Stereochemical Determinants of C-terminal Specificity in PDZ Peptide-binding Domains

A NOVEL CONTRIBUTION OF THE CARBOXYLATE-BINDING LOOP

Jeanine F. Amacher, Patrick R. Cushing, Christopher D. Bahl, Tobias Beck, and Dean R. Madden

From the Department of Biochemistry, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire 03755 and the Department of Structural Chemistry, Georg-August-Universität, Göttingen 37077, Germany

Background: PDZ-peptide binding specificities establish a complex network of protein-protein interactions in the cell. Results: Crystal structures of multiple PDZ-peptide complexes reveal distinct mechanisms for accommodating C-terminal ligand side chains. Conclusion: A residue in the PDZ “XΦ1,Φ2,” signature sequence co-determines peptide carboxylate and C-terminal side-chain binding. Significance: Understanding the stereochemical determinants of peptide binding leads to an improved ability to predict PDZ interaction specificity.

PDZ (PSD-95/Dlg/ZO-1) binding domains often serve as cellular traffic engineers, controlling the localization and activity of a wide variety of binding partners. As a result, they play important roles in both physiological and pathological processes. However, PDZ binding specificities overlap, allowing multiple PDZ proteins to mediate distinct effects on shared binding partners. For example, several PDZ domains bind the cystic fibrosis (CF) transmembrane conductance regulator (CFTR), an epithelial ion channel mutated in CF. Among these binding partners, the CFTR-associated ligand (CAL) facilitates post-maturational degradation of the channel and is thus a potential therapeutic target. Using iterative optimization, we previously developed a selective CAL inhibitor peptide (iCAL36). Here, we investigate the stereochemical basis of iCAL36 specificity. The crystal structure of iCAL36 in complex with the CAL PDZ domain reveals stereochemical interactions distributed along the peptide-binding cleft, despite the apparent degeneracy of the CAL binding motif. A critical selectivity determinant that distinguishes CAL from other CFTR-binding PDZ domains is the accommodation of an isoleucine residue at the C-terminal position (P0), a characteristic shared with the CAL PDZ domain. Comparison of the structures of these two PDZ domains in complex with ligands containing P0 Leu or Ile residues reveals two distinct modes of accommodation for β-branched C-terminal side chains. Access to each mode is controlled by distinct residues in the carboxylate-binding loop. These studies provide new insights into the primary sequence determinants of binding motifs, which in turn control the scope and evolution of PDZ interactomes.

Within each cell, diverse arrays of peptide recognition domains (PRDs) contain binding sites that engage cognate peptide ligands, forming interaction networks that regulate protein trafficking, localization, and activity. The specificity of each PRD is reflected in a linear binding motif that prescribes both the identity and spacing of amino acids required for the interaction. In principle, such preferences are encoded by the PRD primary sequence, which determines the stereochemistry of the peptide-binding cleft, and thus favors interactions with partners that share a complementary motif. Deciphering this “sequence-motif” code would enable us to identify potential interactions and design selective inhibitor sequences in silico (e.g. Ref. 1). Unfortunately, even for highly conserved interactions, connections between PRD sequence determinants and specific target motif preferences often remain obscure.

This challenge is particularly acute for interactions mediated by PDZ domains, which were first characterized in the proteins PSD-95, Dlg, and ZO-1. PDZ domains are now recognized as one of the largest families of PRDs and are responsible for modulating the activity of a wide variety of receptors and enzymes. Multidomain PDZ proteins also form scaffolds that regulate the assembly and composition of supermolecular complexes (2). As a result, PDZ domains play key roles in a number of human diseases, including cancer (3–5), epilepsy and pain (6, 7), and cystic fibrosis (8, 9).

1 This work was supported, in whole or in part, by National Institutes of Health Grant RO1-DK073309 (to D. R. M.) and Training Grants T32-GM08704 (to J. F. A.) and T32-DK07301 (to P. R. C.), Institutional Development Awards P20-RR018787 and P20-GM103413 to the Dartmouth Lung Biology Center of Biomedical Research Excellence, Cystic Fibrosis Foundation Grant RDP STANTO11R0, and a Hitchcock Foundation Program Project Grant.

2 Present address: Dept. of Organic Chemistry, Eidgenössische Technische Hochschule, Zürich, Switzerland.

3 The abbreviations used are: PRD, peptide recognition domain; PDZ, PSD-95, Dlg, and ZO-1; CFTR, cystic fibrosis transmembrane conductance regulator; CAL, CFTR-associated ligand; CALP, CAL PDZ domain; TIP-1, Tax-interacting protein-1; NHERF, Na+/H+ exchanger regulatory factor; SIRAS, single isomorphous replacement with anomalous scattering; I3C, 5-ami-no-2,4,6-triliodoisophthalic acid; SSR5, somatostatin receptor subtype 5; iCAL, inhibitor of CAL; FP, fluorescence polarization; TCEP, tris[2-carboxyethyl]phosphine; MR, molecular replacement; r.m.s., root mean square; PDB, Protein Data Bank.
Although experimental studies have characterized target preferences for a large number of PDZ domains (10, 11), binding information is absent or available only for a limited subset of ligands for many other domains. Furthermore, as for other PRDs, despite an abundance of work investigating recognition patterns and motifs, it is still difficult to predict the binding motif of a given PDZ domain based on its protein sequence or binding-site stereochemistry.

In part, this is due to the surface geometry of PDZ-peptide interactions. All PDZ domains share a conserved tertiary structure, comprised of two α-helices and an antiparallel β-sheet. Canonical target peptides bind in a shallow cleft between the β2 strand and α2 helix, forming an antiparallel β-sheet hydrogen-bonding network and allowing the ligand C-terminal residue to interact with the carboxylate-binding loop on the PDZ domain (12). This loop contains a conserved χ2,ΓΦ2 sequence (Φ = hydrophobic) that facilitates main-chain hydrogen bonding to the peptide carboxylate (13). As a result, most PDZ proteins bind C-terminal target sequences. In addition, this loop fixes the orientation of the C-terminal (P0) side chain to point into a hydrophobic pocket, imposing a primary specificity filter for peptides carrying a hydrophobic residue at the P0 position (14).

PDZ binding specificity is also strongly influenced by the second position from the C terminus (P–1). For class I PDZ domains, an essential hydrogen bond is formed between the Oγ atom of the P–2 residue and the first residue of the α2 helix, α2-1 (His), imposing a requirement for Ser or Thr (15). Subsequent work has revealed additional preferences for many domains, leading to more differentiated motifs (11, 16–18). Nevertheless, the motifs overlap, enabling multiple PDZ proteins to bind a common target, and several domains retain highly degenerate motifs. As a result, selective targeting is a challenge.

We previously reported a peptide inhibitor of the interaction of the cystic fibrosis transmembrane conductance regulator (CFTR) with the PDZ domain of the CFTR-associated ligand (CAL). The resulting inhibitor of CAL (or iCAL36) peptide is a decamer with the sequence ANSRWPTSSII (8, 9). Although the Na+/H+ exchanger regulatory factor (NHERF) proteins also bind CFTR through their PDZ domains, they do not interact with iCAL36 (8, 9).

