JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

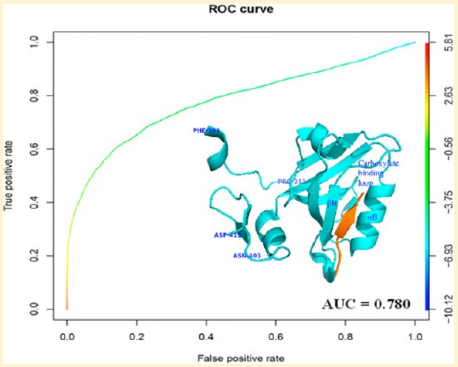
Structure-Based Multiscale Approach for Identification of Interaction Partners of PDZ Domains

Garima Tiwari and Debasisa Mohanty*

Bioinformatics Center, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi—110067, India

Supporting Information

ABSTRACT: PDZ domains are peptide recognition modules which mediate specific protein−protein interactions and are known to have a complex specificity landscape. We have developed a novel structure-based multiscale approach which identifies crucial specificity determining residues (SDRs) of PDZ domains from explicit solvent molecular dynamics (MD) simulations on PDZ−peptide complexes and uses these SDRs in combination with knowledge-based scoring functions for proteomewide identification of their interaction partners. Multiple explicit solvent simulations ranging from 5 to 50 ns duration have been carried out on 28 PDZ−peptide complexes with known binding affinities. MM/PBSA binding energy values calculated from these simulations show a correlation coefficient of 0.755 with the experimental binding affinities. On the basis of the SDRs of PDZ domains identified by MD simulations, we have developed a simple scoring scheme for evaluating binding energies for PDZ−peptide complexes using residue based statistical pair potentials. This multiscale approach has been benchmarked on a mouse PDZ proteome array data set by calculating the binding energies for 217 different substrate peptides in binding pockets of 64 different mouse PDZ domains. Receiver operating characteristic (ROC) curve analysis indicates that, the area under curve (AUC) values for binder vs nonbinder classification by our structure based method is 0.780. Our structure based method does not require experimental PDZ−peptide binding data for training.



INTRODUCTION

PDZ domains are 100 amino acids long modular protein domains, which mediate protein−protein interactions by recognizing short peptide stretches on C-terminus of their interaction partners.[1,2] Hence, PDZ domains are also referred as peptide recognition modules similar to SH3, PTB, WW, etc. Since, PDZ domains are involved in a variety of biologically important functions,[3] disruption of these interactions leads to several human diseases like congenital sensorineural deafness, vestibular dysfunction, blindness and Dejerine−Sottas neuropathy, etc.[4−6] In view of the functional importance of PDZ domains, a number of studies have attempted to decipher the substrate specificities of PDZ domains.[7,8] PDZ domains have been divided into three classes, namely class I, class II and class III,[9−11] based on the sequence motifs present in the C-terminal peptide stretches of their interaction partners. Class I PDZ domains prefer peptides containing X-S/T-X-φ-COOH, class II recognizes X-φ-X-φ-COOH, while class III recognizes X-D/E-X-φ-COOH motifs.[10,12] The crystal structures of PDZ domains in complex with substrate peptides have provided a structural basis for the recognition of above-mentioned sequence motifs by PDZ domains. The structure of PDZ domain consists of five to six β-strands (βA-βF) and two α-helices (αA and αB). The C-terminus of the interaction partner binds as an antiparallel β strand in a groove between βB strand and the αB helix[12−14] (Figure 1). A number of high throughput experimental studies have elucidated the substrate selectivity of a large number of PDZ domains from different organisms. Analysis of binding specificity of PDZ domains of mouse proteome using protein microarrays has provided insight into interaction between 85 mouse PDZ

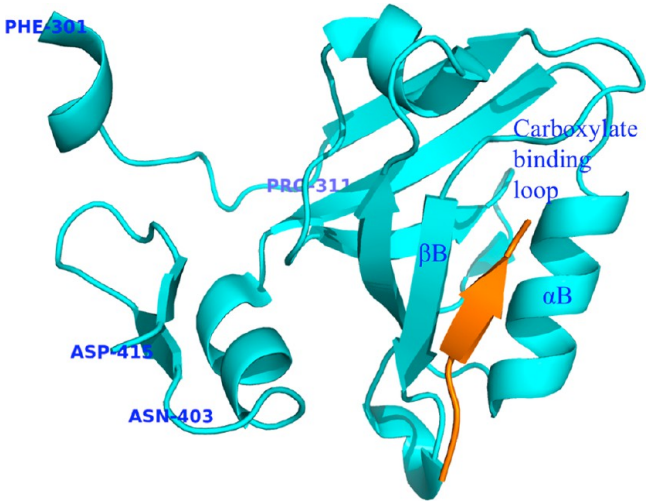


Figure 1. Structure of PDZ3 of PSD-95. Cartoon representation of crystal structure of PDZ3 of PSD-95 protein in complex with the substrate peptide (PDB ID: 1TP3).

Received: October 28, 2013
Published: March 4, 2014

domains and 217 mouse proteome derived peptides.[15] Phage display technique has also been used to map binding specificity of PDZ domains in proteomes of human and *C. elegans*.[8] In order to understand the role of different amino acids at different positions of the substrate peptide, Saro et al.[16] carried out thermodynamic studies to measure the binding affinities of native as well as 27 different mutant peptides for the third PDZ domain of PSD-95. Overall, these studies have revealed that the conventional C-terminal motifs X-S/T-X-φ-COOH, X-φ-X-φ-COOH, and X-D/E-X-φ-COOH are not adequate to explain selectivity of PDZ−peptide interactions. Therefore, on the basis of analysis of available experimental data on PDZ−peptide interactions, different groups have attempted to use computational approaches for deciphering interaction partners of PDZ domains and reconstruct PDZ interaction networks.

Majority of the currently available computational methods for predicting interaction partners of PDZ domains are essentially sequence-based approaches and they utilize structural details of PDZ−peptide complexes to a very limited extent. On the other hand, they utilize the available high throughput experimental data on PDZ−peptide interactions for training. Chen et al.[17] used the mouse peptide array derived interaction data[15] to develop a computational protocol based on Bayesian method to estimate the binding score for PDZ−peptide interaction given the primary sequences of PDZ domain and the peptide. This model implicitly uses the three-dimensional structure of PDZ domain to identify 38 interacting residue pairs between five residues of the substrate peptide and 16 residues in the substrate binding pocket of the PDZ domain. Shao et al.[18] used a regression framework to quantitatively predict the PDZ−peptide interactions using primary sequence.[18] Hui and Bader[19] used peptide array interaction data on mouse PDZ domains and phage display interaction data on human PDZ domains to train a support vector machine (SVM) model for predicting interaction partners of PDZ domains. On the basis of analysis of human PDZ phage display data Gfeller et al. derived multiple position weight matrices (PWMs) for predicting interactions mediated by PDZ domains.[7] Some of the above-mentioned sequence based methods[7,17] have also been used for constructing interaction networks involving PDZ domains.[20,21]

In contrast to the large number of sequence-based approaches for predicting interaction partners of PDZ domains, there are only a limited number of reports on structure-based methods for predicting PDZ−peptide interactions. Very recently, Hui et al.[22] have developed a structure-based predictor that uses SVM trained on mouse and human PDZ interaction data and incorporates various PDZ domain structural features like solvent accessibility, hydrogen bonding, hydrophobicity, etc. apart from interacting residue pairs. Kaufmann et al.[23] have developed a physical model for PDZ−peptide interactions by using leave-one-out cross validation to adjust the ROSETTA energy function for reproducing the experimental binding affinity data on 28 PDZ−peptide complexes. Smith and Kortemme[24] used 17 PDZ−peptide complex structures and protein modeling approach to develop a structure based method for prediction of peptide recognition preferences of natural and synthetic PDZ domains and compared their prediction results to phage display data on human PDZ domains. A molecular dynamics study on PDZ3 domain of PSD-95 protein has suggested that residues from an extra domain helix and βB−βC loop interact with peptide and regulate the binding mechanism.[25] A flexible docking procedure involving simulated annealing, molecular dynamics, and rotamer optimization has also been used for exploring peptide-binding selectivity of PDZ domains.[26]

Since most of the currently available methods for prediction of interaction partners of PDZ domains depend on experimental data for training, these methods often fail in case of PDZ families, which are different from those in the training set. In this work, we have used a novel multiscale approach to develop alternative structure based method, which can potentially identify new interaction partners of PDZ domains which lack homology to known interacting domain families and it is computationally fast enough for genome scale analysis. A similar structure based method has been developed earlier and benchmarked extensively for prediction of substrates for protein kinases and MHC molecules.[27−29] On the basis of analysis of the available structural data in PDB, representative structural templates for modeling PDZ−peptide complexes have been identified. Multiple explicit solvent all-atom MD simulations have been carried out on 28 PDZ−peptide complexes with experimentally known binding affinities to demonstrate that binding free energy values calculated from these explicit solvent MD trajectories using MM/PBSA approach show good correlation with the dissociation constants reported by experimental studies. The specificity determining residues (SDRs) of PDZ domains identified by our all-atom MD simulations on representative PDZ−peptide complexes have been utilized to compute the binding score between the PDZ domain and substrate peptide using a knowledge-based scoring function like Betancourt−Thirumalai residue based statistical potential.[30] This multiscale approach has been benchmarked on mouse PDZ data set[15] by calculating the binding energies for 217 different substrate peptides in binding pockets of 64 different mouse PDZ domains using statistical potential.

