

Characterization of PDZ domain–peptide interactions using an integrated protocol of QM/MM, PB/SA, and CFEA analyses

Feifei Tian · Yonggang Lv · Peng Zhou ·
Li Yang

Received: 25 April 2011 / Accepted: 13 September 2011 / Published online: 1 October 2011
© Springer Science+Business Media B.V. 2011

Abstract Protein–protein interactions, particularly weak and transient ones, are often mediated by peptide recognition domains. Characterizing the interaction interface of domain–peptide complexes and analyzing binding specificity for modular domains are critical for deciphering protein–protein interaction networks. In this article, we report the successful use of an integrated computational protocol to dissect the energetic profile and structural basis of peptide binding to third PDZ domain (PDZ3) from the PSD-95 protein. This protocol employs rigorous quantum mechanics/molecular mechanics (QM/MM), semi-empirical Poisson–Boltzmann/surface area (PB/SA), and empirical conformational free energy analysis (CFEA) to quantitatively describe and decompose systematic energy changes arising from, respectively, noncovalent interaction, desolvation effect, and conformational entropy loss associated with the formation of 30 affinity-known PDZ3–peptide complexes. We show that the QM/MM-, PB/SA-, and CFEA-derived energy components can work together fairly well in

reproducing experimentally measured affinity after a linearly weighting treatment, albeit they are not compatible with each other directly. We also demonstrate that: (1) noncovalent interaction and desolvation effect donate, respectively, stability and specificity to complex architecture, while entropy loss contributes modestly to binding; (2) P₀ and P₋₂ of peptide ligand are the most important positions for determining both the stability and specificity of the PDZ3–peptide complex, P₋₁ and P₋₃ can confer substantial stability (but not specificity) for the complex, and N-terminal P₋₄ and P₋₅ have only a very limited effect on binding.

Keywords PDZ domain · Protein–peptide binding · Noncovalent interaction · Desolvation effect · Conformational entropy loss

Introduction

The interactions between protein domains and their peptide ligands play critical roles in signal transduction and many other key biological processes. Recently, Russell and co-workers estimated that 15–40% of all interactions in the cell are mediated through protein–peptide interactions [1, 2], meaning that nearly every protein is affected either directly or indirectly by peptide–binding events [3]. These interactions are often mediated by peptide recognition modular domains, such as the SH3, SH2, PDZ, and WW domains, which bind to short peptides with specific sequence motifs [4]. The PDZ (PSD-95/Discs-large/ZO-1) domain family is one of the most abundant modular domain families in multicellular proteomes and is present in a variety of proteins, such as phosphatases, tumor suppressors, and regulator proteins, to orchestrate diverse cellular processes. PDZ domains are composed of 80–90

F. Tian and Y. Lv contributed equally to this work.

F. Tian · Y. Lv · L. Yang (✉)
Key Laboratory of Biorheological Science and Technology
Under Ministry of Education, ‘111’ Project Laboratory
of Biomechanics and Tissue Repair, and Bioengineering
College, Chongqing University, Chongqing 400044, China
e-mail: liyang@cqu.edu.cn

F. Tian
School of Life Science and Engineering, Southwest Jiaotong
University, Chengdu 610031, China

P. Zhou
Center of Bioinformatics (COBI), School of Life Science
and Technology, University of Electronic Science
and Technology of China, Chengdu 610054, China

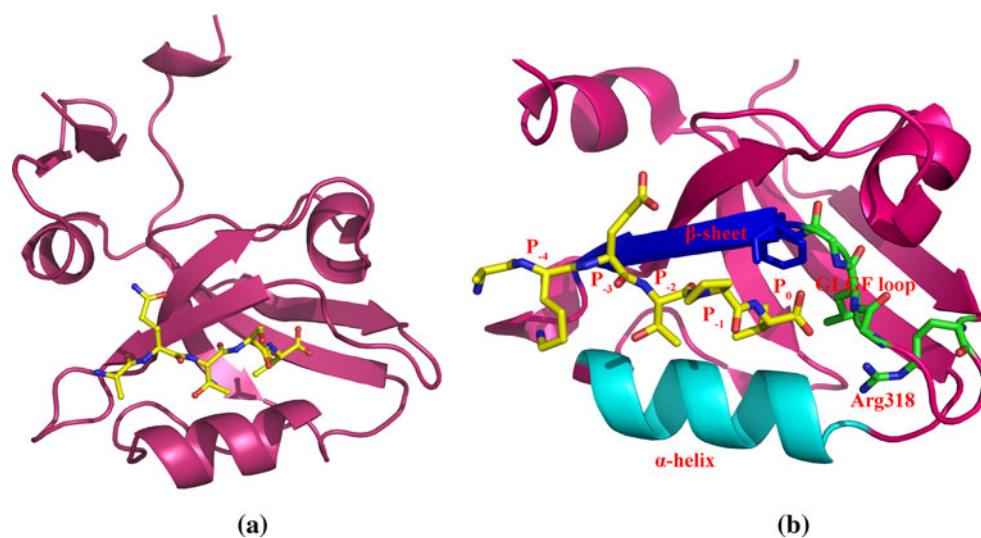
amino acids that define a central bent six-stranded β -sheet surrounded by two α -helices [5], which can specifically bind the C-terminal sequence of partner proteins. Since their initial identification [6], PDZ and PDZ-like domains have been recognized in numerous proteins from organisms as diverse as bacteria, plants, yeast, metazoans, and flies [7]. Diversity of PDZ-containing protein functions is provided by the large number of PDZ proteins distributed throughout nature. This protein family is implicated in many molecular networks from the plasma membrane to the nucleus [8]. The biological importance of PDZ domains is further underscored by the identification of various PDZ proteins as human disease and pathogen effector targets [9], which makes PDZ-involved interactions good candidates for developing small molecule inhibitors [10].

PDZ domains can be categorized into three classes (I, II, and III) in terms of the loop ($\beta 1:\beta 2$) and position -2 of the peptide ligand [11], of which the class I domain is the most-well characterized PDZ domain that has a GL/YGF loop ($\beta 1:\beta 2$) and binds C-terminal peptides with sequence pattern XS/TXV/I/L [12]. High-resolution crystal structures of a class I third PDZ domain (PDZ3) from the postsynaptic density 95 protein (PSD-95) in complex with and in the absence of its peptide ligand KQTSV have been solved by Doyle et al. [13]. The structures revealed that a four-residue C-terminal stretch of the peptide engages the PDZ3 anti-parallel main chain interactions with a β -sheet of the domain. Recognition of the terminal carboxylate group of the peptide is conferred by a cradle of main-chain amides provided by a GLGF loop as well as by an arginine side chain. The binding pocket contains a characteristic hydrophobic loop that binds the peptide ligand through the formation of complicated hydrogen bond networks (Fig. 1).

