ORIGINAL PAPER

A physical model for PDZ-domain/peptide interactions

Kristian Kaufmann · Nicole Shen · Laura Mizoue ·
Jens Meiler

Abstract The PDZ domain is an interaction motif that
recognizes and binds the C-terminal peptides of target
proteins. PDZ domains are ubiquitous in nature and help
assemble multiprotein complexes that control cellular
organization and signaling cascades. We present an opti-
mized energy function to predict the binding free energy
($\Delta\Delta G$) of PDZ domain/peptide interactions computation-
ally. Geometry-optimized models of PDZ domain/peptide
interfaces were built using ROSETTA, and protein and
peptide side chain and backbone degrees of freedom are
minimized simultaneously. Using leave-one-out cross-
validation, ROSETTA's energy function is adjusted to
reproduce experimentally determined $\Delta\Delta G$ values with
a correlation coefficient of 0.66 and a standard deviation
of 0.79 kcal mol^{-1}. The energy function places an
increased weight on hydrogen bonding interactions when
compared to a previously developed method to analyze
protein/protein interactions. Binding free enthalpies
($\Delta\Delta H$) and entropies (ΔS) are predicted with reduced
accuracies of $R=0.60$ and $R=0.17$, respectively. The
computational method improves prediction of PDZ
domain specificity from sequence and allows design of
novel PDZ domain/peptide interactions.

K. Kaufmann · N. Shen · J. Meiler (✉)
Department of Chemistry, Vanderbilt University,
7330 Stevenson Center, Station B 351822,
Nashville, TN 37235, USA
e-mail: jens.meiler@vanderbilt.edu
URL: www.meilerlab.org

L. Mizoue
Department of Biochemistry, Vanderbilt University Medical,
Nashville, TN 37235, USA

L. Mizoue · J. Meiler
Center for Structural Biology, Vanderbilt University,
Nashville, TN 37232, USA

Introduction

Protein/peptide interactions play an important biological
role in an array of cellular processes. One motif frequently
used for such interactions is the well characterized PDZ
(PSD-95, Discs large, Zona occludens 1) domain [1].
Within *Homo sapiens*, *Drosophila melanogaster*, and
Caenorhabditis elegans, Schultz et al. [2] have estimated
the existence of 440 PDZ domains in 259 different proteins,
133 PDZ domains in 86 proteins, and 138 PDZ domains in
96 proteins, respectively. PDZ domains perform critical
roles in signaling cascades of bacteria, yeast, plants, and
animals [3] by acting as intracellular scaffolding proteins
[4, 5]. Pathogens disrupt host-signaling processes using
linear peptide motifs to target PDZ binding sites [6].
Developing inhibitors of these interactions is one avenue
of therapeutic development [7]. The wide-spread presence
of the PDZ domain in nature and its integral role in
numerous biological processes and diseases make it an
ideal focus for studying the specificity of protein/peptide
interactions.

PDZ domains bind peptides through strong backbone hydrogen bonds

PDZ domains are typically composed of 80–90 amino acids
[8] and consist of a central bent six-stranded β-sheet

surrounded by two α-helices. The peptide binding interface (Fig. 1) lies at the edge of the β-sheet. The peptide binds in an extended, antiparallel conformation, using the unsatisfied hydrogen bonding capabilities of PDZ β-strand 2 (β2) to extend the β-sheet by one additional strand. The ligand also engages in side-chain interactions with the second α-helix (α2) of the PDZ domain, which lines the other side of the binding groove. The binding pocket contains a characteristic hydrophobic loop (β1:β2) that binds the peptide carboxy-terminus through the formation of three hydrogen bond interactions. Overall, the interface is characterized by strong backbone–backbone hydrogen bonding contacts within a hydrophobic environment [9].

PDZ domain specificity is governed by side-chain interactions

Although the general binding mode of PDZ domains is the same, different proteins interact with different targets. While specificity in the PDZ family has been studied extensively, an unambiguous classification of the PDZ domain remains a challenge [6]. Generally, PDZ domains have been grouped into three classes (I, II, and III) depending on the characteristics of the β1:β2 loop [6] and position −2 (P_{-2}) of the

ligand (see Fig. 1). Class I domains have a G-L/Y-G-F β1:β2 loop that binds C-terminal peptide residues of sequence X-S/T-X-V/I/L (P_{-3}-P_0) [9]. Additionally, the peptide hydroxyl group at P_{-2} makes an important hydrogen bonding contact with the histidine side chain of $\alpha 2_1$ [6]. Class II proteins have a similar β1:β2 loop sequence of X-L/V-G-F/I/L that binds peptide sequences having a hydrophobic amino acid at position P_{-2} (X-ø-X-ø) [8]. Class III domains are less widespread and have a G-L-G-F β1:β2 loop sequence that binds peptides having an acidic amino acid at P_{-2} (X-D/E-X-ø) [10, 11].

