axis represents positively charged (+), negatively charged (-), polar (P), unique (U), non-polar (NP) and aromatic (A) residues, and X-axis the position inside the SFM. This matrix, color-coded by √(amino acid frequency), pinpointed less (white) and more frequent (red) residues. **(d)** SHROOM3 and USP54 wild-type (WT) sequences and corresponding mutants, including the key residues P1, P4, P8, P9. M2 features mutations N-terminal to the motif, M3 contains mutations within the motif, and M4 includes mutations C-terminal to the motif. M1 incorporates all mutations (N-terminal, within, and C-terminal). Newly mutated positions are labeled in red. **(e)** Quantified relative band intensities from SHROOM3 and USP54 pulldown assays, normalized and averaged across replicates, highlighting differences in binding affinities between WT and mutant peptides. (f) For USP54 M1 addiotinally the affinity to F-actin was determined. Statistical significance is denoted by asterisks: $*p < 0.05$; $**p < 0.01$.

Figure 4: Evolutionary analysis of validated SFM-containing proteins. (a) Sequence logos of SFMs from representative organism classes (e.g., *Actinopteri, Coelacanthiformes*, *Amphibia*, *Mammalia*, *Lepidosauria*, *Testudines*, *Aves*, and *Crocodylia*). Key motif positions P1, P4, P8, and P9 are labeled. Larger letters indicate higher residue conservation, reinforcing the importance of these positions for F-actin binding. **(b)** A dN/dS ratio analysis of all validated SFM-containing peptides is plotted ± SEM, with amino acid position on the x-axis and the dN/dS ratio on the y-axis. The conserved SFM residues (P1, P4, P8, P9) are annotated to highlight their selective pressures. Notably, these positions exhibit lower dN/dS values, suggesting strong purifying selection. **(c)** Phylogenetic tree illustrating the distribution of validated SFM-containing proteins across various species. Symbols mark the most recent common ancestor (MRCA) *Saccharomyces* (ABP140), *Bilateria* (SHROOM3), *Vertebrata* (CGNL1, DIXDC1, FGD4, PPP1R9A), *Euteleostomi* (ARHGEF11, ITPKA), and *Gnathostomata* (CEFI, DENND1C, ESPLN, FGD1, FGD4, USP54).

Figure 5: Iterative SLiMFold pipeline expansion and functional classification of newly identified SFMf-containing proteins. (a) Scheme of three iterative SLiMFold rounds (for detailed description, see Figure 3), including *in vitro* and *in cellulo* validation, yielding 124 peptide-sequences from 103 different genes. **(b)** Cartoon representation of a eukaryotic cell highlighting the functional categories of new SFM-bearing proteins. SFM candidates are grouped according to their principal roles and subcellular localizations, encompassing direct actin modulators, actin regulators, other signaling proteins and enzymes, nuclear structure and RNA processing, chromatin modification and DNA repair, microtubule-associated factors, cilia, and cell-junction-associated proteins. Proteins for which F-actin binding has been attributed to the SFM (e.g., validated by truncation studies) are marked with an asterisk (*), whereas proteins that bind F-actin but lack direct evidence linking the interaction specifically to the SFM are marked with a dagger (†). The cartoon was created with BioRender.com.

Gene	Protein name	Uniprot ID	Start	End	Sequence	ipTM	Iteration	Comment
ABRA	actin-binding Rho-activating protein	Q8N0Z2	242	250	VGNLKGRWQ	0.86	1	In line with our motif, F-actin binding shown by STARS [ABRA] by two regions AA 234–279 and AA 346–375.[1]
AIF1L	allograft inflammatory factor 1-like	Q9BQI0	119	127	VLKLVMMFE	0.78	2	No specific domain/motif analyzed, but F-Actin binding shown.[2]
ANK2	ankyrin-2	Q01484	1990	1998	VRELMKAFQ	0.90	1	
ANK3	ankyrin-3	Q12955	2715	2723	LSSIRLKFE	0.64	3	
ANKLE1	ankyrin repeat and LEM domain-containing protein 2	Q86XL3	729	737	VSDLTVEFD	0.84	3	
ANKRD55	ankyrin repeat domain-containing protein 55	Q3KP44	378	386	VNDIITTFD	0.68	3	
ARHGAP11A	rho GTPase-activating protein 11A	Q6P4F7	808	816	VSDHIQWFN	0.84	3	
ARHGAP42	rho GTPase-activating protein 42	A6NI28	799	807	VAAKAQLFE	0.84	2	
ARHGEF11	rho guanine nucleotide exchange factor 11	O15085	565	573	VRNIIQHFE	0.90	1	Validates our motif, F-actin binding shown by ARHGEF11 through AA 561-585.[3]
ARHGEF15	rho guanine nucleotide exchange factor 15	O94989	150	158	VRRLAGRFE	0.84	1	
BCR	breakpoint cluster region protein	P11274	151	159	VAALRSNFE	0.85	1	
BRD10	uncharacterized bromodomain-containing protein 10	Q5HYC2	773	781	LRELITKIE	0.72	3	
CEFIP	cardiac-enriched FHL2-interacting protein	Q711Q0	143	151	VSTLIKSFD	0.89	1	
C4orf54	uncharacterized protein C4orf54	D6RIA3	1352	1360	VSARAAAFE	0.78	2	No specific domain/motif analyzed, but F-Actin binding shown.[4]
C6orf132	uncharacterized protein C6orf132	Q5T0Z8	563	571	VRQIRNELE	0.81	3	
C9orf152	uncharacterized protein C9orf152	Q5JTZ5	39	47	IQFLRAQYE	0.63	3	
CALD1	caldesmon	Q05682	709	716	VRNIKSMWE	0.85	3	Validates our motif, F-actin binding shown by Gallus CALD1 through AA 666-683. In Gallus CALD1 our Motif corresponds to 667-675.[5]
CAMSAP2	calmodulin-regulated spectrin-associated protein 2	Q08AD1	845	853	LADIKESME	0.72	3	
CASKIN1	caskin-1	Q8WXD9	1001	1009	VKSIAAMLE	0.90	3	
CCDC14	coiled-coil domain-containing protein 14	Q49A88	453	461	IQRLITEME	0.73	3	
CCDC150	coiled-coil domain-containing protein 150	Q8NCX0	828	836	IEALRKQFQ	0.82	3	
CGN	cingulin	Q9P2M7	178	186	VDSLINKFD	0.90	1	Validates our motif, F-actin binding shown by Xenopus CGN through AA 101-294. In Xenopus CGN our Motif corresponds to 267-275.[6]
CGNL1	cingulin-like protein 1	Q0VF96	353	361	VDQLIEKFD	0.89	1	No specific domain/motif analyzed, but F-Actin binding shown.[7]

