Elaboration of the Homer1 Recognition Landscape Reveals Incomplete Divergence of Paralogous EVH1 Domains

Authors: *Avinoam Singer, Alejandra Ramos, Amy E. Keating*

Abstract

Short sequences that mediate interactions with modular binding domains are ubiquitous throughout eukaryotic proteomes. Networks of Short Linear Motifs (SLiMs) and their corresponding binding domains orchestrate many cellular processes, and the low mutational barrier to evolving novel interactions provides a way for biological systems to rapidly sample selectable phenotypes. Mapping SLiM binding specificity and the rules that govern SLiM evolution is fundamental to uncovering the pathways regulated by these networks and developing the tools to manipulate them. Here, we use high-throughput screening of the human proteome to identify sequences that bind to the SLiM-binding Enabled/VASP homology 1 (EVH1) domain of the postsynaptic density scaffolding protein Homer1. In doing so, we expanded our understanding of the determinants of Homer EVH1 binding preferences and defined a new motif that can facilitate the discovery of Homer-mediated interactions. Interestingly, the Homer1 EVH1 domain preferentially binds to sequences containing an N-terminally overlapping motif from the paralogous family of Ena/VASP actin polymerases, and many of these sequences can bind to EVH1 domains from both protein families. We provide evidence from orthologous EVH1 domains in pre-metazoan organisms that the overlap in human Ena/VASP and Homer binding preferences corresponds to an incomplete divergence from a common Ena/VASP ancestor. Given this overlap in binding profiles, promiscuous sequences that can be recognized by both families likely achieve their specificity through extrinsic regulatory strategies. In some cases, these sequences may confer a benefit through their mutual exclusivity, and explain why these paralogs incompletely diverged despite the accessibility of further diverged isoforms.

Introduction

Biological systems encode numerous modular interaction domains that have expanded into families of paralogs [1]. Many of these domains recognize stretches of 3-10 sequential residues within intrinsically disordered regions, termed Short Linear Motifs (SLiMs) [2]. Compared to interactions between globular domains, SLiM-mediated interactions are typically weak (μM) and transient, making them ideal for dynamic processes such as signaling [3]. SLiMs are also used for protein localization and to coordinate molecular assemblies [4].

SLiM Binding Domains (SBD) have characteristic motif preferences that are often described using regular expressions that indicate invariant or common residues found in known interaction partners. Matches to these low-information expressions are prevalent throughout the proteome, suggesting that these short sequences alone are poor predictors of binding. Indeed, high-throughput screens of peptide libraries have shown that although most sequences that bind to a given domain share features of a defining motif, not every sequence that matches that motif will bind. Many case studies illustrate how residues outside the core motif can modulate both the affinity and specificity of binding, demonstrating that motifs provide only a partial picture of domain-SLiM molecular recognition [5], [6].

Nature has harnessed SLiM interaction networks to build increasingly complex systems, as evidenced by the rapid expansion of SBD families in higher eukaryotes [7]. As SBD modules are duplicated, paralogous domains must distinguish their networks to minimize competitive interference and support new functions [8], [9]. Distinct interaction niches can be established in a manner that is extrinsic to the properties of the SBD domain itself, through means such as spatiotemporal compartmentalization of protein expression. Alternatively, this can be achieved intrinsically, at the level of molecular recognition, by changing the sequence to which a domain binds [10].

The dual promise of unraveling SLiM-mediated networks to uncover novel biology and rewiring such networks to build pathways, e.g., for applications in synthetic biology or as molecular therapies, necessitates a clear understanding of how specificity is established. Here, we explore this question by investigating the divergence of Enabled/VASP homology 1 (EVH1) domains. The human genome encodes 5 EVH1-containing protein families: WASP, SPRED, PP4, Ena/VASP, and Homer [11], [12]. Ena/VASP and Homer are thought to have diverged from a common ancestor, and distinct Ena/VASP and Homer proteins in the choanoflagellate *Salpingoeca Rosetta* suggest that this event occurred roughly 600 million years ago [13], [14].

The Ena/VASP family encompasses 3 members: ENAH, VASP, and Evl [15]. These tetrameric proteins regulate the actin cytoskeleton by serving as a scaffold using their EVH1 domains and acting as processive actin polymerases via a C-terminal EVH2 domain [16]–[18]. Ena/VASP proteins are found at sites of actin polymerization, such as filopodia, lamellipodia, and neuronal projections [19]–[22].

Like Ena/VASP proteins, the Homer family contains 3 members: Homer1, Homer2, and Homer3 [23]. The constitutively expressed tetrameric Homer proteins function as scaffolds within the post-synaptic density of neurons [24], [25]. Alternatively spliced, monomeric Homer isoforms dynamically regulate signaling pathways by competing with tetrameric Homer and dissembling these complexes in response to neuronal activity [26], [27]. In this capacity, Homer contributes to synaptic plasticity by regulating dendritic spine morphology,

perhaps through direct association with actin cytoskeleton proteins such as the actin-bunding protein DBN1 [28], [29]. Homer proteins are also expressed in a variety of non-neuronal tissues [26]. For example, Homer3 negatively regulates T-cell activation by competing with calcineurin for binding to NFATc2 [30].

Sequences of known Ena/VASP partners have revealed the interaction motif [FWYL]PxΦP (Φ = FWYPIALV; x = any amino acid) [31], [32]. Several studies have also demonstrated strategies by which regions flanking the core Ena/VASP motif can modulate both affinity and specificity for the EVH1 domain [32]–[34]. Fewer Homer binding partners have been validated, and the sequence preferences are less well-defined. Alignments of Homer binding partners suggest a consensus sequence of PPxxF; peripheral regions that modulate the affinity of this motif are poorly understood [24], [31]. Structural alignments of EVH1 domains in complex with their ligands show that Ena/VASP and Homer families of proteins use partially overlapping binding sites to engage their proline-rich ligands [35], [36].

In this study, we applied high-throughput screening of a library derived from the human proteome to identify Homer1 binding peptides. We found that Homer1 binds to a more general motif than PPxxF and demonstrated how residues flanking the core motif can tune the affinity of this interaction. Notably, compared to the motif alone, Homer1 binds with enhanced affinity to many short segments in the human proteome that match both the Ena/VASP and Homer binding motifs, despite lacking structural features thought to confer these preferences in Ena/VASP proteins. Specifically, we demonstrate that many peptides that contain an N-terminal, overlapping Ena/VASP motif bind with increased affinity to Homer1 while retaining the ability to bind to ENAH. An examination of orthologous EVH1 domains in pre-metazoan organisms suggests that this Homer preference is a vestige of EVH1 domains found in organisms that pre-date the duplication event that gave rise to distinct Ena/VASP and Homer protein families.

Results

Structural Comparison of Related EVH1 Domains

Despite only sharing 30% overall sequence identity, the Ena/VASP and Homer EVH1 domain binding sites exhibit a high degree of structural similarity (Figure 1.1A). Each site is comprised of a proline-binding groove and a neighboring hydrophobic pocket that accommodates the core motif bound by each protein family [35], [36]. (Figure 1.1B).

Within the proline-binding grooves of Ena/VASP and Homer, aromatic residues ($W23_{ENAH}$, $F77_{ENAH}$, $Y16_{ENAH}$; $W24_{Homer1}$, $F74_{Homer1}$, $F14_{Homer1}$) accommodate the distinct polyproline-II helical structure adopted by the proline-tracts of bound peptides ($^{-3}$[FWYL]**PxΦP**^{1}; ^{0}**PP**xxF^{4}). The positioning of aromatic residues within the groove is thought to define the locations of requisite prolines within the preferred motif. Both families conserve aromatic residues at positions 23 (ENAH)/24 (Homer1) and 77 (ENAH)/77 (Homer1). The unique Ena/VASP aromatic residue, $Y16_{ENAH}$ ($I16_{Homer1}$) engages Pro^{-2}, and a preference for Pro at this position is notably absent from the Homer motif. The unique Homer aromatic residue, $F14_{Homer1}$ ($M14_{ENAH}$) engages Pro^{1}, which is also engaged by $F74_{Homer1}$ ($F77_{ENAH}$). Thus, this residue seems to reinforce the shared Ena/VASP and Homer preference for proline at this position.

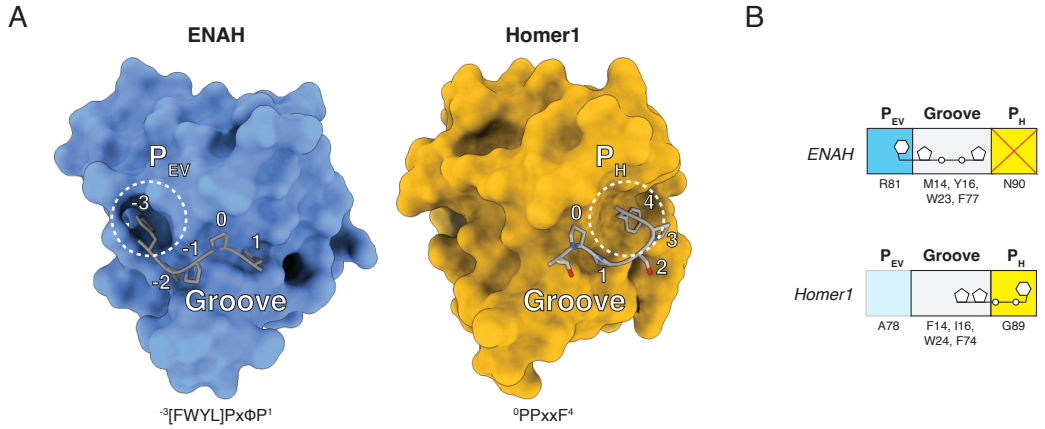


Figure 1.1: (A) Structures of ENAH (1EVH) and Homer1 (1DDV) in complex with representative ligands. Peptide numbering is relative to the Homer PPxxF motif. For Ena/VASP motifs, the numbering corresponds to the position of the overlapping Pro in both motifs [35], [36]. (B) Cartoon representation of the modular binding pockets along the Ena/VASP and Homer EVH1 domains.

The specificity of each domain arises from distinct hydrophobic pockets on opposite sides of the proline-binding groove. The Ena/VASP pocket, which we designate here as P_{EV}, is primarily formed by $K69_{ENAH}$ ($T66_{Homer1}$), $N71_{ENAH}$ ($T68_{Homer1}$), and $R81_{ENAH}$ ($A78_{Homer1}$), which can form a cation-pi interaction when the residue upstream of the motif prolines is aromatic ($^{-3}$**[FWYL]**PxΦP^{1}). While the Homer EVH1 domain lacks a well-formed P_{EV}, the residues in this region do not visibly occlude this pocket. The Homer pocket, which we call P_H, is formed by $G89_{Homer1}$ (^{0}PPxx**F**^{4}). In Ena/VASP proteins, this pocket is obstructed by the corresponding $N90_{ENAH}$.

In addition to the core motif, residues in the surrounding sequence can contribute to the affinity and specificity of a peptide for an SBD [5]. Several affinity-enhancing elements have been discovered for Ena/VASP domains. Acidic residues flanking the $^{-3}$[FWYL]PxΦP^{1} motif strengthen binding due to positively charged patches along the domain [32]. C-terminal proline extensions, such as in a peptide from ABI1 ($^{-3}$FPPPP**PPP**^{4}), position the motif-trailing residues so they can make favorable hydrophobic contacts near the region corresponding to Homer P_H [33]. Finally, Ena/VASP EVH1 domains have been shown to bind a secondary motif at a pseudo-symmetric binding site on the opposite face of the domain (FPPPx_n**FPPPP**) [33], [34]. Few flanking elements have been uncovered for Homer EVH1 domains, though examinations of a handful of known Homer binding partners and mutational studies of a SLiM from the protein DBN1 indicate that Homer prefers an N-terminal Leu **L**$(x)_{1\text{-}2}$PPxxF^{4} and a C-terminal Asp/Asn ^{0}PPxxFx**[DN]** [29].

A Proteome-wide Screen Elaborates the Homer1 Binding Motif

High-throughput screening of peptide libraries can reveal molecular determinants of SBD interactions [37]. The T7-pep library encodes the human proteome as 416,611 36-mer peptides and provides a biologically relevant context for identifying putative SBD binding preferences

Name	Sequence
DBN1	LNFDELPE**PPATF**CDPEEVE
WBP2	**PPPEF**YPGPPMMDGAMGYVQP**PPPPY**PGPMEPPVSG
RAX	GFGPPAQSLPASYTPPP**PPPPF**LNSPPLGPGLQPLA
CAP2	KSSEMNILIPQDGDYREFP**IPEQF**KTAWDGSKLITE
WIPF2	VEDFP**APEEY**KHFQRIYPSKTNRAARGAPPLPPILR

Table 1.1: T7-pep derived and control (DBN1) peptides that bind to Homer1 via bacterial surface display. The putative Homer1 binding site is shown in bold.

and native interaction partners (Figure 1.2A) [38]. Hwang et al. formatted this library for bacterial surface display at the C-terminus of the *Escherichia coli* surface protein eCPX and used screening to identify elements that flank the Ena/VASP motif and modulate binding to ENAH [33].

Preliminary screening of the T7-pep library for peptides that bind the Homer1 EVH1 domain yielded several sequences containing the previously identified PPxxF motif, including peptides from the metabotropic glutamate receptor GRM5, a well-established Homer interaction partner, and WBP2, a protein that has been shown to bind to Homer3 (Table 1.1) [24], [39]. We also discovered matches to this motif in peptides that bind to Homer1 but are not annotated as Homer binding partners, such as the protein RAX, which regulates retinal and neural development [40]. Homer1 EVH1 domain bound peptides listed in Table 1 on the cell surface, as a tetramer, and in biolayer interferometry (BLI) experiments in which the peptide was immobilized on a surface as a monomer, validating the use of the T7-pep library to identify Homer1 binding sequences (Figure 1.2B,C).

In addition to binders with sequences containing the canonical PPxxF motif, we identified several sequences that deviate from this motif by substitution of a small hydrophobic residue at Pro^0 or substitution of Trp or Tyr at Phe^4 (Table 1.2). Testing bacterial clones displaying peptides from the proteins WIPF2 (APEEY) and CAP2 (IPEQF) confirmed binding to Homer1, and these peptides did not bind detectably to established Homer1 mutants with a disrupted proline binding groove ($Homer1^{W24A}$) or with an occluded P_H pocket ($Homer1^{G89N}$), suggesting a shared binding mode with the canonical motif (Figure 1.2B) [35]. BLI measurements showed that these peptides that do not match the consensus motif bind with lower affinity than those including PPxxF sequences (Figure 1.2B, Table 1.2). However, low-affinity motifs may be relevant in avid contexts, such as multimeric complexes or proteins containing several motifs [5]. Notably, full-length CAP2 and WIPF2 contain multiple instances of both canonical and non-canonical Homer motifs.

