A Structural Portrait of the PDZ Domain Family

Andreas Ernst[1], Brent A. Appleton[2], Ylva Ivarsson[1], Yingnan Zhang[2], David Gfeller[1], Christian Wiesmann[2] and Sachdev S. Sidhu[1]

1 - Banting and Best Department of Medical Research and Department of Molecular Genetics, *University of Toronto, The Donnelly Centre, 160 College Street, Toronto, ON M5S 3E1, Canada*
2 - Department of Early Discovery Biochemistry, *Genentech Inc, San Francisco, CA 94080, USA*

Correspondence to Sachdev S. Sidhu: *sachdev.sidhu@utoronto.ca*
http://dx.doi.org/10.1016/j.jmb.2014.08.012
Edited by A. Keating

Abstract

PDZ (*P*SD-95/*D*iscs-large/*Z*O1) domains are interaction modules that typically bind to specific C-terminal sequences of partner proteins and assemble signaling complexes in multicellular organisms. We have analyzed the existing database of PDZ domain structures in the context of a specificity tree based on binding specificities defined by peptide-phage binding selections. We have identified 16 structures of PDZ domains in complex with high-affinity ligands and have elucidated four additional structures to assemble a structural database that covers most of the branches of the PDZ specificity tree. A detailed comparison of the structures reveals features that are responsible for the diverse specificities across the PDZ domain family. Specificity differences can be explained by differences in PDZ residues that are in contact with the peptide ligands, but these contacts involve both side-chain and main-chain interactions. Most PDZ domains bind peptides in a canonical conformation in which the ligand main chain adopts an extended β-strand conformation by interacting in an antiparallel fashion with a PDZ β-strand. However, a subset of PDZ domains bind peptides with a bent main-chain conformation and the specificities of these non-canonical domains could not be explained based on canonical structures. Our analysis provides a structural portrait of the PDZ domain family, which serves as a guide in understanding the structural basis for the diverse specificities across the family.

Introduction

PDZ (*P*SD-95/*D*iscs-large/*Z*O1) domains are among the most common interaction modules in the human proteome, with approximately 270 embedded in more than 150 proteins [1]. PDZ domains are components of scaffolding proteins that bring together proteins at appropriate cellular compartments and thereby organize signaling complexes and localize enzymes with their substrates [2]. PDZ domains generally function by binding to C-terminal peptide stretches, but some may interact with other PDZ domains, phospholipids or intrinsically disordered regions within proteins [3]. The PDZ family members are thus adaptable recognition modules that have evolved into central hubs of complex protein–protein interaction networks. Consequently, significant efforts have been undertaken to understand and predict their binding preferences [4–7], as such knowledge would allow for large-scale protein network analyses that would illuminate our understanding of cell signaling. However, 20 years after their discovery, we still do not fully understand the rules governing binding specificities within the PDZ family.

The PDZ domain structure is composed of a β-sandwich capped by two α-helices, and C-terminal peptides bind in a shallow groove formed by the second α-helix (α2) and the second β-strand (β2). The core PDZ binding motif consists of the four C-terminal amino acids of the binding partner, and these are numbered starting from the last position (P^0) and going backward to P^{-1}, P^{-2} and P^{-3} [8]. PDZ domains have been divided into specificity classes based on the preferred amino acid type at P^{-2}, with class I, class II and class III corresponding to preference for ligands of the type X[T/S]XΦ_{COOH},

$X\Phi X\Phi_{COOH}$ or $X[D/E]X\Phi_{COOH}$ (where "Φ" is a hydrophobic and "X" is any amino acid), respectively [9,10]. Two large-scale studies provided more comprehensive views of specificity within the PDZ family, which showed that PDZ domains can recognize up to seven C-terminal ligand side chains [4,5]. Indeed, clustering of specificity profiles for 54 human PDZ domains derived by peptide-phage display shows that the classification based on P^{-2} alone is overly simplistic [4], as comparison of specificities across the entire binding site produces a specificity tree with many distinct branches (Fig. 1). Notably, similar peptide binding specificities do not necessarily correlate with similarity within the PDZ binding site, as it has been found that highly similar binding preferences can be achieved by binding sites that differ greatly in sequence [11]. Consequently, predictions based solely on primary sequence remain imprecise [12] and a full understanding of the rules governing PDZ domain specificities will require an in-depth structural analysis of high-affinity PDZ–peptide complexes.

As a follow-up to our previous large-scale specificity profiling study of the PDZ family [4], we present a systematic structural survey to elucidate key structural features correlated with PDZ specificity. Although 376 PDZ structures are deposited in the Protein Data Bank (PDB) (Table S1), the majority lack bound ligands and are thus of limited value for understanding PDZ specificity. Moreover, many of the PDZ–ligand complex structures represent suboptimal interactions, and this hampers efforts to derive a meaningful interpretation of the molecular interactions guiding recognition of optimal peptide sequences. Additionally, the current structural database is biased toward class I domains and thus offers limited insights into the binding modes of other specificity classes and unusual family members (Fig. 1).

To address these limitations, we solved four new structures of high-affinity PDZ–peptide complexes