Although the crystallographic information has shed light on the structural basis of PDZ–peptide recognition, there

still exist a number of problems to be solved. In particular, the quantitative energetic knowledge regarding the free energy contributions of peptide chains and residue sites would be fundamentally valuable for understanding the thermodynamic behavior of this specific binding, but it is indeed unavailable from the crystal structures. Alternatively, computational approaches provide a promising way to analyze the binding profile of PDZ–peptide interactions at atomic level. Previously, a number of modeling experiments have attempted to qualitatively classify and quantitatively predict the binding behavior of diverse peptides to different PDZ domains. For instance, Stiffler et al. [14] trained a variation of a position-specific scoring matrix (PSSM) on interaction pairs determined by protein arrays to discriminate the binding specificity of mouse PDZ domains, and Chen et al. [15] later developed a method that incorporated structural information on protein–peptide residue pairs within close proximity of each other. The success of these two studies showed the strength of integrating information from diverse sources. Very recently, by incorporating empirical ROSETTA potentials and structural information, Kaufmann and co-workers have systematically investigated the energetic components involved in the recognition and interaction of PDZ domains with cognate ligands [16]. In order to give deeper insight into the physicochemical and structural implications underlying the specific recognition and interaction between PDZ domains and their peptide ligands, in this study an integrated protocol of hybrid quantum mechanics/molecular mechanics (QM/MM), Poisson–Boltzmann/surface area (PB/SA), and conformational free energy analysis (CFEA) is proposed to accurately characterize the direct nonbonded interactions (by QM/MM) as well as indirect desolvation effect (by PB/SA) and conformational entropy loss (by CFEA) associated with the binding and association of 30

Fig. 1 **a** Stereoview of PDZ3 (ribbon)–peptide KQTSV (stick) complex structure (PDB entry: 1tp3). **b** Close-up view of the PDZ3–peptide binding pocket



affinity-known peptides with the PSD-95 PDZ3. Furthermore, the calculated energy components are correlated to the experimentally measured affinity by using several statistical modeling tools, including linear multiple linear regression (MLR) and nonlinear support vector machine (SVM) and Gaussian process (GP). We also employed rigorous quantum mechanics methods and molecular graphics techniques to dissect the electronic structure characteristics and charge-transfer behavior of noncovalent interactions across the model interface of PDZ3–peptide complexes. On the basis of these calculations, we systematically discuss the thermodynamic nature of the stability and specificity of peptide binding to PDZ3.

Materials and methods

Data set

Saro et al. [17] have designed a hexapeptide KKETEVEV on the basis of preferred binding sequence for class I domains, which exhibits high affinity for PDZ3 ($K_d \approx 2 \mu\text{M}$). This peptide was defined with the position pattern $P_{-5}P_{-4}P_{-3}P_{-2}P_{-1}P_0$ following the nomenclature suggested by Doyle et al. [13]. In a later study, they performed a residue replacement analysis at different positions of the KKETEVEV to construct a diverse mutation profile on this sequence scaffold. As a result, 29 mutated peptides, including 27 hexapeptides plus a pentapeptide and a heptapeptide, were synthesized using a standard F_{moc} -based solid-phase peptide synthesis protocol and purified to single peak homogeneity using reverse-phase HPLC with product masses confirmed by ESI–MS. The thermodynamic parameters associated with the binding of the 29 linear oligopeptides as well as the scaffold sequence KKETEVEV to PSD-95 PDZ3 have been measured using isothermal titration calorimetry (ITC) and shown in Table 1, in which the experimental values are the arithmetic mean of at least two independent assays [18].

Construction of PDZ3–peptide complex models

Recently, two high-resolution crystal structures of PSD-95 PDZ3 separately in complex with peptide ligands KKETWV and KKETPV were solved at 1.54 and 1.99 Å, respectively (PDB entries: 1tp5 and 1tp3), from which the best one (PDZ3–KKETWV complex) was chosen as a template to perform a virtual site-directed mutagenesis to prepare the crude models of other hexapeptide complexes. For the pentapeptide KETEVEV, its complex model was simply obtained by cutting the N-terminal Lys residue from the hexapeptide KKETEVEV in complex with PDZ3, while the heptapeptide KKKETEVEV complex was constructed by manually adding a Lys to the N-terminus of KKETEVEV.

Virtually mutating a peptide residue was implemented in two steps: the side-chain of the residue under mutation was manually deleted from the template KKETWV, and then a new side-chain was added automatically using the rotamer-based SCWRL program [19]. Before the virtual mutagenesis protocol all water molecules and cofactors were removed from the template structure, followed by a hydrogen-adding procedure using the REDUCE strategy [20]. The SCWRL and REDUCE adopted here are because these two programs have been demonstrated good performance in reproducing experimentally determined structures of peptides and proteins [21, 22]. Subsequently, crude complex models were subjected to an energy minimization treatment to eliminate unreasonable collisions and distortions using the AMBER9 force field package [23].

QM/MM, PB/SA, and CFEA analyses of PDZ3–peptide binding energy

Generally, the free energy change associated with three aspects contributing to the binding of a peptide ligand to a protein receptor: noncovalent interactions between protein and peptide, desolvation effect due to the displacement of water molecules from the protein–peptide interface upon binding, and conformational energy loss incurred from the loss of flexibility of rotatable single-bonds during the binding process. In a previous study, we have described a strategy that employs rigorous QM/MM–PB/SA instead of traditional MM–PB/SA to dissect the free energy profile of OppA protein interacting with its cognate ligands [24]. In addition, considering that the PDZ3 ligands are linear peptides that possess large flexibility and hence would bear significant entropy loss during the binding process, we proposed a method called conformational free energy analysis (CFEA) to account for entropic contributions to the binding energy of peptide to PDZ3. Here, we give a detailed description of the QM/MM, PB/SA, and CFEA:

1. Noncovalent interaction energy calculated by QM/MM ($\Delta E_{\text{QM/MM}}$): the direct nonbonded energy arising from, for example, hydrogen bonds, van der Waals contacts, and electrostatic forces between PDZ3 and its peptide ligand in the complex state can be calculated using a two-layered QM/MM scheme, which can be carried out with the ONIOM algorithm [25] implemented in Gaussian03 suite [26]. Briefly, the peptide ligands and corresponding protein residues that specifically interact with the ligand were included in the QM layer and treated with a high level of the AM1 method [27], while the rest of the atoms were assigned to the MM layer and described using a low level of the AMBER96 force field [28]. Li et al. [29] recently demonstrated that use of the semiempirical AM1 method to analyze

Table 1 The thermodynamic data associated with the binding of 30 C-terminal peptides to PDZ3

