



**HAL**  
open science

# Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains

Giulia Morra, Alessandro Genoni, Giorgio Colombo

► **To cite this version:**

Giulia Morra, Alessandro Genoni, Giorgio Colombo. Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains. *Journal of Chemical Theory and Computation*, 2014, 10 (12), pp.5677-5689. 10.1021/ct500326g . hal-02196470

**HAL Id: hal-02196470**

**<https://hal.univ-lorraine.fr/hal-02196470>**

Submitted on 2 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The mechanisms of differential allosteric modulation in homologous proteins:

## Insights from the analysis of internal dynamics and energetics of PDZ domains

*Giulia Morra<sup>1\*</sup>, Alessandro Genoni<sup>1,2,3</sup> and Giorgio Colombo<sup>1\*</sup>*

<sup>1</sup>Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche

Via Mario Bianco 9, 20131, Milano (Italy)

<sup>2</sup> CNRS, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France.

<sup>3</sup> Université de Lorraine, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506,  
France.

\*Authors to whom correspondence should be addressed:

Giulia Morra; [giulia.morra@icrm.cnr.it](mailto:giulia.morra@icrm.cnr.it)

Giorgio Colombo: [giorgio.colombo@icrm.cnr.it](mailto:giorgio.colombo@icrm.cnr.it)

Tel.: +39-02-28500031; Fax: +39-02-28901239.

## **Abstract**

Allostery is a general phenomenon in proteins whereby a perturbation at one site reverberates into a functional change at another one through modulation of its conformational dynamics. Herein, we address the problem of how the molecular signal encoded by a ligand is differentially transmitted through the structures of two homologous PDZ proteins: PDZ2, which responds to binding with structural and dynamical changes in regions distal from the ligand-site, and PDZ3, which is characterized by less intense dynamical variations. We use novel methods of analysis of MD simulations in the unbound and bound states to investigate the determinants of the differential allosteric behavior of the two proteins. The analysis of the correlations between the re-distribution of stabilization energy and local fluctuation patterns highlights the nucleus of residues responsible for the stabilization of the 3D fold, the stability core, as the substructure that defines the difference in the allosteric response: in PDZ2 it undergoes a consistent dynamic and energetic reorganization, while in PDZ3 it remains largely unperturbed. Specifically, we observe for PDZ2 a significant anticorrelation between the motions of distal loops and residues of the stability core and differences in the correlation patterns between the bound and unbound states. Such variation is not observed in PDZ3, indicating that its internal energetics and dynamics are not affected by presence/absence of the ligand. Finally, we propose a model with a direct link between the modulation of the structural, energetic and dynamic properties of a protein and its allosteric response to a perturbation.



## Introduction

Proteins oversee most of the essential functions in cells through finely tuned structural dynamic mechanisms determined by intra and intermolecular interactions. Experimental and theoretical studies have provided fundamental insights into some of the basic principles underpinning the connection between structure, dynamics and function<sup>1-7</sup>. It has been shown that proteins are best described as dynamical entities that explore a range of substates in a finely regulated equilibrium. Conformational dynamics plays a key role in facilitating transitions among different substates, endowed with different functions or recognition properties and compatible with specific cellular requirements<sup>8-10</sup>. One of the central mechanisms that modulate these transitions is allostery, which can be defined as any process in which an event at one site on a protein or complex impacts the function, by altering the dynamics, or distribution of conformations, of another site. More generally, even in the absence of a known functional effect, it is more and more evident that changes in dynamics upon ligand binding are not confined locally at the perturbation site but can extend across the macromolecule, as a global adaptation of the system to compensate the perturbation. In this broader definition, allostery applies to single-domain as well as to multidomain proteins and complexes, where molecular perturbations can be propagated through the structure determining a change in dynamics and possibly in function even in the absence of substantial conformational changes<sup>4,11</sup>.

Such realization has set the basis for renewed interest in understanding the molecular determinants of allostery and discovering allosteric sites, shedding light on fundamental (dis)regulation mechanisms in cells, giving access to new possible sites for drug discovery<sup>12-14</sup>,

and ultimately opening new opportunities for biotechnological<sup>15</sup> and therapeutic exploitation<sup>16</sup>. A broad range of experimental methods has been used to study allostery, including NMR<sup>17, 18</sup> and EPR spectroscopy<sup>19</sup>, X-ray crystallography<sup>20</sup> and fluorescence (FRET) spectroscopy<sup>21</sup>.