PDZ class I, II, and III proteins and their peptides have variable sequence similarities (between 5% and 90%) but are structurally highly similar. Indeed, Stiffler et al. [12] found only a weak correlation between sequence identity and PDZ domain specificity. Instead, Stiffler developed a modified position specific scoring matrix based on the profiles of peptides which bind to a domain. Chen et al. [13] later developed a method that incorporated structural information on protein/peptide residue pairs within close proximity of each other. The model was capable of predicting PDZ domain specificity for multiple species from primary sequences and it was argued that including structural information via the protein/peptide residue
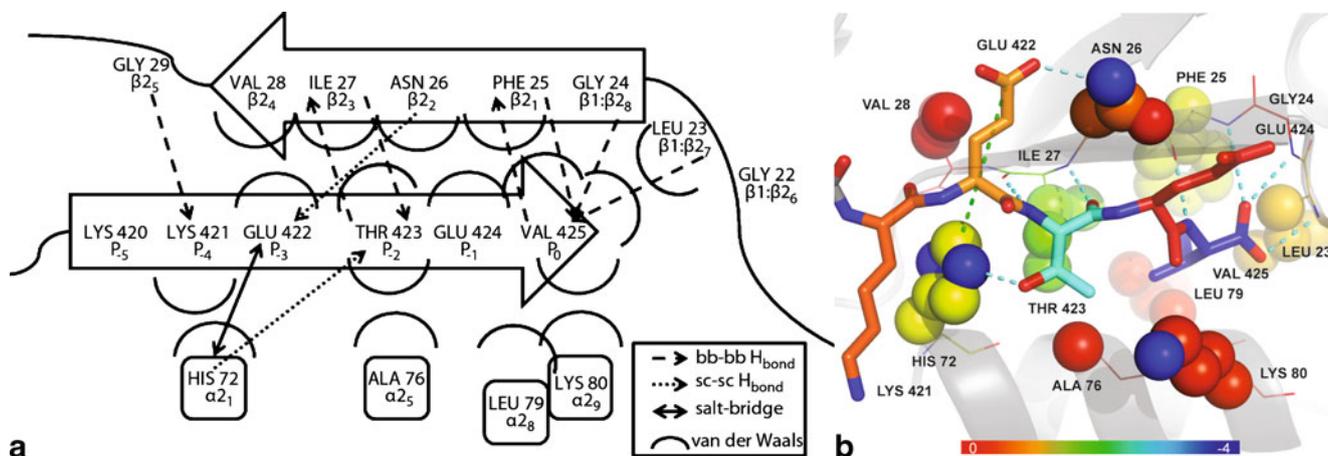

Fig. 1a,b Binding site of PSD-95 a class I domain. PDZ (PSD-95, Discs large, Zona occludens 1) domains bind peptides strongly through backbone–backbone hydrogen bonds; *dashed lines* in (**a**) indicate these interactions while *dotted lines* indicate protein–peptide side chain–side chain hydrogen bonds, and the *arrow* points from the hydrogen-donating nitrogen to the oxygen acceptor. Color in (**b**) illustrates each residue's overall energetic involvement in binding the peptide, summing the weighted ROSETTA energy function of the individual attractive, solvation, repulsive, rotamer, pairwise, and hydrogen bonding energy contributions (ΔΔG values). In (**b**), the strong backbone–backbone hydrogen bonds (shown as *blue dashed lines*) between the PDZ3 protein and peptide residues V425 and T423 highlight the known PDZ3 protein preference for X-Thr/Ser-X-Val-COO(−) peptides [4]. This is further emphasized by E424 and K421's lack of strong backbone-backbone hydrogen bonds and *orange* and *dark yellow* residue colors, which have overall 0.1 and −0.3 ΔΔG

values, respectively. This is in contrast to T423 and V425's *teal* and *dark blue* color and respective −3.0 and −3.1 ΔΔG values. In addition to hydrogen bonding interactions, important salt bridge (shown as *green dashed lines*) and van der Waals interactions (protein side chains contributing to van der Waals interactions shown with *spheres*) are involved in binding the peptide tightly between a beta sheet and alpha helix. H72 of the alpha helix seems to be the most involved in binding the peptide, forming a salt bridge with E422, interacting with T423 through side chain–side chain hydrogen bonding, and engaging in van der Waals interactions with K421. L79 and K80 similarly interact with V425 through van der Waals interactions, experiencing the attractive part of the Lennard-Jones potential energy curve. This probably creates a favorable, hydrophobic environment for the non-polar, valine side chain and amplifies the strength of the hydrogen bonds formed between the peptide carboxy terminus and P25, G24, and L23

position specific interaction matrix was sufficient to predict the specificity of PDZ domains.

PDZ domains display a diverse and finely tuned specificity profile

PDZ domain classification can be extended beyond the three naïve classes discussed here. Specificity within these classes depends upon other differences in the protein/peptide interface that result in a diversified sequence profile. Tonikian et al. [6] performed profiling of 91 point mutants of a model PDZ domain to create a specificity map. Using this map, 82 protein domains of the PDZ family were reclassified into 16 classes distinguished by specificity for peptide residues up to the P_{-6} position. While sequence-based analysis alone reveals diverse specificity profiles, the inclusion of structure-based information should provide a more general model for predicting PDZ specificity. Such a physical model would be a useful tool for PDZ domain classification, specificity prediction, and design.

The ROSETTA protein modeling software predicts specificity of protein/protein interfaces

In a series of experiments, Kortemme et al. [14] demonstrated the power of the knowledge-based energy function of the modeling software ROSETTA to characterize and design protein/protein interfaces. A model for protein/protein binding was created using a data set of alanine mutants at protein/protein interfaces. The model was able to predict successfully the results of alanine scanning experiments on globular proteins (743 mutations) and 19 protein/protein interfaces (233 mutations) with low standard deviations of 0.8 kcal mol^{-1} and 1.1 kcal mol^{-1}, respectively [14]. The model was applied to create new DNase-inhibitor protein pairs with altered specificities that functioned both in vitro and in vivo [15]. It was also used to fuse domains of two homing endonucleases, creating a chimera that recognized a new DNA target and functioned as a highly specific artificial endonuclease [16].

