A proteome-wide screen uncovers diverse roles for sequence context surrounding proline-rich motifs in Ena/VASP molecular recognition

Theresa Hwang[1], Sara S Parker[2], Samantha M Hill[2], Meucci W Ilunga[1], Robert A Grant[1], Venkatesh Sivaraman[1], Ghassan Mouneimne[2], and Amy E. Keating[1,3]*

1. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA
2. Department of Cellular and Molecular Medicine, University of Arizona Cancer Center, University of Arizona, Tucson, AZ
3. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA

*correspondence to keating@mit.edu

Abstract

Protein interactions between intrinsically disordered, short linear motifs (SLiMs) and modular recognition domains are critical elements in many signal transduction pathways. Yet for most interactions, it is still unclear how low-complexity SLiMs discriminate between highly conserved but biologically distinct SLiM-binding proteins. The Ena/VASP family of proteins pose one such specificity problem. Paralogs ENAH, VASP, and EVL bind to a 5-residue SLiM prevalent in the proteome and perform distinct cellular functions despite sharing high sequence and structural similarity. To interrogate how the sequence context of SLiMs impacts Ena/VASP interactions, we performed an unbiased proteomic screen against the ENAH EVH1 domain. We discovered unexpected ways in which local and distal sequence elements flanking native SLiMs modulate binding. Particularly notable is a peptide from PCARE that achieves paralog specificity by stabilizing a unique conformation adopted only by ENAH. A PCARE-derived peptide can selectively recruit ENAH and block its activity in cells, establishing it as a valuable reagent to disentangle the roles of Ena/VASP paralogs and a potential agent for modulating ENAH function in breast cancer. Guided by our analyses of native interactions, we designed the tightest known ENAH EVH1 binder with a dissociation constant of 50 nM and 400-600-fold selectivity over EVL and VASP. Our work demonstrates that sequence context plays a prominent role in dictating SLiM interactions and can inform the design of custom molecules that target SLiM-based interactions.

Main Text

Introduction

Interactions between peptide recognition domains (PRDs) and short linear motifs (SLiMs) direct a broad range of intracellular functions, from protein trafficking to substrate targeting for post-translational modifications. To faithfully propagate these diverse signals, SLiMs must recognize the correct PRD within the crowded cellular environment, but how interaction specificity is achieved is enigmatic. SLiMs are degenerate and have low complexity, meaning they are

defined by just a few key residues or motif features. Crystal structures of PRDs such as SH3, WW, and PDZ domains bound to SLiMs typically reveal just 3-6 residues docked into a shallow groove (1-3). In addition, the explosive expansion of PRD families in metazoan proteomes has led to myriad domains that share overlapping SLiM-binding specificity profiles yet carry out distinct functions in the cell (4). How high-fidelity interactions are maintained between low complexity SLiMs and cognate PRDs remains poorly understood for many pathways.

There is increasing evidence that the sequence context surrounding SLiMs can critically influence interaction affinity and specificity (5-7). However, most SLiM research has centered around defining the “core SLiM”, or the minimal set of amino acids sufficient to bind to a given PRD. High throughput approaches such as peptide phage display have been instrumental in this regard but have not been applied to systematically probe how sequence context affects interactions (8-10). Clarifying the role of sequence context is essential for understanding and manipulating SLiM-based protein-protein interactions (PPIs).

Ena/VASP proteins are cytoskeletal remodeling factors that are recruited to different regions of the cell by binding proline-rich SLiMs via their N-terminal EVH1 domains and by promoting actin polymerization via their C-terminal EVH2 domains. The family is implicated in numerous cellular functions such as axon guidance and cell adhesion (11, 12). The Ena/VASP EVH1 domain recognizes the SLiM [FWYL]PXΦP, where X is any amino acid and Φ is any hydrophobic residue (13). This motif, referred throughout this paper as the *FP4 motif*, adopts a polyproline type II (PPII) helix structure and binds weakly to the EVH1 domain (14). Searching for this core FP4 motif in the human proteome yields 4994 instances. Although not all of these motifs have the opportunity to engage Ena/VASP proteins as spatial, structural, and temporal context are important determinants for cellular interaction (15), the number of potential interaction partners is large and raises the question of whether sequence determinants beyond the FP4 SLiM affect molecular recognition specificity.

Ena/VASP EVH1 domains are part of a larger family with that includes members in WASP, SPRED, and Homer proteins, many of which converge at the actin cytoskeleton and have unique functions but overlapping core SLiM specificity profiles (16). The extent to which these domains cross-react or bind selectively in the cell is unknown. Even within the Ena/VASP family, the three mammalian Ena/VASP paralogs, ENAH (also known as Mena), VASP, and EVL, have distinct functions and binding partners, despite binding the same FP4 SLiM. For example, reduced expression of ENAH or VASP decreases invasive potential in migratory breast cancer cells (17, 18), whereas EVL suppresses breast cancer cell invasion (19). Previous work has explored the development of ENAH EVH1 inhibitors (20, 21), but previously reported molecules lack high paralog selectivity. This is not surprising, given that the three proteins are 100% identical in sequence in the core FP4 binding groove and share 62-72% sequence similarity throughout the EVH1 domain. Nevertheless, examples of paralog specific Ena/VASP EVH1 binders exist (22), raising the question of how common this is and how binding selectivity can be achieved between such similar domains.

In this study, we investigated how the sequence context surrounding the core Ena/VASP FP4 SLiM affects binding specificity in the proteome. We leveraged an unbiased, screening method called MassTitr to identify 36-residue human proteome-derived peptides that bind to the ENAH EVH1 domain with a range of affinities. By analyzing features of high-affinity binders, we identified three distinct ways in which local and distal sequence elements surrounding proteomic FP4 SLiMs impact binding affinity and specificity to ENAH. We also uncovered strategies the ENAH EVH1 domain has evolved to recognize extended SLiMs, including a network of dispersed residues that allows ENAH to adopt a conformation inaccessible to EVL and VASP to promote paralog-specific binding. We combined multiple features used in nature to design "superbinders" with low-nanomolar affinities for ENAH, which is implicated in breast cancer progression (17) Our work demonstrates the importance of considering sequence context when investigating SLiM-mediated interactions, and our experimental and computational pipeline serves as a valuable blueprint to map, predict, and manipulate similar protein-protein interactions.

Results

MassTitr identifies ENAH EVH1 domain binding peptides from the human proteome with a wide range of affinities

To identify ENAH EVH1 binders in the human proteome, we used a screening method called MassTitr. MassTitr is a SORT-SEQ method similar to Tite-seq (23) that is based on fluorescence-activated cell sorting (FACS) of a library of peptide-displaying bacteria and subsequent deconvolution of signals by deep sequencing. Briefly, as shown in Figures 1A and B, peptide-displaying *Escherichia coli* cells are sorted into bins according to their binding signals across a range of protein concentrations, and the binding signal for each peptide at each protein concentration is extracted by deep sequencing each bin. This method allows for the identification of binders with a broad range of affinities.

Using MassTitr, we screened a library of 416,611 36-mer peptides with 7-residue overlaps (the T7-pep library) (24). This library spans the entire protein-coding space of the human genome, and we hypothesized that the long length of the encoded peptides would illuminate the impact of sequence surrounding the FP4 motif in a biologically relevant sequence space. We first prescreened the library for binding to a tetramerized ENAH EVH1 domain, which contained the ENAH EVH1 domain fused to the endogenous ENAH coiled coil as shown in Figure 1A, generating an input library enriched in binders. We then ran MassTitr on the prescreened library, using 8 concentrations of ENAH EVH1 tetramer. After sorting, sequencing, and filtering based on read counts, 108 unique high-confidence binders were identified (Figure 1C, Table S1) and classified as either high-affinity or low-affinity as described in the methods.

We validated 16 peptide hits for binding to monomeric ENAH EVH1 domain, using biolayer interferometry (BLI) to determine dissociation constants that ranged from 0.19 μM to 63 μM (Table S2). With the exception of SHROOM3 and

TENM1 peptides, binders classified as high-affinity by MassTitr bound to ENAH EVH1 domain more tightly than peptides classified as low-affinity. Many newly identified peptides bound with affinities similar to or tighter than a well-studied control peptide from *Listeria monocytogenes* protein ActA, which bound with K_D = 5.2 μM in our BLI assay (Table S2). Prior to this work, this single FP4-motif containing sequence from ActA was the tightest known endogenously derived binder of Ena/VASP EVH1 domains (13). Although the majority of MassTitr hits contained FP4 motifs (Figures 1C, D), 40 out of the 108 high-confidence hits did not (Figure 1C). We validated several of these noncanonical peptides (no FP4 motif) to bind to the ENAH EVH1 with mid-micromolar affinity (Table S2, Figure S1). Our results support increasing evidence that the ENAH EVH1 domain can bind sequences beyond the FP4 motif (22; 25, 26).

MassTitr peptides expand the ENAH signaling network

MassTitr detects ENAH EVH1 domain interactions with short peptides displayed on the cell surface. To highlight putative biologically relevant interaction partners of ENAH, we applied a bioinformatic analysis to identify those motifs that are likely to be accessible and co-localized with ENAH. We filtered our high-confidence hits by disorder propensity (IUPred2A > 0.4) (27) and cytoplasmic subcellular localization (28, 29). This resulted in 33 peptides, of which 14 are derived from interaction partners known to interact or co-localize with an Ena/VASP protein (Table S1). Filtered hits were highly enriched in GO biological process terms including actin filament organization (FDR < 10^{-6}), and positive regulation of cytoskeleton (FDR < .05) (30), which align with documented cellular functions of ENAH. Notably, we also identified proteins localized to the Golgi body and cilia, where Ena/VASP function is not well characterized (Figure 1E) (31,32).

A proline-rich C-terminal flank binds to a novel site on the EVH1 domain to enhance affinity in ENAH interaction partners

We used MassTitr data to identify FP4 SLiM-flanking elements that enhance binding to the ENAH EVH1 domain. A sequence logo made of the high-confidence MassTitr hits shows enrichment of prolines C-terminal to the FP4 motif, and a binomial test confirms that peptides containing FP4 motifs followed by three consecutive prolines are enriched our hit list ($p < 10^{-11}$; Figure 1D). A peptide from ENAH interactor ABI1 (25, 33) was among the highest affinity ligands that we validated by BLI, with K_D = 4 μM (Table S2). ABI1 contains an FP4 motif followed by 4 prolines (referred to here as FP_8). Mutating FPPPPPPPP to FPPPPSSSS in the context of the ABI1 36-mer reduced affinity by approximately 4-fold (Table S3). Although this confirms that the C-terminal prolines enhance affinity, peptide FPPPPPPPP alone binds to the ENAH EVH1 domain with K_D = 29 μM, indicating that additional interactions contribute to the high affinity of the ABI1 36-mer. Previous studies have shown that acidic residues N- and C-terminal to the FP4 motif can enhance affinity (34) and we hypothesized that positively charged patches in ENAH could bind acidic residues

that flank the FP_8 segment in ABI1 (Figure S2). Indeed, truncating the N-terminal or C-terminal acidic flanks of the 36-residue ABI1 peptide decreased affinity by 3- or 2.5-fold, respectively (Table S3).

