PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior

Z. Nevin Gerek,[1,2] Ozlem Keskin,[3] and S. Banu Ozkan[1,2]*

[1] Center for Biological Physics, Arizona State University, Tempe, Arizona
[2] Department of Physics, Arizona State University, Tempe, Arizona
[3] Department of Chemical and Biological Engineering, Koc University, Istanbul, Turkey

ABSTRACT

PDZ domains (PDZs), the most common interaction domain proteins, play critical roles in many cellular processes. PDZs perform their job by binding specific protein partners. However, they are very promiscuous, binding to more than one protein, yet selective at the same time. We examined the binding related dynamics of various PDZs to have insight about their specificity and promiscuity. We used full atomic normal mode analysis and a modified coarse-grained elastic network model to compute the binding related dynamics. In the latter model, we introduced specificity for each single parameter constant and included the solvation effect implicitly. The modified model, referred to as specific-Gaussian Network Model (s-GNM), highlights some interesting differences in the conformational changes of PDZs upon binding to Class I or Class II type peptides. By clustering the residue fluctuation profiles of PDZs, we have shown: (i) binding selectivities can be discriminated from their dynamics, and (ii) the dynamics of different structural regions play critical roles for Class I and Class II specificity. s-GNM is further tested on a dual-specific PDZ which showed only Class I specificity when a point mutation exists on the βA-βB loop. We observe that the binding dynamics change consistently in the mutated and wild type structures. In addition, we found that the binding induced fluctuation profiles can be used to discriminate the binding selectivity of homolog structures. These results indicate that s-GNM can be a powerful method to study the changes in binding selectivities for mutant or homolog PDZs.

Proteins 2009; 77:796–811.

Key words: binding; elastic network model; PDZ domain; dynamics; selectivity.

INTRODUCTION

Protein–protein interactions are the key elements in organizing functional protein complexes and controlling cellular activities. Among the many protein–protein interaction domains, PDZ domains (PDZs) are one of the most commonly found in organisms from bacteria to humans.[1,2] By binding to the specific peptide sequence motif of the target proteins, PDZs assist in assembling these proteins into supramolecular complexes. One important characteristic of PDZs is that they are very promiscuous, that is, they can bind to more than one target sequence motif.[2–4] Yet they are also specific, discriminating which sequence motif to bind. The increase in number of experimental studies related to PDZ domain mediated protein–protein interactions[3,5–20] creates a need for theoretical models that provide quantitative and biologically relevant understanding to the experimental observations as discussed below. Quantitative understanding of the principles that underlies promiscuity and selectivity of PDZ domains will shed light into the recognition and binding mechanisms.

Structural basis of PDZ binding and specificity

All PDZs have similar overall secondary and tertiary structures with an average backbone root mean square deviation of around 1.4 Å.[21] The common structure of PDZ domain comprises six β-strands (βA-βF) and two α-helices (αA and αB), which fold in an overall six-stranded β sandwich [Fig. 1(A)]. They typically recognize specific amino acids in the C-terminal end of peptide motifs or target proteins. In addition to binding to carboxyl peptides, PDZs can interact with internal peptide motifs of target proteins, with another PDZ domain[1–3] and even with lipids.[22,23] The phenomenon of binding to the C-terminal end of peptide motifs is known as the classical or canonical binding mode. C-terminal peptides bind as an antiparallel β-strand in a groove between the βB strand and the αB helix, in essence extending the β-sheet. The conserved sequence Gly-Leu-Gly-

Additional Supporting Information may be found in the online version of this article.
*Correspondence to: S. Banu Ozkan, Arizona State University, Department of Physics, P.O. Box 871504, Tempe, AZ 85287-1504. E-mail: Banu.Ozkan@asu.edu
Received 8 January 2009; Revised 6 May 2009; Accepted 25 May 2009
Published online 2 June 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22492

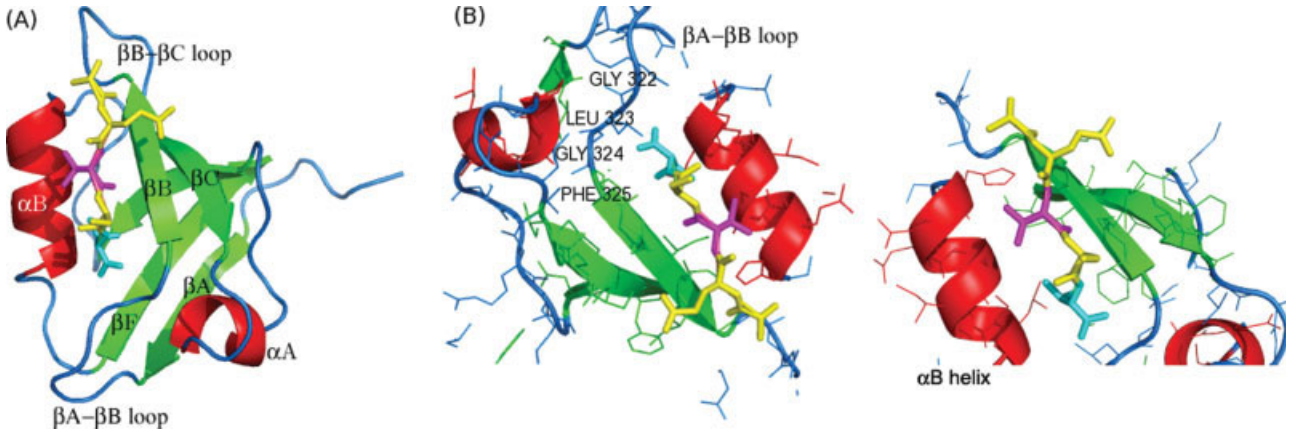

Figure 1

Representative structure of a PDZ domain (PSD-95) in complex with its target peptide (PDB ID: 1BE9). (**A**) The common structure of PDZ domains contains six β strands (green) and two α-helices (red) with the peptide (in stick form). PDZ domain specifically recognizes short C-terminal peptide motifs that bind in an extended groove between second α-helix (αB) and the second β-strand (βB) and (**B**) the interactions of the peptide with the αB helix and GLGF segment of the βA-βB loop (**C**) Details of the binding region. Structural figures were produced with PyMOL.[24]

Leu (or Phe) between βA-βB loop and αB helix is directly involved in peptide recognition and binding [Fig. 1(B)].

PDZs have been generally categorized into three different classes based on the last four residues at the sequence of the C-terminal peptide.[25] Class I PDZs bind to a C-terminal motif with the sequence [Ser/Thr-X-Φ-COOH], Class II PDZs prefer the sequence [Φ-X-Φ-COOH], and Class III types interact with the sequence [Asp/Glu-X-Φ-COOH], where X is any amino acid and Φ is any hydrophobic amino acid. All classes have a preference for a hydrophobic residue at the very C-terminal (p_0) position. In contrast to serine or threonine residues at p_{-2} position in Class I, Class II PDZs prefer bulky hydrophobic residues [Fig. 1(C)]. However, several PDZ-mediated interactions do not conform to these three major canonical types of recognition and new classes are introduced to account for them.[26] Consequently, it has been also shown that the classification based on the chemical properties of residues at the p_0 and p_{-2} positions within these C-terminal targets fail to predict the specificity of binding.[27,28]

As a result, a very interesting and challenging problem comes up: what makes PDZs both promiscuous and selective when they share similar primary and tertiary structures and have a conserved binding site[25]? Moreover, the residues interacting with the C-terminal peptide are mostly conserved. Therefore, the key questions remaining to be understood are: what is the underlying mechanism that determines the interaction with specific peptide sequences and how can we predict the binding selectivities of PDZ domains-C-terminal peptide interactions?

Role of dynamics in PDZs

Experimental[29] and computational studies[30–39] have shown that the dynamics of PDZs must play a crucial role toward their binding selectivity. Basdevant *et al.*[30] analyzed 12 different PDZ domain complexes using molecular dynamics simulations to identify the sources of selectivity and promiscuity underlying their interactions. They found that PDZ interactions favor nonpolar binding contributions when compared with electrostatic ones. Another computational study of three different PDZs (hPTP1E, ZASP, and PSD 95) demonstrated that a single or at most a couple of low-frequency collective domain movements captures the deformation associated with the binding of a peptide.[32] From an evolutionary approach, it has been shown that PDZs use minor conformational changes to propagate signals from one to functionally important distant sites through allosteric regulation.[21] In addition, an experimental study of hPTP1E indicated the existence of a network of residues that has a dynamic response to a ligand binding.[29] Recently, Dhulesia *et al.*[33] extended this study using constrained molecular dynamics simulations and provided a detailed characterization of the relation between structural and dynamical changes upon binding and identified the most connected residues based on a structural network. All these studies have shown that characterization of PDZ domain dynamics can provide insights to the relationship between their promiscuity and selectivity.

In this study, we investigate how the dynamics of unbound states of PDZs contribute to the binding selectivity. Recent NMR dynamic studies suggest that (a) the bound conformation is already present among the conformational sub-states of the protein before a ligand

binds[40] and (b) the stabilization of pre-existing conformations might be the fundamental paradigm for ligand binding.[40–42] Thus, the binding induced dynamics of unbound states can provide insights about binding selectivities of PDZs.

We use full atomic normal mode analysis (NMA) and a coarse-grained elastic network models (ENM) to compute the residue fluctuations of a large set of PDZs. Elastic network models have been effectively used to understand the fluctuation dynamics of proteins for a decade.[43–53] In addition, elastic network models have also shown that it is possible to obtain the ligand-induced conformational changes based on one single slowest fluctuation profile or a combination of some, or all fluctuation profiles of an unbound conformation.[54–60]

In ENMs, the folded structure of a protein is assumed to be a three-dimensional elastic network subject to a uniform, single-parameter harmonic potential between all residue pairs, bonded or nonbonded, located within a certain cutoff distance. If the fluctuations are assumed to be isotropic with no directional preference, the model is referred to as the Gaussian Network Model (GNM).[44,45] In this model, high frequency modes give kinetically important residues (i.e., residues important for folding and binding),[59,61–64] whereas low frequency modes give the global motions of the proteins which are known to be important for the protein function.[44]

GNM is simply an effective model to elucidate the collective motions of proteins based on the structural topology; however, it lacks specificity, which is important to predict binding selectivity differences within the same protein family. Therefore, we modified GNM so that it takes into account the solvent interactions and residue interaction specificity. The extended model, called the specific Gaussian Network Model (s-GNM), is tested on a set of PDZs to predict the classification and the binding specificity based on the collective fluctuations near the native or equilibrium state. We further compare it with full atomic NMA, Anisotropic Network Model (ANM) and GNM. We observe that all-atom NMA calculations are consistent with the coarse-grained s-GNM results. Thus, the remarkable agreement between the two methods allows us to use s-GNM satisfactorily in further analyses.