MATERIALS AND METHODS

Analysis of Available Crystal Structures of PDZ Domain. All the available crystal structures of PDZ−peptide complexes were retrieved from the Protein Data Bank (PDB).[31] Multiple sequence alignment (MSA) of the sequences of PDZ domains in these 60 entries was carried out using the CLUSTALW package.[32] The MSA was provided as input to the MEGA4[33] package for constructing a phylogenetic tree by neighborhood joining approach with bootstrapping. The sequences of the representative structures from each cluster in the dendrogram were pairwise aligned with each other using a locally installed version of NCBI BLAST program[34] and structural similarity was analyzed using the DALI server.[35]

MD Simulations. MD simulations were performed on the crystal structure of the third PDZ domain of PSD-95 protein in complex with 28 different peptides for which binding affinities were known from experimental studies reported by Saro et al.[16] The PDB entry 1TP3 corresponding to the crystal structure of this PDZ domain in complex with the peptide KKETPV was used as template for modeling starting structures for the 27 other PDZ−peptide complexes. The PDZ−peptide complexes corresponding to these 27 other peptides were generated by introducing mutations in the bound peptide in 1TP3 using backbone-dependent rotamer library approach of SCWRL.[36] Explicit solvent MD simulations were carried out using AMBER 9[37] package and ff03 force field.[38] Each of these 28 PDZ peptide complexes was solvated using TIP3P water model[39] in a rectangular box that extended 8 Å from the outermost protein atoms along *x*, *y*, and *z*-axes. The solvated structures were first minimized by steepest descent algorithm to RMS gradient of

0.001 kcal/mol·Å using SANDER module of AMBER 9. After minimization, the temperature of the system was gradually increased to 300 K over 50 ps MD simulations under *NVT* conditions and over the next 50 ps of simulation pressure equilibration was carried to 1 atm using *NPT* conditions. After the equilibration, production dynamics simulations were carried out under *NVT* conditions at 300 K, using a time step of 1 fs. Nonbonded interactions were computed using a cutoff of 8 Å and the particle mesh Ewald (PME) approach[40] was employed to consider the long-range electrostatic interactions. For each of these 28 PDZ−peptide complexes, 6 independent 5 ns MD trajectories were computed starting from different random number seeds to assign different sets of initial velocities in each of these six independent simulations. One simulation was also run for longer duration of 50 ns on the crystal structure 1TP3 corresponding to the structure in complex with the peptide KKETPV. The various MD trajectories were analyzed using PTRAJ module of AMBER package and Perl scripts developed in house. Two-dimensional root mean squared diameter (2D-RMSD) analysis and clustering of conformers were also carried out using PTRAJ.

Free Energy Calculation Using the MM/PBSA Method. From each of the six 5 ns explicit solvent MD trajectories for different PDZ−peptide complexes, the last 1 ns of the MD trajectories was utilized for calculation of the binding free energies. Snapshots at an interval of 10 ps were collected, giving a total of 100 conformations of PDZ−peptide complexes from each trajectory. For the 50 ns simulation, the last 10 ns of the trajectory was taken for binding energy calculation, giving rise to a set of 1000 conformations for the PDZ−peptide complexes. The binding free energy was calculated using MM/PBSA approach,[41] which uses an implicit solvent model. Therefore, coordinates of solvent molecules were removed from each of the extracted snapshots and only the coordinates of the PDZ domain and bound peptide were used for MM/PBSA binding free energy calculation by AMBER 9 package. The free energy was calculated for each molecular species namely, PDZ−peptide complex, receptor (PDZ domain), and ligand (peptide). The peptide binding free energy was computed as the difference.

$$\Delta G_{\text{binding}} = G_{\text{complex}} - G_{\text{PDZ-domain}} - G_{\text{peptide}}$$

The MM/PBSA module of the AMBER 9 package, which uses an implicit solvent model, was used for this purpose.

Binding Pocket Analysis and Calculation of Pair Potential Scores. Snapshots of protein−peptide complexes at an interval of 50 ps were extracted from the last 1 ns of one of the representative trajectory of the six 5 ns MD simulations for a given PDZ−peptide complex. This resulted in 20 conformations of 20 PDZ−peptide complexes for the chosen 5 ns trajectory. In each of these 20 PDZ−peptide complexes, the contacting residue pairs between the PDZ domain and the peptide were identified using the criteria of any two atoms of the residue pair being at a distance less than 3.5 Å. Thus, the binding pockets were identified for various residues of the bound peptide. For identification of dynamically stable binding pockets, only those contacting residue pairs which were present in more than 50% of the PDZ−peptide complexes in last 1 ns of the MD trajectory were considered. Since the residue at position P (−5) of the peptide was exposed to the solvent and did not interact with any of the residues of the PDZ domain, no binding pocket was identified for this residue. So, in total five binding pockets were identified for peptide residues −4 to 0. On the basis of the contacting residue pairs between protein and peptide, the binding energy was evaluated using Betancourt−Thirumalai (BT) statistical pair potential.[30]

Benchmarking of the Protocol for Identification of Interaction Partners of Mouse PDZ Domains. The computational protocol for genome scale identification of interaction partners of mouse PDZ domains was benchmarked on the PDZ−peptide interaction data reported by Stiffler et al.[15] This data set covered interactions between 85 mouse PDZ domains and 217 peptides derived from mouse genome. Out of the 85 mouse PDZ domains in the data set of Stiffler et al., database accession numbers could be obtained for only 82 PDZ domains. In order to identify suitable structural templates for modeling these mouse PDZ domains in complex with bound peptides, the sequences of these 82 mouse PDZ domains were pairwise aligned using the NCBI BLAST program with the sequences of PDZ domains in the data set of the crystal structures of PDZ−peptide complexes. Out of the 82 mouse PDZ domains, structural templates could be identified for only 38 domains based on statistically significant pair BLAST alignments. For the remaining 44 sequences, sequence to structure alignment by threading approach was carried out by using MUSTER[42] for identification of suitable structural templates. However, only 26 out of these 44 mouse PDZ domains aligned with crystal structures of PDZ−peptide complexes in our data set by threading approach. Thus, out of 82 mouse PDZ domains, structural templates could be identified for only 64 PDZ domains, and they corresponded to 24 PDZ−peptide complex crystal structures including 1TP3. Structure-based multiple sequence alignment was carried out for these 24 PDZ domains using PROMALS3D,[43] and the putative binding pockets for peptide residues −4 to 0 were mapped for each of these 24 PDZ structural templates based on alignment with peptide binding pocket residues on 1TP3. Since, the peptide binding pocket was mapped onto each of these 24 structural templates, the putative binding pocket residues for each of the 64 mouse PDZ domains could be identified from the alignment of each of these mouse PDZ domains with their respective structural templates by using programs like CLUSTALW.[32] After identification of binding pocket residues for each of these 64 mouse PDZ domains, interaction energy of any peptide in complex with each of these 64 mouse PDZ domains was evaluated using Betancourt−Thirumalai (BT) residue based statistical pair potential.[30] Using this computational protocol interaction energy of each of these 64 mouse PDZ domains in complex with 217 mouse genome derived peptides was evaluated. By varying the cutoff for the computed BT pair potential binding energy score, each of these 217 peptides were classified as binder or nonbinder for a given PDZ domain. On the basis of comparison with experimental data of Stiffler et al., prediction for each PDZ−peptide pair was classified as true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The statistical significance of our prediction method was evaluated using receiver operating characteristic (ROC) curve analysis. ROC is a plot of true positive rate (TPR) vs false positive rate (FPR) as the cutoff value varied from minimum to maximum. TPR, also known as sensitivity, was calculated as the fraction of true positive among all positives (TPR = TP/(TP + FN)). Specificity was calculated as the fraction of true negatives among all negatives (specificity = TN/(TN + FP)). FPR (also known as (1-specificity)) was calculated as the fraction of false positives among all negatives (FPR = FP/(FP + TN)). The predictive power of our computational protocol was evaluated by calculating the area under the curve

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

(AUC) in the ROC plot. ROC function of R statistical language (http://www.r-project.org) was used for calculation of AUC values. For all classification models Cohen's kappa values and the corresponding errors were calculated by using R package.

RESULTS

Analysis of the Crystal Structures of PDZ−Peptide Complexes. Figure 2 shows the sequence based phylogenetic tree of 60 PDZ domains for which structures are available in complex with peptides. As can be seen from Figure 2, based on sequence similarity of PDZ domains, these 60 PDZ−peptide complexes can be grouped into 11 clusters, out of which 2 clusters contain single structures. Figure 3 shows the results of detailed analysis of sequence and structural similarity between representative PDZ−peptide complexes from each of these 11 clusters. For each pair of structures, the upper diagonal in Figure 3 shows the sequence identity for the aligned segments, while the lower diagonal shows the C^{α} RMSD and the number of aligned residues. As can be seen, even though the PDZ domains in these 11 clusters show sequence identity in the range of 13−41%, the corresponding structures align with RMSD values in the range of 1.3−2.9 Å over at least 60 residues which correspond to at least 70% of their length. Thus, our analysis reveals that despite high sequence divergence PDZ domains adopt a highly conserved structural fold. Further analysis also revealed that if a pair of PDZ domains from these 11 representative clusters is superposed using the coordinates of the PDZ domain alone, the bound peptides also show good structural alignment with low RMSD values. This result suggests that the peptide binding site on these PDZ domains are also conserved and the peptide ligands bind in similar backbone conformations as an antiparallel β-strand as reported by Lee and Zheng.[13]

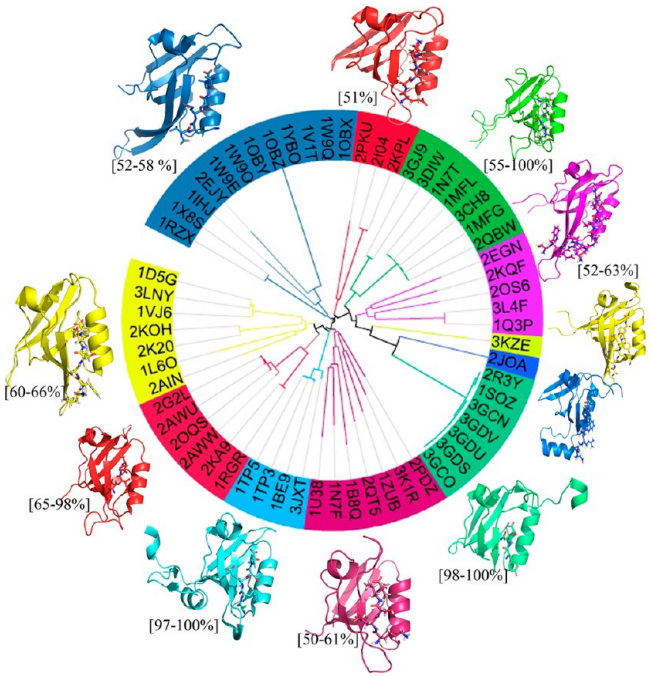


Figure 2. Clustering of PDZ domains. Sequence-based clustering of 60 PDZ domains for which peptide bound structures are available in PDB. Values in square brackets are sequence similarity ranges for the PDZ domains within the cluster.

