



ARTICLE

Comprehensive analysis of all evolutionary paths between two divergent PDZ domain specificities

Joan Teyra¹  | Andreas Ernst² | Alex Singer¹ | Frank Sicheri^{3,4} | Sachdev S. Sidhu^{1,4} 

¹The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada

²Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Frankfurt am Main, Germany

³Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada

⁴Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Correspondence

Sachdev S. Sidhu, The Donnelly Centre University of Toronto, Toronto, ON, Canada.
Email: sachdev.sidhu@utoronto.ca

Funding information

Canadian Institutes of Health Research, Grant/Award Number: MOP-93684

Abstract

To understand the molecular evolution of functional diversity in protein families, we comprehensively investigated the consequences of all possible mutation combinations separating two peptide-binding domains with highly divergent specificities. We analyzed the Erbin PDZ domain (Erbin-PDZ), which exhibits canonical type I specificity, and a synthetic Erbin-PDZ variant (E-14) that differs at six positions and exhibits an atypical specificity that closely resembles that of the natural Pdlim4 PDZ domain (Pdlim4-PDZ). We constructed a panel of 64 PDZ domains covering all possible transitions between Erbin-PDZ and E-14 (i.e., the panel contained variants with all possible combinations of either the Erbin-PDZ or E-14 sequence at the six differing positions). We assessed the specificity profiles of the 64 PDZ domains using a C-terminal phage-displayed peptide library containing all possible genetically encoded heptapeptides. The specificity profiles clustered into six distinct groups, showing that intermediate domains can be nodes for the evolution of divergent functions. Remarkably, three substitutions were sufficient to convert the specificity of Erbin-PDZ to that of Pdlim4-PDZ, whereas Pdlim4-PDZ contains 71 differences relative to Erbin-PDZ. X-ray crystallography revealed the structural basis for specificity transition: a single substitution in the center of the binding site, supported by contributions from auxiliary substitutions, altered the main chain conformation of the peptide ligand to resemble that of ligands bound to Pdlim4-PDZ. Our results show that a very small set of mutations can dramatically alter protein specificity, and these findings support the hypothesis whereby complex protein functions evolve by gene duplication followed by cumulative mutations.

KEYWORDS

C-terminal peptide, in vitro evolution, PDZ domains, peptide-phage display, protein engineering, specificity profile

Abbreviations: DlgA, Drosophila disc large tumor suppressor; ELISA, enzyme-linked immunosorbent assay; GST, glutathione S-transferase; PDZ: Psd-95, post synaptic density protein; PRM, Peptide-recognition module; PWM, position weight matrix; ZO1, zonula occludens-1 protein.

Joan Teyra and Andreas Ernst contributed equally to this work.

1 | INTRODUCTION

Multicellular organisms rely on complex, finely tuned protein interaction networks to regulate biological functions and respond to environmental changes.¹ There is intense interest in understanding the processes whereby protein–protein interactions evolve to tune and re-wire signaling networks.² However, efforts to address the impact of genetic mutations on the molecular details of protein–protein interactions have been stymied by the complex nature of these systems,³ which often depend on dozens of residues directly involved in the interactions and also rely on poorly understood long range effects that are generally nonadditive and have been collectively described as epistasis.⁴ Furthermore, comparative structural analyses of the effects of mutations have been limited by the small number of close homologs in the natural protein repertoire.⁵

Peptide recognition modules (PRMs), key players in the evolution of signaling pathways, recognize short linear sequences in other proteins.⁶ Signal transduction proteins often contain multiple copies of the same or different protein interaction modules. PDZ domains, one of the most abundant PRMs, typically recognize specific C-terminal sequences to assemble protein complexes that transmit signals.⁷ The PDZ family is defined by a common fold, which consists of a six-stranded β -sheet sandwich and two α -helices. C-terminal peptides dock between the second β -strand (β_2) and second α -helix (α_2) and also interact with a conserved carboxylate-binding loop (Figure 1a).⁸ Although the main specificity determinants for PDZ domains have been defined, even subtle changes in the binding site can dramatically alter specificity.⁹

Previously, we used phage-displayed peptide libraries to derive a specificity map covering 82 natural PDZ domains, which shed light on the functional diversity of the family¹⁰ but did not provide a full understanding of the relationships between PDZ domain sequence and specificity. To a large extent, this was because the average sequence identity between PDZ domains is less than 30%, and it is difficult to ascertain which differences are functionally significant and what the consequences are for individual differences.^{10,11}

Confronted by the complexity and limited coverage of natural PDZ domain sequence space, we adapted the Erbin PDZ domain (Erbin-PDZ) as a model system to explore protein evolution in a more systematic manner. We constructed a large combinatorial library of “synthetic” Erbin-PDZ variants by randomizing 10 positions within the peptide-binding site, and we used phage display to select for stable domains.¹² Remarkably, one-quarter of the domains selected for structure alone proved to be functional for recognition of C-terminal

