Alteration of the C-Terminal Ligand Specificity of the Erbin PDZ Domain by Allosteric Mutational Effects

Javier Murciano-Calles [1], Megan E. McLaughlin [2], Ariel Erijman [3], Yogesh Hooda [2], Nishant Chakravorty [4], Jose C. Martinez [1], Julia M. Shifman [3] and Sachdev S. Sidhu [2]

*1 - **Departamento de Química Física e Instituto de Biotecnología,** Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain*
*2 - **Banting and Best Department of Medical Research and Department of Molecular Genetics,** University of Toronto, The Donnelly Centre, 160 College Street, Toronto, Ontario, M5S 3E1, Canada*
*3 - **Department of Biological Chemistry,** The Alexander Silberman Institute of Life Sciences, Edmond J. Safra Campus Givat Ram, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel*
*4 - **Institute of Health and Biomedical Innovation,** Queensland University of Technology, Brisbane QLD 4000, Australia*

***Correspondence to Julia M. Shifman and Sachdev S. Sidhu:** jshifman@mail.huji.ac.il; sachdev.sidhu@utoronto.ca*
http://dx.doi.org/10.1016/j.jmb.2014.05.003
Edited by A. Keating

Abstract

Modulation of protein binding specificity is important for basic biology and for applied science. Here we explore how binding specificity is conveyed in PDZ (*p*ostsynaptic density protein-95/*d*iscs large/*z*onula occludens-1) domains, small interaction modules that recognize various proteins by binding to an extended C terminus. Our goal was to engineer variants of the Erbin PDZ domain with altered specificity for the most C-terminal position (position 0) where a Val is strongly preferred by the wild-type domain. We constructed a library of PDZ domains by randomizing residues in direct contact with position 0 and in a loop that is close to but does not contact position 0. We used phage display to select for PDZ variants that bind to 19 peptide ligands differing only at position 0. To verify that each obtained PDZ domain exhibited the correct binding specificity, we selected peptide ligands for each domain. Despite intensive efforts, we were only able to evolve Erbin PDZ domain variants with selectivity for the aliphatic C-terminal side chains Val, Ile and Leu. Interestingly, many PDZ domains with these three distinct specificities contained identical amino acids at positions that directly contact position 0 but differed in the loop that does not contact position 0. Computational modeling of the selected PDZ domains shows how slight conformational changes in the loop region propagate to the binding site and result in different binding specificities. Our results demonstrate that second-sphere residues could be crucial in determining protein binding specificity.

© 2014 Elsevier Ltd. All rights reserved.

Introduction

PDZ (*p*ostsynaptic density protein-95/*d*iscs large/*z*onula occludens-1) domains are among the most abundant protein–protein interaction domains in multicellular organisms. They help to organize signaling complexes and regulate trafficking of receptors and ion channels by acting as protein scaffolds with diverse binding partners. PDZ domains are composed of seven β-strands and one or two α-helices, and they generally recognize protein C termini, which bind along a groove formed by the α2 helix and the β2 strand (Fig. 1). In early studies, two specificity classes were postulated based on two ligand positions, class 1 [X-(T/S)-X-φ_{COOH}] and class 2 (X-φ-X-φ_{COOH}), where "X" is any amino acid and "φ" is a hydrophobe [1,2]. Thus, according to the accepted nomenclature, the initially postulated specificity classes were based on the recognition of ligand position 0 (C-terminal residue) and position −2. However, a later study showed that any of the last seven ligand residues can potentially interact with the PDZ domain, and, accordingly, PDZ domains were grouped into at least 16 specificity classes [3]. This study also suggested that specificities for positions 0 and −2 were mainly affected by mutations at PDZ positions close to the ligand residue, whereas specificities for positions −1 and

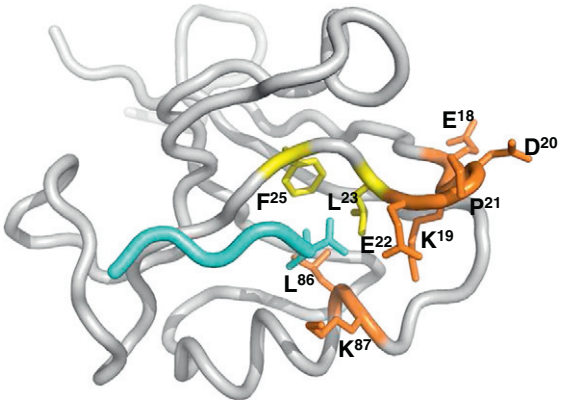

Fig. 1. Erbin-PDZ library design. The Erbin-PDZ and peptide ligand ($TGWETWV_{COOH}$) main chains are shown as gray and cyan tubes, respectively. The C-terminal side chain of the peptide ligand is shown. The Erbin-PDZ residues that were diversified in the library are labeled and shown as sticks colored orange or yellow if they were allowed to vary as all 20 genetically encoded amino acids or as predominantly hydrophobic amino acids, respectively. The figure was generated using PyMOL (http://www.pymol.org/) using NMR structure coordinates (PDB entry 1N7T), and residue numbering corresponds to that in the PDB file.

–3 were affected by mutations throughout the peptide-binding site. Thus, it seemed likely that the structural basis for ligand specificity may be easiest to decipher for sites 0 and –2, given that local effects mediated by direct contacts are easier to rationalize and engineer.