While this model proved successful in modeling protein/protein interfaces, the derived parameterization is not optimal for protein/peptide interfaces as these are characterized by distinct features that require a tailored parameterization, such as smaller hydrophobic surface area and a greater dependence of hydrogen bonding interactions. Sood and Baker explored the use of ROSETTA to design elongated p53 and dystroglycan-based peptides that bind with increased affinity to Mdm2 oncoprotein and dystrophin, respectively. These studies included backbone flexibility and allowed side-chain

flexibility through repacking of a rotamer library but used the standard ROSETTA energy function with a packing score derived from the change in solvent accessible surface area [17]. Sood and Baker found that sampling of the backbone conformation improved recovery of sequence diversity in designed peptides and in cases where the algorithm fails, insufficient sampling of backbone degrees of freedom explains the error.

A ROSETTA parameterization tailored for PDZ domain/peptide interfaces

It is the objective of the present work to develop a model for predicting the specificity of PDZ domains using the protein structure prediction program ROSETTA. Saro et al. [18] conducted isothermal titration calorimetry (ITC) measurements on a series of peptides binding the third PDZ domain (PDZ3) of postsynaptic density 95 protein (PSD-95), a class I domain. They recorded the thermodynamic properties $\Delta\Delta G$, $\Delta\Delta H$, and $T\Delta S$ for a series of six-residue peptides of sequence (X-X-X-T-X-V), with different X amino acids influencing binding. We parameterize ROSETTA to accurately predict these thermodynamic parameters.

Methods

Dataset for energy function parameterization

The dataset contains free energy ($\Delta\Delta G$), enthalpy ($\Delta\Delta H$), and entropy ($T\Delta S$) measurements for binding of 28 peptides to the PDZ3 domain of PSD-95 (Table 1) [18]. The crystal structure of the PDZ3 domain of PSD-95 with the highest resolution (1.54 Å) from the PDB was used for structural modeling (PDBID 1TP5).

The crystal structure was determined in complex with the peptide KKETWV.

Introduction of mutations and initial minimization of structural models

ROSETTADESIGN [19] protocols allow in silico mutation of amino acids. Briefly, the side chain of the amino acid in question is removed and replaced with the side chain of the target amino acid. The conformation of the introduced amino acid is chosen from a backbone-dependent rotamer library [20] to minimize the ROSETTA energy function. First, the tryptophan at position P_{-1} of 1TP5 was reverted to a glutamate to match the base peptide KKETEV used in the study by Saro et al. [18]. Following this modification, the 28 PDZ domain/peptide complexes were built (Table 1, Fig. 2b). All models underwent gradient minimization

Table 1 Thermodynamic parameters determined experimentally by Saro et al. [18]. Binding energy changes to point mutations on the native peptide, KKETEV were determined using isothermal titration calorimetry (ITC) and represent the average of at least two independent experiments

Peptide	K_d (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)
1. KKETEV	1.9±0.1	−7.8±0.1	−6.2±0.1	1.6±0.1
2. KKETEA	91.0±2.0	−5.5±0.1	−4.6±0.2	0.9±0.2
3. KKETEL	7.9±1.3	−7.0±0.1	−4.1±0.3	2.9±0.2
4. KKETEI	7.7±1.2	−7.0±0.1	−4.3±0.2	2.7±0.1
5. KKETEM	21.0±2.0	−6.4±0.1	−6.8±0.2	−0.4±0.1
6. KKETEF	57.0±2.0	−5.8±0.1	−4.4±0.4	1.4±0.4
7. KKETET	105.0±6.0	−5.4±0.1	−5.9±0.2	−0.5±0.2
8. KKESEV	6.6±0.9	−7.1±0.1	−4.8±0.1	2.3±0.2
9. KKECEV	72.0±7.0	−5.7±0.1	−1.7±0.1	4.0±0.2
10. KKESEL	33.0±2.0	−6.1±0.1	−4.0±0.1	2.1±0.1
11. KKESEI	24.0±6.0	−6.3±0.2	−5.0±0.2	1.3±0.4
12. KKESEF	98.0±16.0	−5.5±0.1	−3.1±0.1	2.4±0.1
13. KKETGV	2.4±0.0	−7.7±0.1	−5.7±0.2	2.0±0.2
14. KKETAV	0.5±0.1	−8.7±0.1	−5.3±0.4	3.4±0.4
15. KKETVV	1.3±0.2	−8.1±0.1	−5.9±0.1	2.2±0.1
16. KKETLV	1.8±0.3	−7.8±0.1	−3.7±0.4	4.1±0.3
17. KKETPV	0.9±0.2	−8.2±0.1	−4.3±0.1	3.9±0.2
18. KKETWV	2.8±0.4	−7.6±0.1	−3.5±0.2	4.1±0.1
19. KKETDV	20.0±2.0	−6.4±0.1	−4.1±0.3	2.3±0.3
20. KKETKV	1.2±0.0	−8.1±0.1	−5.6±0.6	2.5±0.6
21. KKGTEV	80.0±3.0	−5.6±0.1	−2.7±0.1	2.9±0.1
22. KKATEV	21.0±4.0	−6.4±0.1	−2.4±0.1	4.0±0.2
23. KKQTEV	4.0±0.0	−7.4±0.1	−4.9±0.3	2.5±0.3
24. KKDTEV	85.0±12.0	−5.6±0.1	−3.9±0.3	1.7±0.2
25. KKKTEV	27.0±4.0	−6.2±0.1	−2.7±0.3	3.5±0.4
26. KKGTGV	273.0±30.0	−4.9±0.1	−2.6±0.3	2.3±0.2
27. KKATAV	8.3±1.5	−6.9±0.1	−3.0±0.1	3.9±0.2
28. YKETEV	1.2±0.1	−8.1±0.1	−6.9±0.1	1.2±0.2

using ROSETTA to remove initial clashes (Fig. 2c) [21]. The protocol involves eight rounds of gradient-based minimization of all torsional degrees of freedom, which is alternated with side chain repacking using a rotamer library. The all-atom RMSD of the structure changed by 0.40 Å on average with a maximum of 0.51 Å observed for complex 6 containing the KKETEF mutant peptide.