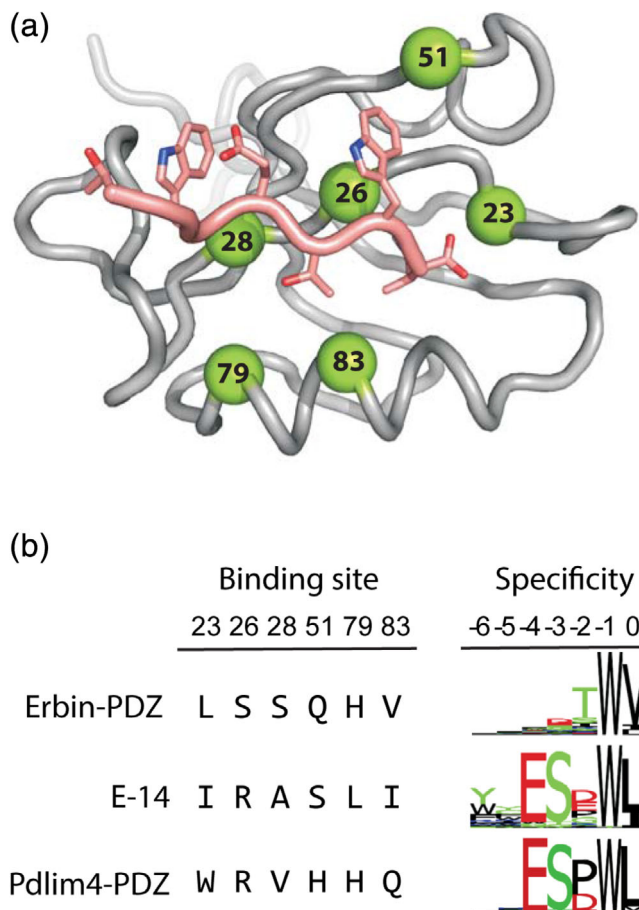


FIGURE 1 Structure and function of PDZ domains analyzed in this study. (a) The structure of Erbin-PDZ bound to an optimal peptide ligand (TWETWV_{COOH}). The Erbin-PDZ and peptide main chains are shown as a gray or salmon tubes, respectively. Peptide side chains are shown as sticks colored by atom type as follows: carbon (salmon), oxygen (red), nitrogen (blue). The six positions that differ between Erbin-PDZ and E-14 are shown as green spheres labeled according to the PDB file 1N7T.²⁹ The structure was rendered in Pymol (Schrödinger, LLC) using PDB coordinates 1N7T.²⁹ (b) PDZ binding site sequences and specificity profiles defined by peptide-phage display. Sequences are shown for the six positions that differ between Erbin-PDZ and E-14. Binding site positions are numbered according to PDB file 17NT and specificity profile positions are numbered according to the standard nomenclature for PDZ ligands in which the C-terminal residue is numbered “0” and preceding residues are numbered with negative integers.³⁰

peptides, and we found that a set of 61 functional domains represented at least 14 distinct specificity types, including some that closely matched those of natural domains and others that were not observed in nature.¹²

One particular variant, named E-14, exhibited specificity that was virtually identical to that of the atypical PDZ domain from the actin cytoskeleton associated protein Pdlim4 (Pdlim4-PDZ) (Figure 1b).¹⁰ Whereas Erbin-

PDZ and Pdlim4-PDZ share only 23% sequence identity and differ at 71 of 92 positions, E-14 differs from Erbin-PDZ at only six positions. Consequently, E-14 represents an ideal model system for understanding the minimal changes required to transition between the very different natural specificities exhibited by Erbin-PDZ and Pdlim4-PDZ, and consequently, to delineate precisely the minimal molecular changes required to evolve distinct biological functions from a common protein framework.

Here, we have comprehensively explored all evolutionary pathways between Erbin-PDZ and E-14 to define the minimal changes responsible for the two distinct functions. To accomplish this, we assembled a panel of 64 PDZ domains representing all mutational transitions between the two domains and subjected the panel to specificity profiling by C-terminal peptide-phage display. We complemented the functional analysis with crystal structures for Erbin-PDZ, E-14, and a key intermediate in complex with optimal ligands and compared these to the crystal structure of Pdlim4-PDZ in complex with its optimal ligand. Together, the structural and functional data provide a unique and comprehensive view of protein evolution at the molecular level.

2 | RESULTS AND DISCUSSION

2.1 | Specificity profiles for Erbin-PDZ and its variants

To understand in comprehensive detail how the six differences between Erbin-PDZ and E-14 alter specificity, we purified a panel of 64 Erbin-PDZ variants representing all possible transitions between the two. To facilitate high-throughput purification and analysis, the PDZ domains were produced as fusions to the C-terminus of glutathione *S*-transferase (GST). We used a random C-terminal peptide-phage library¹² representing all possible genetically encoded heptapeptides to select ligands for each variant and used clonal phage ELISAs to identify positive peptides that bound to the GST-PDZ fusion but not to GST. Although all 64 variants could be purified in a stable and soluble form, 10 did not select any positive peptides, suggesting that these variants are folded but likely not able to recognize C-terminal peptides with high affinity. DNA sequencing of positive clones for the other 54 variants revealed many unique ligands, and we assembled a data set of 1,166 peptides with each peptide validated as a ligand for a particular Erbin-PDZ variant (Table S1).

Ligands for each variant were aligned and used to calculate a position weight matrix (PWM), which was represented as a sequence logo¹³ (Figure S1). By grouping together variants that exhibited similar specificity logos, we identified