We solved a crystal structure of the ENAH EVH1-ABI1 peptide complex at 1.88 Å resolution. Only the FP_8 region was fully resolved in the structure under the crystallization conditions, which included high salt and low pH (Figure 2A). The peptide folds into a PPII helix with prolines 1, 4, and 7 (^{0}F**P**PP**P**PP**P**P^{8}) contacting the EVH1 surface (Figure 2B). The FP4 portion of the peptide binds the canonical FP4 groove, as observed in other structures (14, 35) whereas the 7th proline docks into a previously uncharacterized site on ENAH composed of Ala12, Phe32, and the aliphatic part of the side chain of Asn90 (Figure 2C). Notably, a similar binding site at the analogous location is used by the Homer EVH1 domain to bind the phenylalanine of PPXXF motifs (36). This site is relatively shallow in ENAH and modelling large aromatic acids at this position on the ABI1 peptide using Pymol leads to severe steric clashes. Homer contains a smaller Gly89 at the site of Asn90 in ENAH and can consequently accommodate the bulky Phe of the PPXXF motif (Figure 2D). These results reveal how individual EVH1 family members have evolved distinct pockets to recognize proline-rich sequences that extend beyond the core SLiM.

Distal sequence elements enhance ENAH EVH1 binding through multivalent interactions

Another enriched feature of MassTitr-identified binders, relative to the pre-screened input library, is the presence of multiple FP4 motifs (binomial test, $p < 10^{-22}$). Multi-motif hits highlight preferred spacings of approximately 5 or 15 residues between FP4 motifs (Figure 3A). Multiple motifs are also enriched in MassTitr high-affinity hits relative to all hits ($p < 0.02$), supporting the hypothesis that multiple FP4 motifs enhance affinity. We confirmed this experimentally by showing that binding was reduced by 3-6 fold when 36-mer sequences from LPP and NHSL1, which contain two FP4 motifs, were truncated to leave only one motif (Table 1, Figure 3B).

Zyxin, which contains four clustered FP4 motifs, has been shown to bind to the VASP EVH1 domain by contacting both the canonical FP4 site and a noncanonical site on the opposite side of the EVH1 domain (37). Interestingly, a crystal structure of the ENAH EVH1 domain bound to a single-FP4 motif peptide at the canonical site also contains a second single-FP4 peptide bound to the region corresponding to the noncanonical binding site in VASP (PDB 5NC7, 38). To test whether multi-FP4 peptides engage this noncanonical site, we designed ENAH EVH1 R47A. In the ActA peptide-bound structure (PDB 5NC7), ENAH Arg47 forms a bidentate hydrogen bond with a carbonyl on the PPII helix backbone of a single FP4-motif peptide. In addition, the analogous VASP Arg48 exhibits significant NMR HSQC chemical shifts upon titration with a multi-FP4 motif zyxin peptide (37). Thus, we predicted that the R47A mutation would disrupt back-site binding. Indeed, while the affinities of single-FP4 motif peptides from ActA and PCARE were minimally affected by this mutation, peptides

containing two FP4 motifs from zyxin, LPP, and NHSL1 experienced a 7-15-fold reduction in affinity to ENAH R47A relative to wild type (Table S4, Figure 3C).

Next, we investigated the stoichiometry of multi-motif peptide binding. Using isothermal titration calorimetry (ITC), we confirmed that dual-FP4 motif peptides from LPP and NHSL1 and the single-FP4 motif peptide from ActA fit well to a 1:1 binding model (Figure 3D). Interestingly, the ITC analysis showed that binding of the ActA-derived single-FP4 motif peptide was driven by favorable entropy, whereas binding of the NHSL1 and LPP dual-motif peptides was enthalpically driven. ActA, LPP, and NHSL1 peptides have similar binding free energies, but the entropic contribution to the dual motif interactions is ~10-fold less favorable (Figure 3E). These data are consistent with a model in which long, disordered dual-motif peptides pay an entropic penalty to wrap around the EVH1 domain and engage two sites but gain enthalpic binding energy from additional interactions across the EVH1 interface.

Finally, we examined the motif-spacing requirements of bivalent binding. We used Rosetta to build a peptide chain to connect single FP4-motif peptides bound to the canonical and noncanonical sites of the ENAH EVH1 domain. There are two different orientations that preserve the directionalities of the bound FP4 motifs observed in structure 5NC7 (Figure S3). 10 residues were required to span the two motifs in orientation 1, whereas 9 residues were sufficient in orientation 2 (Figure S3A).

FP4 motif flanking elements in retinal protein PCARE confer high affinity by inducing non-canonical binding

The highest affinity peptide that we discovered was from cilia-associated protein PCARE (K_D = 0.19 μM for 36-residue peptide PCARE$^{813-848}$; Table S2). Successive truncations of this peptide identified the minimal region for high affinity binding (PCARE$^{826-848}$, which we call PCARE B, K_D = 0.32 μM, Figure S4A). We solved a crystal structure of ENAH EVH1 domain fused to the 36-mer PCARE sequence to 1.65 Å resolution. 21 residues of the 36-mer PCARE peptide were fully resolved in the electron density (PCARE$^{828-848}$, Figure 4A) and led to the surprising discovery that the LPPPP motif in PCARE binds at the expected canonical site, but in the opposite orientation from previously observed Ena/VASP EVH1 domains engaged with proline-rich peptides (13, 14, 34) (Figure 4B, C). PCARE$^{828-848}$ uses a 14-residue alpha helix-rich extension C-terminal to the LPPPP motif to make additional contacts in a novel pocket on the EVH1 domain, explaining its high affinity.

PCARE residues Phe843 and Leu846 make hydrophobic interactions with Ala83 and Pro65 on ENAH. The side chain of Asp840 on PCARE docks into a polar pocket on ENAH made up of the backbone of ENAH Lys69 and Arg81. Notably, the backbone NH and side chain hydroxyl group of Ser842 on PCARE form hydrogen bonds with the side chain of Asp840, positioning Asp840 to hydrogen bond with a water molecule that is coordinated by the backbone NH of ENAH Arg81. Most intriguing is the interaction between Val837 on PCARE and the hydrophobic groove in ENAH that typically binds to large aromatic such as

Phe. The alpha helical structure of PCARE$^{828-848}$ buries the smaller Val837 in the same site where Phe can bind, but in a different orientation (Figure 4A).

PCARE achieves paralog selectivity by stabilizing an ENAH EVH1 domain specific conformation with a novel FP4 flanking sequence element

Surprisingly, we found that PCARE B has a 70-140-fold preference for binding to ENAH over the VASP or EVL EVH1 domains (Figure 4D), yet our structure of ENAH EVH1 domain bound to PCARE$^{828-848}$ shows that most of the ENAH residues contacting PCARE are conserved in VASP and EVL (Figure 4C). Comparing the structures of ENAH EVH1 domain bound to PCARE$^{828-848}$ vs. FPPPP (PDB 1EVH) shows that a loop composed of ENAH residues 80-86 is shifted by 3 Å, forming the extended PCARE binding site (Figure 4E). EVL and VASP have bulkier residues in this region, and modeling on the ENAH-PCARE$^{828-848}$ structure predicted severe steric clashes if EVL or VASP were to adopt the PCARE$^{828-848}$-bound conformation (Figure 4E). We hypothesized that this conformational change in ENAH provides a mechanism for achieving specificity.

To identify the ENAH residues responsible for PCARE binding specificity, we used the structure-based modeling method dTERMen (39). dTERMen is a protocol for scoring the compatibility of a sequence with a backbone structure. Energies are computed based on the frequencies with which combinations of residues are found in tertiary motifs in known proteins structures. As expected, the EVL sequence scored considerably worse than ENAH on the PCARE$^{828-848}$-bound ENAH EVH1 template (Figure S5). Guided by the dTERMen score, we introduced increasing numbers of residues from ENAH into the sequence of EVL. Seven replacements were sufficient to recapitulate dTERMen energies similar to that for ENAH in the PCARE$^{828-848}$-bound conformation (Figure S5). These residues are distributed across the EVH1 domain, and several are distant from the PCARE$^{828-848}$ binding site (Figure 4F). We made a mutated version of EVL with the 7 corresponding residues from ENAH. This EVL mutant ($EVL_{swapped}$ in Figure 4F) bound as tightly to PCARE B as ENAH (K_D = 0.35 vs. 0.32 μM). Given that wild-type EVL and ENAH differ at 29 sites, and there are 1.56 million potential residue swaps of 7 residues, it is particularly notable that dTERMen quickly led us to a combination of residues sufficient to transfer binding specificity.

PCARE binds selectively to ENAH in cells and inhibits ENAH-dependent cell adhesion

To test whether PCARE B could selectively bind to ENAH in cells, we used Ena/VASP-family-deficient cell line MV^{D7} (40) and expressed the individual paralogs ENAH, EVL, and VASP to evaluate their interactions with PCARE. MV^{D7} are embryonic fibroblasts derived from *ENAH*$^{-/-}$ *VASP*$^{-/-}$ mice, with low *EVL* expression (41). We used an shRNA against *EVL* to further decrease residual EVL (Figure S5). We tagged PCARE B with mRuby2 (mRuby2-PCARE B) and co-expressed this construct with GFP, or GFP fusions to ENAH, an EVH1

deletion mutant of ENAH (ΔEVH1-ENAH), EVL, or VASP. ENAH, EVL, and VASP all robustly localized to focal adhesions as previously observed (42) while ΔEVH1-ENAH localization was cytoplasmic. mRuby2-PCARE B exhibited diffuse cytoplasmic localization under all conditions except when co-expressed with ENAH, in which case mRuby2-PCARE B was moderately enriched at focal adhesions (Figure 5A). Plots of fluorescence intensity along a line passing through focal adhesions show co-localization of PCARE B and ENAH (Figure 5B).

To test the interaction specificity of PCARE B and ENAH, PCARE B was tagged with a mitochondrial localization sequence, generating Mito-mRuby2-PCARE. In $MV^{D7}{}_{shEVL}$ cells with exogenously expressed Ena/VASP paralogs, ENAH was strongly colocalized with Mito-mRuby2-PCARE B at mitochondria, in addition to being enriched at focal adhesions, while EVL and VASP remained entirely localized to focal adhesions and did not localize to mitochondria (Figures 5C, D). Together, these findings indicate that PCARE B recruits ENAH, but not EVL or VASP, to artificial localization sites such as the mitochondria, supporting interaction specificity between PCARE B and ENAH in cells.

We also tested whether expression of PCARE B could disrupt the function of endogenous ENAH. We examined focal adhesion maturation as a readout of ENAH function in MCF7 breast cancer cells, which express significant levels of ENAH (Figure 5E). We compared the effects of PCARE B expression to the effects of ENAH knockdown by shRNA by assessing cell adhesion, using paxillin immunofluorescence labelling to delineate focal adhesions (Figure 5E). As expected, ENAH knockdown was associated with diminished adhesion and smaller focal adhesion size as compared to cells expressing non-targeting shRNA (Figures 5F-H). Intriguingly, expression of PCARE B resulted in a similar decrease in adhesion, suggesting suppression of ENAH function at focal adhesions (Figures 5F-H). Importantly, expression of PCARE B in MCF7 cells shifted the enrichment of ENAH from focal adhesions to membrane protrusions. ENAH enrichment at protrusions was not observed in MCF7 cells expressing Mito-PCARE B, which exhibited strong mitochondrial localization of ENAH and markedly decreased adhesion. This suggests that in MCF7 cells, blockade of the EVH1 domain by cytosolic PCARE B liberates ENAH from focal adhesions, while permitting other EVH1-independent interactions elsewhere in the cell. In contrast, Mito-PCARE B serves to sink ENAH away from its normal sites of action at focal adhesions and at the cell membrane.

