Unusual binding interactions in PDZ domain crystal structures help explain binding mechanisms

Jonathan M. Elkins, Carina Gileadi, Leela Shrestha, Claire Phillips, Jing Wang, João R.C. Muniz, and Declan A. Doyle*

Structural Genomics Consortium, Oxford University, Oxford, OX3 7DQ, United Kingdom

Received 2 November 2009; Revised 30 December 2009; Accepted 31 December 2009
DOI: 10.1002/pro.349
Published online 29 January 2010 proteinscience.org

Abstract: PDZ domains most commonly bind the C-terminus of their protein targets. Typically the C-terminal four residues of the protein target are considered as the binding motif, particularly the C-terminal residue (P0) and third-last residue (P-2) that form the major contacts with the PDZ domain's "binding groove". We solved crystal structures of seven human PDZ domains, including five of the seven PDLIM family members. The structures of GRASP, PDLIM2, PDLIM5, and PDLIM7 show a binding mode with only the C-terminal P0 residue bound in the binding groove. Importantly, in some cases, the P-2 residue formed interactions outside of the binding groove, providing insight into the influence of residues remote from the binding groove on selectivity. In the GRASP structure, we observed both canonical and noncanonical binding in the two molecules present in the asymmetric unit making a direct comparison of these binding modes possible. In addition, structures of the PDZ domains from PDLIM1 and PDLIM4 also presented here allow comparison with canonical binding for the PDLIM PDZ domain family. Although influenced by crystal packing arrangements, the structures nevertheless show that changes in the positions of PDZ domain side-chains and the αB helix allow noncanonical binding interactions. These interactions may be indicative of intermediate states between unbound and fully bound PDZ domain and target protein. The noncanonical "perpendicular" binding observed potentially represents the general form of a kinetic intermediate. Comparison with canonical binding suggests that the rearrangement during binding involves both the PDZ domain and its ligand.

Keywords: PDZ domain; X-ray structure; binding mode; substrate selectivity; binding mechanism

Introduction

The main mode of action of PDZ domains is binding of protein C-termini to assemble protein–protein complexes. PDZ domains show a varied selectivity for up to seven C-terminal residues.[1,2] The canonical binding mode involves the C-terminus of the protein target binding to the PDZ domain in a shallow groove, with the side-chains of the C-terminal residue (P0) and third-last residue (P-2) pointing toward the groove and being the most significant for recognition of the protein target by the PDZ domain.[3] Various classification schemes have been devised to aid in the prediction of target protein identity based on the sequence of PDZ domains. For example, in the classical "Class I" mode, where P0 is a hydrophobic residue and P-2 is a serine or threonine, the hydroxyl side-chain of P-2 binds to a histidine located as the first residue on helix αB (αB1) of the PDZ domain, and the presence of a histidine in the

Grant sponsors: Canadian Institutes for Health Research; Canadian Foundation for Innovation; Genome Canada (Ontario Genomics Institute); GlaxoSmithKline; Karolinska Institutet; Knut and Alice Wallenberg Foundation; Ontario Innovation Trust; Ontario Ministry for Research and Innovation; Merck & Co. Inc.; Novartis Research Foundation; Swedish Agency for Innovation Systems; Swedish Foundation for Strategic Research; Wellcome Trust.

*Correspondence to: Declan A. Doyle, Trinity College Dublin, School of Biochemistry and Immunology, College Green, Dublin 2, Ireland. E-mail: declanadoyle@googlemail.com

conserved αB1 position is an indicator of a preference for serine or threonine at P-2 of a target protein. The classification schemes all assume a canonical binding, however, systematic mutation of the ERBB2IP-1 PDZ domain showed that residues not immediately adjacent to the binding groove influence the selectivity.[1] Furthermore, kinetic analysis of the second PDZ domain of PTP-BL showed evidence for two stages in the binding mechanism, with initial formation of a weaker complex followed by a conformational change to a lower-energy state[4,5]; the nature of the identified intermediate was unknown.

Although most PDZ domains bind their target sequences in the canonical manner, exceptions to this pattern have been observed. The examples of PDZ domains binding internal peptide sequences are obvious exceptions,[6,7] but there are also exceptions to the canonical binding arrangement when binding protein C-termini. For instance, the recent structure of Dvl2 PDZ domain in complex with a noncanonical C-terminal sequence (WKWYGWF) revealed a conformation of the glycine residue at P(-2), which allowed the P(-3) tyrosine residue to occupy the binding position occupied by a P(-2) residue in the canonical arrangement.[8]

We have previously shown that addition of appropriate C-terminal extensions to constructs of PDZ domains greatly assists in the formation of PDZ domain crystals, with each PDZ domain binding the C-terminus of a crystallographically adjacent molecule.[9] The C-terminal extensions can be chosen either based on the sequence of a known binding target of the PDZ domain or based on a likely target sequence according to the class of the PDZ domain. This produces crystal structures of PDZ domains self-bound to their own C-termini and allows analysis of the binding for that PDZ domain—C-terminus interaction. Here we have applied this technique to the GRASP (Tamalin) PDZ domain, the PDLIM PDZ domains, and the MAST4 PDZ domain.

The primary function of GRASP is to link receptors such as mGluR1 and mGluR5 to neuronal proteins.[10] The structure of GRASP PDZ domain from *Rattus norvegicus* was previously determined by Sugi *et al.* in three different forms: The PDZ domain alone, with C-terminal extensions representing the C-termini of GRASP itself, and of mGluR5.[11] Interestingly, in their structure with the C-terminus of GRASP ("auto-inhibited" structure) two of the four molecules in the asymmetric unit displayed an unusual "perpendicular," noncanonical, mode of binding in which only the C-terminal P0 residue was bound in the binding groove. This binding arrangement resembled that observed by NMR for the auto-inhibited X11/Mint PDZ domain protein.[12] We determined a crystal structure of human GRASP (97% identical over the PDZ domain) simultaneously with the publication by Sugi *et al.* Our structure, which also shows noncanonical binding, is presented here.

