PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Characterization of PDZ domain-peptide interaction interface based on energetic patterns

Nan Li,[1] Tingjun Hou,[2] Bo Ding,[1] and Wei Wang[1]*

[1] Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093-0359

[2] Institute of Functional Nano and Soft Materials (FUNSOM), Jiangsu Key Laboratory for Carbon-Based Functional Materials and Devices, Soochow University, Suzhou, Jiangsu 215123, People's Republic of China

ABSTRACT

PDZ domain is one of the abundant modular domains that recognize short peptide sequences to mediate protein–protein interactions. To decipher the binding specificity of PDZ domain, we analyzed the interactions between 11 mouse PDZ domains and 2387 peptides using a method called MIEC-SVM, which energetically characterizes the domain-peptide interaction using molecular interaction energy components (MIECs) and predicts binding specificity using support vector machine (SVM). Cross-validation and leave-one-domain-out test showed that the MIEC-SVM using all 44 PDZ-peptide residue pairs at the interaction interface outperformed the sequence-based methods in the literature. A further feature (residue pair) selection procedure illustrated that 16 residue pairs were uninformative to the binding specificity, even though they contributed significantly (∼50%) to the binding energy. If only using the 28 informative residue pairs, the performance of the MIEC-SVM on predicting the PDZ binding specificity was significantly improved. This analysis suggests that the informative and uninformative residue interactions between the PDZ domain and the peptide may represent those contributing to binding specificity and affinity, respectively. We performed additional structural and energetic analyses to shed light on understanding how the PDZ-peptide recognition is established. The success of the MIEC-SVM method on PDZ domains in this study and SH3 domains in our previous studies illustrates its generality on characterizing protein-peptide interactions and understanding protein recognition from a structural and energetic viewpoint.

Proteins 2011; 79:3208–3220.

Key words: protein-peptide interaction; protein binding specificity; molecular interaction energy component (MIEC); support vector machine (SVM); molecular dynamics (MD).

INTRODUCTION

Protein–protein interactions play critical roles in biological processes of every living organism and they are often mediated by weak and transient domain-peptide interaction.[1] These modular domains are involved in many signal transduction processes.[2] PDZ domain is one of the abundant modular domains[3–6] and it plays critical roles in the assembly of subcellular structures, such as neuronal postsynaptic density[7] and epithelial junction,[8] and the maintenance of cell polarity.[9] The biological function of PDZ domains relies on the specific recognition of short peptides from the carboxyl terminus of their substrate proteins. Previous studies have shown that PDZ-binding peptides can be divided into three classes based on the three residues at the carboxyl terminus: class I (X[T/S]XΦCOOH), class II (XΦXΦCOOH), and class III (X[E/D]XΦCOOH), where Φ-hydrophobic residue; X-any residue; T-threonine; S-serine; E-glutamate; D-aspartate.[10–13] Peptides recognized by PDZ domains are about five amino-acid long,[14] while occasionally residues farther away from the carboxyl terminus also contribute to binding specificity.[15]

High-throughput technologies such as phage display[16–19] and protein microarray[20,21] have been exploited to investigate the binding specificity of PDZ domains and to identify the peptide sequence motif recognized by each domain. For example, phage-display experiments on human and worm PDZ domains by Tonikian *et al.*[18] revealed 16 specificity classes, conserved between the two species. Stiffler *et al.*[21] have purified the majority of mouse PDZ domains and conducted protein microarray experiments to determine their binding specificity against 217 peptides taken from the C-terminal of mouse proteins. A principal component analysis on the peptide sequences recognized by these PDZ domains suggested that PDZ-peptide interactions are continuous in the selectivity space, which challenged the previous view of

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant numbers: R01GM085188.

*Correspondence to: Wei Wang, Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0359.
E-mail: wei-wang@ucsd.edu

Received 29 November 2010; Revised 12 July 2011; Accepted 14 July 2011
Published online 4 August 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23157

these interactions falling into discrete classes. A variant of position-specific scoring matrix that emphasizes discrimination between binding and nonbinding peptides was employed to predict PDZ-peptide interactions. A more sophisticated model was trained on the mouse binding data in a follow-up study by Chen *et al.*[22] to predict PDZ-peptide interactions in *D. melanogaster* and *C. elegant*.

These high-throughput experiments, such as microarray and phage-display, have generated significant amount of data to determine the binding preferences of PDZ domains.[16–21] However, interpretation of such data is often solely based on amino acid sequence without structural insight to understand the molecular basis of interaction specificity. Although more and more PDZ-peptide complex structures are being solved,[15,23–29] the number of available complex structures is still a few orders of magnitude less than that of the interacting domain-peptide pairs determined in the high-throughput experiments. Several template complexes shed light on understanding domain-peptide interactions in general, but lack of complex structures for each individual domain hinders illustration of how exactly the binding specificity is achieved. Numerous computational methods have also been developed to predict pairwise PDZ-peptide interactions[21,22,30,31] or peptide sequence preference of individual PDZ domains.[32] Most of these methods rely on peptide sequence information such as MDSM,[21,22,31] or amino acid trigram,[30] with limited structural information and/or lack of energetic characterization of PDZ-peptide recognition.[22,31]

We recently developed a computational method based on molecular interaction energy components (MIECs)[33–35] that aims to decipher and predict binding specificity between protein domains and peptides. In this method, thousands of domain-peptide complexes are modeled based on template structures and then optimized using molecular mechanics. Energetic interaction profile (MIECs) of each domain-peptide complex is obtained from free energy calculation, which includes van der Waals, electrostatic, and desolvation energy components for interacting residues on the binding interface. Combined with either regression method (e.g., partial least squares, PLS) or classification method (e.g., support vector machine, SVM), MIEC-PLS or MIEC-SVM can predict binding affinity of a domain-peptide pair or classify peptides to binder/nonbinder category for a given domain. In the previous studies, these methods have been successfully applied to the SH3-peptide recognition[33,34] and HIV protease-small molecule interactions.[35]

In this study, we applied MIEC-SVM method to PDZ-peptide interactions. The main purpose of this study is to obtain a generic MIEC-SVM model that can discriminate binder and nonbinder peptides for PDZ domains and quantitatively reveal the molecular basis of the binding specificity from the SVM-MIEC model. To achieve this goal, 2387 PDZ-peptide complexes were modeled. Their MIECs were calculated and used in generating the MIEC-SVM model. The initial cross-validation and leave-one-domain-out (LODO) test showed that the MIEC-SVM model outperformed the previous MDSM-based method.[21,22] A follow-up feature selection procedure, that is, selecting the MIECs most predictive of binding specificity, further boosted the performance of our model. This procedure also revealed that the protein-peptide residue interaction pairs at the binding interface fall into informative and uninformative categories. Further analysis on residue pairs showed that the energy distribution between the informative and uninformative pairs was biased, which suggests that these two categories may play different roles in the PDZ-peptide recognition.

MATERIALS AND METHODS

PDZ domains and binding peptides

In the study of Stiffler *et al.*,[21] 85 PDZ domains were observed to interact with at least one peptide, among which 12 domains have available structures in the PDB database.[36] Four out of 12 domains, PTP-BL(2/5) (1VJ6),[26] MAGI-1(2/6) (2I04),[25] α1-syntrophin(1/1) (2PDZ),[27] and TIP1(1/1) (3DIW),[28] have complex structures with peptides. The rest eight domains, PAR-6B(1/1) (1NF3),[37] Harmonin(2/3) (1V6B), PAR3B(1/3) (1WG6), RGS3(1/1) (1WHD), PDZK11(1/1) (1WI2), ZO-2(1/3) (2CSJ), NHERF-3(1/4) (2D90), and NHERF-3(3/4) (2EDZ), only have domain (PDZ-only) structure without binding peptides (Supporting Information Table S1). PDZK11(1/1) (PDB code: 1WI2) was left out in the free energy calculations because the molecular dynamics simulation could not reach equilibrium, as manifested by significant RMSD fluctuation (data not shown).

The binding peptides for the above 11 PDZ domains were taken from the study of Stiffler *et al.*,[21] in which total 217 peptides (Supporting Information Table S2) were tested in the protein microarray experiments. Eighty-six PDZ-peptide interactions observed between the 217 peptides and the 11 domains were used as the binder (positive) set (Supporting Information Table S3) to train the MIEC-SVM model. For each binding peptide, six nonbinding peptides were randomly selected to form the nonbinder (negative) set that contained 516 noninteractions. The ratio of 6 between nonbinder and binder was chosen to be consistent with that in the test set of Stiffler *et al.*'s data.[21]

Modeling PDZ-peptide complex structures

The structures of the 11 PDZ domains were first preprocessed as follows. For X-ray structures, water and ions

were removed and only one PDZ chain was kept in the cases that multiple chains were present. For NMR structures, the α carbon RMSD was calculated between each structure and the average structure; then the NMR structure with the smallest RMSD to the average structure was selected for the following analysis.

