THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 31, pp. 22312–22320, August 4, 2006

Comparative Structural Analysis of the Erbin PDZ Domain and the First PDZ Domain of ZO-1

*INSIGHTS INTO DETERMINANTS OF PDZ DOMAIN SPECIFICITY**

Received for publication, March 28, 2006, and in revised form, May 26, 2006 Published, JBC Papers in Press, May 31, 2006, DOI 10.1074/jbc.M602901200

Brent A. Appleton[‡], Yingnan Zhang[‡], Ping Wu[‡], Jian Ping Yin[‡], Walter Hunziker[§], Nicholas J. Skelton[¶], Sachdev S. Sidhu[‡1], and Christian Wiesmann[‡2]

From the Departments of [‡]Protein Engineering and [¶]Medicinal Chemistry, Genentech, Inc., South San Francisco, California 94080 and the [§]Epithelial Cell Biology Laboratory, Institute of Molecular and Cell Biology, Singapore 138673, Republic of Singapore

We report a structural comparison of the first PDZ domain of ZO-1 (ZO1-PDZ1) and the PDZ domain of Erbin (Erbin-PDZ). Although the binding profile of Erbin-PDZ is extremely specific ([D/E][T/S]WV$_{COOH}$), that of ZO1-PDZ1 is similar ([R/K/S/T][T/S][W/Y][V/I/L]$_{COOH}$) but broadened by increased promiscuity for three of the last four ligand residues. Consequently, the biological function of ZO-1 is also broadened, as it interacts with both tight and adherens junction proteins, whereas Erbin is restricted to adherens junctions. Structural analyses reveal that the differences in specificity can be accounted for by two key differences in primary sequence. A reduction in the size of the hydrophobic residue at the base of the site^0 pocket enables ZO1-PDZ1 to accommodate larger C-terminal residues. A single additional difference alters the specificity of both site^{-1} and site^{-3}. In ZO1-PDZ1, an Asp residue makes favorable interactions with both Tyr^{-1} and Lys/Arg^{-3}. In contrast, Erbin-PDZ contains an Arg at the equivalent position, and this side chain cannot accommodate either Tyr^{-1} or Lys/Arg^{-3} but, instead, interacts favorably with Glu/Asp^{-3}. We propose a model for ligand recognition that accounts for interactions extending across the entire binding site but that highlights several key specificity switches within the PDZ domain fold.

PDZ[3] (PSD-95/Discs-large/ZO-1) domains are compact globular modules that typically recognize specific C-terminal motifs and, in so doing, assemble multicomponent protein complexes inside eukaryotic cells (1, 2). PDZ domains are usually embedded in larger multidomain scaffold proteins that often include multiple PDZ domains and other protein-binding modules. Thus, the biological function of each PDZ domain is determined by its intrinsic ligand specificity and also by the context in which it interacts with other cellular components. To understand how the diverse members of the PDZ domain family have adapted to their particular biological roles, it will be important to define accurately the specificity of each domain and the structural basis for ligand recognition. To this end, we have conducted comparative structural and functional studies of the PDZ domains of the human LAP (leucine-rich repeats and PDZ domains) family member Erbin (3, 4) and the MAGUK (membrane-associated guanylate kinase) family member zonula occludens-1 (ZO-1) (5), as these two proteins offer an opportunity to examine both the similarities and differences between domains that have adapted to similar but distinct biological functions.

In an accompanying article (6), we used phage-displayed peptide libraries to define the binding specificities of the single PDZ domain of Erbin (Erbin-PDZ) and two of the three PDZ domains of ZO-1, along with those of the other human LAP family members Densin-180 and Scribble (or Scrib) (Fig. 1). We found that Erbin-PDZ exhibits a strict preference for particular types of residues at each of the last four ligand positions[4] and thus recognizes a highly conserved core binding motif ([E/D][T/S]WV$_{COOH}$). Interestingly, the first PDZ domain of ZO-1 (ZO1-PDZ1) recognizes a core binding motif ([T/S/K/R][T/S][W/Y][V/L/I]$_{COOH}$) that resembles that of Erbin-PDZ but also exhibits significant differences. In particular, it appears that the ligand binding profile of ZO1-PDZ1 is expanded in comparison with that of Erbin-PDZ by virtue of increased promiscuity for three of the last four ligand residues. Specifically, both domains accept ligands containing Val at position^0, but only ZO1-PDZ1 is able to accept the larger Leu/Ile residues with comparable affinity. At site^{-1}, Erbin-PDZ exhibits a strict preference for Trp side chains, whereas ZO1-PDZ1 is bispecific, accepting either Trp or Tyr but discriminating against other residues. Finally, the two domains differ dramatically in their preferences at site^{-3}. Erbin-PDZ exhibits a preference for acidic residues, whereas ZO1-PDZ1 exhibits bispecific character at this site, preferring either basic residues or small, polar Thr/Ser residues.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (codes 2H2B, 2H2C, 2H3L, 2H3M) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] To whom correspondence may be addressed: Dept. of Protein Engineering, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080. Tel.: 650-225-1056; Fax: 650-225-3734; E-mail: sidhu@gene.com.

[2] To whom correspondence may be addressed: Dept. of Protein Engineering, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080. Tel.: 650-225-7484; Fax: 650-225-3734; E-mail: chw@gene.com.

[3] The abbreviations used are: PDZ, PSD-95/Discs-large/ZO-1; GUK, guanylate kinase; LAP, leucine-rich repeats and PDZ; MAGUK, (membrane-associated guanylate kinase; PDB, Protein Data Bank; Scrib-PDZ*n*, the *n*th PDZ domain of human Scribble; ZO, zonula occludens; ZO1-PDZ*n*, the *n*th PDZ domain of human ZO-1; r.m.s.d., root mean square deviation.

[4] The C terminus of a PDZ domain ligand is designated as position^0, and the remaining positions are numbered with negative integers for which the absolute value decreases toward the C terminus. The corresponding binding sites on the PDZ domain are numbered in a corresponding manner as site^0, site^{-1}, etc.