ROSETTADOCK generation of structural models for protein/peptide complexes

To generate minimized models for energy evaluation, all bound structures underwent a small perturbation protocol applied to the transformational degrees of freedom in the protein/peptide complex using ROSETTADOCK (Fig. 2e)

[22]. This rigid body motion is complemented by a simultaneous optimization of side chain coordinates through a fast repacking protocol. The backbone coordinates of protein and peptide are held fixed in the process.

The protocol is setup in an iterative fashion. First a random small perturbation of up to 0.1 Å translation and up to 2° rotation is made to the rigid body degrees of freedom. Then the side chain conformations are allowed to change
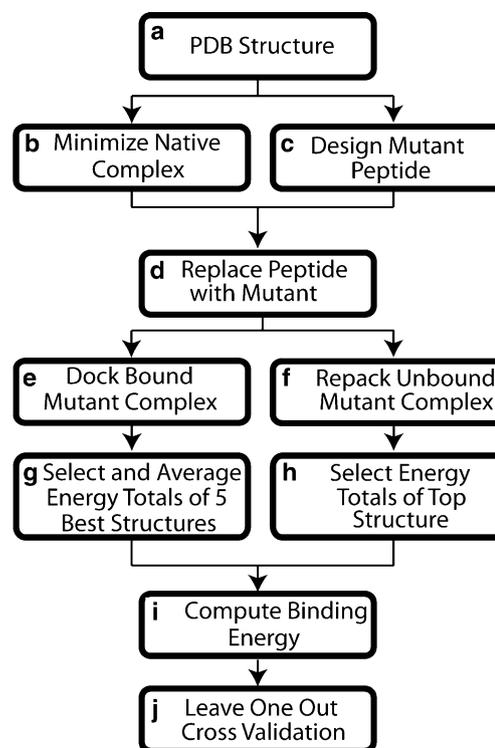

Fig. 2 Procedural flowchart. Protein and peptide mutants of the PSD-95 PDZ3 domain, 1TP5, were each downloaded (**a**) and minimized (**c**) in ROSETTA to remove initial clashes. Experimental mutants were reproduced in the computational analysis using design (**b**). Mutant proteins and peptides were combined with minimized peptides and proteins, respectively (**d**). These bound mutants were docked (**e**), yielding 100 decoys, of which the 5 structures with the lowest energy were selected and averaged over selected energy components (**g**), including attractive, solvation, repulsive, Dunbrack, residue pair electrostatics, hydrogen bonding, amino acid reference energies. The protein structure with the peptide far removed was repacked over selected residues determined from docking the structure (**f**). Because the 100 unbound structures had the same total ROSETTA energy, a single structure's energy values were used rather than the average of five structures (**h**). The unbound energy values for each structure were subtracted from the corresponding, averaged bound energy values of the structure (**i**), producing $\Delta\Delta E$ values for each energy term, which were weighted and summed to produce the overall change in energy due to the protein binding the peptide, which was correlated to the experimentally determined binding energy, yielding the best correlation with particular weights (**j**). For the specificity analysis, 17 PDB files were downloaded (**a**) and separated into their fundamental protein, peptide components. All possible protein–peptide complexes were combined and minimized (**c**). Steps (**d**) through (**i**) as previously discussed were followed. Energy terms from (**i**) were weighted using weights determined from the mutational investigations (**j**)

by substituting discrete rotamers from a library of conformations commonly seen in the PDB. If the substitution results in a lower total energy, ROSETTA keeps the new conformation of the protein. If the energy is higher, ROSETTA may still accept the substitution with a probability inversely proportional to the energy increase (Metropolis criterion). On average, around 50 of these iterations are completed in order to find the best combination of amino acid side chain conformations. The output model is the lowest energy complex observed throughout the entire trajectory. Lastly, a gradient-based minimization on the rigid body degrees of freedom moves the final model into the nearest local minimum in the ROSETTA energy landscape. A total of 100 bound models were generated for each complex. The 5 models with the lowest overall energy were selected for further analysis (Fig. 2g).

Modeling apo structures in ROSETTA

The unbound (apo) structures were created by removing the peptide from the binding pocket and away from the protein by a distance sufficiently large to prevent any interaction (> 100 Å). The side chains, which were allowed to move during the docking protocol, were allowed to rearrange using repacking algorithms [19]. A total of 100 models was generated for each of the mutants, and the total ROSETTA energy was used to select a single most favorable unbound conformation for each of the 28 complexes (Fig. 2h).

Calculation and evaluation of binding free energy

The ROSETTA energy function contains six energy terms. Van der Waals energies are modeled using a Lennard-Jones 12-6 potential. The potential is split into an attractive (atr) and a repulsive (rep) component. ROSETTA introduces a solvation energy (sol) that imposes a penalty for polar atoms buried in the core of a protein accounting for the exposure preferences of polar and non-polar atoms [23]. Side chain conformational probabilities are reflected by an energy (dun) derived from rotamer probabilities [20]. Electrostatic interactions are mimicked by a knowledge-based pair-wise potential (pair) derived from statistics over the PDB. Hydrogen bonds (hbnd) are captured by an orientation dependent potential [24]. Note that, in the past, hydrogen bonds have been classified into three classes: long-range backbone–backbone (lr-bb), backbone–side chain (bb-sc), and side chain–side chain (sc-sc) hydrogen bonds [14].

