A distributed residue network permits conformational binding specificity in a conserved family of actin remodelers

Theresa Hwang[1], Robert A Grant[1], and Amy E Keating[1,2]*

1. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA
2. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA

*correspondence to keating@mit.edu

Abstract

Metazoan proteomes contain many protein families wherein paralogs with high sequence and structural similarity have evolved unique functions and binding profiles. We uncovered a region from ciliary protein PCARE is highly specific to Ena/VASP paralog ENAH, but not VASP and EVL (Hwang et al., 2021). Here, we show that despite binding at a site that is identical between paralogs, PCARE stabilizes a conformation of the EVH1 domain of ENAH that is inaccessible to family members VASP and EVL to achieve its high affinity and ~100-fold specificity. Structure-based modeling rapidly identified seven residues distributed throughout EVL that, when mutated to residues of ENAH, are sufficient to confer high-affinity binding of PCARE. By exploiting the ENAH-specific EVH1 conformation, we rationally designed the tightest and most selective ENAH binder to date, providing a tool for dissecting paralog-specific Ena/VASP functions in processes including cancer cell invasion. Our work uncovers a mechanism of interaction specificity that distinguishes paralogs that share high sequence identity and many common binding partners.

Introduction

Metazoan signal transduction networks have evolved a high degree of complexity using adapter proteins that are specialized to make many interactions and/or highly specific interactions (Rowland et al., 2017; Zarrinpar et al., 2003). Signaling complexity arises in part from a plethora of interaction domain families such as the SH3, SH2, and PDZ domains. The facile recombination and insertion of modular domains to generate diverse protein architectures have enabled the evolution of new signaling circuits (Pawson and Nash, 2000).

Many modular interaction domains bind to short linear motifs (SLiMs), which occur as stretches of 3-10 consecutive amino acids in intrinsically disordered regions of proteins. The SLiM-binding specificity profiles of different paralogous members in a family of domains are often highly overlapping, yet individual members can in some cases engage in highly selective interactions (Xin et al., 2014; Hause et al., 2012). For example, a SLiM in Pbs2 binds only the SH3 domain of Sho1 out of the 27 SH3 domains in yeast to activate the high-osmolarity stress response pathway (Zarrinpar et al., 2003). Conversely, the actin assembly protein Las17 binds promiscuously to many

SH3 domains, including Sho1, to drive actin-related processes such as endocytosis in yeast (Kelil et al., 2016; Robertson et al., 2009).

The Ena/VASP proteins are a family of actin regulators involved in functions ranging from T-cell activation to axon guidance (Kwiatkowski et al., 2003). There are three paralogs in mammals: ENAH, VASP, and EVL. All three proteins contain an N-terminal EVH1 interaction domain responsible for subcellular localization and a C-terminal EVH2 domain that polymerizes actin. The two domains are connected by a linker, predicted to be largely disordered, that contains binding motifs for other proteins. The EVH1 domain binds SLiMs with the consensus motif [FWYL]PXΦP, where X is any amino acid and Φ is any hydrophobic residue (Ball et al., 2000). This sequence, referred to here as the FP4 motif, binds the EVH1 domain as a polyproline type II (PPII) helix, as shown in Figure 1A.

Ena/VASP proteins have evolved both overlapping and paralog-specific cellular functions. On the one hand, ENAH, VASP, and EVL can all bind to FP4 motifs in lamellipodin to promote actin assembly at the leading edge (Krause et al., 2004; Hansen and Mullins, 2015). Single deletions of Ena/VASP paralogs lead to mild phenotypic defects in mice, indicating that the paralogs can functionally compensate for each other, whereas triple mutant mice die after proceeding to late embryogenesis (Aszódi et al., 1999; Lanier et al., 1999; Kwiatowski et al., 2007). On the other hand, the three paralogs participate in distinct pathways. ENAH, alone, promotes haptotaxis of breast cancer cells through fibronectin gradients and regulates translation of specific mRNAs in developing axons, implicating it in functions beyond its role in actin polymerization (Oudin et al., 2016; Vidaki et al., 2018). In addition, whereas ENAH and VASP promote the invasive potential of migratory breast cancer cells, EVL suppresses breast cancer invasion (Roussos et al., 2011; Zhang et al., 2009; Mouneimne et al., 2018).

The FP4-binding pocket of the EVH1 domain is 100% conserved across ENAH, VASP, and EVL, and the paralogs share 62-72% sequence identity over the entire domain (Figure 1A). Consequently, the three EVH1 domains recognize many common binding partners. Nevertheless, some proteins bind selectively to certain paralogs. The LIM3 domain of testin binds specifically to the ENAH EVH1 domain in a region adjacent to the highly conserved FP4-binding pocket; this is the only example of an endogenous Ena/VASP EVH1 binding partner where the mechanistic basis for specificity is defined (Boëda et al., 2007). Testin makes use of a generally well-characterized strategy to obtain binding specificity by contacting surface residues that have diverged across paralogs to form distinct interfaces (Skerker et al., 2008; Bardwell et al., 2009; Schreiber and Keating, 2011).

We previously identified a peptide from ciliary protein PCARE that binds with 70-140-fold higher affinity to ENAH than to EVL or VASP, despite containing an FP4 motif that can engage the perfectly conserved FP4-binding site (Hwang et al., 2021). Here we describe the surprising mechanistic basis behind this selectivity. An epistatic residue network in the ENAH EVH1 domain allows ENAH to adopt a conformation unattainable by paralogs VASP and EVL. An alpha-helical extension C-terminal to the FP4 motif in PCARE engages and stabilizes this ENAH-specific conformation using a noncanonical binding mode. These observations revealed a strategy to obtain binding specificity in a highly conserved family that must also make promiscuous interactions, and we

demonstrate how this information can be leveraged to design synthetic peptides with unprecedented affinity and specificity for ENAH.