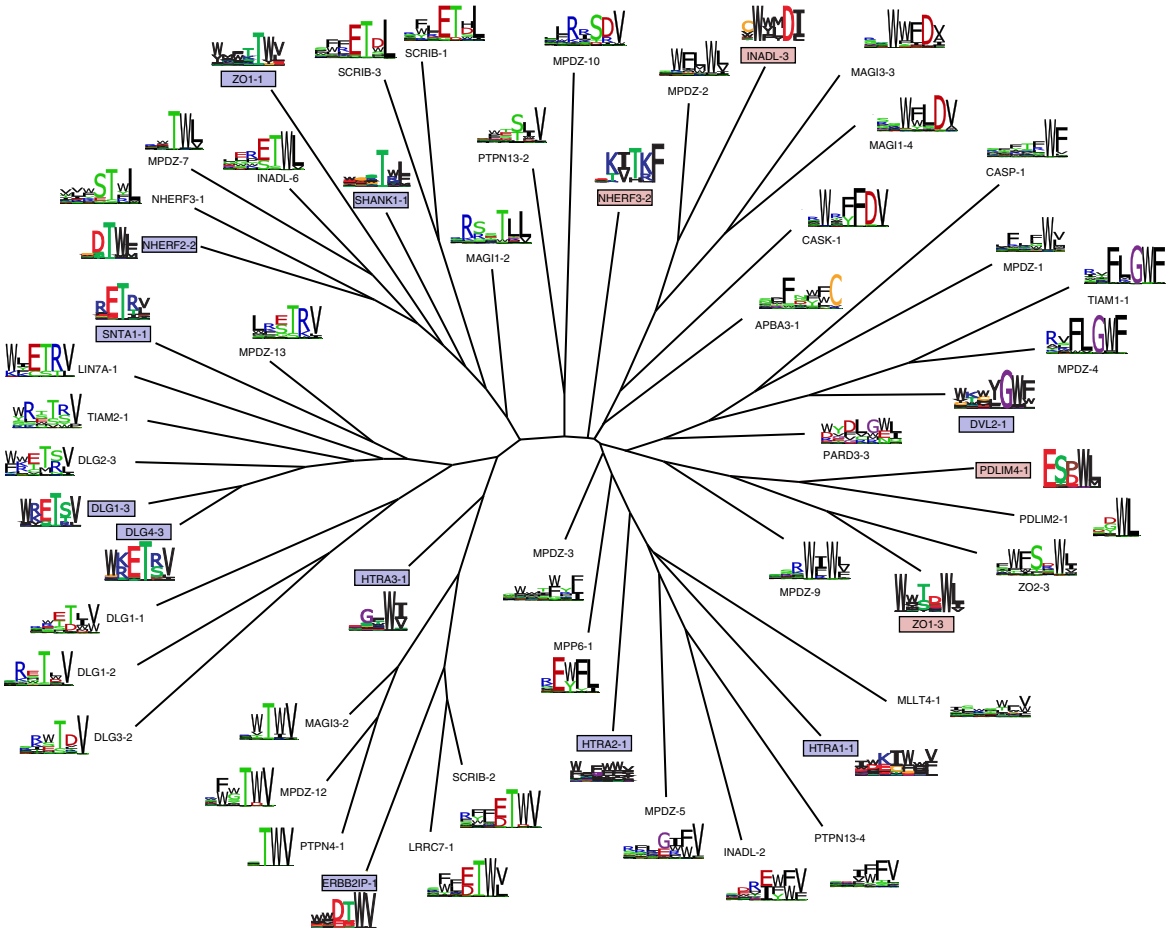

Fig. 1. The specificity tree of PDZ domains. The position weight matrices (PWMs) were derived from C-terminal peptide ligands isolated from phage-displayed libraries, as previously described [4], and were clustered on the basis of overall similarity. The colored boxes indicate PDZ domains for which structures are available in complex with high-affinity ligands, either from previous studies (blue) or from the current study (red). The PWM shown for SHANK1-1 was derived from peptides binding to the close homolog SHANK3-1.

representing PDZ specificities that are not represented in the structural database. Taking advantage of the new structural information, we present a detailed structural analysis of the specificity determinants of PDZ–peptide interactions. We also demonstrate the utility of our analysis by altering the binding specificity of a model PDZ domain by rational design. Our study demonstrates that PDZ domain specificity depends on both direct and indirect interactions, and it provides a comprehensive framework for understanding structure–function relationships within the PDZ family.

Results and Discussion

New structures expand structural coverage of the PDZ family

At the time of this analysis, 28 ligand-bound PDZ structures were deposited in the PDB (Table S2), of which 16 represent 12 distinct PDZ domains in complex with near-optimal ligands that closely resemble specificity profiles defined by peptide-phage display (Figs. 1 and 2). Eight of these structures were solved by our group with ligands designed on the basis of peptide-phage display experiments [8,13–16] and the other eight were solved by other groups with ligands designed on the basis of confirmed natural interactions [17–22]. Twelve of the 16 structures represent class I interactions. Only three structures represent class II interactions and these all belong to the HTRA PDZ domain subfamily, whose members are structurally closer to bacterial PDZ domains and are permuted relative to most metazoan PDZ domains. A single structure of the dishevelled-2 PDZ domain (DVL2-1) represents a non-canonical binding mode in which the peptide adopts an unusual conformation [15]. Thus, the current structural database is heavily biased in favor of class I PDZ domains and lacks coverage of the other diverse branches of the specificity tree (Fig. 1).

To expand our structural understanding of PDZ–peptide interactions, we used the specificity tree to identify branches not represented in the structural database and complemented the available dataset by solving the crystal structures of four PDZ–peptide complexes representing distinct specificities (Figs. 1 and 2a–d). These complex structures included the class II third PDZ domain of the scaffolding protein INADL (*i*nactivation *n*o *a*fter potential *D-l*ike protein) (INADL-3; Fig. 2A) and the class III third PDZ domain of the tight junction protein ZO1 (ZO1-3; Fig. 2B). We also solved structures of two PDZ–peptide complexes that exemplify unusual features, namely, the second PDZ domain of the Na^+/H^+ exchange regulatory cofactor 3 (NHERF3-2), which prefers lysine residues at P^{-1} and P^{-4} (Fig. 2C), and the PDZ domain of the LIM domain protein 4 (PDLIM4-1), which prefers proline at P^{-2} (Fig. 2D).

Each PDZ domain was crystallized with its cognate peptide ligand by extending the C terminus of the PDZ domain with a short flexible linker followed by the peptide ligand sequence (Table 1 and Fig. S1) using an approach previously employed for other PDZ–peptide complexes [15,19]. All bound peptide extensions showed a well-defined electron density, thus allowing a precise description of the molecular details mediating specificity (Fig. S2). INADL-3 (Fig. S1a) and ZO1-3 (Fig. S1b) crystallized as homodimers with the peptides inter-locking the two domains. NHERF3-2 crystallized in a trimeric form (Fig. S1c). In the case of PDLIM4-1, the asymmetric unit shows an elongated chain of three domain dimers, which are connected by a disulfide bond between the Cys^{44} side chains (Fig. S1d). Similar to INADL-3 and ZO1-3, the peptide extension of each adjacent PDLIM4-1 monomer is bound to its neighboring molecule and forms a head-to-tail polymer [22]. Overall, the structures share a typical PDZ domain fold composed of a six-stranded antiparallel β-sandwich and two α-helices. As seen previously in other PDZ–peptide complexes (Fig. 2), the ligand peptide binds in the shallow groove between strand β2 and helix α2, and the "carboxylate-binding" loop that precedes strand β2 coordinates the C-terminal carboxyl group. Side chains located on β2, β3 and α2 interact directly with the ligand peptide, and importantly, the side-chain interactions differ between PDZ–peptide complexes, which accounts for differences in ligand specificity. With the addition of the four new PDZ–peptide complexes, we now have structural examples of most branches of the PDZ specificity tree (Fig. 1), which allows detailed investigation of the structural basis for ligand discrimination by the PDZ family.

Structural determinants of site^0 specificity

The major determinant for C-terminal peptide recognition is the conserved carboxylate-binding loop [17] whose main-chain amides hydrogen bond to the C-terminal carboxyl group and anchor the ligand in the binding site. The C-terminal residue is docked into a hydrophobic cavity lined by the side chains of the seventh residue in the loop between strands β1 and β2 (β1:β2-7), the first residue in strand β2 (β2-1) and the eighth residue in helix α2 (α2-8) (Fig. 3) (PDZ residues are numbered according to a previously described structure-based nomenclature [23]). Consequently, PDZ domains prefer ligands that terminate in hydrophobic residues, but nonetheless, many PDZ domains show exquisite discrimination with regard to the chemical nature of the C-terminal side chain, indicating that the site^0 pockets are often fine-tuned for recognition of particular hydrophobic side chains [4].

