PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Signaling pathways of PDZ2 domain: A molecular dynamics interaction correlation analysis

Yifei Kong[1] and Martin Karplus[1,2]*

[1] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

[2] Laboratoire de Chimie Biophysique, Institut de Science et d' Ingenierie Supramoleculaires, Universite Louis Pasteur, 8 rue Gaspard Monge, F-67000 Strasbourg, France

ABSTRACT

PDZ domains are found in many signaling proteins. One of their functions is to provide scaffolds for forming membrane-associated protein complexes by binding to the carboxyl termini of their partners. PDZ domains are thought also to play a signal transduction role by propagating the information that binding has occurred to remote sites. In this study, a molecular dynamics (MD) simulation-based approach, referred to as an interaction correlation analysis, is applied to the PDZ2 domain to identify the possible signal transduction pathways. A residue correlation matrix is constructed from the interaction energy correlations between all residue pairs obtained from the MD simulations. Two continuous interaction pathways, starting at the ligand binding pocket, are identified by a hierarchical clustering analysis of the residue correlation matrix. One pathway is mainly localized at the N-terminal side of helix α1 and the adjacent C-terminus of loop β1-β2. The other pathway is perpendicular to the central β-sheet and extends toward the side of PDZ2 domain opposite to the ligand binding pocket. The results complement previous studies based on multiple sequence analysis, NMR, and MD simulations. Importantly, they reveal the energetic origin of the long-range coupling. The PDZ2 results, as well as the earlier rhodopsin analysis, show that the interaction correlation analysis is a robust approach for determining pathways of intramolecular signal transduction.

Proteins 2009; 74:145–154.

Key words: PDZ2 domain; molecular dynamics simulation; signal transduction; allosteric effects; interaction correlation.

INTRODUCTION

PDZ domains constitute a class of modules that are important for mediating the interactions between partners along signal transduction pathways.[1–3] They are one of the most abundant interaction domains and have been identified in several hundred eukaryote proteins.[1] They are often associated with receptors (e.g., NMDA and AMPA/kainite) as well as membrane channels (e.g., Shaker K^+ and TRP Ca^{2+}) and are involved in clustering proteins into functional complexes at the plasma membrane.[1–4] Usually, the PDZ-containing protein has several PDZ domains (up to 10 domains[5]), which can concurrently interact with different protein, and thereby form functional protein complexes. For example, InaD contains five PDZ domains that contribute to the formation of the cluster of eight proteins involved in the *Drosophila* phototransduction cascade,[1] whereas PSD-95 has three PDZ domains that interact with glutamate receptors at synapses (e.g., in terms of numbering, PDZ3, which has been studied extensively, is the third PDZ domain of PSD-95).[1,6,7] Moreover, PDZ domains can synergistically bind to inositol phospholipids with pleckstrin homology domains.[8] Clearly, an understanding of the dynamic properties of PDZ domains and their possible role in allosteric effects is of interest.

Structures of several PDZ domains are available through either X-ray crystallography (PDZ1[9,10]), NMR studies (PDZ4[11]), or both (PDZ2[12–16]). PDZ domains adopt an overall globular structure containing two α-helices and six β-strands [Fig. 1(a)]. Most PDZ-mediated interactions occur between the C-terminal peptide of the target protein and the binding pocket between the α2 helix and the β2 strand.[1,6] Figure 1(b) shows an enlarged view of the binding site of the PDZ2 domain with the ligand, which fills a hydrophobic pocket between strand β2 and helix α2 and is capped by the loop β1-β2. The last three residues (P_0, P_{-1}, and P_{-2}) of the peptide form an antiparallel β-sheet with the β2 strand of PDZ2 [Fig. 1(b)] (P_0 denotes the C terminal residue of the bound

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

Yifei Kong's current address is Monsanto Company, 800 N. Lindbergh Blvd. St. Louis, Missouri 63167

Grant sponsors: National Institute of Health, CHARMM Development Project.

*Correspondence to: Martin Karplus, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138. E-mail: marci@tammy.harvard.edu.

Received 25 January 2008; Revised 2 April 2008; Accepted 28 April 2008
Published online 10 July 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22139

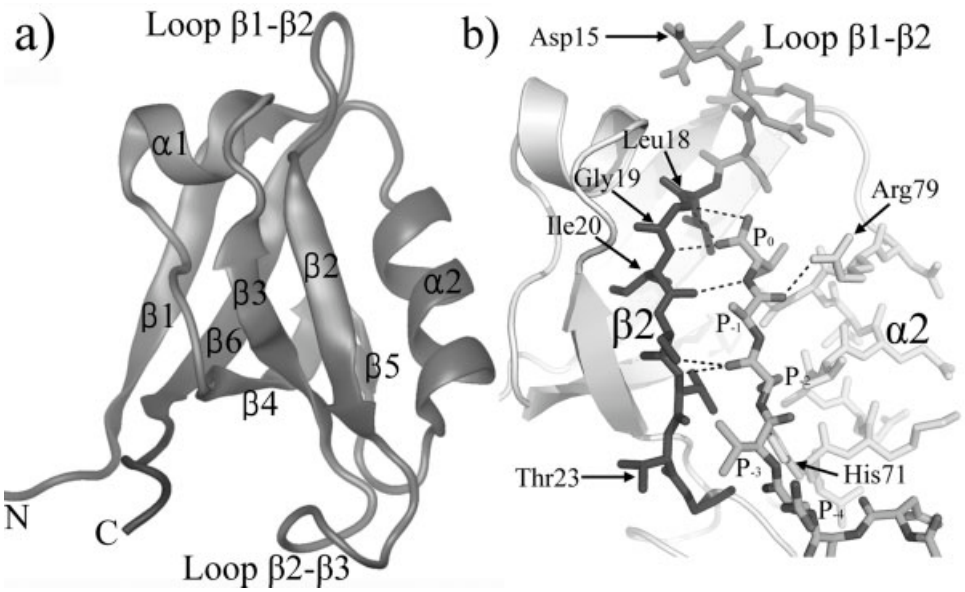

Figure 1

The structure of PDZ2 domain. (**a**) The overall structure of PDZ2 domain (PDB ID 3PDZ). Image was made with VMD.[17] (**b**) An enlarged view of the ligand peptide and the binding pocket (PDB ID 1D5G). Helix α2, strand β2, and loop β1-β2 are colored white, black and gray, respectively. Hydrogen bonds between ligand peptide and PDZ2 domain are shown as dash lines. Images were made by Pymol.[18]