We performed an enrichment screen of the T7-pep library to identify additional Homer1 binders. Using an expanded Homer binding motif of ^{0}[PIALV]Pxx[FWY]^4, we identified 360 weakly enriching and 62 strongly enriching motif-containing sequences, as defined in the methods (Figure S1.9A, Supplementary Table 1). Subsequent validation using BLI identified peptides that bound with affinities ranging from 15-100 μM, which is within the range of established Homer binding sequences (Table 1.2). Interestingly, a peptide from the protein PSORS1C2, which lacks a canonical motif but contains the sequence APPLF, bound with $K_D = 15.2 \pm 0.5$ μM, which is similar to the affinity of peptides containing the canonical

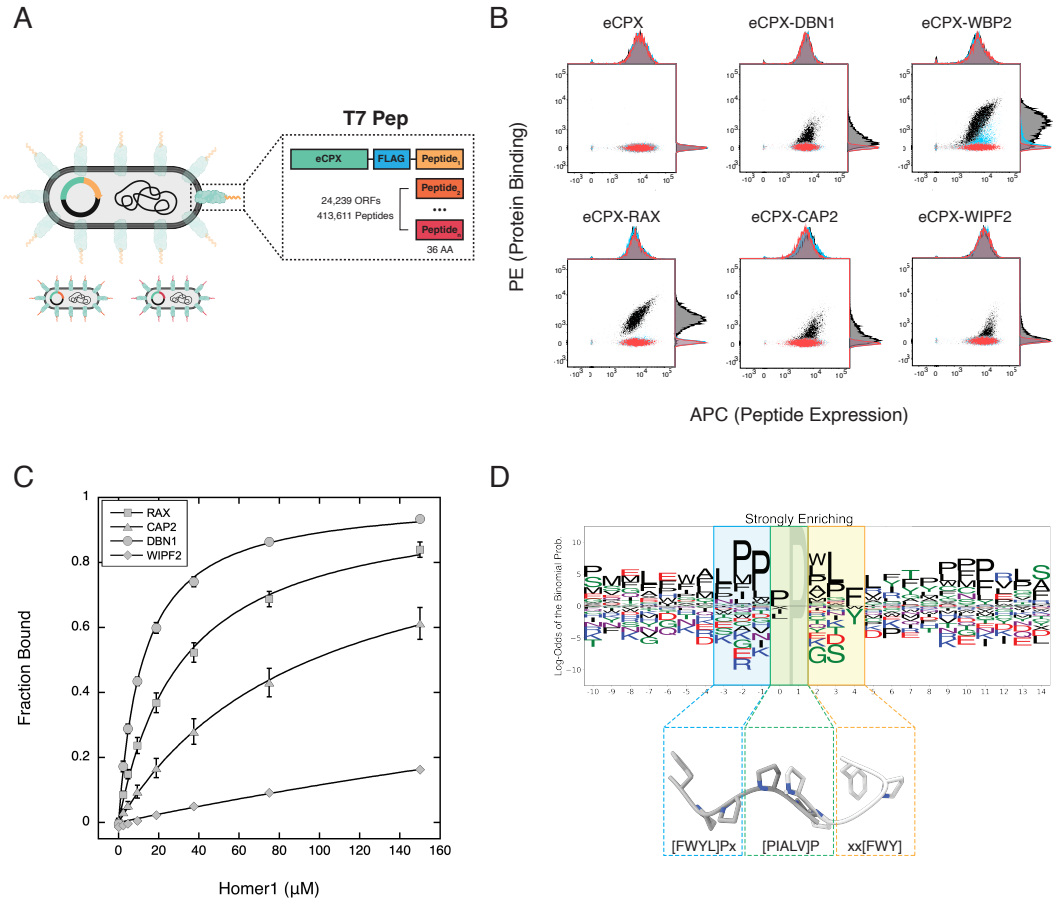


Figure 1.2: (A) The T7-pep library encodes 36-residue fragments of the human proteome as C-terminal extensions of the *E. coli* surface protein eCPX. (B) Single clone surface display measurements of control (eCPX, DBN1) and T7-pep derived (WBP2, RAX, CAP2, WIPF2) peptides against Homer1 (black) and binding site mutatants (Red = Homer1^{W24A}, Blue = Homer1^{G89N}). (C) BLI measurements of control (DBN1) and T7-pep derived (RAX, CAP2, WIPF2) peptides. (D) A logo of T7-pep sequences that strongly enrich for Homer1, aligned by their motif. Structural superposition of ENAH (1EVH) and Homer1 (1DDV) EVH1 domains (not shown) in complex with representative peptides demonstrates the Homer1 preference for an N-terminal, Ena/VASP-like flank (blue = Ena/VASP flank, green = overlapping prolines, yellow = Homer-specific motif residues) [35], [36].

Name	Sequence	K_D (μM)
SMR3B	YPPGPLA**PPQPF**GPGFVPPP**PPPPY**GPGRIPPPPPA	14.7 ± 0.8
PSORS1C2	PPAEDREEAGSPTLPQGPPVPGDPWPG**APPLF**EDPP	15.2 ± 0.5
HOXB2	ETFQTSSIKESTLIP**PPPPF**EQTFPSLQPGASTLQR	17.4 ± 0.4
WASF2	FSEDNLPP**PPAEF**SYPVDNQRGSGLAGPKRSSVVSP	18.4 ± 0.3
RAX	GFGPPAQSLPASYTPPP**PPPPF**LNSPPLGPGLQPLA	32.3 ± 4.2
GRID2IP	PLSPPPP**PPLPF**HDAKPSSRSSDGSRGPAQALAKRL	61.0 ± 3.9
LSR	SEERRRPHKEEEEEAYYPP**APPPY**SETDSQASRERR	62.4 ± 0.9
CAP2	KSSEMNILIPQDGDYREFP**IPEQF**KTAWDGSKLITE	97.5 ± 18.0
NEDD4L	EVVDSNDSASQHQEELPPPP**LPPGW**EEKVDNLGRTY	100 ± 1.6
SH2B1	MNGAPSPEDGASPSSPPLPPP**PPPSW**REFCESHARP	112.6 ± 3.6
WIPF2	VEDFP**APEEY**KHFQRIYPSKTNRAARGAPPLPPILR	>150
GPR179	MGTRGAVM**PPPMW**GLLGCCFVCAWALGGPRPIRSLP	>150
ROCK2	ADAKE**IPRIF**QILYANEGESKKEQEFPVEPVGEKSN	>150
NFASC	MARQ**PPPPW**VHAAFLLCLLSLGGAIEIPMDPSIQNE	>150
DVL3	PLPHPGAAPWPMAFPYQYPP**PPHPY**NPHPGFPELGY	46.3 ± 3.7
$DVL3^{Y(-3)A}$	PLPHPGAAPWPMAFPYQAPP**PPHPY**NPHPGFPELGY	103.6 ± 4.3
$DVL3^{P(-2)A}$	PLPHPGAAPWPMAFPYQYAP**PPHPY**NPHPGFPELGY	145.1 ± 12.2
$DVL3^{P(-1)A}$	PLPHPGAAPWPMAFPYQYPA**PPHPY**NPHPGFPELGY	83.6 ± 8.9
$DVL3^{P0A}$	PLPHPGAAPWPMAFPYQYPP**APHPY**NPHPGFPELGY	121.3 ± 9.7
$DVL3^{P1A}$	PLPHPGAAPWPMAFPYQYPP**PAHPY**NPHPGFPELGY	>150
$DVL3^{Y4F}$	PLPHPGAAPWPMAFPYQYPP**PPHPF**NPHPGFPELGY	25.1 ± 0.7
$DVL3^{Y4W}$	PLPHPGAAPWPMAFPYQYPP**PPHPW**NPHPGFPELGY	63.3 ± 2.3

Table 1.2: Binding affinities between Homer1 and peptides derived from the T7-pep enrichment screen, measured by BLI. Values represent the average and standard deviation of 3 replicates.

motif.

An N-terminal Ena/VASP Motif Enhances Homer1 Binding

Approximately 97% of sequences matching the ^{0}[PIALV]Pxx[FWY]^{4} motif that were present in the T7-pep library failed to enrich, underscoring that the motif alone is insufficient for persistent binding, even in the avid conditions of the screen. To uncover elements that modulate motif affinity for Homer1, we visualized the screen results using pLogo, which scales residue height according to over- and underrepresentation relative to a background set of amino acid frequencies [41]. Weakly and strongly enriching sequences were aligned by their motifs and compared to a background of all Homer motif instances in the input library (Figure 1.2D, Figure S1.9B).

Notably, the logo of strongly enriching sequences unveiled a preference for the residues $^{-3}$[LF]PP^{-1} immediately N-terminal to the Homer motif ($^{-3}$**[LF]PP**[PIALV]Pxx[FWY]^{4}). These residues resemble the Ena/VASP motif, often encountered as the sequence $^{-3}$[FWYL]PPPP^{1}. We refer to the motif $^{-3}$[LF]PP[PIALV]Pxx[FWY]^{4} as the extended Homer motif. Compar-

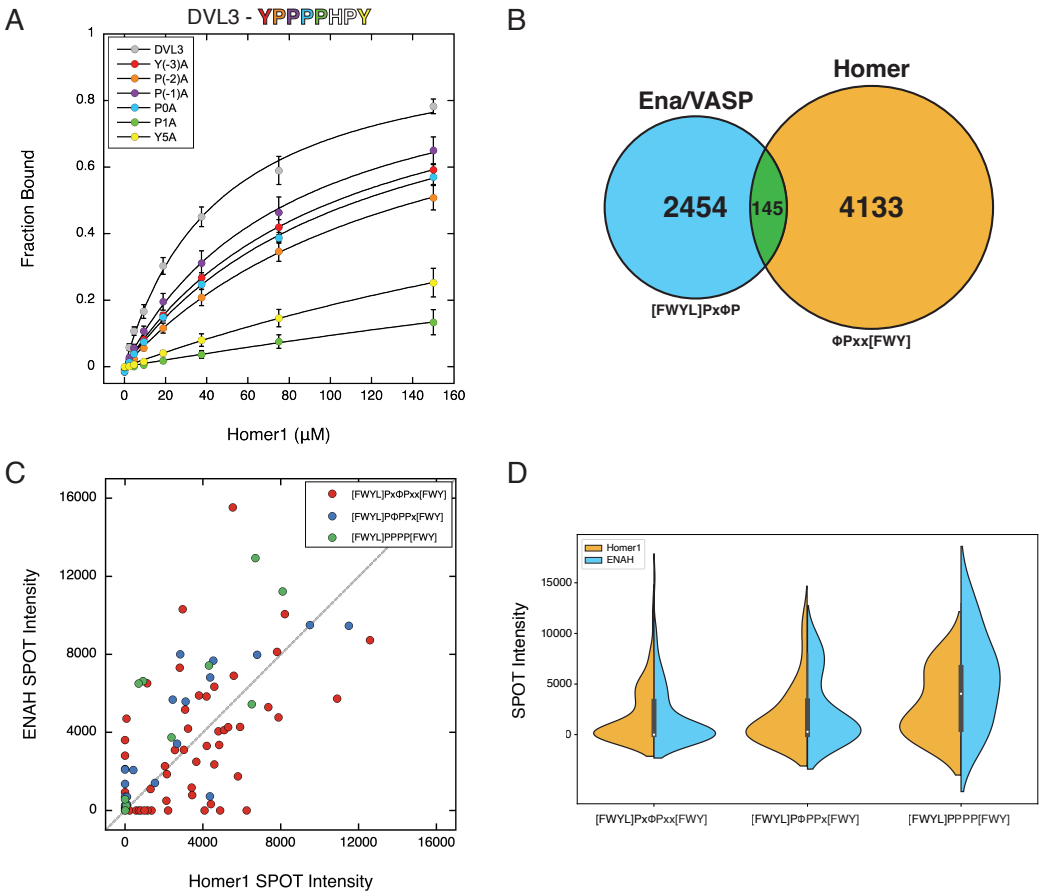


Figure 1.3: (A) Alanine mutations throughout the Homer1 motif and N-terminal Ena/VASP-like flank diminish binding to Homer1. (B) Bioinformatic search for Ena/VASP (blue), Homer (gold), and overlapping motifs (green) within disordered regions of the human proteome. (C,D) SPOT intensities for ENAH and Homer1 binding to overlapping motifs from the human proteome.

ison of structures of Ena/VASP (PDB: 1EVH) vs. Homer (PDB: 1DDV) EVH1 domains bound to motif-matching peptides show that the two terminal prolines of the Ena/VASP motif ($^{-3}$[FWYL]PP**PP**^{1}) bind to Ena/VASP in a pose similar to how the first two prolines of the (^{0}**PP**STF^{4}) motif bind to Homer (Figure 1.2D) [35], [36]. This structural alignment predicts that the N-terminal flanking residues that we note in the sequence logo may dock with Homer in a manner similar to the binding of a canonical Ena/VASP ligand to Ena/VASP EVH1, despite the absence of a well-formed P_{EV} pocket or the residue $Y16_{ENAH}$, which is used by Ena/VASP proteins for Pro^{-2} docking.

Segment polarity protein disheveled homology DVL3, a key player in the Wnt signaling pathway, is reported to bind to Homer3 in a yeast two-hybrid assay [42]. While DVL3 does not contain a canonical PPxxF motif, a peptide containing the non-canonical motif and an Ena/VASP-like N-terminal flank ($^{-3}$YPPPPHPY^{4}) enriched in our screen and bound with $K_D = 46.3 \pm 3.7$ μM. Alanine substitutions throughout this sequence were tolerated to

varying degrees (Table 1.2). Within the core motif, the substitution of Pro^1 or Tyr^4 with Ala resulted in peptides without appreciable binding to Homer1, while substitutions of Pro^0 caused nearly a 3-fold reduction in binding affinity (Figure 1.3A). Mutations within the Ena/VASP-like flank had a similar impact on overall binding affinity as Pro^0 substitution or the aromatic residue at position 4, highlighting the extent to which motif-external elements can influence binding. Interestingly, most of the BLI-validated sequences contained at least a partial match to the Ena/VASP-like flank, which was absent from non-binding sequences with non-canonical motifs. These results demonstrate that an Ena/VASP-like flank can enhance Homer1 binding and compensate for a suboptimal core Homer motif.

The Homer1 and ENAH EVH1 Domains have Incompletely Diverged

The identification of an N-terminal Ena/VASP motif as a Homer affinity-enhancing flanking element is perplexing. Homer proteins are highly expressed in neurons, where Ena/VASP proteins play important roles in neuronal migration and axon guidance [22]. If Ena/VASP EVH1 domains are similarly capable of binding these overlapping motifs, that would imply the potential for competitive binding between the two protein families that have different functions.

The intrinsically disordered regions (IUPRed > 0.45) of the human proteome contain 103 instances of a Homer motif flanked by a potentially affinity-enhancing Ena/VASP-like sequence, based on sequences enriched in our screen: $^{-3}$**[FWYL]Px**ΦPxx[FWY]^4 (Figure 1.3B, Supplementary Table 2) [43]. To determine whether these peptides that match the core motifs of both domains bind to both Ena/VASP and Homer EVH1 domains, we tested the interaction of these peptides with tetrameric EVH1 constructs using a peptide SPOT array (Figure 1.3C, Figure S1.10). Many peptides bound to both domains and the distributions of intensities for all peptides were remarkably similar for Homer1 and ENAH (Figure 1.3D). BLI measurements confirmed that many peptides bind to both ENAH and Homer1 with similar affinities (Table 1.3). A further 42 Homer motif-matching peptides contain an N-terminally overlapping Ena/VASP motif in a frame that does not structurally align to the Ena/VASP binding mode, that is, a match to $^{-2}$**[FWYL]P**ΦPPx[FWY]^4 or ^{1}**[FWYL]**PPPP^4 (Figure 1.3B). While many peptides within these classes were similarly capable of binding Homer1, these peptides showed a slight bias toward binding ENAH based on their distribution of SPOT intensities (Figure 1.3C, Figure 1.3D).