Seven PDZ domains, 1NF3, 1V6B, 1WG6, 1WHD, 2CSJ, 2D90, and 2EDZ do not have available complex structures in the PDB database. Their complex structures were built by superimposing these PDZ-only structures to other PDZ-peptide complexes and acquired the peptide coordinates from the complex structure (Supporting Information Table S1). Six PDZ-peptide complex structures[15,23–25,28,29] (Supporting Information Table S4) were selected from PDB as the templates, because these structures were aligned best to the query PDZ proteins indicated by a minimal Cα RMSD (Supporting Information Table S1) and a maximum number of residues aligned. The structure superimposition was performed by Align Structure module in Discovery Studio.[38] Because the peptide in the template complex may not be a binder to the query PDZ domain, the peptide was then mutated using *SCWRL4*[39] to the sequence in the 217 peptides tested in Stiffler *et al.*'s article[21] that had the strongest binding affinity to the query PDZ domain.

The 11 PDZ complexes were optimized by molecular dynamics simulations using AMBER10[40] and the AMBER03 force field.[41] First, *tleap* was used to process the structure and generate the AMBER topology and coordinate files. Missing atoms and hydrogen atoms were added by tleap. The structure was solvated in a rectangle box of TIP3P water[42] extended 12 Å away from any solute atom. Counter ions of Na^+ or Cl^-were placed in a grid surrounding the complex structure based on the Columbic potential to neutralize the system. Second, 5000 steps of energy minimization were performed to relax the system, consisting of 1000 steps of steepest descent followed by 4000 steps of conjugate gradient minimization. Third, after the minimization, molecular dynamics simulation was performed using program *sander*. The first 20 ps heated the system from 10 to 300 K with a constant volume periodic boundary; the time step was set to 0.5 fs. Subsequently, another 20-ps simulation was performed to adjust the density of water molecules using a constant pressure periodic boundary; SHAKE[43] was applied to constrain bonds involving hydrogen atoms; the time step was set to 1 fs. Next, 2-ns simulation was performed for system equilibration and data collection; the constant pressure periodic boundary and SHAKE procedure were applied; the time step was 1 fs. Throughout the simulation, Particle Mesh Ewald[44] was employed to calculate long-range electrostatics interactions and the weak-coupling algorithm was applied to maintain the temperature of the system. The last snapshot of the simulation was relaxed by 5000 steps of energy minimization. The minimized conformation was saved as the complex template for modeling other PDZ-peptide interactions.

To model interactions between the 11 PDZ domains and other peptides, we mutated the peptide in the template complexes using *SCWRL4*[39] to other peptides in the positive and negative sets. Restrained by the computational cost, only 5000 steps of energy minimization were performed to optimize each complex. The first 1000 steps used steepest descent method and the rest used conjugate gradient method. TIP3P water[42] was added to solvate the system.

Energy decomposition analysis

Energy decomposition was performed on the minimized PDZ-peptide complexes using *mm_pbsa* in the AMBER package. For each residue pair between the PDZ domain and the peptide, its interaction energy was decomposed into: van der Waals interaction energy ΔE_{vdw}; electrostatic interaction energy ΔE_{ele}; polar contribution to the desolvation energy ΔE_{gb}; nonpolar contribution to the desolvation energy ΔE_{sa}. The dielectric constant of 1 was used to calculate ΔE_{ele}. ΔE_{gb} was calculated using the GB model with parameters developed by Onufriev *et al.*[45,46] The interior and the exterior dielectric constants in the GB calculation were set to 1 and 80, respectively. The ΔE_{sa} was estimated according to the solvent accessible surface area (SASA) as $\Delta E_{sa} = 0.0072 \times$ SASA.

The MIEC profile and the MIEC-SVM model

The MIECs (Fig. 1) were used to capture the energetic characteristics of interactions between peptides and a specific domain family.[33–35] Here, to construct a generic model for all the PDZ domains, we needed to identify a set of interacting domain-peptide residue pairs at the interface that may determine the binding specificity. Sequence alignment for the 11 PDZ domains (see Fig. 2) was generated by Clustal W.[47] The result from Clustal W was manually adjusted based on the structure to ensure correct alignment of the PDZ binding region,[14] which contains the carboxylate-binding GLGF loop (residue 39–42), β2 (residue 43–47), and α2 (residue 105–113).[14]

The interaction residue pairs were defined by a 6-Å distance cutoff between any two residues at the domain-peptide interface, and total 44 domain-peptide residue pairs were so identified. In this study, we also included residue pairs between adjacent peptide residues in the positions from 0 to −4 and these interactions reflected the conformational preference of the peptide. Therefore, there were 48 interaction residue pairs in total. For each interaction residue pair, ΔE_{vdw}, ΔE_{ele}, ΔE_{gb}, and ΔE_{sa} were calculated using energy decomposition analysis. The number of MIECs for one domain-peptide was $(44 + 4) \times 4 = 192$.

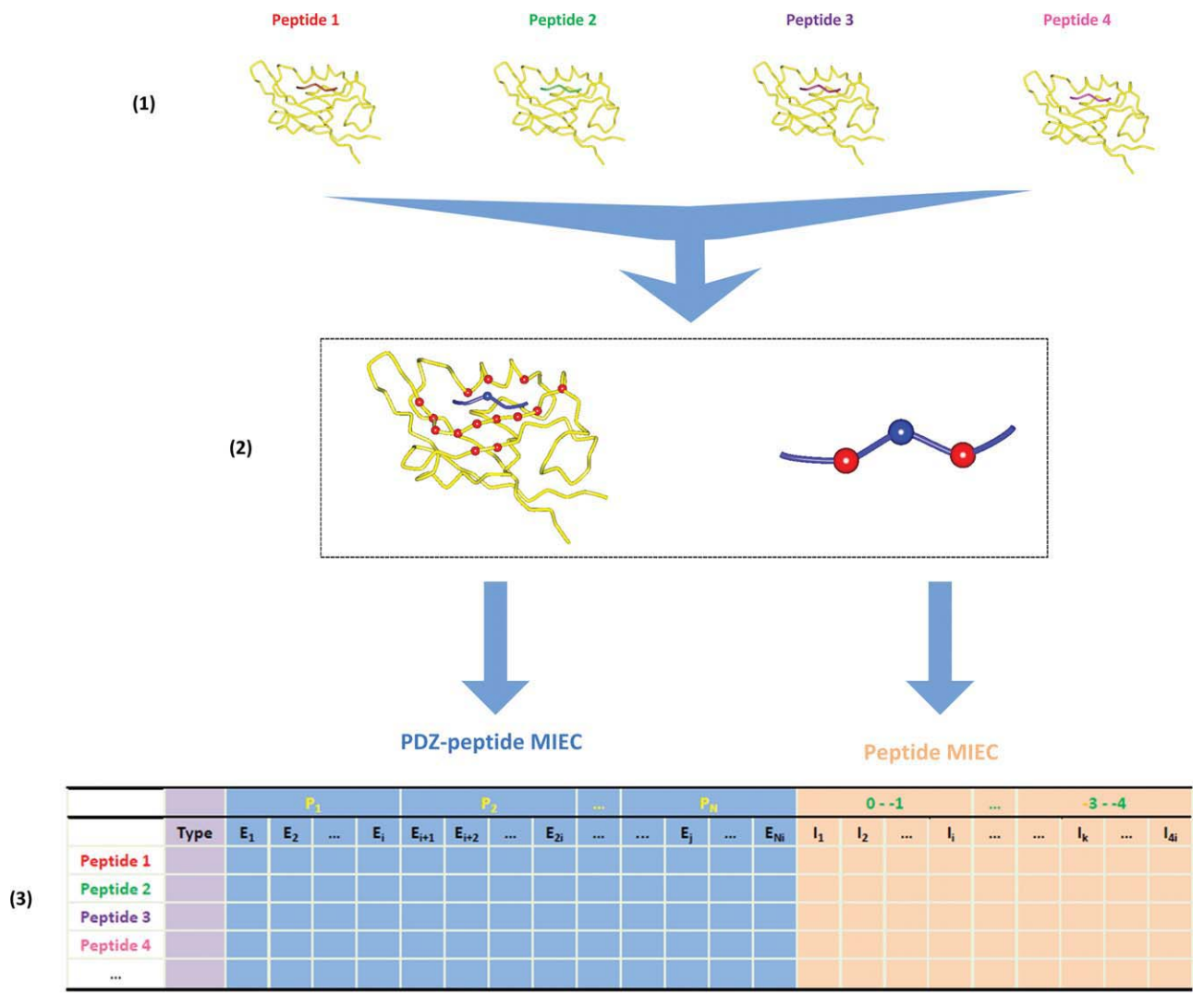

Figure 1

The MIEC-SVM method. (1) The complex structures formed between PDZ and different peptides were modeled by homology modeling and MD simulations. (2) Residues close to the interaction interface were selected. PDZ residues: red balls. Peptide residues: blue balls. To consider the conformational preference of peptides, the interactions of adjacent peptide residues were also calculated. (3) Molecular interaction energy components (MIECs) were computed for all binders and nonbinders by energy decomposition analysis. The fully filled MIEC matrix was then used to train a SVM model to classify peptides into binders or nonbinders. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

SVM was chosen to classify peptides to binders/nonbinders based on the MIECs. In this study, the software package LIB-SVM[48] was used. We have conducted three tests to assess the performance of the MIEC-SVM models. First, 500 runs of threefold cross validation were performed on all the data to select the best-performed MIECs and assess the contributions of individual positions (leave-one-site-out test). Second, 500 runs of fivefold cross validation were used to compare the prediction accuracy of our model with the Stiffler *et al.*'s study.[21] Third, 500 runs of LODO cross validation were carried out to assess the generality of the MIEC-SVM model.