The intrinsic fluctuations of a set of unbound Class I and II PDZs are compared with their ligand induced experimental conformational changes. By clustering the fluctuation profiles of Class I and Class II type PDZs, we also analyze the dynamics that induces Class I and Class II specificities. Our results show that the binding specificity of PDZs can be related to their dynamics[30–33,37,39] and the mobility of different structural regions play important roles for binding to Class I and Class II peptides. In addition, we observe that changes in binding selectivity upon mutation can also be captured by the fluctuation profiles obtained by s-GNM. We have also investigated how s-GNM discriminates the binding selectivity when it is applied to homology models and found that the binding induced fluctuation profiles of homolog models can be used to discriminate their binding selectivity.

MATERIAL AND METHODS

The set of PDZ domain structures analyzed in this study is listed in Table I. The bound (closed) and unbound (open) structures of PDZs, obtained from the Protein Data Bank (PDB),[67] are listed in the first column. The second column displays the names of the corresponding proteins. The sequences of the binding peptides, the class of PDZ domain according to their binding specificity and the backbone root-mean square deviation (RMSD) between unbound-bound structures are listed in the following columns of the table. In this study, we focus on the most canonical classification of Class I and Class II type of interactions due to the fact that we can find many Class I and Class II type PDZs whose experimental structures are available. On the other hand, there are only a few Class III PDZs; and it is hard to draw conclusions from such scarce data.

Table I
List of PDZ Domain Proteins Analyzed

PDB code	Name of the protein	Peptide sequence	PDZ class	RMSD (Å)[a]
(unbound–bound)				
1g9o–1gq4	NHERF-PDZ1	-NDSLL	I	1.51
1g9o–1gq5	NHERF-PDZ1	-EDSFL	I	1.94
1g9o–1i92	NHERF-PDZ1	-QDTRL	I	1.98
1bfe–1be9	PSD-95-PDZ3	-KQTSV	I	0.77
1q3o–1q3p	Shank	-EAQTRL	I	0.91
1n7e–1n7f	GRIP-PDZ6	-ATVRTYSC	II	0.58
1nte–1obx	Syntenin-PDZ2	-(ETLE)DSVF	I	0.48
1nte–1oby	Syntenin-PDZ2	-TNEFYA	II	0.42
2h3l–1n7t[b]	Erbin	-TGWETWV	I	1.59
2h3l–1mfg	Erbin	-EYLGLDVPV	II	0.59
1z86[b]–2pdz[b]	α1-Syntrophin	-GVKESLV	I	2.50
1t2m[b]–2ain[b]	AF-6	-LFSTEV	I	1.49
2ego–2egn	Tamalin	-SSSSL	I	3.25
1u37–1u38	X11/Mint	-PVYI	III	1.60
1gm1[b]–1vj6[b]	PTP-BL-PDZ2	-KRHSGSYLVTSV	I	1.77
2awx–2aww	SAP97	-ATGL	I	3.51
3pdz[b]–1d5g[b]	hPTP1E-PDZ2	-EQVSAV	I	2.14
2ev8[b]–2ejy[b]	Erythrocyte p55	-RKEYCI	II	2.54
(only unbound)				
1kwa[b]	hCASK	-EYYV -EFYA	II	–
2fe5[b]	DLG-PDZ2	-ETDV	I	–
2f5y[b]	RGS3	-YYKV	II	–
1x45[b,c]	X11-M1	-GEPL	III	–
2yt7[b,c]	X11-M3	-GEPL	III	–
(only bound)				
2gzv	PICK1	-YYKV	II	–

[a]Backbone root-mean square deviation (RMSD) between unbound-bound structures based on the α carbons.
[b]Structures solved by NMR.
[c]The bound structures of these proteins are not available but there are experimental studies of peptide sequences that bind to hCASK,[8] DLG-PDZ2,[25] RGS3,[65] X11-M1,[66] and X11-M3.[66]

Full atomic normal mode analysis

In this study, the VIBRAN module in the CHARMM simulation program[68] is used to calculate the normal modes of the structures. The full atomic Hessian matrix of each structure is generated with the all-atom CHARMM22 force field. The solvent effect is introduced into NMA using a distance-dependent dielectric factor (RDIE) with eps = 4r or effective energy function 1 (eef1.1).[69] The mass weighted Hessian matrix is diagonalized using the DIAG module implemented in CHARMM. The structure is relaxed by energy minimization, including 10,000 steepest descent minimization steps followed by 10,000 steps of adapted basis Newton-Raphson (ABNR) minimization algorithm[70] with gradually decreasing harmonic constraints. This is followed by 20,000 steps of ABNR minimization algorithm until an energy gradient reached a value of 10^{-10} kcal/mole Å. There are no modes with negative eigenvalues and six modes with near-zero frequency.

Anisotropic network model (ANM)

ANM is an extension of the original Gaussian network model (GNM) that incorporates the anisotropic fluctuations of a protein's residues.[43] It is equivalent to a NMA with an elastic network model at the C^{α} level and the Hessian is based on a harmonic potential form.[71] ANM provides information on the directions of fluctuations. The detailed theory of the ANM can be found in Atilgan *et al.*[43] Here, we used a cutoff value of 10 Å.

Specific-Gaussian network model (s-GNM)

The details of Gaussian network model (GNM) have been given in different references.[44,45,72] In the model, each residue is represented with its α-carbon atom. The topology of the network of N residues is defined by the adjacency or Kirchhoff matrix, the elements of which are given by,

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \quad \text{and} \quad R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \quad \text{and} \quad R_{ij} > r_c \\ -\sum_{j,j\neq i} \Gamma_{ij}, & \text{if } i = j \end{cases} \qquad (1)$$

where r_c is the cut-off distance and R_{ij} is the distance between ith and jth α-carbons. Because GNM (i) assigns a uniform strength for each interaction without considering the specificity and (ii) treats deeply buried residues and those exposed to the solvent equally, it is nonspecific to protein amino acid sequence and solvent effects. In this study, we modify GNM by introducing specificity to predict the binding mechanism more precisely. How well GNM predicts the features of collective protein motion is usually measured by the comparison of the fluctuation profiles obtained by GNM with experimental crystallographic B-factors and NMR relaxation data. Recent work shows that GNM yields a correlation of 0.59 with crystallographic B-factors[73] and higher correlation (0.75) with NMR data. However, the correlation decreases drastically for the structures with a large difference between the number of buried and solvent exposed residues.[48] In the case of PDZs, the correlation between the fluctuation profiles computed by GNM and experimental data is low (average correlation coefficient of 0.52; and even as low as 0.17 in some PDZs). To modify GNM, first, we introduce an additional term (i.e., additional node in the interaction network) that will take into account solvent interactions based on average solvent accessibility of each residue. In the original matrix, the nodes are connected with springs, if the pairs of residues are located within an interaction range, or cutoff distance. In the modified version of GNM, called specific GNM (s-GNM), there are $N + 1$ nodes for a protein of N amino acids. This extra node denotes the solvent interaction of each node (i.e., each amino acid), and the strength of the spring that connects the $(N + 1)$th node with rest of the nodes are determined based on the solvent accessible surface area of the amino acid. Accessible surface area values are determined using Surface Racer program[74] with a probe radius of 1.4 Å, corresponding to the size of a water molecule. They are normalized by the total accessible surface area of the protein. The contribution of the solvation interactions is controlled with an adjustable parameter α. The value of α is obtained by optimizing the correlation between the experimental (i.e., the fluctuations in different models of NMR structures or B-factors in X-ray structures) and the theoretical fluctuation profiles. The range of the optimized α parameter has been found between 0.003 for X-ray structures and 0.001 for NMR structures.

Second, we adjust the pairwise residue interaction strength specific to type of interaction using statistical contact potentials (Miyazawa and Jernigan[75] or Thomas and Dill[76] statistical potentials). In a recent study, Phillips and coworkers[48] introduced the chemical network model (CNM) where residue interactions are divided into bonded and nonbonded types. Similarly, to assign specific interaction strength parameters for the springs, we also define two different spring constants in the residue interaction network for bonded and nonbonded interactions, γ_b and γ_{nb}, respectively.

$$\Gamma_{ij} = \begin{cases} -\gamma_b, & \text{if } ij \text{ bonded} \quad \text{and} \quad R_{ij} \leq r_c \\ -\gamma_{nb}, & \text{if } ij \text{ nonbonded} \quad \text{and} \quad R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \quad \text{and} \quad R_{ij} > r_c \\ -\sum_{j,j\neq i} \Gamma_{ij}, & \text{if } i = j \end{cases} \qquad (2)$$

The strength parameter of nonbonded interactions is further adjusted based on the inter-residue interaction potentials of Miyazawa and Jernigan[75] or Thomas and Dill.[76] In this study, we tested our model using both the Miyazawa-Jernigan and the Thomas-Dill knowledge-based potentials. Although there are differences in developing

these two statistical potentials, our results do not change with the choice of the statistical potential. The results using the Thomas-Dill potentials are reported here.

The statistical inter-residue interaction potential between the ith and jth nonbonded residues, ε_{ij}, is incorporated in a discretized way:

$$\gamma_{nb} = \begin{cases} -1.5^{*}\gamma_{nb}, & -2.0 < \varepsilon_{ij} < -0.6 \\ -\gamma_{nb}, & -0.6 < \varepsilon_{ij} < -0.01 \\ -0.5^{*}\gamma_{nb}, & -0.01 < \varepsilon_{ij} < 0.01 \\ 0, & \varepsilon_{ij} > 0.01 \end{cases} \quad (3)$$

The optimal cutoff distance is taken as 7.5 Å. The spring constant of the bonded part (γ_b) is taken as 1. For the nonbonded part (γ_{nb}), 0.1 is found as the optimized value by maximizing the correlation between the computed fluctuations and experimental crystallographic B-factors. The residue fluctuations determined by the statistical mechanical average of overall fluctuations are obtained from the inverse of the Kirchhoff matrix.