A Single Motif-Flanking Residue Modulates Binding to Homer1

Of the 140 peptides that contain overlapping Ena/VASP and Homer motifs that we tested, 58 bound detectably to Homer1 and 55 to ENAH. Most peptides were not highly specific for either domain, with 46 peptides binding to both. Several overlapping peptides exhibited modest selectivity in the SPOT array assay, such as FAT1 (FPPPPEDF) and ASPP2 (YPPPPY), which bound to ENAH 2 to 3-fold better than Homer1. FAT1 also demonstrated a near 10-fold preference for ENAH ($K_D = 13.7 \pm 0.3$ μM) over Homer1 ($K_D = 98.2 \pm 7.1$ μM) via BLI, corroborating the differences in SPOT intensity (Table 1.3). The weak affinity of Homer1 for FAT1 is surprising given the presence of both a canonical PPxxF motif and an Ena-VASP-like flank, and we hypothesized that features in the motif-surrounding sequence

Name	**Sequence**	$\mathbf{K_D}^{\mathbf{ENAH}}$ **(μM)**	$\mathbf{K_D}^{\mathbf{Homer1}}$ **(μM)**
ABI3	GDELG**LPPPPPGF**GPDEPSW	25.7 ± 0.6	3.9 ± 0.1
ABI3^{G5P}	GDELG**LPPPPPGF**PPDEPSW	40.2 ± 5.7	141.8 ± 9.9
DBN1	LNFDE**LPEPPATF**CDPEEVE	57.3 ± 0.9	12.2 ± 0.8
DBN1^{C5P}	LNFDE**LPEPPATF**PDPEEVE	64.9 ± 6.3	>150
WASF2	FSEDN**LPPPPAEF**SYPVDNQ	17.9 ± 0.4	13.4 ± 0.7
APBB1IP (2)	LDDPE**LPPPPPDF**MEPPPDF	8.2 ± 0.1	24.6 ± 1.4*
WBP2	PPGYP**YPPPPPEF**YPGPPMM	11.8 ± 1.8*	16.6 ± 2.2*
CEBPB	PACLP**LPPPPPAF**KSMEVAN	57.2 ± 8.8	36.9 ± 3.6
IQSEC3	THQPP**LPPPPPPY**NHPHQFC	57.1 ± 6.1	37.7 ± 4.0
FAT1	DIESD**FPPPPEDF**PAADELP	13.7 ± 0.3	98.2 ± 7.1
FAT1^{P5G}	DIESD**FPPPPEDF**GAADELP	16.8 ± 1.1	15.4 ± 0.6
APBB1IP (1)	TGGGG**LPAPPDDF**LPPPPPP	25.9 ± 0.5*	>150
PCDH15	PPSIP**LPLPPPTF**FPLSVST	107.8 ± 28.4	>150
DSCAM	GTLQI**FPFPPSSF**STLIHDN	>150	55.3 ± 4.0
RARA	GHLNG**YPVPPYAF**FFPPMLG	>150	98.9 ± 21.5
PRRC2B	HVDSV**LPVPPIEF**GVSPKDS	>150	103.5 ± 17.4
KANK4	PNHLP**LPGPPFSF**QNVLVVL	>150	>150
FLNC	FIAHI**LPAPPDCF**PDKVKAF	>150	>150
SH3BGRL	RPATG**YPLPPQIF**NESQYRG	>150	>150
DVL1	GYPYQ**YPGPPPCF**PPAYQDP	>150	>150

Table 1.3: Binding affinities between bioinformatically-identified peptides containing the extended Homer motif and Homer1 or ENAH. Measurements marked with an asterisk (*) are peptides for which binding was maintained even when the motif was knocked out via SPOT array. Values represent the average and standard deviation of 3 replicates.

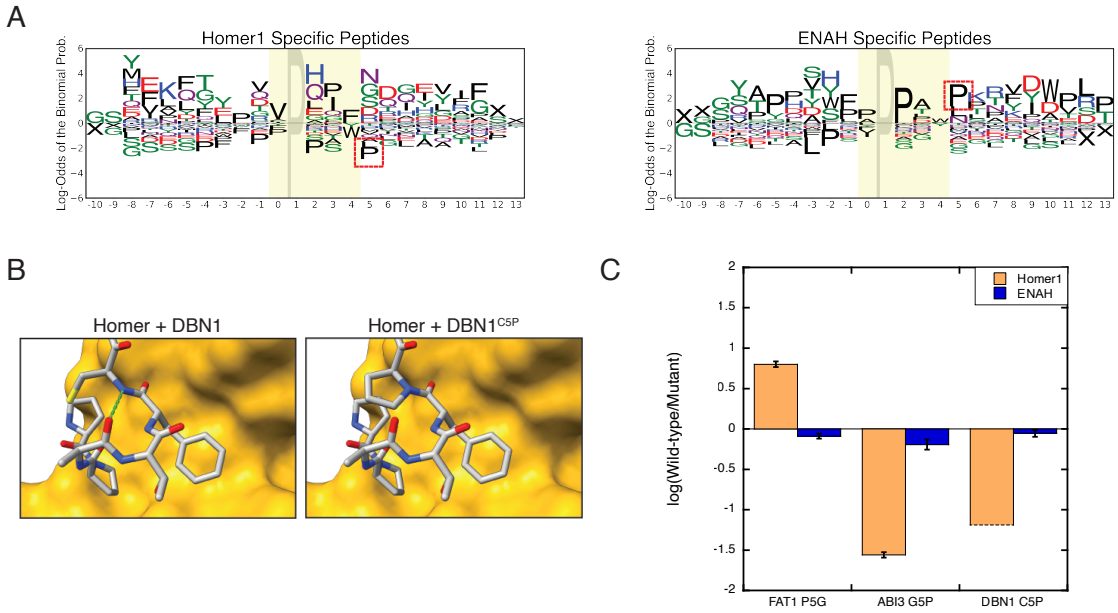


Figure 1.4: (A) Logos for bioinformatically-identified peptides containing the extended Homer motif that preferentially bind to Homer1 or ENAH by SPOT array. Sequences are aligned by their Homer1 motif. The red box indicates the position of the Pro[5] Homer1 negative elements (B) A Structure of the Homer2 EVH1 domain bound to a DBN1 demonstrates that the residue in position 5, when not Pro, can hydrogen bond (green) with the residue in position 3 [29]. (C) The fold-change binding affinity of ENAH and Homer1 for peptides in which a Pro has been removed from ($FAT1^{P5G}$) or introduced to ($ABI3^{G5P}$, $DBN1^{C5P}$) the position following the Homer1 motif.

may disfavor the FAT1-Homer1 interaction.

To identify potential negative design elements, we created logos for the sequences that bound specifically to Homer1 or ENAH. This revealed that Homer1, but not ENAH, disfavored Pro at position 5, immediately C-terminal to the core Homer motif (Figure 1.4A). Notably, both FAT1 and ASPP2 encode a Pro in this position. The crystal structure of Homer2 bound to a peptide from DBN1 (PDB: 5ZZ9) suggests that this residue, when not Pro, can hydrogen bond with the residue in position 3 using its backbone amide hydrogen (Figure 1.4B) [29]. This interaction may orient the aromatic residue in position 4 for docking within P_H, explaining the specificity of this effect for Homer1. Indeed, removing this element in FAT1 (P5G) yielded a peptide that bound six-fold tighter to Homer1, but this mutation had almost no effect on ENAH binding (Figure 1.4C). Similarly, introducing Pro at this position in peptides from ABI3 (G5P) and DBN1 (C5P) substantially decreased the affinity for Homer1 with minimal impact on ENAH binding. The role of proline at position 5 is an elegant example of a structural strategy that can increase the binding specificity of peptide segments that match both the Ena/VASP and Homer binding motifs.

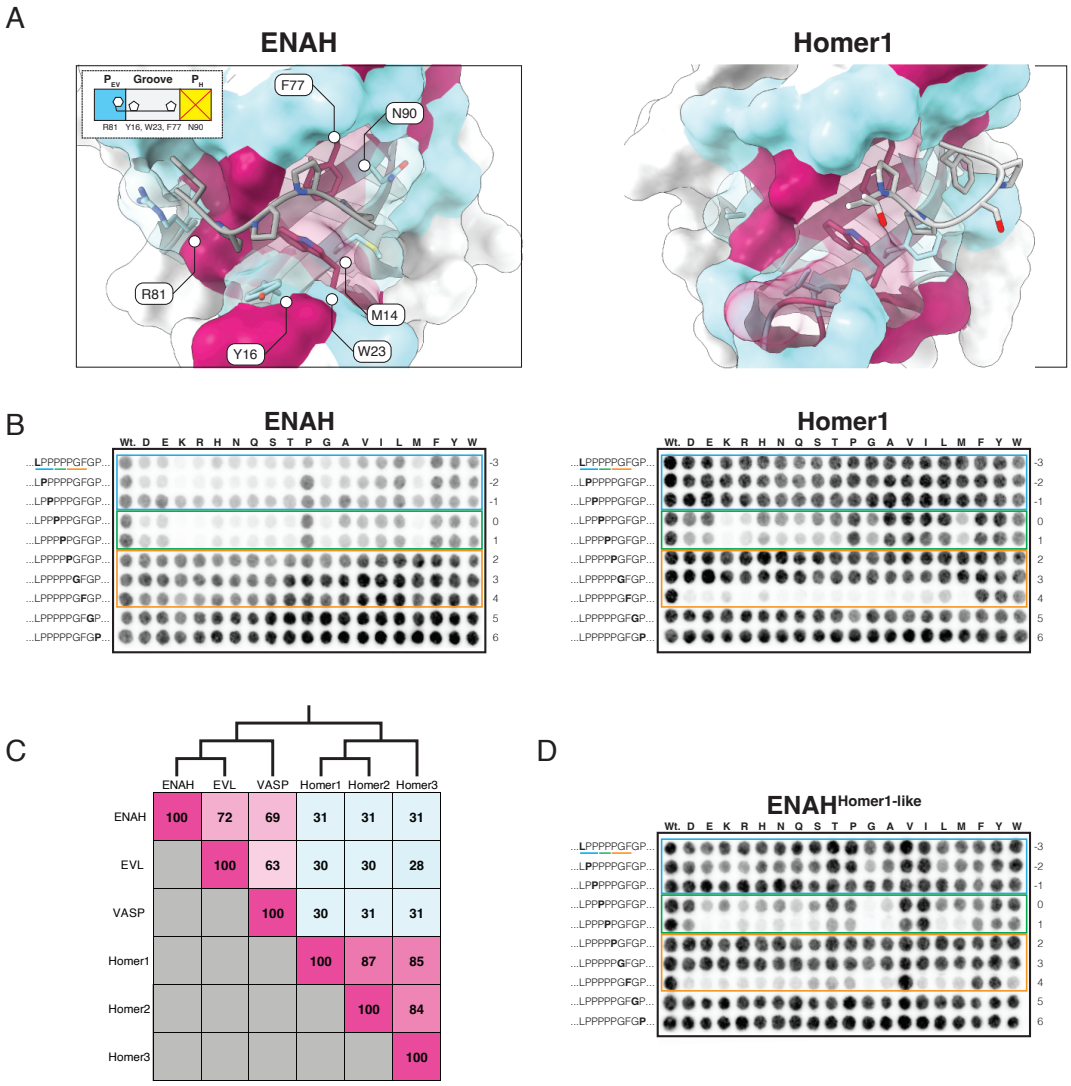


Figure 1.5: (A) ENAH (EVH1) and Homer1 (1DDV) in complex with their representative ligands. Residues within 7Åof the binding site are colored by whether they are shared (maroon) or different (cyan) between these paralogous domains [35], [36]. (B,D) Single substitution SPOT arrays of the overlapping motif in ABI3. In each row, the substituted residue is marked in bold text. (C) Percentage identity matrix of the human Ena/VASP and Homer family EVH1 domains. The tree corresponds to the pairwise sequence identity between domains.

Transformation of EVH1 Binding Preferences

Co-crystal structures of Ena/VASP and Homer EVH1 domains bound to representative ligands have revealed residues within the binding sites that, through their contacts, appear crucial for defining the reported motifs. To identify the extent to which these residues are responsible for conferring these motif preferences, we attempted to convert the binding preferences of Homer1 and ENAH by swapping four binding site residues in ENAH, representing 5% of the difference between these domains ($ENAH^{Homer1\text{-}like}$; M14F/Y16I/R81A/N90G) (Figure 1.5A). As demonstrated by a single-substitution SPOT array of the promiscuous peptide from ABI3, ENAH was most sensitive to substitutions within the Ena/VASP motif ($^{-3}LPPPP^{1}$), whereas Homer1 was sensitive to substitutions within the Homer motif ($^{0}PPPGF^{4}$) (Figure 1.5B). The tolerance for deviations from the canonical motifs for each domain and the inability to resolve the expected contribution of N-terminal residues to Homer1 binding is likely due to the use of a high-affinity ligand in a multivalent context. Notably, $ENAH^{Homer1\text{-}like}$ exhibited Homer1-like binding properties, both by the decreased sensitivity to substitutions in the Ena/VASP motif and the acquired requirement for a large hydrophobic residue at position 4 (Figure 1.5D). Unlike Homer1, $ENAH^{Homer1\text{-}like}$ had a strong preference for valine at position 4, implying that additional differences between these domains are responsible for fine-tuning the motif. Nevertheless, the ability to recapitulate Homer1 preferences despite markedly different background sequences verifies that these residues play a pivotal role in defining EVH1 motifs.

An Ortholog from *Dictyostelium discoideum* Sheds Light on the Evolution of EVH1 Binding Preferences

Ena/VASP and Homer are highly conserved in metazoans, and both families are present in choanoflagellates, the single-cell organisms most closely related to metazoans [14]. Several phylogenetically remote amoebas lack Homer proteins but encode a singular Ena/VASP protein, as defined by the presence of both EVH1 and EVH2 domains. We hypothesized that these orthologs, seemingly from organisms that pre-date the gene duplication event that gave rise to Homer, may offer insight into the evolution of binding specificity within this paralogous family.

In the soil-dwelling amoeba *Dictyostelium discoideum*, the Ena/VASP protein, ddVASP, is required for filopodia formation, chemotaxis, and macroendocytosis [44], [45]. As assessed by sequence identity, the *D. discoideum* EVH1 domain (which we refer to here as ddEVH1) is similarly diverged from the human Ena/VASP (35%) and Homer (31%) EVH1 domains (Figure 1.6A). An AlphaFold2 model of ddEVH1 reveals that the region corresponding to the peptide binding site in Ena/VASP and Homer exhibits a blend of features from both human protein families (Figure 1.6B) [46], [47]. Like Homer, but not Ena/VASP, ddEVH1 lacks a well-formed P_{EV} ($T80_{ddEVH1}$ and $A78_{Homer1}$ vs. $R81_{ENAH}$). Like Ena/VASP, but not Homer, ddEVH1 has an occluded P_H ($N89_{ddEVH1}$ and $N90_{ENAH}$ vs. $G89_{Homer1}$). Additionally, whereas Ena/VASP and Homer EVH1 domains each contain three aromatic residues in the proline binding groove, ddEVH1 contains a tetrad consisting of both the unique Ena/VASP ($Y16_{ddEVH1}$ and $Y16_{ENAH}$ vs. $I16_{Homer1}$) and Homer ($F14_{ddEVH1}$ and $F14_{Homer1}$ vs. $M14_{ENAH}$) aromatic residues. Sequence analysis of the Ena/VASP proteins from other amoebas demon-

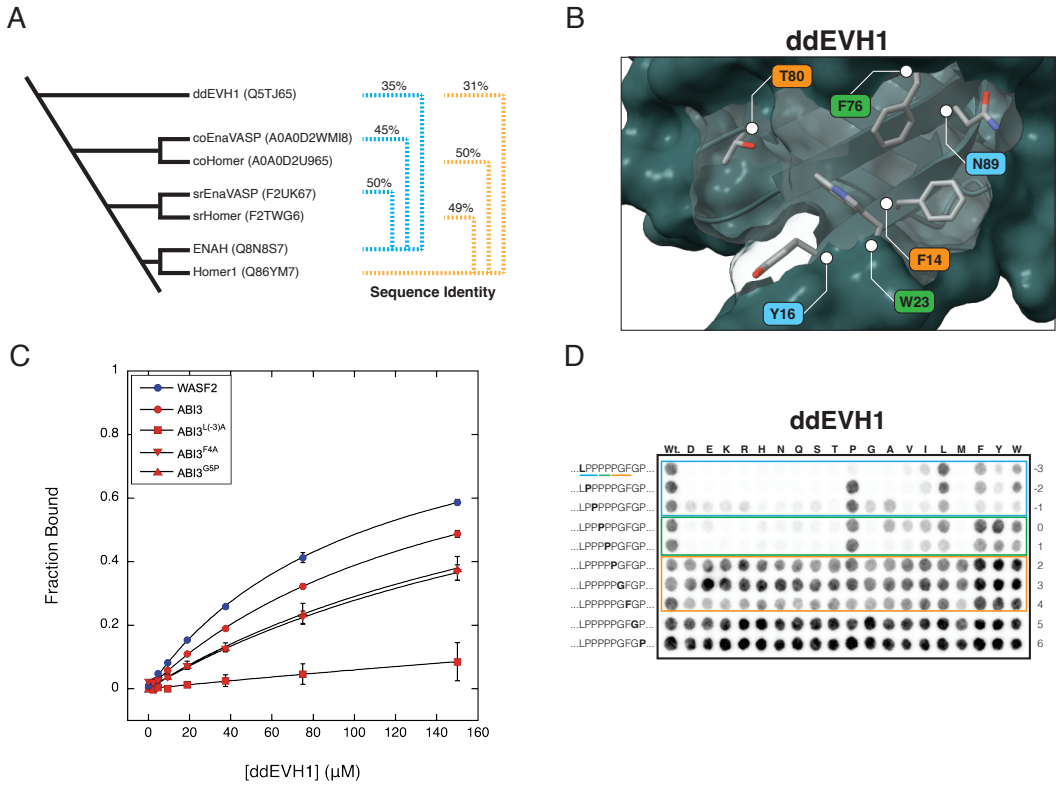


Figure 1.6: (A) Dendrogram representing the positioning of diverse organisms containing Ena/VASP and Homer proteins along the tree of life. The sequence identity for Homer EVH1 domains (and ddEVH1) is shown relative to human Homer1 (yellow), and the sequence identity for Ena/VASP EVH1 domains is shown relative to human ENAH (blue). (B) Alphafold2 structural model of ddEVH1 (Q5TJ65). The color of each label reflects whether the residue is similar to the corresponding residue in Homer (yellow), Ena/VASP (blue), or both (green). (C) BLI measurements for ddEVH1 and promiscuous peptides from WASF2 and ABI3. (D) Single substitution SPOT arrays of the overlapping motif in ABI3. In each row, the substituted residue is marked in bold text.

strates that this binding site architecture is highly conserved (Figure S1.12).