peptide and P_{-n} denotes the *n*th amino acid up stream from it). In particular, loop β1-β2 contains a conservative glycine-rich motif (Leu18, Gly19, and Ile20), whose main chain forms hydrogen bonds with the carboxylate of the P_0 residue. The side chains of residues P_0 and P_{-2} are buried inside the binding groove.[15] Variations in the size and geometry of the hydrophobic binding pockets results in different binding preferences among PDZ domains.[1] However, the hydrophobic recognition is usually not very specific and helps to explain why some PDZ domains can bind peptides with different sequences.[1]

The overall structure of the PDZ2 domains studied here does not undergo large changes on binding a peptide ligand; the RMSD between backbone atoms of ligand-bound and ligand-free structures is about 2.0 Å[14,15] (The four flexible N- and C-terminal amino acids are not included in this value). The groove between strand β2 and helix α2 opens slightly to accommodate the binding of the peptide. Specifically, the amplitude of the opening gradually increases away from the P^0 binding site, e.g., the Cα distances between residue pairs 19/79 and 23/71 increase about 1 Å and 4 Å, respectively. Also the β1-β2 loop (residues 13–17) moves away from the binding pocket upon peptide binding; e.g., the displacement of residue 15 is about 5 Å.

In addition to the fact that PDZ domains recognize and bind the C terminus of target proteins, it has been suggested that they propagate a signal indicative of binding (i.e., that the target protein is bound) toward other parts of the molecular complex not directly connected to the binding pocket.[19–21] A specific example is that the binding of the guanylate kinase domain with PDZ containing microtubule-associated protein 1A has been shown to be enhanced by a variety of PDZ ligands.[22] This type of "allosteric" effect is of intrinsic interest, but it is also of more general biological importance. An understanding of the mechanism of intramolecular signal transduction could be useful for its modification in biologically significant processes. Several analyses concerned with the dynamic response of different PDZ domains have been published; they include the multiple sequence alignment-based statistical analysis,[20] which served to stimulate this study, NMR measurements of the effect of ligands,[12,19,23] a normal mode analysis,[24] an anisotropic thermal diffusion calculation at a low temperature,[25] and a pump-probe molecular dynamics (MD) study.[26] We comment on these alternative approaches in the Discussion section.

The objective of this study is to find communication pathways between the binding region and distant portions of the PDZ molecule that are likely to be involved in interaction with the other part of the complex and binding of other partners. MD simulations of the apo structure PDZ2 domain (96 amino acids) of human phosphatase hPTP1E[14] and of its ligand-bound form (with a 15-amino acid peptide)[15] were performed with the available NMR structures. This PDZ2 domain was chosen for study, not only because both the ligand-free

(PDB ID 3PDZ)[14] and ligand-bound (PDB ID 1D5G)[15] structures were available when this work was initiated, but also because their existed data for the chemical shift changes upon ligand binding for a close homolog.[12] The latter, a mouse PDZ2 domain (PDB ID 1V56) has 95% sequence identity with the human PDZ2 domain that we studied; from the structure, released after much of our work was completed, the RMSD between the two is 2.7 Å. The MD simulations were used to calculate the correlations between residue–residue interaction pairs, based on a recently developed method that has been applied to determine the signaling path in rhodopsin.[27] This analysis revealed two communication pathways (one novel, the other suggested in Ref. 19), which originate from the ligand binding pocket and extend to the opposite surface of PDZ2 domain [i.e., the N-terminal part of helix α1 and the N-terminal of strand β1, see Fig. 1(a)]. Since these pathways are found in the equilibrium simulations of the PDZ2 domain without ligand, the results indicate that pathways are "imprinted" in the structure by evolution. Further, since the coupling is altered by ligand binding, it could serve to signal this fact to other binding partners; i.e., after the ligand binds, the signal could be efficiently propagated toward the target regions through the already existing coupling network.

RESULTS

The interaction correlation method[27] was used to identify the energetically coupled regions and pathways of the PDZ2 domain.[14,15] The method is based on a calculation of the energetic correlations between all interacting residue–residue pairs in an ensemble of protein conformers generated by nanosecond MD simulations. The PDZ2 domain with and without a ligand was studied. A detailed description of the methods used and the MD simulations performed is given in the Methods section.

Dynamically coupled regions identified by interaction correlation

To identify conformational couplings within the PDZ2 domain,[14,15] the equal-time correlation, $C_{i,jlk,l}$, of the nonbonded interaction energies between all residue pairs were calculated (see Methods section). To simplify the analysis and remove the thermal noise, two residues are included only if the average interaction energy is larger than 1 kcal/mol in magnitude. Also, neighboring residues in the sequence that are covalently bonded were not included. Based on 10 trajectories with a total simulation length of more than 50 ns (see Methods section), a 261 × 261 symmetric matrix was generated corresponding to 261 nonredundant residue–residue interactions with absolute values of the interaction energies larger than 1 kcal/mol. Most interaction pairs (there are total of 68,260) have no or very weak correlation, as the average correlation value is 0.087 with a standard deviation of 0.084. To avoid the background noise (see Methods section), we constructed a condensed version of the interaction correlation matrix with a correlation value cutoff, C_{cutoff}, in which all correlations less than C_{cutoff} and larger than $-C_{\text{cutoff}}$ were set to zero, and columns/rows containing all zero elements were not included. In the study of PDZ2 domain, the C_{cutoff} was set to 0.4 (i.e., only values less that −0.4 and greater than +0.4 were included); the resulting interaction energy correlation matrix dimensions are 136 × 136 [Fig. 2(a)]. The index of the residue interaction is shown in Supplemental Table 1. Since each row or column in this matrix corresponds to an individual residue–residue interaction, it is difficult to map the results on the three-dimensional structure and compare them with residue-based experimental data, such as those from mutagenesis and NMR studies. Consequently, the condensed interaction correlation matrix was constructed [see Eq. (4)] to obtain the residue correlation matrix [Fig. 2(b)]; in this matrix, each column and row represents a residue of the protein.