The temperature factor (B-factor) of any residue can be expressed in terms of a weighted sum of the mean-square fluctuations driven by all modes as

$$B_i = \frac{8\pi^2}{3}\left\langle (\Delta R_i)^2 \right\rangle = 8\pi^2 k_B T \left[\Gamma^{-1}\right]_{ii} \quad (4)$$

Both original GNM and s-GNM are written in python. In s-GNM, the scipy optimization module (Powell's method) is used for minimizing the difference between theoretical mean-square fluctuation and experimental B-factors. The experimental B-factors of the X-ray structures and variations in residue positions between NMR models are compared with the residue fluctuations predicted by the GNM and s-GNM. For some PDZs, the agreement between the fluctuation profiles obtained by s-GNM and experimental data is significantly better when it is compared with GNM results (see Table II, columns 4 and 5). Correlations of 0.60, 0.73, and 0.56 are obtained from s-GNM for Na+/H+ exchanger regulatory factor (NHERF), Shank, and Erbin, respectively, whereas GNM provides correlation coefficients of 0.45, 0.73, and 0.17. Overall, the s-GNM increases the correlation (0.58) but not significantly when compared with that of GNM (0.52). We observe that if the correlation coefficient obtained by GNM is greater than 0.6, s-GNM does not improve the correlation. However, it gives statistically significant higher correlation coefficients for the rest. (We obtain a P value of 0.026 for a paired one-tailed t-test). The effects of the crystal environment in B-factors have been shown.[48,57,80–82] We also include crystal effect following the procedure of Kondrashov *et al.*[48] where the crystal contacts are taken into account by adding interactions between residues involved in crystal

Table II
Selected Modes Based on s-GNM, GNM, and NMA Methods and Comparison with Theoretical and Experimental Displacements Between Unbound and Bound Structures and Correlation with B-factors ($R_{B\text{-factor}}$) of the X-ray Structures and Variations in Residue Positions Between NMR Models

PDB code (unbound)	Resolution (Å)	Name of the protein	$R_{B\text{-factor}}$		s-GNM		GNM		NMA		ANM	
			GNM	s-GNM	Mode	R	Mode	R	Mode	R	Mode	R
1g9o	1.50	NHERF	0.45	0.60	1–4	0.64	1–4	0.57	1–9	0.32	1–10	0.33
1bfe	2.30	PSD-95	0.21	0.41	1–6	0.68	1–4	0.47	1–5	0.20	1–6	0.29
1q3o	1.80	Shank	0.73	0.73	1–7	0.47	1–2	0.61	1–2	0.11	1–7	0.61
1n7e	1.50	GRIP	0.54	0.60	1–3	0.56	1–4	0.49	1–3	0.22	1–2	0.59
1nte	1.24	Syntenin-I	0.49	0.55	1–3	0.54	1–2	0.36	1–6	0.65	1–2	0.67
		Syntenin-II[a]			1–3	0.64	1–2	0.57	1–6	0.52	1–2	0.65
2h3l	1.00	Erbin-I	0.17	0.56	1–2	0.60	1–6	0.46	1–10	0.62	1–4	0.53
		Erbin-II[a]			1–2	0.12	2	0.33[a]	1–4	0.31	1–5	0.03
1z86	NMR	α1-Syntrophin	0.71	0.55	1–2	0.63	1–2	0.41	1–7	0.15	1–3	0.54
1t2m	NMR	AF6	0.44	0.48	1–2	0.82	1–2	0.46	1–9	0.40	1–3	0.49
2ego	1.80	Tamalin	0.65	0.65	1–10	0.56	1–7	0.46	1–4	0.09	1–4	0.52
1u37	NMR	X11/Mint	0.63	0.55	1–3	0.53	1–2	0.58	1–4	0.37	1–5	0.38
1gm1[b]	NMR	PTP-BL	0.53	0.55	1–2	0.59	1–2	0.44	–	–	1–8	0.49
2awx	1.80	SAP97	0.67	0.70	1–10	0.45	1–8	0.35	1–10	0.45	1–4	0.43
3pdz	NMR	hPTP1E	0.83	0.76	1–5	0.57	1–6	0.43	1–8	0.37	1–8	0.40
2ev8	NMR	p55	0.45	0.40	1–6	0.52	1–2	0.59	1–7	0.60	1–6	0.59
1kwa[c]	1.93	hCASK	0.12	0.54	1–3	–	–	–	–	–	–	–
2fe5[c]	1.10	DLG-PDZ2	0.81	0.84	1–3	–	–	–	–	–	–	–
2f5y[c]	2.39	RGS3	0.38	0.46	1–16	–	–	–	–	–		–
1x45[c]	NMR	X11-M1	0.47	0.47	1–5	–	–	–	–	–	–	–
2yt7[c]	NMR	X11-M3	0.59	0.61	1–7	–	–	–	–	–	–	–
AVG			0.52	0.58		0.56		0.48		0.36		0.47

[a]Proteins that bind to Class II type peptides.
[b]The structure did not converge in the minimization step of NMA.
[c]The bound structures of these proteins are not available.

contacts (for results, see Table SIII in Supporting Information). For our data set, the average correlation between thermal fluctuations and B-factors does not significantly change, when we include crystal environment (i.e., with the s-GNM, the overall average correlation of the B-factors is 0.59, whereas correlation is 0.56 with the inclusion of crystal contacts). We also note that s-GNM is observed not to improve the correlation coefficients significantly for the NMR structures.

In s-GNM, the motions along different modes are obtained from eigenvalue decomposition ($\Gamma = \mathbf{U}\Lambda\mathbf{U}^{-1}$) of the Kirchhoff matrix. U is the orthonormal matrix of eigenvectors (u_k) of Γ, and Λ is the diagonal matrix of the eigenvalues (λ_k), $1 \leq k \leq N + 1$ (N modes for N residues in the protein and one for the solvent). The eigenvalues λ_k and the eigenvectors $\mathbf{u}_k$ of Γ yield the frequencies and shapes of the N modes of motion ($1 \leq k \leq N$) controlling the dynamics of the system.[44] The fluctuation of residue i from its equilibrium position along the kth principal coordinate can be described with the ith element of $(\mathbf{u}_k)_i$ of the kth eigenvector. The mean-square fluctuation of residue i can be written as a weighted sum of the square fluctuations driven by all modes as

$$\left\langle \Delta \mathbf{R}_i^2 \right\rangle = \sum_k \left[\Delta \mathbf{R}_i^2\right]_k = 3k_B T \sum_k \left[\lambda_k^{-1} (\mathbf{u}_k)_i^2\right] \qquad (5)$$

Displacements along the most cooperative global modes of motion predicted by the model enable us to understand how fluctuations concert at the opposite or same directions as also shown in an elastic network analysis of Tobi and Bahar.[59]

Comparison of theoretical residue fluctuations with experimental ligand-induced conformational changes

To find the binding induced fluctuations, we focus on the contribution of each mode, weighted by the inverse of the corresponding eigenvalue.[83] The issue of how many modes will contribute to the binding dynamics is addressed by evaluating the distribution of mode frequencies (eigenvalues). A subset of modes whose eigenvalues are dispersed from those of the other modes is identified following the procedure also used in the study of Isin *et al.*[84] In this approach, first, the histogram of the eigenvalues is generated then the bin size is computed based on the highest dispersion in eigenvalue spectrum (i.e., the first gap in the eigenvalue spectrum). Eigenvalues corresponding to the first bin are used along with their eigenmodes to compute the weighted sum of the square fluctuations obtained by Eq. (5). The number of modes used for each protein is listed in Table II. The fluctuation profiles are also compared with the experimental conformational changes upon binding.

Experimental conformational changes (residue displacements) are found by superimposing the α-carbons of the unbound and bound structures and calculating the difference between two conformations as

$$\Delta \mathbf{R} = \mathbf{R}_U - \mathbf{R}_B \qquad (6)$$

where $\mathbf{R}_U$ and $\mathbf{R}_B$ are the crystallographic coordinates of the unbound and bound structures. These normalized vectors are compared with normalized residue fluctuations.

Computation of global motions related to binding specificity

NMA-based methods have proven to be useful for identifying the common and distinctive structural and dynamic features of the globin heme binding family,[85] protein superfamilies,[86] protease families,[87] and enzymes.[88] With a similar methodology, we use s-GNM for obtaining the common fluctuation profiles for different classes of PDZs. Because all PDZs have different sequence sizes, in order to compare their dynamics, we first need to find the common regions among the PDZ proteins. Multiple sequence alignments are performed on the segments of the PDZs that are critical for binding: βA-βB loop, βB strand, βB-βC loop, βC strand, and αB helix. Figure S1 in Supporting Information gives the aligned residues of all PDZs (results of the multiple sequence alignments) as obtained using ClustalW for all PDZ domains.[89,90] Second, we form a sub-root-mean-square fluctuation-matrix restricted to common residues, obtained from multiple sequence alignment. The mean-square fluctuation values corresponding to the consecutive aligned residues are extracted using a set of collective modes as explained above [See Eq. (5)]. In this sub-matrix (A), each row corresponds to an aligned residue and each column to a PDZ domain protein. This sub-matrix is formed for both Class I and Class II PDZ domain proteins, separately.

$$A = \left[\begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \cdot \\ \cdot \\ \Delta R_m \end{bmatrix}_1 \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \cdot \\ \cdot \\ \Delta R_m \end{bmatrix}_2 \cdots \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \cdot \\ \cdot \\ \Delta R_m \end{bmatrix}_{n-1} \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \cdot \\ \cdot \\ \Delta R_m \end{bmatrix}_n \right] \qquad (7)$$

where ΔR_i ($i = 1,2 \ldots m$) corresponds to the theoretical root mean square fluctuations of unbound structures from s-GNM, m being the number of aligned residues obtained from ungapped sequence alignments, and n representing the total number of PDZ domains used for the computations.