The prediction accuracy of our models were evaluated using sensitivity (SEN), specificity (SPC), accuracy for positive (ACC+), accuracy for negative (ACC−), and Matthews correlation coefficient (MCC).

$$\mathrm{SEN} = \frac{\mathrm{TP}}{TP+FN},$$
$$\mathrm{SPC} = \frac{\mathrm{TN}}{TN+FP},$$
$$\mathrm{ACC+} = \frac{\mathrm{TP}}{TP+FP},$$
$$\mathrm{ACC-} = \frac{\mathrm{TN}}{TN+FN},$$
$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}}$$

TP, FN, TN, and FP represent the count of true positive, false negative, true negative, and false positive, respectively.

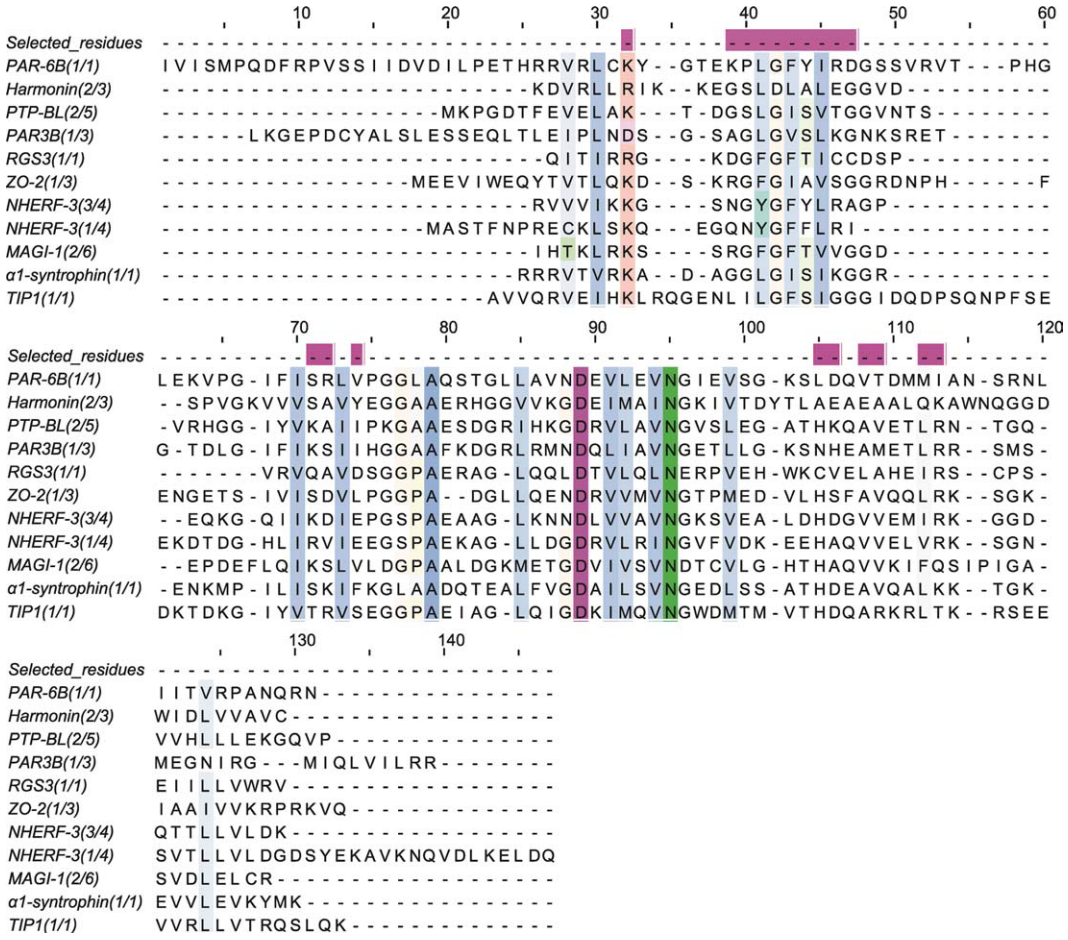

Figure 2

Multiple sequence alignment for the 11 mouse PDZ domains. The multiple sequence alignment was generated by Clustal W with manual adjustment. The alignment is colored using the Clustal X coloring scheme. The 19 selected residue positions in our MIEC study are colored in purple in the *Selected_residues* row. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

RESULTS AND DISCUSSION

Modeling of PDZ-peptide complexes and calculation of MIECs

The structures of the 11 PDZ domains used in this study were obtained from PDB database. Complex structures were built for the seven PDZ domains whose complex structures were not available. All 11 complex structures were then optimized by molecular dynamics. For each optimized PDZ-peptide structure, virtual mutagenesis was performed to mutate the peptide in the optimized structure to one of the 217 peptides, which generated 2387 PDZ-peptide complex structures. These complex structures were then optimized by energy minimization.

MIECs between PDZ-peptide residue pairs at the interaction interface, including van der Waals, electrostatic, nonpolar and polar part of desolvation, were calculated to capture the energetic characteristics of the binding. To select the MIECs for the best prediction of binding specificity, we tried numerous combinations of MIECs. The performance of each combination was evaluated by 500 runs of threefold cross-validations with random partition of binders and nonbinders to training and test sets (Table I, Supporting Information Table S5). Based on the previous studies[33,34] and the test for PDZ-peptide interactions (data not shown), RBF kernel was chosen for SVM because it performed better than the other kernel functions in cross-validations.

Cross-validations were used to select the best combination of MIECs, consisted of van der Waals, electrostatics, polar part of desolvation, nonpolar part of desolvation, and the peptide internal interactions (model 12 in Table I). The sensitivity, specificity, and MCC for this combination were 59.6, 88.5, and 0.438%, respectively (Table I). Note that inclusion of peptide internal interactions is only an empirical option that indirectly accounts for conformational preference of the peptide. As shown in Table I, adding this term only moderately improved the

Table I
Selection of MIEC Components in Predicting PDZ Binding Specificity

Model	MIEC component[a]	SEN	SPC	ACC+	ACC-	MCC
	PDZ-peptide interactions only					
1	ΔE_{vdw}, ΔE_{ele}	0.537	0.883	0.445	0.920	0.390
2	ΔE_{vdw}, ΔE_{ele}, ΔE_{gb}	0.542	0.891	0.463	0.921	0.406
3	ΔE_{vdw}, ΔE_{ele}, ΔE_{sa}	0.576	0.885	0.464	0.926	0.422
4	ΔE_{vdw}, ΔE_{polar}, ΔE_{sa}	0.558	0.897	0.482	0.924	0.428
5	ΔE_{vdw}, ΔE_{ele}, ΔE_{desol}	0.548	0.887	0.458	0.922	0.405
6	ΔE_{vdw}, ΔE_{ele}, ΔE_{gb}, ΔE_{sa}	0.575	0.890	0.475	0.926	0.431
	PDZ-peptide interactions plus intrapeptide (adjacent residue pair) interactions					
7	ΔE_{vdw}, ΔE_{ele}	0.583	0.876	0.449	0.927	0.415
8	ΔE_{vdw}, ΔE_{ele}, ΔE_{gb}	0.584	0.876	0.450	0.927	0.414
9	ΔE_{vdw}, ΔE_{ele}, ΔE_{sa}	0.602	0.881	0.465	0.930	0.436
10	ΔE_{vdw}, ΔE_{polar}, ΔE_{sa}	0.555	0.897	0.483	0.924	0.428
11	ΔE_{vdw}, ΔE_{ele}, ΔE_{desol}	0.572	0.884	0.462	0.925	0.419
12	ΔE_{vdw}, ΔE_{ele}, ΔE_{gb}, ΔE_{sa}	0.596	0.885	0.472	0.929	0.438