One of the earliest observations in the development of iCAL36 was that CAL and NHERF PDZ domains exhibit different secondary preferences for the P0 residue. Although each prefers a Leu residue at this position, CAL also favors either a Val or Ile, whereas the NHERF domains can accommodate a Phe instead.

To understand the stereochemical basis for this difference, we determined a high-resolution structure of the CAL PDZ domain (CALP) in complex with either iCAL36 or a variant bearing a C-terminal Leu (iCAL36L). In addition, we co-crystallized both peptides with the Tax-interacting protein-1 (TIP-1), another PDZ protein that accommodates β-branched C-terminal side chains (19) and binds iCAL36.4 By comparing these complexes to previously studied PDZ domains, we have identified alternative mechanisms of P0 side-chain accommodation. The resulting insights expand our understanding of sequence-motif relationships in this important class of PRDs.

**EXPERIMENTAL PROCEDURES**

Protein Expression and Purification—CALP (Ref. 20) and the NHERF PDZ domains (NHERF1 PDZ1 (N1P1), NHERF1 PDZ2 (N1P2), NHERF2 PDZ1 (N2P1), and NHERF2 PDZ2 (N2P2), Ref. 21) were expressed and purified as previously described. The TIP-1 purification has also been described in detail.4 Briefly, full-length TIP-1 was expressed with an N-terminal His10 tag followed by a modified HRV 3C protease recognition sequence (LEVLFQ*)G. TIP-1 was purified by immobilized metal-affinity chromatography and size-exclusion chromatography, and incubated with HRV 3C protease to remove the polyhistidine tag. CALP and TIP-1 proteins were stored in their respective size-exclusion buffers (20). For fluorescence polarization (FP) experiments, proteins were diazylated overnight into storage buffer (150 mM NaCl, 0.02% (w/v) sodium azide, 25 mM sodium phosphate, pH 8.0, 0.2 mM TCEP, 5% (v/v) glycerol) (20).

Structure Determination of CALP-iCAL36—CALP-iCAL36 complexes were crystallized, and data were collected as previously described (20). Phases were determined by single isomorphous replacement with anomalous scattering (SIRAS), using a CALP-iCAL36 crystal (reservoir solution: 200 mM NaCl, 100 mM Tris, pH 7.5, 32% (w/v) PEG 8000) that had been soaked 2 min in reservoir solution containing 250 mM 5-amino-2,4,6-triiodoisophthalic acid (IC3; lithium salt), and then transferred to cryoprotectant buffer (200 mM NaCl, 100 mM Tris, pH 7.5, 32% (w/v) PEG 8000, 20% (v/v) glycerol) prior to flash cooling in liquid nitrogen. For the anomalous data, crystal diffraction was evaluated at 100 K on beamline X6A at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Oscillation data were collected at λ = 1.7463 Å over a 720° range, using 0.25° frames and an exposure time of 3 s per frame. The two data sets were scaled using SADABS (22). The data were prepared using XPREP (23), and heavy atoms were located using SHELXD (24). Phase extension and density modification were performed with SHELXE including an autotracing algorithm (25). Density modification, including noncrystallographic symmetry averaging, was performed using the CCP4 suite (26) and the in-house modeling program ENVAT (27). Model building and refinement were performed using PHENIX (28, 29), and manual refinement done using COOT (30) (Table 1). After refinement, model geometry was assessed using MOLPROBITY (31) and the PDB validation server (32–34). All residues are found in the favored (95.5%) and allowed (4.5%) regions of the Ramachandran plot. Coordinates and structure factors were deposited in the Protein Data Bank with accession code 4E34.

To locate the position of the I3C compound in the electron density map of the heavy atom substructure, we exploited the defined geometry of the iodine peaks (35). To visualize the interaction of I3C with the protein-peptide complex, a molecular replacement (MR) search of the anomalous dataset was performed in PHENIX (28, 29), using a single proteomer (A) of the refined CALP-iCAL36 structure as the search model. Model
CAL and TIP-1 PDZ \( \Phi \) Binding Stereochemistry

building, refinement, and assessment were performed as described above (Table 1). Previously calculated restraints, derived from a small molecule crystal structure, were used in refinement of the I3C molecule (36). Due to the low occupancy of the I3C compound, density is only seen for the 3 iodine atoms, whose occupancies were set to 0.3. The final \( B \)-factors for the I3C iodine atoms were 67, 76, and 27 Å².

Structure Determination of CALP-iCAL36l—The CALP-iCAL36l complex crystallized by vapor diffusion in hanging-drop format under conditions similar to those for CALP-iCAL36l (20). The iCAL36l peptide was synthesized and purified by the Tufts Peptide Core Facility. CALP was dialyzed into crystallization buffer (10 mM HEPES, pH 7.4, 25 mM NaCl) and adjusted to a final concentration of 5.5 mg ml⁻¹ of protein, and iCAL36l was added to a final concentration of 1 mM. The CALP-iCAL36l solution was mixed in a 1:1 ratio with reservoir solution, total volume was 4 μl, and the drop was allowed to equilibrate by vapor diffusion at 291 K. As seen with the CALP-iCAL36 complex, crystals appeared in 2–4 days and continued to grow for up to 10 days (20). The crystal used for diffraction was obtained using a reservoir solution of 150 mM NaCl, 100 mM Tris, pH 7.5, 33% (w/v) PEG 8000.

Prior to data collection, the crystal was transferred into cryoprotectant buffer (150 mM NaCl, 100 mM Tris, pH 7.5, 33% (w/v) PEG 8000, 20% (v/v) glycerol) and flash cooled by plunging into a liquid nitrogen bath. Crystal diffraction was evaluated at 100 K on beamline X6A (NSLS). Oscillation data were collected at \( \lambda = 1.0000 \) Å over a 180° range, using 0.3° frames and an exposure time of 1 s per frame. Diffraction images were processed using the XDS package (37) (Table 1). An MR search was performed with PHENIX (28, 29), using a CALP-iCAL36l protomer (A), excluding peptide, as a search model. To minimize model bias in either the peptide or the \( \Phi \) and \( \Psi \) residues, initial modeling was performed using a composite-omit strategy (38). As a further control for bias, the procedure was repeated using the other protomer (B) as a search model, and the solutions were compared with ensure that there were no significant differences. Model building and refinement were performed as described for the CALP-iCAL36l structure above. Final refinement statistics are shown in Table 1. Coordinates and structure factors were deposited in the Protein Data Bank with accession code 4E35.

Structure Determination of TIP-1-iCAL36l—Crystals of TIP-1-iCAL36l were obtained by the same protocol described above and elsewhere,4 using a solution of 5.5 mg ml⁻¹ of TIP-1 and 1 mM iCAL36l. The crystal used for data collection was obtained with a reservoir buffer of 250 mM KSCN, 100 mM MES, pH 6.0, 25% (w/v) PEG3350.