Peptide	Experiment [18]		Energy decomposition				Statistical modeling		
	$pK_d^{\text{expl.}}$	$\Delta G^{\text{expl.}}$	$\Delta E_{\text{QM/MM}}$	$\Delta G_{\text{PB/SA}}$	ΔG_{CEFA}	ΔG_{total}	pK_d^{MLR}	pK_d^{SVM}	pK_d^{GP}
KETEV	4.74	-6.5	-11.52	-5.10	2.35	-14.27	4.92	4.89	4.88
KKETEV	5.72	-7.8	-12.27	-5.34	3.26	-14.35	5.18	5.17	5.12
KKKETE V	5.89	-8.0	-13.68	-5.96	3.88	-15.76	5.72	5.68	5.59
KKETEA	4.04	-5.5	-14.03	-4.21	3.07	-15.17	4.86	4.95	4.88
KKETEL	5.10	-7.0	-14.76	-5.71	4.69	-15.78	5.78	5.79	5.67
KKETEI	5.11	-7.0	-13.52	-5.77	4.12	-15.17	5.61	5.59	5.50
KKETEM	4.68	-6.4	-12.60	-4.32	3.18	-13.74	4.71	4.78	4.74
KKETEF	4.24	-5.8	-11.44	-4.45	4.56	-11.33	4.65	4.74	4.70
KKETET	3.98	-5.4	-12.36	-3.87	2.01	-14.22	4.41	4.48	4.47
KKESEV	5.18	-7.1	-14.79	-5.17	4.98	-14.98	5.52	5.58	5.48
KKECEV	4.14	-5.7	-10.11	-3.23	3.15	-10.19	3.79	3.92	3.97
KKESEL	4.48	-6.1	-12.89	-3.71	3.90	-12.70	4.46	4.61	4.57
KKESEI	4.62	-6.3	-11.37	-3.92	4.28	-11.01	4.36	4.48	4.46
KKESEF	4.01	-5.5	-10.36	-4.66	3.69	-11.33	4.57	4.60	4.61
KKETGV	5.62	-7.7	-15.01	-5.73	2.83	-17.91	5.77	5.72	5.62
KKETAV	6.35	-8.7	-13.47	-5.29	2.94	-15.82	5.32	5.31	5.24
KKETVV	5.89	-8.1	-13.63	-5.66	3.18	-16.11	5.54	5.51	5.43
KKETLV	5.74	-7.8	-14.82	-4.38	4.69	-14.51	5.11	5.24	5.13
KKETPV	6.02	-8.2	-13.78	-5.46	4.42	-14.82	5.50	5.52	5.43
KKETWV	5.55	-7.6	-12.56	-4.31	3.01	-13.86	4.69	4.76	4.72
KKETDV	4.70	-6.4	-14.53	-4.38	4.17	-14.74	5.06	5.17	5.07
KKETKV	5.92	-8.1	-12.68	-5.09	4.58	-13.19	5.16	5.20	5.13
KKGTEV	4.10	-5.6	-11.48	-4.42	3.46	-12.44	4.60	4.66	4.65
KKATEV	4.68	-6.4	-12.69	-4.36	3.77	-13.28	4.76	4.85	4.80
KKQTEV	5.40	-7.4	-14.72	-4.98	4.49	-15.21	5.40	5.47	5.36
KKDTEV	4.07	-5.6	-13.40	-3.87	3.14	-14.13	4.60	4.71	4.66
KKKTEV	4.57	-6.2	-14.83	-4.03	3.87	-14.99	4.91	5.04	4.95
KKGTGV	3.56	-4.9	-11.64	-4.79	4.02	-12.41	4.83	4.88	4.85
KKATAV	5.08	-6.9	-12.08	-3.50	3.10	-12.48	4.21	4.35	4.33
YKETE V	5.92	-8.1	-13.41	-4.71	4.77	-13.35	5.08	5.17	5.09

All energy values are in kcal/mol

the weakly bound systems involving relatively large ligands is a good compromise between computational efficiency and accuracy.

According to the PDZ3–peptide recognition mode suggested by Kaufmann et al. [16], only the protein residues His372, Ala376, Leu379, and Lys380 can effectively interact with the side-chains of peptide and provide specific judgement for the recognition. Thus, these four residues were chosen as the key residues and included in the QM layer (Fig. 2). Subsequently, the protein–peptide interaction energy ($\Delta E_{\text{QM/MM}}$) of the complex was predicted according to the strategy proposed by Zhou et al. [30]. This was accomplished by performing a single point energy calculation twice: once on the bound system (E_1) and once on the unbound system in which the protein and peptide were

separated distantly to each other (E_2). In this way, the interaction energy can be calculated as $\Delta E_{\text{QM/MM}} = E_1 - E_2$.

- Desolvation free energy determined by PB/SA ($\Delta G_{\text{PB/SA}}$): protein–ligand binding involves desolvation, which significantly impacts the free energy of the system. To account for solvent effects associated with this process, the empirical solvent accessible surface area (SASA) model [31] and semi-empirical Poisson–Boltzmann/surface area (PB/SA) model [32] are available. Previously, the relatively rigorous PB/SA method was successfully incorporated into QM/MM to investigate the interaction behavior of proteins with their cognate and non-cognate ligands and received satisfactory qualitative results [33, 34]. In the PB/SA procedure total desolvation free energy ($\Delta G_{\text{PB/SA}}$) was

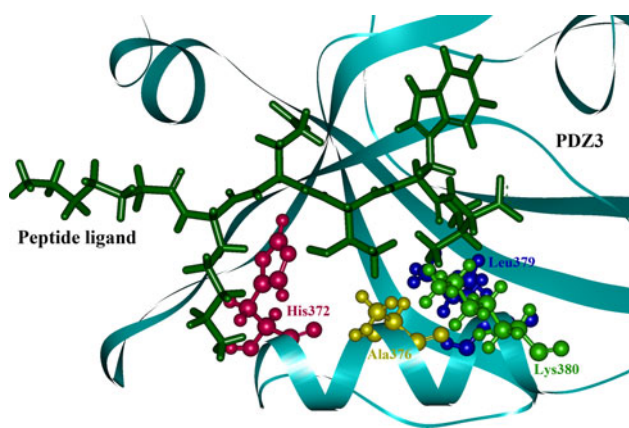


Fig. 2 QM/MM partition scheme for the PDZ3–peptide KKETWV complex (PDB entry: 1tp5). The peptide ligand (*stick*) and the protein residues His372, Ala376, Leu379, and Lys380 (*ball and stick*) that specifically interact with the ligand are in the QM layer, while the remainder (*ribbon*) of this system is in the MM layer

estimated from the polar (electrostatic) desolvation energy (ΔG_{polar}) and the nonpolar desolvation energy ($\Delta G_{\text{nonpolar}}$). The polar component was calculated by finite difference solutions to the nonlinear Poisson–Boltzmann equation, as implemented in DELPHI program [35], while nonpolar contribution was determined by summing up the weighted surface area of whole solute molecule, i.e. $\Delta G_{\text{nonpolar}} = \gamma \Delta A$, where $\gamma = 0.00542 \text{ kcal/mol } \text{Å}^2$ [32], and ΔA is the change in surface area upon the PDZ3–peptide binding, which can be computed with MSMS program [36].

3. Conformational entropy loss estimated by CFEA (ΔG_{CFEA}): accurate determination of entropy change upon biomolecular binding is of great challenge in the computational biology community, since entropy is perhaps the most elusive aspect of biomolecular thermodynamic behavior. Peptides are flexible linear molecules which encounter considerable entropic penalties due to loss of this flexibility during the binding process. The free energy contribution of conformational entropy loss (conformational free energy) to peptide affinity stems from two aspects: increased rigidity of the peptide backbone and side-chains upon binding. In this study, the former could be regarded as a constant if considering that the backbones of peptides investigated here are very similar in their length and arrangement manner in the binding groove of PDZ3. Therefore, we herein only calculated the conformational free energy contribution from increased rigidity of peptide side-chains, which can be computed as the difference between the conformational entropies of peptide side-chains in bound and unbound states, i.e. $\Delta G_{\text{CFEA}} = -T(S_{\text{bound}} - S_{\text{unbound}})$. The conformational entropy of a peptide ligand

side-chain in the bound state was estimated using Boltzmann’s formulation $S_{\text{bound}} = -\sum_i p_i \ln p_i$, where the sum is taken over all conformational states (modeled by penultimate rotamer library [37]) of the side-chain and p_i is the probability of being in state i . In this study, we used an in-house program 2D-GraLab [38] to carry out side-chain conformational entropy analysis in the bound state. For the conformational entropy of peptide side-chains in the unbound state (S_{unbound}) we simply took the values published by Creamer [39], who has performed exhaustive Monte Carlo simulations to give an accurate description for the side-chain behavior of unfolded polypeptides. In addition, entropy loss associated with the increased rigidity of PDZ3’s side-chains at the binding interface can also be calculated using the same protocol as that for peptide side-chains.