We tested ddEVH1 for binding to several human peptides identified in this study (Figure 1.6C). Like ENAH and Homer1, ddEVH1 bound to peptides from ABI3 and WASF2, though the affinity for each was weak ($K_D > 100\mu M$). Alanine substitution of the N-terminal Leu^{-3} ($^{-3}$**L**PPPPPGF^{-4}) in ABI3 (ABI3$^{L(-3)A}$) abrogated ddEVH1 binding, while substitution of the C-terminal Phe^{4} (ABI3^{F4A}), or replacement of the motif-trailing Gly^{5} with Pro (ABI3^{G5P}), had a modest effect on binding. These results suggest that ddEVH1 favors a sequence similar to the human Ena/VASP motif, despite, like Homer1, containing differences in key binding site residues. Nevertheless, ddEVH1 bound poorly to the ENAH-binding peptides from DBN1 and FAT1, indicating that there are distinctions between the binding preferences of the *D. discoideum* and human domains (data not shown).

In agreement with our BLI measurements, the single-substitution SPOT array of ABI3 demonstrated that ddEVH1 exhibits similar binding preferences to human Ena/VASP proteins (Figure 1.6D). In contrast to ENAH, ddEVH1 strongly preferred Leu over an aromatic residue at position -3, as reflected by relative spot intensity. This is likely caused by the absence of cation-pi interactions formed in ENAH by $R81_{ENAH}$ and may explain the weak binding of ddEVH1 to FAT1 ($^{-3}$**F**PPPPEDF^{4}). The substitution array also revealed that ddEVH1 has a stronger preference than ENAH for proline residues at positions -2 through 1. This may indicate that the proline binding groove, with its extra aromatic residue, compensates for a poorly formed P_{EV} pocket. DBN1, which does not bind to ddEVH1, lacks a proline at the -1 position ($^{-3}$LP**E**PPATF^{4}). Lastly, ddEVH1 exhibits a weak sensitivity to the replacement of Phe^{4} with a non-aromatic residue, despite the occlusion of the P_H pocket by $N89_{ddEVH1}$. This is consistent with our finding that alanine substitution of this residue (ABI3^{F4A}) had a modest effect on binding, as measured by BLI (Figure 1.6C).

Our biochemical characterization of ddEVH1 suggests a preference for the motif [LF]PPPP, with less tolerance for deviation within this sequence than is observed in human Ena/VASP orthologs. 113 Dictyostelium proteins contain a [FL]PPPP motif; of these 22 have human orthologs, with several being annotated as participating in actin regulation or neuronal processes (Supplementary Table 3). Furthermore, the ABI protein in *D. discoideum*, a known binding partner of ddEVH1, contains 6 copies of the sequence [LF]PPPP (UniProt ID: Q55FT9) [48]. This, together with our observation that peptides matching this motif bind ddEVH1 an order of magnitude weaker than ENAH and Homer1, suggests that multivalency may be crucial for obtaining physiologically relevant binding affinities in primitive Ena/VASP systems.

Incomplete Divergence Does Not Reflect Evolutionary Inaccessibility of Complete Divergence

The incomplete divergence of the Ena/VASP and Homer binding profiles may imply that evolutionary paths to distinct interactions are long (requiring many mutations) or rare (requiring very specific combinations of mutations), such that fully differentiated binding properties are unlikely to arise even over long time periods. This could be the case, for example, if most binding-site mutations that lead to diverged binding compromise the structural integrity of the domain. The advent of multicellularity provided a new mechanism by which biological systems can spatiotemporally restrict the co-localization of proteins, segregating proteins

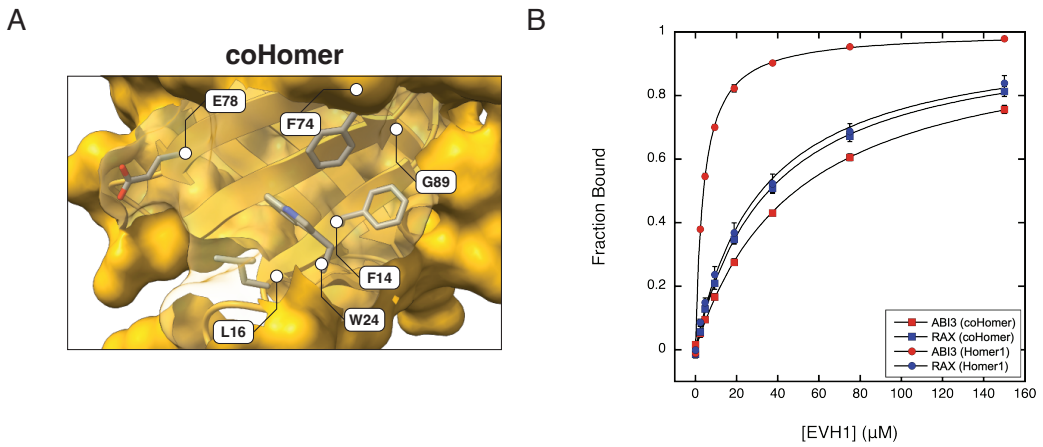


Figure 1.7: (A) Alphafold2 structural model of coHomer (UniProt: A0A0D2U965). (B) BLI binding curves comparing Homer1 (circle) and coHomer (square) binding to peptides from ABI3 and RAX.

into distinct niches without the need for divergence at the level of molecular recognition. Given the potential for deleterious competition between Ena/VASP and Homer, we hypothesized that pre-metazoan unicellular organisms, which lack this regulatory layer, may face greater pressure to diverge their interface residues.

Our search for orthologs in pre-metazoans identified both an Ena/VASP and Homer protein in the single-celled eukaryote *Capsaspora owczarzaki*, a member of the Filasteria clade, the sister group to choanoflagellates and metazoans (Figure 1.6A). The EVH1 domain in the Homer protein, which we call coHomer, shares 50% sequence identity to Homer1. While the architecture of the proline-groove and P_H is near-identical to Homer1, coHomer uniquely incorporates an acidic residue (E78) within the P_{EV} pocket (Figure 1.7A). This mutation may disfavor the engagement of peptides that bind well to Ena/VASP by disrupting the docking of a hydrophobic residue in P_{EV} or by repelling peptides containing acidic residues often found upstream of the motif in Ena/VASP binders [32].

We compared the binding of multiple peptides to both Homer1 and coHomer EVH1 domains using BLI (Figure 1.7B). Peptides that included acidic residues within a 5-residue window N-terminal to the motif (ABI3, DBN1, and FAT1^{P5G}) bound to coHomer with at least a 10-fold reduction in affinity compared to Homer1 (Table 1.4). This effect was more pronounced for DBN1 and FAT1^{P5G}, which contain acidic residues immediately N-terminal to the extended motif. DVL3, which contains an overlapping motif but lacks N-terminal acidic residues, also exhibited a reduction in binding affinity. RAX, which lacks the Ena/VASP P_{EV} docking residue ($^{-3}$**[FWYL]**PxΦP^{1}), bound equivalently to both Homer1 and coHomer. This trend suggests that coHomer disfavors binding to Ena/VASP-binding peptides largely through the presence of N-terminal acidic residues, but also through the inclusion of a P_{EV} docking hydrophobic residue. Introducing this glutamate into the Homer1 EVH1 domain (Homer1^{A78E}) alone was insufficient to evoke a similar change in affinity, and may be due to the rigidity of the glutamate side-chain in coHomer, which is flanked by L85 and V80 (Figure S1.13A,B). While multiple mutations may be required to impart these preferences, our results demonstrate that further diverged molecular recognition profiles of Homer and

Name	Sequence	K_D (μM)
DBN1	LNFDE**LPEPPATF**CDPEEVE	>150
FAT1^{P5G}	DIESD**FPPPPEDF**GAADELP	>150
ABI3	GDELG**LPPPPPGF**GPDEPSW	49.0 ± 2.7
RAX	GFGPPAQSLPASYTPPP**PPPPF**LNSPPLGPGLQPLA	35.8 ± 3.1
DVL3	PLPHPGAAPWPMAFPYQ**YPPPPHPY**NPHPGFPELGY	>150

Table 1.4: coHomer binding affinities measured by BLI. Values represent the average and standard deviation of 3 replicates.

Ena/VASP are evolutionary accessible, but have not been selected for in metazoan lineages.

1.0.4 Discussion

Novel Insight into Homer1 Binding Preferences

Homer, a constitutive tetramer, carries out its biological function by crosslinking (or, in the case of immediate early monomeric Homer isoforms, disassembling) complexes of partners using its EVH1 domain [23]. Identifying these binding partners is key to understanding the biological role of Homer proteins. Because of the relatively weak affinity (low- to mid-micromolar) of EVH1-mediated binding and the dynamic nature of the postsynaptic density, many of these interactions are likely transient and only populated in response to specific stimuli. For this reason, Homer-mediated interactions may evade standard techniques for identifying native protein-protein interactions, and this may explain the low recovery of annotated binding partners in various Homer co-IP mass spectrometry studies [49], [50]. Limitations of these experiments also stem from their tendency to focus on a single tissue type, despite the wider expression of Homer [26].

As such, most Homer binding partners have been identified in low-throughput, candidate-driven studies by virtue of the inclusions of a canonical PPxxF motif, first identified in family members of the metabotropic glutamate receptors [24], [28]. This has led to the circular reinforcement of PPxxF as the Homer binding motif, potentially to the exclusion of partners that lack this sequence. Several deviations from this motif have been noted in the literature, though this has not led to a broader exploration of Homer binding preferences [51], [52].

The high-throughput screening of peptide libraries is a powerful technique for mapping SBD binding landscapes in an unbiased way [53]. Proteome-derived libraries such as Pro-PD and T7-pep allow for the identification of candidate binding partners of potential biological relevance [33], [37], [38]. Even for library hits that do not interact in a cell, these binders nonetheless offer insight into the diverse ways that an SBD can engage its ligands. In the case of the T7-pep library, the 36-mer formatting permits the discovery of binding features outside of the core motif, as uncovered for the interaction between the ENAH EVH1 domain and a 17-residue sequence from PCARE [54].

In this study, we screened T7-pep to identify Homer1 binders and uncover features that support binding. Our results demonstrate that the reported core Homer motif of PPxxF is overly restrictive and that sequences matching a more general motif of [PIALV]Pxx[FWY] can bind to the Homer1 EVH1 domain. Additionally, we identified specific examples of

how the sequence context flanking the core motif, such as a leading Ena/VASP motif-like segment or a trailing proline, can alter the affinity and specificity of this interaction. These findings build on and contextualize the previous discoveries that many Homer ligands contain an N-terminal, affinity-enhancing Leu, and that mutation of an N-terminal FPP diminishes the binding of Homer3 to FAT1 [29], [55]. By combining our findings and reported Homer preferences from the literature, we define an updated Homer motif [fwyl]pp[PIALV]P..[FWY][^p][dn], where uppercase letters reflect obligate positions and lowercase letters indicate positive and negative elements that can alter the affinity of this interaction [29], [56].

The Evolution of Binding Preferences in EVH1 Domains

Homer and Ena/VASP share a common ancestor, and these distinct families are present in the pre-metazoan *Salpingoeca Rosetta*, indicating that the family-expanding duplication event predates the emergence of metazoans some 600 million years ago (Figure 1.6A) [13]. Many amoebas, further remote from metazoans along the tree of life, encode a singular Ena/VASP protein but lack Homer orthologs. The absence of Homer further back in the tree suggests that these organisms predate the duplication event that gave rise to Homer. Thus, we hypothesized that these EVH1 domains may resemble the ancestral domain's binding preferences.

The EVH1 domains from amoebas share near-equal sequence identity with domains from both human Ena/VASP and Homer. Furthermore, the binding site architecture of these domains blends features from both human families. While we cannot rule out that the binding preferences of these EVH1 domains may have drifted over time, proteins at the center of interaction networks are buffered against drift due to the extensive co-evolution required to maintain their interactomes [57]–[60]. In support of this, even those EVH1 domains from amoebas with low sequence identity share the binding site residues we identified as key for defining motif preferences, suggesting the persistence of a common binding mode (Supplementary Table 3). Together, these observations beget a model whereby subfunctionalization of the ancestral binding site birthed distinct Ena/VASP and Homer families.

To explore this model, we characterized the binding preferences of an EVH1 domain from the Ena/VASP ortholog in *Dictyostelium discoideum*. We observed that this domain, ddEVH1, binds weakly to the motif [FL]PPPP and exhibits a slight preference for the C-terminal aromatic residue found in the Homer motif. Seemingly, selective reinforcement of pre-existing preferences for hydrophobic residues flanking the prolines, as opposed to *de novo* emergence, established the higher-affinity motifs that are recognized by metazoan EVH1 domains (Figure 1.8). As evidenced by the affinity of Homer1 for sequences containing an Ena/VASP-like N-terminal flank, Homer did not lose its preference for the ancestral motif, which exists independent of a well-formed P_{EV}. Furthermore, negative design elements did not arise to eliminate binding overlap between the human Ena/VASP and Homer families.

Ena/VASP and Homer EVH1 domains may not have diverged further because changes that enhance selectivity compromise binding. However, the sole Homer protein in the pre-metazoan *Capsaspora owczarzaki* robustly discriminates against sequences containing an Ena/VASP moiety and affinity-enhancing acidic residues, demonstrating that further diverged binding profiles are accessible. Why then did the Homer preference for the Ena/-

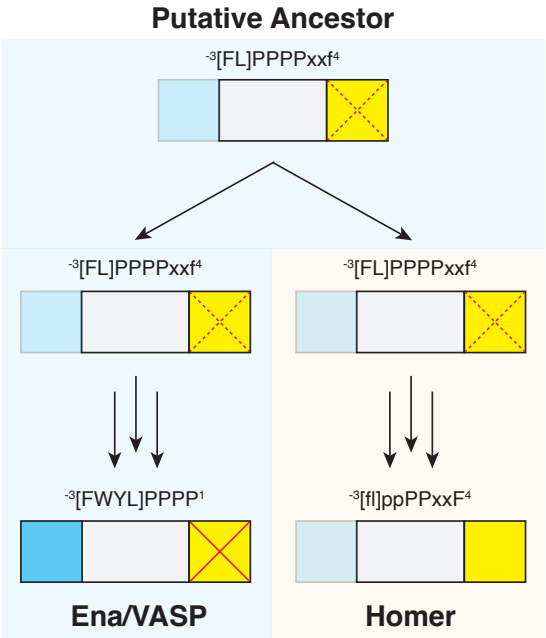


Figure 1.8: An evolutionary model of EVH1 divergence. The putative ancestor of Ena/VASP and Homer likely bound to a motif akin to ddEVH1. Incomplete subfunctionalization of the core binding site resulted in a Homer EVH1 domain that retained its preference for an Ena/VASP motif.

VASP moiety persist in metazoan lineages? We posit that the retention of this preference in Homer is either functionally beneficial or neutral with respect to fitness but preserved due to constraints on drift that arise from high interaction connectivity.