Ligand-free state

The residue correlation matrix of the PDZ2 domain in ligand-free state shows the magnitude of the average coupling between residue pairs obtained from the MD simulations, which were performed at thermal equilibrium. As shown in the residue–residue absolute interaction energy matrix [Fig. 2(c)] and the displacement correlation matrix [Fig. 2(d)], neighboring residues in the PDZ2 domain secondary structure have strong average interaction energies and their displacement correlations are also relatively high. Overall, residues with strong interactions [Fig. 2(b)] also show strong displacement correlations [Fig. 2(c,d)], because the magnitude of interaction is dependent on the distance. Of primary interest, therefore, are the residue couplings present in the residue correlation matrix [Fig. 2(b)] that are absent in the average interaction energy matrix and in the displacement correlation matrix. For example, residue 37 has strong coupling with residues 82 and 83 in Figure 2(b) (indicated by white arrow), whereas there is no corresponding signal in Figure 2(c,d). These excess correlations correspond to the coupling between different regions within the PDZ2 domain. To relate the results to the possibility of signaling that a ligand is bound, we calculated the cumulative interaction correlation (binding pocket correlation factor, see Methods section) of each residue with the ligand binding pocket (residues 16–23 and 70–79), still in the absence of ligand. Based on the geometric distribution of residues with a strong binding pocket correlation factor [Fig. 4(a)], two continuous pathways are identified extending from the binding pocket to the opposite side of the PDZ2 domain. Specifically, pathway I starts at the P_0 binding site on strand β2 and extends along the long

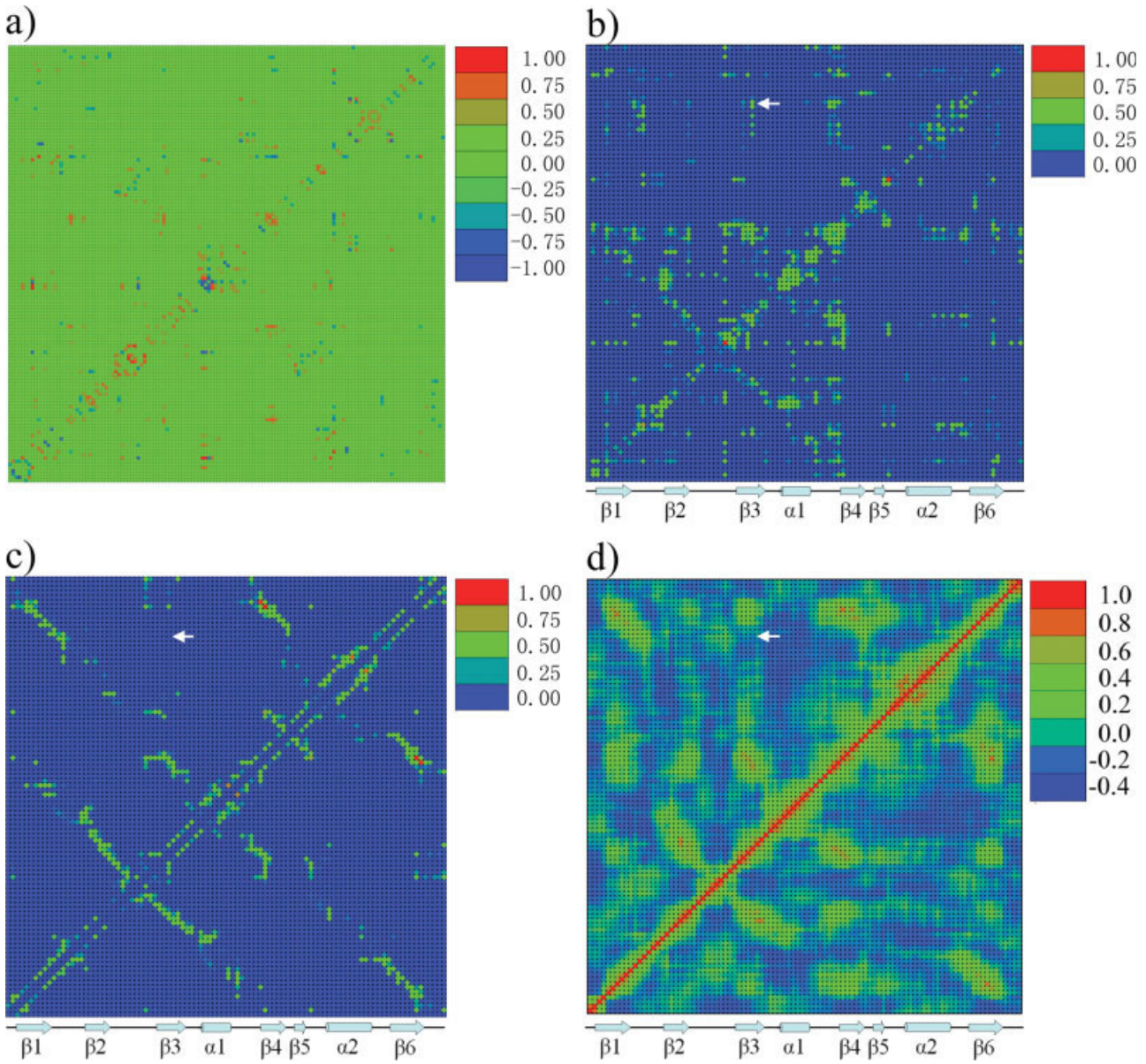

Figure 2

Matrix representation of statistical results for the ligand-free PDZ2 domain. (**a**) The interaction correlation matrix. The dimension of this symmetric matrix is 136 × 136, and the range of interaction correlation value is from −0.93 to 0.92. Each column or row represents a specific residue–residue nonbonded interaction. The color scale is from blue (−1) to red (1), and green indicates zero correlation after C_{cutoff} has been applied (larger than −0.4 and smaller than 0.4). (**b**) The residue correlation matrix. The dimension of this symmetric matrix is 96 × 96, corresponding the number of residues in the protein. Each column or row represents a specific residue of PDZ2 domain. The maximum correlation value was scaled to 1. The color scale is from 0 (blue) to 1 (red). (**c**) The dynamic contact map. This matrix has the same dimension and color scheme as (b). Each spot in this matrix indicates the absolute value of average nonbonded interaction energy between two residues, scaled so that the maximum value in this matrix is equal to 1. (**d**) Residue-based displacement correlation calculated from quasi-harmonic analysis based on MD trajectories with total length of 50 ns (see text and Methods section).