Molecular Basis of the Peptide Binding Specificity of PDZ3 domain of PSD-95. The major objective of our study was to investigate whether structure-based approach can predict peptide binding specificity for PDZ domains. PDZ3, the third PDZ domain of PSD-95, was the most appropriate representative for this analysis because the crystal structure (PDB ID: 1TP3) of this PDZ domain was available in complex with the native substrate peptide KKETPV and also experimental binding affinities of PDZ3 were available from thermodynamic studies for a set of 27 different mutants of the peptide KKETPV. Table 1 summarizes the binding affinity (K_d) values of 28 different peptides for PDZ3 domain of PSD-95. Since crystal structure of the native peptide was available in complex with the PDZ domain, we wanted to first investigate whether binding free energy values computed by residue based statistical pair potentials show correlation with experimentally determined binding free energy values. Supporting Information Figure S1 shows the correlation plot between experimentally determined binding free energy values and pair potential scores calculated by BT matrix. The correlation coefficient is 0.06, which clearly indicates that BT pair potential is not able to discriminate between binders and nonbinders for PDZ domains. This result was per se surprising, because earlier studies had indicated that in case of kinases and MHC domains binding partners could be predicted using such potentials.[28,29] However, careful analysis of the binding modes of peptides to PDZ domains, MHCs, and kinases indicated that unlike MHCs and kinases the PDZ domains are relatively smaller in size and the peptide binds as an antiparallel β strand in a shallow cavity. Hence, conformational flexibilities of the PDZ domain might play a crucial role in substrate recognition.

Therefore, in order to introduce flexibility into the system, multiple nanosecond time scale explicit solvent all atom MD simulations were carried out for each of the 28 PDZ−peptide complexes for which experimental binding affinity data was

	1N7T	1RZX	1TP3	2AIN	2AWW	2KQF	2JOA	2PDZ	2PKU	3GCN	3KZE
1N7T(87 aa)		28%	37%	28%	38%	24%	22%	33%	24%	22%	24%
1RZX(91 aa)	2.1(78)		27%	29%	25%	20%	26%	29%	27%	23%	18%
1TP3(79 aa)	1.4(75)	1.8(77)		38%	39%	28%	21%	42%	24%	17%	17%
2AIN(71 aa)	2.2(61)	2.0(63)	1.9(61)		41%	20%	23%	36%	37%	27%	26%
2AWW(89 aa)	1.3(82)	2.0(81)	1.5(79)	1.7(61)		21%	13%	41%	23%	18%	26%
2KQF(96 aa)	2.0(78)	3.7(86)	1.9(78)	2.6(69)	2.0(80)		19%	24%	24%	25%	23%
2JOA(84 aa)	2.4(67)	3.1(68)	1.9(67)	3.0(53)	2.2(69)	2.3(69)		18%	25%	34%	27%
2PDZ(81 aa)	1.7(76)	2.1(79)	1.4(78)	2.3(61)	1.6(80)	2.2(79)	1.9(66)		28%	18%	19%
2PKU(80 aa)	1.9(76)	1.9(77)	1.5(78)	2.3(62)	1.6(79)	2.1(78)	2.3(67)	1.9(80)		21%	17%
3GCN(73 aa)	1.9(60)	1.6(60)	1.8(63)	1.7(45)	1.7(61)	1.9(61)	1.7(62)	1.7(61)	1.5(61)		19%
3KZE(81 aa)	2.9(75)	2.5(79)	2.1(75)	2.2(61)	2.4(74)	1.9(77)	2.4(67)	2.4(74)	2.6(76)	2.2(62)	

Figure 3. Sequence and structural similarities between representative structures from each of the 11 clusters. First row and column indicate PDB IDS of 11 representative structures with the number of residues in parentheses. The upper diagonal of the table indicates sequence identity for the aligned segment, and the lower diagonal shows RMSD and the number of aligned residues (in parentheses).

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

Table 1. K_d Values Determined Experimentally by Saro et al.,[16] Corresponding Experimental Binding Free Energies, Various Simulations Carried for Each of the PDZ−Peptide Complexes, and Average Binding Free Energies Calculated by MM/PBSA from Molecular Dynamics Trajectories

S. no.	peptide	K_d (μM)	$\Delta G = -RT \ln K_d$ (kcal/mol)	simulation length[a]	MM/PBSA (kcal/mol)[b]
1	KKETEV	1.9 ± 0.1	−7.845	5 ns(6)	−39.04(±3.67)
2	KKETEA	91.0 ± 2.0	−5.541	5 ns(6)	−29.78(±2.11)
3	KKETEL	7.9 ± 1.3	−6.996	5 ns(6)	−36.46(±2.02)
4	KKETEI	7.7 ± 1.2	−7.012	5 ns(6)	−38.42(±5.48)
5	KKETEM	21.0 ± 2.0	−6.414	5 ns(6)	−36.23(±2.65)
6	KKETEF	57.0 ± 2.0	−5.819	5 ns(6)	−40.21(±1.54)
7	KKETET	105.0 ± 6.0	−5.456	5 ns(6)	−32.35(±1.77)
8	KKESEV	6.6 ± 0.9	−7.103	5 ns(6)	−40.01(±8.26)
9	KKECEV	72.0 ± 7.0	−5.68	5 ns(6)	−34.32(±5.59)
10	KKESEL	33.0 ± 2.0	−6.145	5 ns(6)	−36.79(±4.32)
11	KKESEI	24.0 ± 6.0	−6.335	5 ns(6)	−32.62(±4.19)
12	KKESEF	98.0 ± 16.0	−5.497	5 ns(6)	−36.16(±3.21)
13	KKETGV	2.4 ± 0.0	−7.706	5 ns(6)	−38.75(±3.74)
14	KKETAV	0.5 ± 0.1	−8.703	5 ns(6)	−36.74(±2.66)
15	KKETVV	1.3 ± 0.2	−8.071	5 ns(6)	−35.01(±3.54)
16	KKETLV	1.8 ± 0.3	−7.877	5 ns(6)	−39.04(±1.38)
17	KKETPV	0.9 ± 0.2	−8.258	5 ns(6)	−38.53(±2.02)
18	KKETWV	2.8 ± 0.4	−7.614	5 ns(6)	−37.66(±1.33)
19	KKETDV	20.0 ± 2.0	−6.443	5 ns(6)	−37.00(±7.24)
20	KKETKV	1.2 ± 0.0	−8.119	5 ns(6)	−37.96(±6.00)
21	KKGTEV	80.0 ± 3.0	−5.618	5 ns(6)	−35.39(±5.98)
22	KKATEV	21.0 ± 4.0	−6.414	5 ns(6)	−33.75(±1.96)
23	KKQTEV	4.0 ± 0.0	−7.402	5 ns(6)	−37.16(±5.49)
24	KKDTEV	85.0 ± 12.0	−5.582	5 ns(6)	−36.05(±2.31)
25	KKKTEV	27.0 ± 4.0	−6.264	5 ns(6)	−32.54(±6.70)
26	KKGTGV	273.0 ± 30.0	−4.887	5 ns(6)	−32.13(±4.69)
27	KKATAV	8.3 ± 1.5	−6.967	5 ns(6)	−38.52(±7.16)
28	YKETEV	1.2 ± 0.1	−8.119	5 ns(6)	−34.84(±2.18)

[a]Number of repeats. [b]Standard deviation.

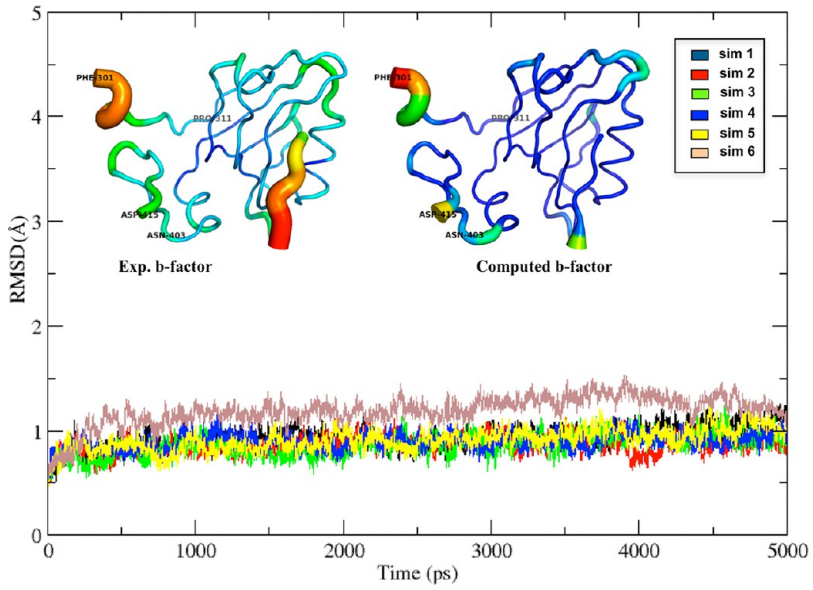


Figure 4. RMSD vs time plot for 5 ns MD simulation of PDZ3−peptide complex. RMSD vs time plot for MD simulations on PDZ3 domain of PSD-95 protein (excluding N- and C-terminal portions) in complex with its cognate peptide. (inset) Cα trace for the crystal structure as well as structure of PDZ3−peptide complex obtained after 5 ns of MD simulations with thickness and color of the chain reflecting the B-factor values of the residues (blue = lowest and red = highest B-factor).

available. Six independent simulations of 5 ns duration were carried out for each of these 28 PDZ−peptide complexes by starting the simulations with different random number seeds for assigning different sets of initial velocity values. This resulted in a total trajectory length of 30 ns for each of the PDZ−peptide complexes. In order to monitor the conformational changes in

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

the various PDZ−peptide complexes, the backbone RMSD with respect to the initial structure was computed for the PDZ−peptide complexes sampled during the MD simulations. Figure 4 shows the variation of the backbone RMSD with respect to the initial structure, for the complex involving the native peptide KKETPV (1TP3). The inset to Figure 4 shows comparison of the B-factor values computed from RMSF values with the experimental B-factors. Detailed analysis of the computed B-factors or RMSF values for the individual residues over each of these six MD trajectories indicated that N-terminus (residues Phe 301 to Pro 311) and C-terminus (Asn 403 to Asp 415) fragments (inset in Figure 4) of the PDZ domain which make relatively fewer contacts with the rest of the PDZ domain or bound peptide show higher flexibility. It may be noted that RMSDs in Figure 4 has been calculated after exclusion of the N- and C-terminus fragments that show higher flexibility, while Supporting Information Figure S2 shows complete RMSD when all the residues were included. As can be seen from Figure 4 in all six simulations the RMSDs converge to values in the range of 0.8−1.2 Å after first few hundred picoseconds of the simulations. These results indicate that the backbone conformations of the core of the PDZ domain as well as the bound peptide remain close to the crystallographic structure during the explicit solvent simulations. Similar analysis of total RMSD and RMSD excluding the terminal regions for the other 27 PDZ−peptide complexes revealed that they also show a very similar behavior. This is indeed expected as these complexes had identical PDZ domain and essentially involved point mutations in the bound peptide.