CTTN	src substrate cortactin	Q14247	347	355	TSNIRANFE	0.76	1	Validates our motif, F-actin binding shown by CTTN by "helical domain". Corresponds approx. to AA 348-401.[8]
CRYBG3	very large A-kinase anchor protein	Q68DQ2	2045	2053	RTDLVHHFE	0.74	3	
DBNDD2	dysbindin domain-containing protein 2	Q9BQY9	114	122	LRERQKFFE	0.78	3	
DENND1A	DENN domain-containing protein 1A	Q8TEH3	998	1006	VEQLRKQWE	0.76	1	
DENND1C	DENN domain-containing protein 1C	Q8IV53	792	800	VADLKKCFE	0.76	1	Validates our motif, F-actin binding shown by DENND1C through AA 786-801.[9]
DENND2C	DENN domain-containing protein 2C	Q68D51	21	29	IKQKISQWE	0.87	3	
DIXDC1	dixin	Q155Q3	198	206	VRALVQQYE	0.88	1	Validates our motif, F-actin binding shown by DIXDC1 through AA 127-300.[10]
DSP	desmoplakin	P15924	1203	1211	MSNLRNKYE	0.77	1	
DZIP1L	cilium assembly protein DZIP1L	Q8IYY4	671	679	VQSMVKNLE	0.84	2	
EML1	echinoderm microtubule-associated protein-like 1	O00423	33	41	VTDRIASLE	0.74	3	
ENKD1	enkurin domain-containing protein 1	Q9H0I2	74	82	VGDVLLQLE	0.72	1	
ESPN	espin isoform	A0A1B0GUN9	949	957	VQHLRATYE	0.88	2	
ESPNL	espin-like protein	Q6ZVH7	652	660	VRTLRGNFE	0.87	1	
FBXO34	F-box only protein 34	Q9NWN3	279	287	VLDMVAKLE	0.87	1	No specific domain/motif analyzed, but F-Actin colocalization shown.[11]
FGD1	FYVE, RhoGEF and PH domain-containing protein 1	P98174	211	219	VSSLIEKFE	0.89	1	No specific domain/motif analyzed, but F-Actin colocalization shown.[12]
FGD2	FYVE, RhoGEF and PH domain-containing protein 2	Q7Z6J4	12	20	VSNLVTVFE	0.87	1	
FGD4	FYVE, RhoGEF and PH domain-containing protein 4	Q96M96	20	28	VSDLISRFE	0.84	1	Validates our motif, F-actin binding shown by FGD4 through AA 23-28.[3]
FHOD3	FH1/FH2 domain-containing protein 3	Q2V2M9	745	753	VKAFAEKFN	0.67	3	
FYB2	FYN-binding protein 2	Q5VWT5	9	17	FKELRAKFQ	0.63	1	
HCLS1	hematopoietic lineage cell-specific protein	P14317	235	243	TRGLKAKFE	0.71	1	Validates our motif, F-actin binding shown by HCLS1 by "Coiled-coil domain". Corresponds to AA 215-280.[13]
HOOK1	protein Hook homolog 1	Q9UJC3	517	525	IRELQQQIE	0.61	3	
IER5	immediate early response gene 5 protein	Q5VY09	266	274	VANLISIFG	0.78	1	
IL16	pro-interleukin-16	Q14005	839	847	IRQRISSFE	0.76	2	
ITPKA	inositol-trisphosphate 3-kinase A	P23677	31	39	VGELRLLFE	0.87	1	Validates our motif, F-actin binding shown by ITPKA through AA 9-52.[14]
ITPKB	inositol-trisphosphate 3-kinase B	P27987	143	151	VNQKVGMFE	0.74	2	Validates our motif, F-actin binding shown by ITPKB through AA 108-170.[15]
ITPRID2	protein ITPRID2	P28290	169	177	VSELLELYE	0.74	2	No specific domain/motif analyzed, but F-Actin colocalization shown.[16]

KIF21B	kinesin-like protein KIF21B	O75037	866	874	VSSIVRQWN	0.90	1	No specific domain/motif analyzed, but F-Actin colocalization shown.[17]
KIF23	kinesin-like protein KIF23	Q02241	722	730	VASCISEWE	0.84	3	
LAD1	ladinin-1	O00515	438	446	VASKRHLFE	0.78	1	No specific domain/motif analyzed, but F-Actin colocalization shown.[18]
LCA5	lebercilin	Q86VQ0	258	266	LAERKRAYE	0.66	3	
LETM1	mitochondrial proton/calcium exchanger protein precursor	O95202	612	620	VQQMIGQID	0.73	2	
LIMA1	LIM domain and actin-binding protein 1	Q9UHB6	194	202	LNRLKMMFE	0.82	1	In line with our motif, F-actin binding shown by LIMA1 by two regions, of which one is located N-terminally (AA 1-260).[19]
LSP1	lymphocyte-specific protein 1	P33241	260	268	VASTKSRWE	0.85	3	Our motif was not analyzed, but F-Actin colocalization shown by other actin binding domains (Caldesmon like binding domain and villin headpiece like binding domain).[20,21]
METTL25	probable methyltransferase-like protein 25	Q8N6Q8	286	294	ISNIRNQME	0.65	2	
MTUS2	microtubule-associated tumor suppressor candidate 2	Q5JR59	506	514	VAENRNLLE	0.71	1	
MYO15B	myosin XVB	Q96JP2	2147	2155	IGNIIRMYQ	0.61	1	
MYO18A	unconventional myosin-XVIIIa	Q92614	1685	1693	IAQLKNQLE	0.69	1	
N4BP1	NEDD4-binding protein 1	O75113	244	252	VSELTKQMD	0.80	1	
NAV2	neuron navigator 2	Q8IVL1	753	761	IADLRQNLE	0.67	1	No specific domain/motif analyzed, but F-Actin binding shown.[22]
NAV3	neuron navigator 3	Q8IVL0	689	697	IADLRQNLE	0.71	1	No specific domain/motif analyzed, but F-Actin binding shown.[22]
NIN	ninein	Q8N4C6	755	763	VRGLTQELE	0.86	3	
OPHN1	oligophrenin-1	O60890	776	784	VASRTRFFE	0.87	2	Validates our motif, F-actin binding shown by Oligophrenin through AA 677-802.[23]
PCDH15	protocadherin-15	Q96QU1	1724	1731	IETERNIFE	0.64	3	
PDZD2	PDZ domain-containing protein 2	O15018	2335	2343	VKQRIKSFE	0.85	2	
PEAK1	inactive tyrosine-protein kinase PEAK1	Q9H792	702	710	VAQKVQEFN	0.79	3	Validates our motif, F-actin binding shown by inactive tyrosine-protein kinase PEAK1 through AA 339-727. [24]
PHC2	polyhomeotic-like protein 2	Q8IXK0	594	602	VGNLKKKYA	0.68	1	
PLCB2	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-2	Q00722	1040	1048	TKEMKKKLE	0.75	3	
PLEKHD1	pleckstrin homology domain-containing family D member 1	A6NEE1	396	404	VSHLKRFFE	0.83	2	
PLEKHG2	pleckstrin homology domain-containing family G member 2	Q9H7P9	817	825	VRELARLYS	0.83	2	
PLEKHG3	pleckstrin homology domain-containing	A1L390	1205	1213	VRNLREKFQ	0.74	1	