Engineered binders engage ENAH EVH1 domain with increased affinity and specificity

Motivated by our observation that PCARE B can selectively sequester ENAH and inhibit its cellular functions, we designed even higher affinity peptides that can be used to further unravel Ena/VASP biology and serve as tools for developing therapies to target ENAH-dependent disease. Our rational design approach involved combining affinity-enhancing elements that target multiple sites on the ENAH EVH1 domain (Figure 6A). We first fused the ABI1 FP_8

sequence, along with its acidic flanking sequence, to a sequence derived from LPP that includes part of the LPP linker and a second FP4 motif (ABI1-LPP; Figure 6B). Our rationale was that the ABI1-derived segment would occupy the canonical FP4 binding site and the LPP linker would wrap along the surface of the EVH1 domain and position a second FP4 motif near the noncanonical FP4 site. The ABI1-LPP fusion peptide bound tighter (K_D = 0.76 μM) than did peptides corresponding to its constituent parts (Figure 6B). We also designed a peptide that fused the high affinity PCARE B sequence, via a short linker, to a second FP4 motif designed to support bivalent binding. Peptide PCARE-Dual bound ~7-fold tighter than PCARE B, with K_D = 50 nM (Figure 6B). The enhancements in affinity for ABI1-LPP and PCARE-Dual come from decreases in off-rate, with PCARE-Dual dissociating ~100-fold slower than dual-motif peptide LPP (Figure 6C). Finally, we found that PCARE-Dual is 400-600-fold selective for ENAH over EVL and VASP, even though VASP is also known to have a noncanonical binding site that can engage a second FP4 motif (36), providing the tightest and most specific known binder to the ENAH EVH1 domain to date (Figure 6D).

Discussion

In recent years, phage display screens of signaling domains against a 16-mer peptide library derived from the human proteome (Pro-PD) have been used to define SLiM specificity profiles and predict novel interaction partners (43-46). In this work, we used MassTitr to screen more than 400,000 36-residue segments of the human proteome against the cytoskeleton regulator ENAH and learned that examining such long fragments provides important insights into the role of SLiM sequence context. We identified distal sequence features up to 15 residues away from the core FP4 SLiM that are important for binding. Our study highlights how low-information SLiMs exploit sequence context to selectively recognize PRDs within the proteome and, by taking this observation into the realm of protein design, we demonstrated how SLiM-flanking sequences can be leveraged to generate tools that selectively perturb SLiM-PRD interactions.

Proteomic FP4 SLiMs exploit sequence context in distinct ways to recognize ENAH

We found multiple ways that sequence flanking the FP4 motif can modulate binding. We first demonstrated that prolines C-terminal to FP4 motifs can enhance binding by contacting a previously uncharacterized hydrophobic patch on ENAH. Both secondary structure and sequence are key to this binding mode, which positions the 7th proline of a ^{0}FPPPPPPP**P**P^{8} peptide to contact ENAH in a shallow groove that we refer to as the Pro7 binding pocket. The relatively flat surface of the EVH1 domain in this region limits the binding energy available from favorable contacts, but PPII helix preorganization minimizes the entropic cost of binding. We anticipate that this binding mode is widely exploited by ENAH cellular interaction partners of ENAH. Multiple previously annotated Ena/VASP interactors, including proteins identified in our screen such as FBLIM1, ZYX, and

LPP (47-49) contain FP4 motifs followed by trailing prolines, with either a leucine or proline in the 7th position (^{0}FPPPPPP**P**P^{8}).

Interestingly, the Homer EVH1 domain uses a site structurally analogous to the ENAH Pro7 binding pocket to accommodate Phe in the SLiM PPXXF (36). Thus, part of the core binding site in the Homer EVH1 domain is used by the ENAH EVH1 domain as a secondary affinity-enhancing site. This provides a striking example of how SLiMs can use flanking sequence to target diverged binding pockets in homologs to increase molecular recognition specificity.

We also demonstrated that multiple FP4 motifs in ENAH binders can enhance affinity. Many known Ena/VASP partners contain multiple FP4 motifs (50) but the mechanisms by which ENAH engages such peptides are not well characterized. Our data and previous work support a model of bivalent binding, where multi-FP4 motif peptides can wrap around a single EVH1 domain (37). Analysis of our multi-FP4 MassTitr hits showed preferential spacing of ~5 or 15 residues between FP4 motifs in a single chain, but FP4 motifs separated by 5 residues probably do not bind simultaneously to a single EVH1 domain, as structural modelling suggests that the minimum chain length required to span the two putative bindings sites is 9 residues. In such cases, it may be that two EVH1 domains bind to two closely spaced motifs (see one possible model in Figure S2B). Another possibility is that clustered FP4 motifs separated by only a few residues bind using mechanisms such as allovalency, where the increased effective concentration of multiple FP4 motifs close together enhance affinity (51).

The sequence requirements for ligands binding to the noncanonical site on ENAH are unknown, but we speculate that diverse sequences, particularly those with propensity to adopt a PPII helix conformation, could occupy this site when present at high effective concentration due to binding of a primary motif at the canonical site. In support of this, we saw that peptides from LPP and NHSL1 that lacked a second FP4 motif but contained at least one proline residue within 10 residues of a single-FP4 bound more tightly to wild-type ENAH EVH1 than to ENAH EVH1 R47A, which disrupts the non-canonical site (Table 1). Moreover, dual-FP4 peptides from LPP and NHSL1 in which the linker was duplicated bound with similar affinities as the WT peptides to the ENAH EVH1 domain, and also bound less tightly to ENAH EVH1 R47A. This is consistent with a model in which the linker itself can provide stabilizing interactions and the second FP4 motif is not essential for bivalent binding (Table 1). Our data suggest that either a second FP4 motif or hydrophobic linker residues can make favorable interactions with the noncanonical site.

SLiM context in PCARE leads to high-affinity binding in a reversed chain orientation

An additional way that sequence context can affect affinity is illustrated by a PCARE-derived peptide, the highest affinity hit from our screen. The structure of a PCARE peptide bound to ENAH shows that the high affinity results from additional contacts made by 14 residues C-terminal to an LPPPP motif. A

dramatic feature of the structure is the reversed orientation of the LPPPP motif in the FP4 binding groove relative to previously observed FP4 ligands. The PPII helix possess two-fold rotational pseudosymmetry, meaning that the side chains and backbone carbonyls are similarly positioned in either the N-to-C- or C-to-N-terminal directions. SH3 domains exploit this pseudosymmetry to bind proline-rich sequences in either direction (52). Although all Ena/VASP and Homer EVH1 domain structures solved so far show the PPII helix engaged in a single direction (16), the WASP EVH1 domain binds its proline-rich ligand with the chain reversed (53). Consequently, it has been hypothesized, although never demonstrated, that Ena/VASP EVH1 domains might also bind FP4 motifs in either direction (Ball et al., 2002). Here we demonstrate that this is the case for ENAH. We speculate that because PCARE contains the motif LPPPP, the smaller leucine contributes less binding energy than larger, aromatic residues (Phe, Tyr, Trp) when binding into the "Phe0" pocket of ENAH, relaxing the polarity of the LP4 motif and increasing the influence of flanking residues. It is possible that other LP4 motif-containing proteins bind to ENAH in the same orientation as PCARE, influenced by flanking sequence.

An epistatic network in ENAH contributes to paralog specificity by stabilizing a unique conformation

Paralogs ENAH, VASP, and EVL are highly conserved in sequence and structure, making it challenging to rationally target one paralog. There are currently two proteins reported to bind specifically to the ENAH EVH1 domain. The LIM domain of Testin binds to ENAH EVH1 with low micromolar affinity (K_D = 3 μM) and does not bind detectably to the EVL or VASP EVH1 domains. The binding site is in a region where residues among the paralogs differ (22). Synthetic miniprotein pGolemi binds to ENAH with K_D = 0.29 μM, moderate selectivity (20-fold) over VASP, and higher selectivity (≥120-fold) over EVL (54). The mechanism behind pGolemi's specificity has not been conclusively determined.

We showed that PCARE achieves high selectivity and affinity for ENAH over VASP and EVL by stabilizing an ENAH-specific conformation. This mechanism is reminiscent of how the cancer drug Gleevec achieves 3000-fold selectivity for Abl over Src, despite the high sequence conservation between the two kinases. The residues in Abl that are important for specificity are distal from the Gleevec binding site and were only identified by ancestral reconstruction (55). Here, we took an alternative approach to discovering residues in ENAH that contribute to PCARE specificity by using structure-based modelling. Scoring mutations for their compatibility with the PCARE-bound conformation of ENAH using dTERMen solved the challenging problem of identifying an epistatic network that contributes to function. Importantly, this method is easily generalizable to many protein systems as long as a structure is available.

The interaction of PCARE with ENAH provides an intriguing example of how nature has evolved proteins that achieve selectivity by exploiting a conformation accessible to only one paralog, rather than by contacting paralog-specific

residues. Our findings raise the question of how widely this mechanism of selectivity is exploited by other paralogous families. Although instances of this type of binding specificity are difficult to anticipate or understand without in-depth structural analysis, when paralog-specific conformations can be identified they provide an attractive route to designing highly specific inhibitors.

MassTitr hits expand ENAH biology

Filtering MassTitr hits for potential interactions based on localization and disorder yielded peptides from 33 putative binding partners. 18 proteins from this list have, to our knowledge, not been reported to associate with Ena/VASP proteins and provide avenues for further investigation. For example, one of our hits was from the protein KIAA1522, which is known to potentiate metastasis in esophageal carcinoma and breast cancer cells (56, 57). We validated that a noncanonical motif in KIAA1522 binds to the ENAH EVH1 domain (Table S2), potentially linking ENAH and KIAA1522 in tumor progression.

PCARE, a protein localized to ciliary structures, has been reported to associate with ENAH through tandem affinity purification mass spectrometry and we have now established the residues that mediate a direct interaction (58). Investigating the biological implications of PCARE's specificity towards ENAH will be of interest. ENAH, VASP, and EVL have been shown to induce distinct actin remodeling phenotypes based on their differential preferences for profilin isoforms (19). Thus, paralog selectivity could be necessary when paralog-dependent actin dynamics and structures are required. PCARE is expressed in retinal photoreceptor cells and interacts with actin modulator WASF3 (58). This interaction localizes actin remodeling to the base of the outer segment, a highly modified primary cilium, to help initiate new outer segment disk formation. Our results suggest that ENAH, but not VASP or EVL, could promote this process.

Engineered ENAH ligands are tools to probe ENAH-specific biology and to develop targeted therapies

We demonstrated that a peptide derived from PCARE interacts specifically with ENAH but not VASP and EVL. Cytoplasmic expression of this peptide effectively displaces ENAH from focal adhesions but not membrane protrusions, indicating that this peptide could serve as a valuable tool to selectively perturb EVH1-dependent functions of ENAH. In contrast, a mito-tagged version of this peptide selectively sequesters ENAH at the mitochondria, depleting it from sites of normal function and inhibiting ENAH-dependent biological processes such as cell adhesion. This is likely due to the high avidity of these peptides when localized to the mitochondrial surface, which could potentially outcompete interactions between the EVH2 domain of ENAH and actin-based structures (59).