The crystal was transferred into cryoprotectant buffer (250 mM KSCN, 100 mM MES, pH 6.0, 30% (w/v) PEG 400) prior to flash cooling in liquid nitrogen, and diffraction data were collected on beamline X6A (see above) at \( \lambda = 0.9795 \) Å over a 270° range, using 0.2° frames. Data were processed using the XDS package (37) (Table 1). MR was performed using PHENIX (28, 29), with a protomer (A) of TIP-1 from the TIP-1-iCAL36l structure as a search model. Electron density for the \( \Phi \) Leu residue confirmed the MR solution (data not shown). Model building, refinement, and assessment were performed as described above for the CALP-iCAL36l structure. Final refinement statistics are presented in Table 1. Coordinates and structure factors were deposited in the Protein Data Bank, with accession code 4E3B.

Structure Analysis—Alignments of the protomers of CALP-iCAL36l, CALP-iCAL36l₁, and N1P1-CFTR were performed using LSQKAB (39), part of the CCP4 suite of programs (26). To generate the superimposition of multiple Class I PDZ domain-ligand structures, alignments of the \( \beta_2/\alpha_2 \) secondary structure elements were performed using PYMOL (40), with CALP-iCAL36l as the reference structure. Structures were aligned and colored according to carboxylate-binding loop motif for the following PDB entries: 1BE9, 1I92, 1N7T, 1TP5, 1VJ6, 2EGK, 2EGN, 2FCF, 2FNE, 2H2B, 2H2C, 2HE4, 2I01, 2I0L, 2I1N, 2IWN, 2IWQ, 2OCS, 3DIW, 3JXT, 3NGH, 3GJ9, 3QGL, 3QJM, 3QIN, 3RL7, and 3SOE. Analysis of the \( B \)-factors of these structures in the vicinity of the loop suggests that their positions are well defined in all cases (maximum difference 42 Å²).

FP Assays—FP assays were performed as described previously (9, 21), using the following FP buffer: 150 mM NaCl, 0.02% (w/v) sodium azide, 50 mM sodium phosphate, pH 7.4, and 0.1 mM TCEP. The final protein concentration for competition binding experiments was \( 1.5 \times K_d \) for CALP, TIP-1, N1P1, N1P2, and N2P2. The following reporter peptides were used: \( F^-iCAL36l \) (\( F^-ANSRWPTSII \)) for CALP (\( K_d = 0.97 \mu \)M) and TIP-1 (\( K_d = 0.54 \mu \)M); \( F^-CFTR6 \) (\( F^-VQDTRL \)) for N1P1 (\( K_d = 0.37 \mu \)M) and N1P2 (\( K_d = 1.08 \mu \)M); and \( F^-CFTR10 \) (\( F^-TEEVQDTRL \)) for N2P1 (\( K_d = 0.32 \mu \)M) and N2P2 (\( K_d = 0.23 \mu \)M). \( F^- \) corresponds to a fluorescein group linked to the peptide N terminus via an aminohexanoic acid linker. The \( K_d \) values of the reporter peptides were calculated as previously described (21). The \( K_d \) values were calculated using a SOLVER-based least squares fitting algorithm in EXCEL (9, 21). Values for weakly interacting peptides, defined as \( K_d > 1000 \mu \)M, were calculated by comparison with modeled displacement isotherms (9).

RESULTS

Structure Determination of the CALP-iCAL36l Complex—As previously reported, although CALP-iCAL36 co-crystals diffract to 1.4-Å resolution, molecular replacement searches failed to yield a correct phase model (20). To obtain de novo phase information, we soaked CALP-iCAL36 crystals in the presence of I3C (35). This compound was recently developed as a tool to facilitate the incorporation of heavy atoms into protein crystals for phasing by anomalous scattering and/or isomorphous replacement, an attractive option, as CALP contains no methionine residues for selenomethionine labeling. Soaking in cryoprotectant buffer yielded an I3C derivative CALP-iCAL36 co-crystal that diffracted to 1.98-Å resolution and exhibited anomalous differences (SigANO = 2.0 in the 6.99–6.0 Å resolution shell) (37).

Initial phase determination by SIRAS failed using traditional protocols, yielding a figure of merit of 0.50, and pseudo-free correlation coefficient of 53%. However, using the SHELXE autotracing function (25), we identified a solution with an figure of merit value of 0.69 and pseudo-free correlation coefficient of 72%. Phase quality was confirmed by the presence of
side chains that were not present in the polyalanine phasing model (Fig. 1A), as well as clear electron density for the bound peptide (data not shown).

The asymmetric unit contains two CALP:iCAL36 complexes (protomers A and B), bridged by a single I3C molecule "halogen bonding" (41) to the carbonyl oxygens of Ile315 in the A-protomer and His296 in the B-protomer (Fig. 1, B and C). Following iterative rounds of density modification, model building, and maximum likelihood refinement to 1.4-Å resolution, the final model exhibits excellent agreement with the experimental diffraction data, yielding values for $R_{work}$ and $R_{free}$ of 0.174 and 0.197, respectively. It also conforms to standard peptide geometrical constraints (Table 1). Alternative side-chain conformations were exhibited by five solvent-exposed residues. The final model includes 87 residues of each CALP protomer, along with their corresponding decameric peptide ligands. Least squares superposition of the noncrystallographic symmetry-related complexes by main-chain atoms shows they share a highly similar domain structure: 344 atoms align with a root mean square deviation (r.m.s. deviation) of 0.45 Å, and no individual differences exceed a threshold of 3 Å r.m.s. deviation. The largest variation between the two chains is in the carboxylate-binding loop, at residues Gly300 and Ile301, with offsets of 0.9 and 1.2 Å, respectively.

**CALP:iCAL36 Binding Stereochemistry**—The resulting structure provides the first high resolution images of CALP ligand-binding interactions. Consistent with previous NMR structural analyses of the domain (42, 43), both CALP protomers exhibit a standard PDZ tertiary structure (44), consisting of five...
β-strands (β1–β5) and two α-helices (α1–α2) (Fig. 1B). Comparison of the A- and B-protomers to known structures using the DALI distance-alignment matrix algorithm (45) confirmed that the 100 structures most similar to CALP are all PDZ domains. For both the A- and B-protomers, the most similar domain is the A-protomer of α-syntrophin (PDB code IQAV) with Z-scores of 18.2 and 17.8, respectively. Superposition of 85 Cα atoms of either protomer with a previously determined CALP NMR structure (42) also reveals close similarity, with equivalent secondary structure elements and an r.m.s. deviation of 1.4 Å. Main-chain offsets are distributed evenly throughout the entire protein, which may have confounded previous phase determination efforts by MR (20).

In both CALP protomers, the P0 to P−5 residues bind in an extended peptide-binding groove that runs along one surface of the domain, where they engage in canonical Class I interactions (Fig. 1D). Through main-chain hydrogen bonds to the CALP β2 strand, the peptide forms an additional strand of the central anti-parallel β-sheet in the PDZ domain (Fig. 1E). The serine residue at the P−2 position of iCAL36 forms a hydrogen bond with His349 of CALP, whereas the C-terminal oxygen atoms interact with Leu309 and Gly310 in the CALP carboxylate-binding motif. Finally, the P0 side chain is buried in a pocket formed by Gly306 and Ile301, the GΦ2 pair of the carboxylate-binding motif, as well as Ile303 and Val353.