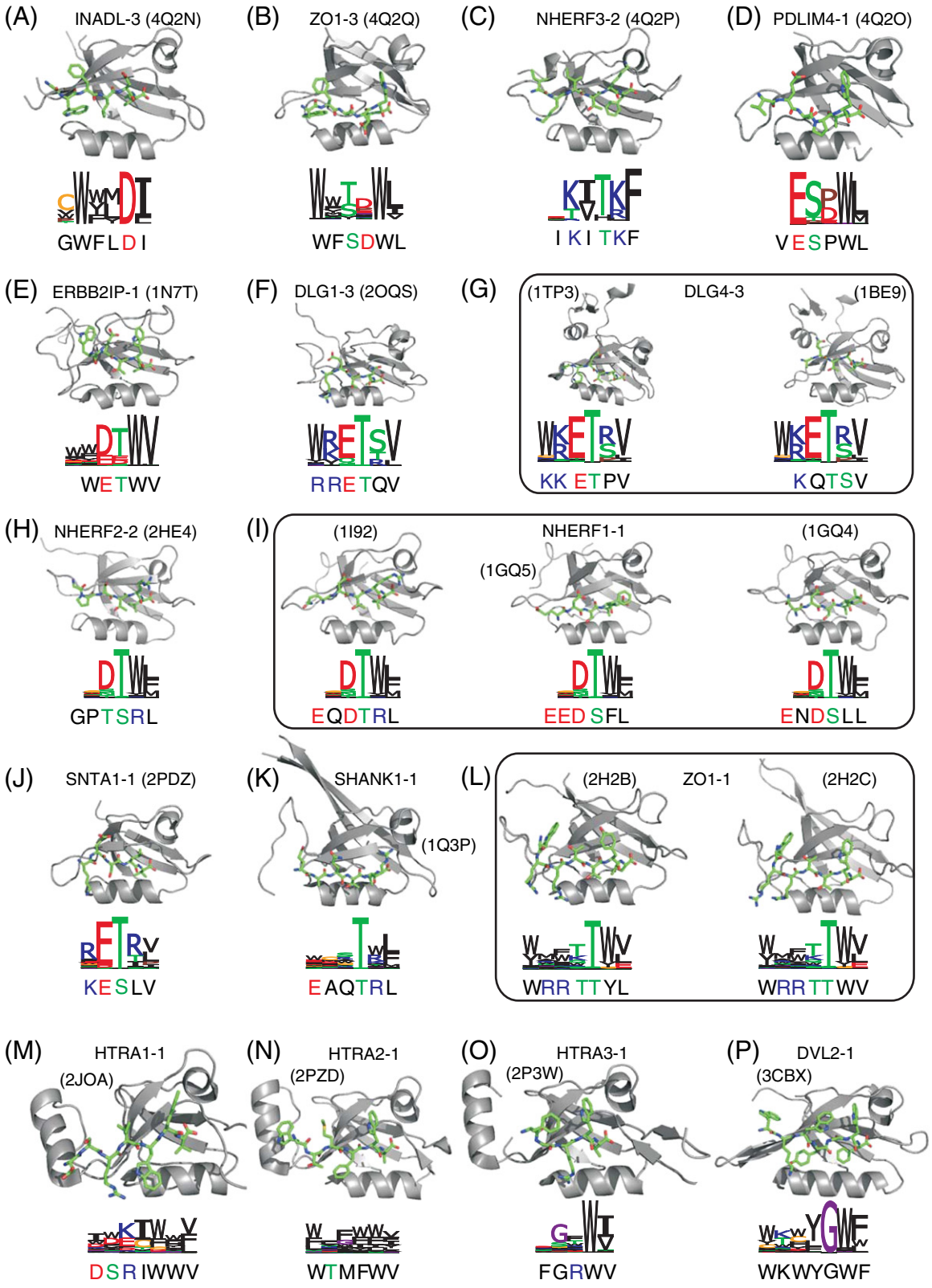

Fig. 2 *(legend on next page)*

Table 1. Data collection and refinement statistics

	INADL-3	PDLIM4-1	NHERF3-2	ZO1-3
Data collection				
Space group	$P2_1$	$P2_1$	$H3$	$P4_12_12$
Cell dimensions				
a, *b*, *c* (Å)	59.0, 65.2, 82.5	33.7, 64.6, 126	153, 153, 34.6	50.6, 50.6, 78.7
α, β, γ (°)	90, 101, 90	90, 93.1, 90	90, 90, 120	90, 90, 90
Resolution (Å)[a]	25–2.00 (2.07–2.00)	50–2.10 (2.18–2.10)	20–2.05 (2.12–2.05)	50–1.45 (1.50–1.45)
R_{sym}	0.090 (0.596)	0.040 (0.384)	0.043 (0.296)	0.032 (0.651)
$I/\sigma I$	17.7 (2.6)	26.1 (3.6)	37.1 (4.4)	58.2 (3.4)
Completeness (%)	98.4 (96.8)	99.3 (99.8)	94.9 (73.6)	99.1 (98.5)
Redundancy	5.1 (5.0)	4.0 (3.6)	5.3 (4.6)	10.1 (9.5)
Refinement				
Resolution (Å)	20–2.00	20–2.10	20–2.05	20–1.45
No. of reflections	38,818	29,444	17,128	17,647
R_{work}/R_{free}	0.201/0.253	0.246/0.291	0.188/0.252	0.160/0.210
No. of atoms				
Protein	4442	3910	2045	721
Ligand/ion	24	0	12	0
Water	326	61	100	94
r.m.s.d.				
Bond lengths (Å)	0.014	0.012	0.015	0.016
Bond angles (°)	1.4	1.4	1.5	1.6
Ramachandran (%)				
Favored	98.6	99.0	99.6	98.9
Outliers	0.2	0.4	0	0

[a] Values for the highest-resolution shell are shown in parentheses.

A comparison of structures of PDZ domains bound to peptides with Val, Leu, Ile or Phe at P^0 reveals that the orientations of residues at positions β2-1 and β1:β2-7 change in order to accommodate the different ligand side chains (Fig. 3). For example, binding pockets for P^0 side chains in ERBB2IP-1 (Fig. 3A) and INADL-3 (Fig. 3B) differ only at the β2-1 position that is occupied by a Phe in ERBB2IP-1 or an Ile in INADL-3, but the structures show that the orientation of the β2-1 side chains differs between the two domains. These differences are sufficient to alter specificity, as ERBB2IP-1 prefers Val^0 while INADL-3 prefers the larger Ile^0. Due to size and orientation differences at the β2-1 position, the site^0 pocket of INADL-3 is deeper than that of ERBB2IP-1, and consequently, it can accommodate bulkier side chains. In PDLIM4-1 (Fig. 3C) and NHERF3-2 (Fig. 3D), which prefer Leu^0 or Phe^0, respectively, the side chain at the β1:β2-7 position points away from the site^0 pocket. In NHERF3-2, this creates a flat, hydrophobic site with specificity for Phe^0. The similarly shaped hydrophobic site of PDLIM4-1 prefers Leu^0, which packs against the flat surface formed by Phe(β2-1) and Ile(α2-8). Thus, as noted previously, the specificity for the P^0 side chain is mediated by the size and shape of the hydrophobic pocket [8,21]. However, we have shown recently that the shape of the site^0 pocket is determined not only by the residues that line the pocket but also by second-sphere residues that alter the pocket through indirect interactions [23].