[a] ΔE_{vdw}, ΔE_{ele}, ΔE_{gb}, and ΔE_{sa} are van der Waals, electrostatic, polar part of desolvation, and nonpolar part of desolvation energies, respectively. ΔE_{polar} is the sum of ΔE_{ele} and ΔE_{gb}. ΔE_{desol} is the sum of ΔE_{gb} and ΔE_{sa}.

performance of the model (models 6 and 12, MCC increased from 0.431 to 0.438). This is most likely because the peptides binding to PDZ domains are short (five residue-long) and do not form specific secondary structure. In contrast, for the SH3 domains, including this term in the model improved the performance more significantly (models 6 and 12 in Table I of Ref. [34] MCC increased from 0.374 to 0.523), which may be due to the left-handed ployproline type II helix formed by the peptides binding to the SH3 domain. It is also worth pointing out that the dataset for cross-validation was unbalanced because nonbinders were six times of binders. Therefore, different weights were assigned to positive and negative classes in SVM. In the current study, weight $k_+ = 4$ was used for the binder class and $k_- = 1$ was remained for the nonbinder class. The optimal k_+ value was selected based on a systematic cross-validation test using k_+ from 1 to 30 (Supporting Information Figure S1). For the LODO test (see next section), k_+ was selected as four by cross-validations on all 11 LODO datasets (Supporting Information Figure S7), which was consistent with the k_+ value selected from the entire dataset.

Generality of the MIEC-SVM model

To assess the generality of the MIEC-SVM model on predicting the binding specificity of PDZ-peptide interactions, we conducted a LODO cross validation. In this test, the interaction data associated with one domain were completely removed from the training set and used as the test set for the model trained on the remaining data. This test mimics the situation in reality to predict the binding specificity of a new PDZ domain that is not included in training. In total, 11 LODO cross validations were performed (Table II, Supporting Information Table S5). The average sensitivity, specificity, and MCC were 63.7, 78.9, and 0.343%, respectively. The average accuracies for the binders and nonbinders were 33.7 and 93.2%, respectively. Compared with our previous work of SH3 domains,[34] the MIEC-SVM model performed slightly worse, which may be due to the much smaller training data available for each PDZ domain (Supporting Information Table S1). On average, the binders and nonbinders for each PDZ domain in this study were 8 and 48, respectively. As a comparison, each SH3 domain in our previous study[34] had 27 and 540 binders and nonbinders, respectively. A larger training dataset no doubt helps characterize the domain-peptide binding specificity better.

Important residues for PDZ-peptide specific binding

To identify residues important for the binding specificity of PDZ-peptide interactions, we conducted leave-one-residue-out (LORO) tests. If leaving out a residue significantly deteriorated the performance of the MIEC-SVM model, we considered this residue important for binding specificity. In each test, one of the 19 PDZ and five peptide residues involved in calculating MIECs was excluded, that is, all the interactions involving this residue were removed from the MIEC matrix. Then cross-validations were performed on the new MIEC-SVM model to assess the importance of the left-out residue. The difference between the MCC of the LORO model and that of the default using all the 24 residues reflected the contribution of the left-out residue to the prediction accuracy.

Not surprisingly, this LORO test showed that four peptide residues at the C-terminal are important for the binding specificity [their MCC differences >0.02, Fig. 3(A) and Supporting Information Table S6]. Indeed, these four residues fit to the binding pocket of the PDZ

Table II
The MIEC-SVM Performance Evaluated by the Leave-One-Domain-Out Cross Validation

Domain name (PDB ID)	Binders[a]	SEN[b]	SPC[c]	ACC+[d]	ACC−[e]	MCC[f]
PAR-6B(1/1) (1NF3)	2	0.995	0.866	0.610	0.999	0.720
NHERF-3(3/4) (2D90)	7	0.939	0.897	0.624	0.989	0.714
MAGI-1(2/6) (2I04)	2	0.880	0.762	0.409	0.979	0.496
NHERF-3(1/4) (2EDZ)	11	0.844	0.782	0.406	0.969	0.483
TIP1(1/1) (3DIW)	6	0.602	0.845	0.418	0.928	0.389
α1-syntrophin(1/1) (2PDZ)	6	0.730	0.764	0.350	0.945	0.381
PAR3B(1/3) (1WG6)	3	0.947	0.574	0.282	0.986	0.373
PTP-BL(2/5) (1VJ6)	20	0.762	0.544	0.218	0.933	0.215
RGS3 (1/1) (1WHD)	14	0.537	0.537	0.163	0.874	0.052
ZO-2(1/3) (2CSJ)	13	0.043	0.954	0.132	0.857	−0.005
Harmonin(2/3) (1V6B)	2	0.000	0.970	0.000	0.853	−0.038
Average	8	0.637	0.789	0.337	0.932	0.343

[a] Number of binders for each PDZ domain.
[b] Sensitivity.
[c] Specificity.
[d] Prediction accuracy for binders.
[e] Prediction accuracy for nonbinders.
[f] Matthews correlation coefficient.

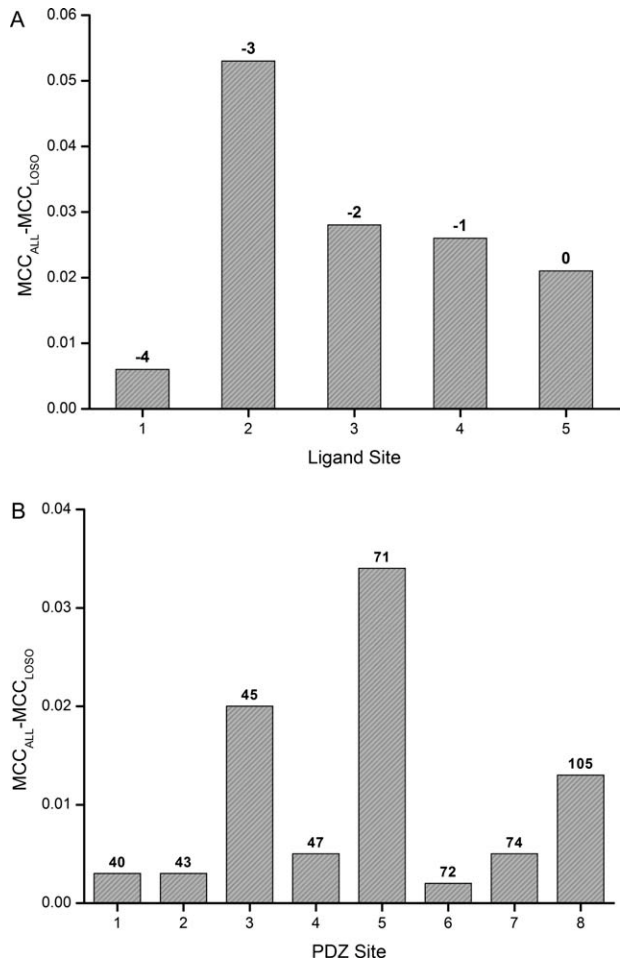

Figure 3

Leave-one-residue-out (LORO) test results for (**A**) peptide residues; (**B**) PDZ residues. The PDZ residue numbers are taken from the multiple sequence alignment (see Fig. 2).

domain and establish the peptide binding specificity [Fig. 4(A,B)], which is consistent with the knowledge that residues at the carboxyl terminal of a peptide are important for PDZ binding.[14,18]

Three residues on the PDZ domain were identified as important for the binding specificity. Their residue numbers in the sequence alignment were 45, 71, and 105 and their MCC differences were 0.020, 0.034, and 0.013, respectively [Fig. 3(B), Supporting Information Table S6]. The structural roles of these three residues in the PDZ-peptide binding are illustrated in Figure 4(C,D). In PDZ domains, residue 45 is often one of the three hydrophobic amino acids: Ile, Leu, and Val (see Fig. 2). It is part of the hydrophobic pocket to stabilize the side chain of peptide residue 0. In addition, the hydrogen bonds between residue 45 and peptide residue −1 are crucial for the formation of the well-known antiparallel sheet between β2 of the PDZ domain and the peptide. Residue 71 is located in β3 strand and tends to interact with the side chain of residue −3 of the peptide[23,24,26,28,29] through either electrostatic[28,29] or hydrophobic[23,26] interactions, which is consistent with the LORO test results. Residue 105 is in the α2 of PDZ domain and is often a histidine (7 out of 11, Fig. 2). The hydrogen bonds formed between residue 105 and Thr/Ser on class I peptide −2 position was known critical to distinguish class I[14,24,26,27,29] and class II peptide binding,[15] which supports the important role of residue 105 in determining binding specificity.