Rationalizing Overlap in the Ena/VASP and Homer Recognition Landscape

SLiMs play pivotal roles in cellular physiology, and mutations that disrupt these sequences can reduce fitness and lead to human disease [4], [61]. Fitness can also be compromised by competition between proteins for binding. To avoid this, paralogs develop specificity niches, either by diverging their molecular recognition profiles or by spatiotemporal compartmentalization [62]–[64]. Yeast SH3 domains exhibit limited specificity *in vitro* [65], [66]. Nevertheless, a peptide from the protein Pbs2 is able to bind to the Sho1 SH3 domain with high specificity, suggesting that their molecular recognition profiles have diverged far enough to permit the encoding of specific interactions [67]. On the other hand, a peptide from Pex14 binds promiscuously to a number of SH3 domains, but it is quarantined with its cognate SH3 domain (Pex13) in the peroxisome. Similarly, the Ena/VASP and Homer EVH1 domains have diverged their recognition profiles sufficiently for specific, non-promiscuous ligands to exist. Perhaps like the Pex14-Pex13 interaction, promiscuous peptides that bind to both Ena/VASP and Homer are spatiotemporally restricted from encountering both simultaneously.

Alternatively, these peptides may convey a beneficial function. The intricate landscape of cellular processes is dependent on conditional interactions, and the mutual exclusivity conferred by overlapping SLiM binding sites presents a strategic mechanism that can orchestrate these precise regulatory events [68], [69]. One way of achieving this type of regulation is through direct competition for binding sites, such as that between Homer3 and calcineurin for binding to NFATc2 [30]. Competitive binding between constitutively expressed tetrameric Homer and the immediate early monomeric isoform is central to Homer function [26], [27]. Ena/VASP proteins, which lack a monomeric isoform, may be able to hijack this mechanism in cases where their binding partner has evolved to overlap with a Homer motif.

Several proteins that contain dual-binding peptides stand out as candidate functional partners for both Ena/VASP and Homer (Figure S1.11). Among these are ABI3 and WASF2, two members of the pentameric WAVE complex that regulates Arp2/3 mediated actin polymerization [70]. The WAVE complex interacts directly with Ena/VASP proteins to drive enhanced actin assembly, and this association is conserved as far back as *D. discoideum* [48], [71]. Given the role of Homer in synaptic plasticity - an actin remodeling event - a direct connection between Homer and actin polymerization machinery is conceivable.

Circumstantial evidence suggests that Homer proteins may also interact with the WAVE complex. In neurons, the WAVE complex plays a crucial role in orchestrating the development of diverse actin-filament protrusions [72]. ABI3 has specifically been shown to regulate dendritic spine morphology and synapse formation, and the peptide from ABI3 is the tightest Homer1 binder identified in this study ($K_D = 3.9 \pm 0.1\ \mu M$) [73]. Through its interaction with various proteins, Homer can influence both the emergence and maturation of dendritic spines [74], [75]. Homer3 knockdown in neutrophils impairs the recruitment of the WAVE2 complex to the leading edge of motile cells [76]. In the cerebral cortex of mice, Homer1 dissociates from a complex containing the Arp2/3 members ARPC2, 4, and 5 in an activity-

dependant manner [49].

The small number of interface residues that constitute SLiM-mediated interactions makes them highly amenable to regulation by post-translational modification. Motif phosphorylation is a well-documented switch for SLiM binding (e.g phosphorylation of the NPxY motif in the integrin β3 subunit shifts binding from talin to Dok1) [68], [77]. Phosphorylation can also modulate the affinity of a motif, and phosphorylation of the non-interface Ser and Thr residues in the mGlur5 Homer binding peptide (**T**PP**S**PF) by Cdk5 increases affinity for Homer1 [78].

Our study uncovered that Homer1 can bind to sequences containing Tyr^4 within the core motif, but is intolerant to acidic residues and presumably $pTyr^4$ at this position. In the context of an overlapping motif, this modification may even enhance binding to Ena/VASP, given the preference for motif-flanking negative charges [32]. Of the 46 dual-binding peptides identified in this study, 20 contain Tyr^4 and 5 of these residues are annotated as phosphorylation sites [79]. One of these proteins, the non-receptor kinase Activated CDC42 kinase 1 (TNK2) has a documented role in neurite migration, and phosphorylation of this Tyr plays a role in glioma oncogenesis [80], [81].

In summary, our results suggest that the Homer interaction landscape remains only partially explored, and we present new motif definitions and candidate binding partners to facilitate the discovery of novel interactions. Furthermore, these findings provide a case study into the evolution of molecular specificity in SLiM binding domains by documenting the incomplete divergence within this paralogous family of domains. We speculate that this observation suggests novel ways in which Homer may have evolved to regulate molecular decision-making at the post-synaptic density, and may have broader ramifications on the design of family-specific EVH1 inhibitors.

Methods

Protein Constructs and Purification (BLI Constructs)

EVH1 domains used for BLI experiments were cloned into a pMCSG7 vector for recombinant expression in Escherichia coli. The plasmid was transformed into Rosetta2(DE3) cells (Novagen) and grown in 1L Terrific Broth (TB) media with 100 μg/ml ampicillin. Cells were grown while shaking at 37 °C until an OD600 of 0.5-0.6, upon which the cultures were placed on ice for 20 min. Following this, protein expression was induced with 0.5mM isopropylthio-β-galactoside (IPTG) and the induced cells were grown by shaking at 18C overnight. Cultures were harvested by centrifugation, and the resulting pellets were resuspended in Lysis buffer (20mM HEPES pH7.6, 500mM NaCl, 20mM Imidazole) at a concentration of 0.2g/mL before being frozen at -80 °C.

The next day, the pellets were thawed and supplemented with 0.2mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor. The cells were lysed by sonication, centrifuged, and the resulting supernatant was incubated with 2mL Ni-NTA resin (washed in Lysis Buffer) while rotating at 4 °C. After 1 hour, the resin was washed 2x with 20mL Lysis buffer. The protein was eluted from the Ni-NTA resin with 8mL Elution Buffer (20mM HEPES pH7.6, 500mM NaCl, 300mM Imidazole). TEV protease was added to the sample at an approximate concentration of 1mg TEV:50mg EVH1. The sample was dialyzed overnight at 4 °C in 1L

of TEV Buffer (50mM HEPES pH7.5, 300mM NaCl, 0.5mM EDTA, 2mM DTT).

The completeness of the TEV reaction was assessed by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS–PAGE). Next, the sample was applied to 2mL Ni-NTA resin and washed with 8mL Lysis Buffer. The flow-through and wash were pooled, filtered, and loaded onto an S75 26/60 sizing column equilibrated in Gel Filtration buffer (20mM HEPES pH7.6, 150mM NaCl, 1mM DTT). Fractions were collected and their purity was assessed via SDS-PAGE. Pure samples were pooled and concentrated prior to being flash-frozen and stored at -80 °C.

Protein Constructs and Purification (Surface Display Constructs)

EVH1 domains used for bacterial cell surface display and SPOT arrays were cloned into a pDW363 vector for recombinant expression in Escherichia coli. The plasmid was transformed into Rosetta2(DE3) cells (Novagen) and grown in 1L Terrific Broth (TB) media with 100 μg/ml ampicillin and 0.05mM D-(+)-biotin for in vivo biotinylation. Protein expression was carried out as above.

Pellets were thawed and supplemented with 0.2mM PMSF prior to lysis by sonication. The lysed sample was centrifuged, and the resulting supernatant was incubated with 2mL Ni-NTA resin (washed in Lysis Buffer) while rotating at 4°C. After 1 hour, the resin was washed 2x with 20mL Lysis buffer. The protein was eluted from the Ni-NTA resin with 8mL of Elution Buffer (20mM HEPES pH7.6, 500mM NaCl, 300mM Imidazole). The eluted sample was filtered and loaded onto a S75 26/60 sizing column equilibrated in Gel Filtration buffer (20mM HEPES pH7.6, 150mM NaCl, 1mM DTT, 5% Glycerol). Fractions were collected and their purity assessed via SDS-PAGE. Pure samples were pooled and concentrated prior to being flash-frozen and stored at -80°C.

Protein Constructs and Purification (SUMO-Peptides)

Peptides were cloned as SUMO fusion constructs and inserted into a pDW363 vector for recombinant expression in Escherichia coli. The construct was transformed into Rosetta2(DE3) cells (Novagen) and grown in Lysogeny Broth (LB) media supplemented with 100 μg/ml ampicillin and 0.05mM D-(+)-biotin for in vivo biotinylation. The cells were grown while shaking at 37 °C until reaching an OD600 of 00.7, upon which protein expression was induced with 1mM IPTG at 37°C for 5 hours. Cultures were harvested by centrifugation, and the resulting pellets were frozen overnight at -80 °C. The next day, the pellets were thawed and resuspended in 4mL B-PER (Thermo Fischer)/g cell pellet and supplemented with 0.2mM PMSF. Samples were incubated at room temperature for 15 minutes, spun down, and the supernatant was loaded into columns containing 0.4mL Ni-NTA resin equilibrated in Wash Buffer (20mM Tris pH8.0, 500mM NaCl, 20mM imidazole). The columns were washed 3x with 1mL Wash Buffer and then eluted from the column with 1.5mL Elution Buffer (20mM Tris, 500mM NaCl, 300mM Imidazole pH8.0). Samples were flash-frozen and stored at -80°C.

Surface Display Constructs

Control peptides for bacterial surface display were cloned as C-terminal fusions to the surface protein eCPX using a vector designed by the Daugherty group [82]. Peptides were flanked

by an N-terminal FLAG tag and a C-terminal c-Myc tag for quantification of peptide expression on the bacterial surface. These plasmids were transformed into chemically competent MC1061 E. Coli cells for FACS analysis.

The T7-pep library, consisting of the same construct design detailed above, was transformed into electrocompetent MC1061 cells (Lucigen) to maximize transformation efficiency [38]. 25μL of cells were pre-chilled on ice for 10 min. These cells were mixed with plasmid and were pulsed in a 0.1cm cuvette at 1.8kV, 600 Ohm, 10μF using a Biorad electroporator. The cuvette was immediately rinsed twice with 1mL of Super Optimal Broth media with 20 mM glucose (SOC) that had been pre-warmed at 37 °C. A serial dilution of the transformed cells was plated on LB+Cm plated to assess transformation efficiency. The cells were recovered at 37 °C for 1 hour and then combined with 40mL LB + chloramphenicol (Cm) + 0.2% glucose for overnight growth. This process was repeated between rounds of enrichment to minimize the possibility of cells developing growth advantages associated with mutations in the chromosomal DNA.

FACS Sample Preparation

The following protocol was adopted from Hwang et al. [33]. All PBS buffers were supplemented with additional NaCl to a final concentration of 300mM. Cultures consisting of 5mL LB + 25 μg/mL chloramphenicol + 0.2% glucose (w/v) were inoculated overnight (37 °C) with MC1061 cells encoding arabinose-inducible eCPX-peptide fusion constructs. The next morning, cells were pelleted and used to inoculate a fresh culture of LB + 25 μg/mL chloramphenicol (no glucose). These cultures were grown until an OD600 of 0.5-0.6 (37 °C). Upon reaching this density, eCPX expression was induced with 0.04% arabinose (w/v) for 1.5 hours. Following expression, the OD600 was measured again, and an appropriate volume of culture was centrifuged to obtain a pellet of 1 x 10^7 cells/FACs analysis sample. In the case of T7-pep library sorting, a total of 1 x 10^8 cells were pelleted. The pellets were washed with PBS pH7.4 + 0.1% BSA and then incubated with anti-FLAG antibody conjugated to allophycocyanin (anti-FLAG:APC) (diluted 1:100 into PBS pH7.4 + 0.1% BSA; 30μL /1 x 10^7 cells) for 30 min at 4 °C. At this point, the biotinylated monomeric EVH1 domain (1μM) to be analyzed was pre-tetramerized by incubation with streptavidin-phycoerythrin (SA-PE) for 30 min. Following this incubation, the cells were pelleted and washed with PBS pH7.4 + 0.1% BSA to remove excess anti-FLAG:APC. Cells were pelleted, resuspended in PBS pH7.4 + 1% BSA at 25μL buffer/1 x 10^7 cells, and incubated with pre-tetramerized EVH1 domain at a 1:1 for a final volume of 50μL/sample for 1 hour at 4 °C. Following this incubation, the cells were transferred to a pre-wet Millipore MultiScreen 0.22μm Hydrophilic Low Protein Binding plate and the buffer was removed by vacuum. The cells were then washed twice with 200μL PBS + 0.1% BSA before being resuspended in 250μL PBS + 0.1% BSA for FACS analysis or sorting.

Flow Cytometry and FACS Analysis

Single clone bacterial surface display analysis was carried out on a FACSCanto instrument (BD Biosciences) with samples prepared as described above. For each sample, a totally of 10,000 events were observed. FACS data was processed out using the commercially avail-

able software FlowJo (FlowJo^{TM} Software for Mac, Version 10.8.1. Ashland, OR: Becton, Dickinson and Company; 2023). Cells were gated on (1) FSC-H vs. SSC-H (2) SSC-W vs. SSC-H. Cells which made it through these gates were analyzed for their APC and PE signal to quantify peptide expression and EVH1 binding, respectively.

Enrichment sorting of the T7-pep library was carried out on a FACSAria instrument (BD Biosciences). Collected cells were gated three times: (1) FSC-H vs. SSC-H (2) SSC-W vs. SSC-H (3) APC vs. PE. A no-peptide negative control and two positive controls of different affinities for Homer1 (CAP2, DBN1) were used to set the final collection gate. With each round of sorting, the PE voltage was adjusted to ensure a similar percentage of control cells fell within the gate. Samples, prepared as described above, were oversampled at least 10-fold relative to the library population. After the initial round of selection against the naïve T7-pep library, we performed a negative, PE-only sort and collected the non-binding population in order to eliminate sequences that were binding to PE. Subsequently, we performed 3 more rounds of positive selection using Homer1, for a total of 4 positive selections. The final round of sorting was performed twice (Sort5a/5b), but for the purpose of analysis, these populations were combined. Sorted cells were collected in SOC media and recovered at 37 °C for 1 hour. The culture was then combined with LB + 25 μg/mL chloramphenicol + 0.2% glucose (w/v) culture for overnight growth (37 °C) and then miniprepped (Qiagen).

Illumina Amplicon Preparation

The plasmids from each round of enrichment sorting were prepared for single-lane Illumina deep sequencing as follows. Miniprepped plasmids from each round of enrichment (100ng DNA/sample) were PCR amplified with a common forward primer to introduce an i5 Illumina anchor (FWD PCR) and a unique reverse primer to introduce the i7 Illumina anchor and a unique 6-nucleotide index for demultiplexing (REV INDEX X). A complete list of these primers and their sequences is provided in (Supplementary X). Twelve cycles of amplification were carried out with Q5 High Fidelity DNA Polymerase (NEB) using an annealing temperature of 68 °C and a 1 min extension at 65 °C. The success of this amplification was verified using a 1.5% agarose gel. Each PCR product was purified using a double-sided SPRI bead cleanup (AMPure XP- Beckman). First, a right-sided selection (0.56x) was used to remove large library fragments. The SPRI beads were resuspended in a solution of 2.5M NaCl and PEG 8000 (20% w/v), mixed with the PCR products, and then incubated for 5 min at room temperature. Next, the beads were magnetically separated and the supernatant removed. This supernatant was then used to perform a left-sided selection (0.85x) to remove library fragments and adapter dimers. A fresh sample of SPRI beads was resuspended in PEG solution and then combined with the supernatant from the right-handed selection. The beads were then magnetically separated, the supernatant discarded and washed twice with 200μL of pre-chilled 80% ethanol. Following this wash, the ethanol was removed and the beads were dried for 10 minutes at room temperature. The final sample was eluted from the beads by resuspension in 11μL of 10 mM Tris pH8.0 and their concentration measured via NanoDrop. The samples were pooled and submitted for sequencing using a NextSeq500 instrument with paired-end reads.