Here we have assessed the range of specificities that can be accommodated within site 0 of a single PDZ domain and we have explored whether specificity can be altered by mutations at positions that are in direct contact with the ligand or by conformational changes caused by mutations at distal positions. We chose the human Erbin PDZ domain (Erbin-PDZ) as the model system, as Erbin-PDZ has been extensively studied structurally and by mutagenesis [4–8]. We constructed a combinatorial phage-displayed library of Erbin-PDZ variants with mutations within and around site 0, and we selected for variants that bind to a set of 19 peptide ligands that differed only at the C terminus and represented all genetically encoded amino acids except proline. Despite extensive efforts, we were only able to evolve Erbin-PDZ variants with selectivity for aliphatic C-terminal side chains, and thus, since wild-type (wt) Erbin-PDZ prefers Val but tolerates Leu and Ile at site 0, only minor changes in specificity were obtained. Surprisingly, we found that changes in specificity appeared to generally arise from conformational changes caused by mutations at positions that do not contact the ligand. Our results suggest that more dramatic changes in PDZ domain site 0 specificity will likely require changes not only at positions that contact the ligand but also at distal positions that influence the main-chain conformation of the domain.

Results

Design and construction of the Erbin-PDZ library

We inspected the structure of Erbin-PDZ in complex with a high-affinity peptide ligand ($TGWETWV_{COOH}$) to identify residues that might influence site 0 specificity and, thus, would be candidates for mutagenesis to alter specificity (Fig. 1). We identified three side chains that make contact with the side chain of the C-terminal Val of the ligand (Leu23, Phe25, Leu86), and we defined this set as the site 0 binding pocket. In addition, we identified five contiguous residues (Glu18–Glu22) that form a loop (loop 18–22) that immediately precedes Leu23, and we reasoned that mutations in this region might alter the conformation of the site 0 pocket through indirect conformational effects. Finally, we identified the solvent-exposed Lys87, which precedes Leu86 in helix α2 and could potentially be recruited for C-terminal ligand recognition if there were substantial main-chain conformational changes in the site 0 pocket region. A phage-displayed combinatorial library of Erbin-PDZ variants was constructed by replacing seven of these nine positions with degenerate codons encoding for all 20 natural amino acids. Positions 23 and 25 were replaced with a degenerate codon encoding mainly for hydrophobic amino acids because these positions reside in the carboxylate binding loop and are conserved as hydrophobes across the PDZ domain family [9].

Binding selections for Erbin-PDZ variants with altered specificities

Phage pools representing the Erbin-PDZ library were cycled through rounds of binding selections with a set of 19 peptides that differed in sequence from a high-affinity Erbin-PDZ ligand ($TGWETWV_{COOH}$) only at the C-terminal position. The 19 peptides together represented sequences with 19 of the 20 natural amino acids at the C terminus. Pro was excluded because the Pro side chain forms a bond with the peptide main chain, disrupting hydrogen bonding interactions between backbone nitrogen on the peptide and the PDZ domain. Because our goal was to select PDZ domains with specificity for a particular amino acid rather than promiscuous domains that could accept a wide variety of C-terminal residues, we performed the selections with each peptide of interest immobilized on a solid support and mixtures of other peptides in solution at high concentration. Under these conditions, we expected that promiscuous domains would bind to the competitor peptides in solution and only domains that were selective for the

peptide of interest would be retained. For most selections, the competitive peptide mixture consisted of equal ratios of peptides terminating with Val, Leu or Ile because the vast majority of PDZ domains (including Erbin-PDZ) prefer these C-terminal side chains [3]. For selections for PDZ domains selective for Val, Leu or Ile, the competitive peptide mixture only contained peptides terminating with the two aliphatic amino acids other than the one being selected for.

After the final round of each binding selection, phage from 48 individual clones were subjected to enzyme-linked immunosorbent assays (ELISAs) with immobilized cognate peptide to detect phage-displayed Erbin-PDZ variants that exhibited positive binding signals, which we arbitrarily defined as optical densities greater than 1 (Fig. 2). With this definition, no positive binding clones were obtained for 14 of the peptides, and the five peptides that yielded positive binding clones all terminated in hydrophobic residues (Val, Leu, Ile, Met or Phe).

DNA sequencing of the positive clones revealed 13 unique sequences each for the Val, Leu and Ile C-terminal peptides and four or two unique sequences for the Met or Phe C-terminal peptides, respectively (Supplementary Fig. 1). However, many of the sequences contained one or two Cys residues, which often give rise to artifactual binding of phage particles due to the formation of intermolecular or intramolecular disulfide bonds between or within displayed protein molecules. In particular, all of the sequences isolated for binding to Phe or Met C-terminal peptides contained two Cys residues, and these PDZ domains could not be purified as free proteins (data not shown). Thus, we removed all Cys-containing sequences from our data set, and we were left with positive binding clones only for peptides terminating in Val, Leu or Ile (Fig. 3). Notably, these three peptides represent the C-terminal ligands that are accepted by wt Erbin-PDZ [8].