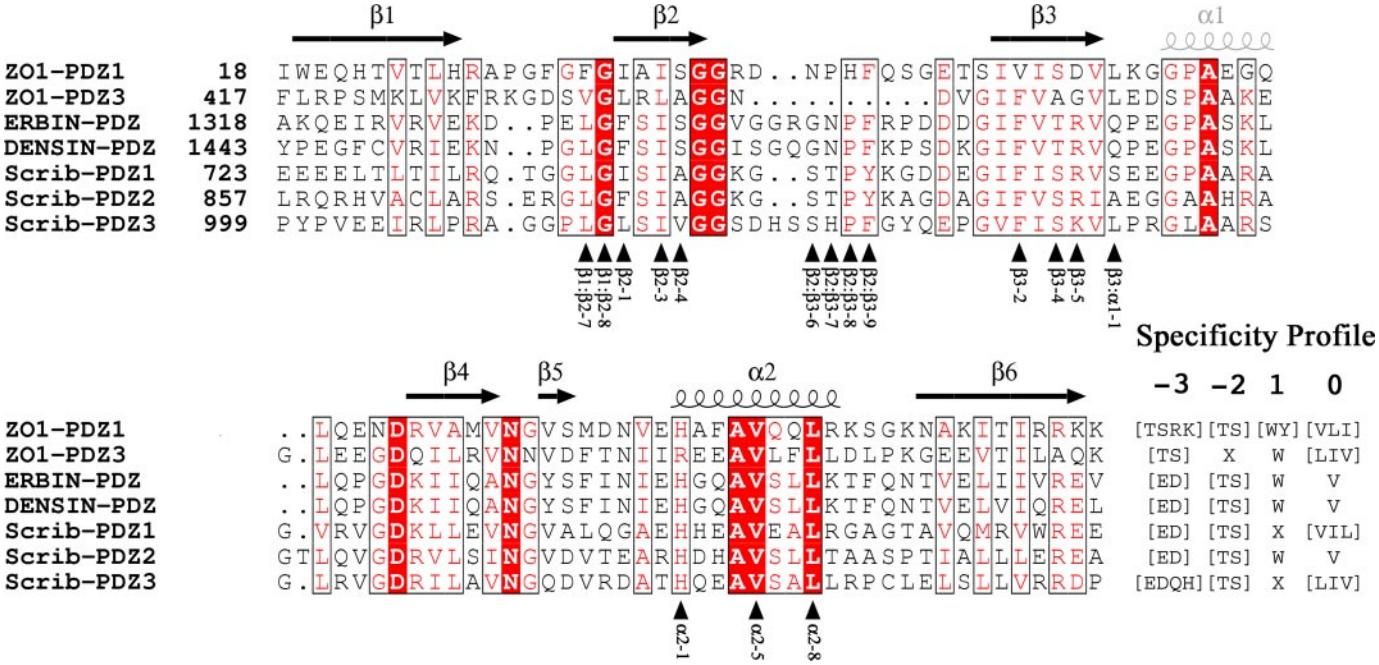

FIGURE 1. **Structure-based sequence alignment and specificity profiles of LAP and ZO-1 PDZ domains.** The sequence alignment delineates several PDZ domains and their ligand binding profiles, which are detailed in an accompanying paper (6). The numbering of the starting position of each domain within the full-length protein sequence is shown following the name. Specificity profiles for the last four residues are *highlighted* at the end of the alignment. Secondary structure elements are indicated *above* the alignment and refer to the ZO1-PDZ1 structure. Helix α1 (*light gray*) is observed in the NMR structures of Scrib-PDZ1 and -PDZ2 (PDB codes 1X5Q and 1UJU, respectively) but is absent in Erbin-PDZ and ZO1-PDZ1. Each residue is denoted and numbered according to its position in an element of secondary structure; within loop regions, the residues are numbered according to the longest loop in the alignment. The *black arrowheads* under the sequence indicate residues that are involved in ligand binding as detailed in Fig. 4. The figure was produced with ESPript (44).

These differences and similarities in ligand specificity reflect the overlapping yet distinct subcellular localizations and biological interactions of Erbin and ZO-1. Erbin and the other LAP family members are involved in the organization of junctional complexes and in the establishment and maintenance of cell polarity; and in polarized epithelia, they are localized on the lateral membrane and at adherens junctions (7–9). Notably, we and others have shown that Erbin-PDZ interacts strongly with three members of the p120-like catenin family (δ-catenin, ARVCF (armadillo protein deleted in velo-cardiofacial syndrome), and p0071) (10–12), which contain a conserved C-terminal sequence ($DSWV_{COOH}$) that closely matches the optimal binding motif defined by phage display. These p120-like catenins also associate with other catenins and E-cadherin to form the cadherin-catenin complex, a major component of adherens junctions (13).

In non-epithelial cells and nonpolarized epithelial cells, ZO-1 also associates with adherens junction markers (14–18), and its localization resembles that of Erbin. Indeed, it has been shown that, under these conditions, ZO-1 co-localizes with E-cadherin, and like Erbin-PDZ, ZO1-PDZ1 binds to the C terminus of ARVCF (11, 16). However, as polarization proceeds and distinct junctional complexes form, ZO-1 segregates away from E-cadherin and localizes exclusively to the tight junctions that form apical to the adherens junctions (14, 19). The recruitment of ZO-1 to the tight junctions is mediated by interactions between ZO1-PDZ1 and the C termini of claudins (20, 21), which are the major structural proteins of tight junctions and for the most part share a common C-terminal motif (YV_{COOH}) (6). In contrast, Erbin and the other LAP family members are excluded from the tight junctions and remain confined to the adherins junctions and the basolateral layer (7–9).

Thus, it appears that the bispecific character of the −1 and −3 sites of ZO1-PDZ1 reflects the bifunctional nature of ZO-1 in the cellular environment. In nonpolarized cells, ZO1-PDZ1 is able to interact at adherens juctions with p120-like catenins that contain Trp^{-1} (16), whereas in polarized cells it is able to interact at tight junctions with claudins that contain Tyr^{-1} (6, 21). On the other hand, Erbin-PDZ has a strict preference for Trp^{-1}, and even in polarized cells, Erbin does not interact with claudins but, rather, remains associated with p120-like catenins at the adherens junctions (10–12). These differences in cellular localization are also likely to be enforced by differences in $site^{-3}$. Erbin-PDZ prefers acidic residues at $position^{-3}$, and all three p120-like catenins described above contain an Asp at this position. In contrast, acidic residues are conspicuously absent at $position^{-3}$ among the 20 known human claudins, which typically contain either basic residues or Thr/Ser at this position, in good agreement with the binding specificity of ZO1-PDZ1 (6).

In contrast with earlier studies that emphasized the importance of $site^{0}$ and $site^{-2}$ (22–24), more recent studies, including the work described above, make it clear that PDZ domains recognize ligands through interactions that extend across the entire surface of the binding groove (25–33). According to the established PDZ domain classification system based on $site^{0}$ and $site^{-2}$, both Erbin-PDZ and ZO1-PDZ1 exhibit typical class I specificities ($X[T/S]X\Phi_{COOH}$, where Φ is a hydrophobic residue). However, it is clear that the ligand-binding specificities of these domains depend critically on interactions with all four C-terminal residues and, importantly, that differences in

the specificities of site^{-1} and site^{-3} are critical determinants of functional convergence and divergence.

Herein we report a comparative structural analysis aimed at elucidating the molecular basis for the similarities and differences between the binding specificities of Erbin-PDZ and ZO1-PDZ1. We solved the unliganded crystal structures of both domains, as well as high resolution structures of ZO1-PDZ1 in complex with two different phage-derived ligands designed to investigate the broadened specificity of ZO1-PDZ1. Together with a previously reported structure of Erbin-PDZ bound to an optimal ligand (29), these structures allow for a detailed comparison of the ligand-binding sites of the two domains. Remarkably, the analysis reveals that the differences in the site^{-1} and site^{-3} specificities of Erbin-PDZ and ZO1-PDZ1 are likely caused by a single amino acid difference between the two domains, and furthermore, the differences in site^{0} can be accounted for by one additional difference. Overall, the results support a model for PDZ domain-ligand interactions that takes into account interactions that extend across the entire binding site but depends critically on several key specificity switches within the PDZ domain fold.