axis of helix α1 and pathway II starts at the P_{-2}/P_{-3} binding sites on strand β2 and goes perpendicularly across strands β2, β3, β4, β6, and β1. Interestingly, some of the residues in pathway I are coupled with residues in pathway II [Fig. 2(b)]. To clearly identify the residue subgroups, a clustered correlation matrix [Fig. 3(a)] was generated by a Markov cluster (MCL) algorithm[28] (see Methods section) from the residue correlation matrix [Fig. 2(b)]. The indexes of rows/columns are rearranged such that residues with more similar coupling patterns have a shorter distance in their indexes. Five major residue clusters with different coupling patterns

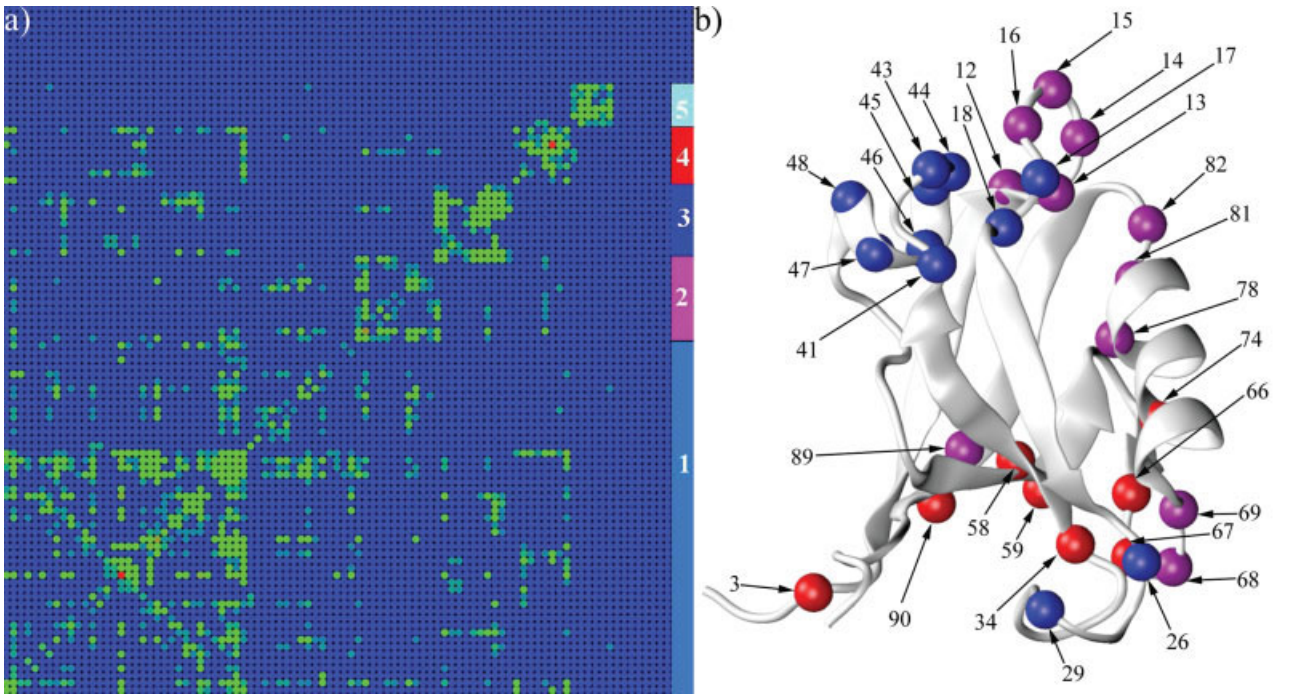

Figure 3

Markov cluster algorithm results. (**a**) The clustered residue correlation matrix (see Methods section). This matrix is the same as the one shown in Figure 2(b), except the column/row indexes are rearranged according to the clustering results. Five clusters were identified and were colored cyan (1), purple (2), blue (3) red (4), and light blue (5) in the right 4 columns. (**b**) Residues in clusters 2, 3, and 4 are displayed as spheres on the PDZ2 domain structure. The colors of the residues are the same as those in (a). Images were made by VMD.[17]

were identified [Fig. 3(a)]. Cluster 1 contains 49 residues and is spread isotropically over the PDZ2 structure (Supplemental Fig. 1a). The coupling among these residues is likely to be due to their contacts and the thermal motions of the protein scaffold. Cluster 5 includes six residues geometrically clustered on the protein surface opposite to the ligand binding pocket (Supplemental Fig. 1b). Cluster 2 (purple, residues 12–16, 68, 69, 78, 81, 82, and 89) is mostly localized at the ligand binding pocket, including loop β1-β2 and N- and C-terminal adjacent loops of helix α2 [Fig. 3(b)]. Cluster 3 (blue) includes helix α1 (residues 41 and 43–48) and two nearby residues (residues 17 and 18) on loop β1-β2 [Fig. 3(b)]. The distribution of cluster 4 shows an approximately linear pattern (it is perpendicular to the central β-sheet), which extends from helix α2 (residue 74), through strand β5 (residues 66 and 67), strand β3 (residue 34), strand β4 (residues 58 and 59), strand β6 (residue 90), and N-terminal loop (residue 3) to the side of PDZ2 domain opposite to the ligand binding pocket [Fig. 3(b)].