Within each structure, all residues were individually evaluated. To obtain the total energy of the model, the sum over all amino acids was computed and averaged over the top five bound structures (Fig. 2g). For the unbound models energies from the single structure with lowest ROSETTA

energy were used directly (Fig. 2h). The binding free energy was computed for each of the above-mentioned terms $\Delta\Delta E_{binding}^{term}$ using:

$$\Delta\Delta E_{binding}^{term} = \frac{1}{5}\sum_{i=1}^{5}\Delta E_{i\ bound}^{term} - \Delta E_{unbound}^{term} \qquad (1)$$

$\Delta E_{i\ bound}^{term}$ is the ROSETTA energy one of the five complex models, respectively; $\Delta E_{unbound}^{term}$ is the ROSETTA energy of the single unbound model.

Multiple linear regression is used to parameterize an overall free energy function

To obtain an energy function optimized for the analysis of protein/peptide interactions (Fig. 2i), a multiple linear regression (MLR) analysis was used. Each of the $\Delta E_{binding}^{term}$ terms is affiliated with a weight w^{term}:

$$\Delta\Delta E_{binding} = w^{atr}\Delta\Delta E_{binding}^{atr} + w^{rep}\Delta\Delta E_{binding}^{rep} + w^{sol}\Delta\Delta E_{binding}^{sol}$$
$$+ w^{pair}\Delta\Delta E_{binding}^{pair} + w^{dun}\Delta\Delta E_{binding}^{dun} + w^{hbnd}\Delta\Delta E_{binding}^{hbnd} + bias \qquad (2)$$

The bias is introduced to account for contributions to the binding free energy not represented in the ROSETTA energy function, such as the loss in entropy. The bias assumes that these contributions are constant, an obvious limitation of the present model.

The weights were determined by performing a leave-one-out (LOO) cross validation analysis. In a round-robin setup, 27 of the 28 mutants with known experimental binding affinities were used to determine an optimal weight set given these 27 data points. Afterwards, the binding free energy of the 28th mutant was predicted and compared with the experiment to enter a correlation analysis. This experiment was repeated for all 28 mutants.

To determine whether an energy term contributes significantly to an optimal energy function for protein/peptide interfaces, energy terms were removed systematically. The subset of energy terms that resulted in the optimal correlation coefficient within the cross-validation experiment was used. The final weight set reported consists of the average weights and standard deviations over of all 28 experiments. The protocol was implemented using the MATHEMATICA software package (Fig. 2j).

Results

The physical model for protein/peptide interactions depends on van der Waals, solvation, and hydrogen bonding

The optimal weight set was determined by a LOO cross validation analysis as described in the Methods section. Of

the six ROSETTA energy terms considered, only van der Waals attraction (atr), solvation (sol), and hydrogen bonding energies (hbnd), contributed to an energy function that optimally reproduced experimentally determined binding free energies:

$$\Delta\Delta E_{binding} = 0.47 \times \Delta\Delta E_{binding}^{atr} + 0.40 \times \Delta\Delta E_{binding}^{sol}$$
$$+ 1.34 \times \Delta\Delta E_{binding}^{hbnd} + 3.90 \qquad (3)$$

The correlation coefficient for the independent dataset is 0.66 (Fig. 3).

Known characteristics of the PDZ binding domain are mirrored within the model

Figure 1 displays the per amino acid changes in free energy upon peptide binding for the PSD-95 PDZ3 in complex with the peptide KKETEV as determined by our model. Strong backbone–backbone hydrogen bonds between the class I domain and the peptide residues $V(P_0)$ and $T(P_{-2})$ agree with the anti-parallel β-strand binding motif of the PDZ domain that forms two backbone hydrogen bonds for every other
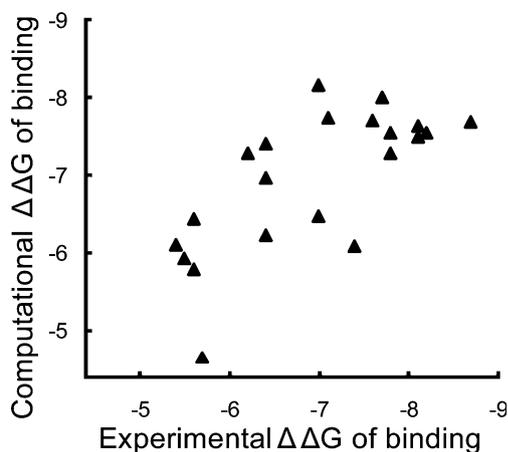

Fig. 3 Correlation of experimentally and computationally measured ΔΔG values over peptide mutants of the PDZ3 domain. Experimentally calculated binding energies were determined using isothermal titration calorimetry (ITC) [18]. Computational binding energies were determined after a leave-one-out (LOO) cross validation analysis of the summed calculation of the various combinations of the weighted changes in the attractive (atr), solvation (sol), repulsive (rep), residue pair electrostatics (pair), dunbrack (dun), and hydrogen bonding (hb and hb_bb) energy terms over all the residues due to the protein binding the peptide. The individual binding term energy changes were calculated using Eq. 1. Different combinations of these terms were weighted and totaled according to Eq. 3. By calculating these weights when each mutant was left out and then applying the determined function, the LOO cross validation analysis measures the weighted energy function's predicting power. The predicted binding energies (y-axis) for the various peptide mutants correlated nicely with the experimentally observed binding energies (x-axis), having an r value equal to 0.66. The overall computational function $f(x) = 0.47*E_{atr} + 0.40*E_{sol} + 1.34 E_{hbnd} + 3.90$ indicates the importance of the attractive, solvation, and side chain hydrogen bonding energy terms

amino acid. In this particular case, the C-terminal amino acid $V(P_0)$ engages in three hydrogen bonds. This alternative pattern is further highlighted by $E(P_{-1})$ and $E(P_{-3})$, which contribute only 0.1 and -0.3 kcal mol^{-1} to the binding free energy, respectively. In contrast, $V(P_0)$ and $T(P_{-2})$ contribute -3.1 and -3.0 kcal mol^{-1}, respectively.