Unexpectedly, the sequences of the variants did not differ greatly at the positions that line the site 0 pocket, regardless of differences in the C-terminal ligand used for the binding selections. In particular, virtually all of the variants contained a Leu-to-Phe

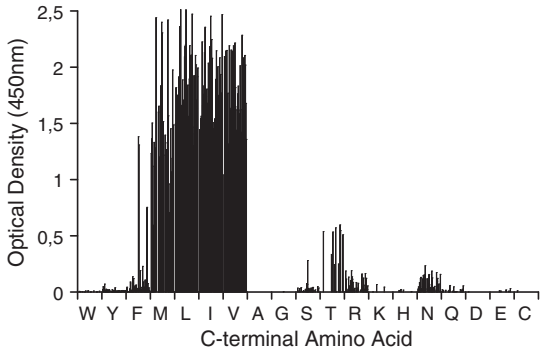

Fig. 2. Binding analysis of phage-displayed Erbin-PDZ variants. Following four rounds of binding selections, phage ELISAs were used to detect binding of 48 individual phage-displayed Erbin-PDZ variants to the peptide against which they were selected (TGWETWX$_{COOH}$, where "X" is indicated on the *x*-axis). Bound phage were detected spectrophotometrically (optical density at 450 nm, *y*-axis).

(a)

	18	19	20	21	22	**23**	**25**	**86**	87
WT	E	K	D	P	E	**L**	**F**	**L**	K
V-1	P	S	R	-	G	**F**	**-**	**-**	R
V-2	G	D	R	R	S	**F**	**-**	**M**	R
V-3	M	T	G	G	L	**F**	**-**	**F**	R
V-4	R	T	R	A	S	**-**	**-**	**-**	R
V-5	M	N	R	P	K	**F**	**-**	**-**	R
V-6	V	P	R	R	T	**F**	**-**	**-**	R

(b)

	18	19	20	21	22	**23**	**25**	**86**	87
WT	E	K	D	P	E	**L**	**F**	**L**	K
L-1	A	N	R	G	R	**F**	**-**	**F**	R
L-2	P	R	S	-	R	**F**	**-**	**F**	R
L-3	P	G	G	R	Y	**F**	**-**	**M**	R
L-4	V	P	R	K	Y	**F**	**-**	**I**	N
L-5	G	R	-	R	K	**F**	**-**	**F**	R
L-6	N	R	N	A	V	**F**	**-**	**F**	R
L-7	G	R	-	R	K	**F**	**-**	**I**	R
L-8	S	M	K	-	R	**F**	**-**	**I**	R
L-9	R	S	R	-	V	**F**	**-**	**F**	-

(c)

	18	19	20	21	22	**23**	**25**	**86**	87
WT	E	K	D	P	E	**L**	**F**	**L**	K
I-1	T	Q	P	R	N	**F**	**-**	**-**	R
I-2	V	P	S	S	R	**F**	**-**	**-**	R
I-3	R	V	H	S	R	**F**	**-**	**-**	R
I-4	R	A	R	R	D	**F**	**-**	**-**	R
I-5	G	S	P	R	G	**F**	**-**	**-**	R
I-6	A	S	R	G	R	**F**	**-**	**-**	R
I-7	T	T	Q	R	N	**F**	**-**	**-**	R
I-8	P	R	G	S	R	**F**	**-**	**-**	R
I-9	S	L	K	R	N	**F**	**-**	**-**	R
I-10	L	G	R	G	R	**F**	**-**	**-**	R
I-11	T	H	R	R	G	**F**	**-**	**-**	R
I-12	F	P	P	R	G	**F**	**-**	**-**	R
I-13	L	D	R	G	R	**F**	**-**	**-**	R

Fig. 3. Sequence alignment of Erbin-PDZ variants. Only those positions that were diversified in the library are shown for Erbin-PDZ variants that were selected and shown to bind peptides containing a C-terminal (a) Val, (b) Leu or (c) Ile residue. Positions that were conserved as the wt are indicated by dashes and residues that contact the C-terminal ligand side chain are boxed. The name of each variant is indicated to the left.

mutation at position 23 and all variants conserved the wt Phe residue at position 25. At position 86, there was greater diversity among variants selected for binding to the Leu0 ligand, as this position was occupied by Phe, Met or Ile. However, this position was also more diverse among variants selected for binding to the Val0 ligand, which contained Leu, Met or Phe at this position. Most surprisingly, all of the variants selected for binding to the Ile0 ligand contained a wt Leu86, and in fact, all of these variants are identical at all three site 0 positions. These site 0 sequences are identical with the most common sequences observed among the variants selected for binding to Val0. Thus, taken together, these results show that alteration of Erbin-PDZ specificity from the wt preference for Val0 to a preference for Ile0 cannot be explained by changes in the direct contacts in site 0.

Among the mutated Erbin-PDZ residues that are not in direct contact with the C-terminal ligand side chain, position 87 is also not involved in altering specificity because essentially all of the variants contain a Lys-to-Arg mutation at this position. In contrast, the loop that precedes the site 0 contact positions 23 and 25 (loop 18–22) is heavily mutated in all of the variants, and the wt sequence occurs only rarely within this stretch. There is no clear consensus among the mutated sequences in this region, but nonetheless, we can conclude that mutations in this region must be responsible for alterations in specificity for variants that prefer Ile0 in contrast to the wt preference for Val0. This must be the case because all of the variants selected for binding to the Ile0 ligand are identical at positions 23, 25, 86 and 87, and these sequences are also identical with those of three variants selected for binding to the Val0 ligand (Fig. 3).