We then apply Singular Value Decomposition (SVD) to find the dominant dynamics profiles of Class I and Class II proteins. The SVD approach has been used effectively to classify protein structures based on their amino acid composition[91] and to classify anticancer agents by examining their growth inhibitory potential values.[92]

SVD transforms the original data points into a new space with principal axes. The variance along the first axis is highest meaning that it captures the highest information of the original matrix. The original data points can be expressed as a linear combination of a few dominant principal components. Using the SVD method, A is decomposed into three matrices,

$$A = SwD^{T} \quad (8)$$

where S and D^{T}, or D transposed, are left and right singular vectors (or principal axes), and w is a diagonal matrix of singular values. S is the orthonormal transformation matrix that defines the relationship between the original coordinate frame and the new SVD frame, D^{T} is a matrix representing the new coordinates of the original data points in the new SVD space. S defines a common pattern of the collective fluctuations among PDZ domains and each right principal axis indicates the distribution of proteins based on their dynamics profile.

Applying s-GNM to mutant proteins

To understand the effect of a mutation on the dynamics and consequently on the selectivity of PDZs, a computational point mutation is introduced into the X-ray structure of the PICK1 (PDZ protein interacting with C kinase 1) via Swiss PDB Viewer.[93] Before applying the s-GNM method, the mutated structure is subjected to an energy minimization of 50 steepest descent iterations followed by 1000 conjugate gradient iterations using AMBER 96 force field[94,95] to relax the side chains. Then, the same procedure described above is followed for the mutated structure to find its dynamics profile.

Applying s-GNM to homology models

In a frontier study of Stiffler *et al.*,[77] 85 PDZ domain sequences (out of 157 mouse PDZs) were found to bind to at least one peptide. We carefully look at the sequences of the 85 PDZ domains and choose sequences that can bind either Class I or Class II peptides. The protein sequences are obtained from the SMART (Simple Modular Architecture Research Tool) database.[96] They are submitted to "The Structure Prediction Meta Server"[97,98] with the option of an automated homology model program ESyPref3D. ESyPref3D is based on a strategy using neural networks to evaluate sequence alignments[99] and uses the program MODELLER[100,101] to build the final structural model. Homology models are constructed with MODELLER with a minimal sequence similarity of 50% to the target and then s-GNM is applied to all structures.

Applying linear discriminant analysis (LDA) for classification

We apply linear discrimant analysis (LDA) to classify the PDZs into three groups: Class I, Class II, and both Class I and Class II (i.e., dual specificity) by training with PDZs of known classes. Linear Discriminant Analysis (LDA) has been utilized as an effective method of choice when the data set is already clustered. LDA maximizes the ratio of between-class distance and minimizing the within-class distance in any particular data set. Then we predict the classes of mutated structures, which represent an independent test set. The classify function in MATLAB is used in the analysis.

RESULTS AND DISCUSSION

Dominant modes of motions obtained from unbound conformation derive bound conformations

We applied four different methods to a set that includes PDZs having Class I, Class II, and both Class I and Class II (dual) type binding specificities. The results of (i) full atomic NMA with empirical force fields, (ii) Anisotropic Network Model (ANM), (iii) Gaussian Network Model (GNM), and (iv) specific-Gaussian Network Model (s-GNM) are compared and discussed in the following sections.

Here, we have analyzed the collective motions of the unbound structures of PDZs by the four aforementioned methods. We compared the collective (slowest) modes [see Methods and Eq. (5) for the selection of how many modes would contribute to the fluctuation dynamics] with the observed conformational changes by a specific peptide binding.

Figure 2 presents the normalized fluctuation profiles obtained by the slowest modes of s-GNM, ANM, NMA, and GNM along with ligand-induced experimental conformational changes for PSD-95 and hPTP1E. The correlation coefficient for these cases can be found in Table II. It also shows the corresponding ribbon diagrams of these two proteins that are colored with respect to their fluctuation profiles from a spectrum of red (for the highest fluctuation) to yellow, green, cyan, and finally blue (the lowest fluctuation). There are five ribbon diagrams per each protein, which are colored with respect to the fluctuation profiles obtained from GNM, s-GNM, NMA, ANM, and the α-carbon displacement vector upon transition from unbound to bound form. The agreement between each fluctuation profile and the conformational change upon binding shows that s-GNM results display comparable agreement with NMA and ANM results and yet a much better correspondence with experimental changes when compared with GNM results. Table II lists the agreement with the experimental conformational change as correlation coefficients for each PDZ and the

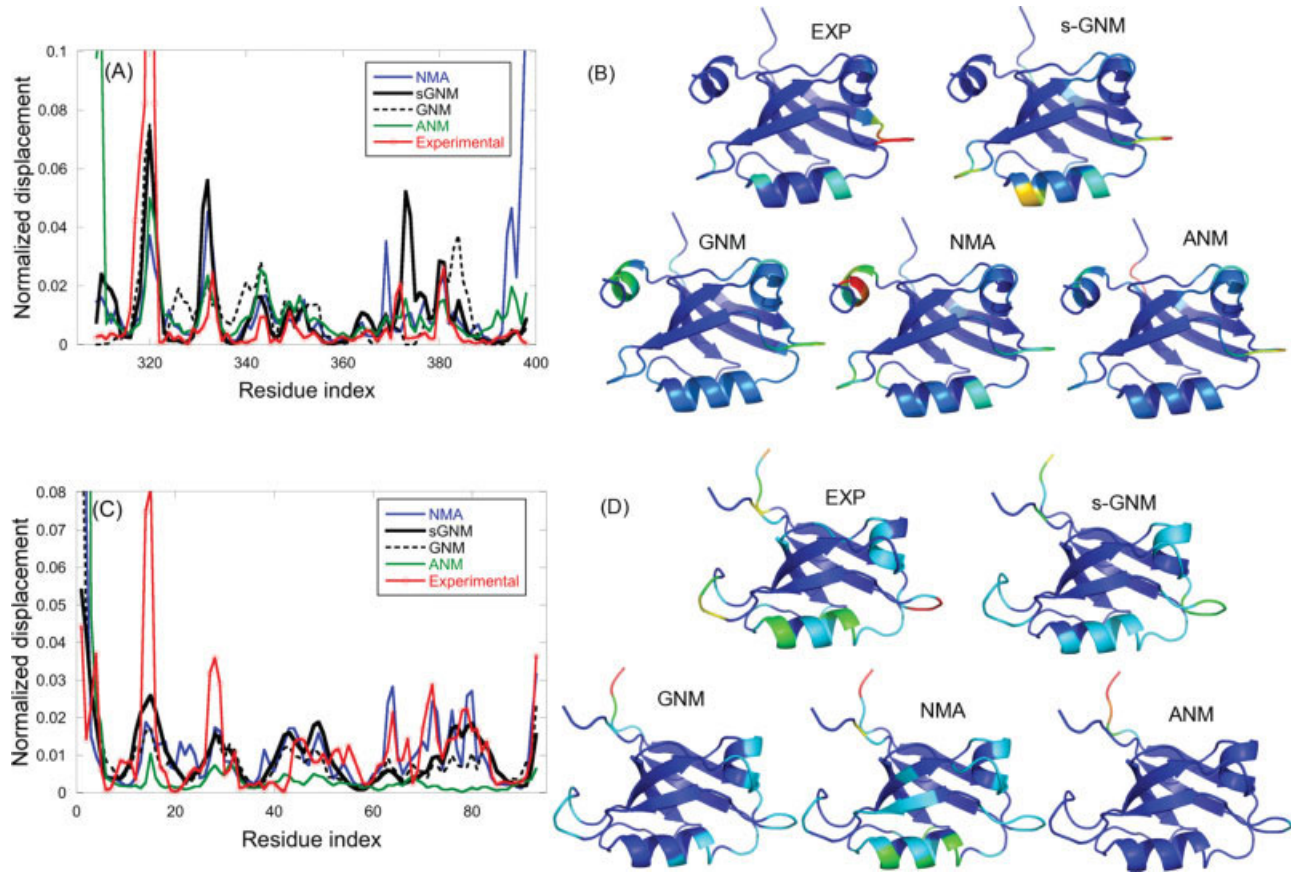

Figure 2

The binding induced fluctuation profiles of (**A**) PSD-95 and (**C**) hPTP1E obtained by NMA, GNM, ANM, s-GNM, and the experimental α-carbon displacement vectors that show average residue displacement upon ligand binding along with the ribbon diagrams colored with respect to fluctuation profiles (**B** and **D**). The dashed curves are obtained from original GNM; blue, black, and green solid curves are the predictions of NMA, s-GNM, and ANM, and solid lines with data points represent the experimental Cα displacements upon binding. Theoretical curves predicted by NMA, ANM, s-GNM, and GNM are normalized to have the same area under the curve with the experimental one. The fluctuation profiles obtained by s-GNM agree better with experimental conformational change (correlation coefficients of 0.68 and 0.57, respectively for protein PSD-95 and hPTP1E can be listed in Table II). There are five ribbon diagrams colored with respect to the fluctuation profiles obtained from GNM, s-GNM, NMA, ANM, and the α-carbon displacement vector upon transition from unbound to bound form. The corresponding ribbon diagrams are colored with respect to their fluctuation profiles from a spectrum of red (for the highest fluctuation) to yellow, green, and finally blue (the lowest fluctuation). The theoretical data obtained by s-GNM and NMA agree better with the experimental data when we compare the agreement between the fluctuation profile from GNM and experimental displacement vector of PSD-95 and hPTP1E.

number of modes used in NMA, ANM, GNM, and s-GNM.

In previous studies, it was shown that the slowest global modes are important to represent the ligand-induced conformational changes in proteins.[56,59,60,102] Likewise, our results show that the weighted average of slowest modes can capture the binding induced conformational changes in this large set of PDZs as also shown earlier for three PDZs.[32] Especially, the conformational changes in peptide binding regions, that are very critical for binding specificity, have significant agreement with the fluctuation profile obtained by s-GNM. Overall, the results have shown that the collective fluctuation profile obtained by s-GNM can capture the binding induced conformational changes with an average correlation of 0.56 as opposed to 0.48, 0.36, and 0.47 for GNM, NMA, and ANM, respectively (see Table II). We have also determined the single eigenmode that gives the best correlation with the experimental conformational changes and this analysis gives a similar qualitative picture as observed in Table II (See Table SI in Supporting Information).