Feature (residue pair) selection to boost the model performance

LORO test revealed that not all the residue pairs were informative for the discrimination of binders and non-binders (Supporting Information Table S6). The initial residue pairs were selected based on the distance between the residues of the PDZ domain and the peptide, that is, any PDZ-peptide residue pair closer than 6 Å was included in the MIEC calculation. To systematically remove noisy pairs from the model, we conducted feature selection using the sequential backward selection method.[50] First, SVM model was trained using all residue pairs. The MCCs for the cross validation and LODO tests were recorded. Second, residue pairs were left out from the MIEC calculation one at a time, and cross validation and LODO tests were performed using the new MIECs. We removed all the residue pairs whose removal improved the performance of the MIEC-SVM model in both cross-validation and LODO tests, as measured by MCC change, that is, MCC_{LORO}-MCC_{ALL} larger than 0 in both cross-validation and LODO tests.

As a result, 16 residue pairs were removed and 28 were kept (Supporting Information Table S7). This feature selection procedure was fast and straightforward, but significantly improved the prediction accuracy. The MCCs for threefold cross validation and LODO test were boosted to 0.487 and 0.426 (Tables III and IV, Supporting Information Table S5), compared to 0.438 and 0.343 without feature selection, respectively. Especially, significant improvement on ZO-2(1/3) (2CSJ) was achieved by feature selection. Our analysis illustrated the importance of feature selection, which also sheds lights on understanding the PDZ-peptide binding specificity (see below).

We noticed that the PDZ domain of Harmonin(2/3) (1V6B) did not perform as well as other domains. We found that the binding site of uncomplexed Harmonin(2/3) (1V6B) is quite different from the other 10 PDZ domains: the β2 strand and α2 helix are much closer to each other in Harmonin(2/3) than in other proteins; GLGF motif region takes a flat conformation in Harmonin(2/3) but not in other PDZs. It is likely that the Harmonin(2/3) (1V6B) PDZ domain may undergo substantial conformational change upon peptide binding, which requires molecular dynamics longer than the one we conducted to model the complex structure.

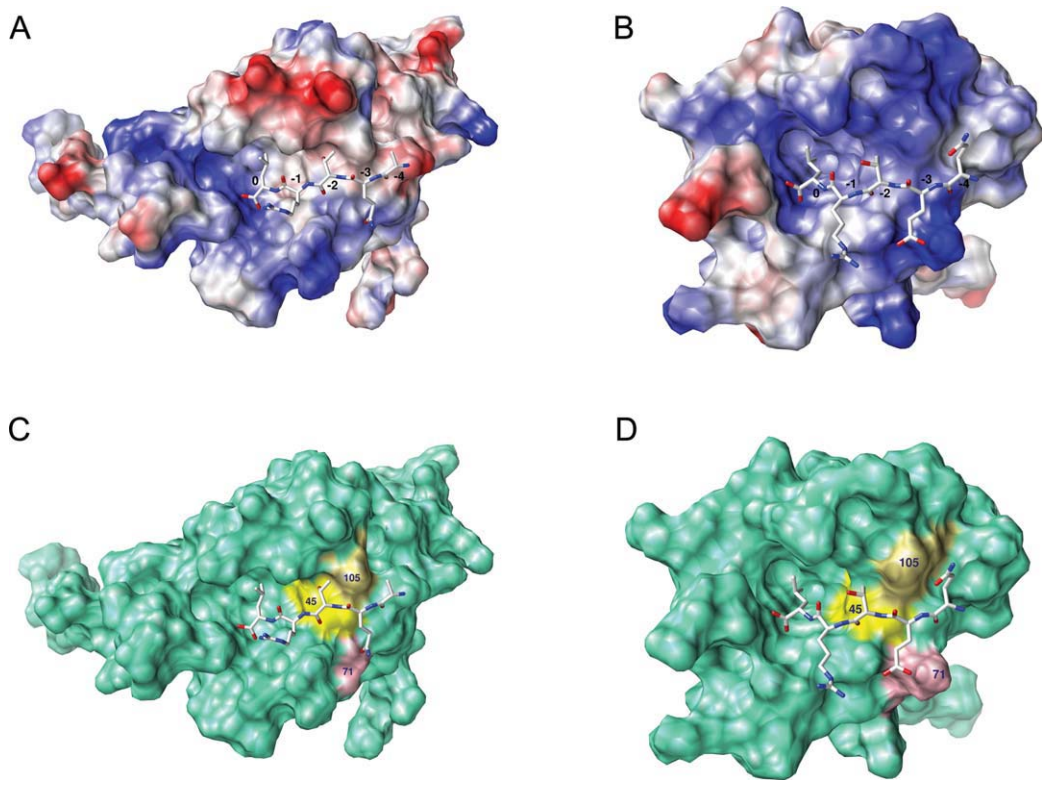

Figure 4

Important residues for the PDZ-peptide interactions. Peptide residues shown in the complex structure of (**A**) PAR-6B(1/1) and (**B**) PTP-BL(2/5). PDZ residue numbers are taken from the multiple sequence alignment. PDZ surface is colored by electrostatic potential and peptide is shown in stick. The important PDZ residues shown in the complex structure of (**C**) PAR-6B(1/1) and (**D**) PTP-BL(2/5). All figures were generated using UCSF Chimera 1.4.1.[49]

To further test the generality of the MIEC-SVM model, we modeled two additional mouse PDZ domains that were arbitrarily selected from the mouse proteome and not included in the initial dataset: PSD95-2 and CHP110-2. We predicted their binding specificity using the model trained from the 11 PDZ domains. The MCCs for the two domains are 0.505 and 0.42, respectively (Table V). Note that the maximal sequence similarity of both PSD95-2 and CHP110-2 are 68.8% to α1-syntrophin (1/1), but they only share less than 1/3 of binding peptides [4 out of 12 between PSD95-2 and α1-syntrophin (1/1) and 5 out of 18 for CHP110-2]. Therefore, the satisfactory performance of the model further validated the generality of our model.

Table III

The MIEC-SVM Performance Evaluated by Cross Validation Before and After Feature Selection

	SEN[a]	SPC[b]	ACC+[c]	ACC−[d]	MCC[e]
No feature selection	0.596	0.885	0.472	0.929	0.438
After feature selection	0.648	0.890	0.504	0.938	0.487
Chen et al.'s model	0.605	0.948	0.359	0.980	0.433

[a]Sensitivity.
[b]Specificity.
[c]Prediction accuracy for binders.
[d]Prediction accuracy for nonbinders.
[e]Matthews correlation coefficient.

Table IV

The MIEC-SVM Performance Evaluated by Leave-One-Domain-Out (LODO) Test After Feature Selection

Domain name (PDB ID)	Binders[a]	SEN[b]	SPC[c]	ACC+[d]	ACC−[e]	MCC[f]
PAR-6B(1/1) (1NF3)	2	1.000	0.849	0.578	1.000	0.697
NHERF-3(3/4) (2D90)	7	1.000	0.949	0.786	1.000	0.862
MAGI-1(2/6) (2I04)	2	0.950	0.695	0.377	0.990	0.484
NHERF-3(1/4) (2EDZ)	11	0.762	0.877	0.522	0.959	0.551
TIP1(1/1) (3DIW)	6	0.790	0.757	0.368	0.956	0.420
α1-syntrophin(1/1) (2PDZ)	6	0.755	0.816	0.422	0.953	0.461
PAR3B(1/3) (1WG6)	3	0.823	0.772	0.406	0.964	0.466
PTP-BL(2/5) (1VJ6)	20	0.823	0.630	0.273	0.956	0.321
RGS3 (1/1) (1WHD)	14	0.659	0.521	0.188	0.901	0.127
ZO-2(1/3) (2CSJ)	13	0.479	0.887	0.414	0.912	0.342
Harmonin(2/3) (1V6B)	2	0.000	0.962	0.000	0.852	−0.047
Average	8	0.731	0.792	0.394	0.949	0.426

[a]Number of binders for each PDZ domain.
[b]Sensitivity.
[c]Specificity.
[d]Prediction accuracy for binders.
[e]Prediction accuracy for nonbinders.
[f]Matthews correlation coefficient.

Table V
Prediction Results of Two Additional Mouse PDZ Domains

Domain name	Binders[a]	SEN[b]	SPC[c]	ACC+[d]	ACC−[e]	MCC[f]
PSD-95(2/3)	12	0.438	0.967	0.726	0.912	0.505
CHP110(2/3)	18	0.490	0.922	0.516	0.916	0.420

[a]Number of binders for each PDZ domain.
[b]Sensitivity.
[c]Specificity.
[d]Prediction accuracy for binders.
[e]Prediction accuracy for nonbinders.
[f]Matthews correlation coefficient.