Specificity profiling of Erbin-PDZ variants

We decided to explore in more detail the Erbin-PDZ variants selected for binding to the Ile0 ligand. The wt Erbin-PDZ prefers Val0 over Ile0 by approximately 2 orders of magnitude [8]. Thus, we ascertained whether the selected variants exhibited in fact an altered specificity that favors Ile0 over Val0. We accomplished this by purifying the Erbin-PDZ variants as glutathione *S*-transferase (GST) fusions and using the purified proteins as targets for binding selection experiments with phage-displayed libraries of random C-terminal peptides. We selected phage-displayed peptide ligands for 11 of the Erbin-PDZ variants selected for binding to the Ile0 peptide ligand, and we also analyzed the wt and three Erbin-PDZ variants that were selected for binding to the Val0 ligand and differ in sequence at position 86 (V-1, V-2 and V-3).

In all cases, the specificity profiles were very similar for positions upstream of position 0, indicating that the mutations did not affect the subsites beyond site 0 (Fig. 4). All of the variants selected for binding to the Ile0 ligand exhibited a preference for peptides with a C-terminal Ile residue. In contrast and in agreement with previous results, the wt Erbin-PDZ preferred peptides terminating with a Val residue. Moreover, the three variants selected for binding to the Val0 ligand also exhibited a preference for peptides terminating with a Val residue. Notably, variant V-1 exhibits a strong preference for peptides terminating with a Val residue but, at positions 23, 25, 86 and 87, it is identical with all the variants that exhibit a preference for ligands terminating with an Ile residue. These results show that the selection process was effective in enriching for PDZ domains with selectivity toward specific C-terminal sequences, and they show that altered specificities that favor binding to Ile0 ligands rather than Val0 ligands arise from alterations in loop 18–22, which does not contact the peptide ligand.

Computational modeling

To better understand the specificity profiling results, we attempted to reproduce the phage selection results computationally, using an atomic-based force field to model interactions between the PDZ domains and the peptides [10]. We modeled the structures of wt Erbin-PDZ, two variants that prefer Val0 ligands (V-1 and V-2) and four variants that prefer Ile0 ligands (I-1, I-2, I-3 and I-4). For each PDZ domain, we generated an ensemble of 20 structures by introducing mutations into the NMR structural ensemble of wt Erbin-PDZ bound to a high-affinity ligand (PDB entry 1N7T). We then computationally designed the best possible amino acid sequence for the three last positions of the peptide ligand for each PDZ domain. In each case, the experimentally determined specificity for position 0 was well reproduced by our design calculations. The correct specificity was also reproduced for positions −1 and −2, where our calculations mostly selected Trp or Thr, respectively.

To explore how site 0 binding pockets that are lined with identical residues at positions 23, 25 and 86 can nonetheless exhibit distinct preferences for either Val0 or Ile0, we modeled the structures of the PDZ variants V-1 and I-1 bound to Val0 and Ile0 ligands, respectively. V-1 and I-1 were chosen for modeling since they exhibit the strongest preferences for phage-displayed peptides that terminate in either Val or Ile, respectively (Fig. 4). In our modeling, we first threaded the sequences of the V-1 and I-1 variants and the corresponding peptides onto the same structure of the wt PDZ domain. We then performed a 10-ns molecular dynamics (MD) simulation to see how the two structures diverge to accommodate a different set of mutations. Structural superposition of the two PDZ domain models after MD simulations is shown in Fig. 5 and the last nanosecond of simulation is available as movies in Supplementary Information (Ile.gif and Val.gif for I-1 and V-1, respectively). Our modeling shows that I-1 and V-1 PDZ variants exhibit