Historically, a mito-tagged ActA peptide has served as a valuable sideways knockout tool in dissecting Ena/VASP dependent functions (40). We anticipate that PCARE B, and our engineered ligand PCARE-Dual, will serve in a similar vein, but with the added benefit of selectively perturbing ENAH function. There is emerging evidence in the literature that paralog-specific functions exist within the

Ena/VASP family. For example, ENAH, but not VASP or EVL, has been shown to be a regulator of local translation in neurons (60). In addition, EVL has a specific role in durotactic invasion (42). Our mito-tagged peptides will serve as valuable research reagents in uncovering and characterizing ENAH-specific molecular functions in other tissues, such as the retina.

Barone, et al. have shown that small-molecule Ena/VASP EVH1 domain inhibitors reduce breast cancer-mediated invasion in zebrafish models, and in *in vitro* invasion assays, establishing Ena/VASP EVH1 domains as promising therapeutic targets (21). Yet Ena/VASP paralogs have distinct functions in cancer. ENAH deficiency in mouse models of cancer has been shown to decrease metastasis by reducing tumor cell invasion and intravasation (17). On the other hand, EVL is known to promote the formation of cortical actin bundles that suppress invasion of breast cancer cells (61). Thus, developing high affinity, paralog-specific inhibitors of Ena/VASP proteins could be crucial in reducing pleiotropic side effects when treating metastasis. Our engineered PCARE-Dual ligand exhibits exquisite affinity and specificity for ENAH and is the tightest binder of the ENAH EVH1 domain known to date. We envision our engineered ligands, and our structure that reveals how PCARE binds to a paralog-specific EVH1 conformation, as promising leads for developing therapies to treat ENAH-dependent diseases.

Materials and Methods

FACS sample preparation and analysis

The protocol for sample preparation for FACS analysis and MassTitr sorting was adapted from Foight and Keating (62) and Whitney et al. (manuscript in preparation) and is as follows: 5 mL cell cultures of eCPX plasmid expressing either library or control peptide were grown overnight at 37 °C in LB + 25 µg/ml chloramphenicol + 0.2% w/v glucose. The next day cells were inoculated into fresh TB + 25 µg/mL chloramphenicol and grown at 37 °C. Upon reaching an O.D. 600 of 0.5-0.6, cells were induced with 0.04% w/v arabinose for 1.5 hours at 37 °C. The O.D. 600 was then remeasured, and enough cells were pelleted for analysis (1 x 10^7 cells per FACS analysis sample, 7 x 10^7 cells for library sorting). Cells were resuspended to a concentration of 4 x 10^8 cells/mL, washed in PBS + 0.1% BSA, and then incubated with anti-FLAG antibody conjugated to APC (αFLAG-APC; PerkinElmer) diluted 1:100 in PBS + 0.1% BSA at a ratio of 30 µL labelled antibody:10^7 cells. Tubes wrapped in foil were incubated at 4 °C for 15 minutes, then cells were washed with PBS + 0.1% BSA and pelleted. For each FACS analysis sample, 25 µL of 1 x 10^7 cells in PBS were mixed with 25 µL of a 2x concentration of ENAH tetramer in PBS + 1% BSA + 4 mM DTT (final concentration of 2 mM DTT), and then incubated at 4 °C for 1 hour in foil. After incubation, 50 µL of the mixture was added per well to a 96-well Multi-Screen HTS GV sterile filtration plate (Millipore), buffer was removed by vacuum, and then cells were washed twice with 200 µL of PBS + 0.5% BSA. Each well containing 1 x 10^7 cells was then resuspended in 30 µL of streptavidin-PE (SAV-

PE; ThermoFisher Scientific) diluted 1:100 in PBS + 0.1% BSA, and incubated for 15 minutes at 4 °C, washed with 200 µL of PBS + 0.1% BSA, and resuspended in 250 µL of PBS + 0.1% BSA for subsequent FACS analysis or sorting.

Biolayer interferometry (BLI)

All BLI experiments were performed with 2 replicates on an Octet Red96 instrument (ForteBio). Biotinylated, 6x-His-SUMO-peptide fusions purified in small scale were diluted into BLI buffer (PBS pH 7.4, 1% BSA, 0.1% Tween-20, 1 mM DTT) and immobilized onto streptavidin-coated tips (ForteBio) until loading reached a response level between 0.5-0.6 nm. The loaded tips were immersed in a solution of ENAH EVH1 domain at a relevant dilution series in BLI buffer at an orbital shake speed of 1000 rpm and data were collected until the binding signal plateaued. ENAH-bound tips were subsequently placed into BLI buffer for dissociation and data were collected until the binding signal plateaued. K_D values were obtained through steady state analysis. Briefly, the data were corrected for background by subtracting the signal obtained when doing the same experiment using biotinylated, 6x-His-SUMO lacking any peptide, instead of an immobilized ENAH-binding peptide. The association phases were then fit to a one-phase binding model in Prism and the equilibrium steady-state binding signal values from that fit were plotted against ENAH concentration and fit to a single-site binding model in Prism to obtain dissociation constants. Errors are reported as the standard deviation of two replicates.

Isothermal titration calorimetry (ITC)

ITC experiments were performed with two replicates using a VP-ITC microcalorimeter (MicroCal LLC). To prepare samples for ITC, 2.5 mL of 100 µM ENAH EVH1 domain and 1 mL of 800 µM-1.2 mM of SUMO-peptide fusions were dialyzed against 2 L of ITC Buffer (20 mM HEPES pH 7.6, 150 mM NaCl, 1 mM TCEP) at 4 °C overnight. The concentrations of proteins were remeasured after dialysis on the day of experiment. SUMO-peptide was titrated into ENAH EVH1 domain at 25 °C. Data analysis and curve fitting were performed with the Origin 7.0 software (OriginLab). Errors reported as fitting error.

Crystallography

Crystals of ENAH fused at the C-terminus to ABI1 were grown in hanging drops over a reservoir containing 0.1 M sodium acetate pH 4.5 and 2.90 M NaCl. 1 µL of ENAH-ABI1 (250 µM in 20 mM HEPES, 150 mM NaCl, 1 mM DTT) was mixed with 1 µL of reservoir solution, and 3D crystals appeared in two weeks. Crystals of ENAH fused at the C-terminus to PCARE were grown in hanging drops containing 0.1M Tris pH 8.0 and 3.30 M NaCl. 1.5 µL of ENAH-PCARE (769 µM in 20 mM HEPES, 150 mM NaCl, 1 mM DTT) were mixed with 0.5 µL of reservoir solution, and football shaped crystals appeared in two days. Both crystals were grown at 18 °C. Diffraction data were collected at the Advanced Photon Source at Argonne National Laboratory, NE-CAT beamline 24-IDE. The ENAH-ABI1 and

ENAH-PCARE data were integrated and scaled to 1.88 Å and 1.65 Å with AIMLESS and XDS respectively, and the structures were solved with molecular replacement using ENAH EVH1 structures 5NC7 (for ENAH-ABI1) or 6RD2 (for ENAH-PCARE) as search models. Structures were refined with iterative rounds of model rebuilding with PHENIX and COOT. Tables S6 reports refinement statistics. Structures are deposited in the PDB with identifiers 7LXE and 7LXF. Note that the ABI1 FP_8 peptide is numbered 120-129, and the $PCARE^{828-848}$ peptide is numbered as 133-153 in the PDB file in accordance with the ENAH-ABI1 and ENAH-PCARE fusion protein numbering.

Cell culture

MCF7 and HEK293T cells were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM) base media with sodium pyruvate (Corning) supplemented with 10% fetal bovine serum (FBS; Millipore), 2 mM L-glutamine (Corning), and 100 U/mL penicillin with 100 µg/mL streptomycin (Corning). MV^{D7} cells were cultured in high-glucose DMEM supplemented with 15% FBS, 2 mM L-glutamine, 100 U/mL mouse interferon-γ (Millipore). MCF7 and HEK293T cells were maintained in a 37 °C humidified incubator under 5% CO_2. MV^{D7} cells were maintained in a 32 °C humidified incubator under 5% CO_2. For lentiviral production, second generation lentiviral particles were generated by PEI transfection of 293T cells as previously described (63) with transfer plasmid, pMD2.G, and psPAX2 (Addgene #12259, #12260, gifts from Didier Trono). HEK293T media containing lentiviral particles was collected, filtered, and added directly to cultures with polybrene (Gibco). Puromycin (2 µg/mL final concentration; Thermo Fisher) and blasticidin (4 µg/mL final; Gibco) were used to select for cells stably expressing shRNA sequences or Ena/VASP constructs, respectively, after lentiviral transduction. For immunofluorescence experiments, MCF7 cells were cultured on fibronectin-coated coverslips (10 µg/mL; Corning). For live-cell imaging (MV^{D7}), cells were cultured in fibronectin-coated glass-bottom dishes (Mattek).

Acknowledgements

This project was supported by a NIGMS award R01 GM129007 to A.E. K. and by an NCI award R01 CA196885-01 to G.M. Part of this work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. T.H. was partially supported by NIGMS T32 GM007287 and a fellowship from the Koch Institute for Integrative Cancer Research.
We thank the Koch Institute Flow Cytometry Core facility for assistance with fluorescence-activated cell sorting (FACS), the MIT Structural Biology Core for

assistance with X-ray crystallography, the MIT Biophysical Instrumentation Facility for instrumentation resources, and the MIT BioMicroCenter for assistance with high-throughput sequencing. We thank D. Whitney for on the MassTitr technology, and LRF Backman for help with X-ray data collection. We thank members for the Keating lab and F. Gertler for thoughtful input. We also thank J. Tadros and F. Gertler for reagents. We also thank M. Li, S. Elledge, and Brigham and Women's hospital for the T7-pep library.

Author Contributions: T.H., S.S.P., G.M., and A.E.K. designed the study. T.H., S.S.P., S.M.H., and M.W.I. performed the experiments. R.A.G. helped with X-ray crystallography experiments. T.H., R.A.G., S.S.P, S.M.H., and M.W.I. analyzed data. V.S. built the pipeline to process next generation sequencing data. T.H. wrote the original draft with assistance from S.S.P. A.E.K. and G.M. reviewed and edited the paper. A.E.K. and G.M. provided supervision and funding acquisition. G.M. led and supervised all cell-based experiments, while A.E.K. oversaw all other experimental and computational work reported here.