Comparison with other methods

We compared our method with the study of Stiffler *et al.*[21] Because Stiffler *et al.* used a sequence-based method, they were able to train their model on 16,058 interactions, and tested it on 3552 interactions. Essentially, they conducted a cross-validation of roughly fivefold (the ratio between their training set and the test set was 4.52) because they trained their model on the same PDZ domains used in the test set against different peptides. The sensitivity (48%), specificity (88%), and accuracy for the positive class (38.5% calculated from their reported TP/FP = 0.63) obtained from such cross-validation were satisfactory.[21]

For a comparison, we performed 500 runs of fivefold cross-validations with random partitions of binders and nonbinders. The sensitivity, specificity, and accuracy for positive interactions were 63.2, 88.7, and 50.0%, respectively, which all outperformed the model of Stiffler *et al.*[21] This comparison suggests that the MIEC-SVM model is able to capture the energetic patterns of the interaction interface of PDZ-peptides.

Following Stiffler *et al.*'s work, Chen *et al.*[22] developed a modified sequence-based model combined with the structural information from one selected PDZ domain. We run Chen *et al.*'s model on the 11 PDZ domains in our dataset. For a fair comparison, the ratio of binder and nonbinder was kept as 6. The prediction was repeated 500 times with nonbinder randomly sampled each time. Without feature selection, our MIEC-SVM model had comparable MCC (0.438) as Chen *et al.*'s model (0.433). With feature selection, our MIEC-SVM model outperformed (MCC = 0.487) the Chen *et al.*'s model (Table III). It is worth noting that the interactions between the 11 PDZ and the 217 peptides were included to train the Chen *et al.*'s model.[22]

Sequence and binding specificity similarity of the PDZ domains in the dataset

It is critical to confirm that the generalization capability of our model was not due to presence of similar PDZ domains in our dataset. We checked the similarity of the 11 PDZ domains in terms of their interface sequence (residues involved in the interactions) similarity, full sequence similarity (Fig. 5 and Supporting Information Table S8), structural (RMSD of interface) similarity (Supporting Information Table S8), and binding specificity similarity (overlaps of binding peptides, Fig. 7).

Interface sequence similarity is shown in Supporting Information Table S8 and Figure 5. Sequence similarity is calculated as the percentage of identical residues in the sequence. Interface sequence contains 16 PDZ residues that form 28 informative pairs with five peptide residues. The number of identical residues was obtained from the multiple sequence alignment. The interface sequence similarity is moderate: the average value is 27.7%. Note that the interface contains several conserved residues in most of the PDZ domains (whole GLGF loop, whole β2, and His at the beginning of the α2). As a comparison, the 82 PDZ domains in the Chen *et al.*'s dataset have an average interface sequence similarity of 38.3% (see Fig. 5). We also performed *t*-test for the two distributions of interface sequence similarity. The *P*-value was 1.34e-7, which showed that the means of the two distributions were significantly different. Therefore, the 11 PDZ domains in our dataset have even lower pairwise similarity in the binding sites than all the 82 PDZ domains in the Chen *et al.*'s dataset. The maximal sequence similarity distribution between our dataset and Chen *et al.*'s dataset was also compared (see Fig. 6). The maximal similarity of PDZ domains in our dataset is significantly smaller than those in Chen *et al.*'s study.

The full sequence and RMSD similarity are shown in Supporting Information Table S8. The average full sequence similarity in our dataset is 23.1%, lower than the interface sequence similarity. The RMSD similarity between PDZs was calculated between interface residues on the minimized PDZ structures after the MD simulation. The average RMSD similarity in our dataset is 1.346 Å.

The specificity similarity between two PDZs was calculated using the number of shared binders. As a comparison,

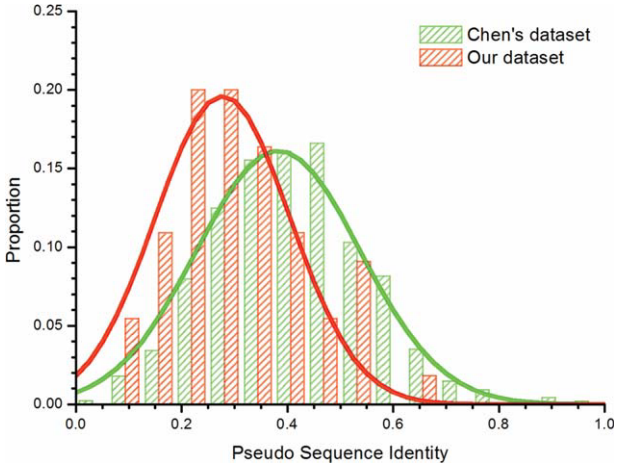

Figure 5
Interface sequence similarity. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

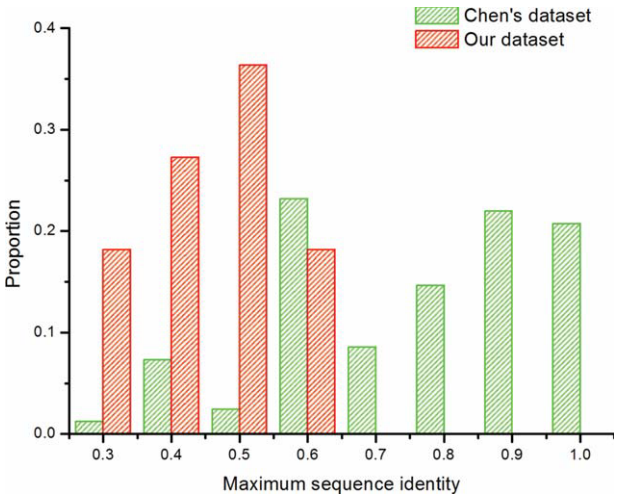

Figure 6

Maximal sequence identity between PDZ domains in our and Chen *et al.*'s datasets. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the average specificity similarity in our dataset and Chen *et al.*'s dataset are 0.618 and 1.332, respectively. We also calculated the percentage of overlapping peptides among the nonredundant binders of any pair of PDZ domains and showed the distribution in Figure 7. It is clear that 11 PDZ domains in our study have lower binding specificity similarity than the 82 PDZ domains in Chen *et al.*'s study.

Insights into PDZ-peptide interaction from MIEC and residue pair selection

We further investigated the MIECs of informative and uninformative PDZ-peptide residue pairs. The average

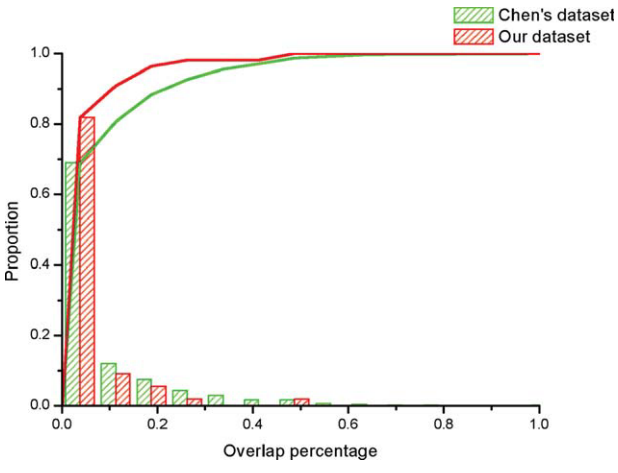

Figure 7

Binding specificity similarity. The bars show the portion of overlapping peptides and the curves are the accumulated counts. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

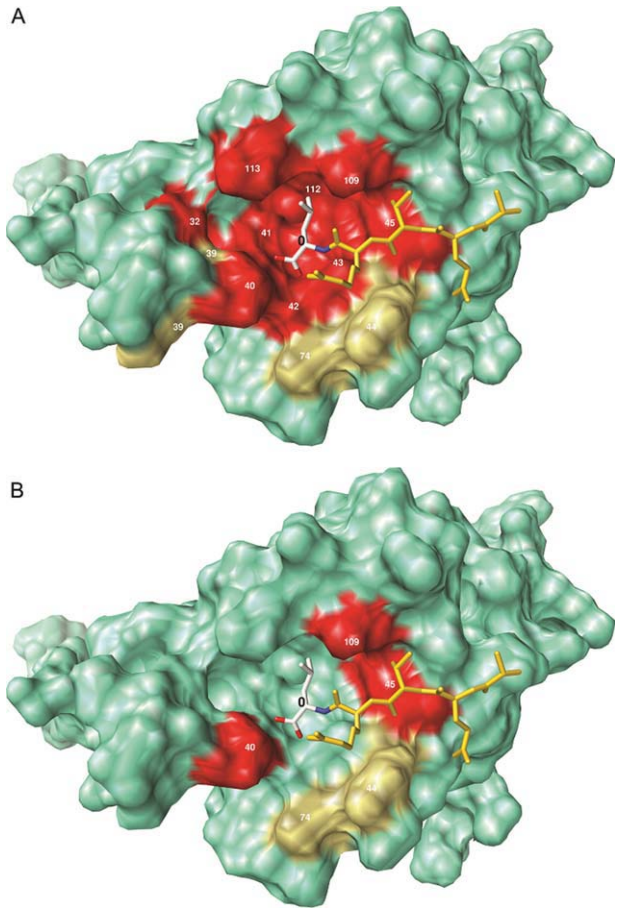

Figure 8

PDZ-peptide residue pairs involving peptide residue 0 (shown on the PAR-6B(1/1) structure). (**A**) PDZ residues interacting with peptide residue 0. (**B**) PDZ residues forming informative residue pairs with peptide residue 0. PDZ residue numbers are taken from the multiple sequence alignment. Residues interacting strongly (<-1.0 kcal/mol) and weakly (≥ -1.0 kcal/mol) with peptide residue 0 are colored in red and yellow, respectively. Figures were generated using UCSF Chimera 1.4.1.[49] [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