Illumina Data Processing

Unless noted, an in-house script was used for the following data processing. Forward and reverse reads from Illumina sequencing were merged (BBMerge) using the default overlap requirement of 12 base pairs and a quality score of 20 [83]. Read counts for each pool were collapsed by amino-acid sequence, and we applied an initial quality filter requiring a sequence length of 36 amino acids as well as an input library read count filter >5. The remaining 741,742 sequences were clustered to remove redundancy using CD-Hit, yielding a final count of 311,560 unique sequences [84]. For each sequence, the raw read count was converted to a frequency (F_i) by dividing the total number of reads within the respective pool. Sequences that failed to enrich in the first selection against the naive input library were categorized as "non-enriching". For sequences that persisted through the first two rounds of selection, an enrichment score was calculated by taking the area under the curve of the following transformation:

$$E = \log_2 \left(\frac{F_{i+1}^{\text{seq}}}{F_i^{\text{seq}}} \right) \tag{1.1}$$

The distribution of this enrichment score was used to categorize sequences as "strongly enriching" or "weakly enriching" with a Z-score cutoff of 1.

Sequence Logos

Matches to the core Homer motif were identified in sequences from the T7-Pep library using the regular expression [PAILV]P..[FWY]. These sequences were aligned by the motif with both an N- and C-terminal padding of 10 residues. In sequences where the motif was within 10 residues of the N-terminus, we appended the sequence "xxxxxxialr", and for those within 10 residues of the C-terminus we appended "riarxxxxxx". These residues correspond to non-proteomic sequences within the surface display construct and were displayed in lowercase to differentiate them from the proteomic residues within the logo. Position-specific scoring matrices (PSSMs) were generated using the publicly available software pLogo, with matches to the motif in all three populations (strongly enriching, weakly enriching, and non-enriching) serving as the background [41]. PSSMs for sequences derived from the proteomic SPOT array were similarly aligned by the Homer motif. Final logos were generated from these PSSMs using the Logomaker python library [85].

BLI Assay

BLI experiments were carried out on an Octet Red96 instrument (ForteBio). All assay steps were carried out in a 1:1 mixture of EVH1 Gel Filtration Buffer (20mM HEPES pH7.6, 150mM NaCl, 1mM DTT) and PBS Buffer (1x PBS pH7.6, 1mM DTT, 1% BSA, 0.1% Tween-20). Data was collected at room temperature using a shake speed of 1000 rpm at the default sampling rate. Streptavidin-coated tips (ForteBio) were equilibrated in the above buffer for 10 min. Biotinylated SUMO-peptides were loaded on the streptavidin tips to a target optical density between 0.5-1 nm and subsequently washed in the buffer for 60s. The loaded tips were immersed in a solution of EVH1 domain to collect association data for 100s. Subsequently, EVH1-bound tips were transferred to a well-containing buffer, and

dissociation data was collected for 100s. This process was done iteratively, using the same tip, with increasing concentrations of EVH1 domain, for a total of 8 concentrations.

Due to the fast kinetics of these interactions, we opted to calculate K_D values using the equilibrium signal of the association step at each concentration of the EVH1 domain. To account for noise, this signal was calculated as the average signal over the final 10 seconds of association, minus the average of the final 10 seconds of the dissociation step. The association data of a SUMO-only control was subtracted from that of the SUMO-peptides. The binding curve was fit and K_D value was extracted in Kaleidagraph using non-linear least-squares fitting to the following equation:

$$y = \frac{C[\text{EVH1}]}{K_D + [\text{EVH1}]} \tag{1.2}$$

Where y is the signal and C is the max signal at saturation. K_D values and their respective errors are reported as the average and standard deviation on 3 technical replicates.

Fold change error in Figure 4C was calculated by error propagation using the formula:

$$\sqrt{\left(\frac{SD_a}{\mu_a}\right)^2 + \left(\frac{SD_b}{\mu_b}\right)^2} \tag{1.3}$$

Proteome Bioinformatic Analysis

Bioinformatic analysis of the human (UP000005640; Organism 9606) and Dictyostelium discoideum (UP000002195; Organism 44689) proteomes was performed using the one protein per gene datasets available on UniProt. Regions were classified as disordered using IUPRED [43]. Human orthologs for Dictyostelium proteins were identified using the publicly available InParanoiDB9 (Human: 9606; Dictyostelium 44689) [86].

SPOT Array

ABI3 substitution and overlapping motif SPOT arrays were synthesized at the MIT Biopolymer Facility using an Intavis SPOT synthesis peptide arrayer system. At the N-terminus, the 20-mer peptides were capped with a Gly-Ser linker and acetylated. The peptides were linked to the cellulose membrane by a C-terminal PEG spacer. For the overlapping motif array, because many of the peptide sequences contain additional Ena/VASP and Homer motifs, each peptide was also synthesized as a control where the internal proline tract was mutated to poly-Glycine. The sequences of each peptide are provided in Supplementary X. The protocol for blocking and probing the membrane was adapted from Frank and Dubel [87]. The membrane was hydrated in 100% ethanol, washed three times with TBS (50 mM Tris pH 7.0, 137 mM NaCl, 2.7 mM KCl, 1mM DTT) and then blocked overnight in MBS (TBS +0.2% Tween-20, 2% dry milk) at room temperature. The following day, the array was washed in T-TBS (TBS + 0.05% Tween-20) and then incubated with the protein sample in MBS (+1μM DTT) for 3 hours at room temperature. All array experiments were carried out with Streptavidin-Cy3 pre-tetramerized EVH1 at a final monomer concentration of 0.5μM. The membrane was washed three times in T-TBS prior to scanning on a GE Amersham Imager 680. SPOT intensities were calculated using the ImageJ analysis software [88]. No

background correction was performed. For the overlapping peptides, the final intensity was calculated as the difference between the intensity of the wild-type and motif knockout control. An intensity threshold of 1000 was used to classify a peptide as having bound. Overlapping motifs that bound to both ENAH and Homer1 via SPOT array were assessed as candidate binding partners based on localization (gTEX/UniProt subcellular localization), interaction (HURI/BioGRID), and GO term data (Supplemental Figure1.11) [42], [89], [90]. GO term analysis was performed using the SLiMSearch 4 "shared functional annotations" tool [91]. A p-value threshold of 0.001 for any paralogous Ena/VASP or Homer family member was used as a cutoff for marking GO terms as being shared.

Contributions: Avinoam Singer - Conceptualization, Data Curation, Formal analysis, Writing - Original Draft, Writing - Review and Editing. Alejandra Ramos - Data Curation. Amy E. Keating - Conceptualization, Funding Acquisition, Supervision, Writing - Original Draft, Writing - Review and Editing.

Supplemental Figures

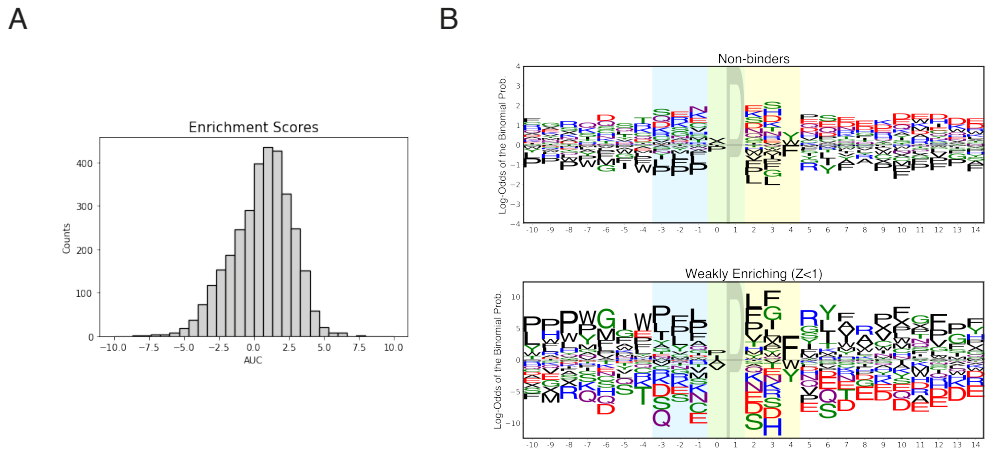


Figure 1.9: (A) Distribution of enrichment scores as described in the methods (B) Sequence logos for non-enriching and weakly enriching sequences.

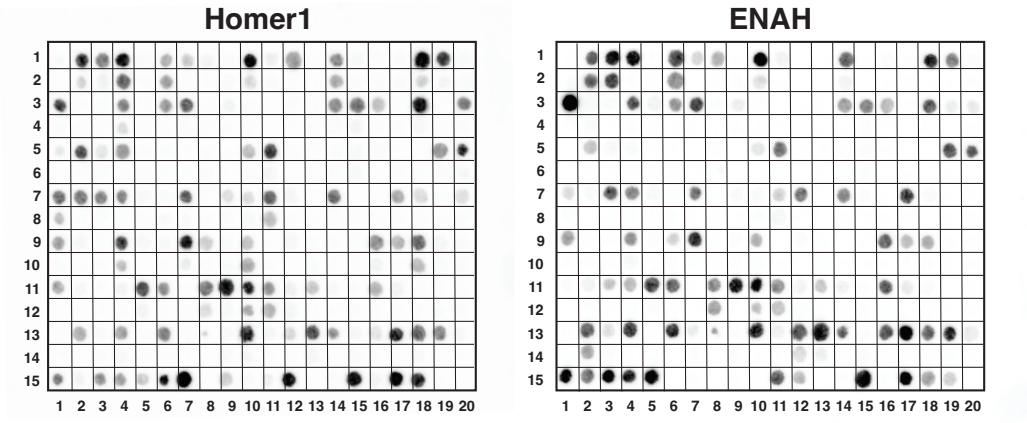


Figure 1.10: Overlapping motif SPOT arrays for Homer1 and ENAH.

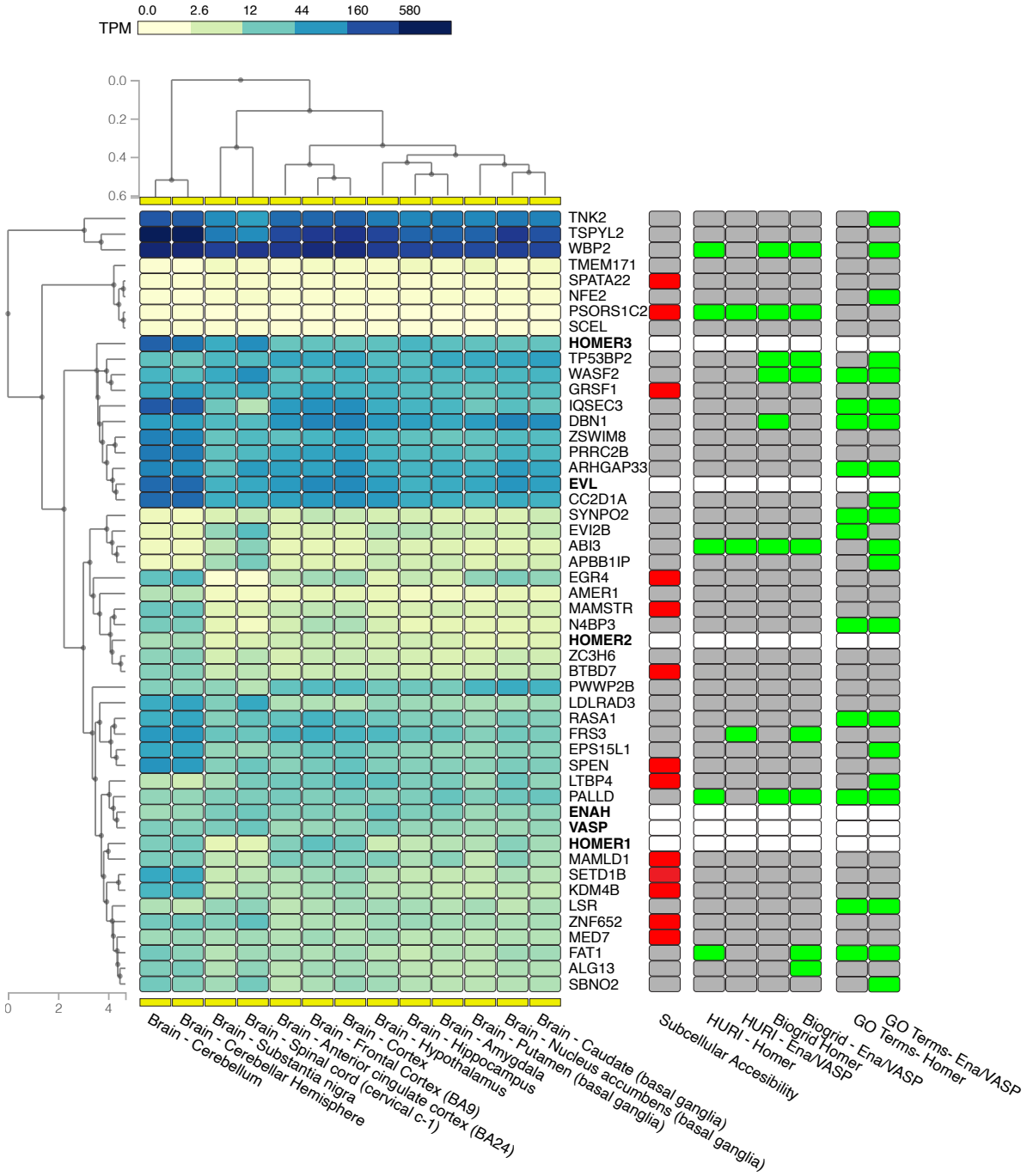


Figure 1.11: Expression, localization (gTex), and cataloged interactions of proteins containing promiscuous SLiMs (gTex).

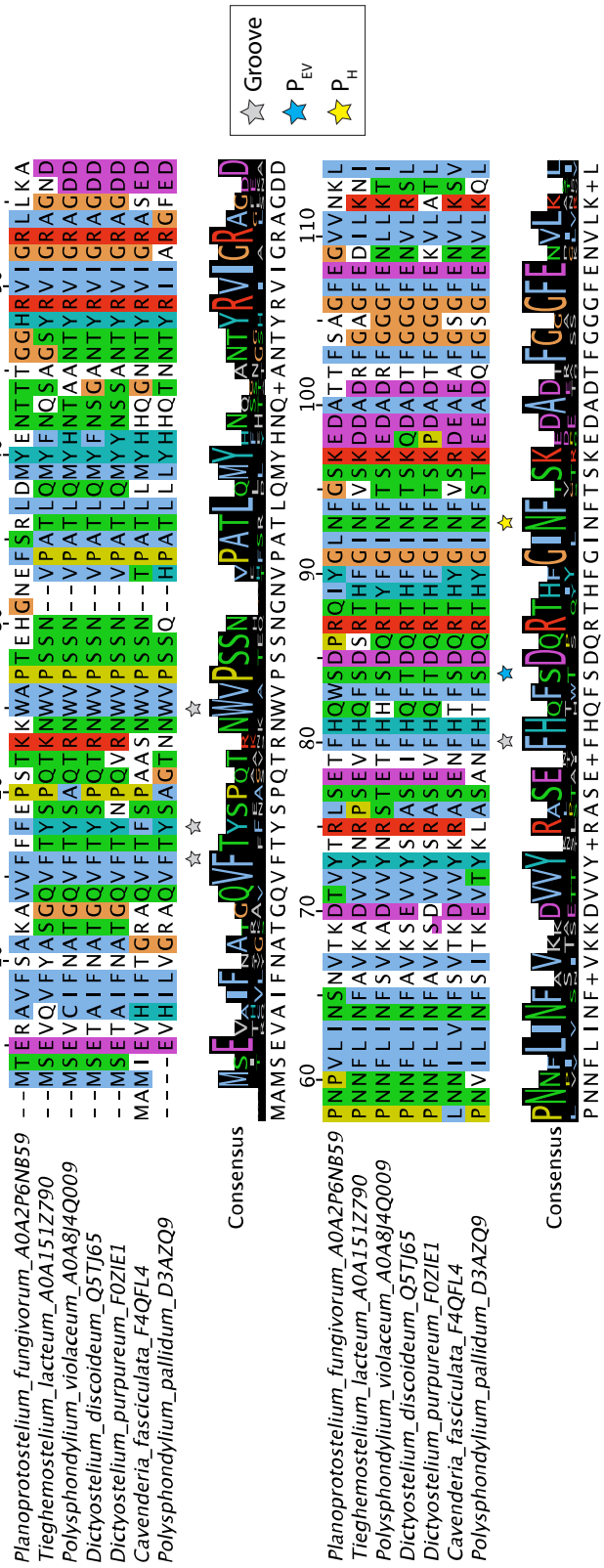
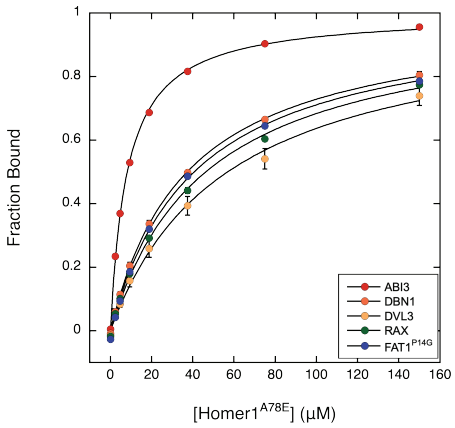


Figure 1.12: Sequence alignment of Ena/VASP EVH1 domains from amoebas.