Results

FP4 motif-flanking elements in ciliary protein PCARE confer high affinity by inducing noncanonical binding

Previously, we performed a proteomic screen to identify peptides that bind to the ENAH EVH1 domain with dissociation constants (K_D) primarily in the low- to mid-micromolar range (2 μM – 60 μM). The highest affinity hit from our screen was a 36-residue peptide from ciliary protein PCARE (PCARE$^{813-848}$) that bound to ENAH with a K_D of 0.19 μM (Hwang et al., 2021). To understand the structural basis for this high-affinity interaction, we solved a crystal structure of ENAH EVH1 domain fused to the 36-mer PCARE sequence to 1.65 Å resolution. Twenty-one residues of the 36-mer PCARE peptide were fully resolved in the electron density (PCARE$^{828-848}$, Figure 1A) 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 (Ball et al., 2000; Prehoda et al., 1999; Federov et al., 1999) (Figure 1A, B). PCARE$^{828-848}$ uses a 14-residue alpha helix-rich extension C-terminal to the LPPPP motif to make additional contacts to an extended region on the EVH1 domain, explaining its high affinity.

Contacts between the extended, alpha-helical region of PCARE and ENAH are shown in Figure 1C. 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 atoms of ENAH residues 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 further coordinated by the backbone NH of ENAH Arg81. Most intriguing is the interaction between Val837 on PCARE and the hydrophobic groove in ENAH, which typically engages large aromatic residues. The alpha-helical structure of PCARE$^{828-848}$ buries the smaller Val837 in the same site where phenylalanine can bind. Collectively, our results reveal a noncanonical mode of binding where the FP4 motif of PCARE binds to the ENAH EVH1 domain in a reversed N-to-C orientation and makes extra contacts to achieve high affinity.

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

We previously demonstrated that a 23-residue segment of the PCARE peptide (PCARE$^{826-848}$, which we call PCARE B) is the minimal region required for high-affinity binding to ENAH and that PCARE B exhibits a 70-140-fold preference for binding to ENAH over the VASP or EVL EVH1 domains (Figure 2A). Interestingly, our structure of ENAH EVH1 domain bound to PCARE$^{828-848}$ shows that 16 of the 18 residues that are within 4 Å of PCARE in ENAH are identical in VASP and EVL; the binding sites are highly conserved. Residue 63 is alanine in ENAH and VASP, and the corresponding

residue 64 in EVL is serine. Modeling serine at position 63 in the PCARE-bound structure of ENAH shows that the side-chain hydroxyl group can be readily accommodated in a solvent-facing conformation without interfering with PCARE binding. On the other hand, residue 65 is proline in ENAH and the corresponding residue is valine in EVL and VASP. ENAH Pro65 makes extensive contacts with PCARE and is largely occluded at the domain-peptide interface (Figure 1C). We speculated that proline vs. valine might contribute to the ENAH-binding preference of PCARE. To test this, we made EVL with a valine-to-proline mutation, with the expectation that this would increase PCARE B binding affinity. Surprisingly, EVL V65P bound to PCARE B 5-fold weaker than did wild-type EVL (K_D = 112.1 μM vs. 22 μM, Table 1). In contrast, EVL V65P bound to an FP4-containing ActA peptide, which does not contact Val65 (Barone et al., 2020), with the same affinity as wild-type EVL (K_D = 2.4 μM vs. 2.7 μM, Table 1), indicating that the V65P mutation does not lead to global disruption of the domain structure.

Comparing the structures of ENAH EVH1 domain bound to PCARE$^{828\text{-}848}$ vs. the peptide FPPPP (PDB 1EVH; Prehoda et al., 1999) shows a conformational difference in ENAH: a loop composed of residues 80-86, which forms part of the extended PCARE binding site, is shifted by 3 Å (Figure 2B). Structure gazing suggested that hydrophobic core residues Tyr63 in EVL (Cys62 in ENAH), and Trp89 and Leu15 in VASP (Tyr87 and Val15 in ENAH), are incompatible with this conformational change (Figure 2C). To test this, we made EVL V65P Y62C. This EVL double mutant bound to PCARE B with K_D = 2.2 μM, which is 56-fold lower than the K_D for binding to EVL V65P (Table 1). This striking enhancement in affinity indicates strong coupling between these two mutated positions in EVL. However, these two mutations alone enhanced binding to EVL by only 10-fold over wild type, whereas the difference in binding affinity between ENAH and EVL is 70-fold (Figure 2A). We concluded that a broader set of residues must contribute to stabilizing the ENAH-specific conformation, but it was not readily apparent which residues these might be.

To identify the ENAH residues responsible for PCARE binding specificity, we used the structure-based modeling method dTERMen (Zhou et al., 2017). 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 protein structures. As expected, when scoring different sequences on the structure of ENAH bound to PCARE, the EVL sequence scored considerably worse than the sequence of ENAH itself (Figure 2C). Guided by the dTERMen score, we introduced increasing numbers of residues from ENAH into EVL. Seven replacements were sufficient to recapitulate dTERMen energies similar to that for ENAH in the PCARE$^{828\text{-}848}$-bound conformation (Figure 2C). These residues are distributed across the EVH1 domain, and several are distant from the PCARE$^{828\text{-}848}$ binding site (Figure 2D). We made a mutated EVL EVH1 domain with the 7 corresponding residues from ENAH. This protein, $EVL_{swapped}$, bound as tightly to PCARE B as did ENAH EVH1 (K_D = 0.35 vs. 0.32 μM) (Figure 2D). 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, in just a single attempt, to a combination of residues sufficient to transfer binding specificity.

Engineered binders engage ENAH EVH1 domain with increased affinity and specificity

Motivated by our previous observation that PCARE B can selectively sequester ENAH in cells and inhibit its functions (Hwang et al., 2021), we aimed to design even higher affinity, ENAH-selective peptides. To this end, we took a rational design approach that involved combining affinity-enhancing elements that we previously characterized to target multiple sites on the ENAH EVH1 domain (Figure 3A). Our strategy relied on designing peptides that can simultaneously engage two binding sites on ENAH EVH1: the canonical FP4-binding site that is occupied by FP4 peptides in many Ena/VASP EVH1 crystal structures (Prehoda et al., 1999; Federov et al., 1999, Barone et al., 2020), and a noncanonical site previously identified in VASP EVH1 that we have shown is also important for certain ENAH-peptide complexes (Hwang et al., 2021; Acevedo et al., 2017). We previously created a structural model for how FP4 motifs in a dual-FP4 peptide can simultaneously engage the canonical and noncanonical sites, based on PDB structure 5NC7, which shows ENAH EVH1 bound to two short peptides (Hwang et al., 2021; Barone et al., 2020). As part of this study, we made and tested different combinations of binding motifs and linkers that we predicted could bridge these two sites.