Structural determinants of site^{-1} specificity

At site^{-1}, the most common preference is for hydrophobic side chains, but some PDZ domains prefer either negatively or positively charged side chains. Phage-derived PDZ domain ligands are often

Fig. 2. Structures of PDZ domains in complex with high-affinity C-terminal peptides. (A) INADL-3 (PDB entry: 4Q2N), (B) ZO1-3 (PDB entry: 4Q2Q), (C) NHERF3-2 (PDB entry: 4Q2P), (D) PDLIM4-1 (PDB entry: 4Q2O), (E) ERBB2IP-1 (PDB entry: 1N7T), (F) DLG1-3 (PDB entry: 2OQS), (G) DLG4-3 (PDB entries: 1TP3 and 1BE9), (H) NHERF2-2 (PDB entry: 2HE4), (I) NHERF1-1 (PDB entries: 1I92, 1GQ5 and 1GQ4), (J) SNTA1-1 (PDB entry: 2PDZ), (K) SHANK1-1 (PDB entry: 1Q3P), (L) ZO1-1 (PDB entries: 2H2B and 2H2C), (M) HTRA1-1 (PDB entry: 2JOA), (N) HTRA2-1 (PDB entry: 2PZD), (O) HTRA3-1 (PDB entry: 2P3W) and (P) DVL2-1 (PDB entry: 3CBX). In each panel, the PDZ domain is shown as a gray ribbon and the peptide is shown as sticks colored according to atom type, as follows: carbon (green), nitrogen (blue) and oxygen (red). Below each structure shown is the PWM derived by peptide-phage display and the sequence of the bound peptide [4]. The PWMs shown for SHANK1-1 and NHERF1-1 were derived from peptides binding to the close homologs SHANK3-1 or NHERF2-2, respectively. All structure figures were generated using PyMOL (http://www.pymol.org/).

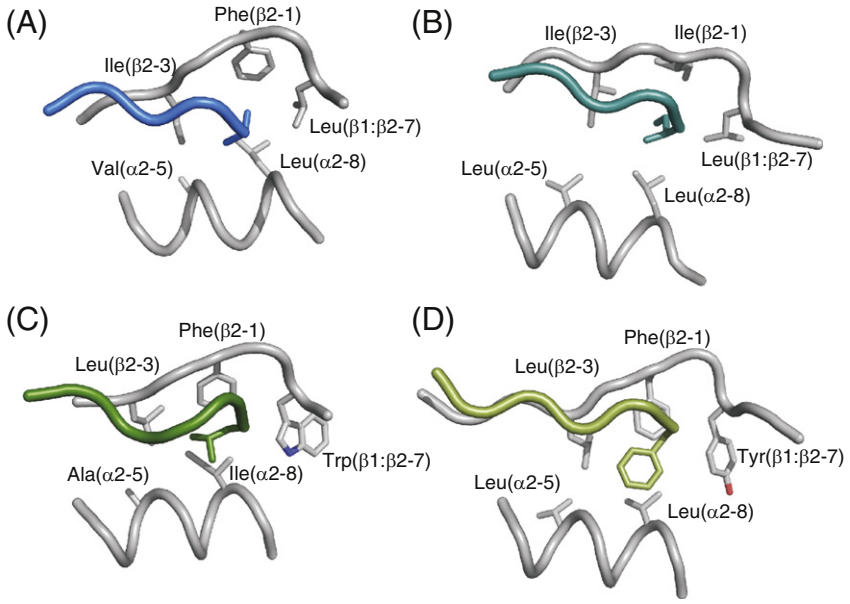

Fig. 3. Structural determinants of site^0 specificity. (A) ERBB2IP-1 (PDB entry: 1N7T), (B) INADL-3, (C) PDLIM4-1 and (D) NHERF3-2. PDZ domains are gray and side chains that form the site^0 pocket are labeled and shown as sticks. Peptide ligands are colored and C-terminal side chains are shown as sticks.

biased toward Trp residues at this position, while this is less frequently observed among natural ligands [24]. Our analysis reveals that specificity at site^{-1} depends on interactions with both main chains and side chains in strands β2 and β3, and different PDZ domain specificities arise from differing interactions with these strands (Fig. 4). As shown by the structure of ERBB2IP-1 (Fig. 4A), a strong preference for Trp at P^{-1} in many PDZ domains is explained by the fact that the bulky Trp side chain interacts mainly with the main chain of the β2 strand and does not interact specifically with PDZ side chains [16]. The preference of INADL-3 for aspartate at P^{-1} is explained by a salt bridge between the acidic ligand side chain and Arg(β2-2) (Fig. 4B). For domains that prefer basic residues, we observe two distinct structural mechanisms. NHERF3-2 crystallized as a homotrimer, and the distances between the ε-NH2 group of Lys^{-1} and Asp:β3-5 vary from 5.3 to 8.5 Å among the three complexes in the trimer (Fig. 4C). While these distances are too far for a direct salt bridge, the Lys^{-1} and Asp:β3-5 side chains likely make favorable electrostatic interactions. In contrast, the recognition of Arg^{-1} by NHERF1-1 relies on a salt bridge with a Glu side chain at the β3:α1-1 position (Fig. 4D). The structure of NHERF1-1 has also been solved in complex with ligands containing either Phe (Fig. 4E) or Leu at P^{-1} (Fig. 4F), and in these cases, there are no obvious favorable interactions between the P^{-1} side chain and PDZ side chains. Instead, the P^{-1} side chain appears to interact with the main chain of strand β2 in a manner similar to that seen for Trp^{-1} interacting with ERBB2IP-1 (Fig. 4A).

As exemplified by ERBB2IP-1, the preference of many PDZ domains for ligands containing Trp at P^{-1} can be considered a "default" specificity that does not depend on favorable interactions with PDZ side chains but, rather, that arises from interactions with the main chain of strand β2. Thus, we hypothesized that, to achieve specificity for charged ligand side chains at site^{-1}, it is necessary, but not sufficient, to introduce favorable counter charges within the PDZ domain. In addition, it is likely necessary to disrupt the favorable interactions between Trp^{-1} and the PDZ main chain by placing bulky side chains within strand β2. Thus, according to our "break-and-make" hypothesis, achieving specificity for charged side chains at site^{-1} requires both destabilization (i.e., breaking) of the Trp^{-1} interaction by disruption of the interactions with strand β2 and stabilization (i.e., making) of favorable electrostatic interactions between ligand and PDZ side chains. This supposition was supported by our previous finding that replacement of Ser(β2-4) with Ile or Val altered the specificity of ERBB2IP-1 at site^{-1} [4], and moreover, the domains that prefer charged ligand side chains at site^{-1} all contain bulky residues at the β2-4 position (Fig. 4).