MIECs of all the 86 PDZ-peptide interacting complexes (Supporting Information Table S7) were calculated for all the 44 PDZ-peptide residue pairs. Obviously, residue pairs did not contribute equally to the binding energy. We found that 22 pairs with the most favorable interaction energies contributed about 90% of the total binding energy. The 28 informative pairs contributed 53% to the total binding energy compared to 47% contributed by the 16 uninformative ones. For all the peptide positions, residue 0 alone contributed about 45% of the total binding energy, while residue -1, -2, -3, -4 contributed 13, 19, 14, and 9%, respectively.

Peptide residue 0 was involved in 13 interaction pairs [Fig. 8(A)]. Although it contributes significantly to the total binding energy, most of its residue pairs (7 out of 13) were uninformative (Supporting Information Table S7). It is owing to many nonspecific backbone interactions

formed between peptide residue 0 and PDZ residues. Note that seven uninformative pairs contribute significantly to the total binding energy ($\Delta G_{total} < -1.0$ kcal/mol) and their total contribution consists about 33% of the average total energy, compared to 12% contributed by the six informative pairs. Specifically, PDZ residues 40, 41, 42, and 43, known as GLGF motif[13,14] or carboxylate-binding loop,[14] interact strongly with peptide residue 0, manifested by the MIECs of these four residue pairs: large polar ($\Delta G_{polar} < -1.5$ kcal/mol) and nonpolar interaction energies ($\Delta G_{nonpolar} < -0.8$ kcal/mol, except residue 42). Only residue pair 0-40 was identified as informative pair [Fig. 8(B)]. For the uninformative pairs 0-41, 0-42, and 0-43, previous studies showed that hydrogen bonds are formed between their backbone atoms,[15,23–29,51] which is consistent with their insignificant contribution to the binding specificity. PDZ residue 32 and 113, mostly Lys or Arg for the 11 PDZ domains form strong polar interactions with peptide residue 0 ($\Delta G_{polar} = -3.1$ and -3.0 kcal/mol, respectively). Previous structural analysis showed that these two residues stabilize carboxyl group of the peptide residue 0 through salt bridges with the backbone atoms,[29,52,53] which supports the uninformative pairs of 0-32 and 0-113 found by feature selection. We also observed that such polar interactions were not always formed in all the 86 PDZ-binding complexes, which suggests that these interactions may be sensitive to conformational flexibility. PDZ residues 43, 45, 109, and 112 are likely to be hydrophobic, such as Phe, Val, Leu, and Ile, as shown in the multiple sequence alignment (see Fig. 2). Strong nonpolar interactions have been observed for residue pairs 0-43, 0-45, 0-109, and 0-112 in this group ($\Delta G_{nonpolar} = -2.3$, -1.2, -1.1, and -1.7 kcal/mol, respectively). Previous structural analysis showed that these four residues form a hydrophobic pocket to interact with the hydrophobic side-chain of peptide residue 0,[24,26,27,29,54] which is normally Leu, Val, or Ile. Forty-five and 109 formed informative interaction pairs with peptide residue 0 probably because these two residues are located at the edge of the binding pocket [Fig. 8(B)].

Peptide residue -1 interacts with 10 PDZ residues (Supporting Information Figure S2a). All the interaction pairs were identified as informative except -1-74. The interaction between -1 and 74 is moderate (average $\Delta G_{total} = -1.5$ kcal/mol, Supporting Information Table S7). Multiple sequence alignment showed that the position 74 contains diverse amino acids, including aliphatic, aromatic, polar, and positive charged amino acids (see Fig. 2). These two observations may explain why this pair is uninformative. Contradictory to the reports that the side chains of -1 and 74 form hydrophobic contacts,[27,29] the -1-74 pair showed weak nonpolar (average $\Delta G_{nonpolar} = -0.5$ kcal/mol, Supporting Information Table S7) and for a few cases strong polar interactions (average $\Delta G_{nonpolar} = -1.0$ kcal/mol, Supporting Information Table S7). The nine informative pairs (Supporting Information Figure S2b) contribute about 11% of the average total binding energy, compared to 2% contributed by the uninformative pair -1-74. Amino acid choice at residue -1 is quite broad and no general pattern has been found.[13,14] In the MIEC-SVM model, the energetic characterization avoids the arbitrariness of summarizing occurrence of amino acids in a peptide or PDZ position. Regardless of amino acid types, as long as an amino acid forms favorable MIECs, it contributes to the PDZ-peptide binding.

Peptide residue -2 are involved in 10 interaction pairs (Supporting Information Figure S3a), including four informative ones: -2-45, -2-46, -2-71, and -2-105 (Supporting Information Figure S3b). Residue 45 shows both strong nonpolar and polar energies ($\Delta G_{nonpolar} = -1.5$ kcal/mol, $\Delta G_{polar} = -3.9$ kcal/mol) due to the hydrogen bonds formed between the backbone atoms and side chains,[55] which explains its roles in binding specificity determination. Residue 105 shows a strong interaction with peptide residue -2 ($\Delta G_{total} = -1.7$ kcal/mol), and this interaction is indeed important in determining the PDZ-peptide binding specificity[18]: residue 105 is critical to form hydrogen bond with residue -2 for recognizing class I peptides[14,24,26,27,29] and it is also part of the hydrophobic pocket crucial for the binding of class II peptides.[51,55] In total, the four informative pairs contribute about 11% of the total average binding energy, while the six uninformative pairs contribute about 8%.

All six interaction pairs involving residue -3 were identified as informative (Supporting Information Figure S4), indicating the important role of residue -3 in determining the binding specificity. These six informative pairs contribute about 14% of the total average binding energy. The strong polar interactions between peptide residue -3 and PDZ residue 71 ($\Delta G_{polar} = -2.4$ kcal/mol) stabilizes the side-chain of residue -3, which is consistent with the previous observations.[18,24,27,29] Residues 44 and 105 were also reported to affect the amino acid choice in position -3.[18]

Among the five interaction pairs formed by peptide residue -4, three of them (-4-47, -4-105, and -4-106) were identified as informative and they contribute about 6% of the total binding energy, compared with 3% contributed by the two uninformative pairs (Supporting Information Figure S5). It is worth pointing out that residue -4 forms strong nonpolar interaction with residue 105 ($\Delta G_{nonpolar} = -1.7$ kcal/mol), which suggests another important interaction formed by residue 105 besides its specific recognition of residue -2. Our analysis here is resonant to the previous studies on the roles of residue -4 in binding specificity.[24,29]

CONCLUSION

In this study, MIEC-SVM was applied to PDZ domain and successfully discriminated binding peptides from

nonbinding peptides. Compared with sequence-based methods of predicting PDZ binding specificity, MIEC-SVM takes advantage of structural information to better characterize the PDZ-peptide interactions. The MIECs are free energy terms and the matrix is fully filled regardless of what amino acids interact with each other. In contrast, the contact matrix used in other methods has to consider the combinations of amino acids, for example, Ala at peptide position −1 contacts with Ala/Ile/Leu etc. at PDZ position 1. Because of the availability of limited binding peptides, the contact matrix is often sparse because not each combination of amino acid pairs is observed in the relatively small samples of binders. Therefore, fully filled in MIEC matrix require less data to train the SVM model and thus avoids over-fitting problem, which reduces noise and improves prediction accuracy. Compared with energy-only-based calculations, MIEC-SVM employs the SVM to filter out noisy interactions as well as to better capture the local interacting environment such as the interior dielectric constant variation.[33–35] The satisfactory performance of MIEC-SVM on various systems (PDZ, SH3, and HIV protease[33–35]) illustrates its generality on deciphering binding specificity of proteins from a structural and energetic viewpoint.