We also wanted to investigate the effect of simulations for longer duration on the core structure of the PDZ domain and conformation of the bound peptide. Therefore, one of the six 5 ns simulations on 1TP3 was extended up to 50 ns. Figure 5 shows the plot for total backbone RMSD as well as the backbone RMSD for the core region (after exclusion of the terminal regions) over the entire 50 ns trajectory. As can be seen, the RMSD of the core region remains stable around 1.5 Å throughout the 50 ns trajectory, while the total RMSD including the terminal regions shows small fluctuations around the value of 2.5 Å. For a detailed

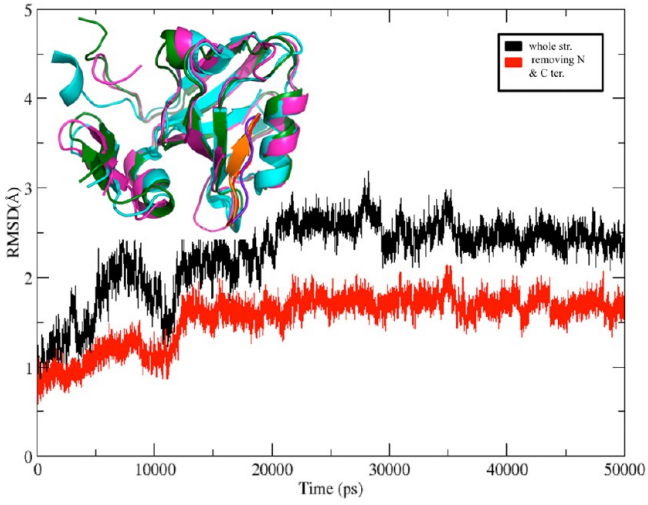


Figure 5. RMSD vs time plot for 50 ns MD simulation of PDZ3−peptide complex. RMSD vs time plot for 50 ns MD simulation of PDZ3 of PSD-95 protein in complex with its cognate peptide (PDB ID: 1TP3). (inset) Superposition of crystal structure (PDZ cyan, pep orange) with final structures obtained from 5 ns (PDZ green, pep red) and 50 ns (PDZ magenta, pep violet) MD simulations.

comparison of the conformations sampled during the longer 50 ns simulation with those sampled during multiple 5 ns simulations, snapshots at an interval of 50 ps were extracted from the various trajectories of 1TP3, resulting in a total of 1500 structures. RMSD between all possible pairs of structures were calculated and represented in the form of 2D-RMSD plot (Figure 6A). As can be seen from Figure 6A, some of the conformations sampled during initial stages of 50 ns trajectory show low RMSD with the structures sampled during the later part of the 50 ns trajectory. Similarly conformers from shorter 5 ns trajectories also show low RMSD with the structures sampled during longer simulation. This suggested existence of clusters of conformations which are common to the shorter and the longer simulations. Clustering of these 1500 structures by ptraj module of AMBER revealed existence of a total of five clusters, out of which two were highly populated (Figure 6B). The inset to Figure 6B shows superposition of representative structures from each of the five clusters. As can be seen from Figure 6B, the core of the PDZ domain which harbors the peptide binding pocket superpose well across the five different clusters and major conformational changes between the clusters occur in the N- and C-terminal regions and extra domain helix. Figure 7 shows the superposition of representative structures from each of the five clusters on the crystal structure 1TP3. As can be seen, even though core of the PDZ domain does not show large conformational differences in different clusters, there are subtle local changes in the conformation of the bound peptide, carboxylate binding loop, and loop connecting the secondary structural elements βB and αB. This is clearly depicted in Supporting Information Figure S3 that shows the RMSF plots obtained from 5 ns and 50 ns MD simulation with the inset showing the flexible loop regions (colored in orange). Interestingly, carboxylate binding loop interacts with the C-terminal residue of the substrate peptide and earlier studies have also suggested that conformational changes in the carboxylate binding loop can facilitate recognition of internal peptides by certain PDZ domains.[44] Similarly the role of the loop region connecting the secondary structural elements βB and αB in substrate recognition of PDZ domains have also been highlighted in earlier studies.[45]

Since the results from the 50 ns simulations on 1TP3 were very similar to those obtained from 5 ns trajectories, we did not carry out longer simulations for the complexes involving mutant peptide. As mentioned in the Material and Methods section, the complexes involving 27 mutant peptides were generated from the PDB entry 1TP3 by introducing appropriate mutations on the bound peptide KKETPV using the rotamer library approach of SCWRL.[36] All the complexes obtained after introducing the mutations in the peptide (present in 1TP3) were first minimized in vacuum for 100 000 steps, and then after solvation, they were minimized using steepest descent algorithm until the RMS gradient value converged to 0.001 kcal/mol·Å. After minimization, for all 28 PDZ−peptide complexes, the same procedure of heating the system up to 300 K and pressure equilibration was carried out prior to production dynamics. We also analyzed whether the modeled peptides of different PDZ−peptide complexes remained in the binding pocket throughout the MD simulation. Supporting Information Figure S4 shows the C^{α} RMS deviation of the peptide from its initial position in the binding pocket. For this analysis, the frames were extracted from the 5 ns trajectories at an interval of 1 ns and the deviation from the native peptide conformation in the crystal structure was calculated by superposing the frames onto 1TP3 using coordinates of the PDZ domain alone. It is clear from Figure

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

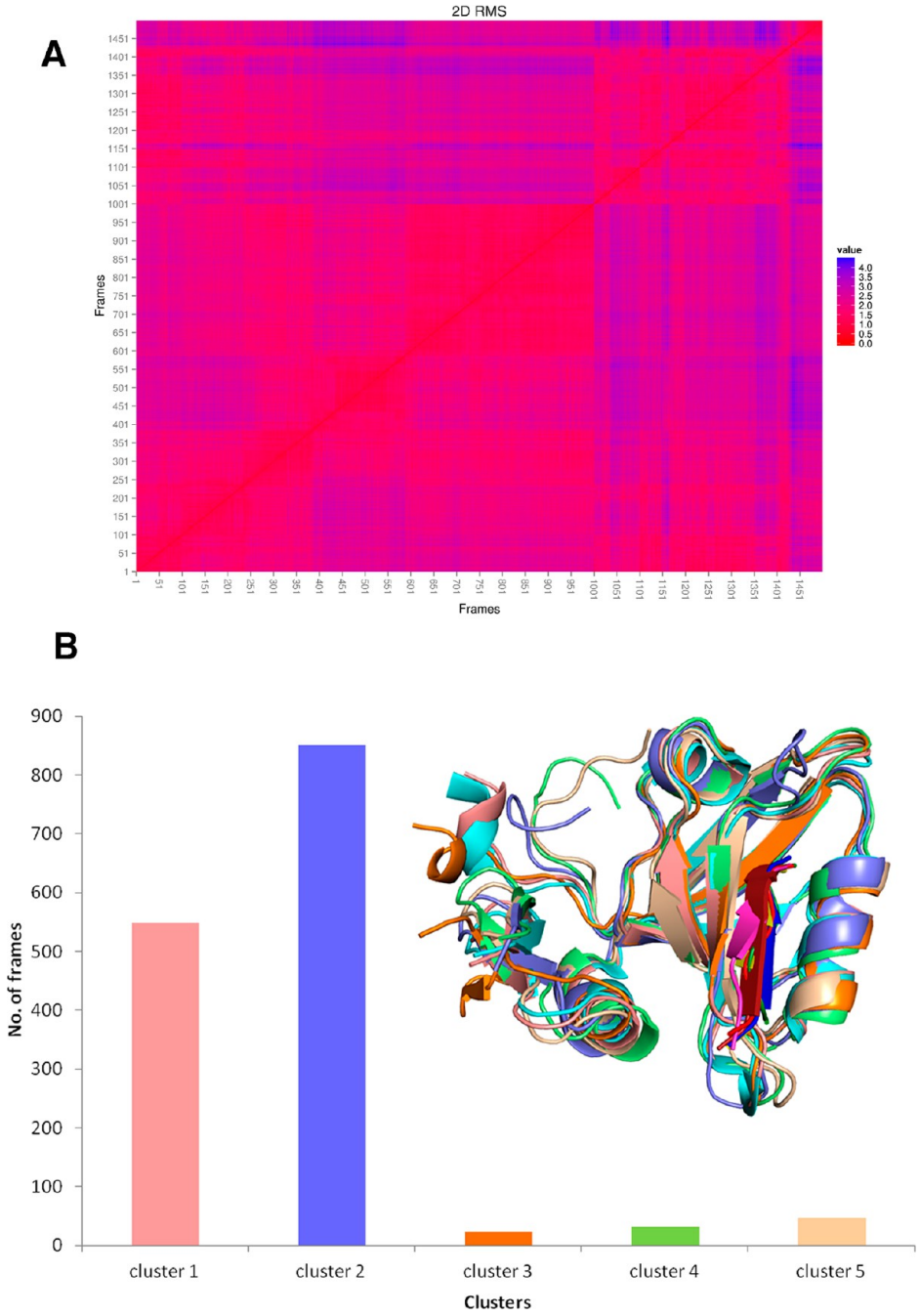


Figure 6. Clustering of PDZ−peptide complexes sampled during the MD simulations on 1TP3. (A) 2D-RMSD plot for 1500 snapshots obtained from one 50 ns and five 5 ns simulations at an interval of 50 ps. (B) Clustering of 1500 structures by average linkage algorithm and distribution of number of structures in various clusters. (inset) Superposition of representative structures from each of the five clusters.