EXPERIMENTAL PROCEDURES

Protein Purification—Erbin-PDZ (residues 1314–1422) was purified as described previously (29). For ZO1-PDZ1, a DNA fragment encoding for residues 18–110 of human ZO-1 was cloned into the NdeI/BamHI sites of the pET22d expression vector, and this created an open reading frame encoding for ZO1-PDZ1 with an N-terminal His tag and a thrombin cleavage site. In addition, standard molecular biology techniques were used to fuse 10-residue extensions to the C terminus of ZO1-PDZ1 to produce open reading frames encoding for ZO1-PDZ1-WV (extension, GGGWRRTTWV) and ZO1-PDZ1-YL (extension, GGGWRRTTYL). For each ZO-1 protein fragment, *Escherichia coli* BL21(DE3) cultures harboring the expression plasmid were grown at 37 °C to mid-log phase (A_{600} = 0.8). Protein expression was induced with 0.4 mM isopropyl 1-thio-β-D-galactopyranoside, and the culture was grown at 27 °C for 16 h. The bacteria were pelleted by centrifugation at 4,000 × *g* for 15 min, washed twice with 20 mM Tris-HCl (pH 8.0), and frozen at −80 °C for 8 h. The pellet was resuspended in 100 ml of buffer A (50 mM Tris-HCl (pH 8.0) and 500 mM NaCl), and the bacteria were lysed by passing through Microfluidizer® processing equipment. The cell lysate was loaded onto a nickel-nitrilotriacetic acid-agarose column (Qiagen). The column was washed with buffer A plus 20 mM imidazole, and the protein was eluted with 250 mM imidazole in buffer A. Fractions containing the protein of interest were pooled, thrombin was added (1 unit/mg protein), and the sample was dialyzed overnight against phosphate-buffered saline at 4 °C. The protein sample was concentrated and further purified over a Superdex-75 column in 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 5 mM β-mercaptoethanol.

Crystallization and Structure Determination—Protein samples were concentrated to 20 mg/ml (Erbin-PDZ) or 6–7 mg/ml (ZO1-PDZ1 proteins). Crystals were grown by vapor diffusion at 19 °C from sitting drops composed of 1 μl of protein and 1 μl of the appropriate reservoir solution. The following reservoir solutions were used: Erbin-PDZ (0.1 M Tris-HCl, 35% polyethylene glycol 4000 (pH 7.5–8.5)), ZO1-PDZ-WV (0.1 M sodium acetate, 2.0 M sodium formate (pH 4.6)), ZO1-PDZ1 and ZO1-PDZ1-YL (0.1 M sodium acetate, 0.2 M ammonium sulfate, 30% polyethylene glycol 2000 monomethyl ether (pH 4.6)). For cryoprotection, each crystal was soaked briefly in a mixture of 75% reservoir solution and 25% ethylene glycol and subsequently flash-frozen in liquid nitrogen.

Diffraction data were collected at beamline 5.0.1 (Advanced Light Source, Berkeley, CA) for the ZO1-PDZ1-WV and ZO1-PDZ1-YL crystals, at beamline 9-2 (Stanford Synchrotron Radiation Laboratory, Palo Alto, CA) for the Erbin-PDZ crystal, or on an in-house Cu-Kα source for the ZO1-PDZ1 crystal (Table 1). Data were processed with DENZO and SCALEPACK from the HKL suite (34). The structures were solved by molecular replacement with AMoRe (35) or PHASER (36) using the PDZ domain from a liganded Erbin-PDZ structure (PDB code 1MFG) (37) as the search model for the unliganded Erbin-PDZ structure, our refined Erbin-PDZ structure as the search model for the unliganded ZO1-PDZ1 structure, and the refined coordinates of the ZO1-PDZ1 structure as the search model for the liganded ZO1-PDZ1 structures. Atomic models were built with multiple rounds of manual inspection using XtalView (38) or O (39) and refined by the maximum likelihood method with REFMAC (40).

Each of the three ZO1-PDZ1 structures contains one molecule per asymmetric unit; the Erbin-PDZ structure contains two molecules per asymmetric unit. The ZO1-PDZ1-WV and ZO1-PDZ1-YL structures form dimers. The 2-fold axes relating the molecules within these dimers to each other coincide in both cases with a crystallographic 2-fold axis. In each dimer, the two PDZ monomers are arranged in a head-to-tail fashion where each PDZ domain binds to the C-terminal peptide of its symmetry-related binding partner. For all of the ZO1-PDZ1 structures, the entire PDZ domain is well defined in the electron density. In the Erbin-PDZ structure, the entire PDZ domain is visible in the first molecule, but the last residue of the C terminus is disordered in the second.

RESULTS

Crystallographic Analysis of Erbin-PDZ and ZO1-PDZ1—We determined the crystal structures of unliganded ZO1-PDZ1 and Erbin-PDZ, and the structures of ZO1-PDZ1 in complex with two high affinity heptapeptides (WRRTTWV-$_{COOH}$ and WRRTTYL$_{COOH}$) (Table 1) identified by C-terminal phage display (6). The heptapeptides differ at the two C-terminal positions and were chosen for structural analysis to investigate the subtle differences between the specificities of ZO1-PDZ1 and Erbin-PDZ. To facilitate structural studies, we used a construct in which the peptides of interest were fused to the C terminus of ZO1-PDZ1 via a tri-glycine linker. This strategy was used previously to determine the crystal structures of various PDZ domain-ligand complexes (27, 28, 37) and, in the case of ZO1-PDZ1, resulted in the formation of crystallographic dimers in which each fused ligand binds to the apposing PDZ domain (Fig. 2*A*). The crystal structures for the resulting fusion proteins are subsequently referred to as ZO1-PDZ1-WV or ZO1-PDZ1-YL for

TABLE 1
Data collection and refinement statistics
ALS, Advanced Light Source; SSRL, Stanford Synchrotron Radiation Laboratory; ASU, asymmetric unit.

	ZO1-PDZ1-YL	ZO1-PDZ1-WV	ZO1-PDZ1	Erbin-PDZ
Data collection				
Source	ALS 5.0.1	ALS 5.0.1	Home source	SSRL 9-2
PDB code	2H2B	2H2C	2H3M	2H3L
Resolution (Å)[a]	50-1.6 (1.7-1.6)	50-2.0 (2.1-2.0)	30-2.8 (2.9-2.8)	30-1.0 (1.04-1.00)
Space group	$P4_12_12$	$P3_121$	H32	$P2_1$
Cell parameters	*a, b* = 52.4 *c* = 92.9	*a, b* = 51.6 *c* = 88.3	*a, b* = 62.8 *c* = 154	*a* = 38.9 *b* = 44.0 *c* = 57.3 β = 106.2°
Molecules/ASU	1	1	1	2
Unique reflections	17,606	9,658	3,037	100,042
Redundancy	7.3	5.8	9.4	3.2
Completeness (%)[a]	99.6 (98.4)	99.8 (100)	99.8 (98.0)	99.4 (97.6)
R_{sym}[a,b]	0.052 (0.401)	0.059 (0.310)	0.076 (0.544)	0.071 (0.517)
$\langle I/\sigma I\rangle$[a]	31.9 (3.7)	29.8 (4.7)	34.2 (2.6)	19.5 (1.6)
Refinement				
Resolution	30-1.6	40-2.0	30-2.9	30-1.0
R_{cryst}, R_{free}[c]	0.203, 0.232	0.209, 0.225	0.227, 0.294	0.148, 0.169
Non-hydrogen atoms				
Protein	824	825	722	1,542
Solvent	113	43	10	269
Average B-factor (Å^2)	27.6	52.4	45.6	12.9
r.m.s.d. bonds (Å)	0.012	0.012	0.010	0.011
r.m.s.d. angles (°)	1.36	1.16	1.15	1.52
Ramachandran plot (%)[d]	90.9/9.1/0/0	89.8/10.2/0/0	89.7/10.3/0/0	90.1/8.6/1.2/0

[a] Values in parentheses refer to data in the highest resolution shell.
[b] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where $\langle I\rangle$ is the average intensity of symmetry-related observations of a unique reflection.
[c] $R_{cryst} = R_{free} = \Sigma|F_o - F_c|/\Sigma F_o$, where R_{free} represents 5% of the data selected randomly.
[d] Values represent the percentage of residues in the most favored, additionally allowed, generously allowed, and disallowed regions of the Ramachandran plot, respectively (43).