The existence of cluster 2 indicates that residues in the binding pocket are allosterically coupled (i.e., residues 15 and 68) even before the ligand binding, which makes the binding a cooperative process. Residues within clusters 3 and 4 are located in the geometrically identified pathways I and II [Fig. 4(a)], respectively. This analysis indicates that the PDZ2 domain has two internal signal transduction pathways, which not only have physical connections in the protein structure, but are also energetically coupled. In other words, after the ligand binds to the pocket between helix α2 and strand β2, the signal of binding can propagate effectively toward the target regions (i.e., the C-terminal part on helix α1 and the N-terminal part on strand β1)[8] because of the preexistence of this long-range coupling.

Ligand-bound state

The structure of PDZ2 domain complexed with its ligand is very similar to its ligand-free form (RMSD 1.2 Å),[19] so that a direct comparison can be made of the dynamic average interaction results. The ligand-bound state interaction correlation of the PDZ2 domain was characterized by the same procedure as used for the ligand-free state; 10 MD trajectories with a total length of 56 ns were used (see Methods section). There are 124 residue–residue interaction pairs with average interaction energies larger than 1 kcal/mol, and the average correlation for 124 × 124 interaction correlation matrix is 0.084 with a standard deviation of 0.077. The residue

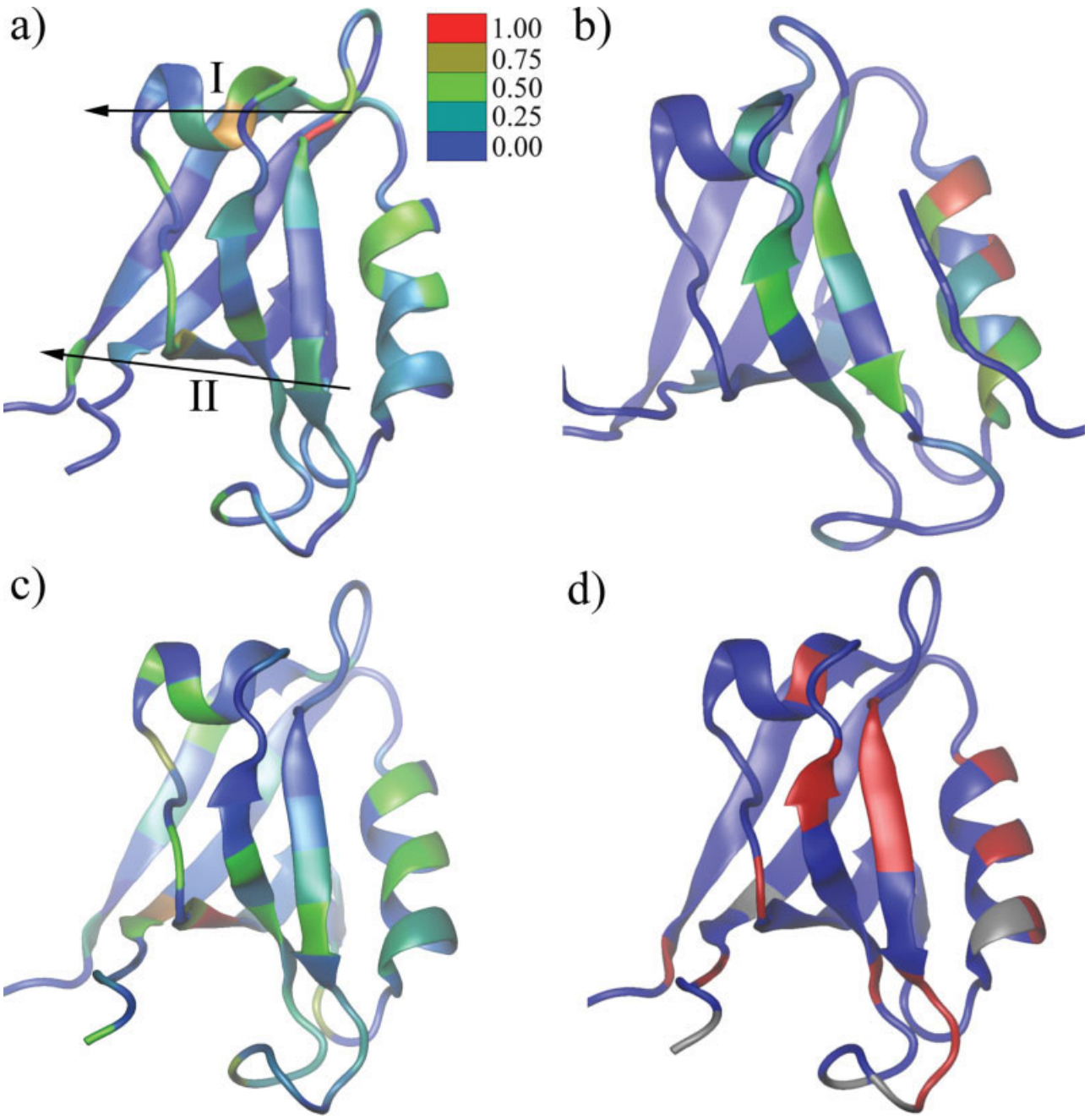

Figure 4

Comparison of interaction correlation results with NMR results and cumulative interaction energy difference on binding. (**a**) The residue correlation factor with the binding pocket (see Methods section) calculated from the PDZ2 ligand-free dynamics. Two proposed signal transduction pathways (I and II) are labeled. (**b**) The residue correlation factor with the binding pocket calculated from PDZ2 ligand-bound dynamics. (**c**) Residues are colored by their cumulative interaction energy difference between the ligand-free and ligand-bound states (see Methods section). (**d**) The backbone hydrogen and nitrogen chemical shift of PDZ2 domain upon titration of human Fas receptor.[12] The blue-to-red gradient represents smaller to larger changes in chemical shift. Grey coloring indicates no data was available. In (a), (b), and (d), the data set was scaled so that the maximum value is equal to 1 and the same color scale is applied. The residue color indicates the coupling strength from blue (none) to red (strongest). All images were made by VMD.[17]

correlation matrix is shown in Supplemental Figure 2 and the binding pocket correlation factor is illustrated in Figure 4(b). As expected, the coupling among residues in the binding pocket becomes stronger (e.g., residues on helix $\alpha 2$) in the presence of ligand. The overall long-distance coupling to the binding pocket, however, becomes weaker than that in the free PDZ2 domain, although residues 40, 44, 45, and 46 on pathway I and residues 35,

36, and 90 on pathway II show similar coupling patterns [cf. Fig. 4(a,b)]. When the ligand is in the binding pocket between helix α2 and sheet β2, it decreases the flexibility of the system and the signal that a ligand is bound appears as a reduction in the remote coupling.