In addition to hydrogen bonding interactions, important salt bridges and van der Waals interactions are involved in binding the peptide tightly. $H(\alpha 2_1)$ is the most important residue within α2 for binding the peptide as it forms a hydrogen bond with $T(P_{-2})$ through side chain hydrogen bonding, and engages in van der Waals interactions with $K(P_{-4})$. $L(\alpha 2_8)$ and $K(\alpha 2_9)$ interact with $V(P_{-0})$ through van der Waals attractive interactions. This creates a favorable, hydrophobic environment for the non-polar valine side chain and amplifies the strength of the hydrogen bonds formed between the peptide carboxyl terminus and $F(\beta 2_1)$, $G(\beta 1:\beta 2_8)$, and $L(\beta 1:\beta 2_7)$.

Enthalpic and entropic contributions to the binding free energy map to different components of the ROSETTA energy function

The investigation was extended to other thermodynamic characteristics of protein/peptide binding including enthalpy and entropy (Table 2). The independent correlation observed for the binding enthalpy $\Delta\Delta H_{binding}$ is with 0.60 only slightly reduced from the value observed for the Gibbs binding free energy (0.66, Fig. 3). In contrast, when correlating with respect to experimentally measured entropy changes the independent correlation drops to 0.17.

Specificity prediction for 12 PDZ domains with available crystal structures

For the specificity analysis, a set of 12 PDZ protein/peptide complexes with available crystal structures was used (Table 3). All experimentally determined structures with resolutions of 2.30 Å or better were considered (PDBID 1BE9 [10], 1N7F [25], 1OBY [26, 27], 1RZX [28], 1TP3, 1TP5, 1V1T [29], 1W9E [29], 1W9O [29], 1W9Q [30, 31],

Table 2 Weighted energy terms over thermodynamic binding properties

correlation	atr	rep	sol	hbnd	rotamer	pair
ΔΔG 0.66	0.47± 0.04	0.00± 0.00	0.40± 0.06	1.34± 0.07	–	–
ΔΔH 0.60	–	–	–	2.25± 0.12	–	1.28± 0.16
ΔΔS 0.17	–	–	–	0.74± 0.10	0.36± 0.03	–

Table 3 Specificity data set

PDB ID	PDZ class	Peptide sequence[a]	Resolution (Å)	Domain
1TP5	1	K**KETWV**	1.54	PSD95-3
1BE9	1	**KQTSV**	1.82	PSD95-3
1TP3	1	**KKETPV**	1.99	PSD95-3
1RZX	1	V**KESLV**	2.10	Par-6B
2I04	1	RR**RETQV**	2.15	MAGI1-1
2QT5	1	NNLQ**DGTEV**	2.30	GRIP1-12
1N7F	2	ATV**RTYSC**	1.80	GRIP1-6
1W9E	2	**TNEFYF**	1.56	Syntenin-2
1W9Q	2	**TNEFAF**	1.70	Syntenin-2
1V1T	2	**TNEYKV**	1.80	Syntenin-2
1W9O	2	**TNEYYV**	2.25	Syntenin-2
1OBY	2	**TNEFYA**	1.85	Syntenin-2

[a] Bold letters indicate amino acids that were used for specificity prediction (P_0-P_{-4})

2I04 [32], 2QT5 [33]). Structures used in the specificity analysis were initially separated into their protein and peptide components. Peptides were truncated to include five carboxy-terminal residues. All possible combinations between PDZ domains and peptides were created yielding a total of 144 complexes. Each complex was refined using the protocol described above (Fig. 2).

The binding energies for each complex were then computed using the PDZ optimized weight set. The heat map in Fig. 4a shows that the PDZ optimized weight set captures specificity within each PDZ class. The complexes group into two blocks reflecting the two classes of PDZ domains. Figure 4b shows the receiver operating characteristics (ROC) curve where a complex is regarded as a true complex if both peptide and protein come from the same PDZ class. The area under the curve is 78%, 28% better than a random predictor.

Discussion

Energy function weights from LOO analysis are stable

The deviations from a perfect correlation are attributed to imperfection in the ROSETTA energy function, which is simplified to only contain pair-wise decomposable energetic terms [34]. The small standard deviations observed for the individual weights (Table 4) demonstrate internal consistency as the analysis of all 28 complexes yielded very similar weight sets.

Table 4 compares the weight set obtained in the present analysis with a weight set optimized for protein/protein interfaces [35] and the default weight set. Specializing the energy function should allow prediction of binding affinities at a higher accuracy than seen for a more generalized function such as that put forth by Kortemme [14]. The weights derived here deviate from the default ROSETTA

weights, highlighting the importance of a customized computational model for protein/peptide binding. The weight set is optimized to predict the binding free energies PDZ domains. In particular, the hydrogen bonding weight is increased substantially relative to other weights. This result can be explained in part by the backbone hydrogen bonds between peptide and PDZ domain. These hydrogen bonds contribute significantly to the stability of the PDZ-domain/peptide interface. However, as these hydrogen bonds are present in all PDZ domain/peptide complexes, they do not govern specificity but contribute an approximately equal amount to all interfaces studied.

Our results indicate that a high weight on side chain hydrogen bonds is particularly important for accurate specificity prediction. Interestingly, a holistic weighting with a single hydrogen bonding weight gave the best results. This is in contrast to the earlier reported optimal weight set for protein/protein interfaces where hydrogen bonds contributed differently depending on the level of solvent exposure [35]. Beyond this aspect, changes in the weight set are small.