	family G member 3		926	934	VYQLARQYS	0.84	3	Validates our motif. F-actin binding suggested for AA 910 to 940. The others could not be found due to limited truncation design; By elongating the N terminus from 1 to 910 (instead of 1- to 600), one could have possibly still observed an actin binding (probably mediated by region 750 - 758). The same accounts for the C terminal truncations, where a construct from 940 - 1219 would also probably still show actin binding (instead of 940 - 1100).[25]
			750	758	VRNSISRFN	0.87	3	
PPP1R9A	neurabin-1	Q9ULJ8	139	147	FTETRKMFE	0.88	1	Validates our motif, F-actin binding shown by Neurabin-1 through AA 1-144.[26]
			202	210	VSQLSAVFE	0.87	1	
			120	128	VSERISRFD	0.85	1	
PPP1R9B	neurabin-2	Q96SB3	144	152	LQETRKLFE	0.62	1	Validates our motif, F-actin binding shown by Neurabin-2 through 2 regions, AA 1-154 and AA 151-282.[27]
			208	216	VSQLSAVFE	0.88	1	
PPRC1	peroxisome proliferator-activated receptor gamma coactivator-related protein 1	Q5VV67	1172	1180	LSSLLEQFE	0.60	1	
PXN	paxillin	P49023	217	225	VESLLDELE	0.79	3	
RAMAC	RNA guanine-N7 methyltransferase activating subunit	Q9BTL3	4	12	TAEAVPKFE	0.85	1	
RAPH1	ras-associated and pleckstrin homology domains-containing protein 1	Q70E73	862	870	VKQIASQFP	0.72	3	
RCSD1	capZ-interacting protein	Q6JBY9	20	28	VAQLAGRFR	0.84	1	
ROCK1	rho-associated protein kinase 1	Q13464	769	777	VKNLTLQLE	0.70	3	
SEPTIN9	septin-9 isoform 7	Q9UHD8-7	5	13	YSALKRSFE	0.60	1	Validates our motif, F-actin binding shown by Septin-9 through AA 1-143.[28]
SETD2	histone-lysine N-methyltransferase SETD2	Q9BYW2	2118	2126	TEERRKLFE	0.79	1	No binding shown, but methylates actin.[29]
SH2D4B	SH2 domain-containing protein 4B	Q5SQS7	247	255	VAGLSSMFR	0.87	1	
SH2D7	SH2 domain-containing protein 7	A6NKC9	119	127	LAELVHHYQ	0.63	3	
SHROOM3	protein Shroom3	Q8TF72	781	789	VLEKVSKFE	0.86	1	In line with our motif, F-actin binding shown by Shroom3 throughout the ASD1 region, which incorporated one part of our identified motif (AA 1054 - 1062).[30,31]
			1054	1062	VADRRRLFE	0.84	1	
SORBS1	sorbin and SH3 domain-containing protein 1 isoform 26	Q9BX66-12	657	665	VRDLASHFE	0.85	2	
			691	699	VSSRITAFE	0.87	1	
SORBS2	sorbin and SH3 domain-containing protein 2	O94875	473	481	VKSRILQYE	0.87	3	No specific domain/motif analyzed, but F-Actin colocalization shown.[32]
			506	514	VPTRISEFE	0.89	3	

SPECC1	cytospin-B	Q5M775	820	828	VKSLIKSFD	0.86	1	F-Actin colocalization shown, but thought to be mediated over other motifs (WH2).[33]
SPECC1L	cytospin-A	Q69YQ0	841	849	VKTLIKSFD	0.89	1	F-Actin colocalization shown, but thought to be mediated over other motifs (WH2).[33]
SPEG	striated muscle preferentially expressed protein kinase	Q15772	400	408	ILDKLQFFE	0.76	1	
SSH2	protein phosphatase Slingshot homolog 2	Q76I76	866	874	VRRATLEFE	0.83	3	No specific domain/motif analyzed, but F-Actin binding shown.[34]
			749	757	VKEIVTEIE	0.83	3	
SSH3	protein phosphatase Slingshot homolog 3	Q8TE77	612	620	VANRTQAFQ	0.68	2	No specific domain/motif analyzed, but F-Actin binding shown.[34]
STRN	striatin	O43815	329	337	ITKLKEQYK	0.82	3	
SVIL	supervillin	O95425	594	602	VAQLRSAFL	0.69	1	F-Actin colocalization shown, but thought to be mediated over other domains.[35–37]
SYCP1	synaptonemal complex protein 1	Q15431	218	226	IEKMITAFE	0.70	2	
TET3	methylcytosine dioxygenase TET3	Q8BG87	715	723	LEELIRQFE	0.75	1	
TIAM2	rho guanine nucleotide exchange factor TIAM2	Q8IVF5	296	304	LSSLRELYK	0.70	3	
TJP1	tight junction protein ZO-1	Q07157	1264	1272	VLTRVKMFE	0.90	3	Validates our motif, F-actin binding shown by tight junction protein ZO-1 through AA 1257-1284.[38]
TPRN	taperin	Q4KMQ1	133	141	VSRLLERFD	0.72	1	
USP54	ubiquitin carboxyl-terminal hydrolase 54	Q70EL1	1129	1137	VRSLAEQFQ	0.88	1	
XIRP1	xin actin-binding repeat-containing protein 1	Q702N8	693	701	VSRQKEVFQ	0.82	3	Validates our motif: defined in this study as a novel, repetitive actin-binding motif present in at least two different muscle proteins. [39]
			123	131	VQATSRKFE	0.88	2	
			91	99	VQCMRWIFE	0.86	3	
			342	350	VQQQQHLFE	0.79	1	
			438	446	**VKTFKNLFE**	0.87	2	
			1491	1499	**VQALRRLFE**	0.79	3	
			547	555	VGTARWLFE	0.85	1	
			153	161	**VRAARWLFE**	0.85	3	
			591	599	VQTIRWLFE	0.89	1	
XIRP2	xin actin-binding repeat-containing protein 2	A4UGR9	399	407	**VQSIRWIFE**	0.88	2	

			´1243	1251	**VKTSRQFFE**	0.87	3
			187	195	**TKLLKEQFE**	0.82	3
			364	372	**VQQARYVFE**	0.79	3
XIRP2	xin actin-binding repeat-containing protein 2 isoform 4	A4UGR9	114	122	**LDELRSVFE**	0.78	1

Table 1. Proteins identified by the SLiMFold pipeline as containing the short linear F-actin–binding motif (SFM). Each entry includes the gene symbol, protein name, UniProt ID, motif boundaries (Start–End), SFM sequence, the mean ipTM (averaged from the top three SLiMFold models), the SLiMFold iteration in which it was identified, and a brief comment providing functional insight or references.