There are seven human PDLIM proteins possessing an N-terminal PDZ domain and at least one C-terminal LIM domain. PDLIM1 (also known as CLP-36), PDLIM2 (Mystique), PDLIM3 (ALP), and PDLIM4 (RIL) each have a single C-terminal LIM domain, whereas PDLIM5 (Enigma homologue), PDLIM6 (Cypher), and PDLIM7 (Enigma) each have three C-terminal LIM domains. PDLIM1, 2, 3, 4, and 6 all bind α-actinin via their PDZ domain and are involved with recruitment of LIM-binding proteins to the cytoskeleton.[13–18] PDLIM7 is known to bind β-tropomyosin via its PDZ domain and, therefore, also acts to direct LIM-binding proteins to the cytoskeleton.[19] A similar association has not yet been observed for PDLIM5, which is, however, of interest due to its potential as a susceptibility gene for schizophrenia.[20]

The NMR structures of the apo PDZ domains from human PDLIM4 [RIKEN structural genomics initiative (RSGI)] and PDLIM6[21] have been solved previously, as well as the NMR structures of mouse PDLIM3 (94% sequence identity to human PDLIM3 over the PDZ domain) and PDLIM6 (100% identical) also from RSGI (PDB IDs 1EEG, 1RGW, 1V5L, 1WJL, respectively). This work completes the structural coverage of this family, presenting the X-ray crystal structures of the PDZ domains from human PDLIM1, 2, 4, 5, and 7, three of which display noncanonical binding. Finally, we present the structure of the MAST4 PDZ domain, which shows binding flexibility relevant to a discussion of noncanonical binding, and discuss the implications of noncanonical binding for a two-state kinetic binding mechanism.

Results

Construct design, protein purification, and crystallization

Multiple constructs were designed for each of the target PDZ domains according to the methodology we described previously, in which each construct had a different number of residues between the PDZ domain and the self-binding C-terminal extension, for multiple attempts at obtaining crystals.[9] Protein was over-expressed in *Escherichia coli* from these constructs, and purified using standard methods.

For each target PDZ domain, the constructs represented a range of different extension lengths and different C-terminal extensions, and protein from more than one construct of each target was put into crystallization, but not all of the constructs generated crystals. Crystals of one construct were followed through until structure determination, as detailed in Table I. In the case of PDLIM7, the

Table I. *Constructs Used for Protein Purification and Structure Determination*

	Constructs used in crystallization	Constructs giving crystals	Residue range[a]	C-terminal extension[a]	No. of residues between domain and extension[a,b]
PDLIM1	7	2	Met1–His86	ESDL	4
PDLIM2	7	4	Met1–Arg82	ITSL	1
PDLIM4	4	1	Met1–Gly85	ESDL	4
PDLIM5	2	2	Met1–Arg83	ESDL	1
PDLIM7	5	1	Met1–Ala84	ITSL	2
GRASP	3	1	Gln97–Tyr188	SSTL	2
MAST4	2	1	Gln1216–Thr1307	ETSV	5

[a] For the construct used for the structure.
[b] The number of residues between the end of the final β-strand (equivalent to Ala82 in PDLIM1) and the four residues of the C-terminal PDZ-binding motif.

construct contained the same C-terminus (ITSL) as β-tropomysin, its known binding partner.[19]

All of the structures were determined by molecular replacement starting with models of previously determined PDZ domains, followed by alteration to the correct sequence and addition of coordinates for missing atoms including the C-terminal extensions and solvent molecules where appropriate.

Description of structures

The structures revealed each PDZ domain in the standard conformation of five beta-strands and two alpha-helices. In the structures of the GRASP and MAST4 PDZ domains, each asymmetric unit contained two molecules, with root mean square (r.m.s.) deviations between the two molecules of 0.67 Å for GRASP and 0.84 Å for MAST4, over 80 and 84 Cα atoms, respectively. The structures of PDLIM2, PDLIM4, PDLIM5, and PDLIM7 had r.m.s. deviations to PDLIM1 of between 0.5 and 0.8 Å (the sequence identities to PDLIM1 over the PDZ domain are between 43% and 71%). These structures collectively show two kinds of binding: Either the canonical binding for a class I PDZ domain or in the case of PDLIM2, PDLIM5, PDLIM7, and one chain of GRASP, a "perpendicular" mode of binding. These two modes are illustrated by a comparison of PDLIM1 and PDLIM2 in Figure 1. The individual structures are discussed later.

GRASP PDZ domain structure

In the structure of GRASP PDZ domain, chain A binds the C-terminal residues of a crystallographic symmetry-related molecule in the canonical manner for a "Class I" PDZ domain, as seen in our previous structures generated by this method[9] and other structures deposited in the PDB. The P0 leucine is bound with its side-chain in the hydrophobic pocket at the end of the binding groove, and the expected interaction for a Class I PDZ domain between the P-2 serine and the histidine at αB1 (first residue of the αB helix) is present. In addition, there are hydrogen bonds between the P-3 serine and the side-chains of Gln117 and Arg137. However, chain B binds the C-terminal residues in a very different conformation, such that only the P0 residue approximates its position as bound to chain A; an overlay of the two chains shows the substantial difference in

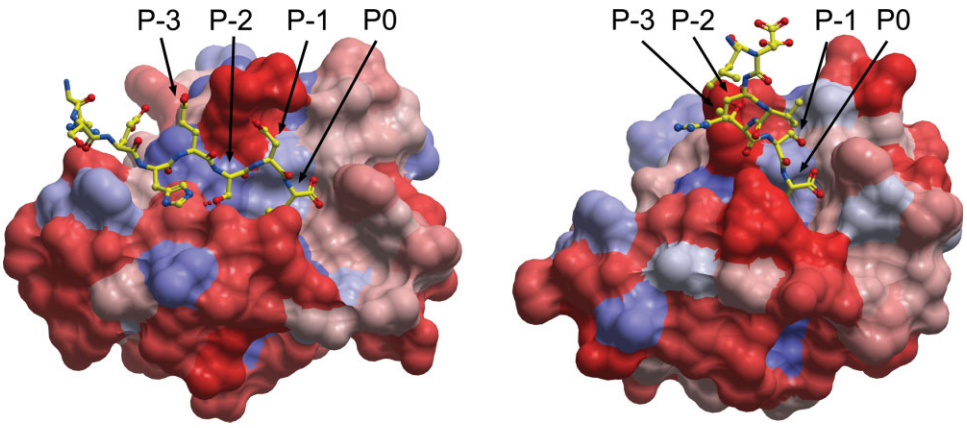

Figure 1. Comparison of binding for PDLIM1 (on the left) and PDLIM2 (on the right). The surface is colored by hydrophobicity, with red, the least, and blue, the most hydrophobic. The ligand C-terminus is shown as a yellow ball-and-stick model with the residues from the C-terminus to the N-terminus labeled as P0 (C-terminal residue), P-1, P-2, and P-3. PDLIM1 has a canonically bound C-terminus, whereas PDLIM2 binds a C-terminus in a "perpendicular" manner.