A



Name	Sequence	K_D (μM)
DBN1	LNFDE**LPEPPATF**CDPEEVE	37.3 ± 1.7
FAT1^{P5G}	DIESD**FPPPPEDF**GAADELP	40.7 ± 1.4
ABI3	GDELG**LPPPPPGF**GPDEPSW	8.2 ± 0.2
RAX	GFGPPAQSLPASYTPPP**PPPPF**LNSPPLGPGLQPLA	46.4 ± 1.4
DVL3	PLPHPGAAPWPMAFPYQ**YPPPPHPY**NPHPGFPELGY	57.4 ± 8.0

B

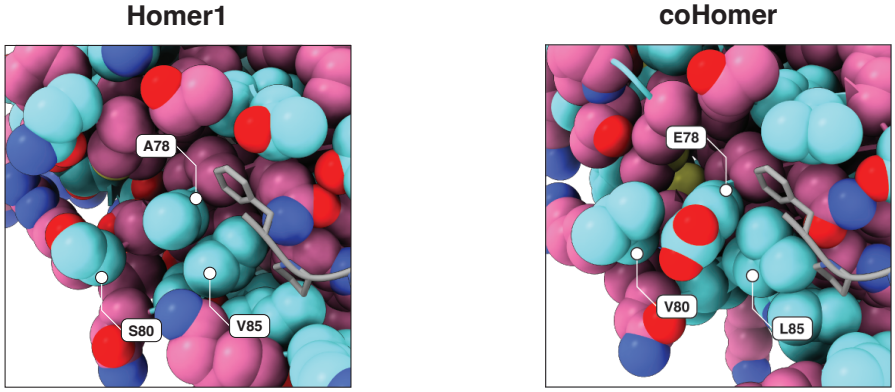


Figure 1.13: (A) BLI measurements for Homer^{A78E} (B) Comparison of the regions corresponding to Homer1 (1DDV) and coHomer P_{EV}.

References

[1] R. P. Bhattacharyya, A. Reményi, B. J. Yeh, and W. A. Lim, "Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits," *Annual Review of Biochemistry*, vol. 75, pp. 655–680, 2006, ISSN: 00664154. DOI: 10.1146/annurev.biochem.75.103004.142710.

[2] P. Tompa, N. E. Davey, T. J. Gibson, and M. M. Babu, "A Million peptide motifs for the molecular biologist," *Molecular Cell*, vol. 55, no. 2, pp. 161–169, 2014, ISSN: 10974164. DOI: 10.1016/j.molcel.2014.05.032. [Online]. Available: http://dx.doi.org/10.1016/j.molcel.2014.05.032.

[3] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nature Reviews Molecular Cell Biology*, vol. 16, no. 1, pp. 18–29, 2015, ISSN: 14710080. DOI: 10.1038/nrm3920.

[4] K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, and N. E. Davey, "Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation," *Chemical Reviews*, vol. 114, no. 13, pp. 6733–6778, 2014, ISSN: 15206890. DOI: 10.1021/cr400585q.

[5] K. Bugge, I. Brakti, C. B. Fernandes, J. E. Dreier, J. E. Lundsgaard, J. G. Olsen, K. Skriver, and B. B. Kragelund, "Interactions by Disorder – A Matter of Context," *Frontiers in Molecular Biosciences*, vol. 7, no. June, pp. 1–16, 2020, ISSN: 2296889X. DOI: 10.3389/fmolb.2020.00110.

[6] G. W. Foight and A. E. Keating, "Comparison of the peptide binding preferences of three closely related TRAF paralogs: TRAF2, TRAF3, and TRAF5," *Protein Science*, vol. 25, pp. 1273–1289, 2016, ISSN: 1469896X. DOI: 10.1002/pro.2881.

[7] C. Vogel and C. Chothia, "Protein family expansions and biological complexity," *PLoS Computational Biology*, vol. 2, no. 5, pp. 370–382, 2006, ISSN: 15537358. DOI: 10.1371/journal.pcbi.0020048.

[8] N. E. Davey, M. S. Cyert, and A. M. Moses, "Short linear motifs - Ex nihilo evolution of protein regulation Short linear motifs - The unexplored frontier of the eukaryotic proteome," *Cell Communication and Signaling*, vol. 13, no. 1, pp. 9–11, 2015, ISSN: 1478811X. DOI: 10.1186/s12964-015-0120-z. [Online]. Available: http://dx.doi.org/10.1186/s12964-015-0120-z.

[9] G. C. Conant and K. H. Wolfe, "Turning a hobby into a job: How duplicated genes find new functions," *Nature Reviews Genetics*, vol. 9, no. 12, pp. 938–950, 2008, ISSN: 14710056. DOI: 10.1038/nrg2482.

[10] N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, and T. J. Gibson, "Attributes of short linear motifs," *Molecular BioSystems*, vol. 8, no. 1, pp. 268–281, 2012, ISSN: 1742206X. DOI: 10.1039/c1mb05231d.

[11] F. C. Peterson and B. F. Volkman, "Diversity of polyproline recognition by EVH1 domains," *Front Biosci (Landmark Ed)*, vol. 14, pp. 833–846, 2009. DOI: 10.2741/3281. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/19273103.

[12] Y. Ueki, T. Kruse, M. B. Weisser, *et al.*, "A Consensus Binding Motif for the PP4 Protein Phosphatase," *Molecular Cell*, vol. 76, no. 6, 953–964.e6, 2019, ISSN: 10974164. DOI: 10.1016/j.molcel.2019.08.029. [Online]. Available: https://doi.org/10.1016/j.molcel.2019.08.029.

[13] F. J. Ayala, A. Rzhetsky, and F. J. Ayala, "Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 2, pp. 606–611, 1998, ISSN: 00278424. DOI: 10.1073/pnas.95.2.606.

[14] P. Burkhardt, M. Grønborg, K. McDonald, T. Sulur, Q. Wang, and N. King, "Evolutionary insights into premetazoan functions of the neuronal protein Homer," *Molecular Biology and Evolution*, 2014, ISSN: 15371719. DOI: 10.1093/molbev/msu178.

[15] J. Faix and K. Rottner, "Ena/VASP proteins in cell edge protrusion, migration and adhesion," *Journal of cell science*, vol. 135, no. 6, 2022, ISSN: 14779137. DOI: 10.1242/jcs.259226.

[16] C. Bachmann, L. Fischer, U. Walter, and M. Reinhard, "The EVH2 domain of the vasodilator-stimulated phosphoprotein mediates tetramerization, F-actin binding, and actin bundle formation," *Journal of Biological Chemistry*, vol. 274, no. 33, pp. 23 549–23 557, 1999, ISSN: 00219258. DOI: 10.1074/jbc.274.33.23549. [Online]. Available: http://dx.doi.org/10.1074/jbc.274.33.23549.

[17] S. Hüttelmaier, B. Harbeck, N. O. Steffens, T. Meßerschmidt, S. Illenberger, and B. M. Jockusch, "Characterization of the actin binding properties of the vasodilator-stimulated phosphoprotein VASP," *FEBS Letters*, vol. 451, no. 1, pp. 68–74, 1999, ISSN: 00145793. DOI: 10.1016/S0014-5793(99)00546-3.

[18] J. E. Bear and F. B. Gertler, "Ena/VASP: Towards resolving a pointed controversy at the barbed end," *Journal of Cell Science*, vol. 122, no. 12, pp. 1947–1953, 2009, ISSN: 00219533. DOI: 10.1242/jcs.038125.

[19] K. Rottner, B. Behrendt, J. V. Small, and J. Wehland, "VASP dynamics during lamellipodia protrusion," *Nature Cell Biology*, vol. 1, no. 5, pp. 321–322, 1999, ISSN: 14657392. DOI: 10.1038/13040.

[20] L. M. Lanier, M. A. Gates, W. Witke, A. S. Menzies, A. M. Wehman, J. D. Macklis, D. Kwiatkowski, P. Soriano, and F. B. Gertler, "Mena is required for neurulation and commissure formation," *Neuron*, vol. 22, no. 2, pp. 313–325, 1999, ISSN: 08966273. DOI: 10.1016/S0896-6273(00)81092-2.

[21] M. Reinhard, K. Jouvenal, D. Tripier, and U. Walter, "Identification, purification, and characterization of a zyxin-related protein that binds the focal adhesion and microfilament protein VASP (vasodilator-stimulated phosphoprotein)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 17, pp. 7956–7960, 1995, ISSN: 00278424. DOI: 10.1073/pnas.92.17.7956.

[22] F. Drees and F. B. Gertler, "Ena/VASP: proteins at the tip of the nervous system," *Current Opinion in Neurobiology*, vol. 18, no. 1, pp. 53–59, 2008, ISSN: 09594388. DOI: 10.1016/j.conb.2008.05.007.

[23] "The Homer family proteins," *Genome Biology*, vol. 8, no. 2, pp. 1–12, 2007, ISSN: 14747596. DOI: 10.1186/gb-2007-8-2-206.

[24] J. C. Tu, B. Xiao, J. P. Yuan, A. A. Lanahan, K. Leoffert, M. Li, D. J. Linden, and P. F. Worley, "Homer Binds a Novel Proline-Rich Motif and Links Group 1 Metabotropic Glutamate Receptors with IP3 Receptors Figure 1. Deletion Analysis Defines a Region in the C Terminus of mGluR5 Essential for Binding Homer Figure 2. Point Mutagenesis Defines Homer Lig," *Neuron*, vol. 21, pp. 717–726, 1998. [Online]. Available: https://ac.els-cdn.com/S0896627300805899/1-s2.0-S0896627300805899-main.pdf?_tid=bb2fc94d-09e7-4c3b-bc6b-377a479a6f8b&acdnat=1531297208_345d985e711b9c3d89b47b67acaee95d.

[25] J. C. Tu, B. Xiao, S. Naisbitt, *et al.*, "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins," *Neuron*, vol. 23, no. 3, pp. 583–592, 1999, ISSN: 08966273. DOI: 10.1016/S0896-6273(00)80810-7.

[26] B. Xiao, J. C. Tu, R. S. Petralia, J. P. Yuan, A. Doan, C. D. Breder, A. Ruggiero, A. A. Lanahan, R. J. Wenthold, and P. F. Worley, "Homer regulates the association of group 1 metabotropic glutamate receptors with multivalent complexes of Homer-related, synaptic proteins," *Neuron*, vol. 21, no. 4, pp. 707–716, 1998, ISSN: 08966273. DOI: 10.1016/S0896-6273(00)80588-7.

[27] N. Clifton, S. Trent, K. Thomas, and J. Hall, "Regulation and Function of Activity-Dependent Homer in Synaptic Plasticity," *Complex Psychiatry*, vol. 5, no. 3, pp. 147–161, 2019, ISSN: 2673-3005. DOI: 10.1159/000500267.

[28] P. R. Brakeman, A. A. Lanahan, R. O'Brien, K. Roche, C. A. Barnes, R. L. Huganir, and P. F. Worley, "Homer: A protein that selectively binds metabotropic glutamate receptors," *Nature*, 1997, ISSN: 00280836. DOI: 10.1038/386284a0.

[29] Z. Li, H. Liu, J. Li, *et al.*, "Supplemental Information Homer Tetramer Promotes Actin Bundling Activity of Drebrin," vol. 27, 2019.

[30] G. N. Huang, D. L. Huso, S. Bouyain, *et al.*, "NFAT binding and regulation of T cell activation by the cytoplasmic scaffolding homer proteins," *Science*, vol. 319, no. 5862, pp. 476–481, 2008, ISSN: 00368075. DOI: 10.1126/science.1151227.

[31] H. Dinkel, S. Michael, R. J. Weatheritt, *et al.*, "ELM - The database of eukaryotic linear motifs," *Nucleic Acids Research*, vol. 40, no. D1, pp. 242–251, 2012, ISSN: 03051048. DOI: 10.1093/nar/gkr1064.

[32] K. Niebuhr, F. Ebel, R. Frank, M. Reinhard, E. Domann, U. D. Carl, U. Walter, F. B. Gertler, J. Wehland, and T. Chakraborty, "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 Journal*, vol. 16, no. 17, pp. 5433–5444, 1997, ISSN: 02614189. DOI: 10.1093/emboj/16.17.5433.

[33] T. Hwang, S. S. Parker, S. M. Hill, R. A. Grant, M. W. Ilunga, V. Sivaraman, G. Mouneimne, and A. E. Keating, "Native proline-rich motifs exploit sequence context to target actin-remodeling Ena/VASP protein ENAH," *eLife*, vol. 11, pp. 1–21, 2022, ISSN: 2050084X. DOI: 10.7554/eLife.70680.

[34] L. A. Acevedo, A. I. Greenwood, and L. K. Nicholson, "A Noncanonical Binding Site in the EVH1 Domain of Vasodilator-Stimulated Phosphoprotein Regulates Its Interactions with the Proline Rich Region of Zyxin," *Biochemistry*, vol. 56, no. 35, pp. 4626–4636, 2017, ISSN: 15204995. DOI: 10.1021/acs.biochem.7b00618.

[35] J. Beneken, J. C. Tu, B. Xiao, M. Nuriya, J. P. Yuan, P. F. Worley, and D. J. Leahy, "Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition," *Neuron*, 2000, ISSN: 08966273. DOI: 10.1016/S0896-6273(00)81145-9.

[36] K. E. Prehoda, D. J. Lee, and W. A. Lim, "Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly," *Cell*, vol. 97, no. 4, pp. 471–480, 1999. DOI: 10.1016/s0092-8674(00)80757-6. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/10338211.

[37] N. E. Davey, M. H. Seo, V. K. Yadav, *et al.*, "Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome," *FEBS Journal*, vol. 284, no. 3, pp. 485–498, 2017, ISSN: 17424658. DOI: 10.1111/febs.13995.

[38] H. B. Larman, Z. Zhao, U. Laserson, *et al.*, "Autoantigen discovery with a synthetic human peptidome," *Nature Biotechnology*, vol. 29, no. 6, pp. 535–541, 2011, ISSN: 10870156. DOI: 10.1038/nbt.1856. [Online]. Available: http://dx.doi.org/10.1038/nbt.1856.

[39] S. Chen, Y. Zhang, H. Wang, Y. Y. Zeng, Z. Li, M. L. Li, F. F. Li, J. You, Z. M. Zhang, and C. M. Tzeng, "WW domain-binding protein 2 acts as an oncogene by modulating the activity of the glycolytic enzyme ENO1 in glioma article," *Cell Death and Disease*, vol. 9, no. 3, 2018, ISSN: 20414889. DOI: 10.1038/s41419-018-0376-5. [Online]. Available: http://dx.doi.org/10.1038/s41419-018-0376-5.

[40] Y. Muranishi, K. Terada, and T. Furukawa, "An essential role for Rax in retina and neuroendocrine system development," *Development Growth and Differentiation*, vol. 54, no. 3, pp. 341–348, 2012, ISSN: 00121592. DOI: 10.1111/j.1440-169X.2012.01337.x.

[41] J. P. O'Shea, M. F. Chou, S. A. Quader, J. K. Ryan, G. M. Church, and D. Schwartz, "PLogo: A probabilistic approach to visualizing sequence motifs," *Nature Methods*, vol. 10, no. 12, pp. 1211–1212, 2013, ISSN: 15487091. DOI: 10.1038/nmeth.2646.