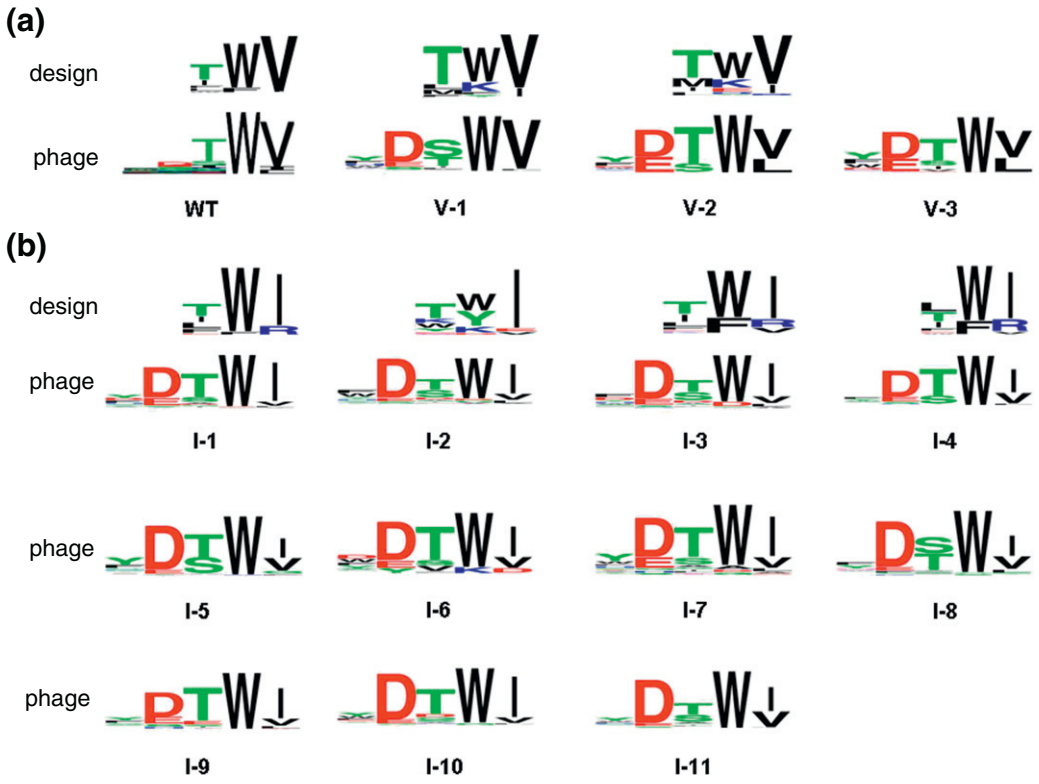

Fig. 4. Specificity profiles for Erbin-PDZ variants. The specificity profiles are shown as sequence logos, determined either by computational modeling (design) or by peptide-phage display (phage) for Erbin-PDZ variants selected for binding to peptides containing a C-terminal (a) Val or (b) Ile residue. The name of each Erbin-PDZ variant is shown below the logos.

similar backbone conformations for most of the protein. However, substantial conformational changes occur in the loop 18–22 region where the mutations were introduced. In addition, loop 51–59, which contacts loop 18–22, assumes different conformations and dynamics. The side chains of the site 0 pocket residues 25 and 86 exhibit similar conformations in both variants. However, Phe23, the other residue that directly contacts the C-terminal ligand side chain, shows a distinct rotation of 60° between the two PDZ variants (Fig. 5 and Supplementary Fig. 2). Phe23 rearrangement seems to be the most crucial change in defining the binding specificity of the PDZ domain for Val0 or Ile0. This re-orientation of Phe23 allows it to make better van der Waals interactions with the neighboring residues 93 and 19, which do not contact the ligand itself. Position 93 is occupied by the same amino acid (Val) in both PDZ variants, but the side chain adapts different conformations in the two models. Position 19 is occupied by Gln or Ser in the I-1 or V-1 variants, respectively, and these side chains make contacts with the side chain of Phe23. Thus, the models suggest that the second-sphere residues 19 and 93 bracket Phe23 and the sequence difference at position 19 may be critical for altering the conformation of the Phe23 side chain and consequently the specificities of the different variants. Thus, our simulations help to elucidate how different ligand preferences at site 0 arise from alterations in second-sphere residues that are not in direct contact with the ligand but, nonetheless, create favorable binding pockets for Val0 or Ile0 ligands through allosteric effects that involve loop reorganization and consequent alterations of side-chain conformations within the site 0 pocket.

Discussion

We attempted to engineer Erbin-PDZ to obtain a variety of binding specificities for the most C-terminal ligand position. In our experiments, we explored the sequences of billions of PDZ domains that varied at positions directly contacting the ligand position 0 or in a loop adjacent to the site 0 binding pocket. We found that only three different C-terminal amino acids (Val, Ile, Leu) could be selectively recognized by the PDZ domains in our library, and these three aliphatic residues are very similar in size and chemical properties. Notably, the same three residues could be accommodated by the wt Erbin-PDZ, although Val is preferred [8]. The inability to switch the specificity to

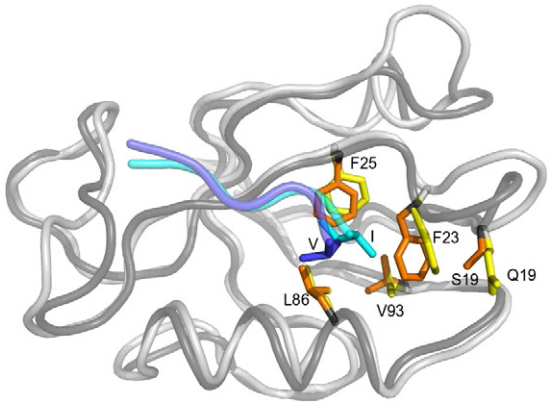

Fig. 5. Superposition of structures for Erbin-PDZ variants V-1 and I-1 bound to Val0 and Ile0 peptides, respectively, generated by the MD simulation. The main chains of V-1 and its peptide ligand are colored gray and blue, respectively, while those of I-1 and its peptide ligand are colored white and cyan, respectively. Side chains are shown as sticks for the C-terminal residue of each peptide ligand and for Erbin-PDZ residues that line site 0 (positions 23, 25 and 86) or contact Phe23 (positions 19 and 93). Side chains of residues at positions 23, 25, 86, 19 and 93 are colored orange or yellow in V-1 or I-1, respectively. Particular structures for the I-1 and V-1 variants were chosen from the main cluster of MD-generated structures with the lowest RMSD of C^{α} atoms to the initial structures for a better graphical representation. For more details, see movies Ile.gif and Val.gif in the supplementary information.

polar or charged amino acids could be explained by the highly hydrophobic nature of the site 0 pocket, which was not altered in our experiments. The incapacity to accommodate larger aromatic amino acids might be explained by the size limit of the binding pocket. In a previous work, we produced a library of Erbin-PDZ variants by mutating residues across the entire peptide-binding site and, again, we found that specificity at site 0 was restricted to hydrophobes and heavily favored aliphatic residues [6].