To test the break-and-make hypothesis, we constructed ERBB2IP-1 variants rationally designed to alter specificity at site^{-1} to resemble that of NHERF1-1, and we determined specificity profiles by using C-terminal peptide-phage libraries. We first introduced an Asp or Glu in place of Gln at position β3:α1-1 (Fig. 4G) and, as predicted, these ERBB2IP-1 variants retained the wild-type preference for Trp^{-1}.

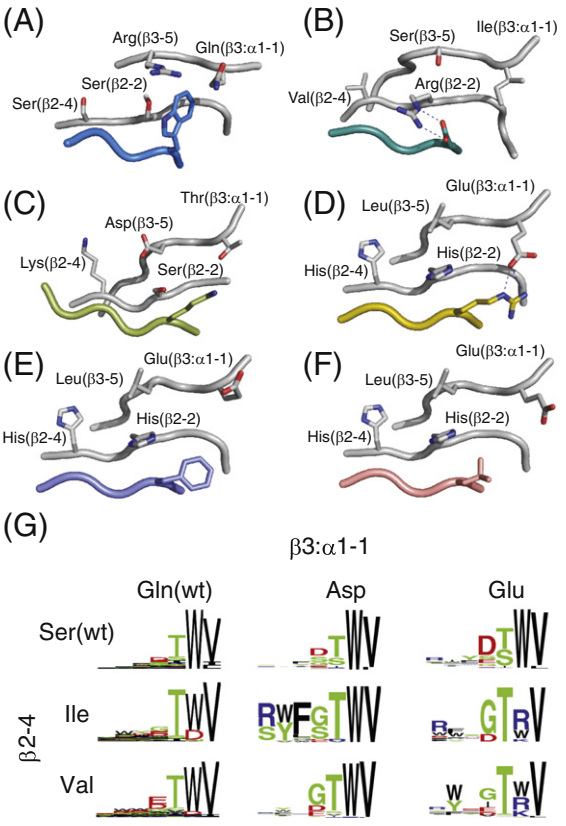

Fig. 4. Structural determinants of site^{-1} specificity. (A) ERBB2IP-1 (PDB entry: 1N7T), (B) INADL-3, (C) NHERF3-2, (D) NHERF1-1 (PDB entry: 1I92), (E) NHERF1-1 (PDB entry: 1GQ5) and (F) NHERF1-1 (PDB entry: 1GQ4). (G) PWMs for ERBB2IP-1 variants derived from C-terminal peptide ligands isolated from phage-displayed libraries. PDZ domains are gray and side chains that contribute to specificity at site^{-1} are labeled and shown as sticks. Peptide ligands are colored and P^{-1} side chains are shown as sticks.

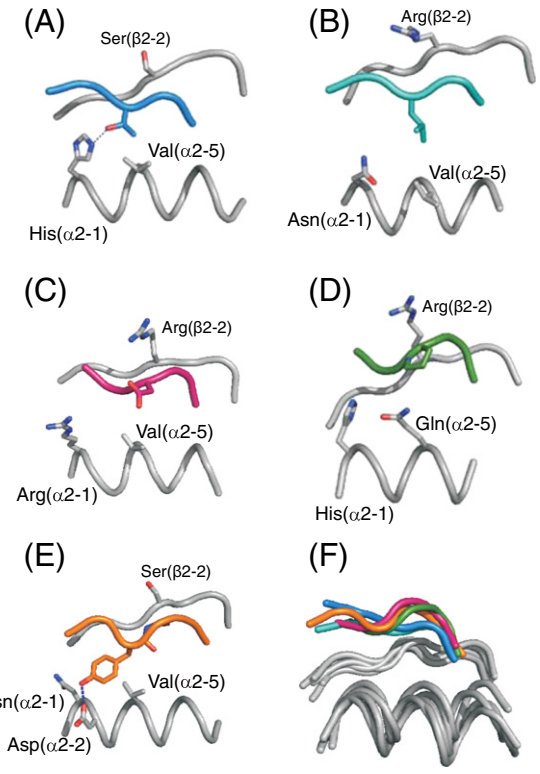

Fig. 5. Structural determinants of site^{-2} specificity. (A) ERBB2IP-1 (PDB entry: 1N7T), (B) INADL-3, (C) ZO1-3, (D) PDLIM4-1 and (E) DVL2-1 (PDB entry: 3CBX). (F) Superposition of the structures from (A) to (E). PDZ domains are gray and side chains that contribute to specificity at site^{-2} are labeled and shown as sticks. Peptide ligands are colored and P^{-2} side chains are shown as sticks. In the DVL2-1 ligand, the Tyr^{-3} side chain is shown.

Introduction of Ile or Val in place of Ser at position β2-4 broadened the specificity of ERBB2IP-1 to accept both Trp and Asp at site^{-1}. The broadened specificity was likely due to changes in the peptide main-chain conformation that destabilized the interaction with Trp^{-1} and allowed for favorable interactions between an Asp^{-1} side chain and the nearby Arg(β3-5) side chain (Fig. 4A). Consistent with our hypothesis, only the double substitution of Glu at position β3:α1-1 and Ile or Val at position β2-4 produced an ERBB2IP-1 variant with a marked preference for ligands containing Arg^{-1}. The fact that the specificities of double mutants containing Asp at position β3:1-1 resembled that of the wild type likely reflects a requirement for precise spatial positioning of side chains and suggests a suboptimal distance between Asp(β3:α1-1) and positively charged side chains at P^{-1}. Taken together, these results support our hypothesis that specificity for P^{-1} relies on interplay between residues in the β2 and β3 strands and involves both the establishment of favorable domain-ligand, side-chain interactions and the introduction of unfavorable interactions that disrupt the interaction between Trp^{-1} and the PDZ main chain.