[42] K. Luck, D. K. Kim, L. Lambourne, *et al.*, "A reference map of the human binary protein interactome," *Nature*, vol. 580, no. 7803, pp. 402–408, 2020, ISSN: 14764687. DOI: 10.1038/s41586-020-2188-x. [Online]. Available: http://dx.doi.org/10.1038/s41586-020-2188-x.

[43] B. Mészáros, G. Erdös, and Z. Dosztányi, "IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding," *Nucleic Acids Research*, 2018, ISSN: 13624962. DOI: 10.1093/nar/gky384.

[44] Y. H. Han, C. Y. Chung, D. Wessels, S. Stephens, M. A. Titus, D. R. Soll, and R. A. Firtel, "Requirement of a vasodilator-stimulated phosphoprotein family member for cell adhesion, the formation of filopodia, and chemotaxis in Dictyostelium," *Journal of Biological Chemistry*, vol. 277, no. 51, pp. 49 877–49 887, 2002, ISSN: 00219258. DOI: 10.1074/jbc.M209107200. [Online]. Available: http://dx.doi.org/10.1074/jbc.M209107200.

[45] S. Körber and J. Faix, "VASP boosts protrusive activity of macroendocytic cups and drives phagosome rocketing after internalization," *European Journal of Cell Biology*, vol. 101, no. 2, 2022, ISSN: 16181298. DOI: 10.1016/j.ejcb.2022.151200.

[46] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, "ColabFold: making protein folding accessible to all," *Nature Methods*, vol. 19, no. 6, pp. 679–682, 2022, ISSN: 15487105. DOI: 10.1038/s41592-022-01488-1.

[47] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, ISSN: 14764687. DOI: 10.1038/s41586-021-03819-2. [Online]. Available: http://dx.doi.org/10.1038/s41586-021-03819-2.

[48] C. Litschko, J. Linkner, S. Brühmann, T. E. Stradal, T. Reinl, L. Jänsch, K. Rottner, and J. Faix, "Differential functions of WAVE regulatory complex subunits in the regulation of actin-driven processes," *European Journal of Cell Biology*, vol. 96, no. 8, pp. 715–727, 2017, ISSN: 16181298. DOI: 10.1016/j.ejcb.2017.08.003. [Online]. Available: http://dx.doi.org/10.1016/j.ejcb.2017.08.003.

[49] M. Stillman, J. D. Lautz, R. S. Johnson, M. J. MacCoss, and S. E. P. Smith, "Activity dependent dissociation of the Homer1 interactome," *Scientific Reports*, vol. 12, no. 1, pp. 1–10, 2022, ISSN: 2045-2322. DOI: 10.1038/s41598-022-07179-3. [Online]. Available: https://doi.org/10.1038/s41598-022-07179-3.

[50] S. P. Goulding, K. K. Szumlinski, C. Contet, M. J. MacCoss, and C. C. Wu, "A mass spectrometry-based proteomic analysis of Homer2-interacting proteins in the mouse brain," *Journal of Proteomics*, vol. 166, pp. 127–137, 2017, ISSN: 18767737. DOI: 10.1016/j.jprot.2017.07.008. [Online]. Available: http://dx.doi.org/10.1016/j.jprot.2017.07.008.

[51] R. Rong, J. Y. Ahn, H. Huang, E. Nagata, D. Kalman, J. A. Kapp, J. Tu, P. F. Worley, S. H. Snyder, and K. Ye, "PI3 kinase enhancer-Homer complex couples mGluRI to PI3 kinase, preventing neuronal apoptosis," *Nature Neuroscience*, vol. 6, no. 11, pp. 1153–1161, 2003, ISSN: 10976256. DOI: 10.1038/nn1134.

[52] J. P. Yuan, K. Kiselyov, D. M. Shin, *et al.*, "Homer binds TRPC family channels and is required for gating of TRPC1 by IP3 receptors," *Cell*, vol. 114, no. 6, pp. 777–789, 2003, ISSN: 00928674. DOI: 10.1016/S0092-8674(03)00716-5.

[53] C. G. Ullman, L. Frigotto, and R. N. Cooley, "In vitro methods for peptide display and their applications," *Briefings in Functional Genomics*, vol. 10, no. 3, pp. 125–134, 2011, ISSN: 20412649. DOI: 10.1093/bfgp/elr010.

[54] T. Hwang, S. S. Parker, S. M. Hill, M. W. Ilunga, R. A. Grant, G. Mouneimne, and A. E. Keating, "A distributed residue network permits conformational binding specificity in a conserved family of actin remodelers," *eLife*, vol. 10, pp. 1–21, 2021, ISSN: 2050084X. DOI: 10.7554/eLife.70601.

[55] D. Schreiner, K. Müller, and H. W. Hofer, "The intracellular domain of the human protocadherin hFat1 interacts with Homer signalling scaffolding proteins," *FEBS Letters*, vol. 580, no. 22, pp. 5295–5300, 2006, ISSN: 00145793. DOI: 10.1016/j.febslet.2006.08.079.

[56] M. Barzik, U. D. Carl, W. D. Schubert, R. Frank, J. Wehland, and D. W. Heinz, "The N-terminal domain of Homer/Vesl is a new class II EVH1 domain," *Journal of Molecular Biology*, vol. 309, no. 1, pp. 155–169, 2001, ISSN: 00222836. DOI: 10.1006/jmbi.2001.4640.

[57] T. Makino and T. Gojobori, "Evolution of Protein-Protein Interaction Network Protein-Protein Interaction Network as a Typical Example of Biological Networks," *Genome Dyn. Basel, Karger*, vol. 3, pp. 13–29, 2007.

[58] H. B. Fraser, D. P. Wall, and A. E. Hirsh, "A simple dependence between protein evolution rate and the number of protein-protein interactions," vol. 6, pp. 6–11, 2003.

[59] M. W. Hahn and A. D. Kern, "Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks," *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 803–806, 2005, ISSN: 07374038. DOI: 10.1093/molbev/msi072.

[60] H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman, "Evolutionary rate in the protein interaction network," *Science*, vol. 296, no. 5568, pp. 750–752, 2002, ISSN: 00368075. DOI: 10.1126/science.1068696.

[61] B. Uyar, R. J. Weatheritt, H. Dinkel, N. E. Davey, and T. J. Gibson, "Proteome-wide analysis of human disease mutations in short linear motifs: Neglected players in cancer?" *Molecular BioSystems*, vol. 10, no. 10, pp. 2626–2642, 2014, ISSN: 17422051. DOI: 10.1039/c4mb00290c.

[62] C. J. McClune and M. T. Laub, "Constraints on the expansion of paralogous protein families," *Current Biology*, vol. 30, no. 10, R460–R464, 2020, ISSN: 18790445. DOI: 10.1016/j.cub.2020.02.075.

[63] D. A. Ghose, K. E. Przydzial, E. M. Mahoney, A. E. Keating, and M. T. Laub, "Marginal specificity in protein interactions constrains evolution of a paralogous family," *Proceedings of the National Academy of Sciences*, vol. 120, no. 18, e2221163120, 2023. DOI: 10.1073/pnas.2221163120. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2221163120. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.2221163120.

[64] C. J. McClune, A. Alvarez-Buylla, C. A. Voigt, and M. T. Laub, "Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space," *Nature*, vol. 574, no. 7780, pp. 702–706, 2019, ISSN: 14764687. DOI: 10.1038/s41586-019-1639-8. [Online]. Available: http://dx.doi.org/10.1038/s41586-019-1639-8.

[65] C. Landgraf, S. Panni, L. Montecchi-Palazzi, L. Castagnoli, J. Schneider-Mergener, R. Volkmer-Engert, and G. Cesareni, "Protein interaction networks by proteome peptide scanning," *PLoS Biology*, vol. 2, no. 1, pp. 94–103, 2004, ISSN: 15449173. DOI: 10.1371/journal.pbio.0020014.

[66] G. Cesareni, S. Panni, G. Nardelli, and L. Castagnoli, "Can we infer peptide recognition specificity mediated by SH3 domains?" *FEBS Letters*, vol. 513, no. 1, pp. 38–44, 2002, ISSN: 00145793. DOI: 10.1016/S0014-5793(01)03307-5.

[67] A. Zarrinpar, S. H. Park, and W. A. Lim, "Optimization of specificity in a cellular protein interaction network by negative selection," *Nature*, vol. 426, no. 6967, pp. 676–680, 2003, ISSN: 00280836. DOI: 10.1038/nature02178.

[68] K. Van Roey, T. J. Gibson, and N. E. Davey, "Motif switches: Decision-making in cell regulation," *Current Opinion in Structural Biology*, vol. 22, no. 3, pp. 378–385, 2012, ISSN: 0959440X. DOI: 10.1016/j.sbi.2012.03.004. [Online]. Available: http://dx.doi.org/10.1016/j.sbi.2012.03.004.

[69] K. Van Roey, H. Dinkel, R. J. Weatheritt, T. J. Gibson, and N. E. Davey, "The switches.ELM resource: A compendium of conditional regulatory interaction interfaces," *Science Signaling*, vol. 6, no. 269, pp. 1–11, 2013, ISSN: 19450877. DOI: 10.1126/scisignal.2003345.

[70] K. Rottner, T. E. Stradal, and B. Chen, "WAVE regulatory complex," *Current Biology*, vol. 31, no. 10, R512–R517, 2021, ISSN: 18790445. DOI: 10.1016/j.cub.2021.01.086.

[71] X. J. Chen, A. J. Squarr, R. Stephan, B. Chen, T. E. Higgins, D. J. Barry, M. C. Martin, M. K. Rosen, S. Bogdan, and M. Way, "Ena/VASP proteins cooperate with the WAVE complex to regulate the actin cytoskeleton," *Developmental Cell*, vol. 30, no. 5, pp. 569–584, 2014, ISSN: 18781551. DOI: 10.1016/j.devcel.2014.08.001. [Online]. Available: http://dx.doi.org/10.1016/j.devcel.2014.08.001.

[72] S. H. Soderling, E. S. Guire, S. Kaech, J. White, F. Zhang, K. Schutz, L. K. Langeberg, G. Banker, J. Raber, and J. D. Scott, "A WAVE-1 and WRP signaling complex regulates spine density, synaptic plasticity, and memory," *Journal of Neuroscience*, vol. 27, no. 2, pp. 355–365, 2007, ISSN: 02706474. DOI: 10.1523/JNEUROSCI.3209-06.2006.

[73] J. Bae, B. H. Sung, I. H. Cho, S. M. Kim, and W. K. Song, "Nesh regulates dendritic spine morphology and synapse formation," *PLoS ONE*, vol. 7, no. 4, 2012, ISSN: 19326203. DOI: 10.1371/journal.pone.0034677.

[74] C. Sala, V. Piech, N. Wilson, M. Passafaro, G. Liu, and M. Sheng, "Regulation of dendritic spine morphology and synaptic function by shank and homer," *Neuron*, vol. 31, no. 1, pp. 115–130, 2001. DOI: 10.1016/s0896-6273(01)00339-7.

[75] S. Yoon, N. H. Piguel, N. Khalatyan, L. E. Dionisio, J. N. Savas, and P. Penzes, "Homer1 promotes dendritic spine growth through ankyrin-G and its loss reshapes the synaptic proteome," *Molecular Psychiatry*, 2021, ISSN: 14765578. DOI: 10.1038/s41380-020-00991-1. [Online]. Available: http://dx.doi.org/10.1038/s41380-020-00991-1.

[76] J. Wu, A. Pipathsouk, A. Keizer-Gunnink, F. Fusetti, W. Alkema, S. Liu, S. Altschuler, L. Wu, A. Kortholt, and O. D. Weiner, "Homer3 regulates the establishment of neutrophil polarity," *Molecular Biology of the Cell*, 2015, ISSN: 19394586. DOI: 10.1091/mbc.E14-07-1197.

[77] C. L. Oxley, N. J. Anthis, E. D. Lowe, I. Vakonakis, I. D. Campbell, and K. L. Wegener, "An integrin phosphorylation switch: The effect of $\beta 3$ integrin tail phosphorylation on Dok1 and talin binding," *Journal of Biological Chemistry*, vol. 283, no. 9, pp. 5420–5426, 2008, ISSN: 00219258. DOI: 10.1074/jbc.M709435200. [Online]. Available: http://dx.doi.org/10.1074/jbc.M709435200.

[78] L. R. Orlando, R. Ayala, L. R. Kett, A. A. Curley, J. Duffner, D. C. Bragg, L. H. Tsai, A. W. Dunah, and A. B. Young, "Phosphorylation of the homer-binding domain of group i metabotropic glutamate receptors by cyclin-dependent kinase 5," *Journal of Neurochemistry*, vol. 110, no. 2, pp. 557–569, 2009, ISSN: 00223042. DOI: 10.1111/j.1471-4159.2009.06139.x.

[79] P. V. Hornbeck, J. M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, and M. Sullivan, "PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse," *Nucleic Acids Research*, vol. 40, no. D1, pp. 261–270, 2012, ISSN: 03051048. DOI: 10.1093/nar/gkr1122.

[80] J. Zhang, T. Chen, Q. Mao, *et al.*, "PDGFR-β-activated ACK1-AKT signaling promotes glioma tumorigenesis," *International Journal of Cancer*, vol. 136, no. 8, pp. 1769–1780, 2015, ISSN: 10970215. DOI: 10.1002/ijc.29234.

[81] A. La Torre, M. Del Mar Masdeu, T. Cotrufo, R. S. Moubarak, J. A. Del Río, J. X. Comella, E. Soriano, and J. M. Ureña, "A role for the tyrosine kinase ACK1 in neurotrophin signaling and neuronal extension and branching," *Cell Death and Disease*, vol. 4, no. 4, pp. 1–14, 2013, ISSN: 20414889. DOI: 10.1038/cddis.2013.99.

[82] J. J. Rice and P. S. Daugherty, "Directed evolution of a biterminal bacterial display scaffold enhances the display of diverse peptides," *Protein Engineering, Design and Selection*, vol. 21, no. 7, pp. 435–442, 2008, ISSN: 17410126. DOI: 10.1093/protein/gzn020.

[83] B. Bushnell, J. Rood, and E. Singer, "BBMerge – Accurate paired shotgun read merging via overlap," *PLoS ONE*, vol. 12, no. 10, pp. 1–15, 2017, ISSN: 19326203. DOI: 10.1371/journal.pone.0185056.

[84] W. Li and A. Godzik, "Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences," *Bioinformatics*, vol. 22, no. 13, pp. 1658–1659, 2006, ISSN: 13674803. DOI: 10.1093/bioinformatics/btl158.

[85] A. Tareen and J. B. Kinney, "Logomaker: Beautiful sequence logos in Python," *Bioinformatics*, vol. 36, no. 7, pp. 2272–2274, 2020, ISSN: 14602059. DOI: 10.1093/bioinformatics/btz921.

[86] E. Persson and E. L. Sonnhammer, "InParanoiDB 9: Ortholog Groups for Protein Domains and Full-Length Proteins," *Journal of Molecular Biology*, vol. 435, no. 14, p. 168 001, 2023, ISSN: 10898638. DOI: 10.1016/j.jmb.2023.168001. [Online]. Available: https://doi.org/10.1016/j.jmb.2023.168001.

[87] R. Frank and S. Dübel, "Analysis of Protein Interactions with Immobilized Peptide Arrays Synthesized on Membrane Supports," *Cold Spring Harbor Protocols*, vol. 2006, no. 4, pdb.prot4566, 2006, ISSN: 1940-3402. DOI: 10.1101/pdb.prot4566.

[88] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with imageJ," *Biophotonics International*, vol. 11, no. 7, pp. 36–41, 2004, ISSN: 10818693. DOI: 10.1201/9781420005615.ax4.

[89] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets.," *Nucleic acids research*, vol. 34, no. Database issue, pp. 535–539, 2006, ISSN: 13624962. DOI: 10.1093/nar/gkj109.

[90] J. Lonsdale, J. Thomas, M. Salvatore, *et al.*, "The Genotype-Tissue Expression (GTEx) project," *Nature Genetics*, vol. 45, no. 6, pp. 580–585, 2013, ISSN: 10614036. DOI: 10.1038/ng.2653.

[91] I. Krystkowiak and N. E. Davey, "SLiMSearch: A framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions," *Nucleic Acids Research*, vol. 45, no. W1, W464–W469, 2017, ISSN: 13624962. DOI: 10.1093/nar/gkx238.