Common fluctuation profiles for Class I, Class II specificity

We have investigated the extent to which PDZ domains could be classified on the basis of their binding-induced dynamics and global motions shared between

different members of PDZs. Thus, we analyze the regions that undergo the highest conformational changes (i.e., which residues exhibit high amplitude fluctuation upon binding) and see if there are patterns for those changes with respect to binding selectivity. First, using the slowest modes, we obtain the fluctuation profiles of all PDZs listed in Table I with s-GNM. Second, we perform multiple sequence alignments using the sequence segments: βA-βB loop, βB strand, βB-βC loop, βC strand, and αB helix of PDZs, which are observed to be in direct contact or in the nearby region of the peptides. The mean-square fluctuations corresponding to these ungapped regions of aligned sequences are then collected and a representative matrix is built by the procedure described in the Methods section: the columns of this matrix represent different PDZ proteins, whereas the rows correspond to the common (sequentially aligned) residues. Therefore, the *ij*th element of the matrix defines the mean-square fluctuations of the *i*th residue in the *j*th protein. We construct two mean square fluctuation matrices for Class I and II proteins separately. Analysis of these two matrices can be used to obtain the common fluctuation profiles that are shared within the members of Class I and Class II type PDZs.

In an attempt to find the common fluctuation profiles responsible for Class I and Class II type binding selectivity, we apply the SVD technique to the fluctuation matrices (see the Methods). An SVD transformation defines a new set of principal axes forming the SVD space, such that the original data is well represented as a linear combination of a few dominant SVD axes. Through singular value decomposition, right principal axes (d^T) provide the new SVD space and left principal axes (s) give the common characteristics of the fluctuation profiles. Figure 3 displays the common profiles of Class I and Class II PDZ domains in the first two principal SVD vectors (A) s_1, and (B) s_2. The first principal axis (s_1) of Class I and Class II PDZs exhibits similar dynamics profiles and it shows the common fluctuation profiles of all PDZ domains due to the similarity of Class I and Class II structures and functions [see Fig. 3(A)]. The two β strands (βB and βC) act as the least mobile regions in this mode. Apparently, this mode is not important for the selectivity of Class I versus Class II types.

Interestingly, we observe differences in the fluctuation profiles of Class I and Class II types in the second principal axis (s_2). The corresponding ribbon diagrams of Class I and Class II type PDZs shown in Figure 3(B) highlight the regions that are critical for binding selectivity. The βA-βB loop and βB-βC loop exhibit high amplitude anti-correlated fluctuations in Class I PDZs. In addition, both loops show similar mobility (same amplitude fluctuations). Thus, the dynamics of both loops are equally critical in Class I type peptide binding. The mutational studies on βA-βB loop and βB-βC loops have also indicated the importance of these two loops in Class I type binding.[29,103–110] In agreement with our findings, it has been shown that the binding affinity of PSD-95 (Class I type PDZ domain) is reduced by introducing specific mutations at the βB-βC loop.[106,108] Another Class I type PDZ domain protein, hPTP1E also supports the importance of βB-βC loop for ligand binding. NMR experiments on the hPTP1E protein have pointed out that Asn27, Thr28, Arg31, and Gly34 in βB-βC loop have significant relaxation contribution from conformational exchange in the unbound form.[29] The importance of the carboxylate binding loop (βA-βB loop) has been discussed for α1-syntrophin[105] and synapse-associated protein-97 (SAP97)[110] that bind to Class I type peptides. The computational studies applied to Class I PDZs showed that βA-βB loop and the αB-helix undergoes the largest deformations upon binding[32] and the region of residue Val30, which is in the βB-βC loop and residues in the region between Val61 and Ala69 become more flexible upon binding.[33] On the other hand, βB-βC loop and αB-helix show high anti-correlated fluctuations in the case of Class II type PDZs. The importance of the αB-helix in class II type binding is corroborated by mutational studies. Changing His71 to Leu in Erbin[111] and Lys 83 to Val and His in PICK1[107] improve the binding to Class II type peptides. Furthermore, analysis displaying the cross-correlation maps of Class I and Class II PDZs agrees with these findings (See Supporting Information Fig. S2 where dark blue represents the anti correlated motions).

To understand the organization and distribution of PDZ selectivities at large, we map the fluctuation profiles of different PDZs onto the first three principal axes of SVD. In contrast to our previous analysis where we construct two matrices for Class I and II proteins separately, here we construct a single matrix including all PDZs. As discussed above, columns of our original data matrix represents different types of PDZs characterized by fluctuation profiles, thus each row represents the mean square fluctuation value of a sequentially aligned common residue. After decomposing the original matrix using SVD, the columns of the new matrix (D^T) are the same PDZ proteins expressed in the new SVD frame, which characterizes the similarities/differences between the proteins, based on their fluctuation profiles. In other words, when the fluctuation profile matrix that includes the fluctuation pattern of various Class I and Class II type PDZ domains are decomposed using SVD, the right principal axes (D^T) of the singular matrix shows the binding-induced fluctuations in the new SVD space for different PDZ domain proteins (See details in the Materials and Methods). The distributions of binding induced fluctuation profiles of unbound PDZ domains (including all PDZs: Class I, Class II, and dual specificity ones) are obtained from the SVD. They are projected into the space spanned by the first three principal axes of the SVD space (See Supporting Information Fig. S3). Con-

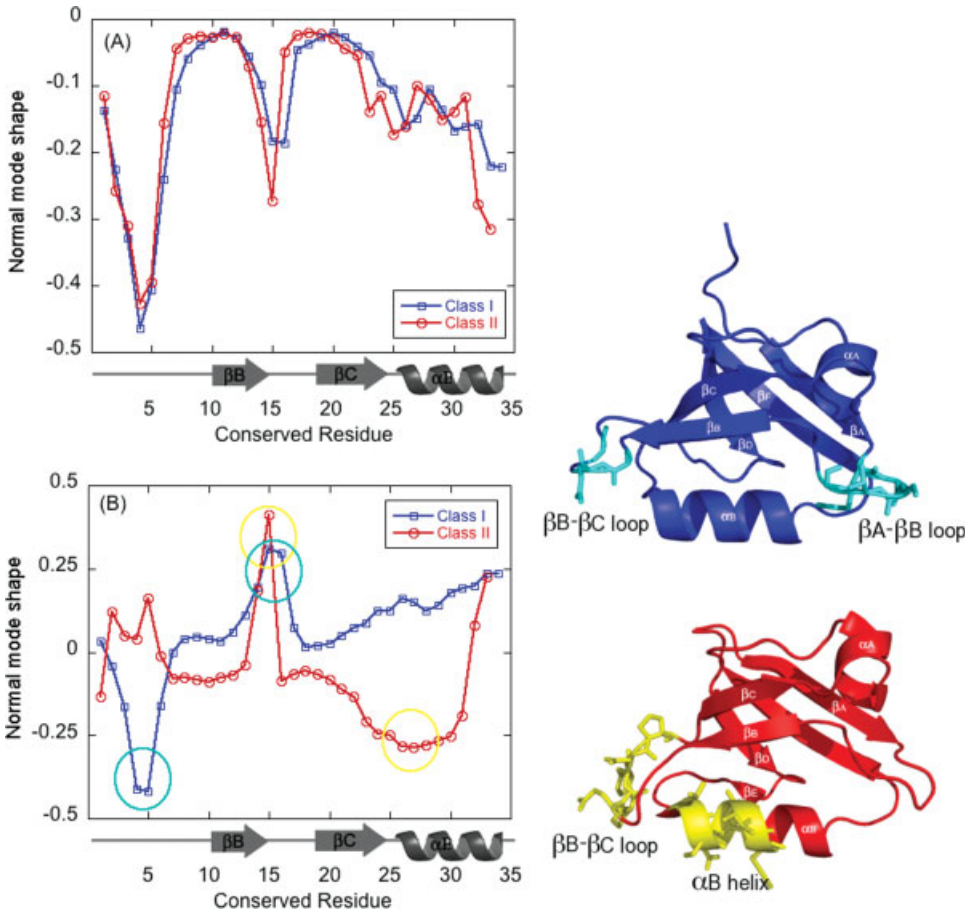

Figure 3

Common motion characteristics of Class I and Class II type PDZ domain interactions obtained by singular value decomposition of the fluctuation profile matrix. (**A**) The shape of the corresponding representative first SVD vector for Class I type and Class II type PDZ domain proteins. Sequence alignment includes only βA-βB loop, βB strand, βB-βC loop, and αB helix that are crucial regions for peptide binding. The first SVD vector shows the common fluctuation of PDZ domains irrespective of their binding specificity. Therefore, Class I type and Class II type PDZ domain proteins show similar fluctuation profile. (**B**) The second SVD vector shows different mobilities between the Class I and Class II type PDZ domain proteins. The dynamics of βA-βB and βB-βC loops are more critical in Class I type binding interaction whereas βB-βC loop and αB-helix becomes critical in Class II type PDZ domain proteins.

sistent with a recent experimental study,[77] we observe that the selectivity space is not discrete for Class I and II proteins but rather continuous.

s-GNM analysis on homology models of PDZ domains

We further investigate how s-GNM discriminates the binding selectivity when it is applied to homology models. To investigate this, we start with PDZs whose structures are not available, yet binding selectivities are studied by Stiffler *et al.*[77] The list of proteins modeled is given in Table III along with their sequence similarities with the known PDZ structures. As case studies, we display the color-coded homology structures of Class I in Figure 4(A) and Class II PDZs in Figure 4(B). Here, red and blue distinguish the highest and lowest flexibility regions, respectively. The figures illustrate that the fluctuation profile of Class I homolog structures exhibit the common binding induced fluctuation profile of Class I type PDZs where the βA-βB and βB-βC loops are the most mobile regions. Likewise, the high amplitude fluctuation profiles of αB-helix and βB-βC loop in Class II homolog proteins confirm that they also possess the common fluctuation characteristics of Class II type.

Overall, these findings indicate that: (i) s-GNM can easily be applied to homology models (i.e., sequences with at least 50% sequence similarity to compute binding induced fluctuation profiles), and (ii) the binding induced fluctuation profiles of homolog models may be used to discriminate their binding selectivity. These results show the potential power of s-GNM. However, one should be careful with cases showing dual specificity and low sequence similarity to a structural template.