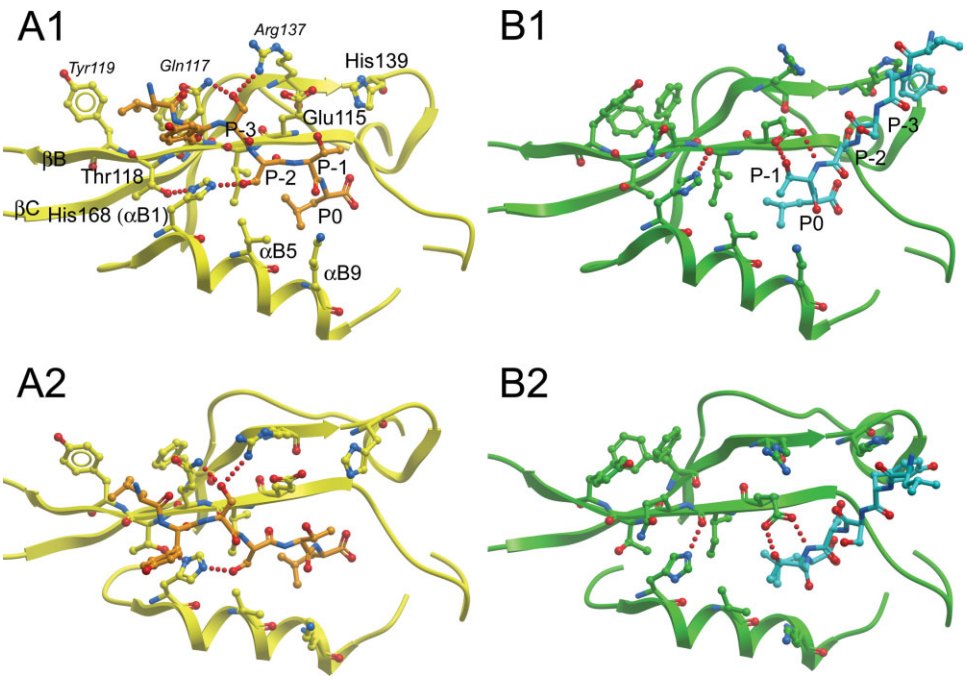

Figure 2. Comparison of binding for the two independent chains of the GRASP PDZ domain structure. A1, A2: View from two angles of chain A, which shows canonical binding. B1, B2: View from the same angles, but of chain B that shows different binding. The side-chain movements associated with the differences in binding can be seen by comparing (A) and (B), especially the rotation of Thr118 and His168 with changes in hydrogen bonding, and the movement of the other labeled side-chains. Labels in smaller italic font indicate residues affected by crystal packing.

direction and binding of the bound C-terminal residues, and many differences in the positions of side-chains (Fig. 2).

Of the seven side-chains on the outside face of the β-strands βB and βC, six are in different conformations (Fig. 2). The only side-chain whose position is maintained is Cys136; this residue is alanine in rat GRASP, the most significant of the sequence differences due to its position close to the binding groove. Of the six side-chains in different conformations, Glu115, from βB, and His139 from βC are not influenced by the crystal packing. The conformations of Gln117 and Tyr119 from βB, and Phe134 and Arg137 from βC are influenced by a combination of crystal packing and the binding arrangement. Phe134 is in two conformations in chain B, but one ordered conformation in chain A.

Considering the bound C-terminus, in the binding arrangement of chain B the P-2 serine is not located in the binding groove and it has hydrogen bonds only to two water molecules, while the P-1 Thr is bound by the side-chain of Glu115. Other interactions in chain B, such as the stacking arrangement of the P-4 Tyr with His139 are further from the binding groove and more likely to be due to the crystal packing.

When comparing the positions of the P0 leucine residue in the two different binding arrangements, the leucine side-chain is rotated 180° around the Cβ—Cγ bond, with the overall rotation of the whole P0 residue meaning that the side-chain still occupies almost the same volume of space. Thus, the only constant in the binding of the two chains in this structure is the position of the C-terminal carboxylate, which forms the anticipated interactions with the backbone nitrogens of the GFGF motif (TFGF in GRASP).

Another interesting difference between the two binding arrangements concerns the histidine at αB1 (His168), the key residue in selectivity for a Thr/Ser at P-2. In chain A, this histidine is in the expected conformation, forming a hydrogen bond from its side-chain NE2 atom to the P-2 serine, as mentioned earlier. However, in chain B, in the absence of the P-2 residue in the binding groove, the side-chain of His168 is rotated 180° around Cβ—Cγ and this movement together with a small rotation of strand βB allows a hydrogen bond to be formed between NE2 and the backbone carbonyl of Ile116 on βB. This movement is illustrated in Figure 2, which also shows how the side-chain of Thr118 is rotated to match the rotation of His168.

Comparisons with previous GRASP PDZ domain structures

The three structures of GRASP PDZ domain published by Sugi *et al.*[11] all show crystal packing interactions through βB and βC to varying extents, which as in our structure makes analysis of the side-chains difficult. Chain A in our structure has a different