Natural PDZ domains recognize invariably hydrophobic residues at site 0, predominantly aliphatic residues, but sometimes Phe, Trp or Cys. Moreover, we have shown that PDZ domain specificity has remained highly conserved across one billion years of evolution that separate worm from human, since specificity profiles of orthologous PDZ domains were very similar for these species [3]. McLaughlin *et al.* have also performed a mutational study of the third PDZ domain of PSD95, and they found that only chemically conservative substitutions for residues in contact with the C terminus of the ligand maintained binding [11]. Thus, based on our current results and these previous studies, we conclude that specificity for aliphatic residues at site 0 is a common, conserved feature of natural PDZ domains, and even with combinatorial mutagenesis, these specificity features can only be altered slightly. Another possible reason for this conserved specificity could be the PDZ fold itself. The site 0 pocket is part of the hydrophobic core of the PDZ domain, and it is likely that altering the pocket might disrupt folding.

However, it must be noted that our failure to obtain more dramatic changes in specificity does not mean that other specificities cannot be accommodated by the PDZ domain fold, since the negative results may be due to the nature of the library design or the selection strategy. In particular, it might be possible to accommodate more polar side chains in the site 0 pocket if the hydrophobic residues in the PDZ core are allowed to vary more freely, although this may require some compromise in stability. Also, it could be that our selection under highly stringent conditions with competitors in solution may have eliminated weak or promiscuous binders that may have been able to recognize polar C-terminal side chains, albeit with reduced affinity or specificity. Indeed, Fig. 2 does show some weak binding signals for PDZ variants selected for binding to ligands terminating in Ser, Thr, Arg or Asn and, in retrospect, it may have been worthwhile to follow up on these clones that may represent low-affinity binders for these peptides. These issues may be addressed in future studies with new library designs and less stringent selection strategies. New library designs will benefit from further insights into the determinants of site 0 specificity that have been gleaned from our recent comprehensive structural survey of the PDZ domain family [12].

Our results also show that specificity switch from Val to Ile occurs not due to mutations in direct contact with the ligand but due to allosteric effects caused by mutations that are adjacent to the site 0 pocket but do not directly contact the peptide ligand. The specificity switch requires only subtle structural changes, including a re-orientation of the Phe23 side-chain and backbone conformation changes in the β1–β2 loop. Interestingly, in the CAL PDZ domain, side-chain orientation of an Ile residue at a position equivalent to Phe23 was found to alter ligand specificity from Ile0 to Leu0 [13]. In another study, substitution of Pro for a Gly residue in the β1–β2 loop of the Par-6 PDZ domain disrupted binding to Cdc42, most likely due to reorganization of the loop [14]. Hence, our results together with these previous results point to the importance of allosteric effects in ligand recognition by PDZ domains [11,15–17].

Many mutagenesis studies aimed at altering binding specificity focused on residues at the binding interface [18–20]. Our experiments confirm previous suggestions that binding specificity not only is a property of the binding interface but also can be mediated by residues that are distal from the binding site [11,16,21]. Such long-range cooperativity has been demonstrated in several protein families with very different folds, including G-protein-coupled receptors, the chymotrypsin class of serine proteases, hemoglobin, guanine nucleotide-binding proteins and

dihydrofolate reductase [21–23]. In our study, we have seen that cooperativity is not due to large changes in backbone structure but rather is caused by subtle conformational changes of side-chain rotamers, as has been reported elsewhere [15,24,25].

Materials and Methods

Construction of an Erbin-PDZ phage-displayed library

The Erbin-PDZ phage-displayed library was constructed using a previously described phagemid and methods [6,8]. A "stop template" version of the phagemid was constructed by substituting a TAA stop codon in place of each codon to be diversified. A mutagenesis reaction was performed with the stop template and mutagenic oligonucleotides designed to replace each TAA stop codon with a degenerate codon. Positions 18, 19, 20, 21, 22, 86 and 87 were replaced by degenerate NNK codons (N = A/C/G/T, K = G/T) that encode for all 20 genetically encoded amino acids, and positions 23 and 25 were replaced by degenerate NYT codons (Y = C/T) that encode for eight amino acids (Phe, Ile, Leu, Val, Pro, Ala, Ser, Thr), resulting in a total of 6×10^{10} possible combinations. The library contained approximately 10^9 unique members and, thus, did not cover all possible combinations. The amino acids are numbered according to the numbering used in the previously reported NMR structure of Erbin-PDZ (PDB entry 1N7T) [8].