Despite the high level of sophistication, experimental techniques are still limited in providing insight at an atomic level into the process of allosteric signal transmission itself. In most cases, allosteric sites and pathways are identified after mutational, structural, and thermodynamic characterization of binding with allosteric modulators (proteins, peptides, or small-molecules) that are either known or discovered empirically<sup>13</sup>. Computational biology approaches offer a valid complement to characterize intramolecular allosteric mechanisms at an atomic level and generate a theoretical framework for the general understanding of the phenomenon.

One of the most used approaches is the Statistical Coupling Analysis introduced by Ranganathan<sup>22, 23</sup> where pairs of residues that are co-mutated in multiple sequence alignments are used to identify coupling between protein sites and their connecting pathways. Alternative sequence-based methods to identify allosteric networks have also been described<sup>24</sup>. Other computational approaches to allosteric mechanisms focus on protein structure and dynamics. The concept underlying most of these methods is that it is possible to detect long-range coordination and propagation pathways connecting the ligand-binding site to the distal responding regions, by analyzing the correlations between the motions of the different parts. Such approaches range from coarse-grained normal mode analysis<sup>25-28</sup>, and elastic network models, such as in PSN-ENM<sup>29, 30</sup>, which make use of a single structure vibrating around a reference native conformation, to coarse-grained nonlinear network models<sup>31</sup>, to ensemble methods based on all-atom descriptions of the proteins and dynamics simulations<sup>32</sup>. In the latter methods, Monte Carlo or MD-based results and ensembles are analyzed using principal components analysis,

quantification of the interatomic correlations by means of information theory approaches, which allows identifying networks via mutual information based methods<sup>33, 34</sup>, dynamical network analysis<sup>35, 36</sup> and the characterization of local folding/unfolding<sup>37, 38</sup>. A complementary approach focuses on energy couplings instead of motions and tries to define pathways connecting the ligand binding site to other protein regions by evaluating the network of non covalent interactions, either by means of structure-based networks<sup>39</sup> or by means of Molecular Dynamics<sup>40</sup>. The detection of modulations in the coupling network or the presence of cross-correlations between couplings is used to define pathways connecting the ligand binding site and distant regions. In most approaches the comparison between the network in the bound and unbound states at equilibrium, illuminate modulation patterns from which a pathway of similarly responding residues is defined. Other methods introduce localized thermal or mechanical perturbations and monitor the subsequent response of residues as the protein relaxes back to equilibrium<sup>41, 42</sup>.

In the literature, a significant number of theoretical approaches aimed at deciphering allosteric phenomena in proteins have been validated using PDZ as a model system, either focusing on fluctuations and on interaction energies. PDZ domains (the acronym comes from the synapse-associated protein PSD-95/SAP90, the septate junction protein Discs-large, and the tight junction protein ZO-1) constitute a family of recognition domains, which are found in several unrelated proteins. Their function consists of binding the C-terminal ends of interacting partners and participating in signal transduction mechanisms or acting as scaffolding element<sup>43, 44</sup>. The fold is very conserved among PDZ domains, in spite of a rather variable sequence. In particular, one of the family members, PDZ2 has been shown to respond to peptide binding with a structural and dynamical modulation of loop regions distal from the ligand site<sup>45</sup>. In contrast, PDZ3, is

characterized by the absence of strong dynamic and structural variations at regions located either near or far from the ligand site <sup>46</sup>. However, it has been interestingly shown that the dynamical modulation upon binding in PDZ3 is recovered by deleting the C-terminal helix H3 <sup>47</sup>.

In particular, a functional role for the structural and dynamical allostery observed in PDZ2 upon binding has been hypothesized. Specifically, it has been observed that the distal region comprising helix H1 and beta strand 1, far from the binding site, interacts with the PDZ1 domain adjacent to PDZ2 in PTP-BL, thereby modulating the peptide binding specificity of PDZ2 <sup>48</sup>.

In this paper, we will explore the allosteric communication in two different representative PDZ domains by using Molecular Dynamics, with the aim of correlating the modulation of fluctuations to the global structural and energetic rearrangements occurring through the protein upon binding. In this context, we will analyze PDZ2 from tyrosine phosphatase PTP-1E, and PDZ3 from the synaptic protein PSD-95. The two domains, which show a sequence identity of 38%, differ for the presence of a C-terminal structural element, namely helix H3, which is present in PDZ3 but not in PDZ2. These proteins represent ideal test systems to analyze the molecular determinants at the basis of differential allosteric responses.