S4 that in most cases the modeled peptides remained in the binding pocket showing RMSD below 2.0 Å. Only in few cases the RMSD increased up to 2.5 Å, and only in the case of two peptides RMSD values were higher than 3.5 Å indicating deviation of the peptide from the initial position.

The analysis of the multiple 5 ns MD trajectories for the 28 PDZ−peptide complexes indicated that the backbone of the core PDZ domain as well as the bound peptide remained close to the starting structure. However, there were subtle conformational rearrangements observed in the side chains of the peptide as well as the residues lining the binding pocket sampled from multiple conformations. We attempted to compute the binding free energy between the peptide and the PDZ domain using MM/PBSA approach. For each of the 28 PDZ−peptide complexes, the conformations sampled during last 1 ns of the six MD trajectories were used for calculation of MM/PBSA binding free energy values. Supporting Information Figure S5 shows the MM/PBSA binding energy values obtained from each of the six MD trajectories for all the 28 PDZ−peptide complexes, while Table 1 summarizes the averaged binding free energies and their standard deviations. It may be noted that MM/PBSA binding energy values reported in Figure S5 and Table 1 have been computed from last 1 ns of the multiple 5 ns simulations for each of the 28 PDZ−peptide complexes. We also investigated the effect of selection of longer trajectory windows on values of MM/PBSA binding energy. The MM/PBSA binding free energy for the

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

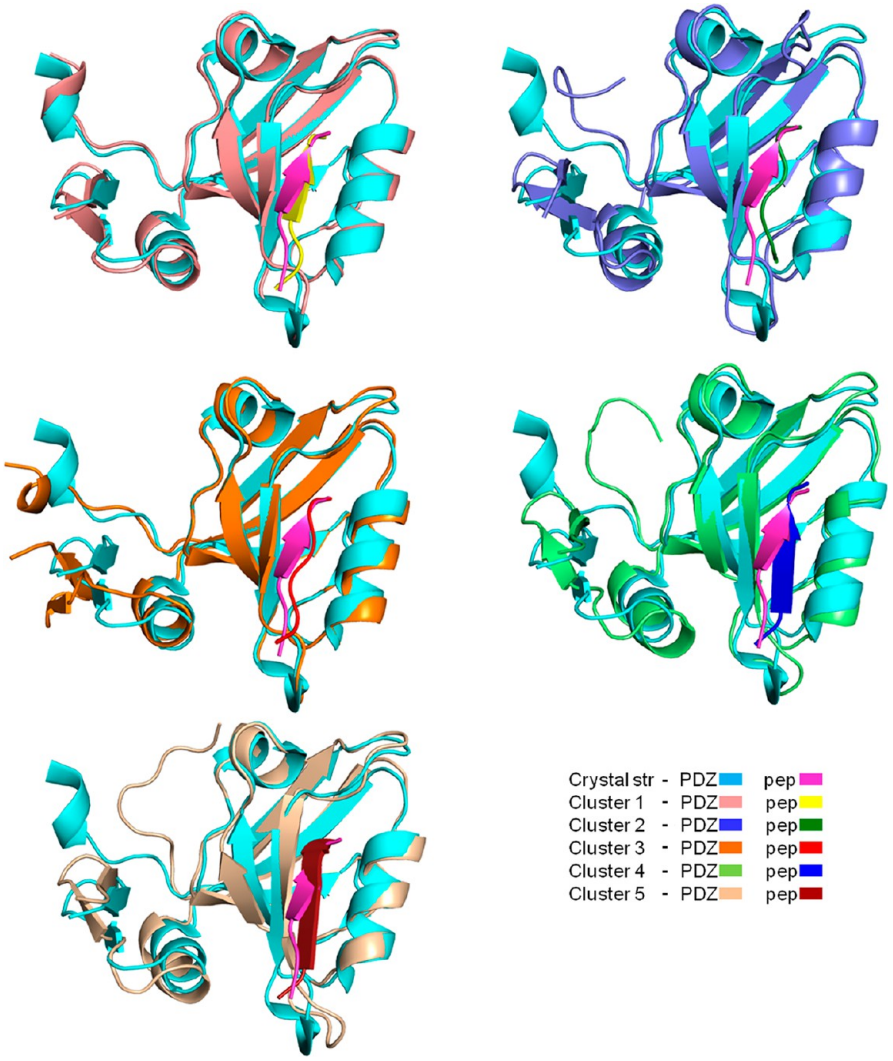


Figure 7. Superposition of representative structures from each of the five clusters on the crystal structure 1TP3. The colors codes for PDZ domain and substrate peptide chains in different clusters are depicted in the inset.

native PDZ−peptide complex was computed by selecting the last 1, 2, and 3 ns of the multiple 5 ns trajectories. Similarly MM/PBSA binding energy was also computed from the last 10 ns of the 50 ns trajectory. Supporting Information Figure S6 shows the comparison of binding energy values obtained from five simulations of 5 ns each, a single 50 ns simulation, and also the effect of variation of length of the trajectory selected for MM/PBSA calculation. This clearly shows that MM/PBSA energies obtained from multiple trajectory widows are very similar. Similarly there are no significant differences between binding energy values calculated from a longer 50 ns simulation over multiple shorter 5 ns simulations. Therefore, for all the 28 PDZ−peptide complexes, MM/PBSA binding energy was computed from last 1 ns of the multiple 5 ns simulations and averaged (Table 1). We then proceeded to analyze the correlations between computed MM/PBSA binding energy and experimentally determined binding free energy values. The lower panel in Figure 8 shows the correlation plot between the MM/PBSA binding energy values obtained from our simulations and experimental binding free energy values obtained from thermodynamic studies.[16] It is encouraging to note that the pearson correlation coefficient is 0.527 (p-value = 0.00394) for all 28 PDZ−peptide complexes and it increases to 0.755 (p-value = 0.00002), if four outliers are excluded from the data set. We also analyzed the predictive power of MM/PBSA binding energy values for binder vs nonbinder prediction in terms of ROC curve (upper panels in Figure 8), by labeling all peptides having experimental binding energy values lower than −6.5 kcal/mol as binders and others as nonbinders. In ROC analysis, the true positive rate (TPR) is plotted against false positive rate (FPR) for predictions by varying the cutoff in computed score (MM/PBSA energy here) over the entire range of values. In the ROC curve shown in Figure 8, the color gradient indicates the MM/PBSA binding energy cutoffs which varies from −27.78 to −40.61 kcal/mol. As can be seen, AUC values are 0.847 for all 28 peptides and it increases to 0.986 after exclusion of 4 outliers. These high AUC values indicate very good prediction accuracy of the MM/PBSA-based binding energy model. The optimum cutoff value for MM/PBSA energy which will have maximal TPR at minimum FPR is −37 kcal/mol. For prediction using the cutoff of −37 kcal/mol in MM/PBSA energy, the confusion matrix which consists of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values is shown in Supporting Information Table S1. We also calculated the Cohen's kappa coefficient and the corresponding error value for this MM/PBSA based prediction of PDZ binders (Table S1). As can be seen when binder vs

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

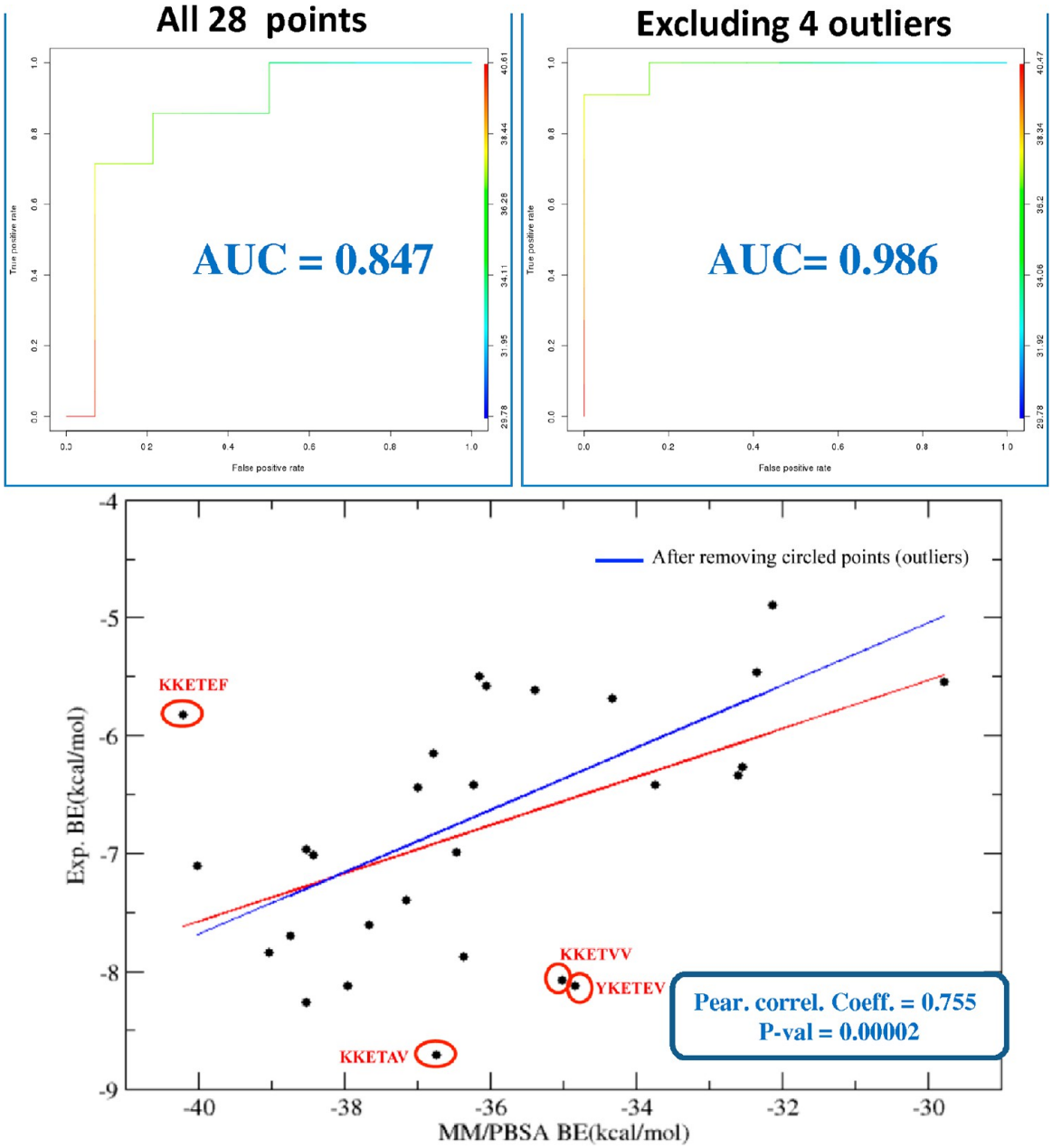


Figure 8. Correlation of MM/PBSA binding energy with experiments. Correlation between experimentally determined binding energy values and binding energies calculated from MM/PBSA for 28 PDZ−peptide complexes (lower panel). (upper panels) ROC curve for binder vs nonbinder prediction for the 28 PDZ−peptide complexes and after exclusion of 4 outliers.