Selection and analysis of peptide-binding Erbin-PDZ variants

Peptide ligands were purified as fusion proteins consisting of peptide sequences of interest fused to the C terminus of GST (GST-peptide). GST-peptide proteins were immobilized on 96-well Maxisorp Immunoplates (NUNC, Rochester, NY) and phage from the Erbin-PDZ library were cycled through six rounds of binding selection, as previously described [6]. To prevent non-specific PDZ domains from binding to the immobilized GST-peptide, we added a 1-μM solution of competing GST-peptide mixture to the phage in solution. The solution-phase GST-peptide solution was an equimolar mixture of Val0, Leu0 and Ile0 ligands for all selections, except the selections with an immobilized GST-peptide representing Val0, Leu0 or Ile0; in these selections, the solution-phase GST-peptide solution did not contain the GST-peptide that was immobilized. After the sixth round, 48 individual phage clones from each selection were assayed for binding to the peptide by phage ELISA [26]. Positive clones were subjected to DNA sequence analysis and unique sequences were aligned.

Purification of Erbin-PDZ variants

Individual Erbin-PDZ variants were expressed and purified as GST fusion proteins, as previously described [6]. In brief, individual variants from the binding selections were used as the template for a PCR that amplified DNA fragments encoding the PDZ domains. The DNA fragments were ligated into an expression phagemid to produce an open reading frame encoding a fusion protein (GST-PDZ) consisting of a hexa-histidine tag, followed by GST followed by an Erbin-PDZ variant. Protein expression and purification was performed in a 96-well format. Bacteria were pelleted and lysed, and the lysates were loaded onto PhyNexus tips containing Ni-NTA resin (PhyNexus, San Jose, CA). After washing, we eluted bound protein with elution buffer [50 mM phosphate (pH 8.0), 300 mM NaCl and 250 mM imidazole]. Protein purities were assessed by SDS-PAGE and were >90%. Protein concentrations were measured using a Bradford assay (Biorad, Hercules, CA) and the yields of purified protein ranged from 0.05 to 0.15 mg per 1.5 ml bacterial culture.

Specificity profiling of Erbin-PDZ variants

Peptide-phage selections were performed using a library of random heptapeptides (10^{11} unique members) fused to the C terminus of the gene-8 major coat protein of M13 phage, as previously described [6]. The binding selections were performed in a 96-well format with one well dedicated to each purified GST-PDZ protein. Phage pools representing the peptide-phage library were cycled through five rounds of binding selection against each immobilized PDZ domain. After five rounds of selection, specific binding of peptide phage was verified by phage ELISA [26] and positive phage were subjected to DNA sequence analysis. For each PDZ domain, unique DNA sequences were used to derive a set of unique peptide ligands that was used to create a binding profile statistical model as a position weight matrix.

Computational modeling

A protein design program ORBIT [10] was used for all protein design calculations. To model the structures of Erbin-PDZ variants, we started from the ensemble of 20 NMR structures for wt Erbin-PDZ bound to a high-affinity ligand (PDB ID 1N7T). In the first calculation, we modeled the structures of the Erbin-PDZ variants selected for binding to Val0 or Ile0 peptide ligands. Mutations were introduced into Erbin-PDZ by fixing the amino acid identity to the specified amino acid and repacking the surrounding residues. The position 0 identity of the peptide ligand was changed to Ile for Erbin-PDZ variants selected for binding to the Ile0 peptide ligand. The structures of the Erbin-PDZ variants were then minimized and served as input for further design calculations and MD simulations. In the second calculation, we designed the best amino acid sequence for positions 0, −1 and −2 of the peptide ligand when the peptide interacts with the corresponding Erbin-PDZ variant. In this calculation, positions 0, −1 and −2 were probed with all amino acids (excluding Pro, Gly and Cys) while all the contacting residues on the PDZ domain and the peptide were allowed to repack their side chains. A rotamer library based on the Dunbrack and Karplus library was used [27]. An empirical energy function containing terms for van der Waals interactions, hydrogen bonding, electrostatic interactions and solvation was used to calculate rotamer–backbone and rotamer–rotamer interactions [10,28,29]. In the energy function, we emphasized intermolecular interactions by using a bias factor of 1.5 as described in our previous study [30]. The lowest-energy peptide sequences were determined utilizing a search algorithm based on the Dead-End Elimination theorem [31]. Web logos were generated

from 20 designed sequences (one for each structure) using a WebLogo server [32].

MD simulations were performed using the GROMACS4 package [33,34]. The starting structures for variants I-1 and V-1 were the same as for the protein design calculations. We performed 1000 steps of steepest decent *in vacuo* minimization followed by 5000 steps of minimization of the solvated system. Solvent and ions were relaxed, keeping the protein atom positions restrained with 100 ps of isothermal–isochoric ensembles followed by 100 ps of isothermal–isobaric ensembles. MD simulations were performed at a temperature of 300 K for 10 ns and convergence was observed after approximately 1.5 ns. The movies showing the last 1000 steps of the MD simulations for variants I-1 and V-1 are available in the supplementary information (Ile.gif for variant I-1 and Val.gif for variant V-1).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.05.003.

Acknowledgements

M.E.M., N.C., Y.H. and S.S.S. were supported by a grant from the Canadian Institutes for Health Research (MOP-93684). J.M.S. acknowledges support from the Israel Science Foundation (1372/10) and the Abisch-Frenkel foundation. J.M.C. and J.C.M. acknowledge the financial support by the Andalusian Regional Government (grant CVI-05915) and the Spanish Ministry of Science and Education and FEDER (grants BIO2009-13261-C02-01 and BIO2012-39922-C02-01) for their financial support in this research. J.M.C. received a postdoctoral contract from the Spanish Ministry of Science and Education.