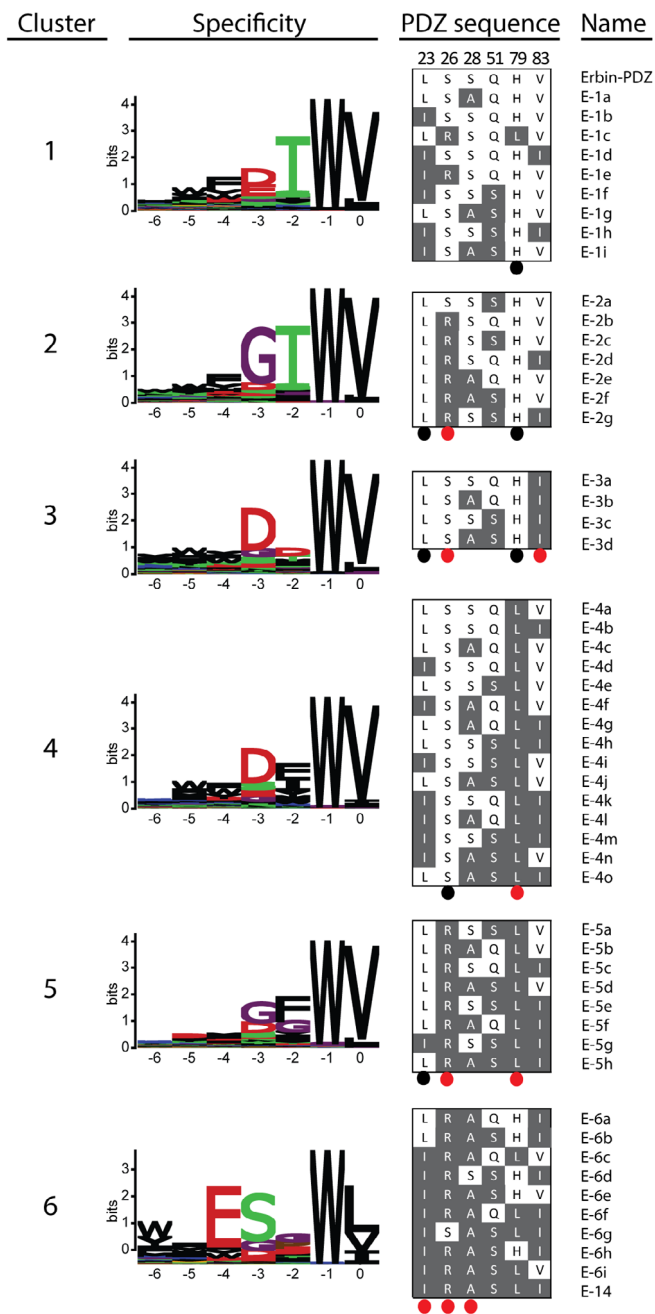


FIGURE 2 Specificity clusters for Erbin-PDZ and its variants. On the left, for each cluster, the aggregate specificity logo is shown for the alignment of all peptide ligands selected for all PDZ domains in the cluster. On the right, the sequences of the positions that differ between Erbin-PDZ (top sequence) and E-14 (bottom sequence) are shown with position numbers at the top and gray shading indicating sequences that match E-14. The clusters are arranged in order of increasing average number of mutations. Black and red filled circles indicate positions in a cluster that are conserved as the Erbin-PDZ or E-14 sequence, respectively. The names of domains are shown to the right of their sequences

six clusters, each with a different specificity. For each cluster, we calculated an aggregate sequence logo using all ligands associated with that cluster (Figure 2). This

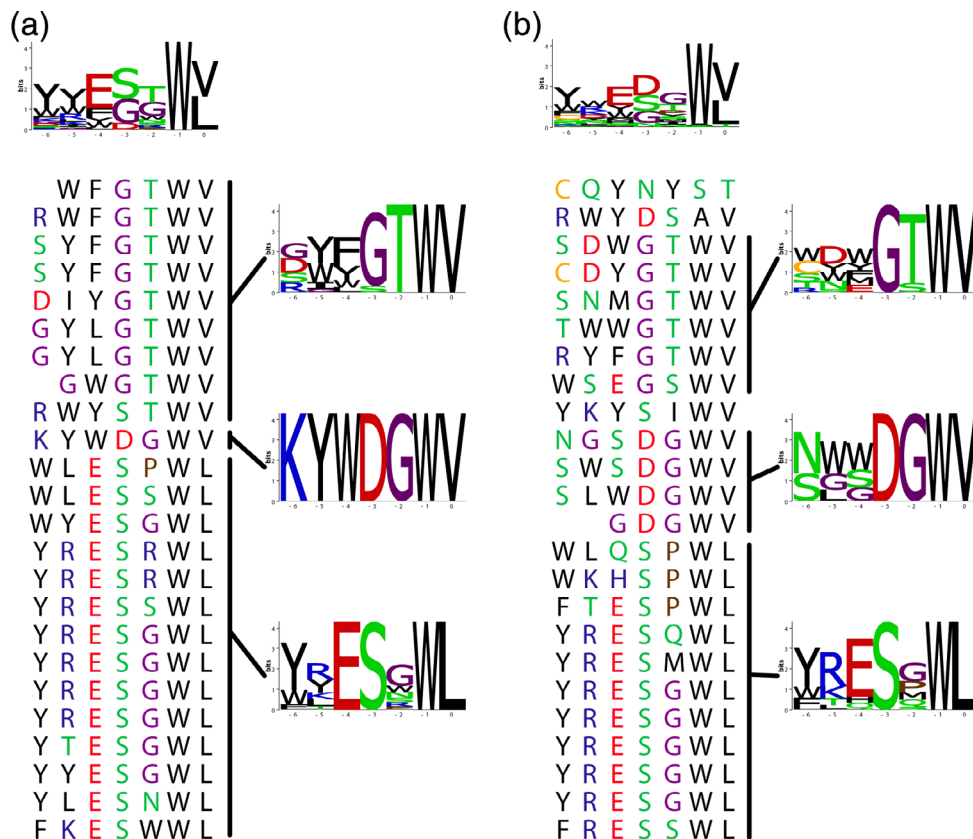


FIGURE 3 Multispecific Erbin-PDZ variants. Specificities are shown for (a) E-6a and (b) E-6b. The aggregate specificity logo for each domain is shown at the top and the peptide ligands used to derive the logo are shown below. The peptides are grouped according to sequence similarities and logos to the right of the peptides are derived from peptide subgroups, as indicated, and either belong to Cluster 2 (top) or Cluster 6 (bottom) or are novel. See Figure 2 for domain sequences and cluster definitions

analysis revealed common substitutions in each cluster, suggesting that these particular substitutions were largely responsible for the specificity changes in each cluster compared with Erbin-PDZ.