References

1. Lim, W. A., Richards, F. M., & Fox, R. O. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **372,** 375–379 (1994).
2. Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., & Oschkinat, H. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**, 646–649 (1996).
3. Schultz, J., Hoffmüller, U., Krause, G., Ashurst, J., Macias, M. J., Schmieder, P., Schneider-Mergener, J., & Oschkinat, H. Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat. Struct. Mol. Biol.* **5**, 19–24 (1998).
4. Bhattacharyya, R. P., Reményi, A., Yeh, B. J., & Lim, W. A. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680 (2006).
5. Palopoli, N., González Foutel, N. S., Gibson, T. J., & Chemes, L. B. Short linear motif core and flanking regions modulate retinoblastoma protein binding affinity and specificity. *Protein Eng. Des. Sel.* **31**, 69–77 (2018).
6. Prestel, A., Wichmann, N., Martins, J. M., Marabini, R., Kassem, N., Broendum, S. S., Otterlei, M., Nielsen, O., Willemoës, M., Ploug, M., Boomsma, W., & Kragelund, B. B. The PCNA interaction motifs revisited: thinking outside the PIP-box. *Cell. Mol. Life Sci.* **76**, 4923–4943 (2019).
7. 7. Li, Z., Liu, H., Li, J., Yang, Q., Feng, Z., Li, Y., Yang, H., Yu, C., Wan, J., Liu, W., & Zhang, M. Homer Tetramer Promotes Actin Bundling Activity of Drebrin. *Structure* **27**, 27–38 (2019).
8. Ivarsson, Y., Arnold, R., McLaughlin, M., Nim, S., Joshi, R., Ray, D., Liu, B., Teyra, J., Pawson, T., Moffat, J., Li, S. S., Sidhu, S. S., & Kim, P. M. Large-scale interaction profiling of PDZ domains through proteomic peptide-phage

display using human and viral phage peptidomes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2542–2547 (2014).
9. Teyra, J., Huang, H., Jain, S., Guan, X., Dong, A., Liu, Y., Tempel, W., Min, J., Tong, Y., Kim, P. M., Bader, G. D., & Sidhu, S. S. Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure* **25**, 1598–1610 (2017).
10. Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J. H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., Seshagiri, S., Lasky, L. A., Sander, C., Boone, C., Bader, G. D., & Sidhu, S. S. A specificity map for the PDZ domain family. *PLOS Biol.* **6**, e239 (2008).
11. McConnell, R. E., Edward van Veen, J., Vidaki, M., Kwiatkowski, A. V., Meyer, A. S., & Gertler, F. B. A requirement for filopodia extension toward Slit during Robo-mediated axon repulsion. *J. Cell Biol.* **213**, 261–274 (2016).
12. Scott, J. A., Shewan, A. M., den Elzen, N. R., Loureiro, J. J., Gertler, F. B., & Yap, A. S. Ena/VASP proteins can regulate distinct modes of actin organization at cadherin-adhesive contacts. *Mol. Biol. Cell* **17**, 1085–1095 (2006).
13. Ball, L. J., Kühne, R., Hoffmann, B., Häfner, A., Schmieder, P., Volkmer-Engert, R., Hof, M., Wahl, M., Schneider-Mergener, J., Walter, U., Oschkinat, H., & Jarchau, T. Dual epitope recognition by the VASP EVH1 domain modulates polyproline ligand specificity and binding affinity. *EMBO J*. **19,** 4903–4914 (2000).
14. Prehoda, K. E., Lee, D. J., & Lim, W. A. Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. *Cell* **97**, 471–480 (1999).
15. Bugge, K., Brakti, I., Fernandes, C. B., Dreier, J. E., Lundsgaard, J. E., Olsen, J. G., Skriver, K., & Kragelund, B. B. Interactions by Disorder - A Matter of Context. *Front. Mol. Biosci.* **7**, 110 (2020).
16. Peterson, F. C., & Volkman, B. F. Diversity of polyproline recognition by EVH1 domains. *Front. Biosci.***14,** 833–846 (2009).
17. Roussos, E. T., Wang, Y., Wyckoff, J. B., Sellers, R. S., Wang, W., Li, J., Pollard, J. W., Gertler, F. B., & Condeelis, J. S. Mena deficiency delays tumor progression and decreases metastasis in polyoma middle-T transgenic mouse mammary tumors. *Breast Cancer Res.***12**, R101 (2010).
18. Gkretsi, V., Stylianou, A., & Stylianopoulos, T. Vasodilator-Stimulated Phosphoprotein (VASP) depletion from breast cancer MDA-MB-231 cells inhibits tumor spheroid invasion through downregulation of Migfilin, β-catenin and urokinase-plasminogen activator (uPA). *Exp. Cell Res.* **352**, 281–292 (2017).
19. Mouneimne G, Hansen SD, Selfors LM, Petrak L, Hickey MM, Gallegos LL, Simpson KJ, Lim J, Gertler FB, Hartwig JH, Mullins RD, Brugge JS. Differential remodeling of actin cytoskeleton architecture by profilin isoforms leads to distinct effects on cell migration and invasion. *Cancer Cell* **22**, 615-30 (2012).

20. Opitz, R., Müller, M., Reuter, C., Barone, M., Soicke, A., Roske, Y., Piotukh, K., Huy, P., Beerbaum, M., Wiesner, B., Beyermann, M., Schmieder, P., Freund, C., Volkmer, R., Oschkinat, H., Schmalz, H. G., & Kühne, R. A modular toolkit to inhibit proline-rich motif-mediated protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5011–5016 (2015).
21. Barone, M., Müller, M., Chiha, S., Ren, J., Albat, D., Soicke, A., Dohmen, S., Klein, M., Bruns, J., van Dinther, M., Opitz, R., Lindemann, P., Beerbaum, M., Motzny, K., Roske, Y., Schmieder, P., Volkmer, R., Nazaré, M., Heinemann, U., Oschkinat, H., … Kühne, R. Designed nanomolar small-molecule inhibitors of Ena/VASP EVH1 interaction impair invasion and extravasation of breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A*. **117**, 29684–29690 (2020).
22. Boëda, B., Briggs, D. C., Higgins, T., Garvalov, B. K., Fadden, A. J., McDonald, N. Q., & Way, M. Tes, a specific Mena interacting partner, breaks the rules for EVH1 binding. *Mol. Cell* **28**, 1071–1082 (2007).
23. Adams, R. M., Mora, T., Walczak, A. M., & Kinney, J. B. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife* **5**, e23156 (2016).
24. Larman, H. B., Zhao, Z., Laserson, U., Li, M. Z., Ciccia, A., Gakidis, M. A., Church, G. M., Kesari, S., Leproust, E. M., Solimini, N. L., & Elledge, S. J. Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* **29**, 535–541 (2011).
25. Chen, X. J., Squarr, A. J., Stephan, R., Chen, B., Higgins, T. E., Barry, D. J., Martin, M. C., Rosen, M. K., Bogdan, S., & Way, M. Ena/VASP proteins cooperate with the WAVE complex to regulate the actin cytoskeleton. *Dev. Cell* **30**, 569–584 (2014).
26. Menon, S., Boyer, N. P., Winkle, C. C., McClain, L. M., Hanlin, C. C., Pandey, D., Rothenfußer, S., Taylor, A. M., & Gupton, S. L. The E3 Ubiquitin Ligase TRIM9 Is a Filopodia Off Switch Required for Netrin-Dependent Axon Guidance. *Dev. Cell* **35**, 698–712 (2015).
27. Mészáros, B., Erdos, G., & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46,** W329–W337 (2018).
28. Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
29. Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., … Lundberg, E. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
30. Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **14**, 703–721 (2019).

31. Kannan, R., Kuzina, I., Wincovitch, S., Nowotarski, S. H., & Giniger, E. The Abl/enabled signaling pathway regulates Golgi architecture in Drosophila photoreceptor neurons. *Mol. Biol. Cell* **25**, 2993–3005 (2014).
32. Tang, D., Zhang, X., Huang, S., Yuan, H., Li, J., & Wang, Y. Mena-GRASP65 interaction couples actin polymerization to Golgi ribbon linking. *Mol. Biol. Cell* **27**, 137–152 (2016).
33. Tani, K., Sato, S., Sukezane, T., Kojima, H., Hirose, H., Hanafusa, H., & Shishido, T. Abl interactor 1 promotes tyrosine 296 phosphorylation of mammalian enabled (Mena) by c-Abl kinase. *J. Biol. Chem.* **278**, 21685–21692 (2003).
34. Niebuhr, K., Ebel, F., Frank, R., Reinhard, M., Domann, E., Carl, U. D., Walter, U., Gertler, F. B., Wehland, J., & Chakraborty, T. A novel proline-rich motif present in ActA of Listeria monocytogenes and cytoskeletal proteins is the ligand for the EVH1 domain, a protein module present in the Ena/VASP family. *EMBO J.* **16**, 5433–5444 (1997).
35. Fedorov, A. A., Fedorov, E., Gertler, F., & Almo, S. C. Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. *Nat. Struct. Mol. Biol.* **6**, 661–665 (1999).
36. Beneken J, Tu JC, Xiao B, Nuriya M, Yuan JP, Worley PF, Leahy DJ. Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition. *Neuron* **26**, 143-54 (2000).
37. Acevedo, L. A., Greenwood, A. I., & Nicholson, L. K. A Noncanonical Binding Site in the EVH1 Domain of Vasodilator-Stimulated Phosphoprotein Regulates Its Interactions with the Proline Rich Region of Zyxin. *Biochemistry* **56**, 4626–4636 (2017).
38. Barone, M. Structure-optimized ProM scaffolds address Ena/VASP as a possible antimetastatic target. [Doctoral dissertation, Freie University Berlin] (2017).
39. Zhou, J., Panaitiu, A. E., & Grigoryan, G. A general-purpose protein design framework based on mining sequence-structure relationships in known protein structures. *Proc. Natl. Acad. Sci. U.S.A*. **117**, 1059–1068 (2020).
40. Bear JE, Loureiro JJ, Libova I, Fässler R, Wehland J, Gertler FB. Negative regulation of fibroblast motility by Ena/VASP proteins. *Cell* **101**, 717-28 (2000).
41. Damiano-Guercio, J., Kurzawa, L., Mueller, J., Dimchev, G., Schaks, M., Nemethova, M., Pokrant, T., Brühmann, S., Linkner, J., Blanchoin, L., Sixt, M., Rottner, K., & Faix, J. Loss of Ena/VASP interferes with lamellipodium architecture, motility and integrin-dependent adhesion. *eLife* **9**, e55351 (2020).
42. Puleo, J. I., Parker, S. S., Roman, M. R., Watson, A. W., Eliato, K. R., Peng, L., Saboda, K., Roe, D. J., Ros, R., Gertler, F. B., & Mouneimne, G. Mechanosensing during directed cell migration requires dynamic actin polymerization at focal adhesions. *J. Cell Biol.* **218**, 4215–4235 (2019).
43. Davey, N. E., Seo, M. H., Yadav, V. K., Jeon, J., Nim, S., Krystkowiak, I., Blikstad, C., Dong, D., Markova, N., Kim, P. M., & Ivarsson, Y. Discovery of