Electronic structure analysis of noncovalent interactions at PDZ3–peptide interface

The electronic structure characteristics of noncovalent interactions between PDZ3 and the peptide residues of interest were dissected by using the “atoms in molecules (AIM)” theory of Bader [40], which is based on a topological analysis of the electron charge density and its Laplacian. The AIM theory has proved itself a valuable tool to conceptually define atoms and, above all, bonds from a quantum mechanical standpoint [41]. Our investigation is conducted by means of the natural bond orbital (NBO) theory of Weinhold and co-workers [42], which allows us to quantitatively evaluate the charge transfer (CT) and bond order (BO) involving the formation of nonbonded interactions of PDZ3 with peptides.

The wave functions fed to AIM and NBO analyses were generated at the stringent Møller–Plesset second order perturbation level of theory in conjunction with Dunning’s augmented correlation consistent basis set, MP2/aug-cc-pVDZ [43], as implemented in the Gaussian03 suite [26], and the following AIM and NBO analyses were carried out with programs AIM2000 [44] and NBO5.0 [45], respectively.

Statistical modeling

The multivariate relationship between QM/MM-PB/SA-CEFA-derived energy components ($\Delta E_{\text{QM/MM}}$, $\Delta G_{\text{PB/SA}}$, and ΔG_{CFEA}) and experimentally determined affinity ($\text{p}K_{\text{d}}^{\text{expl}}$) was further explored by using linear MLR and nonlinear SVM and GP. The MLR method builds a weighted linear formula to correlate the energy terms with the binding affinity of peptide ligand:

$$pK_d = b_0 + b_1 \Delta E_{\text{QM/MM}} + b_2 \Delta G_{\text{PB/SA}} + b_3 \Delta G_{\text{CEFA}} \quad (1)$$

where b_0 is a constant term characterizing the additional, invariable contribution from other unknown factors that were not considered in the QM/MM-PB/SA-CEFA energy components, and b_1 , b_2 , and b_3 are the weights of energy terms $\Delta E_{\text{QM/MM}}$, $\Delta G_{\text{PB/SA}}$, and ΔG_{CEFA} , respectively, in contribution to peptide affinity. Compared with MLR, SVM and GP can handle data with strongly nonlinear, noisy, and collinear variables, and thus might be more suitable for mining complicated dependences involved in the PDZ3–peptide system. Detailed descriptions of MLR, SVM, and GP are found in Refs. [46–48], and these algorithms can be easily manipulated with an in-house program ZP-explore [49], which is running on MATLAB platform.

Results and discussion

Reconstruction of PDZ3–KKETPV complex structure

In order to examine the reliability of the virtual site-directed mutagenesis protocol in producing the structure model of PDZ3–peptide complexes, we herein used this method to reconstruct a structure-known complex, the PDZ3 in complex with peptide KKETPV, which has been elucidated at 1.99 Å resolution with X-ray crystallography (PDB entry: 1tp3). Superposition of reconstructed and crystal structures of peptide ligand in the binding groove of PDZ3 is shown in Fig. 3. As can be seen, two counterparts share a consensus binding motif that their backbones are fairly aligned to each other and the reconstructed side-chains of P₀, P₋₁, P₋₂, and P₋₃, the most important residues in the binding, are also basically placed properly as compared to that of crystal

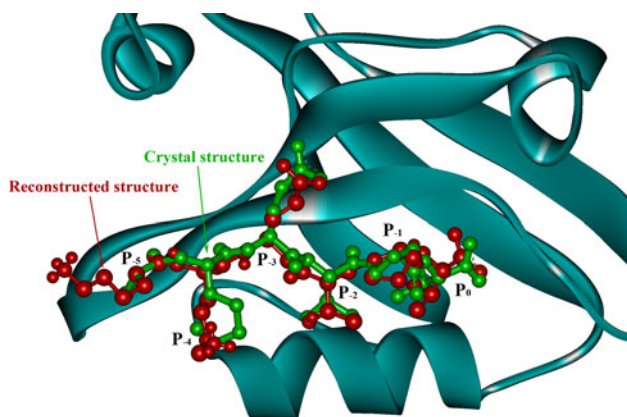


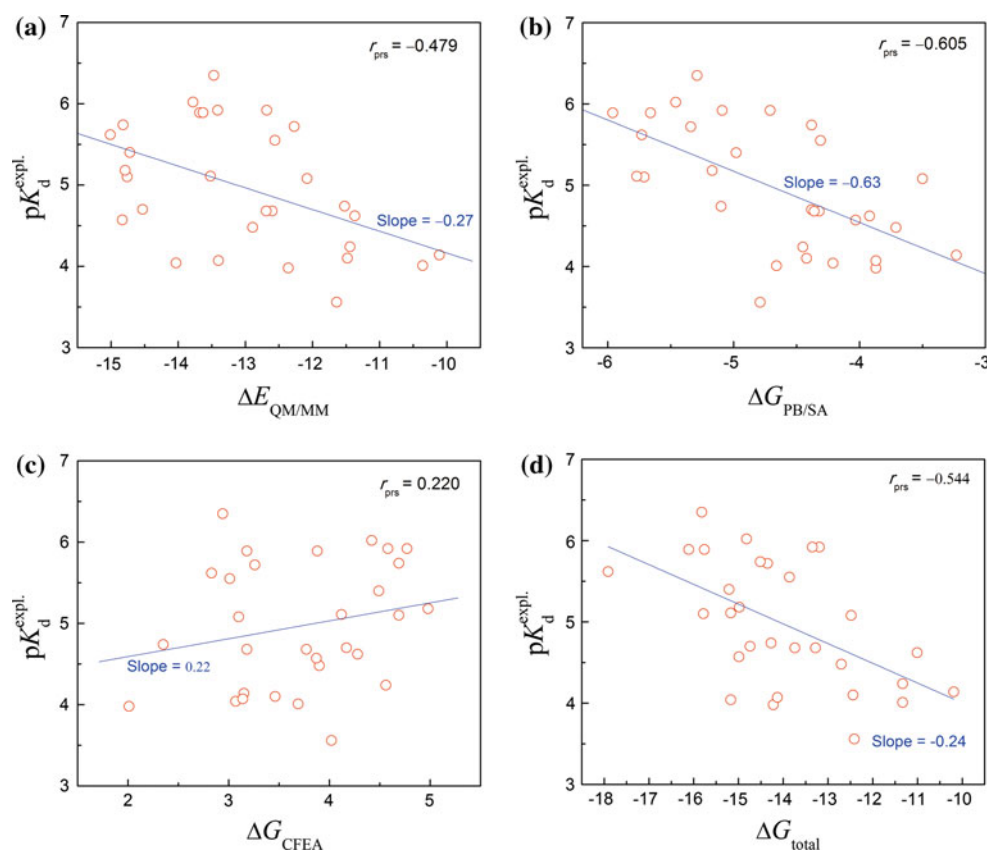
Fig. 3 Superposition of reconstructed and crystal structures of peptide KKETPV in the binding groove of PDZ3. Note that the side-chain of C-terminal Lys residue (P₋₅) is missing in the crystal structure (PDB entry: 1tp3)

structure. Only the Lys residue at P₋₄ in the reconstructed peptide has a relatively large deviation from the corresponding position of native structure. In addition, the missing side-chain of C-terminal Lys residue (P₋₅) in the crystal structure was predicted as an extended conformation along the peptide backbone, which can be confirmed by another complete X-ray-determined PDZ3–KKETWV complex structure (see Fig. 2). Generally speaking, the virtual mutagenesis protocol can properly model the binding mode and backbone arrangement of unknown peptides, but may not be capable of accurately locating the atomic positions of some side-chains (especially those with significant flexibility, such as Lys and Glu) in the groove.