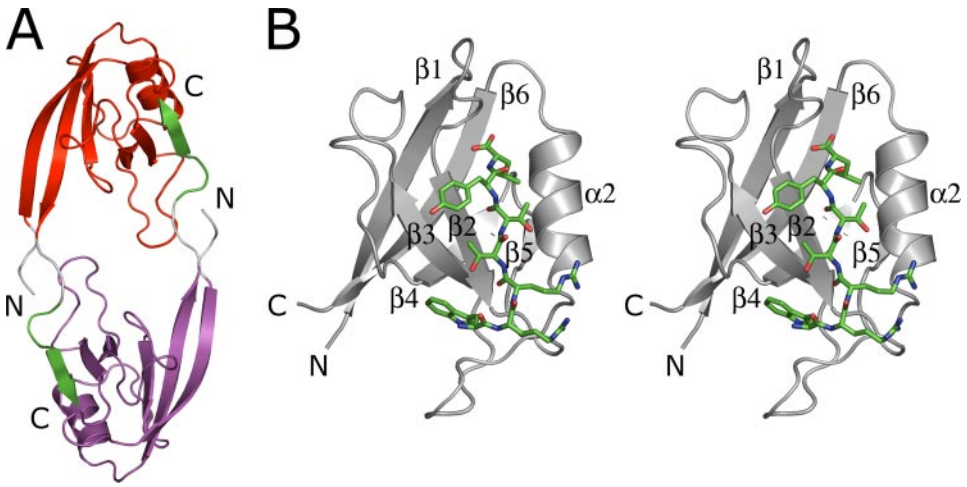

FIGURE 2. **Overall structure of ZO1-PDZ1.** *A*, crystallographic dimer of ZO1-PDZ1-YL with the PDZ domains colored *red* and *magenta* and the heptapeptide ligands colored *green*. Regions in *gray* represent the tri-glycine linker between the PDZ domain and the ligand and a tetrapeptide that was fused to the N terminus as a result of the cloning procedures. *B*, stereoscopic representation of ZO1-PDZ1-YL with secondary structure elements labeled (PDZ domain, *gray*; heptapeptide, *green*). Structural figures were produced with PyMOL (DeLano Scientific, San Carlos, CA).

the structures of the complex with $WRRTTWV_{COOH}$ or $WRRTTYL_{COOH}$, respectively. All four structures presented here were solved by molecular replacement and are well refined with *R*-values ranging from 14.8 to 22.7% and with good geometry. In particular, the ZO1-PDZ1-YL and -WV structures were solved at 1.6 and 2 Å (Table 1), respectively, and thus provide high enough resolution to allow commenting on critical differences of the bound ligands. This enables us to suggest how specificity among related PDZ domains is achieved on a structural basis.

Overall Structure—The canonical PDZ domain fold is characterized by a six-stranded β-sandwich that is capped with two α-helices. Helix α1 usually follows strand β3. However, similar to previously reported structures of Erbin-PDZ (29, 37), ZO1-PDZ1 lacks helix α1 (Fig. 2*B*). Helix α2 follows strand β5 and is present in all PDZ domains. Helix α2 and strand β2 form either side of the conserved ligand-binding cleft within the PDZ domain fold. In the liganded ZO1-PDZ1 structures, the heptapeptides are bound in this cleft and are well defined in the electron density. The temperature factors of the bound ligands are among the lowest in the entire structure with the exception of the guanidinium groups of Arg^{-5} and Arg^{-6}, which are mobile in both structures.

To facilitate the comparison of PDZ structures, we have adopted a naming convention that recognizes structurally equivalent residues (Fig. 1). In this numbering scheme, residue β2-1 for example, represents the first residue in strand β2, whereas β2:β3-1 indicates the first amino acid in the loop between strands β2 and β3. Because of numerous insertions and deletions within the loop regions of many PDZ domains, it is not possible to unequivocally assign structurally equivalent residues in these regions. Therefore, the numbering represents the lengths of the longest loops used in our alignment.

At the high protein concentrations used during crystallization, the ligand-binding grooves of PDZ domains are often occupied by the C termini of symmetry-related molecules

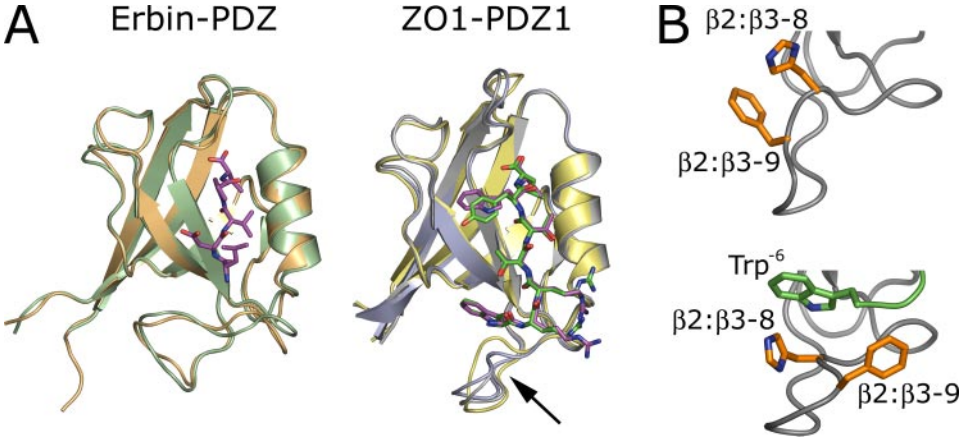

FIGURE 3. **Superposition of ZO1-PDZ1 and Erbin-PDZ structures.** *A*, the *left panel* shows the superposition of structures for unliganded (*light orange*) and liganded Erbin-PDZ (PDZ, *green*; peptide, *magenta*; PDB 1MFL). The *right panel* displays the superposition of structures for unliganded (*yellow*) and liganded ZO1-PDZ1 (ZO1-PDZ-YL: PDZ, *gray*; peptide, *green*; ZO1-PDZ-WV: PDZ, *light blue*; peptide, *magenta*). The *arrow* indicates the position of the β2:β3 loop. *B*, the conformation of the β2:β3 loop of ZO1-PDZ1 differs between the unliganded (*top*) and liganded (*bottom*) forms. This difference is due to rotation about the peptide backbone at residue β2:β3-9 and results in the formation of a hydrophobic pocket that is occupied by Trp^{-6} in the liganded form.

within the crystal lattice (23, 26, 41). Fortunately, the peptide-binding grooves in the unliganded ZO1-PDZ1 and Erbin-PDZ structures presented here are empty and thus provide an opportunity to compare the unliganded and liganded forms of these domains. Crystallographic analysis of the sixth PDZ domain of glutamate receptor-interacting protein (GRIP-PDZ6) revealed that ligand binding induces a reorientation of helix α2 (26), suggesting that PDZ domains might undergo conformational changes to accommodate ligands. In contrast, superposition of the unliganded and liganded structures of the third PDZ domain of PSD-95 (22) and those of the Shank1 PDZ domain (25) did not show significant conformational changes upon ligand binding. Similar to the Shank1 PDZ domains, our unliganded Erbin-PDZ structure has root mean square deviation (r.m.s.d.) values of 0.6 to 0.7 Å when compared with two Erbin-PDZ structures in complex with low affinity peptides (Fig. 3*A*) (37), indicating that no major structural changes are induced upon ligand binding. Similarly, the three ZO1-PDZ1 structures presented here superimpose with root mean square deviation values of only 0.6 to 0.9 Å for 93 equivalent Cα atoms, and aside from subtle movement in the β2:β3 loop, there are no structural differences. The movement in the β2:β3 loop is induced by the rotation of Phe(β2:β3-9) around the peptide backbone, which results in a 5-Å movement of its Cα atom (Fig. 3*B*). In summary, it seems that ligand binding does not induce conformational changes in most PDZ domains studied thus far.