short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *FEBS J*. **284**, 485–498 (2017).
44. Ueki, Y., Kruse, T., Weisser, M. B., Sundell, G. N., Larsen, M., Mendez, B. L., Jenkins, N. P., Garvanska, D. H., Cressey, L., Zhang, G., Davey, N., Montoya, G., Ivarsson, Y., Kettenbach, A. N., & Nilsson, J. A Consensus Binding Motif for the PP4 Protein Phosphatase. *Mol. Cell* **76**, 953–964 (2019).
45. Jespersen, N., Estelle, A., Waugh, N., Davey, N. E., Blikstad, C., Ammon, Y. C., Akhmanova, A., Ivarsson, Y., Hendrix, D. A., & Barbar, E. Systematic identification of recognition motifs for the hub protein LC8. *Life Sci. Alliance* **2**, e201900366 (2019).
46. Wigington, C. P., Roy, J., Damle, N. P., Yadav, V. K., Blikstad, C., Resch, E., Wong, C. J., Mackay, D. R., Wang, J. T., Krystkowiak, I., Bradburn, D. A., Tsekitsidou, E., Hong, S. H., Kaderali, M. A., Xu, S. L., Stearns, T., Gingras, A. C., Ullman, K. S., Ivarsson, Y., Davey, N. E., … Cyert, M. S. Systematic Discovery of Short Linear Motifs Decodes Calcineurin Phosphatase Signaling. *Mol. Cell* **79**, 342–358 (2020).
47. Zhang, Y., Tu, Y., Gkretsi, V., & Wu, C. Migfilin interacts with vasodilator-stimulated phosphoprotein (VASP) and regulates VASP localization to cell-matrix adhesions and migration. *J. Biol. Chem.* **281**, 12397–12407 (2006).
48. Drees B.E., Friederich E., Fradelizi J., Louvard D., Beckerle M.C., Golsteyn R.M. Characterization of the interaction between zyxin, and members of the Ena/vasodilator-stimulated phosphoprotein family of proteins. *J. Biol. Chem.* **275**, 22503-22511 (2000).
49. Petit, M. M., Fradelizi, J., Golsteyn, R. M., Ayoubi, T. A., Menichi, B., Louvard, D., Van de Ven, W. J., & Friederich, E. LPP, an actin cytoskeleton protein related to zyxin, harbors a nuclear export signal and transcriptional activation capacity. *Mol. Biol. Cell* **11**, 117–129 (2000).
50. Hansen, S. D., & Mullins, R. D. Lamellipodin promotes actin assembly by clustering Ena/VASP proteins and tethering them to actin filaments. *eLife* **4,** e06585 (2015).
51. Levchenko A. Allovalency: a case of molecular entanglement. *Curr. Biol.* 13, R876–R878 (2003).
52. Zarrinpar, A., Bhattacharyya, R. P., & Lim, W. A. The structure and function of proline recognition domains. *Sci. Signal* 2003, RE8 (2003).
53. Volkman, B. F., Prehoda, K. E., Scott, J. A., Peterson, F. C., & Lim, W. A. Structure of the N-WASP EVH1 domain-WIP complex: insight into the molecular basis of Wiskott-Aldrich Syndrome. *Cell* **111**, 565–576 (2002).
54. Golemi-Kotra, D., Mahaffy, R., Footer, M. J., Holtzman, J. H., Pollard, T. D., Theriot, J. A., & Schepartz, A. High affinity, paralog-specific recognition of the Mena EVH1 domain by a miniature protein. *J. Am. Chem. Soc.* **126**, 4–5 (2004).
55. Wilson, C., Agafonov, R. V., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D. L., & Kern, D. Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347**, 882–886 (2015).

56. Xie, Z. H., Yu, J., Shang, L., Zhu, Y. Q., Hao, J. J., Cai, Y., Xu, X., Zhang, Y., & Wang, M. R. KIAA1522 overexpression promotes tumorigenicity and metastasis of esophageal cancer cells through potentiating the ERK activity. OncoTargets Ther. **10**, 3743–3754 (2017).
57. Li, Y., Wang, Y., Fan, H., Zhang, Z., & Li, N. miR-125b-5p inhibits breast cancer cell proliferation, migration and invasion by targeting KIAA1522. *Biochem. Biophys. Res. Commun.* **504**, 277–282 (2018).
58. Corral-Serrano, J. C., Lamers, I., van Reeuwijk, J., Duijkers, L., Hoogendoorn, A., Yildirim, A., Argyrou, N., Ruigrok, R., Letteboer, S., Butcher, R., van Essen, M. D., Sakami, S., van Beersum, S., Palczewski, K., Cheetham, M. E., Liu, Q., Boldt, K., Wolfrum, U., Ueffing, M., Garanto, A., … Collin, R. PCARE and WASF3 regulate ciliary F-actin assembly that is required for the initiation of photoreceptor outer segment disk formation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9922–9931 (2020).
59. Bachmann, C., Fischer, L., Walter, U., & Reinhard, M. The EVH2 domain of the vasodilator-stimulated phosphoprotein mediates tetramerization, F-actin binding, and actin bundle formation. *J. Biol. Chem.* **274**, 23549–23557 (1999).
60. Vidaki, M., Drees, F., Saxena, T., Lanslots, E., Taliaferro, M. J., Tatarakis, A., Burge, C. B., Wang, E. T., & Gertler, F. B. A Requirement for Mena, an Actin Regulator, in Local mRNA Translation in Developing Neurons. *Neuron* **95**, 608–622 (2017).
61. Padilla-Rodriguez, M., Parker, S. S., Adams, D. G., Westerling, T., Puleo, J. I., Watson, A. W., Hill, S. M., Noon, M., Gaudin, R., Aaron, J., Tong, D., Roe, D. J., Knudsen, B., & Mouneimne, G. The actin cytoskeletal architecture of estrogen receptor positive breast cancer cells suppresses invasion. *Nature Commun.* **9,** 2980 (2018).
62. Foight, G. W., & Keating, A. E. Comparison of the peptide binding preferences of three closely related TRAF paralogs: TRAF2, TRAF3, and TRAF5. *Protein Sci.* **25**, 1273–1289 (2016).
63. Yang S., Zhou X., Li R., Fu X., Sun P. Optimized PEI-based transfection method for transient transfection and lentiviral production. *Curr. Protoc. Chem. Biol.* **9**,147–157 (2017) .
64. Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188–1190 (2004).

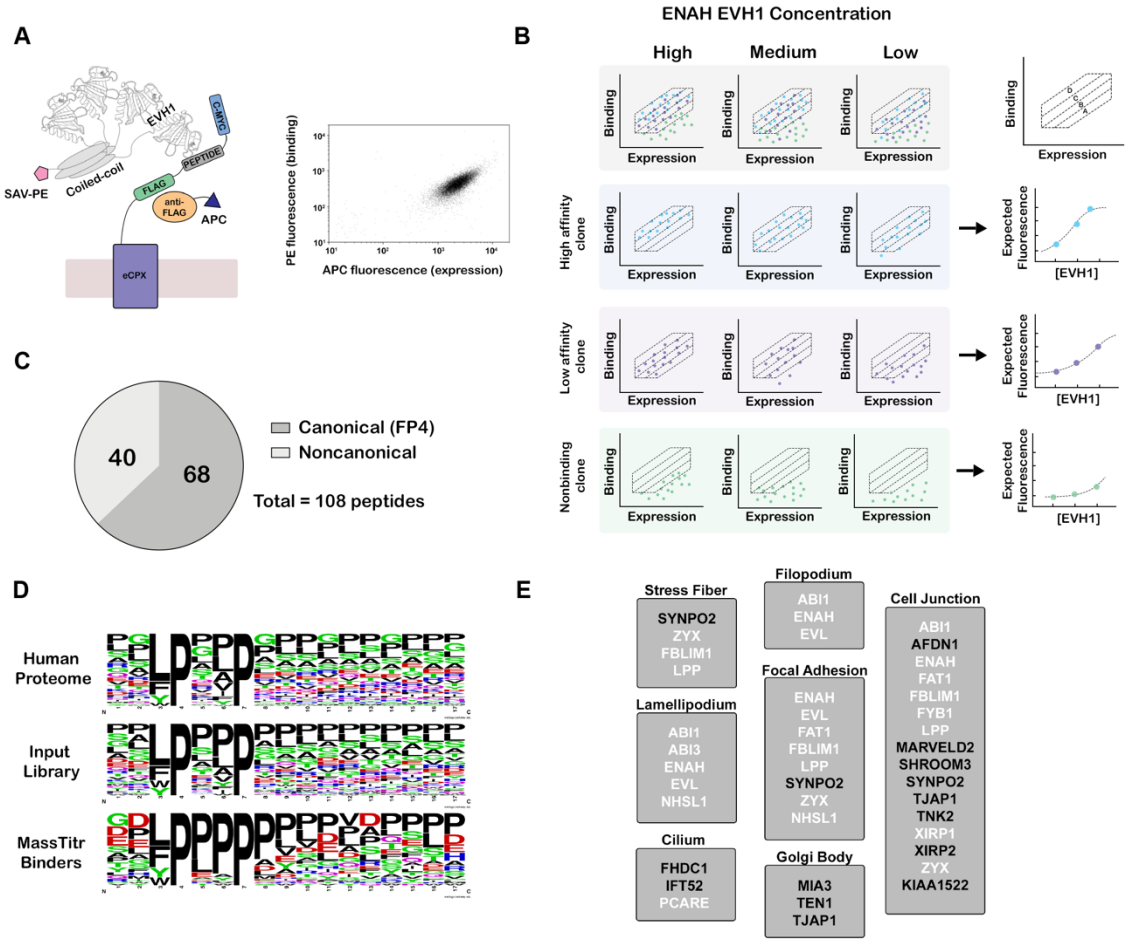

Figure 1. MassTitr screening identifies biologically relevant ENAH EVH1 ligands. (A) At left, bacterial surface display schematic. Library peptides flanked by a FLAG tag and a c-myc tag were expressed as fusions to the C-terminus of eCPX on the surface of *E. coli.* Cells were labelled with anti-FLAG-APC to quantify expression and then incubated with tetrameric ENAH EVH1 domain, which was detected by streptavidin conjugated to phycoerythrin (SAV-PE). At right, a FACS plot for surface displayed ActA peptide binding to 10 μM ENAH EVH1 tetramer. (B) MassTitr schematic. The top row represents a library of three clones, blue, purple, and green, sorted into four gates at three concentrations of ENAH. The rows highlighted in blue, purple, and green illustrate reconstruction of the concentration-dependent binding of each clone based on deep sequencing. This experiment sorted a pre-enriched library of clones into four gates at eight concentrations. (C) Distribution of MassTitr hits after filtering; 68 peptides contained a canonical FP4 motif matching the regular expression [FWYL]PX[FWYLIAVP]P. (D) Frequency plot made from the sequences that match the FP4 motif in the human proteome, the input library, and the MassTitr binders using Weblogo (64). (E) Annotated subcellular localizations where at least two MassTitr hits predicted to be disordered and localized in the cytoplasm are present. White text denotes previously known Ena/VASP interactions.