Structural determinants of site^{-2} specificity

We define a canonical PDZ–ligand interaction as one in which the ligand main chain adopts a β-strand conformation by interacting in an antiparallel fashion with strand β2. This places the P^{-2} side chain proximal to the side chains at positions α2-1 and α2-5, which serve to determine the ligand residue types that are favored at this position. The residue at position β2-2 also contributes to site^{-2} specificity by helping to orient the peptide in the binding groove. Class I PDZ domains such as ERBB2IP-1 (Fig. 5A), which prefer ligands containing Thr/Ser^{-2}, almost invariably contain a His side chain at position α2-1, which acts as a hydrogen bond acceptor for the hydroxyl group of the ligand side chain (Fig. 6). Moreover, most class I domains contain a Val side

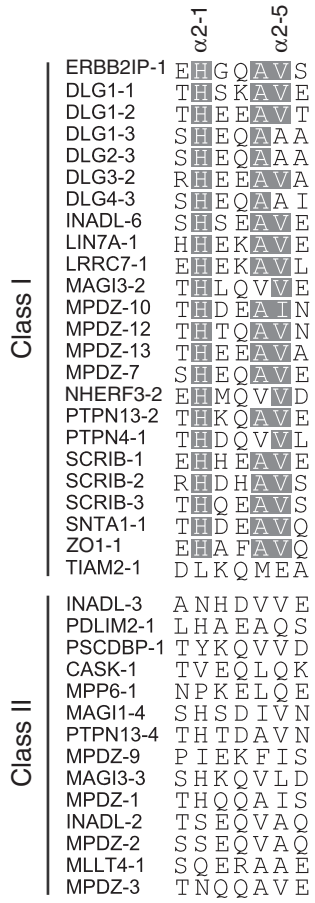

Fig. 6. Sequence alignment of PDZ domain α2 helices. Conserved residues are shaded gray.

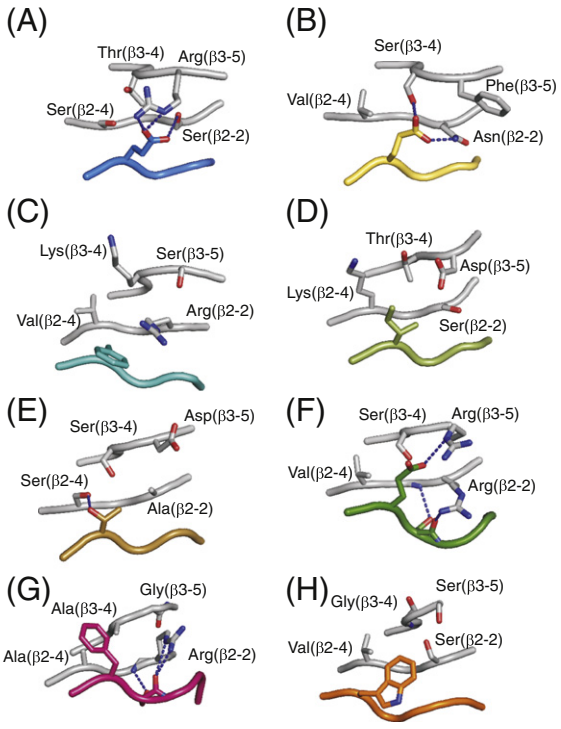

Fig. 7. Structural determinants of site^{-3} specificity. (A) ERBB2IP-1 (PDB entry: 1N7T), (B) DLG4-3 (PDB entry: 1TP3), (C) INADL-3, (D) NHERF3-2, (E) ZO1-1 (PDB entry: 2H2B), (F) PDLIM4-1, (G) ZO1-3 and (h) DVL2-1 (PDB entry: 3CBX). PDZ domains are gray and side chains that contribute to specificity at site^{-3} are labeled and shown as sticks. Peptide ligands are colored and the side chains of residues P^{-3} (A–E), P^{-3} and P^{-4} (F and G) or P^{-4} (H) are shown as sticks.

chain at position α2-5, which makes favorable van der Waals contacts with the aliphatic portion of a Thr^{-2} side chain, thus explaining the preference of most class I domains for Thr over Ser at P^{-2} (Figs. 1 and 2).

In class II domains, hydrophobic P^{-2} side chains are accommodated in a shallow hydrophobic pocket, but the α2-1 and α2-5 side chains that line this pocket are not conserved across the class (Fig. 6). The structure of INADL-3 in complex with the peptide $WFLDI_{COOH}$ shows that binding of Leu^{-2} is guided by hydrophobic interactions with Val(α2-5), which shields site^{-2} from solvent (Fig. 5B). These hydrophobic interactions are rather non-specific, and thus, INADL-3 exhibits broad specificity for ligands containing hydrophobic P^{-2} side chains (Fig. 2). Indeed, class II domains in general exhibit lower specificity for P^{-2} than do class I domains [4].

ZO1-3 can be assigned to class III, as defined by a preference for Asp^{-2} (Fig. 5C), but the specificity profile exhibits significant promiscuity for P^{-2} (Fig. 2) [4]. Arg residues occupy the α2-1 and β2-2 positions of ZO1-3, suggesting the possibility of salt bridges with the Asp^{-2} side chain. However, in the structure of ZO1-3 in complex with the peptide $LWFSDWL_{COOH}$ (Fig. 2C), the side chain of Asp^{-2} points toward solvent and does not interact directly with either of the Arg side chains (Fig. 5C). Instead, the side chain of Arg(β2-2) forms a hydrogen bond with the main chain at P^{-2} and pulls the ligand toward strand β2, resulting in a gap of 7.4 Å between the ligand main chain and helix α2. Consequently, the Arg at position α2-1 does not interact directly with the Asp^{-2} side chain. Moreover, the peptide adopts a non-canonical bent conformation, distinct from the extended β-sheet conformation of canonical ligands (Fig. 5F). The bend in the ligand main chain is supported by the side chain of Trp^{-1}, which binds in the hydrophobic pocket formed by the aliphatic portion of the Arg(β2-2) side chain and the Leu(β3:α1-1) side chain.

The peptide $VESPWL_{COOH}$ bound to PDLIM4-1 also adapts a non-canonical bent conformation (Fig. 5D). Despite containing a His at position α2-1, typical of class I domains, PDLIM4-1 prefers ligands that contain Pro or Asp at P^{-2} (Fig. 2). Thus, the specificity of PDLIM4-1 more closely resembles that of class III domains, and for ligands that contain an Asp at P^{-2}, it is likely that the binding mode is similar

to that observed for ZO1-3 bound to the peptide $LWFSDWL_{COOH}$ (Fig. 5C). In the case of the ligand containing Pro^{-2} (Fig. 5D), the structure shows that the Arg(β2-2) side chain of PDLIM4-1 interacts with the main chain of the P^{-3} residue and packs against the side chain of Trp^{-1}. The bend in the main chain of the peptide ligand is supported by the polar carboxamide group of Gln(α2-5), which appears to repulse the hydrophobic ring of Pro^{-2}.

A bent main-chain conformation was also observed in the structure of DVL2-1 bound to the peptide $WKWYGWF_{COOH}$ [15] (Figs. 2P and 5E and F). Gly^{-2} adopts a positive Φ angle, which would be energetically unfavorable for L-amino acid side chains. Consequently, the direction of the Tyr^{-3} side chain is flipped and, in contrast with canonical PDZ–ligand interactions, $site^{-2}$ is occupied by the Tyr^{-3}–Gly^{-2} pair rather than by a single P^{-2} residue. The unusual ligand conformation at $site^{-2}$ appears to be reinforced by a hydrogen bond between the side chains of Tyr^{-3} and Asp(α2-2). Our previous study showed that the DVL2-1 binding site is unusually flexible and can accommodate both C-terminal and internal ligands [15]. Three distinct structures of DVL2-1 complexed with internal ligands revealed two non-canonical binding modes and a binding mode resembling a canonical class II interaction with a Val^{-2} side chain residing in a shallow pocket formed by α2-1 and α2-5 side chains that are identical with those found in INADL (Fig. 5B and E).