To gain an atomistic understanding of the molecular reasons that determine the differential behavior, we focus on the internal energetics and the fluctuation patterns of the conformational ensembles in the unbound and bound states of each protein. We observe that in PDZ2 the structure is partly rearranged and the stabilization energy is significantly redistributed in the absence of the ligand compared to the bound state, indicating a structural reorganization of the residues forming the energy core. In contrast, in PDZ3 the stability core is less modulated by the ligand and the pattern of important stabilizing energetic interactions is highly conserved in the



bound and unbound cases. Interestingly, the stability core is in direct contact with the ligand in both cases. At the same time distal loop fluctuations are observed to vary in PDZ2, upon changing the binding state, indicating that such dynamic variations may be coupled to the energetic reorganization of the core.

The analysis of the correlations between the time evolution of stabilizing energy and residue fluctuations shows for PDZ2 a significant modulation of the motions of dynamic loops anticorrelated to the energy core residues, which we compare to existing experimental data.

Based on these observations we propose a direct atomistic link between the modulation of the global energetic properties and the internal dynamics of a protein in determining its allosteric response to a certain perturbation. While based on PDZ domains as test systems, we envisage our method as fully general and immediately transferable to other different systems.

## **Results and Discussion**

PDZ2 from protein tyrosine phosphatase PTP-1E , in its free state (PDB structure: 3LNX) and in complex with the RAGEF2 C-terminal peptide, (sequence: ENEQVSAV) (PDB structure: 3LNY) <sup>49</sup> were simulated for 400ns in explicit solvent. The same protocol was applied to PDZ3 from the synaptic protein PSD-95, both in complex with the pentapeptide CRIPT (sequence: KQTSV), and in the unbound state (PDB structures: 1BE9 and 1BFE <sup>50</sup>).

Moreover, two additional groups of simulations were performed to investigate the role of helix H3 and the terminal beta sheet (the additional structural elements characterizing only PDZ3) in the stabilization and binding properties of PDZ3: to this end, a PDZ3 construct truncated at helix

H3 was modeled in both the peptide-bound and unbound states, and each new construct was simulated for 400ns.

For every system, two independent trajectories were produced as described in Methods.

The conformational and dynamic properties of the different systems were first characterized by the analysis of a set of classical structural parameters.

**Structural changes.** Overall, the secondary structures of all systems are stable during MD, indicating convergence of the simulations. Interestingly, the C-alpha RMSD values show a marked increase especially in PDZ2, up to 5 Å in the unbound state, suggesting a significant conformational modulation. When comparing the average structures, calculated over the whole trajectory, with each other and with the crystal structure of the unbound state (Figure 1A, 1B, 1C), a significant rigid rearrangement of helix H2 is visible in one of the two unbound trajectories (C-alpha RMSD vs crystal structure = 3.9 Å), while the other one remains closer to the starting structure (RMSD below 3 Å). The bound state trajectories (Figure 1D, 1E, 1F) stay overall closer to the crystal structure. However, both the bound and the unbound states of PDZ2 show an early unfolding of helix H1, occurring within 10ns after equilibration and coupled to the formation of an aromatic interaction between His53 and Phe7 (see Supplemental Information SI9). A few refolding events are also observed during the MD. This dynamic behavior of helix H1 can therefore be related to the high mobility of the N terminal strand. Interestingly, these results are corroborated by the recent observation of helix H1 unfolding in MD simulations of PDZ2 from protein tyrosine phosphatase PTP-1E<sup>51</sup>.

Cluster analysis (Supplemental Information SI1) confirms a population of different conformations for PDZ2, particularly in the absence of ligand. In the unbound state, the loop next to the binding site of PDZ2, L23, explores different arrangements, and frequently contacts

the strand preceding H2. Coupled to this is the variability in the orientation of helix H2. Moreover, upper loops L12 and L56 populate different conformations, with L12 bending toward helix H2. Conversely, in the bound state loop L23 interacts with the ligand, which also coordinates H2, reducing its orientation freedom (see Supplemental Information SI1).