ABI1 is an Ena/VASP interaction partner that contains the sequence FP_8. We have shown that proline residues C-terminal to the FP4 motif, as well as surrounding acidic residues, enhance affinity for the ENAH EVH1 domain (Hwang et al., 2021). LPP is another ENAH binding partner that we have shown engages the noncanonical EVH1 binding site. We fused a 17-residue segment of ABI1 to part of the LPP linker and a second FP4 motif to make ABI1-LPP (Figure 3B). 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 with K_D = 0.76 μM, which is 3-fold tighter than the ABI1 portion alone and 5-fold tighter than the LPP portion alone (Figure 3B), supporting the success of this general approach.

We then 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 while stabilizing the ENAH-specific conformation. Peptide PCARE-Dual bound ~7-fold tighter than PCARE B, with K_D = 50 nM (Figure 3B). 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 3C). 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 (Acevedo et al., 2017), providing the tightest and most specific known binder to the ENAH EVH1 domain to date (Figure 3D).

Discussion

Although they have some distinct cellular functions, the paralogous proteins ENAH, VASP, and EVL are highly conserved in sequence and structure. The EVH1 domains are 100% identical in sequence in the core FP4-binding groove and share 62-72% sequence identity through the rest of the EVH1 domain. FP4-motif peptides engage a

small and relatively flat surface on EVH1; previously solved structures have shown how short peptides bind to this core region or, in some cases, to an immediately adjacent hydrophobic patch (Barone et al., 2020). The limited contacts with a highly conserved, shallow site make it challenging to achieve high-affinity or paralog-selective binding.

The role of ENAH in cancer metastasis has motivated work to identify inhibitors of its EVH1-mediated interactions, and extensive structure-based design and chemical optimization recently led to a high-affinity molecule (K_D = 120 nM) that reduced breast cancer cell extravasation in a zebrafish model. This compound and its analogs mimic the PPII conformation of an FP4 motif. However, because the molecule binds to the highly conserved FP4-binding site, this inhibitor has similar affinity for ENAH, EVL, and VASP (Barone et al., 2020).

There are currently two proteins reported to bind specifically to the ENAH EVH1 domain. The LIM domain of testin binds to ENAH EVH1 with a K_D of 3 μM and does not bind detectably to the EVL or VASP EVH1 domains. Testin achieves paralog specificity by making contacts with ENAH outside of the canonical FP4 groove, at surface sites where the paralogs differ (Boëda et al., 2007). Synthetic mini-protein pGolemi binds to ENAH with K_D = 0.29 μM, moderate selectivity (20-fold) over VASP, and higher selectivity (≥120-fold) over EVL (Golemi-Kotra et al., 2004). The mechanism behind the specificity of pGolemi has not been determined, and mutational data do not readily rationalize a model (Holtzman et al., 2007).

We have now shown that the protein PCARE contains residues adjacent to an FP4 motif that confer both high affinity and selectivity for ENAH over VASP and EVL by stabilizing an ENAH-specific conformation. A dramatic feature of the PCARE-bound ENAH structure is that the LPPPP motif of PCARE binds in a reversed orientation, relative to previously observed FP4 ligands. The PPII helix possesses 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 (Zarrinpar et al., 2003). Although all Ena/VASP and Homer EVH1 domain structures solved so far show the PPII helix engaged in a single direction, the WASP EVH1 domain binds its proline-rich ligand in the opposite direction (Volkman et al., 2002). Consequently, it has long been hypothesized, although never demonstrated, that Ena/VASP EVH1 domains might also bind FP4 motifs in either direction (Ball et al., 2002). Here we show that this is the case for ENAH. Interestingly, the large effect on PCARE binding that we observed for EVL V65P compared to wild-type EVL EVH1 suggests that EVL also binds PCARE in a reversed orientation (relative to FP4 ligands such as ActA), albeit weakly, because this is the orientation that best explains contacts with position 65.

We are not aware of other examples of such dramatic conformational specificity, in which a natural ligand binds selectively to one paralog despite almost complete (~89%) conservation of the binding site amongst its family members. However, this mechanism is reminiscent of how the cancer drug Gleevec achieves 3000-fold selectivity for Abl over Src, despite the fact that the two proteins share ~46% sequence identity across the kinase domain, and ~86% identity in the Gleevec binding site (Seeliger et al., 2007). The mechanism for the selectivity of Gleevec long eluded explanation. As we found here, a few residue swaps based on sequence alignments and structure gazing failed to rescue the affinity of Src for Gleevec, because these mutations failed to account for the

higher-order epistatic interactions that contribute to specificity (Seeliger et al., 2007). Interestingly, because evolution has sampled kinase sequence space under the constraints of epistasis, retracing the evolutionary trajectories that led to extant Src and Abl kinases, using ancestral reconstruction, allowed Wilson et al. to identify residues involved in a hydrogen-bonding network distant from the Gleevec binding interface that are key for selective binding to Abl (Wilson et al., 2015).

We used an alternative approach to discover a set of residues that contribute to specificity, turning to structure-based modeling. Using dTERMen, we scored the compatibility of different sequences with the ENAH-PCARE$^{828\text{-}848}$ structure template and identified residues in EVL that are incompatible with this binding mode. Our method solved the challenging problem of identifying mutations within an epistatic network that contribute to function. Importantly, this facile method is easily generalizable to many protein systems as long as a structure is available.

Our results demonstrate an intriguing example of how nature has evolved a protein that achieves selectivity by exploiting a conformation accessible to only one paralog, rather than by making contacts with paralog-specific residues. Our findings raise the question of how widely this mechanism of selectivity is exploited by other paralogous families, especially within modular interaction domain families. As evidenced by residues identified by dTERMen in our residue swap experiments, hydrophobic core packing plays a key role in enabling new ligand-bound conformations. A recent study demonstrated that randomizing hydrophobic core residues of the SH3 domain from human Fyn tyrosine kinase could effectively switch its affinity and specificity to different ligands (Ben-David et al., 2018). In addition, directed evolution experiments that varied only the hydrophobic core residues of ubiquitin yielded a conformationally stabilized ubiquitin variant that was specific for the deubiquitinase USP7 (Zhang et al., 2012). These results hint that such mechanisms of conformational specificity could be widespread.