Received 18 November 2013;
Received in revised form 10 April 2014;
Accepted 2 May 2014
Available online 9 May 2014

Keywords:
protein interaction domains;
phage display;
protein engineering;
computational modeling;
binding cooperativity

Abbreviations used:
GST, glutathione S-transferase; wt, wild type; MD, molecular dynamics.

References

[1] Nourry C, Grant SGN, Borg J-P. PDZ domain proteins: plug and play! Sci STKE 2003;2003:re7.
[2] Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, et al. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 1997;275:73–7.
[3] Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh J-H, Reva B, et al. A specificity map for the PDZ domain family. PLoS Biol 2008;6:e239.
[4] Appleton BA, Zhang Y, Wu P, Yin JP, Hunziker W, Skelton NJ, et al. Comparative structural analysis of the Erbin PDZ domain and the first PDZ domain of ZO-1. Insights into determinants of PDZ domain specificity. J Biol Chem 2006;281:22312–20.
[5] Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, et al. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol Biosyst 2010;6:1782–90.
[6] Ernst A, Sazinsky SL, Hui S, Currell B, Dharsee M, Seshagiri S, et al. Rapid evolution of functional complexity in a domain family. Sci Signaling 2009;2:ra50.
[7] Laura RP, Witt AS, Held HA, Gerstner R, Deshayes K, Koehler MF, et al. The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of delta-catenin and ARVCF. J Biol Chem 2002;277:12906–14.
[8] Skelton NJ, Koehler MF, Zobel K, Wong WL, Yeh S, Pisabarro MT, et al. Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. J Biol Chem 2003;278:7645–54.
[9] Ivarsson Y. Plasticity of PDZ domains in ligand recognition and signaling. FEBS Lett 2012;586:2638–47.
[10] Dahiyat BI, Mayo SL. *De novo* protein design: fully automated sequence selection. Science 1997;278:82–7.
[11] McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature 2012;491:138–42.
[12] Ernst A, Appleton BA, Ivarsson Y, Zhang Y, Gfeller D, Wiesmann C, et al. A structural portrait of the PDZ domain family. J Mol Biol 2014 (submitted for publication).
[13] Amacher JF, Cushing PR, Bahl CD, Beck T, Madden DR. Stereochemical determinants of C-terminal specificity in PDZ peptide-binding domains: a novel contribution of the carboxylate-binding loop. J Biol Chem 2013;288:5114–26.
[14] Peterson FC, Penkert RR, Volkman BF, Prehoda KE. Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. Mol Cell 2004;13:665–76.
[15] Dhulesia A, Gsponer J, Vendruscolo M. Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. J Am Chem Soc 2008;130:8931–9.
[16] Gianni S, Haq SR, Montemiglio LC, Jorgens MC, Engstrom Ak, Chi CN, et al. Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity. J Biol Chem 2011;286:27167–75.
[17] Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. Proc Natl Acad Sci 2009;106:18249–54.
[18] Joachimiak LA, Kortemme T, Stoddard BL, Baker D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface. J Mol Biol 2006;361:195–208.
[19] Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, et al. Computer-aided design of a PDZ domain to recognize new target sequences. Nat Struct Biol 2002;9:621–7.
[20] Yosef E, Politi R, Choi MH, Shifman JM. Computational design of calmodulin mutants with up to 900-fold increase in binding specificity. J Mol Biol 2009;385:1470–80.
[21] Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Mol Biol 2003;10:59–69.

[22] Hatley ME, Lockless SW, Gibson SK, Gilman AG, Ranganathan R. Allosteric determinants in guanine nucleotide-binding proteins. Proc Natl Acad Sci USA 2003;100:14445–50.
[23] Reynolds KA, McLaughlin RN, Ranganathan R. Hot spots for allosteric regulation on protein surfaces. Cell 2011;147:1564–75.
[24] Cilia E, Vuister GW, Lenaerts T. Accurate prediction of the dynamical changes within the second PDZ domain of PTP1e. PLoS Comput Biol 2012;8:e1002794.
[25] Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. J Mol Biol 2006;364:337–51.
[26] Deshayes K, Schaffer ML, Skelton NJ, Nakamura GR, Kadkhodayan S, Sidhu SS. Rapid identification of small binding motifs with high-throughput phage display: discovery of peptidic antagonists of IGF-1 function. Chem Biol 2002;9:495–505.
[27] Dunbrack RL, Karplus M. Backbone-dependent rotamer library for proteins application to side-chain prediction. J Mol Biol 1993;230:543–74.
[28] Gordon DB, Marshall SA, Mayot SL. Energy functions for protein design. Curr Opin Struct Biol 1999;9:509–13.
[29] Street AG, Mayo SL. Computational protein design. Structure 1999;7:R105–9.
[30] Shifman JM, Mayo SL. Exploring the origins of binding specificity through the computational redesign of calmodulin. Proc Natl Acad Sci USA 2003;100:13274–9.
[31] Gordon DB, Hom GK, Mayo SL, Pierce NA. Exact rotamer optimization for protein design. J Comput Chem 2003;24:232–43.
[32] Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004;14:1188–90.
[33] Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 1995;91:43–56.
[34] Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–47.