We used the cumulative interaction energy difference (see Methods section) to quantify how each residue is different in its interaction in the ligand-free and ligand-bound states. The results are shown in Figure 4(c), in which the difference for each residue is mapped to the PDZ2 domain structure. Most residues in the binding pocket are significantly perturbed due to the presence of peptide, as expected. Importantly, residues on helix α1 (pathway I) and the continuous residue cluster crossing strands β2, β3, β4, β6, and β1 (pathway II) show different dynamic behavior between two states. The results support the existence of two intramolecule signal transduction pathways, which were identified without reference to the PDZ2 domain structure or dynamics in the ligand-bound form.

Comparison with NMR experiments

A comparison with NMR studies of Walma *et al.*[12] is of interest. They studied the changes of a mouse PDZ2 domain (95% sequence identity with the human PDZ2 used in the current analysis) backbone hydrogen and nitrogen chemical shifts upon binding of a 12-residue peptide. The sequence of this peptide corresponds to the C-terminal end of the human Fas receptor [Fig. 4(d)]. The observed changes were interpreted as an alteration of the chemical environment of a given residue, although no quantitative analysis of the origin of the chemical shifts was made. To compare our computational results to the NMR experiments, the binding pocket correlation factor (CF) was used [see Methods section, Eq. (5)]. The correlation factor of residue i is defined as the sum of the coupling magnitudes between residue i and all residues in the binding pocket. This parameter provides a measure of the coupling of each residue to the ligand binding site. Thus, the amplitude of the CF_i parameter for a residue is expected to be approximately proportional to the perturbation of a residue in response to ligand binding. Figure 4(a) shows the values of the CF_i on the PDZ2 structure. Large signals are observed for residues in or near the binding pocket (helix α2 and strand β2), and the correspondence between the NMR shifts [Fig. 4(d)] and the calculated correlation [Fig. 4(a)] is very good. For regions outside the binding pocket, the correlation is also quite good, i.e., relatively strong signals were observed in the N-terminal region of helix α1, the N-terminal region of strand β1, and the C-terminal region of strand β6 in both the NMR and the simulations. However, there are some differences; a correlation was found between the loop region near the C terminus of helix α1 and the binding pocket, but no chemical shift changes were seen. In another study[23] of the same PDZ domain, a series of mutations was made within the hydrophobic core to probe their effect on ligand binding. Among them, 11 mutations showed different ligand association rates, dissociation rates, or both. The results are consistent with this study, because seven of the 11 mutated positions are included in or are very close to (within two amino acids in sequence) pathways I or II [Figs. 3(b) and 4(a)]. This agreement indicates that the interaction correlation analysis provides a more complete description of the allosteric coupling pathways, since it determines the correlation between all possible residue pairs rather than only the selected mutation sites.

It is noteworthy that in both the computational study and the NMR experiment,[12,23] little change was observed on the side of PDZ2 away from the ligand binding pocket (i.e., the middle portions of strand β1, β4, and β6), suggesting that the signal is rather well-defined and does not spread isotropically.

CONCLUDING DISCUSSION

A primary function of PDZ domains is to bring together different proteins to form multiprotein complexes, but it has been suggested[22] that they also play a signal transduction (allosteric) role.

The processes of signal transduction often involve significant conformational changes in the proteins involved. Normal mode analysis[29,30] and principal component analysis (PCA)[31] are valuable in identifying the functionally related motions by indicating the deformations that require less energy,[32,33] particularly, those involving the relative motion of "rigid" domains with hinges.[32,33] In cases where no significant conformational changes occurs (i.e., activation of G protein-coupled receptors and ligand binding to PDZ domains), NMA and PCA are less useful. Although the PDZ2 domain structures in the ligand-free and ligand-bound state are very similar, a recent study[24] has applied a normal mode analysis to obtain information about the changes in dynamics measured by NMR. In that study, a mode (9th) obtained from an elastic network model of the PDZ2 domain was selected as being the one with the largest overlap with the (small) conformational changes. It was pointed out that the residues with the largest changes in dynamics from NMR are in regions where the fluctuation amplitude was large for that particular mode. This result is interesting but its meaning is not clear since the actual fluctuations predicted by a normal mode model correspond to the sum of the contributions over all the modes. The identified residues were all in or near the binding pocket.

Two interesting nonequilibrium perturbation MD methods, referred as anisotropic thermal diffusion[25] and Pump-Probe MD,[26] have been used to study the signaling pathway of PDZ domains. In the first study, the

protein was simulated at a very low temperature (10 K) and the signal initiation, ligand binding, was introduced by increasing the kinetic energy of a specific residue in the ligand binding pocket to that corresponding to 300 K; the low temperature was used to reduce the realistic thermal noise that would be present if the protein were simulated at 300 K so as to make the effect of the perturbation more obvious. In the second study, the protein was simulated at 300 K and ligand binding was mimicked by applying strong oscillating forces to selected atoms or residues; the propagation of the perturbation was described in terms of increased RMS fluctuations to other residues. Both studies used the PDZ3 domain and perturbed His76 in the binding pocket, as in the study of Lockless and Ranganathan.[20] They identified a coupling pathway similar to pathway I, but did not find pathway II of our work. We suggest several possible reasons for the difference. First and most likely is the fact that both methods simulated the ligand binding by perturbing a single residue as already mentioned, whereas the ligand binding process involves all residues of the binding pocket. Second, the perturbations used are much larger than the realistic effect of binding, so that the dynamic response of the protein may not follow the normal pathway. In this article, the analysis is based on simulations under thermal equilibrium conditions so that these problems do not exist. In addition, we do not exclude the possibility that pathway II is specific to PDZ2 domain, so studies on PDZ3 domain or the PDZ family do not find such a pathway.