Cluster 1 corresponded to the wt Erbin-PDZ specificity, and the wt sequence was dominant at all PDZ positions, aside for a conservative Leu to Ile substitution at position 23. Cluster 2 was characterized by a preference for Gly rather than Asp at ligand position⁻³ and was associated with the substitution of Arg for Ser at PDZ position 26, which resides in the $\beta 2$ strand, where bulky residues have been shown to perturb recognition of the position⁻³ residue.⁹ Cluster 3 was characterized by a lack of specificity for ligand position⁻² and was associated with the substitution of Ile for Val at PDZ position 83, which resides in the $\alpha 2$ helix. Position 83 is close to position⁻² and it appears that a bulkier residue at this position compromises the ability for specific recognition of position⁻² residues. In cluster 4, the position⁻² specificity was for hydrophobes, which corresponds to a class II PDZ domain specificity, rather than the class I specificity for Ser/Thr⁻² residues displayed by Erbin-PDZ. All domains in cluster 4 contain Leu in place of His at position 79, which resides in the $\alpha 2$ helix and is known to be a major determinant for position⁻² specificity. These results are in agreement with sequences of natural PDZ domains, since class I and class II domains contain His or a hydrophobe, respectively, at position 79.¹⁰ In

cluster 5, the specificity profile appears to be a composite of the changes observed in clusters 2 and 4, as position⁻³ is dominated by Gly (similar to cluster 2) and position⁻² is dominated by hydrophobes (similar to cluster 4). Notably, the domains in cluster 5 also appear to be composites of the domains in clusters 2 and 4, as they all contain Ser to Arg substitutions at position 26 (similar to cluster 2) and His to Leu substitutions at position 79 (similar to cluster 4). These results show that positions 26 and 79 are major determinants of specificity for positions ⁻³ or ⁻², respectively, and substitutions at these PDZ positions work independently and exert additive effects on specificity. Finally, cluster 6, which contained the most substitutions relative to Erbin-PDZ on average, was characterized by specificity similar to that of E-14. The specificity of cluster 6 differs dramatically from that of all other clusters for positions ⁻², ⁻³, and ⁻⁴ and, also differs significantly for position 0. Unlike the other clusters, the dramatically different specificity of cluster 6 cannot be explained by individual changes in the PDZ domains, but rather, appears to derive from cooperative interactions between multiple substitutions.

2.2 | Multispecific Erbin-PDZ variants

Aggregate specificity logos derived from ligands for domains with similar specificities were useful for clustering similar

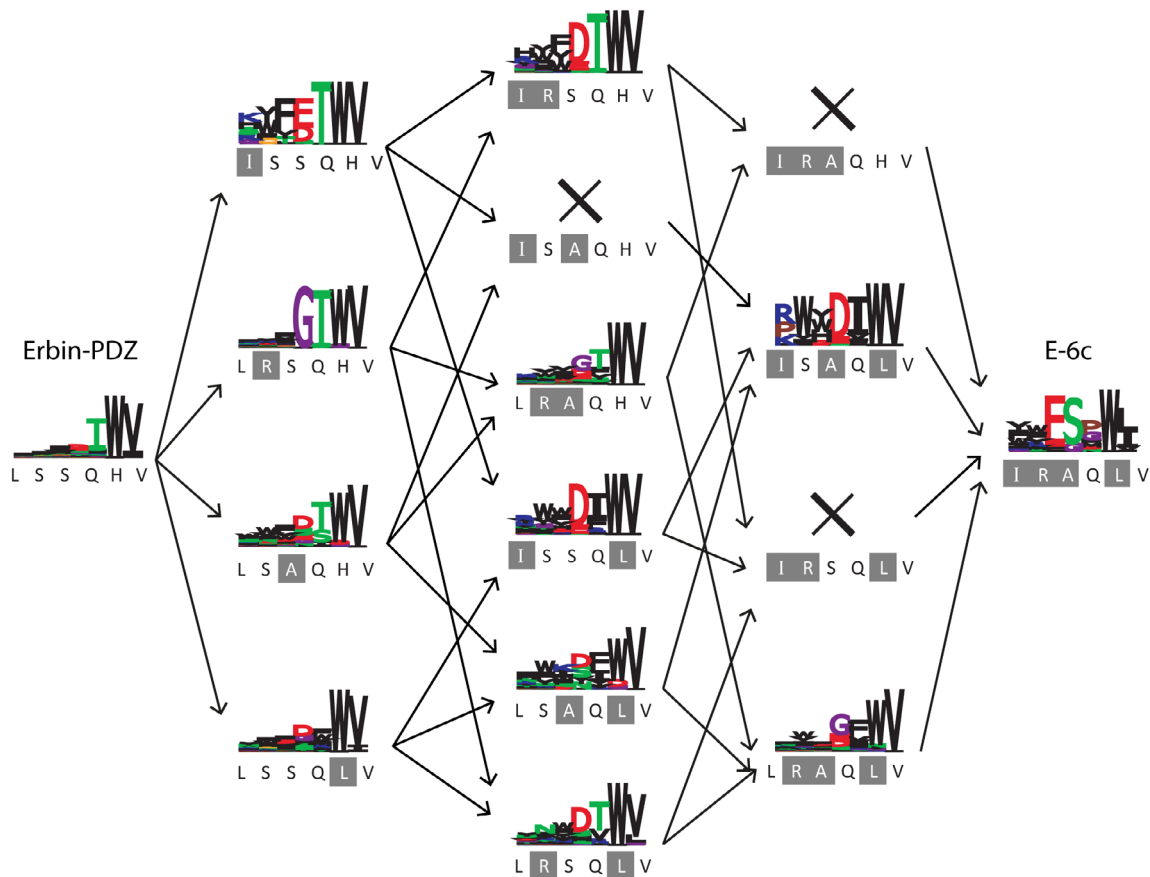


FIGURE 4 Evolutionary paths between Erbin-PDZ and variant E-6c. The specificity logo for Erbin-PDZ and E-6c are on the far left and far right, respectively. Below each logo, the sequence is shown for positions that differ between Erbin-PDZ and E-14 (see Figure 1 for numbering), and grey shading indicates residues that differ between Erbin-PDZ and E-6c. In between, from left to right, are shown the logos and sequences for single, double, and triple substitutions progressing from the sequence of Erbin-PDZ to that of E-6c. Lines connect variants that differ by one substitution and arrowheads point to the variant with an additional substitution relative to Erbin-PDZ. “X” indicates that no logo could be derived because the variant did not select any peptide ligands