References

1. Arai, A., Spencer, J. A. & Olson, E. N. STARS, a Striated Muscle Activator of Rho Signaling and Serum Response Factor-dependent Transcription. *Journal of Biological Chemistry* **277**, 24453–24459 (2002).
2. Schulze, J. O. *et al.* Structural and functional characterization of human Iba proteins. *The FEBS Journal* **275**, 4627–4640 (2008).
3. Banerjee, J., Fischer, C. C. & Wedegaertner, P. B. The Amino Acid Motif L/IIxxFE Defines a Novel Actin-Binding Sequence in PDZ-RhoGEF. *Biochemistry* **48**, 8032–8043 (2009).
4. Boeynaems, S. *et al.* Aberrant phase separation is a common killing strategy of positively charged peptides in biology and human disease. Preprint at https://doi.org/10.1101/2023.03.09.531820 (2023).
5. Zhan, Q. Q., Wong, S. S. & Wang, C. L. A calmodulin-binding peptide of caldesmon. *Journal of Biological Chemistry* **266**, 21810–21814 (1991).
6. D'Atri, F. & Citi, S. Cingulin interacts with F-actin in vitro. *FEBS Letters* **507**, 21–24 (2001).
7. Chrifi, I. *et al.* Cgnl1, an endothelial junction complex protein, regulates GTPase mediated angiogenesis. *Cardiovascular Research* **113**, 1776–1788 (2017).
8. Scherer, A. N., Anand, N. S. & Koleske, A. J. Cortactin stabilization of actin requires actin-binding repeats and linker, is disrupted by specific substitutions, and is independent of nucleotide state. *Journal of Biological Chemistry* **293**, 13022–13032 (2018).
9. Marat, A. L., Ioannou, M. S. & McPherson, P. S. Connecdenn 3/DENND1C binds actin linking Rab35 activation to the actin cytoskeleton. *MBoC* **23**, 163–175 (2012).
10. Wang, X. *et al.* DIXDC1 isoform, l-DIXDC1, is a novel filamentous actin-binding protein. *Biochemical and Biophysical Research Communications* **347**, 22–30 (2006).
11. Zhao, B.-W. *et al.* FBXO34 Regulates the G2/M Transition and Anaphase Entry in Meiotic Oocytes. *Front. Cell Dev. Biol.* **9**, 647103 (2021).
12. Estrada, L. Fgd1, the Cdc42 guanine nucleotide exchange factor responsible for faciogenital dysplasia, is localized to the subcortical actin cytoskeleton and Golgi membrane. *Human Molecular Genetics* **10**, 485–495 (2001).

13. Hao, J.-J., Zhu, J., Zhou, K., Smith, N. & Zhan, X. The Coiled-coil Domain Is Required for HS1 to Bind to F-actin and Activate Arp2/3 Complex. *Journal of Biological Chemistry* **280**, 37988–37994 (2005).
14. Johnson, H. W. & Schell, M. J. Neuronal IP_3 3-Kinase is an F-actin–bundling Protein: Role in Dendritic Targeting and Regulation of Spine Morphology. *MBoC* **20**, 5166–5180 (2009).
15. Brehm, M. A., Schreiber, I., Bertsch, U., Wegner, A. & Mayr, G. W. Identification of the actin-binding domain of Ins(1,4,5) *P* 3 3-kinase isoform B (IP3K-B). *Biochemical Journal* **382**, 353–362 (2004).
16. Inokuchi, J. *et al.* Deregulated expression of KRAP, a novel gene encoding actin-interacting protein, in human colon cancer cells. *J Hum Genet* **49**, 46–52 (2004).
17. Rivera Alvarez, J. *et al.* The kinesin Kif21b regulates radial migration of cortical projection neurons through a non-canonical function on actin cytoskeleton. *Cell Reports* **42**, 112744 (2023).
18. Abé, T. *et al.* Ladinin-1 in actin arcs of oral squamous cell carcinoma is involved in cell migration and epithelial phenotype. *Sci Rep* **14**, 22778 (2024).
19. Maul, R. S. *et al.* EPLIN regulates actin dynamics by cross-linking and stabilizing filaments. *The Journal of Cell Biology* **160**, 399–407 (2003).
20. Wong, M. J., Malapitan, I. A., Sikorski, B. A. & Jongstra, J. A cell-free binding assay maps the LSP1 cytoskeletal binding site to the COOH-terminal 30 amino acids. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1642**, 17–24 (2003).
21. Cervero, P., Wiesner, C., Bouissou, A., Poincloux, R. & Linder, S. Lymphocyte-specific protein 1 regulates mechanosensory oscillation of podosomes and actin isoform-based actomyosin symmetry breaking. *Nat Commun* **9**, 515 (2018).
22. Sandeep, P. *et al.* Neuron navigators: A novel frontier with physiological and pathological implications. *Molecular and Cellular Neuroscience* **127**, 103905 (2023).
23. Fauchereau, F. The RhoGAP activity of OPHN1, a new F-actin-binding protein, is negatively controlled by its amino-terminal domain. *Molecular and Cellular Neuroscience* **23**, 574–586 (2003).

24. Kelber, J. A. & Klemke, R. L. PEAK1, a Novel Kinase Target in the Fight Against Cancer. *Oncotarget* **1**, 219–223 (2010).

25. Nguyen, T. T. T. *et al.* PLEKHG3 enhances polarized cell migration by activating actin filaments at the cell front. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10091–10096 (2016).

26. Oliver, C. J. *et al.* Targeting Protein Phosphatase 1 (PP1) to the Actin Cytoskeleton: the Neurabin I/PP1 Complex Regulates Cell Morphology. *Molecular and Cellular Biology* **22**, 4690–4701 (2002).

27. Barnes, A. P., Smith, F. D., VanDongen, H. M., VanDongen, A. M. J. & Milgram, S. L. The identification of a second actin-binding region in spinophilin/neurabin II. *Molecular Brain Research* **124**, 105–113 (2004).

28. Smith, C. *et al.* Septin 9 Exhibits Polymorphic Binding to F-Actin and Inhibits Myosin and Cofilin Activity. *Journal of Molecular Biology* **427**, 3273–3284 (2015).

29. Maas, M. N. *et al.* Examining prestructured β-actin peptides as substrates of histidine methyltransferase SETD3. *Sci Rep* **14**, 26439 (2024).

30. Hildebrand, J. D. & Soriano, P. Shroom, a PDZ Domain–Containing Actin-Binding Protein, Is Required for Neural Tube Morphogenesis in Mice. *Cell* **99**, 485–497 (1999).

31. Yeo, N. C. *et al. Shroom3* contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res.* **25**, 57–65 (2015).

32. Fredriksson-Lidman, K., Van Itallie, C. M., Tietgens, A. J. & Anderson, J. M. Sorbin and SH3 domain-containing protein 2 (SORBS2) is a component of the acto-myosin ring at the apical junctional complex in epithelial cells. *PLoS ONE* **12**, e0185448 (2017).

33. Mehta, V. *et al.* SPECC1L binds the myosin phosphatase complex MYPT1/PP1β and can regulate its distribution between microtubules and filamentous actin. *Journal of Biological Chemistry* **299**, 102893 (2023).

34. Niwa, R., Nagata-Ohashi, K., Takeichi, M., Mizuno, K. & Uemura, T. Control of Actin Reorganization by Slingshot, a Family of Phosphatases that Dephosphorylate ADF/Cofilin. *Cell* **108**, 233–246 (2002).