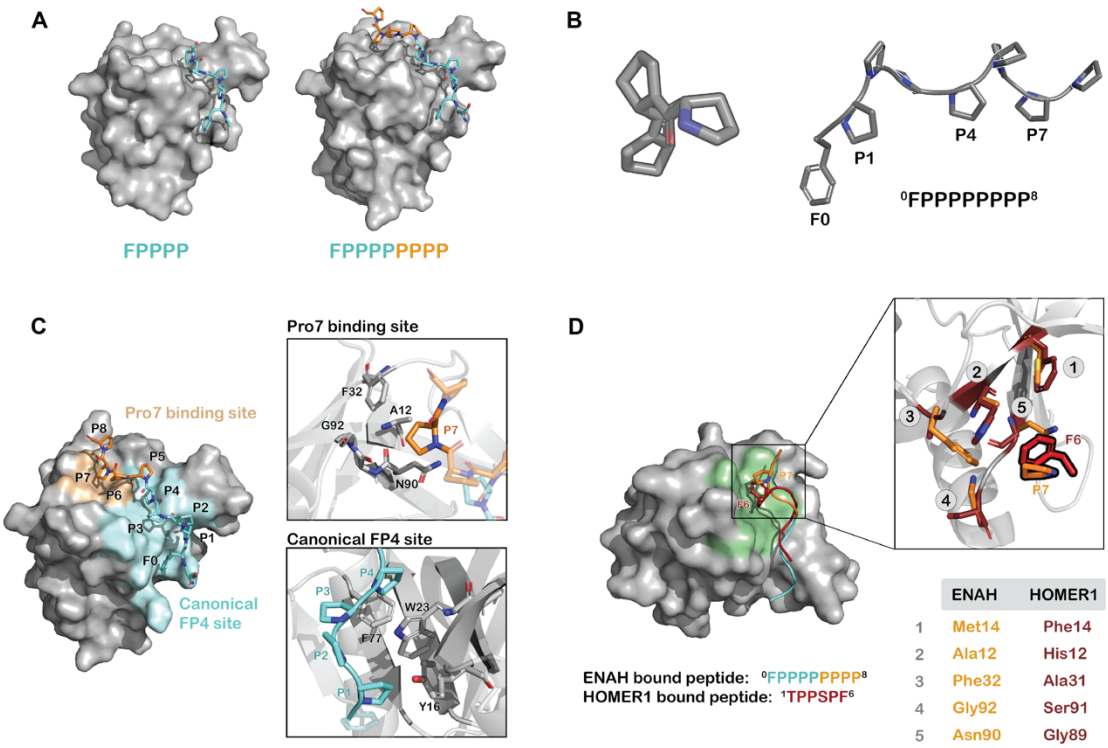

Figure 2. Prolines C-terminal to FP4 can engage a novel ENAH binding site. (A) Surface representation of the ENAH EVH1 domain bound to FP_4 (PDB 1EVH) or FP_8. The core FP4 motif is light blue, the P_4 flank is orange. (B) Axial view of a polyproline type II helix highlighting three-fold symmetry (left); a side view shows P1, P4, and P7 facing the same side (right). (C) Surface representation of ENAH EVH1 domain bound to FP_8 with the canonical binding site in light blue and the extended site (Pro7 site) in orange; insets show details of the interactions. (D) On left, surface representation of the HOMER1 EVH1 domain (PDB 1DDV) aligned to the ENAH EVH1 domain, bound to peptide FP_8 or TPPSPF. Region corresponding to the Pro7 binding pocket in HOMER1 colored in green. Inset: magnified region of Pro7 binding pocket, with ENAH EVH1 residues in orange and HOMER1 EVH1 residues in red. Residues of peptides bound to ENAH or HOMER1 are outlined in black.

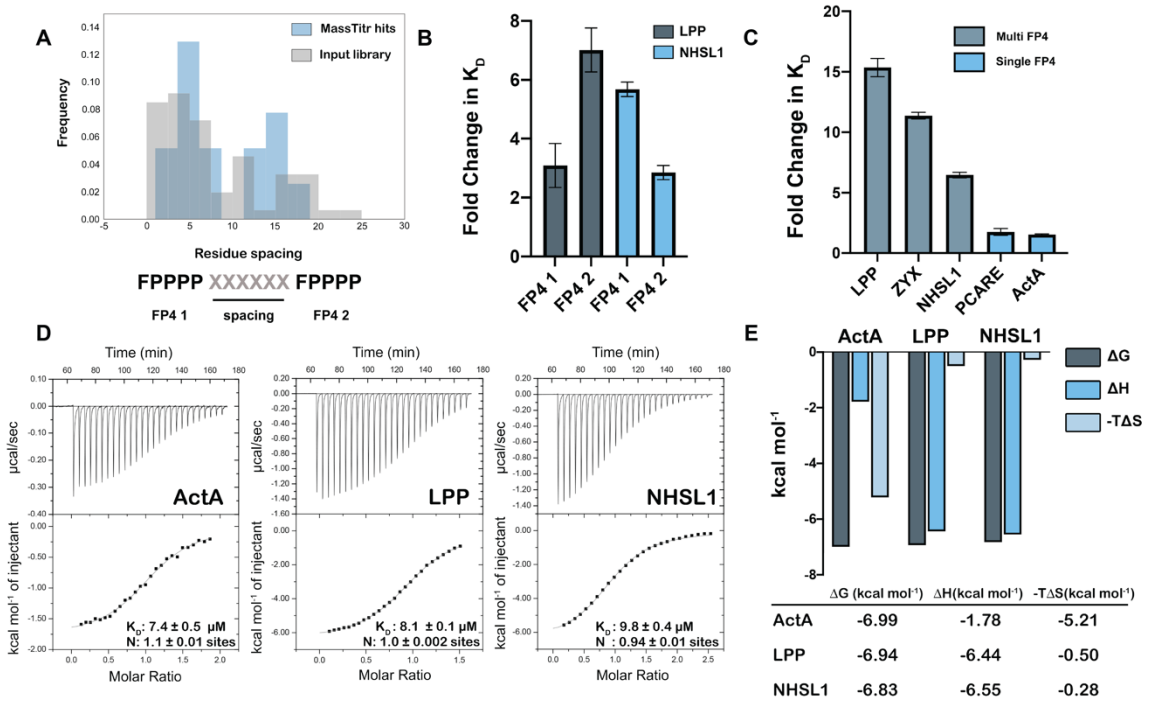

Figure 3. Multiple FP4 motifs enhance peptide binding affinity. (A) Spacing of FP4 motifs in the input library and in high-confidence hits. (B) Fold change in K_D for peptide variants relative to tighter-binding 36-mer library peptides for LPP, NHSL1, or zyxin; see Table 1 for sequences. (C) Fold change in K_D for 36-mer peptides binding to ENAH EVH1 R47A relative to tighter-binding ENAH EVH1 WT. (D) ITC binding curves for 36-residue peptides from ActA, LPP, or NHSL1; (E) The entropic and enthalpic contributions to binding determined using data in panel D. Fold-change errors in (B) and (C) were calculated by propagating the error from two affinity measurements. Sequences for peptides referenced in this figure are given in Tables 1, S3, and S4.

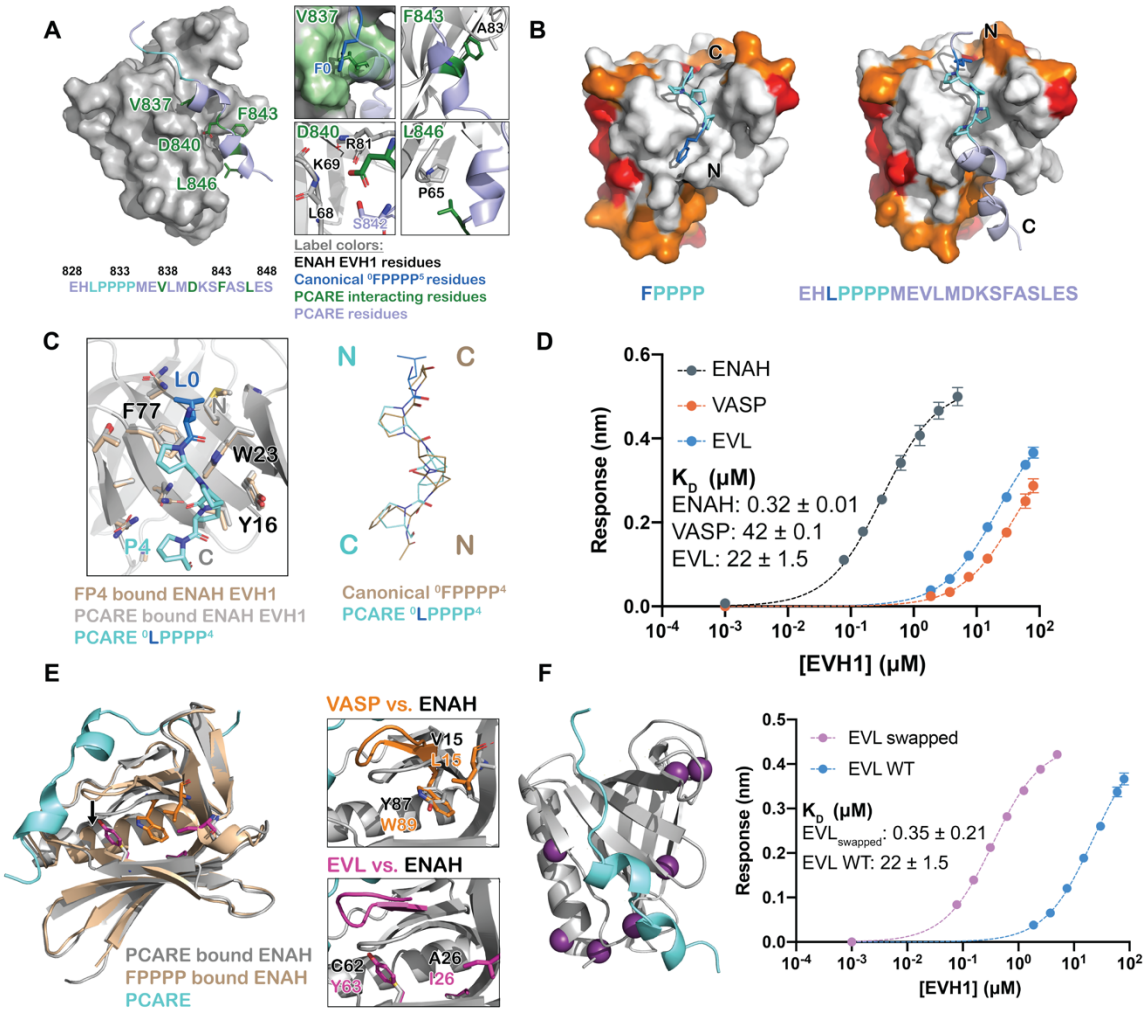

Figure 4. A peptide from PCARE binds with high affinity and specificity to the ENAH EVH1 domain using a noncanonical binding mode. (A) Surface representation of ENAH EVH1 domain bound to the PCARE peptide. The LP_4 residues are light blue and other EVH1-interacting residues are green; insets show details of the interactions. Note that the PCARE$^{828\text{-}848}$ peptide is numbered as 133-153 in the PDB file. (B) Surface representation of ENAH EVH1 highlighting conservation among Ena/VASP paralogs. Residues shared between all three paralogs - white; residues shared by two paralogs - orange, residues that differ in each paralog - red. On the left is ENAH EVH1 bound to FP_4 (PDB 1EVH); on the right is ENAH EVH1 bound to the PCARE peptide. (C) View comparing the orientations of a FP_4 peptide and the LP_4 region of PCARE. Side chains of the ENAH EVH1 domain are shown as sticks using tan for the FP_4 complex and grey for the PCARE complex. (D) BLI binding curves for PCARE B binding to ENAH, EVL, or VASP EVH1 domains. (E) Superposition of ENAH EVH1 domains bound to PCARE or FP_4 peptide (1EVH). The black arrow highlights a 3 Å shift in a loop that forms part of the binding pocket. Insets show residues that differ between ENAH and VASP or EVL near this loop. (F) ENAH EVH1 domain bound to PCARE peptide, with residues that were swapped into the EVL EVH1 domain to rescue affinity marked as purple spheres. On the right are binding curves for WT EVL EVH1 domain and $EVL_{swapped}$ EVH1 domain binding to PCARE B.