nonbinder prediction was carried out for 28 peptides using MM/PBSA, kappa was 0.57 ± 0.16 and it increased to 0.83 ± 0.11 upon removal of 4 outlier points. Thus, these results demonstrate that it is possible to quantitatively predict the binding affinity of PDZ domains with reasonable accuracy, provided the conformational flexibility of the ligand as well as the binding pocket is incorporated in the model. Since binding energy computed from static crystal structures using statistical pair potentials did not show correlation with experimental data but MM/PBSA calculations showed good correlations, we analyzed the MD trajectories of various PDZ peptide complexes to identify changes in the peptide-binding pocket during MD simulations vis-a-vis the binding pocket in the static crystal structures.

Binding Pocket Analysis. We identified the number of contacting residues between PDZ domain and the peptide in various PDZ−peptide complexes before and after MD simulations. For constructing the binding pocket after simulation, snapshots were taken at an interval of 50 ps from the last 1 ns of the trajectory and percentage occupancy of the contacts were calculated, and only those contacts for which occupancy was more than 50% were included in the binding pocket. The pockets were numbered −5 to 0 according to their respective contacting peptide residues. Figure 9A and B shows the binding pockets for the native peptide on PDZ3 domain highlighting the residues involved before and after MD simulations respectively. Figure 9C depicts separately all five subpockets corresponding to peptide residues P 0, P −1, P −2, P −3, and P −4. The left panels correspond to the binding pocket obtained from static structure, while right panels show binding pocket residues obtained from MD simulations (Figure 9C). As

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

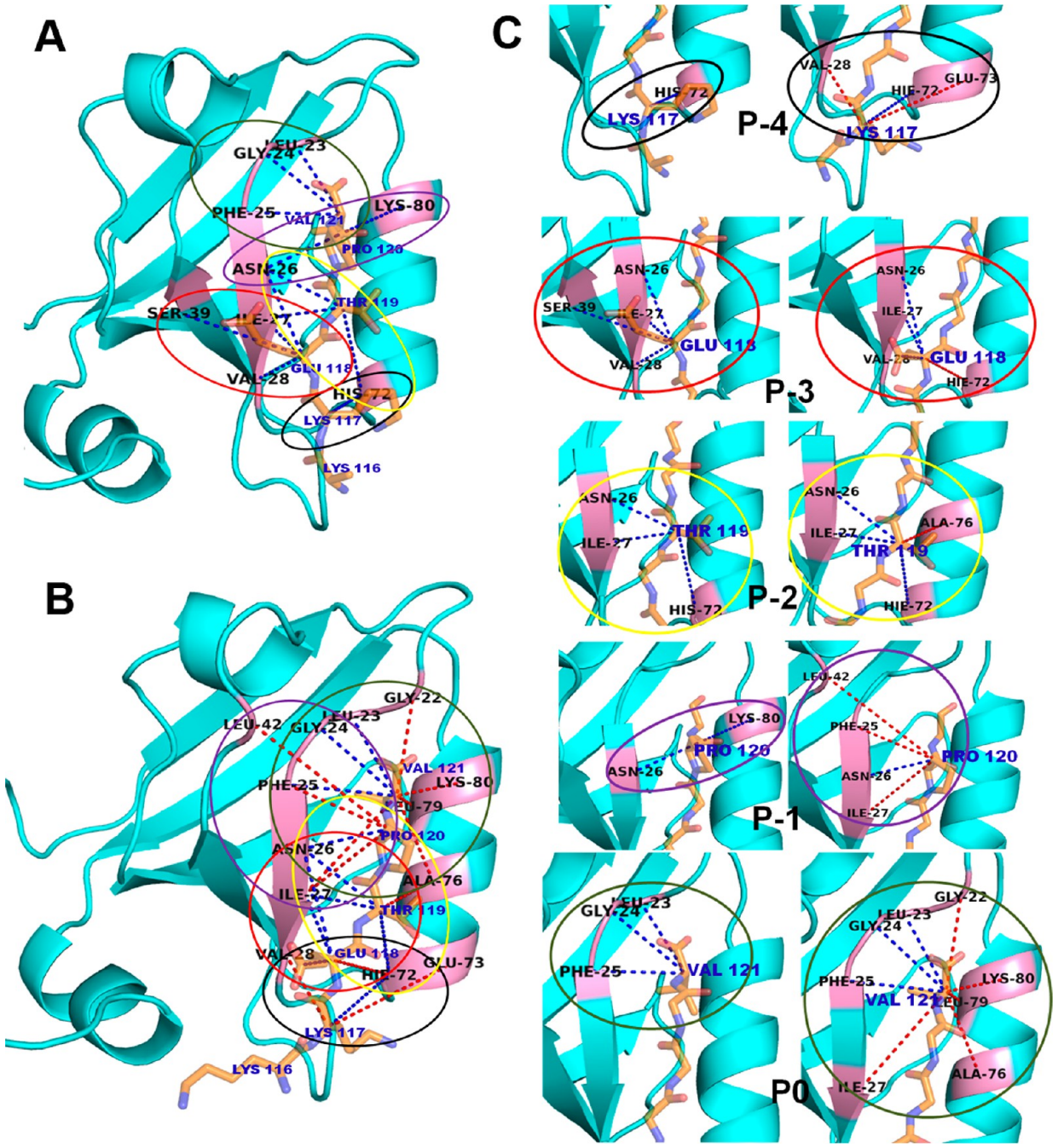


Figure 9. Effect of simulations on number of contacts between protein and peptide (1TP3). The residues of PDZ domain which interact with the residue of the peptide have been colored in pink on the ribbon diagram of PDZ domain. Peptide residue in each subpocket has been labeled in blue, while binding pocket residues of PDZ domain are labled in black. For the sake of clarity, side chains of the binding pocket residues are not shown. (A) Depiction of binding pocket residues before simulation. (B) Depiction of binding pocket residues after simulation. (C) Depiction of binding pocket residues in each of the five subpockets. (left panels) Binding pocket before simulation. (right panels) Binding pocket residues obtained after MD simulations.

can be seen compared to the number of contacts present in the static structure, several new contacts have formed in all the pockets −4 to 0. No contacts were seen with the residue at −5 position on the peptide, as the corresponding Lys was oriented toward the solvent during MD simulation. Table 2 lists the contacting residues in various binding pockets for the native PDZ−peptide complex and their respective occupancy times as calculated from last 1 ns of the 5 ns trajectories. Detailed analysis of the percentage occupancy of contacting residue pairs between peptide and PDZ domain in the 50 ns trajectory also confirmed that binding pocket residues listed in Table 2 also show high occupancy during longer simulation. We wanted to investigate if binding pocket identified by MD simulations is used to define contacting residue pairs between PDZ domain and the peptide and whether binding energy scores computed by BT statistical pair potentials would show better correlation with experimental results. Figure 10 shows the correlation plot between BT matrix binding energy scores based on new binding pockets and experimentally determined binding free energy values. It was really encouraging to note that the correlation coefficient

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

Table 2. Occupancy Time (%) of Contacting Residue Pairs between Domain and Peptide Across last 1 ns of 5 ns MD Trajectory with a Distance Cutoff of 3.5 Å (Any Atom−Any Atom)

res no.	res	res no.	res	% occ
116	LYS	100	PHE	45.00%
		31	GLU	35.00%
		99	ARG	10.00%
		28	VAL	10.00%
		29	GLY	5.00%
117	LYS	72	HIS	100.00%
		28	VAL	90.00%
		73	GLU	80.00%
		29	GLY	45.00%
		31	GLU	15.00%
118	GLU	27	ILE	100.00%
		28	VAL	100.00%
		72	HIS	100.00%
		26	ASN	100.00%
		40	PHE	35.00%
		39	SER	10.00%
119	THR	27	ILE	100.00%
		26	ASN	100.00%
		72	HIS	100.00%
		76	ALA	90.00%
		80	LYS	40.00%
120	PRO	25	PHE	100.00%
		26	ASN	100.00%
		42	LEU	85.00%
		27	ILE	50.00%
		80	LYS	25.00%
121	VAL	24	GLY	100.00%
		25	PHE	100.00%
		23	LEU	100.00%
		79	LEU	100.00%
		22	GLY	100.00%
		80	LYS	95.00%
		27	ILE	90.00%
		76	ALA	85.00%
		26	ASN	35.00%

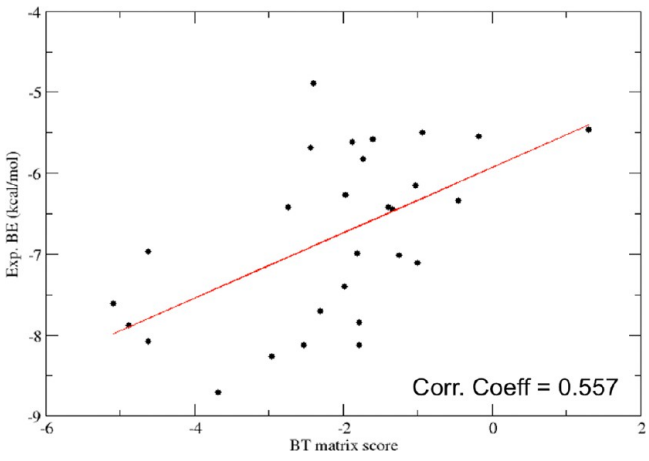


Figure 10. Correlation of BT matrix score with experiments. Correlation plot between experimentally determined binding energy values and pair potential scores evaluated by BT matrix for 28 PDZ−peptide complexes.

improved significantly from 0.06 (in Supporting Information Figure S1) to 0.557, when binding pockets identified by MD simulations were used. This result suggests that if correct binding pocket is identified for a given PDZ domain, it will be possible to search for its interaction partners using less compute intensive statistical pair potentials. Thus, a multiscale approach involving atomically detailed simulations on representative PDZ−peptide complexes can potentially identify correct binding pockets and subsequently statistical pair potential based approach can be used for genome scale search of interaction partners.