Energy function components capture enthalpic but not entropic contributions

The weighted energy terms vary significantly when correlated to the different thermodynamic binding properties. Enthalpy is best predicted from hydrogen bonding (hbnd) and electrostatic (pair) interactions; entropy correlates best with a combination of hydrogen bonding (hbnd) and rotamer probability. Overall we expected that ROSETTA-derived energy terms correlate best with binding free energies. Their knowledge-based character can be well aligned with the definition of free energy in statistical thermodynamics. Hence, every one of the ROSETTA energy terms contains both entropic contributions and enthalpic contributions. However, the term can be dominated by one
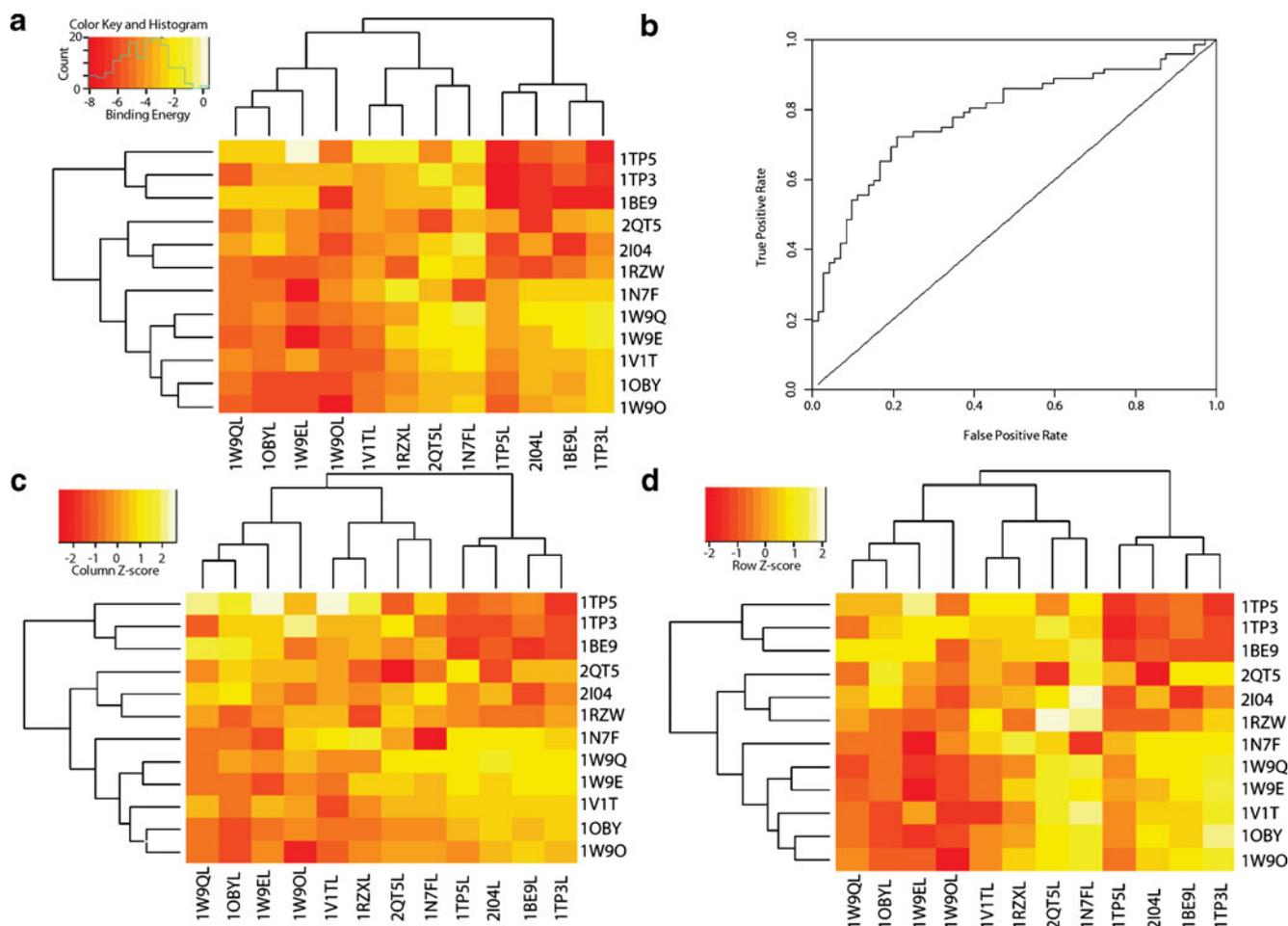
Fig. 4 Specificity based on computed binding energy. Each column shows the computed binding energies of the peptide from a structure to each of the PDZ domains. Each row displays the computed binding energies of each peptide to a given PDZ structure. **a** Heat map with colors scaled according to the raw computed binding energy. **b** Receiver operating characteristics (ROC) curve for PDZ classification. **c** Heat map of the binding energies colored by the z-score computed to the peptide group (i.e., within column). **d** Coloring scaled according to the z-score computed by the PDZ structure group (within row)

of the two if it is better represented by the simplified two-body equations used within ROSETTA. Our results demonstrate that entropic contributions are least accurately reflected and prevent ROSETTA from predicting to higher degrees of accuracy.