Comparison of Site^0 and Site^{-2} of Erbin-PDZ and ZO1-PDZ1—To better understand the molecular basis for ligand recognition and selectivity by PDZ domains, we compared the liganded structures of ZO1-PDZ1 bound to peptides $WRRTTWV_{COOH}$ or $WRRTTYL_{COOH}$ with an NMR structure of Erbin-PDZ bound to a high affinity peptide ($TGWETWV_{COOH}$) (29). In each structure, the PDZ domain forms extensive contacts with the bound ligand. Some of these interactions are observed in most PDZ domains, whereas other contacts seem to be specific to class I PDZ domains preferring Thr/Ser at position^{-2}. In all three structures, the Thr^{-2} hydroxyl group is able to form a hydrogen bond to $N^{\epsilon 2}$ of His(α2-1), whereas the aliphatic portion of the side chain is in van der Waals distance to Val(α2-5) (Fig. 4*A*). These interactions are consistent with the functional analyses of these PDZ domains; they are also in agreement for the PDZ domains of the other LAP family members Densin-180 and Scribble, which also contain His and Val at α2−1 and α2-5, respectively, and show a strong preference for Thr/Ser^{-2} (6). Notably, the third PDZ domain of ZO-1 (ZO1-PDZ3) exhibits no specificity at site^{-2}; this is the only domain that contains an Arg at α2-1 (Fig. 1).

At site^0, all three structures are very closely related in the way that the C-terminal carboxylate group of the ligand forms hydrogen bonds with the main chain amides of the carboxylate binding loop (β1:β2-7, β1:β2-8, and β2-1). Interactions of this type are conserved among all PDZ domains that engage ligands with free C termini (22, 28, 29). However, the structures differ significantly in terms of the interactions that mediate recognition of the C-terminal side chain (Fig. 4*A*). In each liganded ZO1-PDZ1 structure, the side chain of Val^0 or Leu^0 is buried in a deep hydrophobic pocket that is formed predominantly by residues in the β1:β2 loop and strand β2 (Phe(β1:β2-7), Ile(β2-1), Ile(β2-3)) but also involves residues in helix α2 (Val(α2-5), Leu(α2-8)). However, depending on the bound ligand, Ile(β2-3) adopts different rotamers around the Cα–Cβ bond to accommodate either Val^0 or Leu^0. Erbin-PDZ also contains an Ile at β2-3, but position β2-1 is occupied by a larger Phe rather than an Ile. In comparison with ZO1-PDZ1, the bulkier Phe(β2-1) side chain reduces the volume of the site^0 pocket of Erbin-PDZ and restricts the movement of the Ile side chain at β2-3, making the pocket less accommodating for ligands with larger side chains at position^0. It is striking that the PDZ domains that have a stringent preference for a Val^0 side chain (Erbin-PDZ, Densin-PDZ, and Scrib-PDZ2) contain a Phe in position β2-1, whereas those that can accommodate a larger Leu/Ile at site^0 (Scrib-PDZ1, Scrib-PDZ3, ZO1-PDZ1, and ZO1-PDZ3) contain a Leu/Ile at position β2-1 (Fig. 1). Thus, it appears that generally the nature of the hydrophobic residue at β2-1 is a key determinant of the site^0 specificity of PDZ domains.

Comparison of Site^1 and Site^{-3} of Erbin-PDZ and ZO1-PDZ1—Erbin-PDZ displays a strict preference for Trp at site^{-1}, whereas in contrast, ZO1-PDZ1 is bispecific and can accommodate Trp or Tyr equally well. A comparison of the various PDZ domain structures bound to ligands suggests a likely explanation for this biologically important difference in specificity (Fig. 4*B*). In both domains, Trp^{-1} occupies a very similar position, and the side chain extends across strand β2 and inserts between the side chains of β3-5 and β3:α1-1. Alanine-