In contrast, the overall arrangement of PDZ3, including the loops, is in general conserved during the trajectories of both the bound and the unbound state (see Figure 2), as confirmed by cluster analysis (Supplemental Information SI2). A notable exception is given by the C-terminal beta sheet, which is found in different conformations in the unbound state, and again by loop L12, which is more variable in the bound state. However, comparison to the MD average conformation and X-ray structure of the apo PDZ3 (PDB: 1BFE<sup>50</sup>) yields an RMSD value of 2.3 Å for the bound state and 3.1 for the unbound state average conformation, respectively (2.6 if excluding the last 20 residues). Compared to the crystal and also to the average peptide bound MD structures, the main difference between bound and unbound states arises from the arrangement of loop L12 and from a tilting of helix H2 towards the interior of the binding site in the unbound state simulation. The different orientation of the helix was already discussed as a ligand modulated event in previous studies<sup>52</sup>. However, our data show that the rigid displacement of helix H2 is much more pronounced in PDZ2 than in PDZ3. Interestingly, recent NMR data and MD simulations of the same PDZ2 system in the presence of a photoswitch in the binding site probed the rigid conformational dynamics of H2 mimicking ligand binding<sup>51</sup>.

Moreover, a recent MD-based study on the promiscuity of different PDZ domains<sup>53</sup> highlighted BTPL-PDZ2 as the family member whose binding site, albeit rigid, mostly deviates from the bound state when simulated in the apo conformation, suggesting a significant induced-fit component upon ligand binding.

**Fluctuations.** The RMSF fluctuation profile was calculated considering residue motions with respect to the average position and updating the average structure each 10ns of trajectory, both for the bound and for the unbound state. The resulting profile in Figure 3, shows the profiles obtained for bound state (red) and unbound state (black), respectively. In PDZ2, the presence (absence) of the ligand reverberates in a modulation of the fluctuations in several regions. Overall, fluctuations are higher in the absence of ligand, around residues 20-27 (binding site, beta strand 2 and loop L23), 36, 38-39, 52-60 (L34), and 73,79 (H2). In contrast, they increase in the presence of the ligand at residues 43,50.

The modulation pattern observed in PDZ2 over 400ns MD shows agreement with experimentally measured dynamical changes at the single residue level, derived in particular from NMR measurements of methyl relaxation<sup>7,49,54</sup>. In those studies, distal surface 1 comprising methyl containing residues 61-64, 66, 69, 81, 87, corresponding to beta strand 4, loop L45, beta 6, as well as distal surface 2, made of residues 39 and 40 before helix H1 were shown to be modulated on the ps-ns scale upon binding. Restrained Molecular Dynamics<sup>55</sup>, based on the same NMR-derived restraints, was able to reproduce a dynamic modulation pattern corresponding to the experimental data.

Some ligand effects on residue fluctuations are also visible in PDZ3 (Figure 3). There is an increase in fluctuations in the bound system at residues 317-320 (L12) as well as at residues 343-345 at strand 3 and helix H1. Moreover, the C-terminal residue 405, below H3, is significantly more rigid with the ligand. The remaining part of the protein profile does not show a univocal ligand-dependent behavior and, in particular, the local fluctuations minima, corresponding to the most rigid spots, are not modulated by the ligand. This points to a difference between PDZ2 and

PDZ3 domains in terms of their dynamical response to the ligand state, namely a higher ligand-dependent plasticity for PDZ2 and a modulation localized in the C-terminal end for PDZ3. These data are overall corroborated by experimental observations based on a combination of NMR, ultrafast kinetics and mutation analysis performed by Gianni et al <sup>56</sup>, who showed that PDZ3 does not undergo the conformational changes and structural modulations observed upon binding for PDZ2.

**Ligand-protein interactions.** We first examined the energetic impact of the ligand on each residue by calculating the average non-bonded interactions, namely van der Waals and electrostatic terms, between the bound peptide and the protein amino acids (Figure 3, green profile). Residues sensing the ligand will show a non-zero energy coupling during the MD trajectory and are located at short distance from the peptide.

The average ligand binding hotspots observed for PDZ2 and PDZ3 comprise beta strand 2 (res. 17 to 28 in PDZ2 and residues 320 to 330 in PDZ3) and helix H2 (70 to 80 in PDZ2 and 372 to 380 in PDZ3) forming the binding groove. Interestingly, also beta strand 3, adjacent to the binding strand (residues 32-43 in PDZ2 and 335-345 in PDZ3) experiences a direct interaction with the ligand.