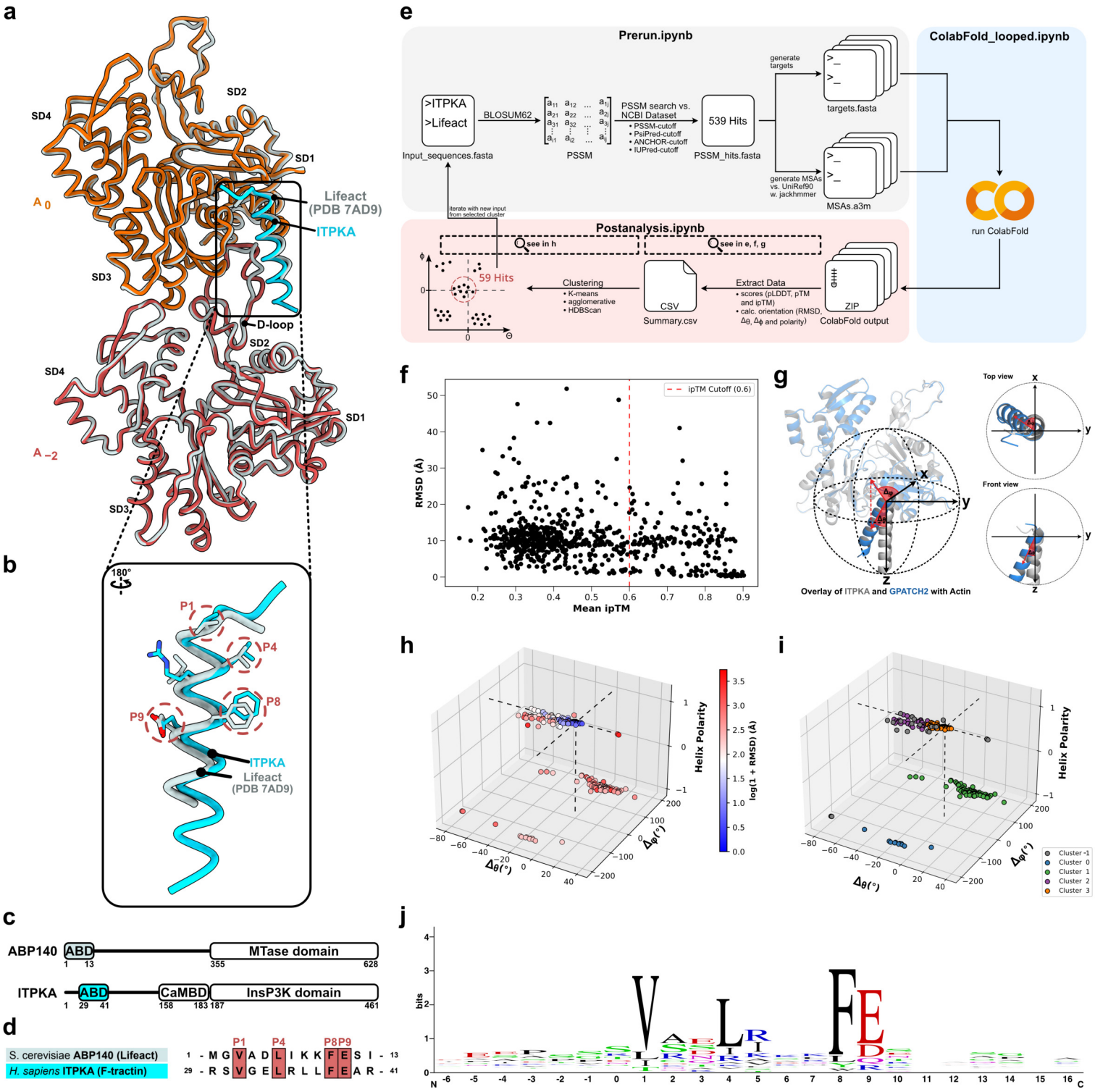
35. Chen, Y. *et al.* F-actin and Myosin II Binding Domains in Supervillin. *Journal of Biological Chemistry* **278**, 46094–46106 (2003).

36. Pestonjamasp, K. N., Pope, R. K., Wulfkuhle, J. D. & Luna, E. J. Supervillin (p205): A Novel Membrane-associated, F-Actin–binding Protein in the Villin/Gelsolin Superfamily. *The Journal of Cell Biology* **139**, 1255–1269 (1997).

37. Wulfkuhle, J. D. *et al.* Domain analysis of supervillin, an F-actin bundling plasma membrane protein with functional nuclear localization signals. *Journal of Cell Science* **112**, 2125–2136 (1999).

38. Belardi, B. *et al.* A Weak Link with Actin Organizes Tight Junctions to Control Epithelial Permeability. *Developmental Cell* **54**, 792-804.e7 (2020).

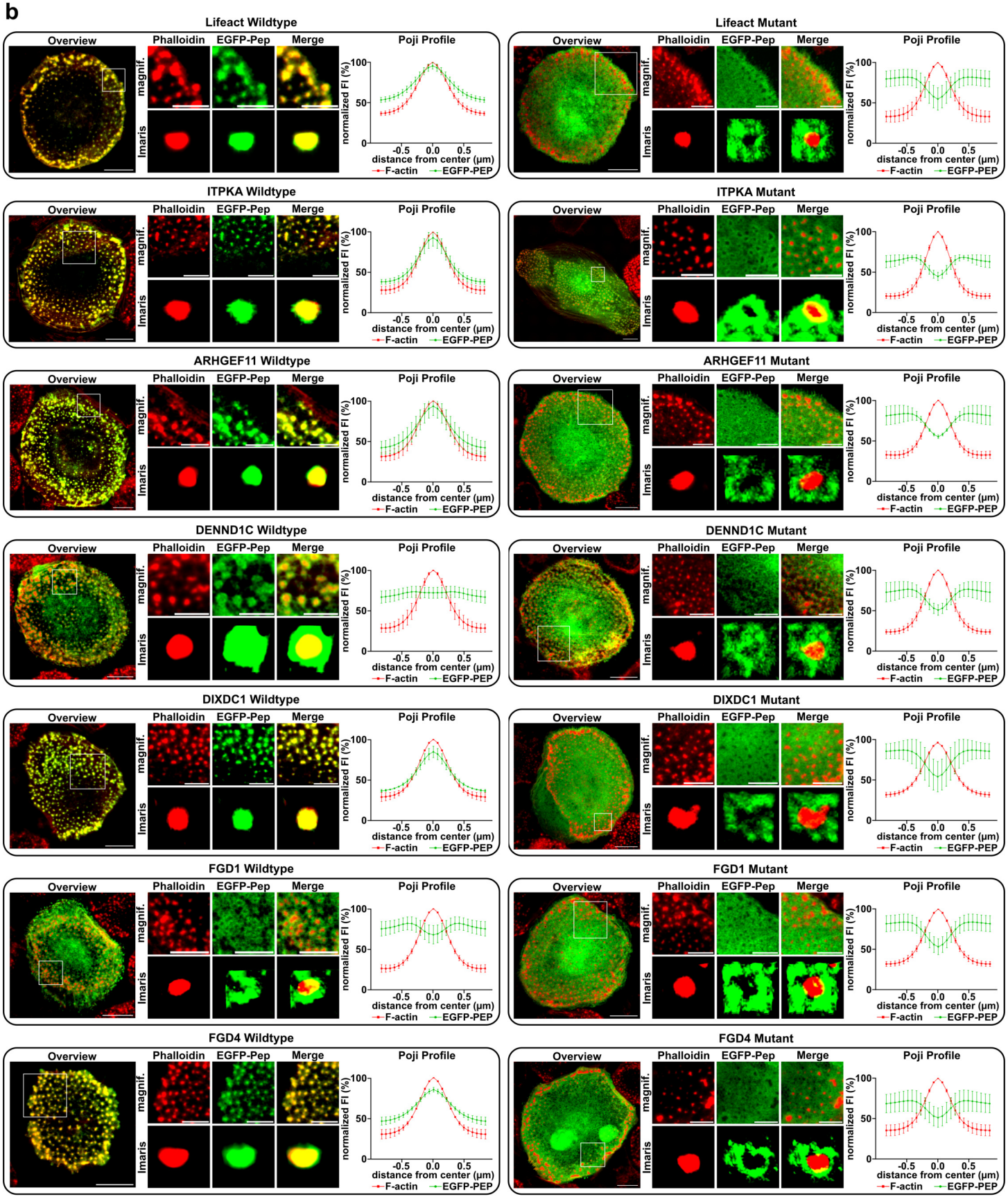
39. Pacholsky, D. *et al.* Xin repeats define a novel actin-binding motif. *Journal of Cell Science* **117**, 5257–5268 (2004).