domains together to reveal common sequence features associated with particular specificities (Figure 2). However, detailed inspection of ligands of individual domains also revealed an interesting subset of domains, which were able to bind peptides that clustered into two or more groups with very different logos, indicating that these are multispecific domains capable of binding specifically to two or more types of ligands (Table S1). Notably, E-6a and E-6b from Cluster 6 recognized some peptides that matched the specificity of E-14 but also recognized peptides that were more similar to ligands for other Erbin-PDZ variants (Erbin-PDZ variants are named E-nx, where “n” is a numeral indicating the cluster that the domain belongs to and “x” is a letter signifying its position in the ranking in Figure 2). In particular, both domains recognized a significant number of peptides that matched the specificity of Cluster 2 and also a few peptides that defined a novel specificity distinct from the six clusters (Figure 3). Notably, these two variants share three substitutions (S26R, S28A, and V83I), which appear to be the minimal changes in Erbin-PDZ that are required to acquire the

specificity of E-14/Pdlim4-PDZ, although intriguingly, both domains retain the capacity to also bind ligands that are more similar to those recognized by Erbin-PDZ. Thus, E-6a and E-6b appear to be intermediate domains that retain specificity similar to that of Erbin-PDZ but have also acquired specificity similar to that of E-14/Pdlim4-PDZ.

2.3 | Evolutionary paths between divergent specificities

A comprehensive set of specificities for all variants between Erbin-PDZ and E-14 provided an ideal opportunity to survey all evolutionary paths between the two specificities. Here we surveyed the evolution of E-6c, a variant that contains only four substitutions relative to Erbin-PDZ yet possesses a specificity that closely resembles that of E-14 (Figure 2). However, the same comprehensive analysis could be performed for any of the other variants in Cluster 6, including E-14 itself with six

TABLE 1 Data collection, phasing, and refinement statistics for PDZ/peptide complexes

	Erbin-PDZ/GYETWV	E-14/YYESDWL	E-6a/YYESGWL
PDB ID	6Q0N	6Q0M	6Q0U
Data collection			
Space group	<i>P1</i>	<i>P1</i>	<i>P1</i>
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	33.1, 33.4, 38.7	34.0, 36.0, 38.7	34.6, 38.1, 38.4
α , β , γ (°)	83.9, 75.5, 69.0	84.3, 77.0, 66.7	83.0, 75.0, 63.7
Wavelength	0.97918	0.97918	0.97918
Resolution (Å)	30.06–1.16 (1.18–1.16)	37.45–1.16(1.18–1.16)	37.14–1.89(1.99–1.89)
R_{sym} or R_{merge}^a	0.051(0.517)	0.041 (1.110)	0.051 (0.232)
$I/\sigma I$	12.4 (1.8)	15.5 (1.0)	10.4 (3.4)
Completeness (%)	86.7 (71.5)	86.5 (67.0)	91.1 (84.8)
Redundancy	3.5 (3.3)	3.7 (3.1)	2.4 (2.4)
Refinement			
Resolution (Å)	30–1.18	30.60–1.20	37.1–1.89
No. reflections	45,554 (4604)	45,815(5029)	12,486(1250)
$R_{\text{work}}^b/R_{\text{free}}^c$	15.5/18.3	15.6/18.5	17.8/23.7
B-factors			
Overall	26.5	21.6	31.9
Protein	25.3	19.7	31.5
R.m.s deviations			
Bond lengths (Å)	0.009	0.008	0.006
Bond angles (°)	1.01	1.04	0.84
Ramachandran			
Percentage of favored	97.9	97.4	97.4
Percentage of outliers	0.5	0.0	0.0

Note: Values in parentheses are for the highest-resolution shell.

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$.

^b $R_{\text{work}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are observed and calculated structure factors, respectively.

^c R_{free} calculated using 5% of total reflections randomly chosen and excluded from the refinement.

substitutions relative to Erbin-PDZ, which would involve many more possible evolutionary pathways.

In plotting the pathways from Erbin-PDZ to E-6c (Figure 4), we observed that the four single substitutions and five of the six double substitutions had only minor effects on specificity, as the logos resembled the Erbin-PDZ logo except that some exhibited class II specificity at position⁻² and two exhibited a preference for Gly rather than Asp/Glu at position⁻³. One of the double substitutions produced a domain that was apparently nonfunctional for C-terminal peptide recognition, as no binding peptides were selected. Two of the four triple substitutions were also apparently nonfunctional, whereas the other two exhibited specificity profiles that resembled those of class II PDZ domains. Strikingly, the addition of a fourth substitution to yield E-6c resulted in a specificity that differed dramatically from the specificities of the two functional domains bearing triple substitutions, and

notably, generated functional domains from the two apparently nonfunctional domains with triple substitutions. These results show that the evolution of specificity in this case is not a cumulative or predictable process, as the specificities of the intermediates between Erbin-PDZ and E-6c differ greatly from that of E-6c.

2.4 | Structural analysis of Erbin-PDZ, E-14, and an intermediate variant

To gain insights into the molecular basis for altered specificity, we crystallized Erbin-PDZ, E-14, and the intermediate E-6a in complex with optimal peptides. We solved the X-ray crystal structures of Erbin-PDZ bound to GYETWV_{COO-}, E-14 bound to YYESDWL_{COO-}, and E-6a bound to YYESGWL_{COO-} at 1.2, 1.2, and 2.0 Å resolution, respectively