Although predicting paralog-specific conformations is difficult when only sequences or single structures are available, uncovering such mechanisms can inform the design of paralog-specific inhibitors in highly conserved families, as demonstrated by Gleevec (which was discovered through high-throughput screening) and by our success in creating PCARE-Dual, which binds ENAH with unprecedented affinity and specificity. Given the antagonistic roles of ENAH and EVL in promoting and suppressing breast cancer metastasis (Roussos et al., 2011; Padilla-Rodriguez et al., 2018), PCARE-Dual is an ideal starting scaffold to engineer anti-metastatic therapies without pleiotropic side effects.

Methods

Biolayer interferometry (BLI), protein purification, and cloning are as described in Hwang, et al. 2021.

Protein constructs

Constructs below were cloned into a pMCSG7 backbone (gift from F. Gertler) which encodes an N-terminal 6xHis-TEV site. Amino acid sequences of proteins are as given:

ENAH EVH1
MHHHHHHSSGVDLGTENLYFQSNAMSEQSICQARAAVMVYDDANKKWVPAGGSTGF
SRVHIYHHTGNNTFRVVGRKIQDHQVVINCAIPKGLKYNQATQTFHQWRDARQVYGLN
FGSKEDANVFASAMMHALEVL*
VASP EVH1
MHHHHHHSSGVDLGTENLYFQSNAMSETVICSSRATVMLYDDGNKRWLPAGTGPQA
FSRVQIYHNPTANSFRVVGRKMQPDQQVVINCAIVRGVKYNQATPNFHQWRDARQV
WGLNFGSKEDAAQFAAGMASALEALE*
EVL EVH1
MHHHHHHSSGVDLGTENLYFQSNAMSEQSICQARASVMVYDDTSKKWVPI**K**PGQQG
FSRINIYHNTASSTFRVVGVKLQDQQVVINYSIVKGLKYNQATPTFHQWRDARQVYGL
NFASKEEATTFSNAMLFALNIMNSQE*
EVL EVH1 V65P[#]
MHHHHHHSSGVDLGTENLYFQSNAMSEQSICQARASVMVYDDTSKKWVPIPGQQGF
SRINIYHNTASNTFRVVGVKLQDQQVVINYSIPKGLKYNQATPTFHQWRDARQVYGLN
FASKEEATTFSNAMLFALNIM*
EVL EVH1 V65P Y62C[#]
MHHHHHHSSGVDLGTENLYFQSNAMSEQSICQARASVMVYDDTSKKWVPIPGQQGF
SRINIYHNTASNTFRVVGVKLQDQQVVINCSIPKGLKYNQATPTFHQWRDARQVYGLN
FASKEEATTFSNAMLFALNIM*
EVL EVH1 swapped (7-residue swapped)[#]
MHHHHHHSSGVDLGTENLYFQSNAMSEQSICQARASVMVYDDTSKKWVPAGGQQG
FSRINIYHNTASNTFRVVGVKLQDQQVVINCSIPKGLKYNQATPTFHQWRDARQVYGL
NFASKEEATTFANAMLFALEIL*
ENAH EVH1-PCARE (for crystallography)
MHHHHHHSSGVDLGTENLYFQSNAMSEQSICQARAAVMVYDDANKKWVPAGGSTGF
SRVHIYHHTGNNTFRVVGRKIQDHQVVINCAIPKGLKYNQATQTFHQWRDARQVYGLN
FGSKEDANVFASAMMHALEVLGGSGSGAAKSEELSCEMEGNLEHLPPPPMEVLMDK
SFASLES*
SUMO-ActA
MAGGLNDIFEAQKIEWHEDTGGSSHHHHHHGSGSGSDSEVNQEAKPEVKPEVKPET
HINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLTFLYDGIEIQADQTPEDLD
MEDNDIIEAHREQIGGGFNAPATSEPSSFEFPPPPTEDELEIIRETASSLDS*
SUMO-PCARE B
MAGGLNDIFEAQKIEWHEDTGGSSHHHHHHGSGSGSDSEVNQEAKPEVKPEVKPET
HINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLTFLYDGIEIQADQTPEDLD
MEDNDIIEAHREQIGGSGSGNLEH**LPPPP**MEVLMDKSFASLES
[#] These EVL constructs had a single-residue deletion of WT EVL residue Lys 27, which was removed to make the lengths of ENAH EVH1 and the mutated EVL EVH1 domains equal. In PDB structure 1QC6, Lys27 is at the end of a loop that is disordered and has a high B-factor.

Crystallography
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 at 18 °C. 1.5 µL of ENAH-PCARE (769

μM in 20 mM HEPES, 150 mM NaCl, 1 mM DTT) was mixed with 0.5 μL of reservoir solution, and football-shaped crystals appeared in two days. Diffraction data were collected at the Advanced Photon Source at Argonne National Laboratory, NE-CAT beamline 24-IDE. The ENAH-PCARE data were integrated and scaled to 1.65 Å with XDS, and the structure was solved with molecular replacement using the ENAH EVH1 structure 6RD2 as a search model. The structure was refined with iterative rounds of model rebuilding with PHENIX and COOT. Table 2 reports refinement statistics. The structure is deposited in the PDB with the identifier 7LXF. Note that the PCARE$^{828\text{-}848}$ peptide is numbered as 133-153 in the PDB file in accordance with the ENAH-PCARE fusion protein numbering.

dTERMen

The dTERMen scoring function and protocol are described in Zhou et al., 2020. The method requires that a template-specific scoring function be computed, based on statistics derived from structures in the PDB. After that, the inputs to dTERMen are the backbone coordinates of a structure and a sequence. dTERMen returns a score for the input sequence adopting the input structure; lower scores correspond to lower energies. Side-chain positions are not modeled explicitly. To score the EVL sequence on the ENAH-PCARE backbone template, we generated pairwise alignments of the EVH1 domains of ENAH and EVL to determine how to map sequence to structure. The EVL EVH1 domain is longer than that of ENAH by one residue, so Lys27 was removed from EVL. Note that Lys27 was also removed from the manually chosen EVL V65P and EVL V65P Y62C mutants.