Structural determinants of $site^{-3}$ specificity

At $site^{-3}$, the predominant specificity is for hydrophobic residues, but many domains prefer Asp/Glu and some prefer Thr/Ser (Fig. 1). Our structural database contains examples for each of these specificities and, moreover, provides insights into how $site^{-3}$ interactions differ between canonical and non-canonical domain-ligand complexes (Fig. 2).

Within canonical PDZ domains, $site^{-3}$ is always occupied by the P^{-3} ligand residue. Four of the five domains in complex with peptides that contain negatively charged side chains at P^{-3} use a similar mechanism for ligand recognition (Fig. 2E, F, G, I and J). As exemplified by ERBB2IP-1 (Fig. 7A), Arg(β3-5) establishes a salt bridge with the Glu^{-3} side chain, which also forms a hydrogen bond with Ser(β2-2). In contrast, DLG4-3 contains a Phe at β3-5 and the Glu^{-3} side chain instead forms hydrogen bonds with Ser(β3-4) and Asn(β2-2) (Fig. 7B). For INADL-3 and NHERF3-2, a preference for hydrophobic P^{-3} residues is explained by a hydrophobic $site^{-3}$, but the sites differ and prefer either aromatic or aliphatic side chains, respectively (Fig. 2A and B). In the case INADL-3, Phe^{-3} is bound in a hydrophobic pocket composed of the aliphatic portions of the side chains of Lys(β3-4), Arg(β2-2) and Val(β2-4) (Fig. 7C). The phenol ring of Phe^{-3} forms a cation–π interaction with the guanidinium group of Arg(β2-2), which likely accounts for the preference for aromatic ligand side chains. NHERF3-2 prefers β-branched aliphatic side chains at P^{-3}, which dock in a shallow hydrophobic groove formed by the side chains of the residues at positions β2-2, β2-4 and β3-4 (Fig. 7D). The aliphatic portion of Lys(β2-4) aligns with the ethyl group of Ile^{-3} and the aliphatic portions of Ser(β2-2) and Thr(β3-4) also contribute to the formation of the hydrophobic pocket. The structures of ZO1-1 explain the preference for a Thr^{-3} side chain, which makes a hydrogen bond with the side chain of Ser(β2-4) (Fig. 7E).

In the non-canonical PDZ domains, the bent conformation of the ligand main chain causes $site^{-3}$ to be occupied by the P^{-4} ligand residue. Nonetheless, in the case of the PDLIM4-1 structure, in which $site^{-3}$ is occupied by a Glu^{-4} side chain (Fig. 7F), the interactions are remarkably similar to those observed between $site^{-3}$ of ERBB2IP-1 and a Glu^{-3} side chain (Fig. 7A). Glu^{-4} forms a salt bridge with Arg(β3-5), while the side chain of Ser^{-3} resides outside of $site^{-3}$ and forms a hydrogen bond with the main chain of Leu(β2-3). Furthermore, Arg(β2-2) contributes to binding by forming a hydrogen bond to the main chain of P^{-3}. In ZO1-3, Arg(β2-2) interacts with Ser^{-3} in the same fashion as in PDLIM4-1. However, ZO1-3 contains a Gly at position β3-5 and prefers aromatic side chains at P^{-4} (Fig. 2B), and the structure shows that Phe^{-4} docks between Ala(β2-4) and Ala(β3-4) (Fig. 7G). Interestingly, the $site^{-3}$ of DVL2-1 shares some characteristics with that of ZO1-3 and binds the aromatic side chain of Trp^{-4}, which docks between Ser(β2-2) and Val(β2-4) (Fig. 7H).

Conclusions

We have dissected the molecular determinants of PDZ–peptide specificity by organizing existing structures and elucidating new structures to cover most branches of the PDZ specificity tree. Our structural database of near-optimal PDZ–peptide complexes shows that specificity is largely governed by ligand interactions with PDZ residues in helix α2 and strands β2 and β3. However, the situation is complicated by the fact that both main-chain and side-chain interactions are important for binding specificity. Moreover, other studies have shown that PDZ specificity can be modulated by residues outside the binding cleft [1] and by allosteric long-range interactions that alter protein dynamics [3,23,25,26]. Our study further advances the understanding of factors that mediate protein interactions at the molecular level and shows how functional specificity profiling and structural studies can be combined to derive an in-depth understanding of binding specificity.

We have provided a detailed description of how PDZ domains discriminate among C-terminal peptide ligands. However, we believe that further research will be required to elucidate the molecular paths PDZ

domains may take to evolve divergent binding specificities [11,27]. Furthermore, some PDZ domains can bind productively to internal regions of proteins [3,29,30] but only a few structures are available for such interactions and these binding mechanisms are poorly understood [15,30]. Finally, the recent finding that many PDZ domains may interact with membrane phospholipids raises new questions about how these interactions are structurally accommodated and what impact the lipid ligand may have on peptide binding [31–33]. Thus, many challenges remain before we can claim to possess a full understanding of PDZ interactions, and we foresee that a combination of specificity profiling, structural analysis, protein engineering and cell biology will be required to address these questions.

Materials and Methods

Crystallization and structure solution

For each crystal structure, a recombinant fusion protein was produced consisting of the PDZ domain, followed by a three- to five-residue linker, followed by the peptide ligand. Recombinant proteins were produced in *Escherichia coli*, purified as previously described [8] and dialyzed into 20 mM Tris (pH 7.5), 200 mM NaCl and 0.5 mM TCEP at concentrations listed in Table S3. Crystals were grown by the vapor diffusion method at 19 °C, utilizing a 1:1 protein-to-well solution ratio, transferred to cryosolution and flash frozen in liquid nitrogen. Data were collected at 100 K on an in-house Cu-Kα source (INADL-3 and NHERF3-2) or at beamline 5.0.1 (PDLIM4-1 and ZO1-3) of the Advanced Light Source (Berkeley, CA) and processed with HKL2000 [34]. The dataset for PDLIM4-1 was severely anisotropic and an ellipsoidal truncation was applied after scaling as previously described [35]. Structures were solved by molecular replacement with PHASER [36] (Table S3) [8,19,37,38]. Structures were built with Coot [39] and refined with REFMAC5 [40]. All four structures showed clear density for the peptides in the electron density maps (Fig. S2) and are well refined with good geometry as assessed by MolProbity [41]. Detailed refinement statistics are listed in Table 1.