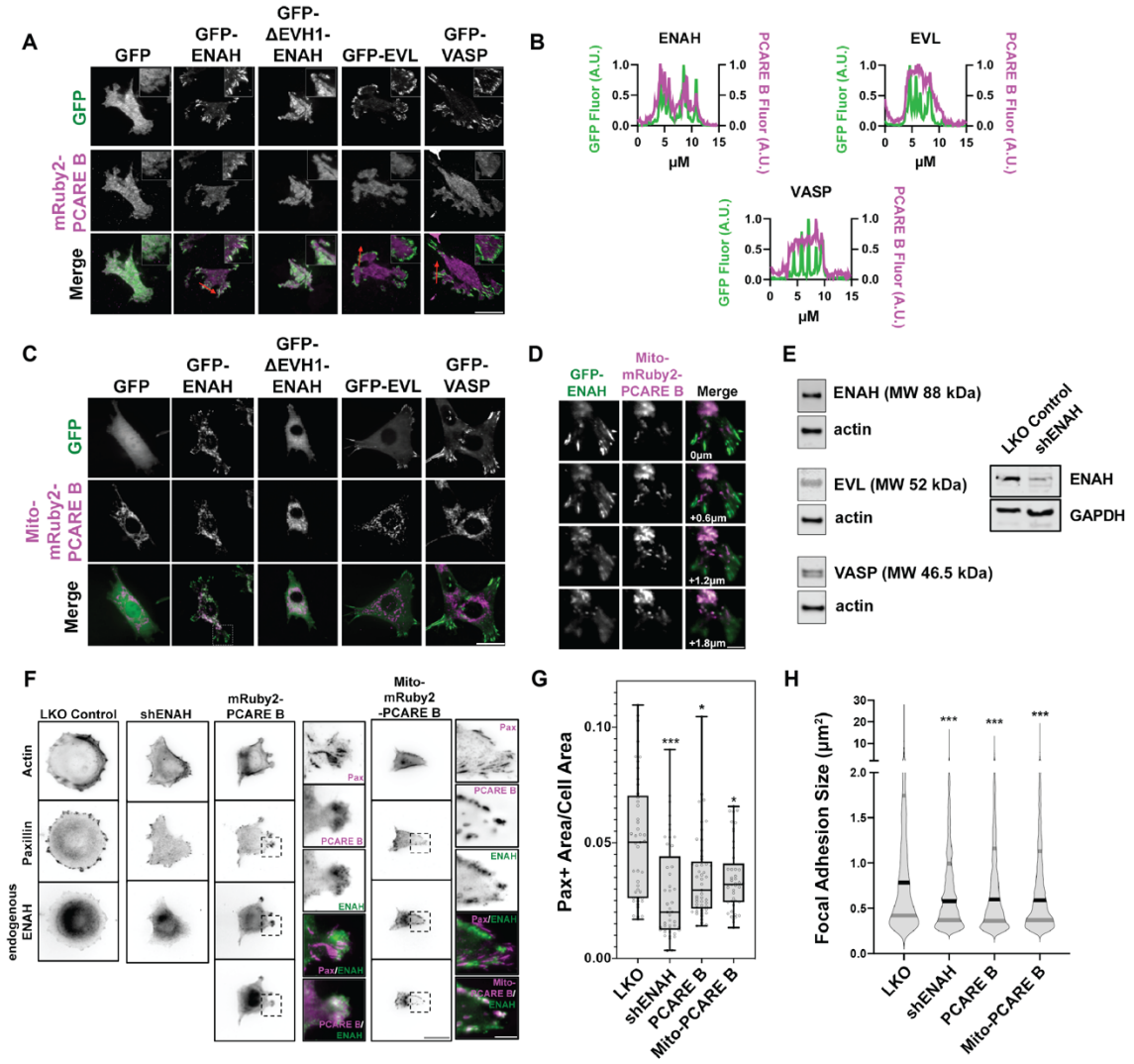

Figure 5. A PCARE-derived peptide selectively recruits ENAH in cells. (A) Live MV^{D7}_{shEVL} cells expressing mRuby2-PCARE B with GFP, GFP-ENAH, GFP-dEVH1-ENAH, GFP-EVL, or GFP-VASP, imaged using TIRF microscopy. Scale bar = 25 μm. (B) Normalized fluorescence intensity of GFP signal (left axis) and mRuby2-PCARE B signal (right axis) along a line drawn through focal adhesions, indicated by red arrow in (A). (C) Live MV^{D7}_{shEVL} cells expressing Mito-mRuby2-PCARE B with GFP, GFP-ENAH, GFP-dEVH1-ENAH, GFP-EVL, or GFP-VASP. Image is a maximum intensity projection of z-stack acquired using widefield fluorescence microscopy. Scale bar = 25 μm. (D) Magnified region of interest indicated by box in (C), showing co-localization of Mito-mRuby2-PCARE B and GFP-ENAH at increasing depths in the cell. Scale bar = 5 μm. (E) At left, representative western blot showing expression of ENAH, EVL, and VASP in MCF7 cells. At right, representative western blot showing magnitude of knockdown using *ENAH*-targeted shRNA. (F) Immunofluorescence labeling of MCF7 cells expressing non-targeting LKO control, *ENAH*-targeting shRNA, mRuby2-PCARE B, or Mito-mRuby2-PCARE B. Cells were fixed eight hours after plating, and immunolabeled for focal adhesion marker paxillin and endogenous ENAH, and additionally stained with phalloidin for F-actin. Box indicates positions of magnified regions of interest (ROI). Scale bar = 25

μm, magnified ROI scale bar = 5 μm. (G) Box-and-whisker plot of total paxillin-positive area per cell normalized to total cell area, for indicated conditions. N= 3 biological replicates, n= 45-49 cells. (H) Violin plot of individual focal adhesion size (for adhesions greater than 0.25 μm^2) for indicated conditions. Central black line indicates median, peripheral gray lines indicate interquartile range. N = 3 biological replicates, n = 1452-2159 individual adhesions.

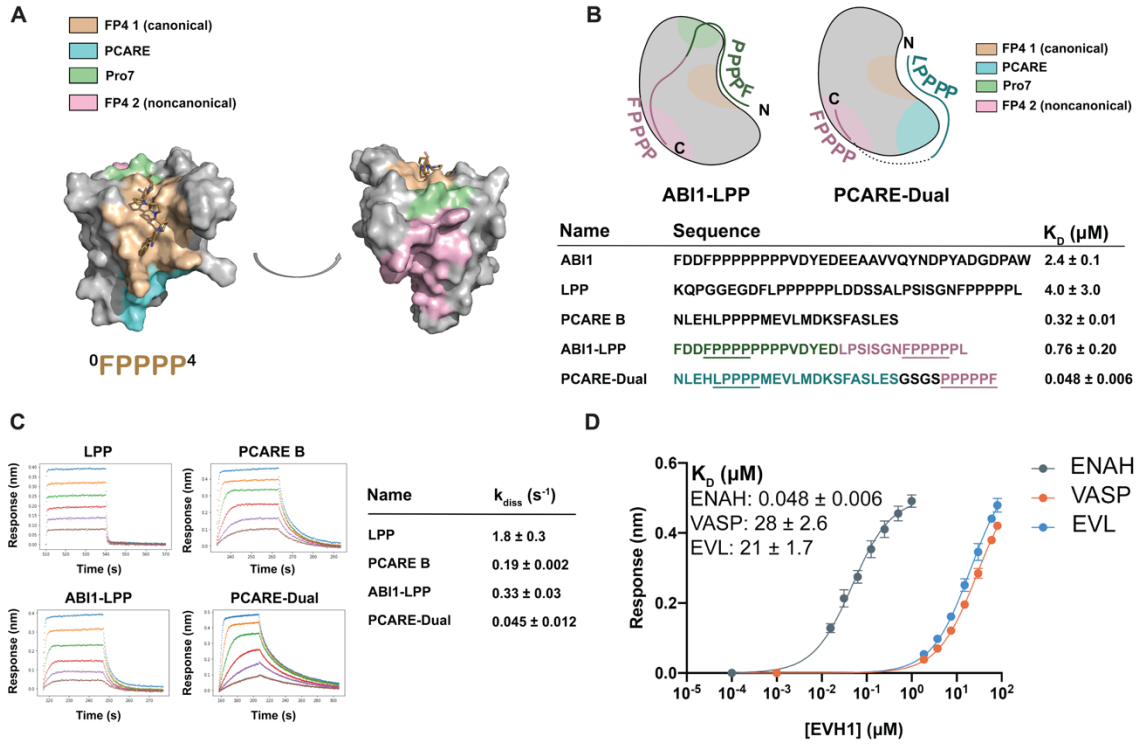

Figure 6. Engineered peptides bind to the ENAH EVH1 domain with high affinity and specificity (A) Surface representation of ENAH EVH1 with binding sites characterized in this study indicated. (B) Design scheme for high affinity binders ABI1-LPP and PCARE-Dual. (C) BLI binding and dissociation curves. Blue, orange, green, red, purple, and brown curves denote EVH1 concentrations in descending order. For LPP: 80, 36, 16, 7.0, 3.1, 1.4 μM. PCARE B and ABI1-LPP: 2.5, 1.3, 0.63, 0.31, 0.16, 0.078 μM. PCARE-Dual: 0.50, 0.25, 0.0625, 0.031, 0.016, 0.0078 μM. (D) BLI binding curves for PCARE-Dual binding to the EVH1 domains of ENAH, VASP, or EVL.

Name	Sequence	WT K_D (μM)	R47A K_D (μM)
NHSL1	ADRSPF**LPPPP**PVTDCSQGSPLPHSPV**FPPPP**PEAL	8.6 ± 2.0	55.4 ± 1.0
NHSL1 FP4 1	ADRSPF**LPPPP**PVTDCSQGSPLPHSPV	48.7 ± 4.0	99.1 ± 17.0
NHSL1 FP4 2	PVTDCSQGSPLPHSPV**FPPPP**PEAL	24.4 ± 1.2	51.0 ± 3.1
NHSL1 Duplicated	ADRSPF**LPPPP**PVTDCSQGSPLPHSPVPVTDCSQGSPLPHSPV**FPPPP**PEAL	18.4 ± 0.1	70.0 ± 1.0
LPP	KQPGGEGDF**LPPPP**PPLDDSSALPSISGN**FPPPP**PL	4.0 ± 3.0	61.9 ± 6.1
LPP FP4 1	KQPGGEGDF**LPPPP**PPLDDSSALPSISGN	12.4 ± 0.3	61.3 ± 0.6
LPP FP4 2	PPLDDSSALPSISGN**FPPPP**PL	28.3 ± 0.9	75.2 ± 7.6
LPP Duplicated	KQPGGEGDF**LPPPP**PPLDDSSALPSISGNDDSSALPSISGN**FPPPP**PL	6.7 ± 0.3	54.2 ± 2.7

Table 1. Affinities of dual FP4 motif peptides and their variants for ENAH EVH1 WT or ENAH EVH1 R47A.