We then used dTERMen to score all possible combinations of residue swaps between EVL and ENAH, up to 6 possible positions. Residue swap combinations that led to the minimum energy score were recorded. The best 6 mutations were sufficient to nearly recapitulate the energy score of the native ENAH sequence on the ENAH PCARE template. We also included an I26A mutation based on manual inspection of the ENAH-PCARE$^{828\text{-}848}$ structure, which also lowered the dTERMen energy. We cloned, overexpressed, and purified this swapped EVL sequence, as described in Hwang et al., 2021, to test for binding to PCARE B

Acknowledgments

This project was supported by NIGMS award R01 GM129007 to A.E.K. 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 beamline is funded by 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 MIT Structural Biology Core for assistance with X-ray crystallography and the MIT Biophysical Instrumentation Facility for instrumentation resources. We thank LRF Backman for help with X-ray data collection. We thank F. Gertler and J. Tadros for

constructs. We thank members of the Keating lab, Mouneimne lab, and F. Gertler for their thoughtful input on the manuscript.

References

1. Acevedo LA, Greenwood AI, Nicholson LK. 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).
2. Aszódi A, Pfeifer A, Ahmad M, Glauner M, Zhou XH, Ny L, Andersson KE, Kehrel B, Offermanns S, Fässler R. The vasodilator-stimulated phosphoprotein (VASP) is involved in cGMP- and cAMP-mediated inhibition of agonist-induced platelet aggregation, but is dispensable for smooth muscle function. *EMBO J.* **18**, 37-48 (1999).
3. Ball LJ, 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).
4. Ball LJ, Jarchau T, Oschkinat H, Walter U. EVH1 domains: structure, function and interactions. *FEBS Lett.* **513**, 45–52 (2002).
5. Bardwell AJ, Frankson E, Bardwell L. Selectivity of docking sites in MAPK kinases. *J Biol Chem*. **284**, 13165-13173 (2009).
6. Ben-David M, Huang H, Sun MGF, Corbi-Verge C, Petsalaki E, Liu K, Gfeller D, Garg P, Tempel W, Sochirca I, Shifman JM, Davidson A, Min J, Kim PM, Sidhu SS. Allosteric Modulation of Binding Specificity by Alternative Packing of Protein Cores. *J Mol Biol.* **43**, 336-350 (2018).
7. Boëda B, Briggs DC, Higgins T, Garvalov BK, Fadden AJ, McDonald NQ, Way M. Tes, a specific Mena interacting partner, breaks the rules for EVH1 binding. *Mol. Cell* **28**, 1071–1082 (2007).
8. Fedorov AA, Fedorov E, Gertler F, Almo SC. 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).
9. Golemi-Kotra D, Mahaffy R, Footer MJ, Holtzman JH, Pollard TD, Theriot JA, Schepartz A. High affinity, paralog-specific recognition of the Mena EVH1 domain by a miniature protein. *J. Am. Chem. Soc.* **126**, 4–5 (2004).
10. Hansen SD, Mullins RD. Lamellipodin promotes actin assembly by clustering Ena/VASP proteins and tethering them to actin filaments. *elife.* **4**, e06585 (2015).
11. Hause RJ, Leung KK, Barkinge JL, Ciaccio MF, Chuu C pin, Jones RB: Comprehensive Binary Interaction Mapping of SH2 Domains via Fluorescence Polarization Reveals Novel Functional Diversification of ErbB Receptors. *PLoS One* **7**, e44471 (2012).
12. Hwang TH, Parker SS, Hill SM, Ilunga MW, Grant RA, Sivaraman V, Mouneimne G, Keating AE. Native proline-rich motifs exploit sequence context to target actin-remodeling Ena/VASP proteins. *Co-submitted to eLife* (2021).
13. Kelil A, Levy ED, Michnick SW. Evolution of domain-peptide interactions to coadapt specificity and affinity to functional diversity. *Proc Natl Acad Sci U S A* **5**, E3862-3871 (2016).

14. Krause M, Leslie JD, Stewart M, Lafuente EM, Valderrama F, Jagannathan R, Strasser GA, Rubinson DA, Liu H, Way M, Yaffe MB, Boussiotis VA, Gertler FB. *Dev. Cell* **7**, 571-583 (2004).
15. Kwiatkowski AV, Gertler FB, Loureiro JJ. Function and regulation of Ena/VASP proteins. *Trends Cell Biol.* **13**, 386-92 (2003).
16. Kwiatkowski AV, Rubinson DA, Dent EW, Edward van Veen J, Leslie JD, Zhang J, Mebane LM, Philippar U, Pinheiro EM, Burds AA, Bronson RT, Mori S, Fässler R, Gertler FB. Ena/VASP Is Required for neuritogenesis in the developing cortex. *Neuron*. **56**, 441-455 (2007).
17. Lanier LM, Gates MA, Witke W, Menzies AS, Wehman AM, Macklis JD, Kwiatkowski D, Soriano P, Gertler FB. Mena is required for neurulation and commissure formation. *Neuron.* **22**, 313-325 (1999).
18. 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).
19. Oudin MJ, Jonas O, Kosciuk T, Broye LC, Guido BC, Wyckoff J, Riquelme D, Lamar JM, Asokan SB, Whittaker C, Ma D, Langer R, Cima MJ, Wisinski KB, Hynes RO, Lauffenburger DA, Keely PJ, Bear JE, Gertler FB. Tumor Cell-Driven Extracellular Matrix Remodeling Drives Haptotaxis during Metastatic Progression. *Cancer Discov.* **6**, 516-31 (2016).
20. Padilla-Rodriguez M, Parker SS, Adams DG, Westerling T, Puleo JI, Watson AW, Hill SM, Noon M, Gaudin R, Aaron J, Tong D, Roe DJ, Knudsen B, Mouneimne G. The actin cytoskeletal architecture of estrogen receptor positive breast cancer cells suppresses invasion. *Nature Commun.* **9,** 2980 (2018).
21. Pawson T, Nash P. Protein-protein interactions define specificity in signal transduction. *Genes Dev.* **14**, 1027-47 (2000).
22. Prehoda KE, Lee DJ, Lim WA. 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).
23. Vidaki M, Drees F, Saxena T, Lanslots E, Taliaferro MJ, Tatarakis A, Burge CB, Wang ET, Gertler FB. A Requirement for Mena, an Actin Regulator, in Local mRNA Translation in Developing Neurons. *Neuron* **95**, 608–622 (2017).
24. Robertson AS, Allwood EG, Smith AP, et al. The WASP homologue Las17 activates the novel actin-regulatory activity of Ysc84 to promote endocytosis in yeast. *Mol Biol Cell* **20**, 1618-1628 (2009).
25. Roussos ET, Wang Y, Wyckoff JB, Sellers RS, Wang W, Li J, Pollard JW, Gertler FB, Condeelis JS. Mena deficiency delays tumor progression and decreases metastasis in polyoma middle-T transgenic mouse mammary tumors. *Breast Cancer Res.***12**, R101 (2010).
26. Rowland MA, Greenbaum JM, Deeds EJ. Crosstalk and the evolvability of intracellular communication. *Nat Commun*. **8**, 16009 (2017).
27. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol*. **21**, 50–61 (2011).