At position P−4, there is weak evidence for an alternative proline pucker in the B-protomer complex. However, when both conformers were modeled, the alternative conformation was refined to an occupancy value ≈20%. Although a proline pucker can influence main-chain orientation (46), in this case, the effects on neighboring residues are on the scale of the maximum likelihood coordinate error (0.16 Å). Thus, our final model includes only the dominant P−4 conformation.

Upstream of the P−6 residue (i.e. closer to the peptide N terminus), the conformation of both peptides appears to be influenced by crystal lattice contacts, involving P−8 Asn and P−9 Ala in the A-protomer, and the P−6 Arg and P−8 Asn residues in the B-protomer. However, the P0 to P−5 positions of each protomer are very similar (r.m.s. deviation <0.16 Å), and the stereochemical contacts of the peptide residues with the PDZ binding site are conserved.

Previous analyses indicate that up to seven ligand residues can interact with the PDZ domain, substantially extending classical motifs (11). Although the CAL binding motif shows clear preferences only at the P0 and P−2 positions, studies of CAL-binding sequences identified affinity contributions extending past the P−3 position (9). Our CALP-iCAL36 complex structure supports the idea that multiple non-motif positions along the binding cleft contribute to the interaction. For example, the Trp residue at P−5, the most N-terminal position not affected by lattice contacts, interacts with a mostly hydrophobic ledge on the CALP surface, composed of residues His309, Gly310, and Val311 (Fig. 2A). In addition, the P−4 Pro residue interacts with a groove formed by CALP residues Gly305, His309, and His349 (Fig. 2B).

P0 Contributions to CAL/NHERF Selectivity—One of the earliest observations from peptide array experiments was that both CALP and NIP1 share a preference for a Leu residue at P0, but that only CALP binds peptides containing a P0 Ile. In a sequence derived from the somatostatin receptor subtype 5 (SSR5) C terminus, replacement of the wild-type P0 Leu with Ile caused only a 1.6-fold weakening of CALP affinity, but a 47-fold weakening of NIP1 affinity (9). To test whether these distinct P0 preferences were dependent on the overall sequence of the peptide, we measured binding affinities for two additional pairs of peptides, with each pair differing only at the C-terminal position.

### Table 1

Data collection and refinement statistics

|                     | CALP-iCAL36 | Anomalous data | CALP-iCAL36 | TIP-1-iCAL36 |
|---------------------|-------------|----------------|-------------|--------------|
| **Data collection** |             |                |             |              |
| Space group (number)| P2, 2, 2, (19) | P2, 2, 2, (19) | P2, 2, 2, (19) | P1 (1)       |
| Unit cell dimensions. | 35.9, 47.7, 97.3 | 36.2, 47.5, 97.4 | 36.1, 47.6, 97.8 | 27.1, 33.6, 66.9 |
| α, β, γ (°)         | 90, 90, 90  | 90, 90, 90      | 90, 90, 90  | 80.1, 86.9, 89.2 |
| Resolution (Å)      | 48.65–1.4 (1.45–1.4) | 50–1.98 (2.07–1.98) | 19.5–1.4 (1.49–1.40) | 19.2–1.50 (1.58–1.50) |
| Bond length RMSD (Å) | 0.006 | 0.007 | 0.006 | 0.006 |
| Bond angle RMSD (°) | 1.12 | 1.03 | 1.13 | 0.98 |

*Values in parentheses are for data in the highest resolution shell.

*αRmerge = [Σ(h)l(h)−l(h)]/[Σl(h)], where l(h) values are the i-th and mean measurements of the intensity of reflection h.

*βRmerge = 2[S(h)obs − S(h)calc]/[S(h)calc], h ∈ [working set].

*Rmerge = 2[S(calc)obs − S(calc)calc]/[S(calc)calc], h ∈ [test set].

Includes alternate conformations.

Core/allowed/generally allowed/disallowed.
The sequence of iCAL36 (ANSRWPTSIL) corresponds to that of iCAL36 with a Leu substitution at P0. The sequence of CFTR (TEEEVQDTR) corresponds to the CFTR C terminus (TEEEVQDTRL) with an Ile substitution at P0.

The results obtained with both pairs of sequences confirm the selectivity pattern seen with SSRS. For the CALP domain, FP competition assays show very similar binding of CALP to iCAL36, versus iCAL36, and also to CFTR versus CFTR (Fig. 3A, Table 2). In contrast, for the NHERF1 and NHERF2 PDZ domains (N1P1, N1P2, N2P1, and N2P2), binding affinities are weakened when Ile is substituted for Leu at P0. Specifically, the Kd values are 20- to 35-fold weaker for CFTR than for the Leu-bearing WT CFTR sequence (Table 2). Likewise, none of the NHERF domains exhibit binding to iCAL36, which carries a P0 Ile, but three of four show detectable affinity for iCAL36, (Fig. 3B, Table 2).

Thus, in multiple background sequences, the NHERF PDZ domains exhibit a robust preference for Leu over Ile at P0, whereas CALP appears largely indifferent. These differing P0 preferences are an essential component of the selectivity of iCAL36 for CALP versus the NHERF PDZ domains. iCAL36 binds CALP >170-fold more tightly than any of the four NHERF domains (9). However, if the P0 Ile is replaced with a Leu this difference is largely erased: iCAL36N binds N2P1 (Kd = 77 ± 2 μM) only 3-fold more weakly than it binds CALP (Kd = 24 ± 2 μM).

P0 Specificity Determinants in the Carboxylate-binding Loop—Given this central contribution to target selectivity, we wished to understand the stereochemical differences that enable CALP, but not the NHERF domains, to accommodate a P0 Ile. In the N1P1-CFTR structure (PDB code 1I92, Ref. 47), in silico replacement of the C-terminal Leu with an Ile side chain leads to steric clashes of the Ile β-branch with the narrow entrance of the P0 pocket (data not shown), providing a straightforward explanation for its failure to bind iCAL36. However, previous structures of the CALP domain had yielded conflicting views of the P0 binding pocket. Our NMR structure (PDB code 2DC2, Ref. 43) shows a CALP pocket more constricted than either.

These conflicting observations not only complicated our understanding of Ile accommodation, but also suggested that the CALP pocket could be inherently flexible, and thus capable of induced-fit binding. To visualize the stereochemistry of accommodation, we therefore compared the N1P1-CFTR complex to the co-crystal structure of CALP bound to iCAL36, which includes a P0 Ile side chain. Superposition of the β2 and α2 secondary structural elements that flank the peptide-binding cleft yielded an r.m.s. deviation of 0.62 Å.

For the A-protomer of the CALP-ICAL36 complex, the dimensions of the P0 binding pocket are surprisingly similar to those found in N1P1 (Fig. 4B, orange and green meshes). However, as noted above, the N1P1 pocket geometry does not appear able to accommodate a branched Ile side chain without further changes. Instead, A-protomer binding appears to involve a concerted translation of the carboxylate-binding loop and the P0 residue. Compared with its position in the N1P1-CFTR complex, the Cα atom of the P0 residue in the carboxylate-binding loop has moved nearly 2 Å closer to the pocket in the A-protomer structure (Fig. 4C). A corresponding displacement of the peptide carboxylate group reorients the Ile side chain within the pocket and alleviates the steric clashes near the entrance (Fig. 4B).