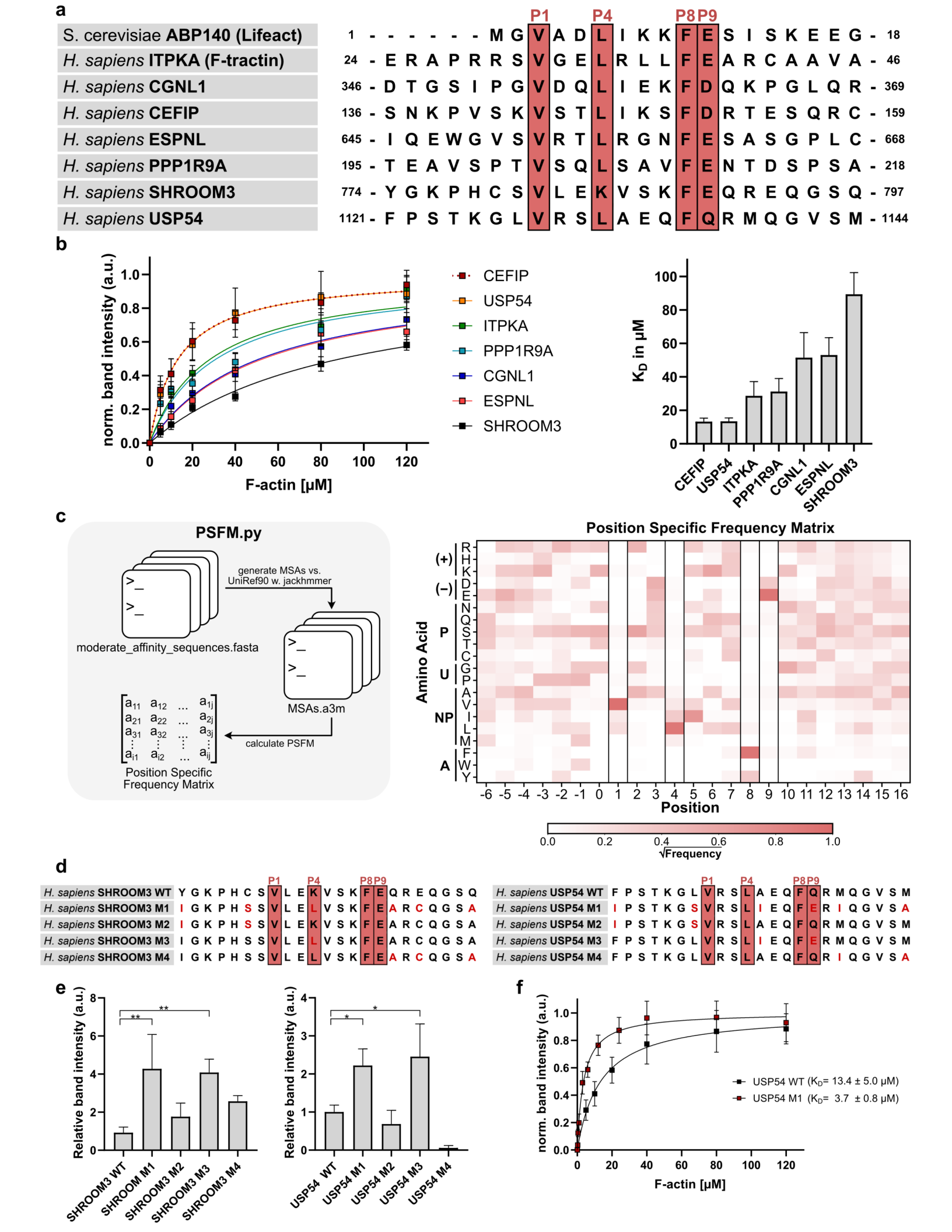

a
SD4
SD2
SD1
A 0
Lifeact
(PDB 7AD9)
ITPKA
SD3
D-loop
SD2
SD4
SD1
A −2
SD3
b
180°
P1
P4
P8
P9
ITPKA
Lifeact
(PDB 7AD9)
e
Prerun.ipynb
>ITPKA
>Lifeact
Input_sequences.fasta
BLOSUM62
PSSM
PSSM search vs.
NCBI Dataset
• PSSM-cutoff
• PsiPred-cutoff
• ANCHOR-cutoff
• IUPred-cutoff
539 Hits
PSSM_hits.fasta
generate
targets
targets.fasta
generate MSAs
vs. UniRef90
w. jackhmmer
MSAs.a3m
ColabFold_looped.ipynb
run ColabFold
iterate with new input
from selected cluster
Postanalysis.ipynb
see in h
see in e, f, g
59 Hits
Clustering
• K-means
• agglomerative
• HDBScan
CSV
Summary.csv
Extract Data
• scores (pLDDT, pTM
and ipTM)
• calc. orientation (RMSD,
Δθ, Δϕ and polarity)
ZIP
ColabFold output
f
ipTM Cutoff (0.6)
RMSD (Å)
Mean ipTM
g
Top view
Front view
Overlay of ITPKA and GPATCH2 with Actin
h
Helix Polarity
log(1 + RMSD) (Å)
Δθ(°)
Δφ(°)
i
Helix Polarity
Cluster -1
Cluster 0
Cluster 1
Cluster 2
Cluster 3
c
ABP140
ABD
MTase domain
ITPKA
ABD
CaMBD
InsP3K domain
d
S. cerevisiae ABP140 (Lifeact)
H. sapiens ITPKA (F-tractin)
P1
P4
P8P9
j
hits