Benchmarking of the Multiscale Approach on Mouse PDZ Data Set. We wanted to investigate whether the binding pockets identified by all atom molecular dynamics simulations can be utilized to predict interaction partners of mouse PDZ domains by using statistical pair potentials. Since the experimental binding affinity data for 85 mouse PDZ domains and 217 mouse proteome derived peptides were available from experimental studies by Stiffler et al.,[15] the predictions by our multiscale approach can indeed be benchmarked. Supporting Information Figure S7 shows the flowchart of the methodology for the identification of binding partners of mouse PDZ domains and descriptions of various steps are given in the Material and Methods section. Figure 11 shows the structure-based multiple sequence alignment of MD simulated 1TP3 and 23 crystal structures of other PDZ domains which can be used as structural templates for 64 out of the 82 mouse PDZ domains. The alignment also highlights the binding pocket residues on 1TP3, and thus, equivalent residues on the other 23 structural templates could be found to map the binding pockets on to each of the structural templates. The binding pocket residues for each of the 64 mouse PDZ domains could be obtained from alignment with the respective structural templates. By this approach, binding pockets for 64 PDZ domains were constructed and BT pair potential matrix was then used for scoring the binding energy for each of the 217 peptides with these 64 PDZ domains. ROC curve analysis was carried out for the binding energy scores of 13 888 PDZ−peptide pairs by varying the score cutoff for binder vs nonbinder classification and comparing the prediction scores to the experimental results. Peptides having experimental binding affinity values higher than 100 μM were considered as nonbinders. Figure 12 shows the ROC curve for the 13 888 PDZ−peptide pairs arising from 217 peptides and 64 PDZ domains. As can be seen, the AUC value for the ROC curve is 0.780, indicating the statistically significant predictive power of our computational protocol. The ROC curve indicates that the optimum cutoff value for BT pair potential which will have maximal TPR at minimum FPR is −1.5. Supporting Information Table S2 lists the confusion matrix and Cohen's kappa values for pair potential based identification of PDZ binding partners using the cutoff of −1.5. As can be seen on this large data set pair potential based prediction has a kappa value of 0.44 ± 0.007. We also analyzed the predictions for the each of the 64 PDZ domains individually. Supporting Information Figure S8 shows the bar graphs of AUC values obtained from ROC analysis of the binding predictions for 217 peptides for each of the 64 mouse PDZ domains. Figure S8 clearly shows that the AUC values are close to 0.8 for most of these 64 PDZ domains and only in 3 cases the AUC values are between 0.6 and 0.7. These results are indeed encouraging because binding pockets derived from all atom simulations on 1TP3 could successfully predict the binding partners for very diverse classes of PDZ domains for large number of peptides. This is primarily because of the high degree of structural conservation in PDZ domains despite large variation

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143−1156

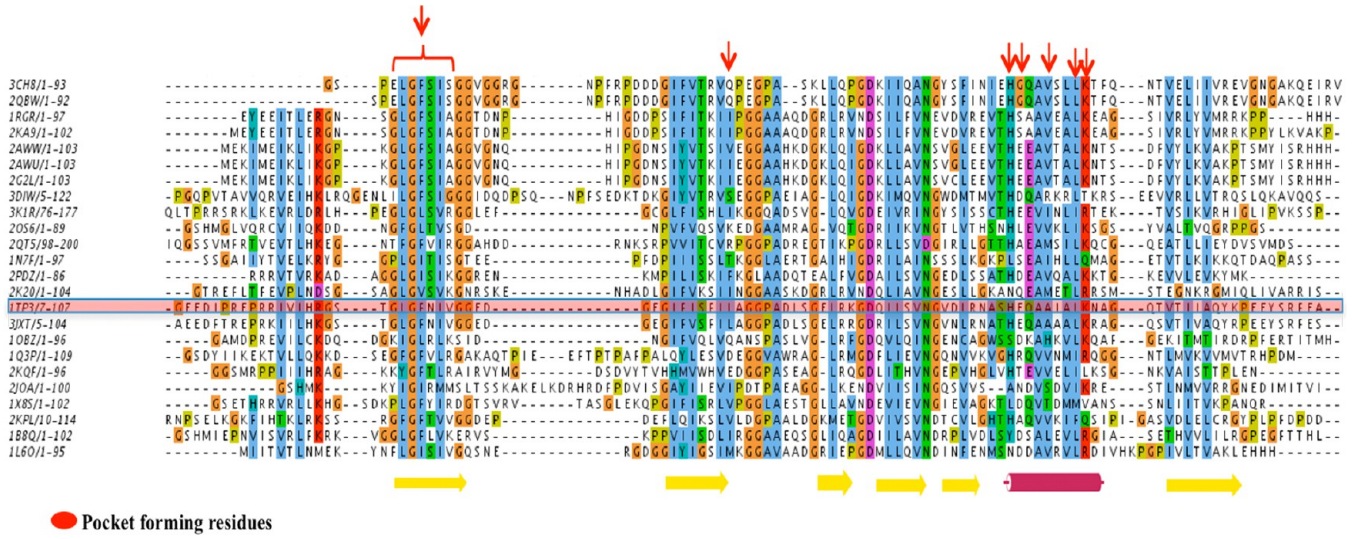


Figure 11. Mapping binding pocket residues on PDZ domains. Structure-based multiple sequence alignment of the crystal structures of 24 PDZ domains, which are used as structural templates for identifying binding pocket residues in the mouse PDZ domains. The residues in various PDZ domains, which align with binding pocket residues (identified by MD simulations) of 1TP3 are marked in red colored arrows, and they constitute the binding pockets in respective PDZ domains.

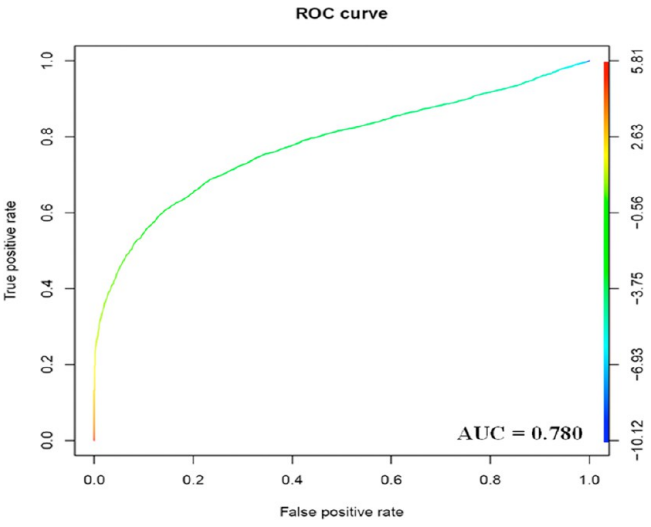


Figure 12. ROC Analysis. ROC plot for all 13 888 PDZ–peptide pairs arising from 217 peptides and 64 PDZ domains.

in sequence. The results from our current study indicates that this multiscale approach for prediction of interaction partners of PDZ domains can potentially be applied to genome scale prediction of PDZ interaction partners on other organisms.

DISCUSSION AND CONCLUSION

The high-resolution structural information on protein–protein interactions is only available for a limited number of proteins from few organisms. Second, the structure based analysis of protein–protein interactions involving molecular dynamics simulations, docking, etc. are extremely computationally intensive and, hence, not specifically suitable for application to genome scale analysis of protein interaction networks. Therefore, it is necessary to develop novel multiscale approaches where crucial specificity determining residues would be identified from atomistic simulations on representative modular interaction domains, and based on these results, simple knowledge based scoring functions can be developed for proteomewide identification of potential interaction partners. Hence, in this work a novel structure based multiscale protocol was developed for predicting binding partners of PDZ domains.

Analysis of available structures of PDZ–peptide complexes in PDB revealed that despite high sequence divergence PDZ domains have highly conserved three-dimensional structure and the substrate peptides bind in a similar conformation and orientation. The structure of third PDZ domain of PSD-95 belonging to one of these 11 clusters was selected as a representative for detailed all atom simulations in complex with different substrate peptides. In order to analyze the conformational flexibility of PDZ–peptide complex, a 50 ns explicit solvent molecular dynamics (MD) simulation was carried out on this representative PDZ–peptide complex. Analysis of the MD trajectory revealed that apart from the residues of the PDZ domain which were in contact with the substrate peptide in the static crystal structure, additional contacts between substrate peptide and PDZ domain were formed within first few nanoseconds and they remained stable during the entire duration of the simulation. Therefore, in order to further understand the molecular basis of binding specificity, multiple explicit solvent all atom MD simulations of 5 ns duration were carried out on third PDZ domain of PSD-95 in complex with 28 different substrate peptides complexes with experimentally known binding affinities. In order to enhance the conformational sampling of the peptide in complex with the PDZ domain, each of the 28 MD simulations of 5 ns duration was repeated six times starting from different initial velocities, thus giving a total simulation of 30 ns for each PDZ–peptide complex. Interestingly, binding free energy values calculated from these explicit solvent MD trajectories using MM/PBSA approach showed good correlation with the experimental dissociation constants. Next, an attempt was made to investigate whether the specificity determining residues (SDRs) of PDZ domains identified by our all atom MD simulations on PSD-95 can be used to develop genome scale approaches involving residue based statistical pair potentials for predicting differential specificities of various PDZ domains for their interaction partners. Our analysis also indicated that binding energy scores between PDZ domains and peptides

dx.doi.org/10.1021/ci400627y | J. Chem. Inf. Model. 2014, 54, 1143–1156

computed by Betancourt and Thirumalai[30] residue based statistical pair potential and MD simulation derived SDRs showed good correlation with experimental binding affinities. This multiscale approach was then benchmarked on mouse PDZ data set by calculating the binding energies for 217 different substrate peptides in binding pockets of 64 different mouse PDZ domains using Betancourt and Thirumalai[30] residue based statistical pair potentials. The robustness of the protocol for prediction of binding partners of PDZ domains was assessed by a rigorous analysis using ROC curves. ROC analysis indicated that the AUC values for binder vs nonbinder classification by our structure-based method is 0.780. This multiscale approach can potentially be applied to human proteome as well to understand the binding selectivity of PDZ domains and reconstruct the PDZ interaction networks.