Table III
List of Homolog Structures Analyzed. The Number of Interactions with the Peptide was Determined Using Fluorescence Polarization (FP) Obtained from the Stiffler et al.[77] study

Class	Protein	Number of interactions	Class I peptide	Class II peptide	Class III peptide	Neither class I, class II or class III peptides	Modeller best template % sequence similarity	s-GNM mode
I	Magi3-pdz5	25	25				65.8	1
I	β1-syntrophin	15	15				82.5	4
I	γ2-syntrophin	4	4				55	2
I	Lrcc7	3	3				73.5	4
I	Shroom	3	3				58.7	2
I	SemCap3-PDZ1	7	7				96.6	2
I	Scrb1-PDZ2	3	3				92.8	1
I	Chapsyn110-PDZ2	18	18				98.8	1
I	Magi3-PDZ1	18	16			2	76.3	4
I	Cipp-PDZ8	10	8		1	1	97.5	3
I	Par6B	2	2				98.9	1
I	Magi2-PDZ5	2	2				98.9	4
I	Erbin-PDZ1	2	2				96.5	5
I	ZO2-PDZ1[a]	13	1		8	4	93	1
I	Interleukin16	5	4			1	52.4	5
I	Mals2	12	12				98.8	3
II	Tiam-1[b]			1			91	2
II	Pdzk3-PDZ2	2		2			54.2	2
II	Mupp1-PDZ11	2		1		1	95	4
II	Cipp-PDZ9	8		6	2		69.6	2

We classified number of interactions as a Class I, Class II, Class III, and none of typical Class I, II, or III peptides.
[a]In the study of Fujii et al.,[78] ZO2-PDZ1 is defined as class I PDZ domain.
[b]Tiam-1 was not included FP analysis in science paper but it is known that it binds to class II type peptides.[2,79]

s-GNM analysis on mutated proteins of PDZ domains

We further investigated our results by examining the dynamics of the wild type and mutant structures of PICK1. The binding characteristics of wild-type PICK1 and PICK1 carrying various PDZ domain mutations have been reported previously; the wild type of PICK1 can bind both Class I (PKCα) and Class II (GluR2) type peptides.[107,109,112] Staudinger *et al.*[109] mutated lysine 27 (K27) alone or together with aspartic acid 28 (D28) to alanine to test the importance of carboxylate-binding loop of PICK1. Mutation of K27 together with the nearby D28 completely disrupted interaction with both GluR2 and PKCα. Furthermore, the experimental study of Dev *et al.*[112] has shown that mutating lysine 27 to glutamic acid, a point mutation on the βA-βB loop, changes the binding selectivity of PICK1 to exhibit only Class I behavior. Another mutation study that replaces the residue in αB helix (lysine 83 to histidine) by Madsen *et al.*[107] showed that the preference of PICK1 reverts to that of a Class I motif.

The wild type we studied is a crystal structure complex with Class II type peptide. We analyzed the dynamics of this structure with s-GNM. Figure 5(A) shows the ribbon diagram of wild type PICK1 and is colored with respect to the fluctuation profile using a spectrum of blue, green, orange, and red based on the increasing order of mobility. The most mobile region of wild type PICK1 is the βB-βC loop. This observation agrees with the dual binding specificity of PICK1 (i.e., both Class I and Class II type of binding) because the mobility of the βB- βC loop plays an important role in both classes. Interestingly, when we apply s-GNM to the point mutated (K27E and K83H) and energy-minimized structure, we observe the signatures of Class I type of binding dynamics, where both the βA-βB and βB-βC loops become equally mobile [See Fig. 5(B,C) for ribbon diagrams of mutated structures (K27E) and (K83H), respectively]. We also applied linear discriminant analysis as explained in the following section to predict the mutant PICK1 preference. The analysis predicts that the mutant PICK prefers Class I type of binding with a probability of 0.82.

Linear discriminant analysis as a prediction tool

By linear discriminant analysis, we found that PDZs showing Class I, Class II, and dual specificity (i.e., both Class I and Class II) can be predicted with an accuracy of 81% based on the fluctuation profiles. (We applied 10-fold cross validation where 90% of the data was randomly selected as training set and the leftover 10% was used as the test case to predict the binding specificities based on singular values of fluctuation profiles. The average of 10 trials with this approach gives an accuracy of 86%.) This prediction tool enables us to suggest mutations that could change a PDZ domain from Class I to Class II or vice versa. The histidine at the first residue of the αB helix is a conserved residue in Class I PDZ

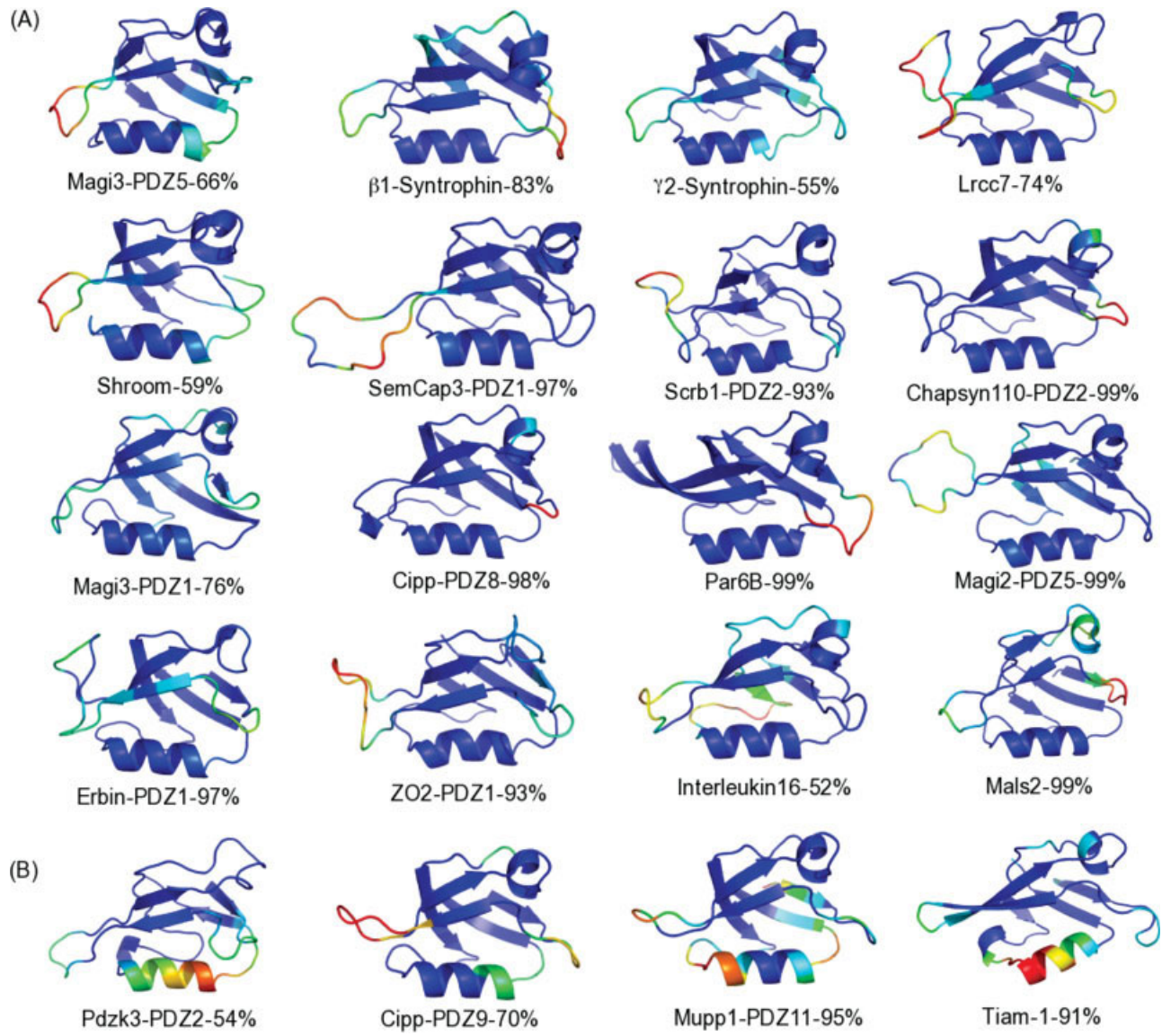

Figure 4

The ribbon diagrams of Class I (**A**) and Class II (**B**) homology models colored with respect to the fluctuation profiles obtained by s-GNM. The most mobile regions are red whereas the least mobile ones are shown in blue within a color spectrum of red, orange, yellow, green, cyan, and blue. The fluctuation profiles of Class I homolog structures exhibit the common binding induced fluctuation profile of Class I type PDZ proteins where βA-βB and βB-βC loops are the most mobile regions. Likewise, the high amplitude fluctuation profiles of αB-helix and βB-βC loop in Class II homolog proteins confirm that they also exhibit the common fluctuation characteristics of Class II type PDZ domain proteins.

domains and is thought to bind threonine or serine in the binding motif.[113] Agreed with this observation, this analysis has shown that mutation of the first residue in αB helix (Asp 25 to His) of Syntenin that has affinity for both classes could change its specificity to Class I type binding. We also tested various mutations in βA-βB loop (from residues 203 through 209) and our predictor tool did not change the specificity towards either Class I or Class II and it stays as dual specificity. Agreed with our prediction, Grootjans *et al.*[104] has also shown experimentally that K203A mutation in βA-βB loop decreases the binding of Class II peptide preference but does not terminate the Class I type preference. They also studied the mutation of the first residue in αB helix and observed that mutation abolishes the interaction between the protein and Class II type of peptide as we also predicted in our analysis. Furthermore, we investigated the binding selectivites of Tiam-1 (91% sequence similarity as a homolog structure) that binds to class II peptide when the first residue (Ser 64) in αB helix is mutated to His, and our prediction tool indicates that Mutant Tiam-1 might bind to both Class I and Class II type of