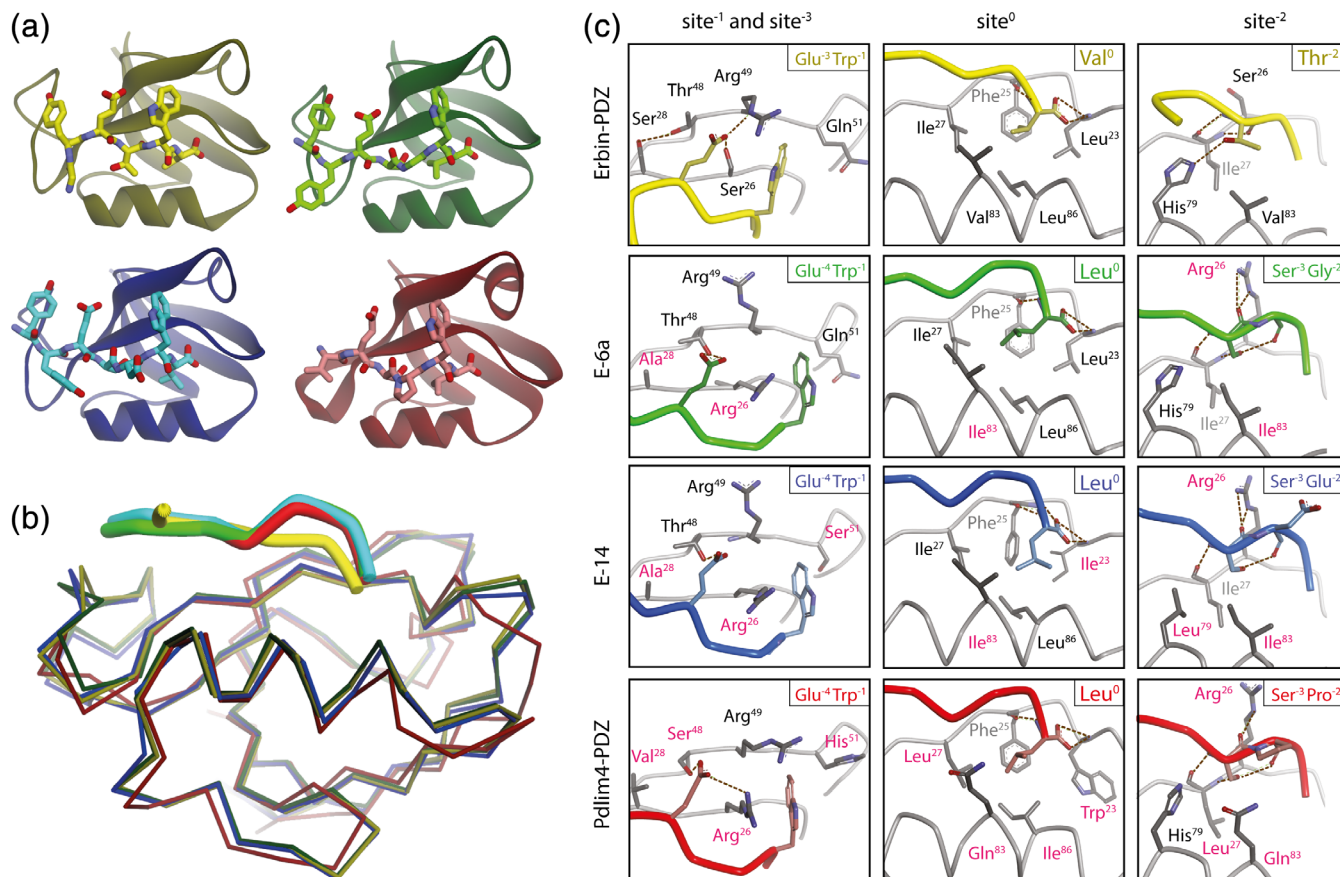


FIGURE 5 Crystal structures of PDZ-peptide complexes. (a) Structures of Erbin-PDZ (yellow), E-6a (green), E-14 (blue), and Pdlim4-PDZ (PDB entry 4Q2O, red)⁹ with peptide ligands depicted in lighter colors. The PDZ domain mainchain is shown as a ribbon and peptide side chains are shown as sticks. (b) Structural superposition of the mainchains of the four PDZ domains (sticks) and peptide ligands (tubes), colored as in panel A. (c) Details of the molecular interactions between PDZ domains and peptide ligands at site⁻¹ and site⁻³ (left), site⁰ (center), and site⁻² (right). PDZ domains are colored gray and side chains contributing to specificity are shown as sticks, with labels according to the PDB file 1N7T²⁹ colored black or pink if the sequences match or do not match the sequence of Erbin-PDZ, respectively. Peptide ligands are colored as in panel A

(Table 1, Figure 5a). The main chains of Erbin-PDZ and its two variants superposed almost perfectly and showed pairwise RMSD values below 0.7 Å, indicating that the main chains were virtually identical, and also superposed well with Pdlim4-PDZ (Figure 5b). As observed in numerous other PDZ-ligand structures,⁹ each peptide ligand bound in a groove between helix $\alpha 2$ and strand $\beta 2$, and the C-terminal carboxylate was coordinated by the “carboxylate-binding” loop that precedes strand $\beta 2$. However, whereas the Erbin-PDZ ligand exhibited a canonical extended main chain conformation, the main chains of E-6a and E-14 exhibited noncanonical bent conformations that closely resembled the conformation of the Pdlim4-PDZ ligand VESPWL_{COO-} (Figure 5b). Like other canonical PDZ domains, the Erbin-PDZ peptide-binding site contains four subsites and each of these interacts specifically with one of the last four residues of the peptide ligand (Figure 5a). In contrast, as observed previously for Pdlim4-PDZ,⁹ the bent main chain of the ligands for E-14 and E-6a causes a change in the register between PDZ subsites and ligand positions, and consequently, unlike

Erbin-PDZ in which site⁻² is occupied only by ligand position⁻², site⁻² is occupied by ligand positions -2 and -3 . This rearrangement of the peptide main chain alters molecular interactions and specificity across the entire binding site.