In contrast, for the B-protomer of the CALP-ICAL36 complex, the P0 binding pocket is surprisingly similar to that of N1P1: i) displacement of the P0 residue within the carboxylate-binding loop (“loop offset”) and/or ii) expansion of the P0 pocket (Fig. 4B, yellow mesh) and enables CALP to accommodate the peptide Ile side chain despite a more modest translation of the carboxylate group.

Overall, it appears that CAL can utilize two distinct modes for expanding its P0 binding repertoire compared with that of N1P1: i) displacement of the P0 residue within the carboxylate-binding loop (“loop offset”) and/or ii) expansion of the P0 pocket via reorientation of the P0 Ile (“pocket expansion”). Each of these strategies involves a rearrangement of one of the two hydrophobic residues of the XΦ1,Φ2 motif.

Influence of P0 Side-Chain Identity on the Binding Site—To determine whether these distinct Φ1 and Φ2 conformations are also observed in the presence of a P0 Leu side chain, we determined the structure of the CALP domain bound to the iCAL36L peptide, using MR protocols designed to avoid model bias. Following refinement, the model exhibits Rwork and Rfree values of 0.184 and 0.198, respectively (Table 1), and also contains two molecules in the asymmetric unit. To compare P0 binding modes, we aligned the iCAL36 and iCAL36L structures by pairwise superposition of the β2 and α2 secondary structural elements, which flank the peptide-binding cleft. Overall, the CALP-ICAL36 and CALP-ICAL36L structures are very similar, with an r.m.s. deviation of 0.04 Å.

Unlike the iCAL36 complex, both of the CALP-ICAL36L complexes in the asymmetric unit exhibit a shared Φ1 and Φ2 conformation, most similar to that seen in the CALP-ICAL36 A-protomer (Fig. 5A). Instead of an expanded pocket, both
placement of the for CALP (reported in Table 2. Determined by least-squares fitting (9, 21). Logistic curve fits (peptides were added to mixtures of the specified PDZ domain and fluorescent reporter peptides (P0-CALP for CALP and TIP-1, and P0-CFTR for N1P1). The average value for a control DMSO titration is also shown (square with intersecting lines). K inhibition constants and fractional reporter occupancy values were determined by least-squares fitting (9, 21). Logistic curve fits (lines) are shown to illustrate offsets in IC50 values. Summary values of triplicate experiments are reported in Table 2.

**TABLE 2**

| Peptide | Sequence | CALP | TIP-1 | N1P1 | N1P2 | N2P1 | N2P2 |
|---------|----------|------|-------|------|------|------|------|
| CFTR    | TEEVQDTRL | 420 ± 80 | >1000 | 0.5 ± 0.2 | 1.8 ± 0.5 | 0.8 ± 0.3 | 0.07 ± 0.05 |
| CFTR(2) | TEEVQDTRI | 490 ± 120 | >1000 | 11 ± 3 | 40 ± 1 | 19 ± 3 | 2.5 ± 0.4 |
| iCAL36  | ANSRWPTSI | 23 ± 8 | >1000 | >1000 | >1000 | >1000 | >1000 |
| iCAL36(2) | ANSRWPTSIL | 24 ± 2 | 0.9 ± 0.4 | 230 ± 100 | >1000 | 77 ± 2 | 530 ± 120 |

**FIGURE 3.** P0 Leu/Ile substitutions affect peptide binding of N1P1, but not CALP or TIP-1 domains. Representative FP displacement isotherms are shown for CALP (A), N1P1 (B), and TIP-1 (C) domains. Increasing concentrations of iCAL36 (circles), iCAL36(2) (dark squares), CFTR (diamonds), and CFTR(2) (triangles) peptides were added to mixtures of the specified PDZ domain and fluorescent reporter peptides (P0-iCAL36 for CALP and TIP-1, and P0-CFTR for N1P1). The average value for a control DMSO titration is also shown (square with intersecting lines). K inhibition constants and fractional reporter occupancy values were determined by least-squares fitting (9, 21). Logistic curve fits (lines) are shown to illustrate offsets in IC50 values. Summary values of triplicate experiments are reported in Table 2.

**TABLE 2**

| Peptide | Sequence | CALP | TIP-1 | N1P1 | N1P2 | N2P1 | N2P2 |
|---------|----------|------|-------|------|------|------|------|
| CFTR    | TEEVQDTRL | 420 ± 80 | >1000 | 0.5 ± 0.2 | 1.8 ± 0.5 | 0.8 ± 0.3 | 0.07 ± 0.05 |
| CFTR(2) | TEEVQDTRI | 490 ± 120 | >1000 | 11 ± 3 | 40 ± 1 | 19 ± 3 | 2.5 ± 0.4 |
| iCAL36  | ANSRWPTSI | 23 ± 8 | >1000 | >1000 | >1000 | >1000 | >1000 |
| iCAL36(2) | ANSRWPTSIL | 24 ± 2 | 0.9 ± 0.4 | 230 ± 100 | >1000 | 77 ± 2 | 530 ± 120 |

show a ~1 Å P0 offset compared with the CALP:iCAL36 B-protomer (Fig. 5B). Thus, the Φ1 and Φ2 conformations of the A-protomers appear relatively unchanged in the presence of Leu and Ile. For the B-protomers, the presence of a C-terminal Ile is associated with the two aforementioned conformational changes in the binding pocket: a Φ2 rotamer shift that expands the P0 pocket (arrow in Fig. 5B), together with a displacement of the Φ1 main chain that repositions the peptide carboxylate and C-terminal side chain (Fig. 5A).

**Accommodation of a P0 Ile without Pocket Expansion**—In an attempt to determine the relative contributions of these two mechanisms to the accommodation of Ile in solution, we first considered a possible influence of the crystal lattice on the observed stereochemistries of binding. In the A-protomers of both CALP:peptide complexes, a lattice hydrogen bond is formed by the main-chain carbonyl oxygen of His288, directly upstream of the 290GLGL293 motif. In contrast, in the B-protomers, no contacts are observed in the vicinity of the Φ1Cα2-loop: the closest symmetry-related atom is at least 6 Å away.

Based on these differential lattice interactions, we initially suspected that the Φ2 reorientation and associated P0 pocket expansion seen in the B-protomers were more likely to reflect the solution-state mechanism for P0 Ile accommodation by the CAL PDZ domain. This idea would be consistent with earlier proposals that the Φ2 residue is a critical determinant of P0 specificity (17).

To test whether Φ2 identity is a crucial component of P0 Ile accommodation, we investigated the binding stereochemistry of the TIP-1 PDZ domain. Like the NHERF domains, the TIP-1 PDZ domain contains a Phe side chain at the Φ2 position, in contrast to the Φ2 Leu found in CALP. Like CALP, TIP-1 also binds iCAL36, and the previously determined TIP-1:iCAL36 crystal structure (PDB code 3SFJ) superimposes well on the CALP domain with an overall Cα r.m.s. deviation value of 1.5 Å. Comparison confirms that the TIP-1 P0 pocket is not expanded relative to the CAL:iCAL36 A-protomer (Fig. 5C).