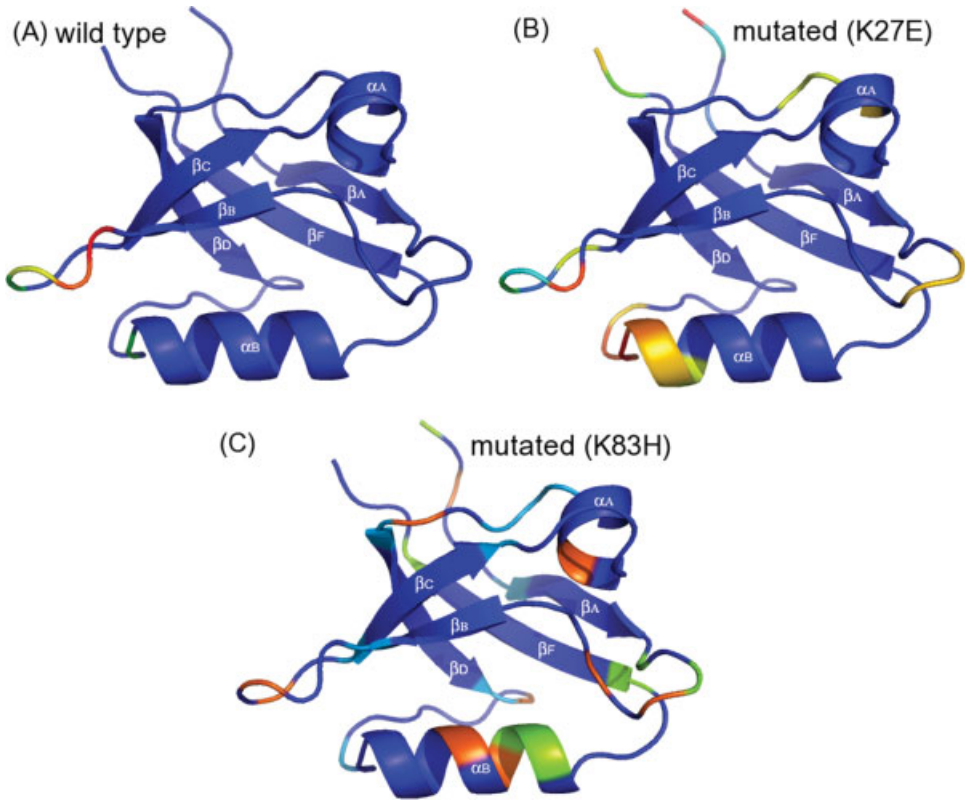

Figure 5

The comparison of the ribbon diagrams corresponding to fluctuation profiles of (**A**) wild type PICK1 (PDB: 2gzv), (**B**) mutated structure (K27E) of PICK1, and (**C**) mutated structure (K83H) of PICK1 colored with respect to their fluctuation profiles: The spectrum of dark blue, green, orange, and red in ribbon diagrams represents the increasing order of mobility. The most mobile region of wild type PICK1 is βB-βC loop, whereas both βA-βB loop and βB-βC loop become equally mobile in mutated structure of PICK1 which is a signature of Class I type binding dynamics.

peptides. In summary, linear discriminant analysis gives us a new tool to make prediction about PDZs' binding specificities upon mutation.

CONCLUSIONS

The potential application range of our new elastic network model, s-GNM, with a set of PDZs has been demonstrated in this work. We are able to determine fluctuation profiles responsible for binding to different type of peptides with the new approach. Further, the general applicability of this method and the validity of optimized parameter sets are tested with a comprehensive data set including homology models of PDZs.

Using singular value decomposition method, we clustered the binding induced fluctuation profiles obtained from s-GNM analysis. Our results show that: (i) the dynamics of PDZs play an important role in their binding selectivity as also shown by other studies,[30–33,37,39] (ii) using s-GNM, we identify the structural regions critical for binding of Class I and Class II peptides, (iii) s-GNM has potential to capture the changes in binding selectivity due to mutations as in the case of PICK1, where the comparison of the binding induced dynamics shows the shift from Class II to Class I type common fluctuation profile upon mutations, in agreement with experimental observation, (iv) s-GNM can be easily and successfully applied to homology models of PDZs. These are very important steps to develop a quantitative tool to alter the binding selectivity of PDZ domains upon mutations and can be used to knock out PDZ-domain interactions for therapeutic interests.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Kingshuk Ghosh, Dr. M. Scott Shell, Dr. Bosco Ho and Art Cerda for their valuable comments and ASU Fulton School of Engineering HPC for the computer time. O.K. is supported by TUBA-Distinguished Young Investigator Award.

REFERENCES

1. Fan JS, Zhang M. Signaling complex organization by PDZ domain proteins. Neurosignals 2002;11:315–321.
2. Nourry C, Grant SG, Borg JP. PDZ domain proteins: plug and play! Sci STKE 2003;2003:RE7.

3. Hung AY, Sheng M. PDZ domains: structural modules for protein complex assembly. J Biol Chem 2002;277:5699–5702.
4. van Ham M, Hendriks W. PDZ domains-glue and guide. Mol Biol Rep 2003;30:69–82.
5. Appleton BA, Zhang Y, Wu P, Yin JP, Hunziker W, Skelton NJ, Sidhu SS, Wiesmann C. Comparative structural analysis of the Erbin PDZ domain and the first PDZ domain of ZO-1. Insights into determinants of PDZ domain specificity. J Biol Chem 2006;281:22312–22320.
6. Birrane G, Chung J, Ladias JA. Novel mode of ligand recognition by the Erbin PDZ domain. J Biol Chem 2003;278:1399–1402.
7. Chen Q, Niu X, Xu Y, Wu J, Shi Y. Solution structure and backbone dynamics of the AF-6 PDZ domain/Bcr peptide complex. Protein Sci 2007;16:1053–1062.
8. Daniels DL, Cohen AR, Anderson JM, Brunger AT. Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition. Nat Struct Biol 1998;5:317–325.
9. Dev KK. PDZ domain protein-protein interactions: a case study with PICK1. Curr Top Med Chem 2007;7:3–20.
10. Elkins JM, Papagrigoriou E, Berridge G, Yang X, Phillips C, Gileadi C, Savitsky P, Doyle DA. Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions. Protein Sci 2007;16:683–694.
11. Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. J Cell Sci 2001;114(Part 18):3219–3231.
12. Hillier BJ, Christopherson KS, Prehoda KE, Bredt DS, Lim WA. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science 1999;284:812–815.
13. Im YJ, Lee JH, Park SH, Park SJ, Rho SH, Kang GB, Kim E, Eom SH. Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. J Biol Chem 2003;278:48099–48104.
14. Im YJ, Park SH, Rho SH, Lee JH, Kang GB, Sheng M, Kim E, Eom SH. Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization. J Biol Chem 2003;278:8501–8507.
15. Kang BS, Cooper DR, Devedjiev Y, Derewenda U, Derewenda ZS. Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. Structure 2003;11:845–853.
16. Karthikeyan S, Leung T, Birrane G, Webster G, Ladias JA. Crystal structure of the PDZ1 domain of human Na(+)/H(+) exchanger regulatory factor provides insights into the mechanism of carboxyl-terminal leucine recognition by class I PDZ domains. J Mol Biol 2001;308:963–973.
17. Karthikeyan S, Leung T, Ladias JA. Structural determinants of the Na+/H+ exchanger regulatory factor interaction with the beta 2 adrenergic and platelet-derived growth factor receptors. J Biol Chem 2002;277:18973–18978.
18. Laura RP, Witt AS, Held HA, Gerstner R, Deshayes K, Koehler MF, Kosik KS, Sidhu SS, Lasky LA. The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of delta-catenin and ARVCF. J Biol Chem 2002;277:12906–12914.
19. Sallee NA, Yeh BJ, Lim WA. Engineering modular protein interaction switches by sequence overlap. J Am Chem Soc 2007;129:4606–4611.
20. Tochio H, Hung F, Li M, Bredt DS, Zhang M. Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95. J Mol Biol 2000;295:225–237.
21. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–299.
22. Zimmermann P. The prevalence and significance of PDZ domain-phophoinositide interactions. Biochim Biophys Acta 2006;1761:947–956.
23. Zimmermann P, Meerschaert K, Reekmans G, Leenaerts I, Small JV, Vandekerckhove J, David G, Gettemans J. PIP2-PDZ domain binding controls the association of syntenin with the plasma membrane. Mol Cell 2002;9:1215–1225.
24. DeLano WL. The PyMOL Molecular Graphics System. Palo Alto, CA: DeLano Scientific; 2002 (http://www.pymol.org).
25. Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, Crompton A, Chan AC, Anderson JM, Cantley LC. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 1997;275:73–77.
26. Bezprozvanny I, Maximov A. Classification of PDZ domains. FEBS Lett 2001;509:457–462.
27. Kurakin A, Swistowski A, Wu SC, Bredesen DE. The PDZ domain as a complex adaptive system. Plos Comput Biol 2007;2:e953.
28. Vaccaro P, Dente L. PDZ domains: troubles in classification. FEBS Lett 2002;512:345–346.
29. Fuentes EJ, Der CJ, Lee AL. Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. J Mol Biol 2004;335:1105–1115.
30. Basdevant N, Weinstein H, Ceruso M. Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. J Am Chem Soc 2006;128:12766–12777.
31. Cecconi F, Rios PDL, Piazza F. Diffusion-limited unbinding of small peptides from PDZ domains. J Phys Chem B 2007;111:11057–11063.
32. De Los Rios P, Cecconi F, Pretre A, Dietler G, Michielin O, Piazza F, Juanico B. Functional dynamics of PDZ binding domains: a normal-mode analysis. Biophys J 2005;89:14–21.
33. Dhulesia A, Gsponer J, Vendruscolo M. Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. J Am Chem Soc 2008;130:8931–8939.
34. Giallourakis C, Cao Z, Green T, Wachtel H, Xie X, Lopez-Illasaca M, Daly M, Rioux J, Xavier R. A molecular-properties-based approach to understanding PDZ domain proteins and PDZ ligands. Genome Res 2006;16:1056–1072.
35. Kempkens O, Medina E, Fernandez-Ballester G, Ozuyaman S, Le Bivic A, Serrano L, Knust E. Computer modelling in combination with in vitro studies reveals similar binding affinities of Drosophila Crumbs for the PDZ domains of Stardust and DmPar-6. Eur J Cell Biol 2006;85:753–767.
36. Niv MY, 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;127:14072–14079.
37. Ota N, Agard DA. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. J Mol Biol 2005;351:345–354.
38. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C. Computer-aided design of a PDZ domain to recognize new target sequences. Nat Struct Biol 2002;9:621–627.
39. Sharp K, Skinner JJ. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. Proteins 2006;65:347–361.
40. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. Nature 2007;449:1078–1082.
41. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D. Intrinsic dynamics of an enzyme underlies catalysis. Nature 2005;438:117–121.
42. Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science 2001;291:2429–2433.
43. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
44. Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. Phys Rev Lett 1998;80:2733–2736.

45. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
46. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry 2002;41:491–501.
47. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78: 2093–2106.
48. Kondrashov DA, Cui Q, Phillips GN, Jr. Optimization and evaluation of a coarse-grained model of protein motion using x-ray crystal data. Biophys J 2006;91:2760–2767.
49. Rader AJ, Vlad DH, Bahar I. Maturation dynamics of bacteriophage HK97 capsid. Structure 2005;13:413–421.
50. Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL. The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. J Chem Theory Comput 2006;2:696–704.
51. Sen TZ, Jernigan RL. Optimizing cutoff distances and spring constants for the Gaussian Network Model of ATP-binding proteins. In: Cui IBaQ, editor. Normal mode analysis: theory and applications to biological and chemical systems. Boca Raton: Chapman and Hall/CRC; 2006. pp 171–186.
52. Tama F, Brooks CL. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. J Mol Biol 2002;318: 733–747.
53. Tama F, Brooks CL. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. J Mol Biol 2005;345:299–314.
54. Delarue M, Sanejouand YH. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 2002;320:1011–1024.
55. Ikeguchi M, Ueno J, Sato M, Kidera A. Protein structural change upon ligand binding: linear response theory. Phys Rev Lett 2005; 94:078102.
56. Keskin O. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations. BMC Struct Biol 2007; 7:31.
57. Kundu S, Jernigan RL. Molecular mechanism of domain swapping in proteins: an analysis of slower motions. Biophys J 2004;86: 3846–3854.
58. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6.
59. Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proc Natl Acad Sci USA 2005;102:18908–18913.
60. Zheng W, Brooks BR. Normal-modes-based prediction of protein conformational changes guided by distance constraints. Biophys J 2005;88:3109–3117.
61. Demirel MC, Atilgan AR, Jernigan RL, Erman B, Bahar I. Identification of kinetically hot residues in proteins. Protein Sci 1998; 7:2522–2532.
62. Haliloglu T, Keskin O, Ma B, Nussinov R. How similar are protein folding and protein binding nuclei? Examination of vibrational motions of energy hot spots and conserved residues. Biophys J 2005;88:1552–1559.
63. Haliloglu T, Seyrek E, Erman B. Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model. Phys Rev Lett 2008;100:228102.
64. Rader AJ, Anderson G, Isin B, Khorana HG, Bahar I, Klein-Seetharaman J. Identification of core amino acids stabilizing rhodopsin. Proc Natl Acad Sci USA 2004;101:7246–7251.
65. Lu Q, Sun EE, Klein RS, Flanagan JG. Ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G protein-coupled chemoattraction. Cell 2001;105:69–79.
66. Long JF, Feng W, Wang R, Chan LN, Ip FC, Xia J, Ip NY, Zhang M. Autoinhibition of X11/Mint scaffold proteins revealed by the closed conformation of the PDZ tandem. Nat Struct Mol Biol 2005;12:722–728.
67. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
68. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm—a Program for macromolecular energy. Minimization, and dynamics calculations J Comput Chem 1983;4:187–217.
69. Masunov A, Lazaridis T. Potentials of mean force between ionizable amino acid side chains in water. J Am Chem Soc 2003; 125:1722–1730.
70. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 1983;80:6571–6575.
71. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
72. Rader AJ, Chennubhotla C, Yang L-W, Bahar I. The Gaussian Network Model: theory and applications. In: Cui Q, Bahar I, editors. Normal mode analysis theory and applications to biological and chemical systems, Chapman and Hall/CRC Mathematical and Computational Biology Series, Taylor and Francis Group: Boca Raton; 2006. pp 41–64.
73. Yang LW, Eyal E, Chennubhotla C, Jee J, Gronenborn AM, Bahar I. Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure 2007;15:741–749.
74. Tsodikov OV, Record MT, Jr, Sergeev YV. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. J Comput Chem 2002;23:600–609.
75. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 1996;256:623–644.
76. Thomas PD, Dill KA. An iterative method for extracting energy-like quantities from protein structures. Proc Natl Acad Sci USA 1996;93:11628–11633.
77. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. Science 2007;317:364–369.
78. Fujii N, Shelat A, Hall RA, Guy RK. Design of a selective chemical probe for class I PDZ domains. Bioorg Med Chem Lett 2007; 17:546–548.
79. Grootjans JJ, Zimmermann P, Reekmans G, Smets A, Degeest G, Durr J, David G. Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. Proc Natl Acad Sci USA 1997;94:13683–13688.
80. Hinsen K. Structural flexibility in proteins: impact of the crystal environment. Bioinformatics 2008;24:521–528.
81. Riccardi D, Cui Q, Phillips GN, Jr. Application of elastic network models to proteins in the crystalline state. Biophys J 2009;96:464–475.
82. Soheilifard R, Makarov DE, Rodin GJ. Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors. Phys Biol 2008;5:26008.
83. Song G, Jernigan RL. vGNM: a better model for understanding the dynamics of proteins in crystals. J Mol Biol 2007;369:880–893.
84. Isin B, Schulten K, Tajkhorshid E, Bahar I. Mechanism of signal propagation upon retinal isomerization: insights from molecular dynamics simulations of rhodopsin restrained by normal modes. Biophys J 2008;95:789–803.
85. Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. Exploring the common dynamics of homologous proteins. Application to the globin family. Biophys J 2005;89:3–13.
86. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An analysis of core deformations in protein superfamilies. Biophys J 2005;88:1291–1299.

87. Chen SC, Bahar I. Mining frequent patterns in protein structures: a study of protease families. Bioinformatics 2004;20(Suppl 1):i77–i85.
88. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 2005;13:893–904.
89. ClustalW, WWW Service at the European Bioinformatics (http://www.ebi.ac.uk/Tools/clustalw2).
90. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
91. Bahar I, Atilgan AR, Jernigan RL, Erman B. Understanding the recognition of protein structural classes by amino acid composition. Proteins 1997;29:172–185.
92. Keskin O, Bahar I, Jernigan RL, Beutler JA, Shoemaker RH, Sausville EA, Covell DG. Characterization of anticancer agents by their growth inhibitory activity and relationships to mechanism of action and structure. Anticancer Drug Des 2000;15:79–98.
93. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
94. Case DA, Cheatham TE, III, Darden T, Gohlke H, Luo R, Merz KM, Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The amber biomolecular simulation programs. J Comput Chem 2005;26:1668–1688.
95. Kollman PA, Dixon R, Cornell W, Vox T, Chipot C, Pohorille A. The development/application of a "Minimalist" organic/biochemical molecular mechanical force field using a combination of ab initio calculations and experimental data. In: van Gunsteren WF, Weiner PK, Wilkinson AJ, editors. Computer simulation of biomolecular systems theoretical and experimental applications. Boston: Kluwer; 1997. pp 83–96.
96. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 1998;95:5857–5864.
97. Bujnicki JM, Elofsson A, Fischer D, Ryclewski L. Structure prediction meta server. Bioinformatics 2001;17:750–751.
98. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
99. Lambert C, Leonard N, De Bolle X, Depiereux E. ESyPred3D: prediction of proteins 3D structures. Bioinformatics 2002;18:1250–1256.
100. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
101. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2507–2524.
102. Yang L, Song G, Jernigan RL. How well can we understand large-scale protein motions using normal modes of elastic network models? Biophys J 2007;93:920–929.
103. Gianni S, Walma T, Arcovito A, Calosci N, Bellelli A, Engstrom A, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. Structure 2006;14:1801–1809.
104. Grootjans JJ, Reekmans G, Ceulemans H, David G. Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin. J Biol Chem 2000;275:19933–19941.
105. Harris BZ, Lau FW, Fujii N, Guy RK, Lim WA. Role of electrostatic interactions in PDZ domain ligand recognition. Biochemistry 2003;42:2797–2805.
106. Imamura F, Maeda S, Doi T, Fujiyoshi Y. Ligand binding of the second PDZ domain regulates clustering of PSD-95 with the Kv1.4 potassium channel. J Biol Chem 2002;277:3640–3646.
107. Madsen KL, Beuming T, Niv MY, Chang CW, Dev KK, Weinstein H, Gether U. Molecular determinants for the complex binding specificity of the PDZ domain in PICK1. J Biol Chem 2005;280:20539–20548.
108. Nonaka M, Doi T, Fujiyoshi Y, Takemoto-Kimura S, Bito H. Essential contribution of the ligand-binding beta B/beta C loop of PDZ1 and PDZ2 in the regulation of postsynaptic clustering, scaffolding, and localization of postsynaptic density-95. J Neurosci 2006;26:763–774.
109. Staudinger J, Lu J, Olson EN. Specific interaction of the PDZ domain protein PICK1 with the COOH terminus of protein kinase C-alpha. J Biol Chem 1997;272:32019–32024.
110. Wang L, Piserchio A, Mierke DF. Structural characterization of the intermolecular interactions of synapse-associated protein-97 with the NR2B subunit of *N*-methyl-D-aspartate receptors. J Biol Chem 2005;280:26992–26996.
111. Jaulin-Bastard F, Saito H, Le Bivic A, Ollendorff V, Marchetto S, Birnbaum D, Borg JP. The ERBB2/HER2 receptor differentially interacts with ERBIN and PICK1 PSD-95/DLG/ZO-1 domain proteins. J Biol Chem 2001;276:15256–15263.
112. Dev KK, Nakanishi S, Henley JM. The PDZ domain of PICK1 differentially accepts protein kinase C-alpha and GluR2 as interacting ligands. J Biol Chem 2004;279:41393–41397.
113. Dev KK. Making protein interactions druggable: targeting PDZ domains. Nat Rev Drug Discov 2004;3:1047–1056.