Analysis of PDZ3–peptide binding energy components

PDZ3–peptide binding is a complicated thermodynamic process in which diverse energy components exert significant effects on free energy of the system. Here, we employed the QM/MM-PB/SA-CEFA scheme to dissect the free energy profile of binding. The calculated energy components for the 30 PDZ3–peptide complexes are tabulated in Table 1, in which $\Delta E_{\text{QM/MM}}$ characterizes the direct nonbonded energy between PDZ3 and peptide, while $\Delta G_{\text{PB/SA}}$ and ΔG_{CEFA} describe the indirect desolvation effect and conformational entropy loss, respectively, due to binding. As might be anticipated, the energy values of $\Delta E_{\text{QM/MM}}$ and $\Delta G_{\text{PB/SA}}$ associated with the 30 complex samples are all negative, indicating favorable contributions of noncovalent interaction and desolvation effect to peptide binding. In contrast, the positive ΔG_{CEFA} for the 30 samples clearly manifest a pronouncedly unfavorable entropic penalty during the PDZ3–peptide recognition process—this is expected, considering that the degrees of freedom of numerous single bonds of the highly flexible peptides are reduced upon binding. In addition, the nonbonded aspect $\Delta E_{\text{QM/MM}}$ exhibits a noticeable potency as compared to desolvation facet $\Delta G_{\text{PB/SA}}$ and, more, to conformational entropy loss ΔG_{CEFA} ; the absolute values of $\Delta E_{\text{QM/MM}}$ are all larger than 10 kcal/mol, whereas those of $\Delta G_{\text{PB/SA}}$ and ΔG_{CEFA} are only at the levels of 5 and 3 kcal/mol, respectively. Although nonbonded energy appears to be dominant in the binding, its correlation with experimental affinity (pK_d^{expl}) are moderate as their Pearson's coefficient r_{prs} is only -0.479 (Fig. 4a). By contrast, the secondary desolvation energy appears to be more relevant to peptide affinity ($r_{\text{prs}} = -0.605$), namely, the variance over the affinity values could be well explained by desolvation term (Fig. 4b). The substantial energy contribution from nonbonded interactions but the high correlation of the desolvation facet reveals that noncovalent interactions provide a larger proportion of systematic stabilization energy, whereas the desolvation effect donates more

Fig. 4 Correlations between calculated energy components and experimental binding affinity of the 30 PDZ3–peptide complexes: **a** pK_d^{expl} versus $\Delta E_{\text{QM/MM}}$, **b** pK_d^{expl} versus $\Delta G_{\text{PB/SA}}$, **c** pK_d^{expl} versus ΔG_{CEFA} , and **d** pK_d^{expl} versus ΔG_{total}



specificity to the interaction. The high stability but low specificity arising from noncovalent interactions are not unexpected, since crystallographic analysis revealed that PDZ3 can form a number of hydrogen bonds with the invariable backbone moiety of peptide ligands [13]. Furthermore, both the magnitude and correlation of conformational free energy conferring to peptide affinity are quite modest (Fig. 4c), indicating that entropy loss upon the binding contributes limitedly to PDZ3–peptide recognition. This is consistent with the results of a previous calorimetry study of PDZ domain interacting with nonproteinogenic peptides [17, 18].

Further, we tried to directly combine the values of the three energy terms ($\Delta E_{\text{QM/MM}}$, $\Delta G_{\text{PB/SA}}$, and ΔG_{CEFA}) together to reproduce the change in systematic free energy (ΔG^{expl}) determined by experiment, but this attempt failed. It is seen from Table 1 that the sums (ΔG_{total}) of calculated energy components deviate significantly from experimentally measured ones, that is, ΔG_{total} values were much larger than ΔG^{expl} values. This deviation could be attributed to the systematic errors existed among the three independent energy terms calculated at different levels of accuracy. Therefore, in the following section we will employ statistical modeling methods to explore linear and nonlinear relationships between these energy components and experimentally determined affinity.

Statistical modeling analysis

Since the contributions of different energetic components of binding may not be identical and linear owing to the heterogeneity between calculation methods and complicated dependences hidden in the investigated system, we further employed several statistical modeling methods, including a linear MLR and two nonlinear machine learning tools SVM [46] and GP [47, 48], to explore the potential relationship between QM/MM-PB/SA-CEFA-derived energy components and the experimentally determined affinity of the 30 PDZ3–peptide complexes. The obtained fitting coefficients of determination R of the three modeling methods are 0.654, 0.647, and 0.667, respectively. These models were further tested via leave-one-out cross-validation (LOOCV), and resulting coefficients Q are 0.509, 0.544, and 0.526, respectively. LOOCV involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. As can be seen from Fig. 5, the scatter plots of experimental versus computational values of peptide affinity are basically around the 45° slope lines and these sample points in the three plots present a common profile of distribution. In addition, the fitting ability (given by R) of the three methods is basically consistent, but the predictive

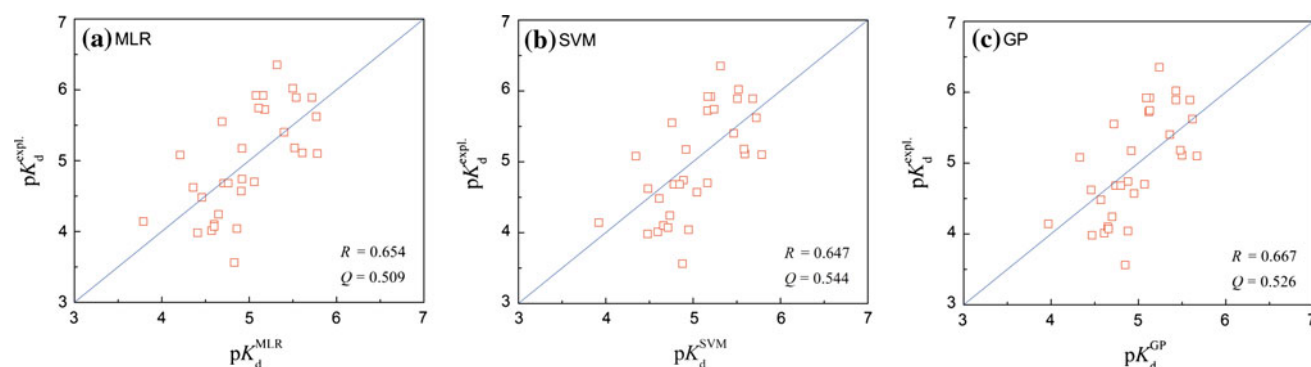


Fig. 5 Plots of experimental (pK_d^{exp}) against fitted (pK_d^{MLR} , pK_d^{SVM} , or pK_d^{GP}) binding affinities for the 30 PDZ3–peptide complexes: **a** MLR model, **b** SVM model, and **c** GP model

power (measured by Q) of SVM and GP seems to be slightly better than that of MLR, indicating that the calculated energy terms are mainly in linear relationship with the affinity, yet some nonlinear components are also involved.