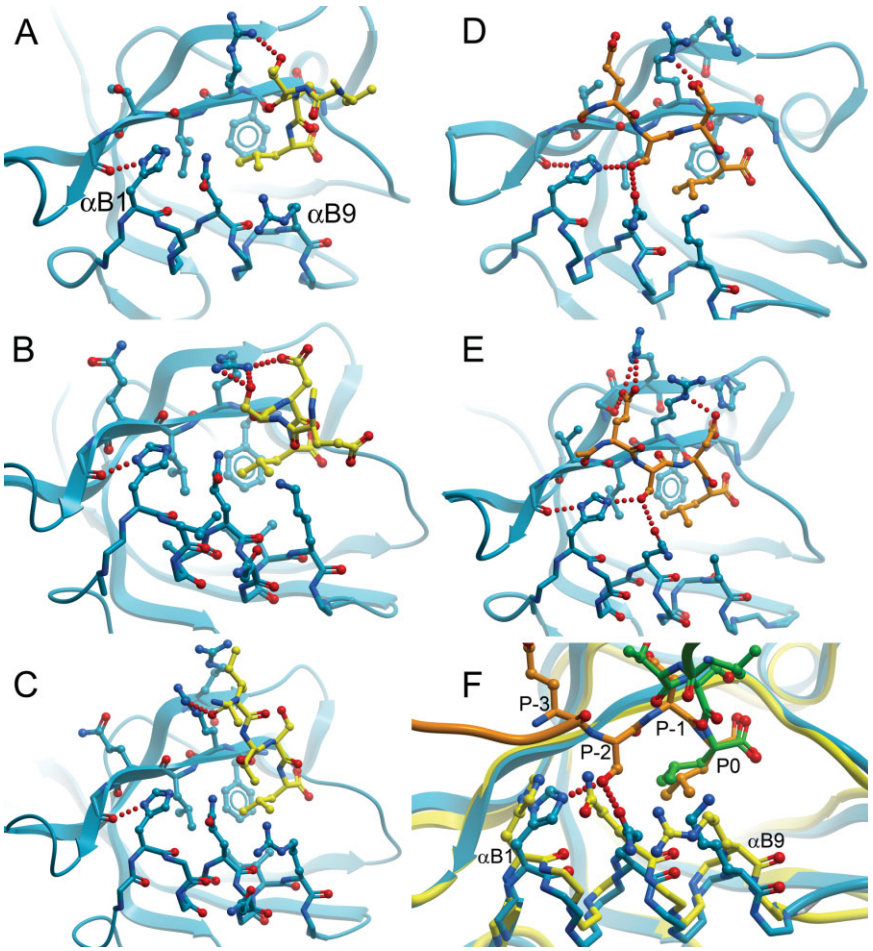

Figure 3. Comparison of the positions of the P0 residues, and of the αB helix, in the binding arrangements of PDLIM PDZ domains. (A) PDLIM2, (B) PDLIM5, (C) PDLIM7, (D) PDLIM1, (E) PDLIM7, (F) Overlay of PDLIM1 and PDLIM2: PDLIM1 is colored blue with its bound C-terminus colored orange, whereas PDLIM2 is colored yellow with its bound C-terminus colored green. Hydrogen bonds are shown for the P-2 threonine of PDLIM1. The lower-occupied alternate confirmation of the PDLIM2 αB9 arginine has been removed for clarity. The rotation of the P0 residue can be seen, as can the movement of the αB helix at the αB9 end.

crystal packing interface, with βB and βC of one molecule interacting with βB and βD from the other face of a second molecule. However, in Chain B the local crystal packing environment on the βB, βC face is very similar to that in the "auto-inhibited" structure of GRASP with the GRASP C-terminus, in which a similar "perpendicular" binding is also seen, although with a different construct, and in a different crystal form.

The variation in position and hydrogen bonding of His168 observed in our structure in the absence of a canonically bound C-terminus is also observed in the structures published by Sugi *et al.* In their auto-inhibited structure, the side-chain of His168 is also rotated 180° around Cβ—Cγ, although it does not form the hydrogen bond with the backbone of βB, instead interacting with a phosphate ion in the crystal lattice. Of particular relevance, however, in their native structure chain A has the rotated His168 side-chain, whereas chain B does not. Superimposition of the two chains shows that as the backbones align very well overall (r.m.s. deviation of 0.60 Å over 87 Cα atoms) with little deviation except at the N- and C-termini, there is a substantial difference in position of the βD-αB loop, including a 1.6 Å difference in position of the backbone at His168 (αB1).

PDLIM PDZ domain structures

For PDLIM1 and PDLIM4, the residues of the C-terminal self-binding extension (ESDL) were bound to the binding groove of a crystallographic symmetry-related molecule in the canonical manner. The side-chain of the P0 residue was buried in the hydrophobic pocket at the end of the binding groove and the side-chain of the P-2 serine/threonine was bound to the side-chain of the histidine residue at αB1 (Fig. 3). However, in the structures of PDLIM2, PDLIM5, and PDLIM7 the C-terminal residues (ESDL/ITSL, Table I) were bound such that only the C-terminal P0 residue was in the canonical position and conformation (Fig. 3), similar to the general

arrangement for the binding to chain B of GRASP. The P-1, P-2, and P-3 residues were not bound in the PDZ domain's binding groove but, as in the binding to chain B of GRASP, did have additional stabilizing interactions with PDZ domain surface side-chains. For the PDLIM PDZ domain structures where there is more than one molecule in the asymmetric unit (PDLIM4, PDLIM5, and PDLIM7), each of the molecules in the asymmetric unit shows a similar manner of binding.

Therefore, as in the two chains of the GRASP structure, binding of the P0 residue is relatively unaffected by the different binding arrangements seen in these crystals. The most extreme positions of the P0 residue are taken by PDLIM2 (in the alternate binding mode) and PDLIM1 (in the normal binding mode). However, even in these the alteration in position of the P0 residue is relatively small, consisting of a modest rotation (Fig. 3). Unlike in the GRASP structure, there is not a rotation of the P0 leucine side-chain about the Cβ—Cγ bond.

Differences between the canonical binding mode and the perpendicular modes

It is harder to compare the positional differences of the surface side-chains on βB and βC than for the comparison between the two chains of the GRASP PDZ domain structure, due to the sequence differences between the PDLIM proteins and the differing crystal packing environments. However, two conserved residues are of interest: The conserved arginine residue at βB2 [second residue of strand βB, immediately after the GYGF motif (PWGF in PDLIMs), Arg16 in PDLIM2 and PDLIM4, Arg17 in PDLIM1, PDLIM5, and PDLIM7], and the αB5 glutamine.

The arginine at βB2 binds the P-1 aspartate in both of the canonical binding structures (PDLIM1 and PDLIM4). In the other structures, it is less well-ordered in general, but nevertheless can be seen to stabilize the bound C-terminus in different ways in each of the other structures; in PDLIM2, it binds the P-2 threonine; in PDLIM5, it binds the P-2 threonine in chain A and the P-1 aspartate in the other four chains, and in PDLIM7, it binds the backbone carbonyl of the P-3 residue in chain A and the backbone carbonyl of the P-2 residue in chain B. The conformations of Arg17 in PDLIM5 are affected by the opportunity for a guanidine-stacking arrangement with Arg45 from the same molecule as the bound C-terminus; there is not a crystal packing influence on this residue in the PDLIM2 or PDLIM7 structures, excepting of course the interactions with the bound C-terminus itself.