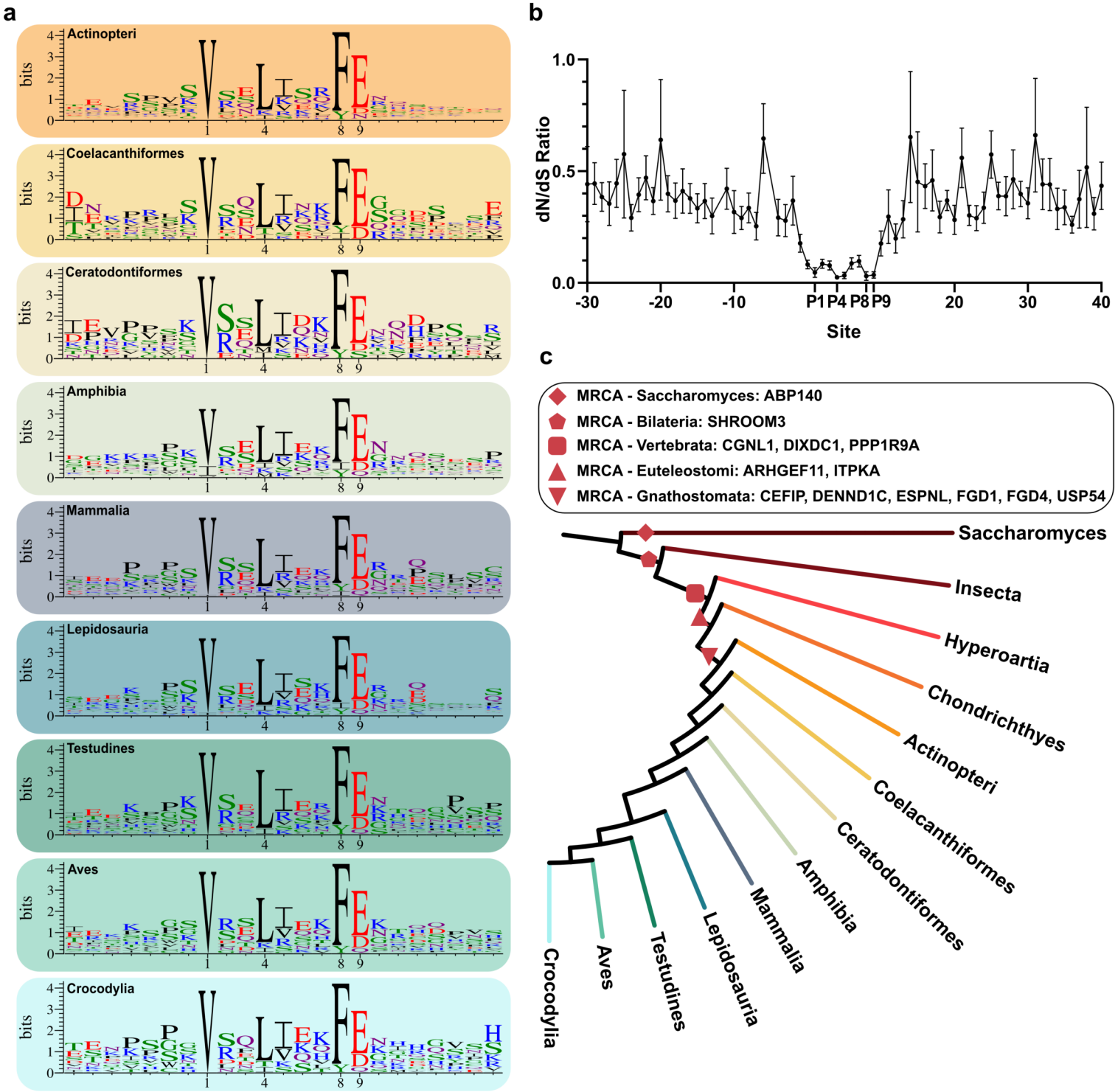
a

		P1	P4	P8 P9	
S. cerevisiae **ABP140 (Lifeact)**	1	- - - - - - M G V A D L I K K F E S I S K E E G -			18
H. sapiens **ITPKA (F-tractin)**	24	- E R A P R R S V G E L R L L F E A R C A A V A -			46
H. sapiens **ARHGEF11**	558	- V E V K P G N V R N I I Q H F E N N Q Q Y D A -			580
H. sapiens **DENND1C**	785	- Q P S S R P R V A D L K K C F E G - - - - - - -			806
H. sapiens **DIXDC1**	191	- I E N P Y W S V R A L V Q Q Y E G Q Q R S P S -			213
H. sapiens **FGD1**	201	- A S P S S A A V S S L I E K F E R E P V I V A -			226
H. sapiens **FGD4**	13	- S K E K P S K V S D L I S R F E G G S S L S N -			35