For a PDZ domain to achieve binding specificity, it needs to distinguish (a) between binding peptides and ones not binding to any PDZ domain; (b) between its own binders and peptides binding to other PDZ domains. The residue pair selection shows that only 28 out of 44 pairs at the interaction interface are informative to binding specificity, although the remaining 16 pairs still contribute significantly to binding. Using only the 28 informative pairs, MIEC-SVM achieved better prediction accuracy. Most of the previous identified important residue pairs for PDZ-peptide binding are informative. Given these observations, we hypothesize that the PDZ-peptide residue pair interactions fall into two categories: binding affinity, that is, the 16 uninformative pairs that discriminate PDZ binding peptides and other peptides; binding specificity, that is, the 28 informative pairs that discriminate among PDZ binding peptides. This hypothesis is waiting for experimental test, which requires point mutations on both PDZ and peptide residues to decipher the contribution of each residue pair to binding specificity. The feature (residue pair) selection in the MIEC-SVM model is straightforward, but provides great insights into understanding the binding specificity. We expect such analysis would shed light on deciphering protein recognition in other systems as well.

ACKNOWLEDGMENTS

Part of the MD simulations was conducted in a computer cluster of Center for Theoretical Biological Physics (CTBP) at UCSD. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

REFERENCES

1. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. Science 2003;300:445–452.
2. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. Annu Rev Biochem 2006;75:655–680.
3. Cho KO, Hunt CA, Kennedy MB. The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. Neuron 1992;9:929–942.
4. Woods DF, Bryant PJ. ZO-1, Dlg A and PSD-95/SAP90: homologous proteins in tight, septate and synaptic cell junctions. Mech Dev 1993;44:85–89.
5. Kim E, Niethammer M, Rothschild A, Jan YN, Sheng M. Clustering of Shaker-type K+ channels by interaction with a family of membrane-associated guanylate kinases. Nature 1995;378:85–88.
6. Jemth P, Gianni S. PDZ domains: folding and binding. Biochemistry 2007;46:8701–8708.
7. Kim E, Sheng M. PDZ domain proteins of synapses. Nat Rev Neurosci 2004;5:771–781.
8. Nagafuchi A. Molecular architecture of adherens junctions. Curr Opin Cell Biol 2001;13:600–603.
9. Altschuler Y, Hodson C, Milgram SL. The apical compartment: trafficking pathways, regulators and scaffolding proteins. Curr Opin Cell Biol 2003;15:423–429.
10. Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, Crompton A, Chan AC, Anderson JM, Cantley LC. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 1997;275:73–77.
11. Nourry C, Grant SG, Borg JP. PDZ domain proteins: plug and play! Sci STKE 2003;2003:RE7.
12. Stricker NL, Christopherson KS, Yi BA, Schatz PJ, Raab RW, Dawes G, Bassett DE, Jr, Bredt DS, Li M. PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. Nat Biotechnol 1997;15:336–342.
13. Hung AY, Sheng M. PDZ domains: structural modules for protein complex assembly. J Biol Chem 2002;277:5699–5702.
14. Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. J Cell Sci 2001;114 (Part 18):3219–3231.
15. Birrane G, Chung J, Ladias JA. Novel mode of ligand recognition by the Erbin PDZ domain. J Biol Chem 2003;278:1399–1402.
16. Fuh G, Pisabarro MT, Li Y, Quan C, Lasky LA, Sidhu SS. Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. J Biol Chem 2000;275:21486–21491.
17. Laura RP, Witt AS, Held HA, Gerstner R, Deshayes K, Koehler MF, Kosik KS, Sidhu SS, Lasky LA. The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of delta-catenin and ARVCF. J Biol Chem 2002;277:12906–12914.
18. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, 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. A specificity map for the PDZ domain family. PLoS Biol 2008;6:e239.
19. Sharma SC, Memic A, Rupasinghe CN, Duc AC, Spaller MR. T7 phage display as a method of peptide ligand discovery for PDZ domain proteins. Biopolymers 2009;92:183–193.
20. Stiffler MA, Grantcharova VP, Sevecka M, MacBeath G. Uncovering quantitative protein interaction networks for mouse PDZ domains using protein microarrays. J Am Chem Soc 2006;128:5913–5922.
21. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. Science 2007;317:364–369.

22. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G. Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol 2008;26:1041–1045.
23. Runyon ST, Zhang Y, Appleton BA, Sazinsky SL, Wu P, Pan B, Wiesmann C, Skelton NJ, Sidhu SS. Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. Protein Sci 2007;16:2454–2471.
24. Kozlov G, Banville D, Gehring K, Ekiel I. Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the β2-β3 loop to PDZ domain-ligand interactions. J Mol Biol 2002;320:813–820.
25. Zhang Y, Dasgupta J, Ma RZ, Banks L, Thomas M, Chen XS. Structures of a human papillomavirus (HPV) E6 polypeptide bound to MAGUK proteins: mechanisms of targeting tumor suppressors by a high-risk HPV oncoprotein. J Virol 2007;81: 3618–3626.
26. Gianni S, Walma T, Arcovito A, Calosci N, Bellelli A, Engstrom A, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. Structure 2006;14:1801–1809.
27. Schultz J, Hoffmuller U, Krause G, Ashurst J, Macias MJ, Schmieder P, Schneider-Mergener J, Oschkinat H. Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. Nat Struct Biol 1998;5:19–24.
28. Zhang J, Yan X, Shi C, Yang X, Guo Y, Tian C, Long J, Shen Y. Structural basis of β-catenin recognition by Tax-interacting protein-1. J Mol Biol 2008;384:255–263.
29. Skelton NJ, Koehler MF, Zobel K, Wong WL, Yeh S, Pisabarro MT, Yin JP, Lasky LA, Sidhu SS. Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. J Biol Chem 2003;278:7645–7654.
30. Kalyoncu S, Keskin O, Gursoy A. Interaction prediction and classification of PDZ domains. BMC Bioinformatics 2010;11:357.
31. Hui S, Bader GD. Proteome scanning to predict PDZ domain interactions using support vector machines. BMC Bioinformatics 2010;11:507.
32. Smith CA, Kortemme T. Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. J Mol Biol 2010;402:460–474.
33. Hou T, Zhang W, Case DA, Wang W. Characterization of domain-peptide interaction interface: a case study on the amphiphysin-1 SH3 domain. J Mol Biol 2008;376:1201–1214.
34. Hou T, Xu Z, Zhang W, McLaughlin WA, Case DA, Xu Y, Wang W. Characterization of domain-peptide interaction interface: a generic structure-based model to decipher the binding specificity of SH3 domains. Mol Cell Proteomics 2009;8:639–649.
35. Hou T, Zhang W, Wang J, Wang W. Predicting drug resistance of the HIV-1 protease using molecular interaction energy components. Proteins 2009;74:837–846.
36. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
37. Garrard SM, Capaldo CT, Gao L, Rosen MK, Macara IG, Tomchick DR. Structure of Cdc42 in a complex with the GTPase-binding domain of the cell polarity protein, Par6. EMBO J 2003;22:1125–1133.
38. Accylrys Inc. Discovery Studio 2.1. San Diego: Accelrys; 2009.
39. Krivov GG, Shapovalov MV, Dunbrack RL, Jr. Improved prediction of protein side-chain conformations with SCWRL4. Proteins 2009;77:778–795.
40. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossvary I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA. AMBER10. San Francisco: University of California; 2008.
41. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003;24:1999–2012.
42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
43. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-integration of Cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. J Comput Phys 1977;23:327–341.
44. Darden T, York D, Pedersen L. Particle Mesh Ewald—an N. Log(N) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
45. Onufriev A, Bashford D, Case DA. Modification of the generalized Born model suitable for macromolecules. J Phys Chem B 2000;104: 3712–3720.
46. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 2004;55:383–394.
47. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
48. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology; 2:27: 1–27, 2001.
49. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612.
50. Theodoridis S, Koutroumbas K. Pattern recognition, 3rd ed. Elsevier; 2006.
51. Kang BS, Cooper DR, Devedjiev Y, Derewenda U, Derewenda ZS. Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. Structure 2003;11:845–853.
52. Karthikeyan S, Leung T, Ladias JA. Structural basis of the Na+/H+ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 2001;276:19683–19686.
53. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 1996;85:1067–1076.
54. Im YJ, Lee JH, Park SH, Park SJ, Rho SH, Kang GB, Kim E, Eom SH. Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. J Biol Chem 2003;278:48099–48104.
55. Im YJ, Park SH, Rho SH, Lee JH, Kang GB, Sheng M, Kim E, Eom SH. Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization. J Biol Chem 2003;278:8501–8507.