As detailed in Figure 5c, the molecular interactions between the Erbin-PDZ variants and their optimal peptides closely resemble those of Pdlim4-PDZ rather than Erbin-PDZ. Amongst the substitutions in the variants relative to Erbin-PDZ, the only one that matches the sequence of Pdlim4-PDZ is the Ser to Arg substitution at position 26, and notably, the Arg²⁶ side chain plays a major role in altering the ligand main chain conformation. In the structures of Pdlim4-PDZ, E-14 and E-6a, the Arg²⁶ side chain interacts with the main chain of the position⁻³ residue and packs against the side chain of Trp⁻¹ to promote a bend in the ligand main chain (Figure 5c, left). In contrast, the small Ser²⁶ side chain of Erbin-PDZ does not interact with Trp⁻¹, and consequently, does not affect the ligand main chain conformation. As a result of the bent ligand main chain, the interactions at site⁻³ are substantially altered in Erbin-PDZ

compared with the other three domains. In each case, site⁻³ is occupied by a glutamate side chain, but this side chain resides in position⁻³ in the case of the Erbin-PDZ ligand and in position⁻⁴ in case of the ligands for the other three domains. At site⁰, Erbin-PDZ prefers a Val⁰ side chain whereas the other domains prefer a larger Leu⁰ side chain, and these preferences can also be attributed to the differences in the ligand main chain conformations. In the case of Erbin-PDZ, the canonical extended conformation of the ligand main chain positions the Val⁰ side chain so that it is buried in a site⁰ pocket that does not appear to be able to accommodate larger side chains. Due to a bent main chain, the ligands for Pdlim4-PDZ, E-14, and E-6a position the Leu⁰ in a conformation where the side chain is oriented towards helix $\alpha 2$ due to different rotamer conformations (Figure 5c, center). Finally, at site⁻², Erbin-PDZ displays canonical Class I specificity, which is defined by a hydrogen-bonding interaction between His⁷⁹ in helix $\alpha 2$ and Thr⁻² in the peptide. In contrast, in the other three PDZ domains, the bent main chain results in non-canonical interactions that place ligand positions -2 and -3 at site⁻² (Figure 5c, left).

3 | CONCLUSIONS

Using synthetic PDZ domains, we have dissected the minimal molecular changes required to transition between the highly divergent specificities of the natural domains Erbin-PDZ and Pdlim4-PDZ. We used peptide-phage display to profile the specificities of all mutational transitions between Erbin-PDZ and E-14, a variant containing six substitutions that endow a specificity virtually identical to that of Pdlim4-PDZ. This analysis revealed that three substitutions create an intermediate variant E-6a that exhibits two distinct specificities, one similar to that of Erbin-PDZ and the other similar to that of E-14/Pdlim4-PDZ. Four or more substitutions are sufficient to fully convert specificity from that of Erbin-PDZ to that of E-14/Pdlim4-PDZ. A survey of all evolutionary paths defined by four substitutions showed that the evolution of E-14/Pdlim4-PDZ specificity from Erbin-PDZ specificity is not an additive or predictable process, but rather, the final specificity arises suddenly upon the introduction of a fourth substitution in addition to three substitutions in domains that either did not bind peptides or exhibited specificities similar to that of Erbin-PDZ.

Our results provide evidence that the evolution of PDZ domain specificity follows an epistatic trajectory in which several nonadditive substitutions must be combined to give rise to a new function. The evolution of protein function through epistatic trajectories has been described for several enzymes,¹⁴⁻¹⁶ including glucocorticoid receptor protein¹⁷ and chalcone isomerase,¹⁸ suggesting that this may be a general principle of protein evolution. We complemented

specificity profiling with structural comparisons of Erbin-PDZ, E-14, E-6a, and Pdlim4-PDZ in complex with optimal ligands. These comparisons showed that the main chains of ligands for E-14 and E-6a exhibited a non-canonical bent conformation that resembled the conformation of the ligand for Pdlim4-PDZ. Notably, these three domains all contained an Arg²⁶ residue, whereas Erbin-PDZ contained a Ser residue at position 26, and the large Arg side chain appeared to be a major contributor to the bent ligand main chain, which in turn altered domain-ligand interactions across the peptide-binding site. Together, the structural and functional data provide a comprehensive view of protein evolution at the molecular level and exemplify how only a few changes in a binding site can dramatically alter specificity, and consequently, how divergent biological functions can evolve rapidly from a common protein framework.

4 | MATERIALS AND METHODS

4.1 | Specificity profiling by phage display

Erbin-PDZ and its variants were purified as GST fusion proteins, which were used as baits for binding selections with a library of random heptapeptides fused to the C-terminus of the gene-8 major coat protein of M13 phage, as described.^{19,20} For each PDZ domain, peptide ligands were aligned using the C-terminus as an anchor position and the alignment was used to derive a position weight matrix (PWM) that described the specificity profile of each domain with a simple statistical model. The PWM was constructed by calculating the distribution of amino acid residues found at each of the seven positions of the ligand and correcting for codon bias in the naïve library using an NNK codon correction, as described in Reference 20 and was visualized as a sequence logo.¹³

4.2 | Crystallography and structure determination

The following fusion protein was expressed and purified for each PDZ domain: a hexaHis tag, followed by GST, followed by a tobacco etch virus (TEV) protease cleavage site, followed by the PDZ domain. Protein was expressed and purified from *Escherichia coli*, as described.⁹ Briefly, cell pellets were lysed by sonication, clarified by centrifugation and purified using NiNTA agarose resin with elution in PBS, 400 mM imidazole. Eluate was dialysed into PBS overnight at 4°C and protein was cleaved with TEV protease. The sample was concentrated and the PDZ domain was separated from GST by gel filtration using a

Superdex 75 10/60 GL column (GE Healthcare) equilibrated in 20 mM Hepes pH 7.5, 200 mM NaCl. PDZ domain fractions were collected and concentrated to 13.6, 11.5, and 6.1 mg/ml for Erbin-PDZ, E-14 or E6a, respectively. Protein samples were flash frozen in liquid nitrogen and stored at -80°C for crystallization trials.