Selection of PDZ domain ligands

ERBB2IP-1 mutants were produced and purified as glutathione *S*-transferase fusion proteins, as previously described [4]. The purified proteins were used as immobilized targets for rounds of binding selections with a phage-displayed library of random C-terminal peptides, as previously described [27,42]. After five rounds of selection, approximately 50 binding clones were sequenced for each PDZ domain and binding profiles were calculated as position weight matrices based on unique peptides using the LOLA software package [4].

Accession numbers

Coordinates and structure factors for INADL-3, PDLIM4-1, NHERF3-2 and ZO1-3 have been deposited in the PDB with accession numbers 4Q2N, 4Q2O, 4Q2P and 4Q2Q, respectively.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.08.012.

Acknowledgments

This work was supported by a grant from the Canadian Institutes for Health Research (MOP-93684).

Received 10 April 2014;
Received in revised form 6 August 2014;
Accepted 12 August 2014
Available online 23 August 2014

Keywords:
protein–protein interaction;
modular domains;
ligand specificity;
peptide recognition;
protein engineering

Abbreviations used:
PDB, Protein Data Bank.

References

[1] Luck K, Charbonnier S, Trave G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. FEBS Lett 2012;586:2648–61.
[2] Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. J Cell Sci 2001;114:3219–31.
[3] Ivarsson Y. Plasticity of PDZ domains in ligand recognition and signaling. FEBS Lett 2012;586:2638–47.
[4] 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.
[5] Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, et al. PDZ domain binding selectivity is optimized across the mouse proteome. Science 2007;317:364–9.
[6] Smith CA, Kortemme T. Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. J Mol Biol 2010;402:460–74.
[7] Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G. Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol 2008;26:1041–5.
[8] 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.
[9] 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.
[10] Stricker NL, Christopherson KS, Yi BA, Schatz PJ, Raab RW, Dawes G, et al. PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. Nat Biotechnol 1997;15:336–42.

[11] 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.
[12] Luck K, Fournane S, Kieffer B, Masson M, Nomine Y, Trave G. Putting into practice domain-linear motif interaction predictions for exploration of protein networks. PLoS One 2011;6:e25376.
[13] Runyon ST, Zhang Y, Appleton BA, Sazinsky SL, Wu P, Pan B, et al. Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. Protein Sci 2007;16:2454–71.
[14] Zhang Y, Appleton BA, Wu P, Wiesmann C, Sidhu SS. Structural and functional analysis of the ligand specificity of the HtrA2/Omi PDZ domain. Protein Sci 2007;16:1738–50.
[15] Zhang Y, Appleton BA, Wiesmann C, Lau T, Costa M, Hannoush RN, et al. Inhibition of Wnt signaling by dishevelled PDZ peptides. Nat Chem Biol 2009;5:217–9.
[16] 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.
[17] Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 1996;85:1067–76.
[18] Liu Y, Henry GD, Hegde RS, Baleja JD. Solution structure of the hDlg/SAP97 PDZ2 domain and its mechanism of interaction with HPV-18 papillomavirus E6 protein. Biochemistry 2007;46:10864–74.
[19] Elkins JM, Papagrigoriou E, Berridge G, Yang X, Phillips C, Gileadi C, et al. Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions. Protein Sci 2007;16:683–94.
[20] Im YJ, Lee JH, Park SH, Park SJ, Rho SH, Kang GB, et al. Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. J Biol Chem 2003;278:48099–104.
[21] Karthikeyan S, Leung T, Ladias JA. Structural determinants of the Na^+/H^+ exchanger regulatory factor interaction with the beta 2 adrenergic and platelet-derived growth factor receptors. J Biol Chem 2002;277:18973–8.
[22] Karthikeyan S, Leung T, Ladias JA. Structural basis of the Na^+/H^+ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 2001;276:19683–6.
[23] Murciano-Calles J, McLaughlin ME, Erijman A, Hooda Y, Chakravorty N, Martinez JC, et al. Alteration of the C-terminal ligand specificity of the Erbin PDZ domain by allosteric mutational effects. J Mol Biol 2014. http://dx.doi.org/10.1016/j.jmb.2014.05.003.
[24] Luck K, Trave G. Phage display can select over-hydrophobic sequences that may impair prediction of natural domain-peptide interactions. Bioinformatics 2011;27:899–902.
[25] Gianni S, Haq SR, Montemiglio LC, Jurgens MC, Engstrom A, 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.
[26] McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature 2012;491:138–42.
[27] 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.
[28] Mu Y, Cai P, Hu S, Ma S, Gao Y. Characterization of diverse internal binding specificities of PDZ domains by yeast two-hybrid screening of a special peptide library. PLoS One 2014;9:e88286.
[29] Wawrzyniak AM, Vermeiren E, Zimmermann P, Ivarsson Y. Extensions of PSD-95/discs large/ZO-1 (PDZ) domains influence lipid binding and membrane targeting of syntenin-1. FEBS Lett 2012;586:1445–51.
[30] Hillier BJ, Christopherson KS, Prehoda KE, Bredt DS, Lim WA. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science 1999;284:812–5.
[31] Chen Y, Sheng R, Kallberg M, Silkov A, Tun MP, Bhardwaj N, et al. Genome-wide functional annotation of dual-specificity protein- and lipid-binding modules that regulate protein interactions. Mol Cell 2012;46:226–37.
[32] Ivarsson Y, Wawrzyniak AM, Wuytens G, Kosloff M, Vermeiren E, Raport M, et al. Cooperative phosphoinositide and peptide binding by PSD-95/discs large/ZO-1 (PDZ) domain of polychaetoid, *Drosophila* zonulin. J Biol Chem 2011;286:44669–78.
[33] Zimmermann P, Meerschaert K, Reekmans G, Leenaerts I, Small JV, Vandekerckhove J, et al. PIP(2)-PDZ domain binding controls the association of syntenin with the plasma membrane. Mol Cell 2002;9:1215–25.
[34] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CW, Sweet RM, editors. Methods Enzymol. New York: Academic Press; 1997. p. 307–26.
[35] Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 2006;103:8060–5.
[36] McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. Acta Crystallogr Sect D Biol Crystallogr 2005;61:458–64.
[37] Elkins JM, Gileadi C, Shrestha L, Phillips C, Wang J, Muniz JR, et al. Unusual binding interactions in PDZ domain crystal structures help explain binding mechanisms. Protein Sci 2010;19:731–41.
[38] Karthikeyan S, Leung T, Birrane G, Webster G, Ladias JA. Crystal structure of the PDZ1 domain of human Na(+)/H(+) exchanger regulatory factor provides insights into the mechanism of carboxyl-terminal leucine recognition by class I PDZ domains. J Mol Biol 2001;308:963–73.
[39] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2126–32.
[40] Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D Biol Crystallogr 1997;53:240–55.
[41] Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375–83.
[42] Tonikian R, Zhang Y, Boone C, Sidhu SS. Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. Nat Protoc 2007;2:1368–86.