The TIP-1:iCAL36 complex also aligns well to the N1P1-CFTR complex (47) (r.m.s. deviation = 0.63 Å), revealing an offset of the carboxylate-binding loop and P0 residue (Fig. 5D). Thus, TIP-1 employs a mechanism for accommodating a P0 Ile that is similar to the loop offset seen in the CAL A-protomer. This result confirms that displacement of the Φ1Cα position is an alternative strategy for accommodating β-branched P0 residues in the presence of Φ2 rotamers (CAL A-protomer) or larger side chains (TIP-1) that restrict the volume of the pocket.

**The Loop Offset Is Also Observed with Ligands Containing a P0 Leu**—A loop offset is not observed exclusively in the presence of a peptide P0 Ile; as described above, it is also seen in the CAL:iCAL36L complex, which contains a P0 Leu. To test
whether this observation applies to the TIP-1 binding site as well, we investigated the stereochemistry of the interaction of the TIP-1 PDZ domain with the iCAL36L peptide. FP competition experiments reveal that TIP-1 binds iCAL36L (K_i = 0.36 μM) about 3-fold more tightly than it binds iCAL36 (K_i = 1.3 μM), confirming a modest preference for a P0 Leu (Fig. 3B). In parallel, we co-crystallized the TIP-1·iCAL36L complex and determined its structure by x-ray diffraction. The final structure was refined to R_work and R_free values of 0.187 and 0.216, respectively, with geometrical parameters as described in Table 1.

The TIP-1·iCAL36L structure contains two very similar molecules within the asymmetric unit, with a C_α r.m.s. deviation of 0.16 Å for main-chain atoms between protomers, and 0.08 Å in comparison to the TIP-1·iCAL36 complex, for β2/α2 elements. In particular, the stereochemistry of the P0 binding pockets and the position of the P0 main-chain atoms within the pockets are highly conserved (Fig. 5E). Comparisons to previously published structures of TIP-1 in complex with C-terminal β-catenin (P0 Leu) or Kir2.3 (P0 Ile) peptides also reveal very similar P0 binding pockets (48, 49) (data not shown), as does alignment to the N1P1-CFTR complex (47) (r.m.s. deviation = 0.64 Å). Consistent with the loop offset mechanism, the distance between the Φ1, C_α position of the N1P1-CFTR complex (Tyr24) and the corresponding position of TIP-1·iCAL36L (Leu30) is 1.7 Å.

This suggests that displacement of the carboxylate-binding loop is not only a prerequisite for binding Ile, or a β-branched P0 ligand side chain. Instead, it most likely represents a fundamental characteristic of the TIP-1 binding site, as it does for CALP.

Taken together, all four TIP-1 and CALP protomers in complex with Leu, and three of four protomers in complex with Ile share a common stereochemistry for the P0 pocket. For these structures, the loop-offset strategy is utilized to accommodate either a P0 Leu or Ile side chain. Because the loop offset is not selectively required by the presence of Ile, or a β-branched P0 ligand side chain, it most likely represents a fundamental conformation for the TIP-1 and CALP binding sites. As a result, although the Leu/Ile switch is associated with an expanded steric volume in the P0 pocket of the CALP B-protomer, such expansion is not a requirement of accommodation.

**Φ1 Residue Identity Influences the Position of the Carboxylate-binding Loop**—Because the C-terminal Ile side chain can be reoriented to fit even within a restricted P0 binding pocket, we wanted to determine which stereochemical factors prevented the N1P1 domain from utilizing the same strategy. The carboxylate-binding motif of N1P1 (GYGE) shares a Phe at the Φ2 position with TIP-1 (ILG). However, at the Φ1 position, N1P1 contains a bulky Tyr residue, in contrast to the Leu residue found in either CAL or TIP-1. This suggested that steric clashes may prevent domains with large Φ1 residues from accessing the loop offset mechanism of P0 Ile accommodation.

To test this hypothesis, we first transformed the TIP-1 carboxylate-binding loop (29GYGF29) in silico to reproduce the conformation observed in the TIP-1·iCAL36L complex (31ILGF31). This operation causes Tyr24 to clash sterically with Ile29, the α2–8 residue, as well as Ala132, Ala135, and Val136 (Fig. 5F). A similar set of clashes was observed when the N1P1 29GYGF29 residues were transformed to reproduce the conformation of the 29GYGF31 carboxylate-binding loop of CALP·iCAL36L (data not shown). Thus, the carboxylate-binding loop of N1P1 cannot adopt the conformation seen in CALP or TIP-1 due to steric constraints at the Φ1 position.

Based on these observations, we extended our analysis to other Class I PDZ domain complex structures. We first analyzed carboxylate-binding loop positioning as a function of the XΦ1Φ2 sequences and observed tight clustering for most variations (Fig. 6, A and B). Although the β1–β2 loop residues upstream of the first position in the motif are highly variable in these comparisons, this variability appears to reflect well determined differences in the upstream conformations, rather than dynamic instability: the average B-factors for residues both upstream and downstream of XΦ1Φ2 are reflective of stable conformations (data not shown).

The carboxylate-binding loop position also appears to be a property of each domain, and is strongly determined by the XΦ1Φ2 sequence. However, we do see slight variability in domains that contain smaller side chains at both the Φ1 and Φ2 positions (GLGI or SLGI; Fig. 6A). This may reflect flexibility.
for this class of domains, analogous to that seen in our CALP structures. We also see a single outlier for domains with a GFGF carboxylate-binding loop sequence. The MAGI3 PDZ3 structure (PDB code 3SOE; Fig. 6B) reveals an atypical 4.9-Å secondary structure offset that expands the overall steric volume of the peptide-binding cleft. As discussed below, this outlier also has implications for P0 motif predictions. To dissect the ability of the key side chains in this interaction to determine loop positioning, we then combined all structures based on the identity of the P0, Cα, or P0 residues. When loop position is analyzed solely by P0 residue identity (Fig. 6C), a spectrum of loop positions is seen for the Leu and Val side chains, but all the structures with Ile side chains contain aliphatic residues at P0, consistent with our hypothesis that P0 stereocchemistry can co-determine P0 side-chain accommodation.

When the loops are superimposed and colored by the P0 character (Fig. 6D), we see a spectrum of carboxylate-binding loop positions for each classification (aliphatic versus aromatic) that reveals no clear pattern. This is supported by our observation that, despite a shared P0 Phe, TIP-1 and N1P1 exhibit a differential Cα offset in this position of 1.3 Å upon P0 Leu binding (Fig. 5D).

In contrast, as seen in Fig. 6E, when the full set of loops is superimposed using the P0-based color scheme from Fig. 6, A and B, distinct clusters are seen. The loops bearing a P0 residue side chain (Fig. 6E, cool colors) are shifted toward the entrance of the P0 binding pocket, as seen in the CALP and TIP-1 loop offset structures. In contrast, the loops bearing an aromatic P0 side chain are shifted away from the entrance (Fig. 6E, warm colors), as seen in the N1P1 structure. As a further test of the role of the carboxylate-binding loop residues in ligand motif prediction, we evaluated the ability of the P0 residues to predict P0 preferences identified for 49 PDZ domains by a high throughput experimental analysis (see supplemental Table S4 in Ref. 11).