In PDLIM1 and PDLIM4, the P-2 serine side-chain also forms a hydrogen bond to the side-chain of the αB5 glutamine residue. This residue is conserved in all of the PDLIM PDZ domains, and in the structures of PDLIM2, PDLIM5, and PDLIM7 is in an alternate conformation closer to the binding groove (Fig. 3). Therefore, this appears to be the favored conformation for this residue in the absence of a protein bound in the binding groove. In PDLIM4, as well as the interaction between the P-1 aspartate and Arg16, there are also binding interactions between the P-1 aspartate and His33 (at the end of βC), and between the P-3 glutamate and Arg16, Ser30, and Arg31.

Difference in position of Helix αB in PDLIM2

In the two structures representing extremes of position of the P0 residue, a comparative difference in position of the αB helix can be seen [Fig. 3(F)]. The movement of the helix relative to that in PDLIM1 results in a wider binding groove for PDLIM1 where the P-2 residue of the bound C-terminus resides in the binding groove in the canonical manner, and a narrower groove for the altered binding of PDLIM2. The movement is greatest at the C-terminal end of the helix, with differences of 1.4 Å in the positions of the Cα atom of αB9 (arginine in both structures), 1.1 Å for the Cα atom of the αB5 glutamines, and 0.7 Å for the Cα atom of the αB1 histidines. This difference in position of αB is specific to PDLIM2 and was not seen in PDLIM5 or PDLIM7.

MAST4 PDZ domain structure

The construct used for the MAST4 PDZ domain structure had five residues between the end of the PDZ domain and the C-terminal self-binding motif ETSV (Table I). There were two molecules in the asymmetric unit. Chain B binds the C-terminus in the canonical manner. In addition to the binding of the P0 valine and the binding of the P-2 threonine to His70 at αB1, there is a salt bridge between the P-3 glutamate and Arg19 at βB4 (residue numbering as in the PDB file, Arg19 corresponds to Arg1130 in the full length protein). Chain A binds the C-terminus in a similar manner overall, including the same salt bridge between P-3 glutamate and Arg19, but with some key differences: The P-2 threonine as bound to chain A is rotated out of the binding groove, and water molecules fill the gap between the P-2 threonine and the αB1 histidine, with hydrogen bonds to the histidine NE2 and P-2 backbone nitrogen (Fig. 4). In chain A, the side-chain of the P-1 Ser points away from the β-strand after the NYGF motif (βB) instead of toward it as in chain B, a rotation around its Cα—Cβ bond. Matching this rotation, Thr17 at βB2 is also rotated around its Cα—Cβ bond.

There is not a substantial difference in the position of helix βB between the two chains, unlike the deviation observed for PDLIM2 compared with PDLIM1, or that observed for the third PDZ domain of MPDZ.[9] Also, there is not an obvious crystal

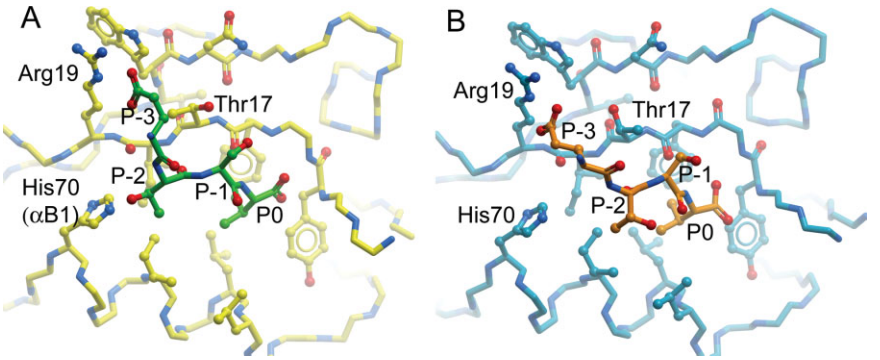

Figure 4. Comparison of binding for the two independent chains of the MAST4 PDZ domain structure. (A) Chain B with canonical binding; (B) Chain A where the P-2 threonine is rotated out of the binding groove, whereas the interaction of the P-3 glutamate with Arg19 is maintained. Note also the rotation of Thr17 that matches the rotation of the P-1 serine.

packing explanation for the difference in binding between chains A and B, with the overall geometry of the two molecules involved in each binding interaction being similar. The conservation of the interactions with the P-3 glutamate suggests that those interactions are more significant than the more common P-2 threonine interaction in this case.

Discussion

Generality of a perpendicular binding mode

It is now becoming clear that the extent of interactions of small protein domains is often wider than in previously recognized models. For example, SH2 domains have been shown to have important interactions via a secondary binding site.[22] C2 domains can bind phosphotyrosine,[23] and a bromodomain was recently shown to bind two acetyl-lysine side-chains simultaneously.[24] In the case of PDZ domains, the observations of binding to phosphoinositides have opened new areas of research.[25,26]

The possibility remains that the observations of the "perpendicular" binding mode are purely induced by crystal packing considerations; however, the number of observations of a perpendicular binding mode, and in particular the observation of this mode by NMR in the X11/Mint PDZ domain protein structure,[12] suggest a more general role for this mode of binding in the interactions of PDZ domains with their ligands. There are even possibilities for it being a necessary binding mode when self-binding is used as a mode of regulation of PDZ domain interactions, if the particular PDZ domain protein does not have the conformational flexibility for canonical binding.[11] The PDLIM and GRASP PDZ domain structures showed a correspondence between the length of the C-terminal self-binding extension and whether in the crystal the C-terminus was bound in the canonical manner or not (Table I). The shorter C-termini may not have had the necessary flexibility to form a canonical binding arrangement, as also allowing the observed crystal packing. These observations together with the conservation of the binding of the P0 residue, despite any possible crystal packing determinants, show the dominance of the P0 residue in recognition and, therefore, selectivity. Interestingly, the MAST4 PDZ domain structure exhibited different binding in each of the molecules in the asymmetric unit, whereas having in principle the flexibility to form a canonical interaction in both.