28. Seeliger MA, Nagar B, Frank F, Cao X, Henderson MN, Kuriyan J. c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. *Structure.* **15**, 299-311 (2007).
29. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. Rewiring the specificity of two-component signal transduction systems. *Cell* **13**, 1043-1054 (2008).
30. Volkman BF, Prehoda KE, Scott JA, Peterson FC, Lim WA. Structure of the N-WASP EVH1 domain-WIP complex: insight into the molecular basis of Wiskott-Aldrich Syndrome. *Cell* **111**, 565–576 (2002).
31. Wilson C, Agafonov RV, Hoemberger M, Kutter S, Zorba A, Halpin J, Buosi V, Otten, R, Waterman D, Theobald DL, Kern D. Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347**, 882–886 (2015).
32. Xin X, Gfeller D, Cheng J, Tonikian R, Sun L, Guo A, Lopez L, Pavlenco A, Akintobi A, Zhang Y, et al.: SH3 interactome conserves general function over specific form. *Mol Syst Biol* **9**, 652–652 (2014).
33. Zarrinpar A, Bhattacharyya RP, Lim WA. The structure and function of proline recognition domains. *Sci. Signal* 2003, RE8 (2003).
34. Zarrinpar A, Park SH, Lim WA. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676-80 (2003).
35. Zhang Y, Han G, Fan B, Zhou Y, Zhou X, Wei L, Zhang J. Green tea (-)-epigallocatechin-3-gallate down-regulates VASP expression and inhibits breast cancer cell migration and invasion by attenuating Rac1 activity. *Eur J Pharmacol*. **606**, 172-179 (2009).
36. Zhang Y, Zhou L, Rouge L, Phillips AH, Lam C, Liu P, Sandoval W, Helgason E, Murray JM, Wertz IE, Corn JE. Conformational stabilization of ubiquitin yields potent and selective inhibitors of USP7. *Nat Chem Biol* **9,** 51–58 (2013).
37. Zhou J, Panaitiu AE, 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).