We first evaluated the predictive value of an aromatic P0 residue, which our studies suggest should disfavor a P0 Ile. Among the 49 human and Caenorhabditis elegans domains studied, 14 have an aromatic P0 residue. Only three of these accommodate a P0 Ile. ZO-1/TIP1 PDZ1 exhibits a weak pre-
A  Aliphatic $\Phi_1$

- GLGF
- ELGF
- SLGI
- GLGI
- ILGF

B  Aromatic $\Phi_1$

- GFGI
- NYGF
- GYGF
- TFGF
- GFGF

C  $\Phi_0$ Leu

- warm + cool colors

D  $\Phi_0$ Val

E  $\Phi_0$ Ile

- only cool colors

D  Aliphatic $\Phi_2$

- Aromatic $\Phi_2$

E  $\Phi_1$ = Leu (cool)

$\Phi_1$ = Aromatic (warm)
CAL and TIP-1 PDZ $P^0$ Binding Stereodynamics

dference for Ile; the carboxylate-binding loop sequence of ZO-1 is GFGI, suggesting a possible $\Phi_2$ rotamer-based mechanism of $P^0$ Ile accommodation. The other exceptions are the MAGI PDZ2 and PDZ3 domains. Fig. 6E confirms that the MAGI3 PDZ3 domain (GFGI) is an outlier, and the only warm colored (olive) loop that clusters with the cool colors. Additional conformational changes may enable the carboxylate-binding loop of MAGI3 PDZ3 to shift, allowing the domain to accommodate a $P^0$ Ile despite the aromatic residue at both the $\Phi_1$ and $\Phi_2$ positions. Overall, in the absence of such additional conformational changes, an aromatic $\Phi_1$ residue appears to strongly disfavor binding of a $P^0$ Ile.

We next evaluated the predictive value of an aliphatic $\Phi_1$ residue, which our studies suggest should permit accommodation of a $P^0$ Ile even in the presence of an aromatic side chain at the $\Phi_2$ position (example GLGF). Previous studies indicate that this combination should be unfavorable for Ile (17). However, of the 28 domains with a $\Phi_2$ Phe, 10 can accommodate a $P^0$ Ile. Of those, 8 carry a small $\Phi_2$ residue, consistent with the loop-ofset mechanism of accommodation seen in TIP-1 (ILGF). The other two are the MAGI3 domains. Overall, these results substantiate the importance of the $\Phi_1$ residue in positioning the $P^0$ ligand side chain within its binding pocket.

**DISCUSSION**

A detailed understanding of PRD binding specificity is a requirement for mapping the switchable protein-protein interactions that regulate cellular functions. Although it is relatively straightforward to identify the PRDs themselves based on sequence alignments, their short and often degenerate recognition sequences are much harder to capture by brute-force analysis of even large-scale sequence datasets (50). Several high-throughput approaches have been employed to provide experimental constraints for this problem (11, 51, 52). Nevertheless, robust biochemical detection is challenging, due to the transient nature of the protein-peptide interactions; the associated affinities are frequently 1–100 $\mu$M or even weaker (12). As an alternative approach, numerous groups have developed algorithms to identify PRD specificity determinants based on the primary sequence of the domain (e.g. Refs. 53–61). Despite the considerable success of these approaches, in many cases, the stereochemical basis of even dominant specificity determinants remains obscure.

A clear example is provided by the secondary $P^0$ motif preferences of the cluster of PDZ proteins that share affinity for the C terminal of CFTR. Published NMR structures had provided conflicting models of how the CAL domain can accommodate $\beta$-branched side chains (42, 43). Our crystal structures of CALP bound to iCAL36 and iCAL36$_1$ address this apparent contradiction. They reveal that the aliphatic CALP $\Phi_1$ and $\Phi_2$ side chains facilitate binding of a $\beta$-branched peptide C-terminal Ile residue by two distinct mechanisms (Fig. 4B) that highlight the central role of these side chains in PDZ binding interactions.

In fact, PDZ domains were initially referred to as “GLGF repeats” (13). The first PDZ domain structures revealed that the GLGF sequence forms critical hydrogen bonds with the peptide main-chain carboxylate (62). In doing so, it positions the C-terminal peptide side chain within a pocket at the bottom of the cleft. As additional PDZ domain family members from multiple genera were identified, it became apparent that sequence variations in 3 of the positions are ubiquitous (15, 63, 64). As a result, the GLGF loop became known as the GFGF$\Phi_1$ or $\Phi_2$G$\Phi_2$ motifs (44).

A recent analysis of the effect of these sequence differences on domain binding suggested that an aromatic residue at the $\Phi_2$ position, which forms one side of the $P^0$ binding pocket, could sterically inhibit binding of a bulky C-terminal peptide side chain (17). However, our structures show that an Ile-bearing peptide can bind to TIP-1 even in the presence of a bulky Phe side chain at $\Phi_2$ (Fig. 5C). Based on analysis of a collection of Class I PDZ domain structures, the loop offset is broadly regulated by the steric volume of the $\Phi_1$ residue (Fig. 6E), which has not previously been implicated in the stereochemistry of $P^0$ binding. Thus, the two hydrophobic residues of the $X\Phi_2G\Phi_2$ loop together co-determine stereocchemical preferences at the C-terminal residue: taking both into account clearly improves prediction of one of the core determinants of PDZ binding specificity.

The ability to connect PDZ domain sequences directly to binding motif preferences will also help to understand the evolution of PDZ-mediated interaction networks. Interaction data suggest that PDZ target preferences are hard-wired into the PDZ domain itself (18, 65, 66). Indeed, one of the hallmarks of organismal complexity is the expansion of the number of PDZ domains and the rewiring of their interactions (66–68). Given the limitations of determining individual PDZ interactomes experimentally, the refinement of sequence-based computational specificity predictions will be critical to understanding these fundamental evolutionary processes. Indeed, understanding the impact of site-directed mutagenesis on PDZ domain binding specificity has been the target of recent computational challenges (e.g. Refs. 55 and 60).