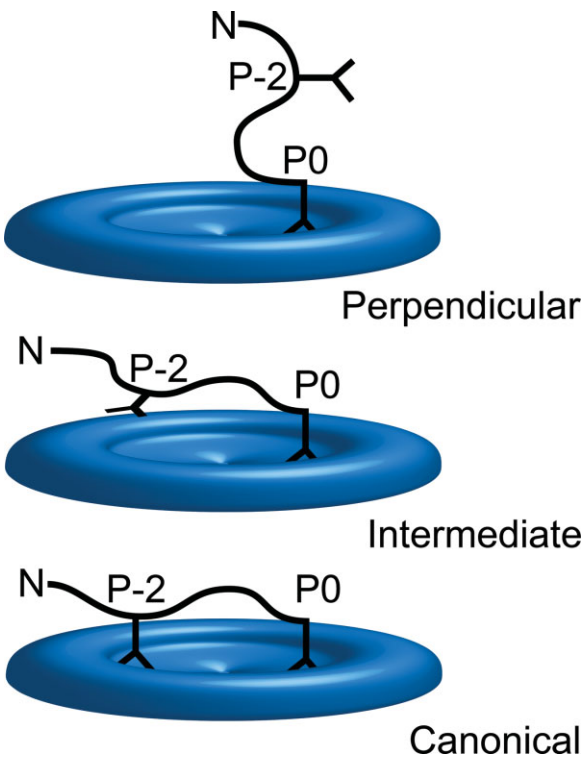

Figure 5. Schematic comparison of two potential alternate binding modes compared with canonical binding.
Each binding mode shows a protein C-terminus bound to the binding groove of a PDZ domain with its P0 residue. In the "perpendicular" binding mode only the P0 residue is bound in the binding groove, whereas an intermediate state has the P-2 residue close to the binding groove, but not bound in a canonical manner.

Table II. *Crystallization Conditions*

	Protein buffer	[Protein] (mg/mL)	Crystallization conditions	Ratio of protein: precipitant in crystallization drop	Cryo-protectant
PDLIM1	50 m*M* Hepes pH 7.4, 500 m*M* NaCl, 0.5 m*M* TCEP	63	40% PEG 300, 0.2*M* calcium acetate, 0.1*M* cacodylate pH 6.5	1:2	None
PDLIM2	50 m*M* Hepes pH 7.4, 500 m*M* NaCl, 0.5 m*M* TCEP	12	30% mPEG 2000, 0.2*M* $(NH_4)_2SO_4$, 0.1*M* acetate pH 4.6	1:2	30% Glycerol
PDLIM4	50 m*M* Hepes pH 7.4, 500 m*M* NaCl, 0.5 m*M* TCEP	6	20% PEG 10000, 0.2*M* $MgCl_2$, 0.1*M* Tris pH 8.5	2:1	15% PEG400
PDLIM5	25 m*M* Hepes pH 7.4, 150 m*M* NaCl, 0.5 m*M* TCEP	13	30% PEG 4000, 0.2*M* $MgCl_2$, 0.1*M* Tris pH 8.5	2:1	20% Ethylene glycol
PDLIM7	50 m*M* Hepes pH 7.4, 500 m*M* NaCl, 0.5 m*M* TCEP	38	20% PEG 4000, 0.1*M* Hepes pH 7.5, 10% isopropanol	1:2	20% Ethylene glycol
GRASP	50 m*M* Hepes pH 7.4, 500 m*M* NaCl, 0.5 m*M* TCEP	50	1.26*M* NaH_2PO_4, 0.14*M* K_2HPO_4	1:1	30% Glycerol
MAST4	25 m*M* Hepes pH 7.5, 150 m*M* NaCl, 0.5 m*M* TCEP	17	0.75*M* NaH_2PO_4, 0.91*M* K_2HPO_4	2:1	25% Ethylene glycol

Binding mechanism

A study on the binding kinetics of PTP-BL PDZ domain and its target peptide showed evidence for a two-step induced-fit mechanism.[4,5] This implied a mechanism of initial binding followed by rearrangement. The kinetics alone identified the presence of a rearrangement but not its cause, in particular whether from a rearrangement of the PDZ domain, the ligand, or both. There is some evidence for rearrangement within the PDZ domain hydrophobic core.[4] Our structures suggest a possible mechanism for induced fit rearrangement of both PDZ domain and ligand, after initial binding of ligand at position P0.

According to such a mechanism, the structures that show only the P0 residue of the ligand bound in the binding groove may, therefore, be representative of one of a number of possible intermediate bound states, before a transition to the final bound state in which the canonical binding would be observed. This transition involves any or all of: Surface side-chain rearrangements, in particular on αB, βB, and βC; rearrangement within the hydrophobic core; and movement of helix αB, as seen in the comparison of PDLIM1 and 2, and also in PTP-BL.[4] All of these transitions are coordinated with the rearrangement of the bound C-terminus. The structure of the MAST4 PDZ domain may represent a transition between perpendicular and canonical binding. The three stages of this mechanism are illustrated schematically in Figure 5 as the "perpendicular", "intermediate", and "canonical" bound states.

Residues may directly influence either the stabilization of the final bound state, through induced-fit cascading side-chain rearrangements, or stabilization of the intermediate bound state, through interactions with the partly bound C-terminus. An individual residue could be involved in either or both of these mechanisms. For example, in GRASP PDZ domain Glu115 on βB seems to be involved in both mechanisms, binding Arg137 (βC) that binds the P-3 Ser in the final bound state, whereas binding the P-1 Thr directly in the intermediate state; His139 (βC) stabilizes the intermediate state through π-stacking with the P-4 Tyr side-chain, but does not directly interact with the bound C-terminus in the final bound state; and Gln117, Tyr119 (βB), and Phe134 (βC) are only involved in stabilizing the final state, which involves movement of all of their side-chains from their positions in the intermediate state.

When considering the influence of residues remote from the binding groove, only residues on the "top" surface of αB, βB, or βC are likely to have direct contact with the bound C-terminus (in any state) and, therefore, only these residues would be expected to have a direct effect on the selectivity. However, other residues may have an indirect effect, even acting through the hydrophobic core of the PDZ domain. For example, the residue equivalent to Ile116 on βB of GRASP, at the centre of the binding groove, is often seen in multiple conformations in crystal structures, and indirect packing effects in the core may alter the range of allowed conformations, affecting selectivity.