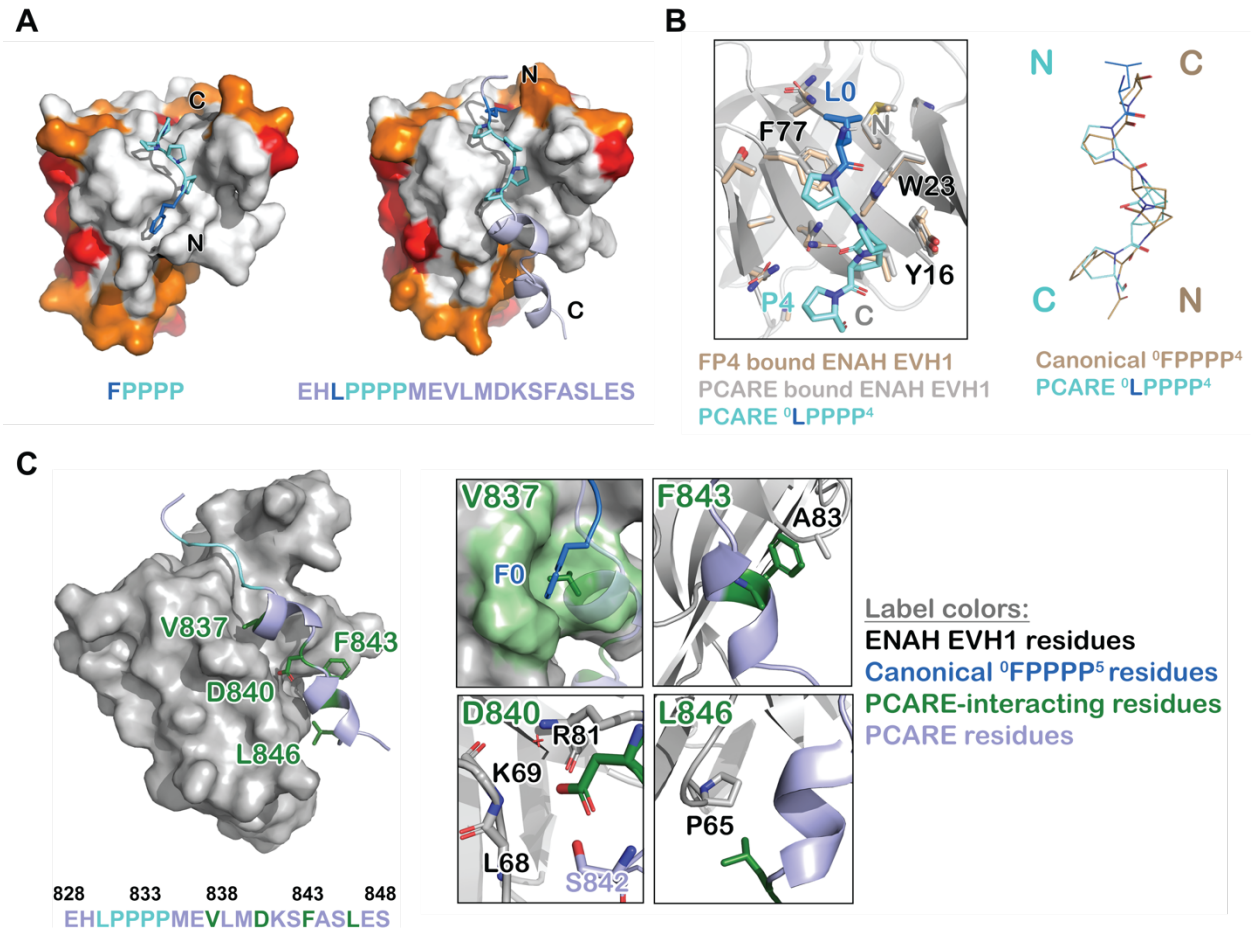
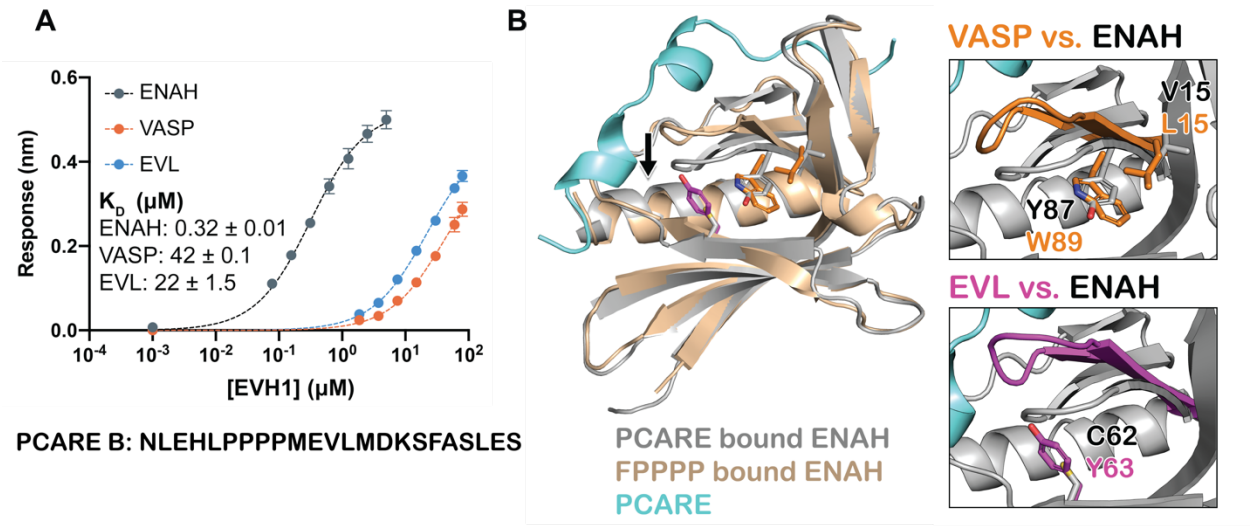


Figure 1. ENAH EVH1 interactions with FPPPP vs. PCARE. (A) Surface representation of ENAH EVH1 highlighting conservation among Ena/VASP paralogs. Residues shared between all three paralogs - white; residues shared by two paralogs - orange, residues that differ in each paralog - red. On the left is ENAH EVH1 bound to FP_4 (PDB 1EVH); on the right is ENAH EVH1 bound to the PCARE peptide (this work). (B) View comparing the orientations of an 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. (C) 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.



C

Number of mutations	Residue Swaps	dTERMen score
0 (Native EVL)		88.3
2	P27G, M111L	78.0
3	P27G, Y62C, M111L	74.9
4	P27G, Y62C, N109E, M111L	72.0
5	P27G, Y62C, S101A, N109E, M111L	69.3
6	P27G, Y62C, V65P, S101A, N109E, M111L	67.2
7	I26A*, P27G, Y62C, V65P, S101A, N109E, M111L	65.2
Native ENAH		63.0

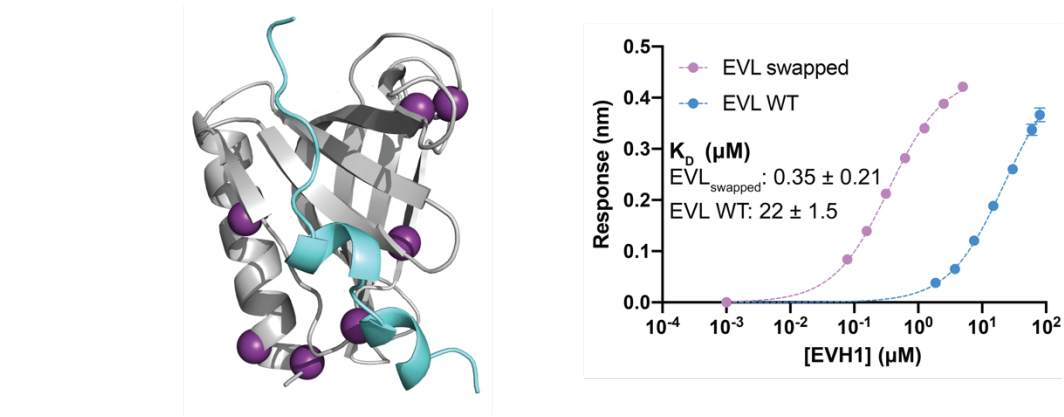


Figure 2. PCARE achieves paralog selectivity by stabilizing an ENAH EVH1 domain specific conformation. (A) Biolayer interferometry (BLI) binding curves for PCARE B binding to ENAH, EVL, or VASP EVH1 domains. (B) Superposition of ENAH EVH1 domains bound to PCARE or FP_4 peptide (1EVH). The black arrow 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. (C) Lowest dTERMen energy obtained when swapping 0-6 residues from ENAH into EVL, when modeling on the structure of ENAH EVH1 bound to PCARE. * indicates the mutation was added based on manual inspection. (D) 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. Data for 2A and 2D reported as the mean $K_D \pm$ SD for two independent replicates.