The statistical analysis of multiple sequence alignment (MSA)[20] was used by Lockless and Ranaganathan to identify coupling pathways based on residue coevolution. They identified a signal transduction pathway in their study of all PDZ domains that corresponds to pathway I of this study. Although the statistically analysis has a conceptual similarity to the present approach, the two methods make use of different information sources and so are complementary. The data set for our method is an ensemble of conformers generated by MD simulations for a single PDZ domain, whereas that for the MSA method is a series of homologous sequences. The partial agreement of the signal transduction pathways found by the two methods supports the validity of both results. That only pathway I was found, as in the two MD methods mentioned earlier, is likely to be due to the fact that His76 alone was used as the perturbation source, rather than the set of binding pocket residues, as in our case. Application of the two approaches to the same system is important because both methods have limitations. The information extracted from MSA-related study represents characteristics of a protein family. Although the sequences are homologous, members of the family have somewhat different structures and functions. Consequently, assuming that the statistical results describe the behavior of individual proteins is likely to include significant noise. An important aspect of the interaction energy correlation analysis is that it identifies the energetic source of the coupling but is computationally much more expensive than the MSA method.

We hope that this analysis and related studies will serve as a stimulus for experimental research to find specific cases where PDZ domains have a signaling function.

METHODS

Molecular dynamics simulation

The constant pressure and constant temperature (CPT) MD simulation[34–36] was performed on both the apo- and ligand-bound PDZ2 domain. The initial coordinates were taken from the Protein Data Bank, PDB ID 3PDZ (ligand free)[14] and 1D5G (ligand bound).[15] Both systems were simulated with periodic boundary in a rectangular water boxes with dimension of $56 \times 52 \times 44$ Å^3 for 3 PDZ and $64 \times 54 \times 45$ Å^3 for 1 D5G. In both systems, sodium and chlorine ions were added to the physiological concentration (0.15*M*). The system of ligand-free PDZ2 domain has 13,498 atoms including 6 Na^+, 5 Cl^-, and 4027 water molecules, and the system of ligand-bound PDZ2 domain contains 16,584 atoms with 13 Na^+, 7 Cl^-, and 4979 water molecules. The CHARMM package[37] was used for the simulations. The all-atom potential function Charmm27[38] was used for both proteins and modified TIP3 water model[39] was used for the water molecules. After 100 ps equilibration, 10 independent simulations were performed on both systems with an integration step-size of 2 fs and length of 5 ns. The temperature was kept at 298 K and pressure at 1 atmosphere pressure. The coordinates from the MD trajectories were saved every 0.2 ps.

Calculation of statistical parameters

The analysis of the dynamics trajectories were started by investigating residue–residue interaction energy (only amino acids have been included). For each recorded frame, the nonbonded interaction energy $E_{i,j}$ between two residues i and j is defined as

$$E_{i,j} = E_{i,j}^{\text{elec}} + E_{i,j}^{\text{vdw}} \quad (|i-j| > 1) \tag{1}$$

where $E_{i,j}^{\text{elec}}$ and $E_{i,j}^{\text{vdw}}$ is the electrostatic and van der Waals interaction energy between residue i and j. The interaction energies between any amino acid residue pairs neighboring in sequence were not included, because the nonbond interaction energy between the covalently bonded residue pair is not sensitive to the conforma-

tional states of the protein. The average interaction energy between residues i and j is defined as follows:

$$\overline{E}_{i,j} = \frac{1}{f}\sum_{t=1}^{f} E_{i,j}^{t} \quad (|i-j|>1), \tag{2}$$

where $E_{i,j}^{t}$ is the interaction energy between residues i and j in frame t, and f is the number or coordinate sets involved in this analysis [Fig. 3(c)]. The correlation between two sets of residue–residue interactions i,j and k,l is defined as follows:

$$C_{i,j|k,l} = \frac{\sum_{t=1}^{f}\left(E_{i,j}^{t} - \overline{E_{i,j}}\right)\left(E_{k,l}^{t} - \overline{E_{k,l}}\right)}{\sum_{t=1}^{f}\sqrt{\left(E_{i,j}^{t} - \overline{E_{i,j}}\right)^{2}\left(E_{k,l}^{t} - \overline{E_{k,l}}\right)^{2}}} \tag{3}$$

A particular interaction would be involved in this correlation matrix only if the absolute value of its average interaction energy is greater than 1 kcal/mol.

Matrix assembly and clustering

From Eq. (3), the interaction energy correlation matrix has columns or rows corresponding to all non-neighboring residue–residue interactions [Fig. 2(a)]. The matrix size is subjected to the correlation cutoff (C_{cutoff}), although the original distribution is not. In this study of PDZ2 domain, the C_{cutoff} was set to 0.4, which generates a 136 × 136 interaction correlation matrix [Fig. 2(a)]. The residue correlation matrix is a projection of the interaction energy correlation matrix on the residue space. The correlation between residue i and j, $\text{RC}_{i,j}$ is defined as follows:

$$\text{RC}_{i,j} = \sum_{m=I}^{N}\sum_{n=I}^{N}\left|C_{m|n}\times\delta_{m|n}^{i,j}\right|, \tag{4}$$

where the dimension of the interaction correlation matrix is N, $\delta_{m|n}^{i,j}$ is equal to 1 only if residues i and j are involved in interactions m and n, or n and m, respectively; otherwise, $\delta_{m|n}^{i,j}$ is set to zero. For example, if the correlation between interaction pairs 10,25|17,30 is 0.35 and 10,14|17,35 is 0.45, the residue correlation between 10 and 17 would be 0.80 based on these two interaction energy correlations. The dimension of this matrix is equal to the number of residues in PDZ2 domain (96).

The residue correlation matrix was clustered by an unsupervised approach, the Markov cluster algorithm.[28] In current implementation, all parameters were used as their default values. The essential idea of this algorithm is to simulate flow within a graph, which can be derived from our residue correlation matrix. By enhancing flow where the current is strong and decreasing flow where the current is weak, the current across borders between different groups will wither away and the clusters in the graph will be revealed.