It may be noted that, in this work we have derived binding pocket profiles of PDZ domains based on atomistic simulations on representative PDZ−peptide complexes and represented them as sets of residue pairs for each pocket. This choice of representation of the binding pocket as residue pairs has been guided by the fact that, with such a representation binding of peptides with different sequences in the pocket can easily be evaluated by statistical pair potentials. Even though this simple representation has successfully captured the specificity profile of mouse PDZ domains, for accurate description of the complex specificity landscape of PDZ domains it will be necessary to represent binding pocket profiles with more structural details. Structure−interaction fingerprints (SIF) have been successfully used to represent binding pocket profiles of kinases and other protein−ligand complexes.[46] Similarly apart of using statistical pair potentials, structure based scoring functions can also be trained using experimental data. It would be interesting to investigate if use of SIFs and other types of scoring functions can help in development of more accurate predictors of PDZ interaction partners.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1: Correlation plot between experimental binding energy and pair potential. Figure S2: RMSD vs time plot for MD simulations on 1TP3. Figure S3: RMSF plot for MD simulations on 1TP3. Figure S4: Peptide deviations from the binding site. Figure S5: Bar graph showing MM/PBSA binding energy values for each simulation. Figure S6: Effect of trajectory length on MM/PBSA energy values. Figure S7: Protocol for identifying interaction partners of mouse PDZ domains. Figure S8: Plot showing AUC values for individual mouse PDZ domains. Table S1 and S2: Confusion matrix for MM/PBSA and pair potential based prediction of PDZ binders. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: deb@nii.res.in. Phone: +91-11-26703749.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work has been supported by grants to NII from Department of Biotechnology (DBT). D.M. also acknowledges financial support from DBT, India, under the BTIS project and National Bioscience Career Development award. The authors thank the Director, NII, for encouragement and support.

■ REFERENCES

(1) Hung, A. Y.; Sheng, M. PDZ domains: structural modules for protein complex assembly. *J. Biol. Chem.* **2002**, *277*, 5699−702.

(2) te Velthuis, A. J.; Sakalis, P. A.; Fowler, D. A.; Bagowski, C. P. Genome-wide analysis of PDZ domain binding reveals inherent functional overlap within the PDZ interaction network. *PloS One* **2011**, *6*, e16047.

(3) Klosi, E.; Saro, D.; Spaller, M. R. Bivalent peptides as PDZ domain ligands. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6147−50.

(4) Verpy, E.; Leibovici, M.; Zwaenepoel, I.; Liu, X. Z.; Gal, A.; Salem, N.; Mansour, A.; Blanchard, S.; Kobayashi, I.; Keats, B. J.; Slim, R.; Petit, C. A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. *Nat. Genet.* **2000**, *26*, 51−5.

(5) Boerkoel, C. F.; Takashima, H.; Stankiewicz, P.; Garcia, C. A.; Leber, S. M.; Rhee-Morris, L.; Lupski, J. R. Periaxin mutations cause recessive Dejerine-Sottas neuropathy. *Amer. J. Human Genet.* **2001**, *68*, 325−33.

(6) Sherman, D. L.; Fabrizi, C.; Gillespie, C. S.; Brophy, P. J. Specific disruption of a schwann cell dystrophin-related protein complex in a demyelinating neuropathy. *Neuron* **2001**, *30*, 677−87.

(7) Gfeller, D.; Butty, F.; Wierzbicka, M.; Verschueren, E.; Vanhee, P.; Huang, H.; Ernst, A.; Dar, N.; Stagljar, I.; Serrano, L.; Sidhu, S. S.; Bader, G. D.; Kim, P. M. The multiple-specificity landscape of modular peptide recognition domains. *Mol. Syst. Biol.* **2011**, *7*, 484.

(8) Tonikian, R.; Zhang, Y.; Sazinsky, S. L.; Currell, B.; Yeh, J. H.; Reva, B.; Held, H. A.; Appleton, B. A.; Evangelista, M.; Wu, Y.; Xin, X.; Chan, A. C.; Seshagiri, S.; Lasky, L. A.; Sander, C.; Boone, C.; Bader, G. D.; Sidhu, S. S. A specificity map for the PDZ domain family. *PLoS Biol.* **2008**, *6*, e239.

(9) Jemth, P.; Gianni, S. PDZ domains: folding and binding. *Biochemistry* **2007**, *46*, 8701−8.

(10) Songyang, Z.; Fanning, A. S.; Fu, C.; Xu, J.; Marfatia, S. M.; Chishti, A. H.; Crompton, A.; Chan, A. C.; Anderson, J. M.; Cantley, L. C. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **1997**, *275*, 73−7.

(11) Stricker, N. L.; Christopherson, K. S.; Yi, B. A.; Schatz, P. J.; Raab, R. W.; Dawes, G.; Bassett, D. E., Jr.; Bredt, D. S.; Li, M. PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. *Nat. Biotechnol.* **1997**, *15*, 336−42.

(12) Nourry, C.; Grant, S. G.; Borg, J. P. PDZ domain proteins: plug and play! *Sci. STKE: Signal Transduction Knowledge Environ.* **2003**, *2003*, RE7.

(13) Lee, H. J.; Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signal.: CCS* **2010**, *8*, 8.

(14) Doyle, D. A.; Lee, A.; Lewis, J.; Kim, E.; Sheng, M.; MacKinnon, R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **1996**, *85*, 1067−76.

(15) Stiffler, M. A.; Chen, J. R.; Grantcharova, V. P.; Lei, Y.; Fuchs, D.; Allen, J. E.; Zaslavskaia, L. A.; MacBeath, G. PDZ domain binding selectivity is optimized across the mouse proteome. *Science* **2007**, *317*, 364−9.

(16) Saro, D.; Li, T.; Rupasinghe, C.; Paredes, A.; Caspers, N.; Spaller, M. R. A thermodynamic ligand binding study of the third PDZ domain (PDZ3) from the mammalian neuronal protein PSD-95. *Biochemistry* **2007**, *46*, 6340−52.

(17) Chen, J. R.; Chang, B. H.; Allen, J. E.; Stiffler, M. A.; MacBeath, G. Predicting PDZ domain-peptide interactions from primary sequences. *Nat. Biotechnol.* **2008**, *26*, 1041−5.

(18) Shao, X.; Tan, C. S.; Voss, C.; Li, S. S.; Deng, N.; Bader, G. D. A regression framework incorporating quantitative and negative interaction data improves quantitative prediction of PDZ domain-peptide interaction from primary sequence. *Bioinformatics* **2010**, *27*, 383−90.

(19) Hui, S.; Bader, G. D. Proteome scanning to predict PDZ domain interactions using support vector machines. *BMC Bioinf.* **2010**, *11*, 507.
(20) Kim, J.; Kim, I.; Yang, J. S.; Shin, Y. E.; Hwang, J.; Park, S.; Choi, Y. S.; Kim, S. Rewiring of PDZ domain-ligand interaction network contributed to eukaryotic evolution. *PLoS Genet.* **2012**, *8*, e1002510.
(21) Luck, K.; Fournane, S.; Kieffer, B.; Masson, M.; Nomine, Y.; Trave, G. Putting into practice domain-linear motif interaction predictions for exploration of protein networks. *PloS One* **2011**, *6*, e25376.
(22) Hui, S.; Xing, X.; Bader, G. D. Predicting PDZ domain mediated protein interactions from structure. *BMC Bioinf.* **2013**, *14*, 27.
(23) Kaufmann, K.; Shen, N.; Mizoue, L.; Meiler, J. A physical model for PDZ-domain/peptide interactions. *J. Mol. Model.* **2010**, *17*, 315−24.
(24) Smith, C. A.; Kortemme, T. Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *J. Mol. Biol.* **2010**, *402*, 460−74.
(25) Mostarda, S.; Gfeller, D.; Rao, F. Beyond the binding site: the role of the beta(2)-beta(3) loop and extra-domain structures in PDZ domains. *PLoS Comput. Biol.* **2012**, *8*, e1002429.
(26) Niv, M. Y.; Weinstein, H. A flexible docking procedure for the exploration of peptide binding selectivity to known structures and homology models of PDZ domains. *J. Am. Chem. Soc.* **2005**, *127*, 14072−9.
(27) Kumar, N.; Mohanty, D. MODPROPEP: a program for knowledge-based modeling of protein-peptide complexes. *Nucleic Acids Res.* **2007**, *35*, W549−55.
(28) Kumar, N.; Mohanty, D. Identification of substrates for Ser/Thr kinases using residue-based statistical pair potentials. *Bioinformatics* **2010**, *26*, 189−97.
(29) Kumar, N.; Mohanty, D. Structure-based identification of MHC binding peptides: Benchmarking of prediction accuracy. *Mol. Biosyst.* **2010**, *6*, 2508−20.
(30) Betancourt, M. R.; Thirumalai, D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **1999**, *8*, 361−9.
(31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−42.
(32) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*; 2002; Chapter 2, Unit 2 3.
(33) Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596−9.
(34) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−10.
(35) Holm, L.; Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **2010**, *38*, W545−9.
(36) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **2003**, *12*, 2001−14.
(37) Case, D. A.; Darden, T. A.; Cheatham, T. E., I; Simmerling, C. L.; Wang J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; Universiy of California: San Francisco, 2006.
(38) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 10.
(40) Darden, T. A.; Pedersen, L. G. Molecular modeling: an experimental tool. *Environ. Health Perspect.* **1993**, *101*, 410−2.
(41) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−97.
(42) Wu, S.; Zhang, Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* **2008**, *72*, 547−56.
(43) Pei, J.; Kim, B. H.; Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **2008**, *36*, 2295−300.
(44) Penkert, R. R.; DiVittorio, H. M.; Prehoda, K. E. Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1122−7.
(45) Tiwari, G.; Mohanty, D. An in silico analysis of the binding modes and binding affinities of small molecule modulators of PDZ−peptide interactions. *PloS One* **2013**, *8*, e71340.
(46) Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein−ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337−44.