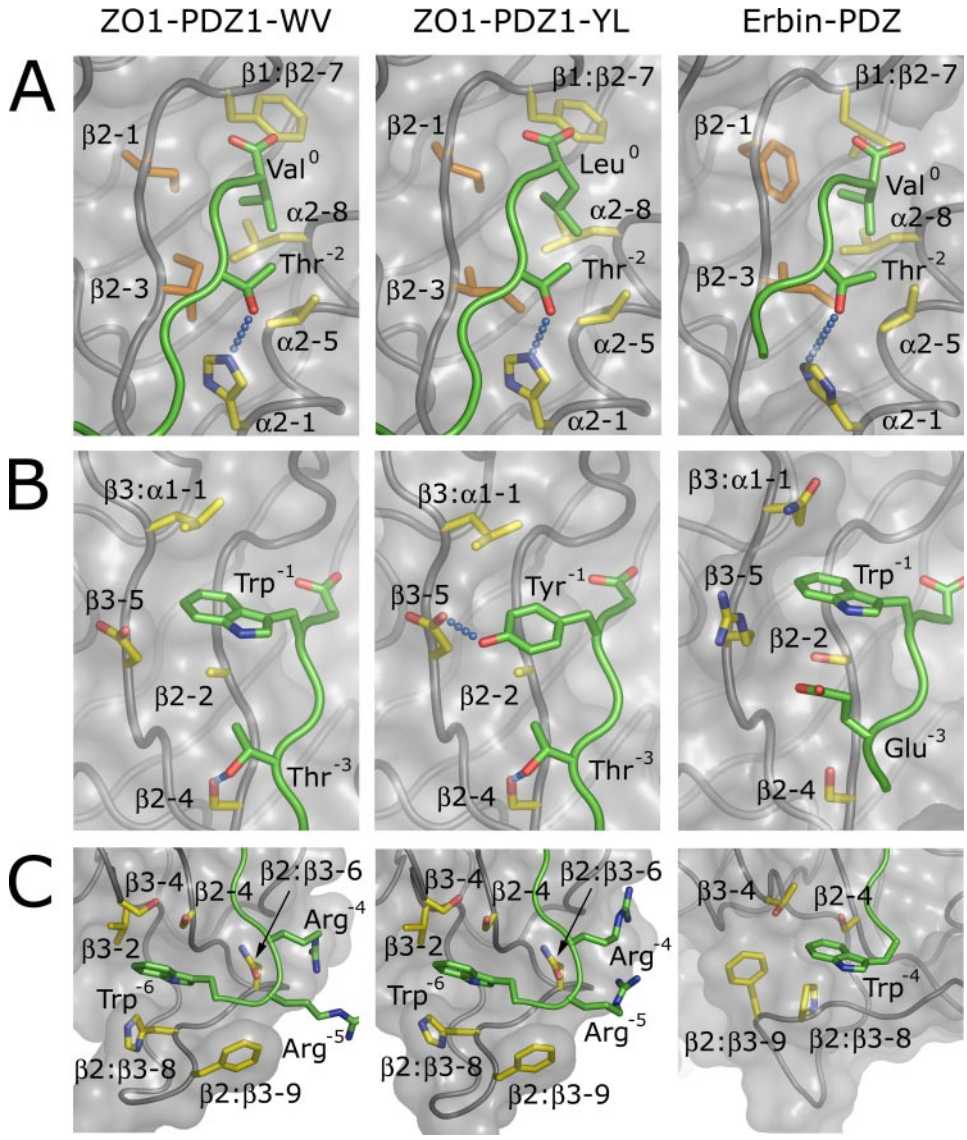

FIGURE 4. **Comparison of the ligand-binding subsites of ZO1-PDZ1 and Erbin-PDZ.** In each horizontal gallery, the three structures are shown in the same relative orientation. Main chain traces are rendered as *gray* or *green tubes* for the PDZ domain or peptide ligand, respectively. Side chains are rendered as *sticks*, and carbon atoms are colored *green* or *yellow* for the ligand or PDZ domain, respectively, except where noted. Side chain oxygen and nitrogen atoms are colored *red* and *blue*, respectively. Hydrogen bonds are shown schematically as *blue dashes*. *A*, the interactions at site^0 and site^{-2} involve residues from the $\beta1$:$\beta2$ loop, strand $\beta2$, and helix $\alpha2$. ZO1-PDZ1, which contains an Ile at $\beta2$-1 (*orange*), can accommodate Val, Leu, or Ile at site^0, whereas Erbin-PDZ (PDB 1N7T), which contains a Phe at $\beta2$-1, shows a strict preference for Val^0. Additionally, Ile($\beta2$-3) (*orange*) of ZO1-PDZ1 adopts different rotamers to accommodate either Val^0 or Leu^0. *B*, both Erbin-PDZ and ZO1-PDZ1 accommodate Trp^{-1}, which extends over strand $\beta2$. However, ZO1-PDZ1 also recognizes Tyr^{-1} because of a hydrogen bond formed with the carboxylate group of Asp($\beta3$-5). In contrast, Erbin-PDZ contains an Arg at $\beta3$-5, which instead mediates a salt bridge with Glu^{-3}. The ZO1-PDZ1 structures also display a hydrogen bond between the hydroxyl atoms of Ser($\beta2$-4) and Thr^{-3}. *C*, ZO1-PDZ1 and Erbin-PDZ display a preference for a Trp at positions −6 and −4, respectively. Although these interactions are dependent on residues in the $\beta2$:$\beta3$ loop of both domains, the molecular details are different (see "Results," under "Contributions of the $\beta2$: $\beta3$ Loop to Ligand Recognition, for details).

scanning mutagenesis of Erbin-PDZ has shown that favorable recognition of the Trp^{-1} side chain does not derive from contacts with PDZ domain side chains but, rather, from contacts with the main chain of strands $\beta2$ and $\beta3$ (29). The crystal structure of ZO1-PDZ1-WV shows as well that most interactions of ZO1-PDZ with the Trp^{-1} side chain are mediated by main chain interactions. In the ZO1-PDZ1-YL structure, the Tyr^{-1} side chain occupies a position that is similar to that of Trp^{-1}; however, as the Tyr side chain is somewhat smaller than the Trp side chain, the overall amount of buried surface and hydrophobic interactions is decreased. In ZO1-PDZ1, this loss in binding energy is compensated for by the formation of a hydrogen bond between the hydroxyl group of Tyr^{-1} to the carboxylate group of Asp($\beta3$-5). In our accompanying article (6), we observed that ZO1-PDZ1 has similar affinities for peptides with Trp^{-1} or Tyr^{-1}. However, the replacement with Phe^{-1} results in a 10-fold loss in affinity, emphasizing the importance of the hydrogen bond between Tyr^{-1} and Asp($\beta3$-5). In contrast, Erbin-PDZ contains an Arg at position $\beta3$-5, which is not suited for the formation of a similar interaction. Thus, it appears that recognition of Tyr^{-1} by ZO1-PDZ1 is dependent on both interactions with the main chain and hydrogen bond formation with the side chain of Asp($\beta3$-5). Among the PDZ domains that we investigated by phage display (Fig. 1), only ZO1-PDZ1 contains an Asp in position $\beta3$-5, and it is the only one with a preference for both Trp^{-1} and Tyr^{-1}. All of the other domains either strongly prefer Trp^{-1} or accommodate any amino acid in this position. More detailed analysis of binding preferences in other PDZ domains will be required to determine whether an Asp in position $\beta3$-5 is a signature for PDZ domains that are bispecific at site^{-1} with a preference for both Trp and Tyr.

Erbin-PDZ and ZO1-PDZ1 differ significantly in their specificity at site^{-3}, with Erbin-PDZ exhibiting a strong preference for acidic residues and ZO1-PDZ1 preferring either Thr/Ser or Arg/Lys (6). A careful analysis of the crystal structures reveals that the specificity of site^{-3} is closely tied to the residue at position $\beta3$-5. This residue, therefore, is involved in the formation of interactions at both site^{-1} and site^{-3} (Fig. 4*B*). In the case of Erbin-PDZ, Arg($\beta3$-5) is well positioned to make favorable electrostatic interactions with a negatively charged ligand side chain at site^{-3}. This explains the preference of this domain and other LAP family PDZ domains to bind peptides that contain negatively charged residues at position^{-3}, as they all contain Arg or

Lys at position β3-5 (Fig. 1). In contrast, the Asp in position β3-5 of ZO1-PDZ1, which is responsible for the bispecific nature of this domain to accommodate either Trp or Tyr in site^{-1}, conducts the preference of this domain toward the binding of a positively charged residue in site^{-3}. Alternatively, ZO1-PDZ readily accepts Ser or Thr in this position, small residues that are able to form hydrogen bonds with the Ser in position β2-4 (Fig. 4*B*). This Ser is conserved in Erbin-PDZ (Fig. 1), but the interaction between Arg(β3-5) and a Glu/Asp^{-3} is likely to be stronger than the formation of a hydrogen bond between Ser(β2-4) and Thr/Ser^{-3}. In summary, the specificity at site^{-3} is conducted by interactions with residues in position β2-4 or β3-5 for short or long ligand side chains, respectively.

Contributions of the β2:β3 Loop to Ligand Recognition—Erbin-PDZ and ZO1-PDZ1 exhibit preferences for ligands that contain hydrophobic side chains as well as arginines and lysines upstream of position^{-3}. In both cases, recognition is mediated predominantly by residues of the β2:β3 loop, but the structural details of these interactions differ considerably (Fig. 4*C*). In the case of Erbin-PDZ, the side chain of Trp^{-4} packs against strand β2 and interacts with residues in strands β2 and β3 (Ser(β2-4), Thr(β3-4)) and the β2:β3 loop (Pro(β2:β3-8) and Phe (β2:β3-9)). In ZO1-PDZ1, the β2:β3 loop forms a number of contacts with side chains at the −4 and −6 positions, and Trp^{-6} appears to be especially important for binding (6). Unlike for Erbin-PDZ, the side chain of Arg^{-4} points toward helix α2, and the side chain of Trp^{-6} lies toward strand β2 and occupies a position similar to that of Trp^{-4} in the Erbin-PDZ ligand.