For crystallization, samples were thawed rapidly and a twofold molar excess of synthetic peptide ligand (Genscript) was added. Crystals were obtained by vapor diffusion in sitting drops at 22°C . Erbin-PDZ/GYETWV_{COO-} complex crystals were grown in a crystallization liquor containing 6% PEG3350, 100 mM MgCl₂, 100 mM sodium acetate, pH 4.5. E-14/YYESDWL_{COO-} complex crystals were grown in a crystallization liquor containing 9% PEG8000, 200 mM MgCl₂, 100 mM sodium acetate, pH 4.5. E-6a/YYESGWL_{COO-} complex crystals were grown in a crystallization liquor containing 14% PEG3350, 200 mM MgCl₂, 100 mM MES, pH 6.0. Crystals were cryoprotected in the same buffer containing 25% ethylene glycol and flash-frozen in liquid nitrogen prior to data collection.

Diffraction data for the Erbin-PDZ/GYETWV_{COO-}, E-14/YYESDWL_{COO-} and E-6a/YYESGWL_{COO-} crystals were collected on beamline 24-ID-E (NE-CAT) at Argonne National Laboratories (Chicago). All data sets were processed with either HKL2000²¹ or MOSFLM,²² and were solved by molecular replacement using Phenix.Phaser²³ within the PHENIX crystallography suite,^{24,25} and subsequent model refinement and water picking was performed either automatically with Phenix.refine within the PHENIX crystallography suite or manually using the graphics program Coot.²⁶ Erbin-PDZ/GYETWV_{COO-}, and E-14/YYESDWL_{COO-} models were fully hydrogenated and anisotropic B-factor refinement was performed for all protein and peptide heavy atoms. However, individual isotropic B-factor refinement with TLS parameterization^{27,28} was used for refinement of the E-6a/YYESGWL_{COO-} model.



ACKNOWLEDGMENTS

This work was supported by an operating grant from the Canadian Institutes of Health Research (MOP-93684) awarded to S.S.S. We are grateful to I. Kurinov for technical assistance on the diffraction experiments performed at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

ORCID

Joan Teyra  <https://orcid.org/0000-0003-2816-339X>
 Sachdev S. Sidhu  <https://orcid.org/0000-0001-7755-5918>

REFERENCES

- Bhattacharyya RP, Reményi A, Yeh BJ, Lim WA. Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem.* 2006;75:655–680.
- Pawson T, Warner N. Oncogenic re-wiring of cellular signaling pathways. *Oncogene.* 2007;26:1268–1275.
- Zhang X, Perica T, Teichmann SA. Evolution of protein structures and interactions from the perspective of residue contact networks. *Curr Opin Struct Biol.* 2013;23:954–963.
- Andreani J, Guerois R. Evolution of protein interactions: From interactomes to interfaces. *Arch Biochem Biophys.* 2014;554:65–75.
- Letunic I, Doerks T, Bork P. SMART 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 2011;40:D302–D305.
- Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. *Science.* 2003;300:445–452.
- Sheng M, Sala C. PDZ domains and the organization of supra-molecular complexes. *Annu Rev Neurosci.* 2001;24:1–29.
- Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci.* 2001;114:3219–3231.
- Ernst A, Appleton BA, Ivarsson Y, et al. A structural portrait of the PDZ domain family. *J Mol Biol.* 2014;426:3509–3519.
- Tonikian R, Zhang Y, Sazinsky SL, et al. A specificity map for the PDZ domain family. *PLoS Biol.* 2008;6:e239.
- Stiffler MA, Chen JR, Grantcharova VP, et al. PDZ domain binding selectivity is optimized across the mouse proteome. *Science.* 2007;317:364–369.
- Ernst A, Sazinsky SL, Hui S, et al. Rapid evolution of functional complexity in a domain family. *Sci Signal.* 2009;2:ra50.
- Schneider TD, Stephens RM. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* 1990;18:6097–6100.
- Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. *Nat Rev Genet.* 2010;11:572–582.
- Starr TN, Thornton JW. Epistasis in protein evolution. *Protein Sci.* 2016;25:1204–1218.
- Miton CM, Tokuriki N. How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.* 2016;25:1260–1272.
- Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. Crystal structure of an ancient protein: Evolution by conformational epistasis. *Science.* 2007;317:1544–1548.
- Kaltenbach M, Burke JR, Dindo M, et al. Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat Chem Biol.* 2018;14:548–555.
- Held HA, Sidhu SS. Comprehensive mutational analysis of the M13 major coat protein: Improved scaffolds for C-terminal phage display. *J Mol Biol.* 2004;340:587–597.
- Tonikian R, Zhang Y, Boone C, Sidhu SS. Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat Protoc.* 2007;2:1368–1386.
- Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997;276:307–326.
- Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW. iMOSFLM: A new graphical interface for diffraction-image

- processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 2011;67:271–281.
23. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst*. 2007;40:658–674.
 24. Adams PD, Grosse-Kunstleve RW, Hung LW, et al. PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*. 2002;58:1948–1954.
 25. Zwart PH, Afonine PV, Grosse-Kunstleve RW, et al. Automated structure solution with the PHENIX suite. *Methods Mol Biol*. 2008;426:419–435.
 26. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr*. 2004;D60:2126–2132.
 27. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr*. 2006;62:439–450.
 28. Urzhumtseva L, Klaholz B, Urzhumtsev A. On effective and optical resolutions of diffraction data sets. *Acta Crystallogr D Biol Crystallogr*. 2013;69:1921–1934.
 29. Skelton NJ, Koehler MFT, Zobel K, et al. Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. *J Biol Chem*. 2003;278:7645–7654.
 30. Aasland R, Abrams C, Ampe C, et al. Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett*. 2002;513:141–144.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Teyra J, Ernst A, Singer A, Sicheri F, Sidhu SS. Comprehensive analysis of all evolutionary paths between two divergent PDZ domain specificities. *Protein Science*. 2020;29:433–442. <https://doi.org/10.1002/pro.3759>