**FIGURE 6. Loop conformation is influenced by $\Phi_1$, residue stereodynamics.** Carboxylate-binding loop conformations are sorted and colored as a function of $\Phi_1$, $G\Phi_2$ sequence (A and B) or clustered as a function of ligand $P^0$ identity (C). $\Phi_2$, character (O), or $\Phi_1$, character (E). Protein and peptide C$_\text{C}$ traces are shown in gray, following least-squares superposition based on a $\alpha$2/$\beta$2 secondary structure main-chain atoms. A–C and E, the carboxylate-binding loop structures are colored based on $\Phi_1$, $G\Phi_2$ sequence. A, sequences bearing an aliphatic $\Phi_1$, residue are in cool colors: GLGF, purple (PDB entry codes 18E9, TIPS; GFGF, 2E28, 2MLG, 2L0L, 2L1N, 3XRT, and 3RL7); ILGF, green (TIP1-iCAL36, PDB entry code 35FJ; TIP-1:iCAL36, PDB entry codes 42E8, 3DW, and 3GA9; GLGI, cyan (CALP-iCAL36, PDB entry code 4E34; CALP:CALP-iCAL36, 4E35); SLGF, marine blue (PDB entry codes 1VJ6, 2FCF, and 21QW); ELGF, dark blue (1NNT); B, sequences bearing an aromatic $\Phi_1$, residue are shown in warm colors: GFGF, olive (PDB entry codes 3QJ3, 3QJN, and 3QOE); GYGF, orange (PDB entry codes 1192, 2HE4, 2OS, 3QGL, and 3QRE; NYGF, yellow (PDB entry code 3NGN); GFGI, sand (PDB entry codes 2HZB and 2LC2); TFGF, red (PDB entry codes 2EGN and 2EGK). The MAGI3 PDZ3 (X$\Phi_1$, G$\Phi_2$ = GFGF; PDB entry code 3SOE) structure represents the largest conformational outlier compared with the other structures bearing the same carboxylate-binding loop sequence (black arrow). C, the structures shown in panels A and B are sorted based on the identity of the $P^0$ residue (Leu, Val, and Ile). Sequences containing either aromatic and aliphatic $\Phi_1$, residues are found in the $P^0$ Leu and Val clusters (left and middle), but only structures containing an aliphatic $\Phi_1$, residue are found in the $P^0$ Ile cluster (right). D, the complete set of carboxylate-binding loop structures colored by the $\Phi_2$, character (aliphatic = cyan, aromatic = orange) shows no consistent influence of the $\Phi_1$, residue on the position of the loop. E, in contrast, when the complete set of structures is superimposed and colored as in panels A and B, the $\Phi_2$, atoms of the $\Phi_1$, residues (spheres) cluster by aliphatic (cool) and aromatic (warm) character (black arrow).
Critically, a more complete understanding of sequence-specificity relationships will not only enhance our ability to predict and interpret cellular interactomes, it will also facilitate the design of specific PDZ domain inhibitors as therapeutic agents, even among domains with overlapping endogenous targets (6, 8, 9, 69). Other studies have shown that the conformation of the carboxylate-binding loop is not determined exclusively by the $\Phi_1$ and $\Phi_2$ residues; it is also affected by surrounding residues, including the $\beta_1$-$\beta_2$ loop and by nearby allosteric interactions, for example, with other PDZ domains (70–73). Thus, one way to target PDZ domains may be through small molecule-mediated allosteric regulation (69). Despite the promiscuous nature of PDZ domain binding, our data suggest that the design of efficacious and specific PDZ inhibitors can exploit the subtle, yet crucial, carboxylate-binding loop determinants that have been carefully calibrated throughout evolution.

Acknowledgments—We thank Abigail Fellows and Jessica St. Laurent for technical assistance. We also thank the entire Madden lab, Ernest Heimsath, Dr. George M. Sheldrick (Georg-August-Universität), and Dr. James Fêthière (University of Montreal) for advice and discussions, as well as Edwin Lazo, Dr. Vivian Stojanoff, and Dr. Jean Jakoncic (NSLS) for assistance with data collection and analysis.

REFERENCES

1. Obenauer, J. C., and Yaffe, M. B. (2004) Computational prediction of protein-protein interactions. Methods Mol. Biol. 261, 445–468
2. Kim, E., and Sheng, M. (2004) PDZ domain proteins of synapses. Nat. Rev. Neurosci. 5, 771–781
3. Babault, N., Cordier, F., Lafage, M., Cockburn, J., Haouz, A., Prehaud, C., Rey, F. A., Delpeyrre, M., Buc, H., Lafon, M., and Wolff, N. (2011) Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death. Structure 19, 1518–1524
4. Shepherd, T. R., Klaus, S. M., Liu, X., Ramaswamy, S., DeMali, K. A., and Fuentes, E. J. (2010) The Tiam1 PDZ domain couples to Syndecan1 and promotes cell-matrix adhesion. J. Mol. Biol. 398, 730–746
5. Lee, H. J., Wang, N. X., Shi, D. L., and Zheng, J. J. (2009) Sulindac inhibits canonical Wnt signaling by blocking the PDZ domain of the protein Dvl. J. Biol. Chem. 284, 31792–31799
6. Babault, N., Cordier, F., Lafage, M., Cockburn, J., Haouz, A., Prehaud, C., Rey, F. A., Delpeyrre, M., Buc, H., Lafon, M., and Wolff, N. (2011) Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death. Structure 19, 1518–1524
7. Thorsen, T. S., Madsen, K. L., Rebola, N., Rathje, M., Anggono, V., Bach, M., and Madsen, K. L. (2011) A stabilizing influence. CAL PDZ inhibition extends the half-life of the Erbin PDZ domain and the first PDZ domain of ZO-1. Insights into determinants of PDZ domain specificity. J. Biol. Chem. 286, 22312–22320
8. Ernst, A., Sazinsky, S. L., Hui, S., Currell, B., Dharsee, M., Seshagiri, S., Bader, G. D., and Sidhu, S. S. (2009) Rapid evolution of functional complexity in a domain family. Sci. Signal. 2, ra50
9. Alewine, C., Olsen, E., Wade, J. B., and Welling, P. A. (2006) TIP-1 has PDZ scaffold antagonist activity. Mol. Biol. Cell 17, 4200–4211
10. Amacher, J. F., Cushing, P. R., Weiner, J. A., and Madden, D. R. (2011) Crystalization and preliminary diffraction analysis of the CAL PDZ domain in complex with a selective peptide inhibitor. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67, 600–603
11. Cushing, P. R., Fellows, A., Villone, D., Boisguerin, P., and Madden, D. R. (2008) The relative binding affinities of PDZ partners for CFTR. A biochemical basis for efficient endocytic recycling. Biochemistry 47, 10084–10098
12. Sheldrick, G. M. (2008) SADABS, Universität Göttingen, Germany
13. Sheldrick, G. M. (1991) XPREP. Space Group Determination and Reciprocal Space Plots. Siemens Analytical X-ray Instruments, Madison, WI
14. Sheldrick, G. M. (2008) A short history of SHELX. Acta Crystallogr. A 64, 112–122
15. Sheldrick, G. M. (2008) Experimental phasing with SHELX/C/E. Combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. 64, 479–485
16. Collaborative Computational Project, Number 4 (1994) The CCP4 suite. Programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 50, 760–763
17. Madden, D. R. (1992) The Structure of the Human Class I MHC Protein HLA-B2705 and Bound Peptides at 2.1-Å Resolution. Ph.D. thesis, Harvard University, Cambridge, MA
18. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221
19. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr. D Biol. Crystallogr. 63, 32–41
20. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501
21. Chen, V. B., Arendall, W. B., Jr, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity. All-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21
22. Feng, J., Westbrook, J., and Berman, H. (1998) PDB Validation Server, NDB-407. Rutgers University, New Brunswick, NJ
23. Laskowski, R., Marczuk, M., Moss, D., and Thornton, J. (1993) PROCHECK. A program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291
24. Vagoune, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr.