In both liganded ZO1-PDZ1 structures, the interactions with the peptide are virtually identical in this region, and it appears that the last three residues of the ligand function as a unit that interacts with the PDZ domain through both side chain and main chain interactions. The main chain nitrogen and oxygen atoms of Arg^{-4} are hydrogen-bonded to O^{δ} and N^{δ} of Asn(β2:β3-6), respectively, whereas the main chain nitrogen of Arg^{-5} participates in coordination of a water molecule with N^{δ} of the same Asn and the main chain oxygen of Pro(β2:β3-7). In addition, the aromatic side chain of Phe(β2:β3-9) is in van der Waals contact with the peptide bond between Arg^{-5} and Trp^{-6}, highlighting the rearrangement within the β2:β3 loop that occurs upon ligand binding (Fig. 3*B*). In contrast to their main chain atoms, the side chains at the −4 and −5 positions are not intimately involved in contacts with ZO1-PDZ1, and the guanidinium groups of both arginines have very high temperature factors, indicating that these side chains are rather mobile. The side chain of Arg^{-5} is completely exposed to solvent and does not form close interactions with the PDZ domain, whereas the aliphatic portion of Arg^{-4} is in proximity to the side chains of Asn(β2:β3-6) and His(α2-1). A lack of interactions involving Arg^{-5} is in good agreement with the promiscuity of ZO1-PDZ1 with respect to this ligand position, but the structural data do not explain the preference for residues with a general hydrophobic character in position^{-4} (6). The aromatic side chain of Trp^{-6} is buried in a deep pocket formed by Val(β3-2) and the aliphatic portions of Ser(β2-4), His(β2:β3-8), and Ser(β3-4). Interestingly, this pocket is not visible in the unliganded structure but, rather, forms upon ligand binding as a result of rearrangements in the β2:β3 loop (Fig. 3*B*).

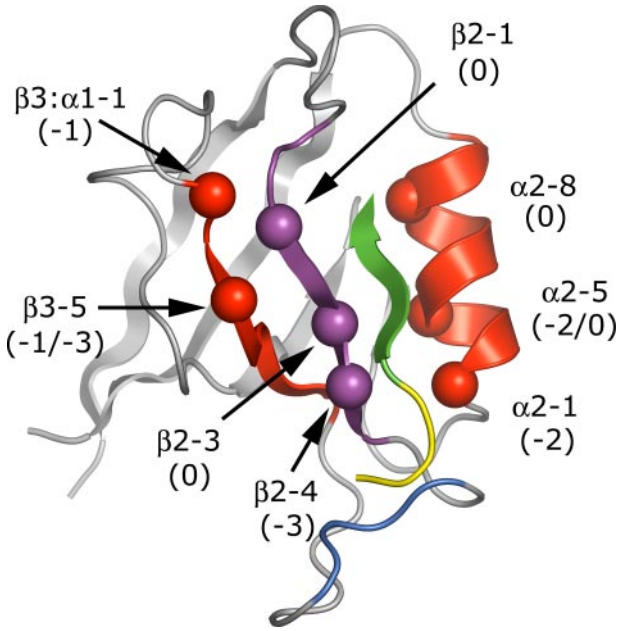

FIGURE 5. **Determinants of PDZ domain specificity.** The structure of ZO1-PDZ1-YL is shown schematically and colored to highlight the functional elements involved in ligand recognition. The peptide ligand is colored *green* (core motif) and *yellow* (auxiliary motif). The PDZ domain functional elements are colored *magenta* (primary), *red* (secondary), and *blue* (tertiary). The *spheres* indicate key side chains of the PDZ domain that contribute to recognition of side chains within the core ligand motif.

DISCUSSION

It has become apparent that PDZ domains recognize C-terminal ligands through interactions with a core ligand motif consisting of the last four residues and that affinity can be further enhanced by auxiliary interactions with residues upstream of the core motif (25–33). According to the structural and functional analyses described here and in the accompanying article (6), we propose a division of the functional elements within the PDZ domain structural framework into three groups on the basis of their contributions to ligand recognition. We have designated these groups as "primary," "secondary," and "tertiary" determinants of specificity (Fig. 5). The primary functional determinants are contained in an eight-residue stretch that includes the carboxylate binding loop and strand β2 (β1:β2-6 to 8 and β2-1 to 5). This functional element forms interactions with the C-terminal carboxylate and the main chain of the four C-terminal residues of the bound peptide. In addition, the hydrophobic side chains at β2-1 and β2-3 contribute to the recognition of the ligand side chain at position^{0}. The main chain of strand β2 is important for recognition of the side chain at the −1 position, and the side chain at β2-4 can also interact with the side chain at postion^{-3}. Importantly, the sequence of this primary functional element is highly conserved among all PDZ domains, and thus it represents an essentially invariant core of PDZ domain function.

The primary functional element is sandwiched on either side by secondary functional elements that account for the remaining interactions necessary for recognition of the core ligand

motif. On one side, the side chain at a key position within strand β3 (β3-5) is poised to contribute to both site^{-1} and site^{-3}. Furthermore, the side chain at β3:α1-1 can contribute to site^{-1} recognition. In the crystal structure of the PDZ domain from Shank1, a salt bridge is observed between Asp(β3:α1-1) and Arg^{-1} of its ligand (25). In essence, interactions involving these side chains expand upon the "default" ligand recognition provided by the primary functional element. For example, Erbin-PDZ and many other PDZ domains prefer a Trp^{-1} (6, 42). This is the default residue in this site, as it forms interactions mostly with the main chain of strand β2 and does not rely on contacts with side chains in strand β3. On the other hand, Erbin-PDZ prefers a Glu/Asp^{-3} because of favorable interactions with Arg(β3-5). In contrast, an exchange of Arg to Asp at position β3-5 in ZO1-PDZ1 relative to Erbin-PDZ expands the specificity of ZO1-PDZ1 site^{-1} to include Tyr^{-1} in addition to Trp^{-1}, as the Asp(β3-5) side chain is able form favorable interactions with the side chain of Tyr^{-1}. This exchange also alters site^{-3} specificity to favor either Thr/Ser^{-3} through interactions with Ser(β2-4) or Arg/Lys^{-3} by the formation of electrostatic interactions with Asp(β3-5). On the other side of the binding groove, the secondary functional elements are contained within helix α2. The residues in positions α2-5 and α2-8 cooperate with the primary functional elements to form the site^{0} pocket, whereas positions α2-1 and α2-5 are the key binding determinants for site^{-2}.

Residues upstream of position^{-3} constitute an auxiliary ligand motif, as interactions involving this region modulate binding affinity for ZO1-PDZ1 and all of the LAP family PDZ domains (6). In both Erbin-PDZ and ZO1-PDZ1 (Fig. 4*C*), and likely many other PDZ domains (31, 33), the recognition of this auxiliary ligand motif is mediated primarily by the β2:β3 loop. This region constitutes a tertiary functional element (Fig. 5) that interacts with the auxiliary motif and supports the binding of the core motif. The β2:β3 loop is highly divergent in sequence, length, and main chain conformation among different PDZ domains, making it difficult to develop general rules that correlate primary sequence with specificity. The structures of Erbin-PDZ and ZO1-PDZ1 explain the particular specificities of each domain for residues upstream of position^{-3} reasonably well. However, these structures are not sufficient to discern the structural and functional roles of β2:β3 loops in other PDZ domains.