Calculation of binding pocket correlation factor

Based on the MD simulation of PDZ2 domain bound with the ligand, residues were categorized as in binding pocket if the absolute value of its average interaction energy with the ligand was larger than 1 kcal/mol. The binding pocket-related residue correlation matrix was constructed by setting all columns/rows of residue correlation matrix to zero, except for the ones that correspond to residues (residues 16–23 and 70–79) in the binding pocket. The correlation of residue i to the binding pocket is defined as follows:

$$\text{CF}_i = \sum_{j=1}^{\text{NRES}} \text{RC}_{i,j}^{\text{bp}} \tag{5}$$

where CF_i is the correlation factor of residue i refer to the binding pocket, $\text{RC}_{i,j}^{\text{bp}}$, is the ith row and jth column of the binding pocket-related residue correlation matrix, and NRES is the number of residue (96) in PDZ2 domain.

Residue deviation between ligand unbound and bound states

The difference between the ligand unbound and bound states is calculated by the cumulative difference in interaction energy of each residue. Based on the average interaction energy matrixes, the interaction energy deviation of residue i, ΔE_i is defined as follows:

$$\Delta E_i = \sum_{j=1}^{\text{NRES}}\left|\overline{E_{i,j}^{\text{apo}}} - \overline{E_{i,j}^{\text{pept}}}\right|, \tag{6}$$

where NRES is the number of residues (96) in the PDZ2 domain, and $\overline{E_{ij}^{\text{apo}}}$ and $\overline{E_{ij}^{\text{pept}}}$ are the average interaction energy between residues i and j in the presence and absence of the ligand, respectively.

ACKNOWLEDGMENTS

The authors thank Professor Gavin MacBeath for careful reading of the manuscript and particularly for suggesting the MCL algorithms. We thank Felix Koziol for managing the computer systems at Harvard University. The authors also thank the National Energy Research Scientific Computing Center for providing computing resource. The research at Harvard was supported, in part, by a grant from the National Institutes of Health.

REFERENCES

1. Sheng M, Sala C. PDZ domains and the organization of supramolecular complexes. Annu Rev Neurosci 2001;24:1–29.
2. Fanning AS, Anderson JM. PDZ domains and the formation of protein networks at the plasma membrane. Curr Top Microbiol Immunol 1998;228:209–233.
3. Fanning AS, Anderson JM. PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. J Clin Invest 1999;103:767–772.
4. Albrecht DE, Froehner SC. Syntrophins and dystrobrevins: defining the dystrophin scaffold at synapses. Neurosignals 2002;11:123–129.
5. Nourry C, Grant SG, Borg JP. PDZ domain proteins: plug and play! Sci STKE 2003;2003:RE7.
6. Hung AY, Sheng M. PDZ domains: structural modules for protein complex assembly. J Biol Chem 2002;277:5699–5702.
7. Zhang M, Wang W. Organization of signaling complexes by PDZ-domain scaffold proteins. Acc Chem Res 2003;36:530–538.
8. Yan J, Wen W, Xu W, Long JF, Adams ME, Froehner SC, Zhang M. Structure of the split PH domain and distinct lipid-binding properties of the PH-PDZ supramodule of α-syntrophin. EMBO J 2005; 24:3985–3995.
9. 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.
10. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 1996;85:1067–1076.
11. Feng W, Shi Y, Li M, Zhang M. Tandem PDZ repeats in glutamate receptor-interacting proteins have a novel mode of PDZ domain-mediated target binding. Nat Struct Biol 2003;10:972–978.
12. Walma T, Spronk CA, Tessari M, Aelen J, Schepens J, Hendriks W, Vuister GW. Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. J Mol Biol 2002;316:1101–1110.
13. Walma T, Aelen J, Nabuurs SB, Oostendorp M, van den Berk L, Hendriks W, Vuister GW. A closed binding pocket and global destabilization modify the binding properties of an alternatively spliced form of the second PDZ domain of PTP-BL. Structure (Camb) 2004;12:11–20.
14. Kozlov G, Gehring K, Ekiel I. Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. Biochemistry 2000;39: 2572–2580.
15. Kozlov G, Banville D, Gehring K, Ekiel I. Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the β2-β3 loop to PDZ domain-ligand interactions. J Mol Biol 2002;320:813–820.
16. Kang BS, Devedjiev Y, Derewenda U, Derewenda ZS. The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography. J Mol Biol 2004;338:483–493.
17. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38;27–38.
18. DeLano WL. The PyMOL Molecular Graphics System on World Wide Web. Available at http://www.pymol.org. 2002.
19. Fuentes EJ, Der CJ, Lee AL. Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. J Mol Biol 2004;335:1105–1115.
20. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–299.
21. Peterson FC, Penkert RR, Volkman BF, Prehoda KE. Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. Mol Cell 2004;13:665–676.
22. Brenman JE, Topinka JR, Cooper EC, McGee AW, Rosen J, Milroy T, Ralston HJ, Bredt DS. Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A. J Neurosci 1998;18:8805–8813.
23. 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.
24. 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.
25. Ota N, Agard DA. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. J Mol Biol 2005;351:345–354.
26. Sharp K, Skinner JJ. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. Proteins 2006; 65:347–361.
27. Kong Y, Karplus M. The signaling pathway of rhodopsin. Structure 2007;15:611–623.
28. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 2002; 30:1575–1584.
29. Brooks B, Karplus M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. Proc Natl Acad Sci USA 1985;82:4995–4999.
30. Brooks CL, III, Karplus M, Pettitt BM. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. Adv Chem Phys 1988;1:1–249.
31. Levy RM, Karplus M, Kushick J, Perahia D. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. Macromolecules 1984;17:1370–1375.
32. Ma J. New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. Curr Protein Pept Sci 2004;5:119–123.
33. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure (Camb) 2005;13:373–380.
34. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 1980;72:2384–2393.
35. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. Phys Rev A 1985;31:1695–1697.
36. Nose S, Klein ML. Constant pressure molecular dynamics for molecular systems. Mol Phys 1983;50:1055–1076.
37. 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.
38. MacKerell AD, Jr, Bashford D, Bellott M, Dunbrack RL, Jr, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, III, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of protein. J Phys Chem B 1998;102:3586–3616.
39. Jorgensen W. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc 1981;103:335–340.