Considering that the linear MLR performed as well and is easy to be interpreted, we herein addressed a further analysis on this model. The weighted formula of MLR model is shown as Eq. 2, in which increases in noncovalent interactions $\Delta E_{\text{QM/MM}}$ and desolvation $\Delta G_{\text{PB/SA}}$ decrease binding affinity, whereas increases in conformational entropy loss ΔG_{CEFA} increase affinity. This equation reveals that the $\Delta G_{\text{PB/SA}}$ term has a dominant effect on the specificity of peptide binding to PDZ3, as which can explain much more variance of observed pK_d than those by $\Delta E_{\text{QM/MM}}$ and ΔG_{CEFA} with respect to the absolute values of regression coefficients associated with these energy terms. The insignificant coefficient value 0.03 gives confirms that the entropy loss has only a modest effect on the binding and hence could be regarded as a secondary aspect affecting peptide affinity. The constant term 0.551 discloses an appreciable contribution of other unknown factors to the binding, albeit this contribution is not very significant (it only increases the K_d by 3.56 μM). The considerable difference between the three regression coefficients (b_1 , b_2 , and b_3) uncovers that the energy components derived from different levels of theory may not be directly compatible with each other, but these values can work together fairly well with a weighting treatment.

$$pK_d^{\text{MLR}} = 0.551 - 0.148 \times \Delta E_{\text{QM/MM}} - 0.508 \times \Delta G_{\text{PB/SA}} + 0.030 \times \Delta G_{\text{CEFA}} \quad (2)$$

Decomposition of total energy components into independent peptide residues

In order to examine in detail the energy contributions from the independent residue sites in the hexapeptide ligand, we broke down the peptide bonds (and capped them with

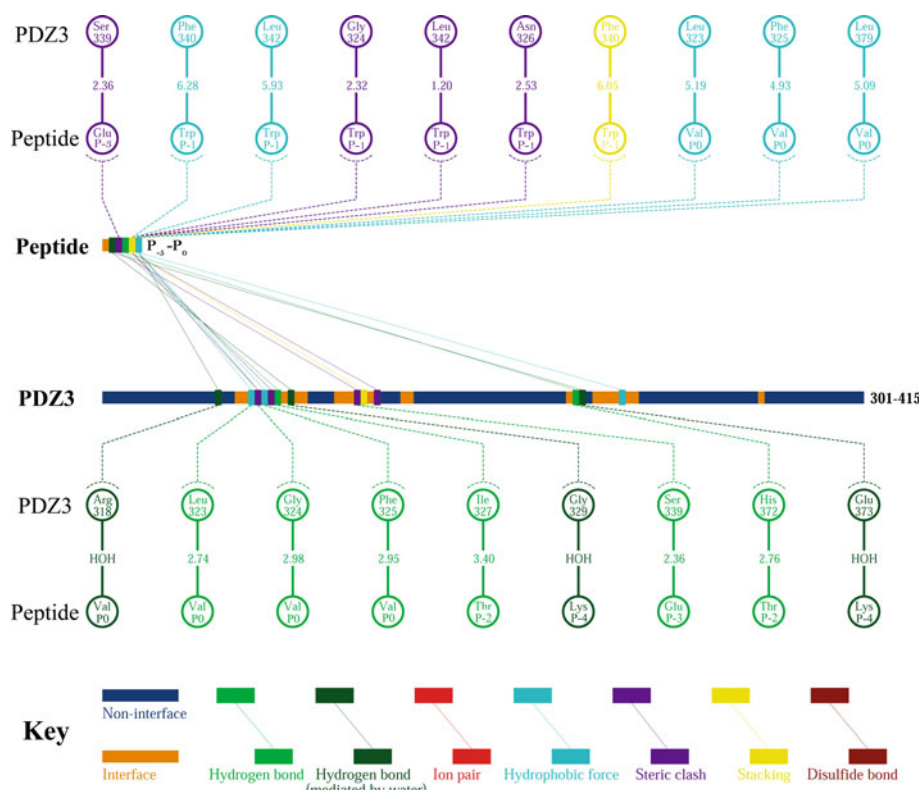
Table 2 Decomposition of total energy components into independent peptide residues

Position	Residue	$\Delta E_{\text{QM/MM}}$	$\Delta G_{\text{PB/SA}}$	ΔG_{CEFA}
P ₋₅	Lys	-0.68	-0.22	0.00
P ₋₄	Lys	-1.98	1.51	0.69
P ₋₃	Glu	-2.66	0.32	0.39
P ₋₂	Thr	-3.87	0.86	1.07
P ₋₁	Trp	-2.49	-3.86	1.36
P ₀	Val	-5.20	-1.89	1.04

hydrogen atoms) of, as a paradigm, KKETWV and separately computed the interaction energy, desolvation effect, and conformational free energy of each residue binding to PDZ3. The protocol used for calculating energy terms for individual residues was the same as that for whole peptide, and the resulting values are listed in Table 2. The sums of decomposed energy terms are not equal to the total energy components derived directly from the whole peptide, which could be attributed to the fact that the cooperation between the peptide residues was overlooked when the energy contributions were calculated on isolated residues. However, this discrepancy is not significant because the sums of individual terms are, albeit not equal to, roughly consistent with the total values. In the following text, we give a further discussion for each position of the hexapeptide KKETWV on the basis of decomposed energy terms, NBO and AIM analyses, as well as the schematic representation of noncovalent interactions across the binding interface of PDZ3–KKETWV complex (Fig. 6).

Position -5. It is known that PDZ domains can bind a C-terminal protein sequence 4–5 amino acids long [5]. In this regard, the P₋₅ and the residues beyond this position are anticipated to exert very limited potency to the binding. This point can be confirmed quantitatively according to the energetic effects associated with P₋₅. As can be seen from Table 2, the values of $\Delta E_{\text{QM/MM}}$, $\Delta G_{\text{PB/SA}}$, and ΔG_{CEFA} at

Fig. 6 Schematic representation of noncovalent interactions across the binding interface of PDZ3–peptide KKETWV complex. This plot is based on a high-resolution crystal structure (PDB entry: 1tp5) and was prepared with the in-house program 2D-GraLab [38]



P_{-5} are quite modest as -0.68 , -0.22 , and 0.00 kcal/mol, respectively. Hence, this position significantly confers neither stability nor specificity of binding.

Position -4 . Usually the P_{-4} of peptide ligand is occupied by a positively charged Lys residue which could impose long-range electrostatic potential on protein receptor. Indeed, an observable quantity (0.175 au) of charge transfer (CT) from this residue to protein during the binding was found with NBO analysis, implying that electrostatic force and CT bond may be the major factors that the residue affects peptide affinity. However, this position is also far always from the main body of PDZ3, leading to a marked discount on the energy contribution from its nonbonded interactions with PDZ3 ($\Delta E_{QM/MM} = -1.98$ kcal/mol). In addition, it is evident that the charged Lys would incur a non-negligible desolvation penalty ($\Delta G_{PB/SA} = 1.51$ kcal/mol) upon the binding owing to the burial of the charge in low-dielectric complex interface from high-dielectric solvent context. As shown in Fig. 6, P_{-4} can further form two water-mediated hydrogen bond bridges with the Gly329 and Glu373 of PDZ3, which would aid the specific recognition between peptide and PDZ3.