In summary, the structural and functional analyses presented here and in the accompanying article (6) provide a detailed view of how subtle changes in PDZ domain structure can have a profound impact on ligand specificity and biological function. In the particular case of Erbin and ZO-1, it is striking how a few key sequence differences have endowed ZO-1 with an expanded binding profile that in turn broadens its range of biological interactions. As a result, ZO-1 assumes a bifunctional role that involves localization to either adherens junctions or tight junctions under different conditions, whereas Erbin is restricted to adherens junctions. In more general terms, our study provides a structural framework for understanding how the diverse members of the PDZ domain family have adapted to different biological roles.

Acknowledgments—We thank Charles Eigenbrot and other members of the structure group for assistance with data collection as well as the beamline support staffs at the Advanced Light Source and the Stanford Synchrotron Radiation Laboratory. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the United States Department of Energy, Office of Basic Energy Sciences. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the United States Department of Energy under Contract DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

REFERENCES

1. Harris, B. Z., and Lim, W. A. (2001) *J. Cell Sci.* **114,** 3219–3231
2. Sheng, M. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7058–7061
3. Bryant, P. J., and Huwe, A. (2000) *Nat. Cell Biol.* **2,** E141–E143
4. Santoni, M. J., Pontarotti, P., Birnbaum, D., and Borg, J. P. (2002) *Trends Genet.* **18,** 494–497
5. Gonzalez-Mariscal, L., Betanzos, A., and Avila-Flores, A. (2000) *Semin. Cell Dev. Biol.* **11,** 315–324
6. Zhang, Y., Yeh, S., Appleton, B. A., Held, H. A., Kausalya, P. J., Phua, D. C. Y., Wong, W. L., Lasky, L. A., Wiesmann, C., Hunziker, W., and Sidhu, S. S. (2006) *J. Biol. Chem.* **281,** 22299–22311
7. Navarro, C., Nola, S., Audebert, S., Santoni, M.-J., Arsanto, J.-P., Ginestier, C., Marchetto, S., Jacquemier, J., Isnardon, D., Bivic, A. L., Birnbaum, D., and Borg, J.-P. (2005) *Oncogene* **24,** 4330–4339
8. Ohno, H., Hirabayashi, S., Iizuka, T., Ohnishi, H., Fujita, T., and Hata, Y. (2002) *Oncogene* **21,** 7042–7049
9. Nakagawa, S., Yaho, T., Nakagawa, K., Takizawa, S., Suzuki, Y., Yasugi, T., Huibregtse, J. M., and Taketani, Y. (2004) *Br. J. Cancer* **90,** 194–199
10. Izawa, I., Nishizawa, M., Tomono, Y., Ohtakara, K., Takahashi, T., and Inagaki, M. (2002) *Genes Cells* **7,** 475–485
11. Laura, R. P., Witt, A. S., Held, H. A., Gerstner, R., Deshayes, K., Koehler, M. F. T., Kosik, K. S., Sidhu, S. S., and Lasky, L. A. (2002) *J. Biol. Chem.* **277,** 12906–12914
12. Jaulin-Bastard, F., Arsanto, J. P., Le Bivic, A., Navarro, C., Vely, F., Saito, H., Marchetto, S., Hatzfeld, M., Santoni, M. J., Birnbaum, D., and Borg, J. P. (2002) *J. Biol. Chem.* **277,** 2869–2875
13. Nagafuchi, A. (2001) *Curr. Opin. Cell Biol.* **13,** 600–603
14. Itoh, M., Nagafuchi, A., Yonemura, S., Kitani-Yasuda, T., and Tsukita, S. (1993) *J. Cell Biol.* **121,** 491–502
15. Itoh, M., Nagafuchi, A., Moroi, S., and Tsukita, S. (1997) *J. Cell Biol.* **138,** 181–192
16. Kausalya, P. J., Phua, D. C., and Hunziker, W. (2004) *Mol. Biol. Cell* **15,** 5503–5515
17. Yonemura, S., Itoh, M., Nagafuchi, A., and Tsukita, S. (1995) *J. Cell Sci.* **108,** 127–142
18. Rajasekaran, A. K., Hojo, M., Huima, T., and Rodriquez-Boulan, E. (1996) *J. Cell Biol.* **132,** 451–463
19. Ando-Akatsuka, Y., Yonemura, S., Itoh, M., Furuse, M., and Tsukita, S. (1999) *J. Cell. Physiol.* **179,** 115–125
20. Ikari, A., Hirai, N., Shiroma, M., Harada, H., Sakai, H., Hayashi, H., Suzuki, Y., Degawa, M., and Takagi, K. (2004) *J. Biol. Chem.* **279,** 54826–54832
21. Itoh, M., Furuse, M., Morita, K., Kubota, K., Saitou, M., and Tsukita, S. (1999) *J. Cell Biol.* **147,** 1351–1363
22. Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996) *Cell* **85,** 1067–1076
23. Daniels, D. L., Cohen, A. R., Anderson, J. M., and Brunger, A. T. (1998) *Nat. Struct. Biol.* **5,** 317–325
24. Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. (1997) *Science* **275,** 73–77
25. Im, Y. J., Lee, J. H., Park, S. H., Park, S. J., Rho, S.-H., Kang, G. B., Kim, E., and Eom, S. H. (2003) *J. Biol. Chem.* **278,** 48099–48104

26. Im, Y. J., Park, S. H., Rho, S.-H., Lee, J. H., Kang, G. B., Sheng, M., Kim, E., and Eom, S. H. (2003) *J. Biol. Chem.* **278,** 8501–8507
27. Karthikeyan, S., Leung, T., and Ladias, J. A. A. (2001) *J. Biol. Chem.* **276,** 19683–19686
28. Karthikeyan, S., Leung, T., and Ladias, J. A. A. (2002) *J. Biol. Chem.* **277,** 18973–18978
29. Skelton, N. J., Koehler, M. F. T., Zobel, K., Wong, W. L., Yeh, S., Pisabarro, M. T., Yin, J. P., Lasky, L. A., and Sidhu, S. S. (2003) *J. Biol. Chem.* **278,** 7645–7654
30. Kang, B. S., Cooper, D. R., Devedjiev, Y., Derewenda, U., and Derewenda, Z. S. (2003) *Structure* (*Camb.*) **11,** 845–853
31. Kozlov, G., Banville, D., Gehring, K., and Ekiel, I. (2002) *J. Mol. Biol.* **320,** 813–820
32. Fuh, G., Pisabarro, M. T., Li, Y., Quan, C., Lasky, L. A., and Sidhu, S. S. (2000) *J. Biol. Chem.* **275,** 21486–21491
33. Kozlov, G., Gehring, K., and Ekiel, I. (2000) *Biochemistry* **39,** 2572–2580
34. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
35. Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50,** 157–163
36. Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) *Acta Crystallogr D. Biol. Crystallogr.* **60,** 432–438
37. Birrane, G., Chung, J., and Ladias, J. A. A. (2003) *J. Biol. Chem.* **278,** 1399–1402
38. McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156–165
39. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
40. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D* **53,** 240–255
41. Karthikeyan, S., Leung, T., Birrane, G., Webster, G., and Ladias, J. A. A. (2001) *J. Mol. Biol.* **308,** 963–973
42. Lu, J., Li, H., Wang, Y., Sudhof, T. C., and Rizo, J. (2005) *J. Mol. Biol.* **352,** 455–466
43. Laskowski, R. A., MacAurthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
44. Gouet, P., Robert, X., and Courcelle, E. (2003) *Nucleic Acids Res.* **31,** 3320–3323