Computed binding energies correctly classify PDZ domains

The correlation of the binding energies within each class is apparent, but the computed binding energies across all PDZ complexes do not accurately rank the complexes. However

Table 4 Weight set optimized for protein/peptide interfaces compared to a weight set optimized for protein/protein interfaces [35] and to the default weight set. *sc* Side chain, *bb* backbone, *atr* attractive component of van der Waals energy, *rep* repulsive component of van der Waals energy, *sol* implicit solvation energy, *hbnd* hydrogen bonding, *rotamer* knowledge based energy for conformation for a side chain

	atr	rep	sol	hbnd			Rotamer
Protein/peptide	0.47±0.04	0.00	0.40±0.06			1.34±0.07	0.00
Protein/protein	0.44	0.07	0.32	sc-bb		0.49	0.28
				sc-sc	exposed	0.16	
				sc-sc	intermediate	0.44	
				sc-sc	buried	0.94	
ROSETTA default	0.42	0.10	0.37			0.24	0.06

when holding either the protein or the peptide constant, the binding energies display a better correlation with specificity as seen in Fig. 4c, d. This may reflect the need to sample a greater conformational space. In fact, Sood and Baker [17] found a better recovery of peptides sequence profiles upon introducing backbone flexibility into their design protocol.

Summary

This study presents a physical model for PDZ domain/peptide interactions. Parameterization of the ROSETTA energy function was achieved by fitting a linear model to experimentally determined binding free energies for 28 PDZ domain/peptide complexes. The energy function is dominated by van der Waals attractive, solvation, and hydrogen bonding interactions. It reproduces well-known determinants of PDZ domain/peptide interactions such as an alternating pattern of backbone hydrogen bonding to the second strand of the PDZ domain (β2) and side chain interactions with the second helix (α2). While the Gibbs free energy correlates well with experimental values (R= 0.66), correlation of enthalpy (R=0.60) and particularly entropy (R=0.17) is reduced. This reduction is attributed to the knowledge-based nature of ROSETTA energy functions, which aligns well with the definition of free energy in statistical mechanics. The resulting weight set was able to classify a given PDZ/peptide complex 28% better than a random predictor.

References

1. Kim E, Sheng M (2004) PDZ domain proteins of synapses. Nat Rev Neurosci 5:771–781
2. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 28:231–234
3. Ponting CP (1997) Evidence for PDZ domains in bacteria, yeast, and plants. Protein Sci 6:464–468
4. Kurschner C, Yuzaki M (1999) Neuronal interleukin-16 (NIL-16): a dual function PDZ domain protein. J Neurosci 19:7770–7780
5. Pawson T, Scott JD (1997) Signaling through scaffold, anchoring, and adaptor proteins. Science 278:2075–2080
6. 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

7. Dev KK (2004) Making protein interactions druggable: targeting PDZ domains. Nat Rev Drug Discov 3:1047–1056
8. Hung AY, Sheng M (2002) PDZ domains: structural modules for protein complex assembly. J Biol Chem 277:5699–5702
9. Nourry C, Grant SG, Borg JP (2003) PDZ domain proteins: plug and play! Sci STKE 2003:RE7
10. 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
11. Stricker NL, Christopherson KS, Yi BA, Schatz PJ, Raab RW, Dawes G, Bassett DE Jr, Bredt DS, Li M (1997) PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. Nat Biotechnol 15:336–342
12. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G (2007) PDZ domain binding selectivity is optimized across the mouse proteome. Science 317:364–369
13. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G (2008) Predicting PDZ domain-peptide interactions from primary sequences. Nat Biotechnol 26:1041–1045
14. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci USA 99:14116–14121
15. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D (2004) Computational redesign of protein–protein interaction specificity. Nat Struct Mol Biol 11:371–379
16. Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10:895–905
17. Sood VD, Baker D (2006) Recapitulation and design of protein binding peptide structures and sequences. J Mol Biol 357:917–927
18. 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
19. Liu Y, Kuhlman B (2006) RosettaDesign server for protein design. Nucleic Acids Res 34:W235–W238
20. Dunbrack RL Jr, Karplus M (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 230:543–574
21. Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. Science 309:1868–1871
22. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D (2003) Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 331:281–299
23. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35:133–152
24. Kortemme T, Morozov AV, Baker D (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. J Mol Biol 326:1239–1259
25. Im YJ, Park SH, Rho SH, Lee JH, Kang GB, Sheng M, Kim E, Eom SH (2003) 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 278:8501–8507
26. Kang BS, Cooper DR, Devedjiev Y, Derewenda U, Derewenda ZS (2003) Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. Structure (Camb) 11:845–853
27. Kang BS, Cooper DR, Jelen F, Devedjiev Y, Derewenda U, Dauter Z, Otlewski J, Derewenda ZS (2003) PDZ tandem of

human syntenin: crystal structure and functional properties. Structure 11:459–468

28. Peterson FC, Penkert RR, Volkman BF, Prehoda KE (2004) Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. Mol Cell 13:665–676

29. Grembecka J, Cierpicki T, Devedjiev Y, Derewenda U, Kang BS, Bushweller JH, Derewenda ZS (2006) The binding of the PDZ tandem of syntenin to target proteins. Biochemistry 45:3674–3683

30. von Ossowski I, Oksanen E, von Ossowski L, Cai C, Sundberg M, Goldman A, Keinanen K (2006) Crystal structure of the second PDZ domain of SAP97 in complex with a GluR-A C-terminal peptide. FEBS J 273:5219–5229

31. von Ossowski L, Tossavainen H, von Ossowski I, Cai C, Aitio O, Fredriksson K, Permi P, Annila A, Keinanen K (2006) Peptide binding and NMR analysis of the interaction between SAP97

PDZ2 and GluR-A: potential involvement of a disulfide bond. Biochemistry 45:5567–5575

32. Zhang Y, Dasgupta J, Ma RZ, Banks L, Thomas M, Chen XS (2007) 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 81:3618–3626

33. Long J, Wei Z, Feng W, Yu C, Zhao YX, Zhang M (2008) Supramodular nature of GRIP1 revealed by the structure of its PDZ12 tandem in complex with the carboxyl tail of Fras1. J Mol Biol 375:1457–1468

34. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci USA 97:10383–10388

35. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein–protein interfaces. Sci STKE 2004:pl2