Position -3 . The P_{-3} Glu of the peptide is a negatively charged residue that can, in a sense, put effective nonbonded force on protein receptor. However, the side-chain of this residue is actually out of the binding groove and thus does not

define any functional interaction with PDZ3. Although the side-chain of Glu is incapable of imposing solid influence on the binding, its backbone was found to form a strong hydrogen bond with PDZ3 Ser339, with a bond length of only 2.36 Å (Fig. 6). The electron density ρ_b and Laplacian of the electron density $\nabla^2\rho_b$ of this hydrogen bond were predicted by AIM method to be, respectively, 0.0243 and 0.0826 au, which characterize a covalence- and electrostatics-hybrid profile for the hydrogen bond [50]. In this way, P_{-3} can give a prominent contribution to complex's stability through forming backbone hydrogen bond ($\Delta E_{QM/MM} = -2.66$ kcal/mol), albeit a fraction of this contribution would be used to pay off the unfavorable desolvation penalty ($\Delta G_{PB/SA} = 0.32$ kcal/mol) and conformational entropy loss ($\Delta G_{CEFA} = 0.39$ kcal/mol) due to the binding of this residue to PDZ3.

Position -2 . A polar Thr residue presents at P_{-2} of peptide ligand, which was observed to be tightly packed in the binding groove of PDZ3 [13], resulting in an appreciable desolvation penalty ($\Delta G_{PB/SA} = 0.86$ kcal/mol). As can be seen from Fig. 6, two hydrogen bonds engage this position: one formed between the Thr backbone and PDZ3 Ile327, another between the Thr side-chain and PDZ3 His372. AIM and NBO analyses indicate that the latter is a strong hydrogen bond, whereas a more dominant van der Waals force makes the former a weak one. Note that the strong side-chain hydrogen bond could determine specificity between the peptide ligand and PDZ3, and is also

responsible for the stability of complex architecture ($\Delta E_{\text{QM/MM}} = -3.87$ kcal/mol).

Position -1. The aromatic Trp residue at P_{-1} appears to be in compact contact with PDZ3. On the one hand, it participates in intensive nonbonded interactions such as π - π stacking and van der Waals collisions with its neighboring residues Gly324, Asn326, Phe340, and Leu342 of PDZ3 (Fig. 6), recruiting complicated noncovalent networks around it ($\Delta E_{\text{QM/MM}} = -2.49$ kcal/mol); on the other hand, the bulky, hydrophobic side-chain of P_{-1} Trp gives rise to a favorable desolvation effect due to its burial in the low-dielectric interface ($\Delta G_{\text{PB/SA}} = -3.86$ kcal/mol). As a result, this position throws marked stabilization energy to the system. The CT phenomenon is also noticeable, as 0.18 au is transferred from the electron-rich aromatic ring of Trp to its surroundings upon binding, which is thought to contribute significantly to stability of the complex. Although a considerable quantity of stabilization energy stems from the P_{-1} Trp, this residue seems not to constitute specificity for the recognition since there are no functional hydrogen bond and/or salt bridge observed.

Position 0. The P_0 is perhaps the most important position for the specific recognition of peptide ligand by PDZ3. Crystallographic analysis revealed that two carboxylate oxygens of P_0 Val participate in hydrogen bond formation with three amide nitrogens separately from Leu323, Gly324, and Phe325 of PDZ3, and a highly ordered water molecule is located between the carboxyl group of P_0 Val and the guanidinium group of PDZ3 Arg318 ($\Delta E_{\text{QM/MM}} = -5.20$ kcal/mol) (Fig. 6), giving rise to a very important functionality of P_0 in PDZ3-peptide recognition [13]. The crystallographic evidence was further confirmed by electronic structure analysis of these hydrogen bonds and water-mediated hydrogen bond; their electron density ρ_b and Laplacian of the electron density $\nabla^2\rho_b$ were calculated to be in the ranges 0.02–0.08 au and -0.4 to 0.03 au, respectively, which follow the quantities recommended for strong (ionic) hydrogen bonds by Nakanishi et al. [51]. A significant CT is observed; the 0.21 au proton shift from PDZ3 to P_0 implies that a notable stabilization effect arises from the CT bonding. In addition, burial of the hydrophobic side-chain of P_0 Val during the binding process could also provide -1.89 kcal/mol free energy contribution to complex's stability, albeit a large part of this would be used to compensate the conformational entropy loss of 1.04 kcal/mol.

Conclusions

An important class of protein interaction in critical cellular processes, such as signaling pathways, involves a domain

from one protein binding to a linear peptide stretch of another [52]. Exploring the highly complex behavior of flexible peptide ligands binding to their cognate receptor is definitely a challenge but can provide much valuable information about thermodynamic properties associated with the recognition process. Previously, Spaller and co-workers have performed a systematic calorimetric analysis on diverse peptide ligands binding to third PDZ domain (PDZ3) from PSD-95 protein, and demonstrated that six residues of peptide ligands are necessary and sufficient to capture maximal affinity [17, 18]. In this study, we have attempted to understand the structural basis and energetic landscape of the PDZ3 domain interacting with its peptide ligands and to enucleate the physicochemical properties and structural implications underlying the interaction. To achieve this, we employed an integrated protocol of QM/MM, PB/SA, and CFEA analyses to ascertain the binding mechanism of 30 affinity-known PDZ3-peptide complexes. We conclude with the following remarks:

1. The stability and specificity of PDZ3-peptide complexes are mainly conferred from, respectively, direct noncovalent interactions and indirect desolvation effect, while conformational entropy loss contributes relatively limited potency to the binding.
2. Due to the calculation of independent energy terms at different levels of accuracy, the QM/MM-, PB/SA-, and CFEA-derived energy components are not compatible with each other directly, but these values can work together fairly well with a weighting treatment.
3. Nonlinear SVM and GP performed slightly better than linear MLR when correlating calculated energy components to experimentally measured affinity of the 30 PDZ3-peptide complexes, indicating that the energy components are primarily in linear relationship with the affinity, and that only slight nonlinear relationships are involved.
4. The P_0 and P_{-2} of the peptide ligand are the most important positions for determining both the stability and specificity of PDZ3-peptide complex, P_{-1} and P_{-3} can confer substantial stability (but not specificity) for the complex architecture, and N-terminal P_{-4} and P_{-5} have only a limited thermodynamic effect on binding.

Acknowledgments The authors would like to express their gratitude to the anonymous reviewers and the editors for their professional, intensive, and useful comments. This work was supported in part by grants from the National Natural Science Foundation of China (11032012 and 30870608), the Key Science and Technology Program of CQ CSTC (2009AA5045), and the sharing fund of Chongqing university's large-scale equipment and the Visiting Scholar Foundation of Key Lab. of Biorheological Science and Technology Under Ministry of Education in Chongqing University.