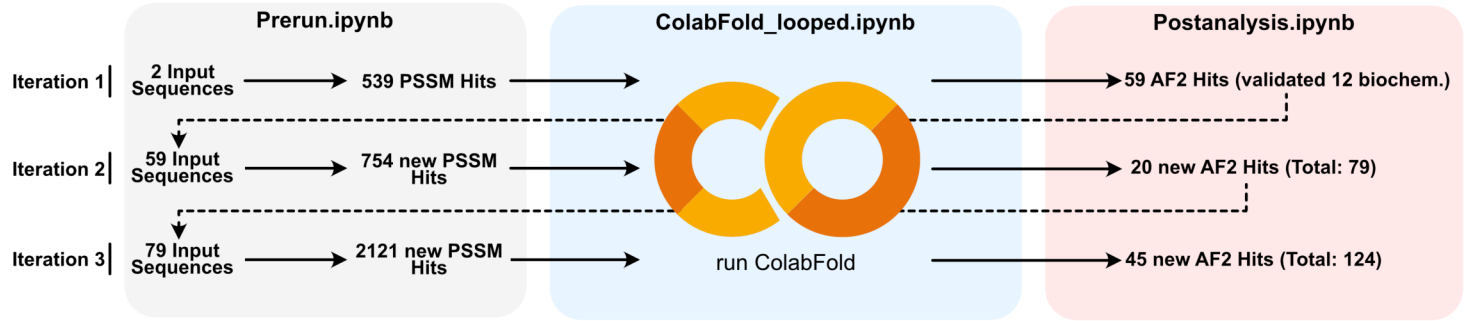
b



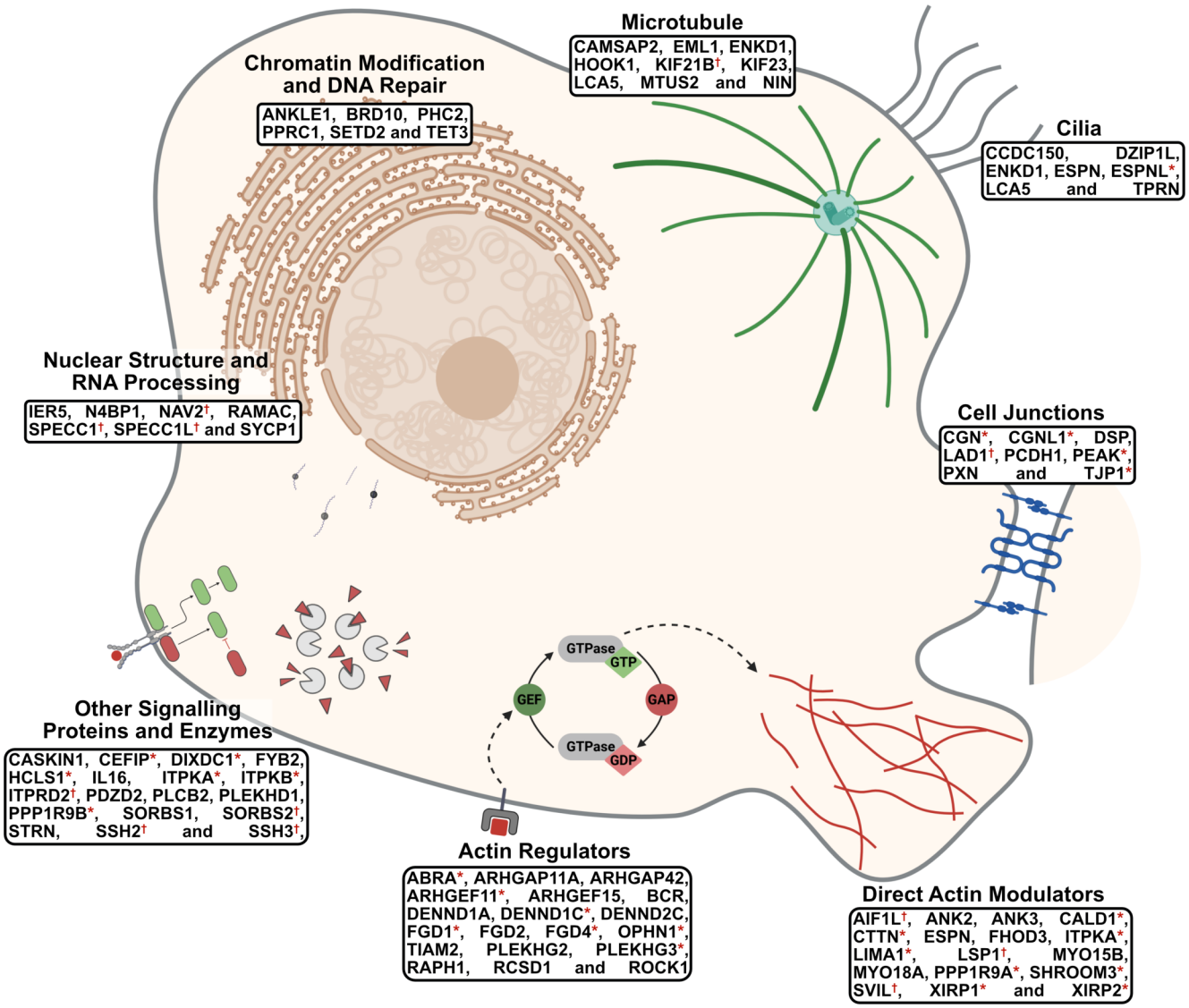

a
P1 P4 P8 P9
S. cerevisiae ABP140 (Lifeact) 1 - - - - - - M G V A D L I K K F E S I S K E E G - 18
H. sapiens ITPKA (F-tractin) 24 - E R A P R R S V G E L R L L F E A R C A A V A - 46
H. sapiens CGNL1 346 - D T G S I P G V D Q L I E K F D Q K P G L Q R - 369
H. sapiens CEFIP 136 - S N K P V S K V S T L I K S F D R T E S Q R C - 159
H. sapiens ESPNL 645 - I Q E W G V S V R T L R G N F E S A S G P L C - 668
H. sapiens PPP1R9A 195 - T E A V S P T V S Q L S A V F E N T D S P S A - 218
H. sapiens SHROOM3 774 - Y G K P H C S V L E K V S K F E Q R E Q G S Q - 797
H. sapiens USP54 1121 - F P S T K G L V R S L A E Q F Q R M Q G V S M - 1144
b
norm. band intensity (a.u.)
F-actin [µM]
CEFIP
USP54
ITPKA
PPP1R9A
CGNL1
ESPNL
SHROOM3
K_D in µM
c
PSFM.py
generate MSAs vs. UniRef90 w. jackhmmer
moderate_affinity_sequences.fasta
MSAs.a3m
calculate PSFM
Position Specific Frequency Matrix
Amino Acid
Position
√Frequency
d
H. sapiens SHROOM3 WT Y G K P H C S V L E K V S K F E Q R E Q G S Q
H. sapiens SHROOM3 M1 I G K P H S S V L E L V S K F E A R C Q G S A
H. sapiens SHROOM3 M2 I G K P H S S V L E K V S K F E A R C Q G S A
H. sapiens SHROOM3 M3 I G K P H S S V L E L V S K F E A R C Q G S A
H. sapiens SHROOM3 M4 I G K P H S S V L E L V S K F E A R C Q G S A
H. sapiens USP54 WT F P S T K G L V R S L A E Q F Q R M Q G V S M
H. sapiens USP54 M1 I P S T K G S V R S L I E Q F E R I Q G V S A
H. sapiens USP54 M2 I P S T K G S V R S L A E Q F Q R M Q G V S M
H. sapiens USP54 M3 F P S T K G L V R S L I E Q F E R M Q G V S M
H. sapiens USP54 M4 F P S T K G L V R S L A E Q F Q R I Q G V S A
e
Relative band intensity (a.u.)
SHROOM3 WT
SHROOM M1
SHROOM3 M2
SHROOM3 M3
SHROOM3 M4
USP54 WT
USP54 M1
USP54 M2
USP54 M3
USP54 M4
f
norm. band intensity (a.u.)
F-actin [µM]
USP54 WT (K_D= 13.4 ± 5.0 µM)
USP54 M1 (K_D= 3.7 ± 0.8 µM)


a
Actinopteri
Coelacanthiformes
Ceratodontiformes
Amphibia
Mammalia
Lepidosauria
Testudines
Aves
Crocodylia
bits
b
dN/dS Ratio
Site
P1 P4 P8 P9
c
MRCA - Saccharomyces: ABP140
MRCA - Bilateria: SHROOM3
MRCA - Vertebrata: CGNL1, DIXDC1, PPP1R9A
MRCA - Euteleostomi: ARHGEF11, ITPKA
MRCA - Gnathostomata: CEFIP, DENND1C, ESPNL, FGD1, FGD4, USP54
Saccharomyces
Insecta
Hyperoartia
Chondrichthyes
Actinopteri
Coelacanthiformes
Ceratodontiformes
Amphibia
Mammalia
Lepidosauria
Testudines
Aves
Crocodylia

a


Prerun.ipynb
ColabFold_looped.ipynb
Postanalysis.ipynb
Iteration 1
2 Input Sequences
539 PSSM Hits
59 AF2 Hits (validated 12 biochem.)
Iteration 2
59 Input Sequences
754 new PSSM Hits
20 new AF2 Hits (Total: 79)
Iteration 3
79 Input Sequences
2121 new PSSM Hits
run ColabFold
45 new AF2 Hits (Total: 124)

b


Microtubule
CAMSAP2, EML1, ENKD1, HOOK1, KIF21B†, KIF23, LCA5, MTUS2 and NIN
Chromatin Modification and DNA Repair
ANKLE1, BRD10, PHC2, PPRC1, SETD2 and TET3
Cilia
CCDC150, DZIP1L, ENKD1, ESPN, ESPNL*, LCA5 and TPRN
Nuclear Structure and RNA Processing
IER5, N4BP1, NAV2†, RAMAC, SPECC1†, SPECC1L† and SYCP1
Cell Junctions
CGN*, CGNL1*, DSP, LAD1†, PCDH1, PEAK*, PXN and TJP1*
GTPase
GTP
GEF
GAP
GTPase
GDP
Other Signalling Proteins and Enzymes
CASKIN1, CEFIP*, DIXDC1*, FYB2, HCLS1*, IL16, ITPKA*, ITPKB*, ITPRD2†, PDZD2, PLCB2, PLEKHD1, PPP1R9B*, SORBS1, SORBS2†, STRN, SSH2† and SSH3†,
Actin Regulators
ABRA*, ARHGAP11A, ARHGAP42, ARHGEF11*, ARHGEF15, BCR, DENND1A, DENND1C*, DENND2C, FGD1*, FGD2, FGD4*, OPHN1*, TIAM2, PLEKHG2, PLEKHG3*, RAPH1, RCSD1 and ROCK1
Direct Actin Modulators
AIF1L†, ANK2, ANK3, CALD1*, CTTN*, ESPN, FHOD3, ITPKA*, LIMA1*, LSP1†, MYO15B, MYO18A, PPP1R9A*, SHROOM3*, SVIL†, XIRP1* and XIRP2*