It is difficult to assign importance to any individual residue in binding mechanisms due to the possibilities of crystal packing influences. However, the overall impression from all of the observations put together is one of concerted adjustment of surface side-chains, secondary structural elements

Table III. *Data Collection and Refinement Statistics*

	PDLIM1	PDLIM2	PDLIM4	PDLIM5	PDLIM7	GRASP	MAST4
Crystal form							
Unit Cell [a, b, c (Å), β/γ (°)]	38.7, 38.7, 246.5	58.8, 58.8, 52.3	87.0, 87.0, 53.9	61.3, 36.5, 88.7, 99.2	46.1, 55.4, 57.3	72.3, 72.3, 163.3, 120.0	100.9, 100.9, 69.5, 120.0
Space group	$P6_522$	$P6_1$	$P42_12$	$P2_1$	$P2_12_12_1$	$P6_522$	$H3$
Number of molecules/asu	1	1	2	5	2	2	2
Data Collection							
Resolution range[a] (Å)	14.02–1.50 (1.58–1.50)	50.00–1.70 (1.80–1.70)	61.55–1.90 (2.00–1.90)	46.40–1.50 (1.60–1.50)	39.84–1.11 (1.15–1.11)	50.00–2.15 (2.23–2.15)	29.14–1.64 (1.73–1.64)
Completeness[a] (%)	97.0 (82.3)	99.7 (98.5)	100.0 (100.0)	79.7 (37.0)	99.4 (100.0)	100.0 (99.9)	99.7 (100.0)
Multiplicity[a]	11.2 (2.5)	6.2 (3.4)	9.3 (9.4)	2.7 (0.8)	5.1 (4.6)	8.8 (6.5)	2.7 (2.7)
R_{merge}[a] (%)	13.5 (40.7)	5.6 (53.2)	10.9 (69.1)	4.0 (38.0)	4.2 (14.4)	6.1 (92.9)	5.0 (52.4)
$\langle I/\sigma(I)\rangle$[a]	15.4 (2.0)	18.1 (1.9)	18.4 (4.2)	17.2 (2.0)	36.6 (11.0)	17.6 (1.4)	7.8 (1.8)
Refinement							
R factor (%)	21.2	19.2	17.3	16.3	13.1	20.9	19.6
R_{free} (%)	23.1	25.3	21.0	21.5	15.7	23.2	22.4
Rmsd bond length (Å) [angle (°)]	0.005 (0.9)	0.009 (1.2)	0.011 (1.4)	0.014 (1.5)	0.014 (1.7)	0.016 (1.5)	0.006 (0.9)
PDB ID	2PKT	2PA1	2V1W	2UZC	2Q3G	2PNT	2W7R

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

(helix αB in PDLIM2, βB in GRASP) and movement of the ligand C-terminus from a partially bound perpendicular state to a canonically (fully) bound state.

In typical analyses of the specificity determinants of C-terminal residues such as that carried out by Tonikian *et al.*,[1] the average specificity determinants of the P-1 and P-3 residues are higher relative to those of P-2 and P0 than one might perhaps expect, considering that P-3 and P-1 do not bind in the "binding groove". However, considering a two-state binding mechanism, the stabilization of an intermediate state would be a contributing factor in the specificity determination.

Materials and Methods

Cloning, expression, and protein purification

DNA for each of the proteins was amplified by PCR from template DNA in the Mammalian Gene Collection. The IMAGE Consortium Clone IDs for PDLIM1, PDLIM2, PDLIM4, PDLIM5, and PDLIM7 are 2985229, 4593207, 4132234, 3345715, and 3506748, respectively. The four amino acid C-terminal extensions were created by inclusion of 12 extra bases in the reverse primer. The PCR products were incorporated into a home-made vector containing an N-terminal hexahistidine tag and TEV protease tag cleavage site (sequence MHHHHHHSSGVDLGTENLYFQSM), by ligation-independent cloning (Table I). The resulting plasmids were transformed into *E. coli* BL21 (DE3) cells containing the pRARE2 plasmid from commercial Rosetta II (DE3) cells. Cultures were grown in shaker flasks containing 1 L of terrific broth medium with the number of flasks per clone being dependent on the level of protein expression (previously determined). The cultures were grown at 37°C until an OD_{600} of 1–3 was reached. The temperature was then reduced to between 20°C and 22°C before induction of protein expression with 0.5–1 m*M* isopropyl-β-D-thiogalactopyranoside. Cells were grown overnight before harvesting by centrifugation. The cells were resuspended in lysis buffer (typically 50 m*M* Hepes pH 7.5, 500 m*M* NaCl, 5 % glycerol, 5 m*M* imidazole, 0.5 m*M* TCEP, and protease inhibitors) and lysed by high pressure homogenization. Polyethyleneimine was added to a concentration of 0.15%, and the insoluble debris removed by centrifugation.

The target protein was purified from the clarified cell extract by immobilized metal ion chromatography: The cell extract was passed over Ni^{2+} resin, and the resin was then washed with lysis buffer containing 25 m*M* imidazole. Protein was eluted with lysis buffer containing 250 m*M* imidazole and the eluted fractions were further purified on an S200 gel filtration column. Removal of the hexahistidine tag was accomplished using TEV protease at 4°C overnight followed by passing the solution over Ni^{2+}

resin. Purified proteins were concentrated to the concentrations shown in Table II and stored at −80°C before crystallization. The final protein buffer solutions are shown in Table II.

Crystallization, data collection, and structure solution

Crystals were grown at 20°C using the sitting drop vapor diffusion method. Protein solution was mixed with crystallization buffer and equilibrated against a reservoir of crystallization buffer. Crystallization conditions and mixing ratios of protein to crystallization buffer are given in Table II. Crystals were cryo-cooled by plunging into liquid nitrogen and X-ray data were collected at 100 K using a nitrogen stream. Where appropriate, the crystals were transferred first into a cryo-protectant solution consisting of crystallization condition plus either PEG400, ethylene glycol, or glycerol (Table II) before being cryo-cooled. All X-ray data was collected on beamline X10A of the Swiss Light Source (see Table III for data collection and refinement statistics).

The data was processed with MOSFLM,[27] HKL2000[28] or XDS,[29] and the CCP4 suite.[30] Structures were solved by molecular replacement using PHASER[31] and homologous models obtained from the PDB. Crystallographic models were rebuilt using O[32] or COOT[33] and refinement was performed using REFMAC5[34] or PHENIX.[35]

Acknowledgment

We thank the members of the SGC biotechnology group for assistance with cloning. We thank the members of the SGC protein crystallography group and the Swiss Light Source for assistance with X-ray data collection.