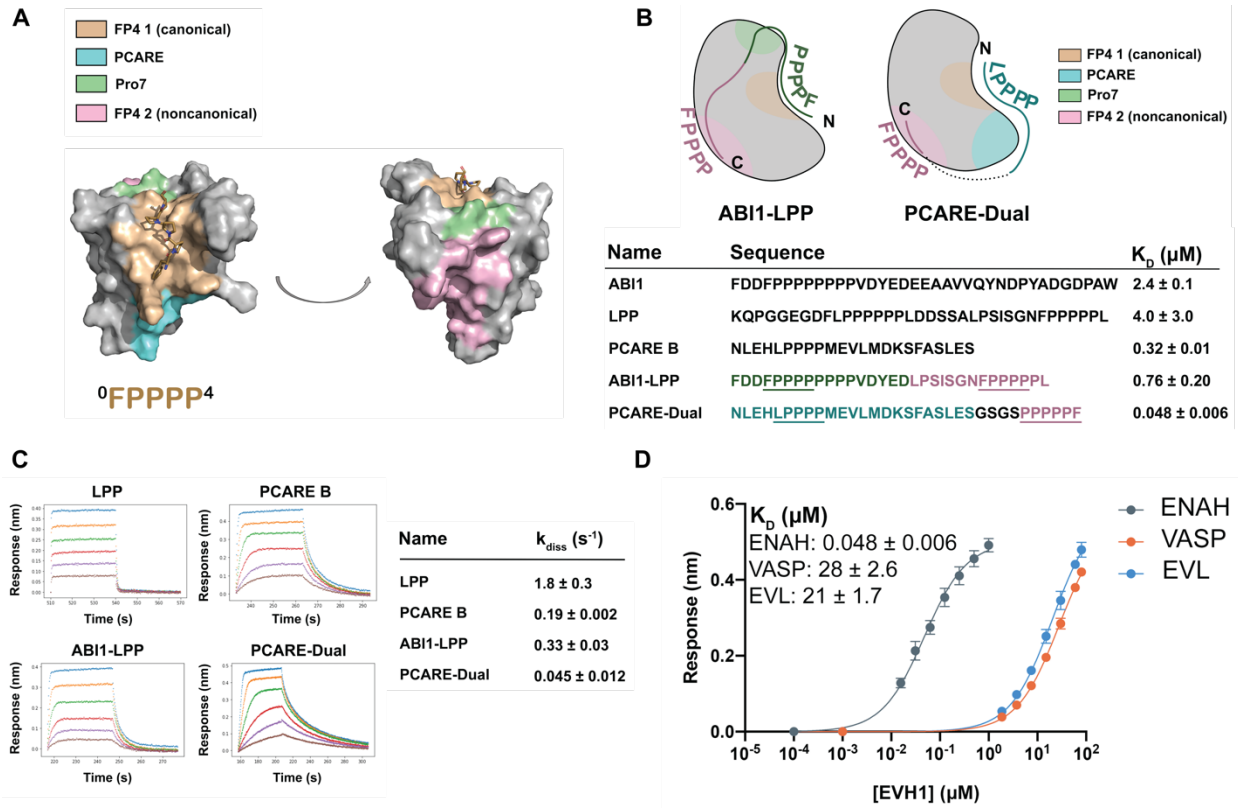


Figure 3. Engineered peptides bind to the ENAH EVH1 domain with high affinity and specificity. (A) Surface representation of ENAH EVH1 with binding sites discussed in this study indicated. (B) Design scheme for high affinity binders ABI1-LPP and PCARE-Dual. (C) BLI binding and dissociation curves. Blue, orange, green, red, purple, and brown curves denote EVH1 concentrations in descending order. 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, 0.063, 0.031, 0.016, 0.0078 μM. Values reported as $k_{diss} \pm$ SD for two independent BLI replicates. (D) BLI binding curves for PCARE-Dual binding to the EVH1 domains of ENAH, VASP, or EVL. Data for 3B and 3D are reported as the mean $K_D \pm$ SD for two independent BLI replicates.

Table 1. Affinity of EVL mutants for ActA peptide and PCARE B.

	ActA K_D (μM)[a, b]	**PCARE B K_D (μM)[a]**
Wild type EVL	2.7 ± 0.3	22 ± 1.5
EVL V65P	2.4 ± 0.1	112.1 ± 9.6
EVL V65P Y62C	5.8 ± 1.0	2.2 ± 0.1

[a] Affinities determined by BLI as described in the methods in Hwang et al., 2021. Data reported as the mean K_D ± SD for two independent BLI replicates.
[b] The ActA peptide has sequence FNAPATSEPSSFE**FPPPP**TEDELEIIRETASSLDS (see methods for the exact construct tested).

Table 2. Refinement table for ENAH-PCARE crystal structure

	ENAH-PCARE (7LXF)
Wavelength	
Resolution range	45.14 - 1.65 (1.71 - 1.65)
Space group	P 61 2 2
Unit cell	52.12 52.12 197.85 90 90 120
Total reflections	742185 (53799)
Unique reflections	20193 (1951)
Multiplicity	36.8 (27.6)
Completeness (%)	99.87 (99.44)
Mean I/sigma(I)	27.78 (0.68)
Wilson B-factor	36.47
R-merge	0.0684 (3.906)
R-meas	0.0694 (3.979)
R-pim	0.0114 (0.750)
CC1/2	1 (0.659)
CC*	1 (0.891)
Reflections used in refinement	20166 (1937)
Reflections used for R-free	1012 (98)
R-work	0.217 (0.346)
R-free	0.235 (0.356)
CC(work)	0.960 (0.786)
CC(free)	0.936 (0.720)
Number of non-hydrogen atoms	1090
macromolecules	1054
solvent	36
Protein residues	134

RMS(bonds)	0.011
RMS(angles)	1.08
Ramachandran favored (%)	96.15
Ramachandran allowed (%)	3.85
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.89
Clashscore	0.48
Average B-factor	48.68
macromolecules	48.85
solvent	43.72

Statistics for the highest-resolution shell are shown in parentheses.

Note that the PCARE$^{828\text{-}848}$ peptide is numbered as 133-153 in the PDB file, based on residue numbers in the domain-peptide fusion.