References

- Neduvu V, Linding R, Su Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol* 3:e405
- Petsalaki E, Russell RB (2008) Peptide mediated interactions in biological systems: new discoveries and applications. *Curr Opin Biotechnol* 19:344–350
- Vanhee P, Stricher F, Baeten L, Verschuere E, Lenaerts T, Serrano L, Rousseau F, Schymkowitz J (2009) Protein–peptide interactions adopt the same structural motifs as monomeric protein folds. *Structure* 17:1128–1136
- Hou T, Zhang W, Case DA, Wang W (2008) Characterization of domain–peptide interaction interface: a case study on the amphiphysin-1 SH3 domain. *J Mol Biol* 376:1201–1214
- Hung AY, Sheng M (2002) PDZ domains: structural modules for protein complex assembly. *J Biol Chem* 277:5699–5702
- Kennedy MB (1995) Origin of PDZ (DHR, GLGF) domains. *Trends Biochem Sci* 20:350
- Ponting CP (1997) Evidence for PDZ domains in bacteria, yeast, and plants. *Protein Sci* 6:46–468
- Noury C, Grant SGN, Borg JP (2003) PDZ domain proteins: plug and play! *Sci STKE* 2003(179):RE7
- 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 (2008) A specificity map for the PDZ domain family. *PLoS Biol* 6:e239
- Dev KK (2004) Making protein interactions druggable: targeting PDZ domains. *Nat Rev Drug Discov* 3:1047–1056
- Giallourakis C, Cao Z, Green T, Wachtel H, Xie X, Lopez-Illasaca M, Daly M, Rioux J, Xavier R (2006) A molecular-properties-based approach to understanding PDZ domain proteins and PDZ ligands. *Genome Res* 16:1056–1072
- Wiedemann U, Boisguerin P, Leben R, Leitner D, Krause G, Moelling K, Volkmer-Engert R, Oschkinat H (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J Mol Biol* 343:703–718
- Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R (1996) Crystal structures of a complexed and peptide-free membrane protein–binding domain: molecular basis of peptide recognition by PDZ. *Cell* 85:1067–1076
- Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaya LA, MacBeath G (2007) PDZ domain binding selectivity is optimized across the mouse proteome. *Science* 317:364–369
- Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G (2008) Predicting PDZ domain–peptide interactions from primary sequences. *Nat Biotechnol* 26:1041–1045
- Kaufmann K, Shen N, Mizoue L, Meiler J (2011) A physical model for PDZ-domain/peptide interactions. *J Mol Model* 17:315–324
- Saro D, Klosi E, Paredes A, Spaller MR (2004) Thermodynamic analysis of a hydrophobic binding site: probing the PDZ domain with nonproteinogenic peptide ligands. *Org Lett* 6:3429–3432
- Saro D, Li T, Rupasinghe C, Paredes A, Caspers N, Spaller MR (2007) A thermodynamic ligand binding study of the third PDZ domain (PDZ3) from the mammalian neuronal protein PSD-95. *Biochemistry* 46:6340–6352
- Krivov GG, Shapovalov MV, Dunbrack RL Jr (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77:778–795
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285:1735–1747
- Knapp B, Omasits U, Schreiner W (2008) Side chain substitution benchmark for peptide/MHC interaction. *Protein Sci* 17:977–982
- Zhou P, Tian F, Lv F, Shang Z (2009) Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* 76:151–163
- 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 (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012
- Tian F, Yang L, Lv F, Luo X, Pan P (2011) Why OppA protein can bind sequence-independent peptides? A combination of QM/MM, PB/SA, and structure-based QSAR Analyses. *Amino Acids* 40:493–503
- Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels–Alder reactions and Pt(P(t-Bu)₃)₂ + H₂ oxidative addition. *J Phys Chem* 100:19357–19363
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (2003) Gaussian 03. Gaussian Inc., Wallingford
- Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc* 107:3902–3909
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
- Li Y, Yang Y, He P, Yang Q (2009) QM/MM study of epitope peptides binding to HLA-A*0201: the roles of anchor residues and water. *Chem Biol Drug Des* 74:611–618
- Zhou P, Zou J, Tian F, Shang Z (2009) Fluorine bonding—how does it work in protein–ligand interactions? *J Chem Inf Model* 49:2344–2355
- Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. *Nature* 319:199–203
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
- Retegan M, Milet A, Jamet H (2009) Exploring the binding of inhibitors derived from tetrabromobenzimidazole to the CK2 protein using a QM/MM-PB/SA approach. *J Chem Inf Model* 49:963–971
- Dubey K, Ojha R (2011) Binding free energy calculation with QM/MM hybrid methods for Abl-Kinase inhibitor. *J Biol Phys* 37:69–78
- Rocchia W, Alexov E, Honig B (2001) Extending the applicability of the nonlinear Poisson–Boltzmann equation: multiple

- dielectric constants and multivalent ions. *J Phys Chem* 105:6507–6514
36. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320
 37. Lovell SC, Word JM, Richardson JS, Richardson DC (2000) The penultimate rotamer library. *Proteins* 40:389–408
 38. Zhou P, Tian F, Shang Z (2009) 2D depiction of nonbonding interactions for protein complexes. *J Comput Chem* 30:940–951
 39. Creamer TP (2000) Side-chain conformational entropy in protein unfolded states. *Proteins* 40:443–450
 40. Bader RFW (1990) *Atoms in molecules: a quantum theory*. UK, Oxford University Press
 41. Wang W, Tian A, Wong NB (2005) Theoretical study on the bromomethane–water 1:2 complexes. *J Phys Chem A* 109:8035–8040
 42. Reed AE, CuritssLA, WeinholdF (1988) Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem Rev* 88:889–926
 43. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. *Phys Rev* 46:618–622
 44. Biegler-Konig F, Schoenbohm J (2002) *AIM2000*, 2.0 ed. Buro fur Innovative Software, Bielefeld, Germany
 45. Glendening ED, Badenhoop JK, Reed AE, Carpenter JE, Bohmann JA, Morales CM, Weinhold F (2001) *NBO 5.0*. Theoretical Chemistry Institute, University of Wisconsin, Madison, WI
 46. Cortes C, Vapnik V (1995) Support vector networks. *Mach Learn* 20:273–293
 47. Zhou P, Tian F, Chen X, Shang Z (2008) Modeling and prediction of binding affinities between the human amphiphysin SH3 domain and its peptide ligands using genetic algorithm-Gaussian processes. *Biopolymers (Pept Sci)* 90:792–802
 48. Zhou P, Chen X, Wu Y, Shang Z (2010) Gaussian process: an alternative approach for QSAM modeling of peptides. *Amino Acids* 38:199–212
 49. Zhou P, Tian F, Lv F, Shang Z (2009) Comprehensive comparison of eight statistical modelling methods used in quantitative structure–retention relationship studies for liquid chromatographic retention times of peptides generated by protease digestion of the *Escherichia coli* proteome. *J Chromatogr A* 1216:3107–3116
 50. Matta CF, Castillo N, Boyd RJ (2005) Characterization of a closed-shell fluorine–fluorine bonding Interaction in aromatic compounds on the basis of the electron density. *J Phys Chem A* 109:3669–3681
 51. Nakanishi W, Hayashi S, Narahara K (2008) Atoms-in-molecules dual parameter analysis of weak to strong interactions: behaviors of electronic energy densities versus Laplacian of electron densities at bond critical points. *J Phys Chem A* 112:13593–13599
 52. Petsalaki E, Stark A, Garcia-Urdiales E, Russell RB (2009) Accurate prediction of peptide binding sites on protein surfaces. 5:e1000335