References

1. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh J-H, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS (2008) A specificity map for the PDZ domain family. PLOS Biol 6:2043–2059.
2. Zhang Y, Yeh S, Appleton BA, Held HA, Kausalya PJ, Phua DCY, Wong WL, Lasky LA, Wiesmann C, Hunziker W, Sidhu SS (2006) Convergent and divergent ligand specificity among PDZ domains of the LAP and zonula occludens (ZO) families. J Biol Chem 281: 22299–22311.
3. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 85: 1067–1076.
4. Gianni S, Walma T, Arcovito A, Calosci N, Bellelli A, Engström Å, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW (2006) Demonstration of long range interactions in a PDZ domain by NMR, kinetics, and protein engineering. Structure 14:1801–1809.
5. Jemth P, Gianni S (2007) PDZ domains: folding and binding. Biochemistry 46:8701–8708.
6. Hillier BJ, Christopherson KS, Prehoda KE, Bredt DS, Lim WA (1999) Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science 284:812–815.
7. Penkert RR, DiVittorio HM, Prehoda KE (2004) Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. Nat Struct Mol Biol 11: 1122–1127.
8. Zhang Y, Appleton BA, Wiesmann C, Lau T, Costa M, Hannoush RN, Sidhu SS (2010) Inhibition of Wnt signaling by Dishevelled PDZ peptides. Nat Chem Biol 5: 217–219.
9. Elkins JM, Papagrigoriou E, Berridge G, Yang X, Phillips C, Gileadi C, Savitsky P, Doyle DA (2007) Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions. Protein Sci 16:683–694.
10. Kitano J, Kimura K, Yamazaki Y, Soda T, Shigemoto R, Nakajima Y, Nakanishi S (2002) Tamalin, a PDZ domain-containing protein, links a protein complex formation of group 1 metabotropic glutamate receptors and the guanine nucleotide exchange factor cytohesins. J Neurosci 22:1280–1289.
11. Sugi T, Oyama T, Muto T, Nakanishi S, Morikawa K, Jingami H (2007) Crystal structures of autoinhibitory PDZ domain of Tamalin: implications for metabotropic glutamate receptor trafficking regulation. EMBO J 26: 2192–2205.
12. Long J-F, Feng W, Wang R, Chan L-N, Ip FCF, Xia J, Ip NY, Zhang M (2005) Autoinhibition of X11/Mint scaffold proteins revealed by the closed conformation of the PDZ tandem. Nat Struct Mol Biol 12:722–728.
13. Vallenius T, Luukko K, Mäkelä TP (2000) CLP-36 PDZ-LIM protein associates with nonmuscle α-actinin-1 and α-actinin-4. J Biol Chem 275:11100–11105.
14. Torrado M, Senatorov VV, Trivedi R, Fariss RN, Tomarev SI (2004) Pdlim2, a novel PDZ-LIM domain protein, interacts with α-actinins and filamin A. Invest Ophthamol Vis Sci 45:3955–3963.
15. Xia H, Winokur ST, Kuo W-L, Altherr MR, Bredt DS (1997) Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs. J Cell Biol 139:507–515.
16. Schulz TW, Nakagawa T, Licznerski P, Pawlak V, Kolleker A, Rozov A, Kim J, Dittgen T, Köhr G, Sheng M, Seeburg PH, Osten P (2004) Actin/α-actinin-dependent transport of AMPA receptors in dendritic spines: role of the PDZ-LIM protein RIL. J Neurosci 24: 8584–8594.
17. Vallenius T, Scharm B, Vesikansa A, Luukko K, Schäfer R, Mäkelä TP (2004) The PDZ-LIM protein RIL modulates actin stress fiber turnover and enhances the association of α-actinin with F-actin. Exp Cell Res 293:117–128.
18. Zhou Q, Ruiz-Lozano P, Martone ME, Chen J (1999) Cypher, a striated muscle-restricted PDZ and LIM domain-containing protein, binds to α-actinin-2 and protein kinase C. J Biol Chem 274:19807–19813.
19. Guy PM, Kenny DA, Gill GN (1999) The PDZ domain of the LIM protein enigma binds to β-tropomyosin. Mol Biol Cell 10:1973–1984.
20. Li C, Tao R, Qin W, Zheng Y, He G, Shi Y, Li X, Guo Z, Chen H, Feng G, He L (2008) Positive association between *PDLIM5* and schizophrenia in the Chinese Han population. Int J Neuropsychopharmacol 11: 27–34.
21. Au Y, Atkinson RA, Guerrini R, Kelly G, Joseph C, Martin SR, Muskett FW, Pallavicini A, Faulkner G, Pastore A (2004) Solution structure of ZASP PDZ

domain: implications for sarcomere ultrastructure and enigma family redundancy. Structure 12:611–622.
22. Bae JH, Lew ED, Yuzawa S, Tomé F, Lax I, Schlessinger J (2009) The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site. Cell 138:514–524.
23. Benes CH, Wu N, Elia AEH, Dharia T, Cantley LC, Soltoff SP (2005) The C2 domain of PKCδ is a phosphotyrosine binding domain. Cell 121:271–280.
24. Morinière J, Rousseaux S, Steuerwald U, Soler-López M, Curtet S, Vitte A-L, Govin J, Gaucher J, Sadoul K, Hart DJ, Krijgsveld J, Khochbin S, Müller CW, Petosa C (2009) Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature 461: 664–669.
25. Wu H, Feng W, Chen J, Chan L-N, Huang S, Zhang M (2007) PDZ domains of Par-3 as potential phosphoinositide signaling integrators. Mol Cell 28:886–898.
26. Zimmermann P, Meerschaert K, Reekmans G, Leenaerts I, Small JV, Vandekerckhove J, David G, Gettermans J (2002) PIP_2-PDZ domain binding controls the association of syntenin with the plasma membrane. Mol Cell 9: 1215–1225.
27. Leslie AGW (1999) Integration of macromolecular diffraction data. Acta Crystallogr D 55:1696–1702.
28. Otwinowski Z, Minor W, Processing of X-ray diffraction data collected in oscillation mode. In: Carter CWJ, Sweet RM, Eds. (1997) Methods in enzymology. Academic Press.
29. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallogr 26: 795–800.
30. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D50:760–763.
31. Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. Acta Crystallogr D 60:432–438.
32. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 47:110–119.
33. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D 60: 2126–2132.
34. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr D 55:247–255.
35. Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger T, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D 58:1948–1954.