In PDZ3, part of the C-terminal domain (residues 400-410) only marginally interacts with the ligand (at res. 399 and 400). Yet its motions are largely increased in the absence of the bound peptide. Similarly residues 317-320, whose fluctuations decrease in the absence of the ligand, do not directly interact with the peptide, as shown by the analysis of the protein-ligand energy coupling profile. An interesting difference in the binding mode of the two proteins involves helix H2 and the conserved Histidine (His 71 in PDZ2, His 372 in PDZ3), whose interaction with the peptide is more intense in PDZ3.

The residue-based profile of ligand interactions can be compared to the mutational profile of ligand affinity changes for PDZ3 in <sup>57</sup>. Most mutations affecting the binding affinity, identified by MacLaughlin *et al.*<sup>57</sup> are indeed residues that directly interact with the ligand in our simulations, with the exception of single residues like 349 and 353.

**Protein internal energy and its modulation by ligand binding.** Here we address the question of how the local interactions between protein and ligand can modulate the stability of the structure, possibly inducing the experimentally observed dynamic modulation, even without being reflected in a substantially different average structure. In order to investigate the effect on the intra protein interaction energy induced by ligand binding, we compare the overall stabilization energy distribution of the proteins in the free and bound states using the Energy Decomposition Approach<sup>58, 59-61</sup>. By means of this method, one is able to extract from a residue-residue coupling energy matrix, containing all pair non-bonded electrostatic and van der Waals terms, the essential interactions responsible for the stability of the fold (see Methods). The first eigenvector profile recapitulates a ranking of the most significant residues involved in stabilizing interactions and its peaks define the *stability core*. Proteins with similar fold share a similar profile, that is largely determined by the residue-residue packing and hence by the structure. The profile is thus a sensitive measure of the architecture of side chain-side chain interactions within the protein.

Here, we evaluate the energy eigenvector profile from the average energy matrix, calculated over the whole trajectory for each of the simulated systems (Figure 4 and Supplemental Information SI3). As explained in the Methods section, we can neglect the ligand when analyzing the bound trajectory and observe the change in the intra-protein interactions, as if the

ligand were an external bias. A change in the eigenvector profile between bound and unbound state indicates therefore a global structural effect of the ligand. In parallel, we can evaluate the profile of the first eigenvector of a matrix where the average ligand-protein interactions are explicitly considered on the diagonal and their contribution is hence accounted for in the first eigenvector. By comparing the bound state to the unbound state eigenvectors, and the bound state eigenvectors where ligand is included or excluded, we can discriminate to what extent ligand binding is associated with a global reorganization of the core in the protein.

In PDZ2, the profile of the first eigenvector for the bound state calculated without considering the energy of the ligand (red line in Figure 4) points to a well defined *stability core* that includes the N and C terminal ends (residues 4-6 and 85-90 downstream of helix H2) and residues Thr23, Tyr36, Asp56 and Arg57. In contrast, in the unbound state (Figure 4, black line), the ranking of the stabilizing contributions is significantly altered, especially in the trajectory characterized by the stronger re-orientation of helix H2, but to a lesser extent, in the second one (see Supplemental Information S1, S3, S11). In particular, the peaks of N and C terminal core residues, as well as 36, 56 and 57 are reduced relative to residues 60-79, including H2. Structurally, the relative average destabilization of the core, in the unbound state is visible as rearrangement of residues, such as Tyr 36 in strand 3, which starts flipping back and forth from the binding site toward the termini after about 50-70ns, and goes along with a population of arrangements of loop L23 and helix H2 (see cluster analysis in Supplemental Information S1). When considered explicitly (green line versus red line), the ligand adds stabilizing energy, i.e. relatively increases the peaks of strand 2 near the binding site, especially of residues 21,22,23-26, but also of the already mentioned core residues on strand 3.

In PDZ3 (Figure 4 and S3 center panel, Figure S11), the profile of the first eigenvector is dominated by the C-terminal region, comprising helix H3 and the subsequent beta sheet. Other relatively relevant peaks include residues 327, 337, 355,356, 358-360, which contribute to the stability core and, if aligned to PDZ2 as in Figure 4, correspond to the same core region (23,36,56,57) found there and stabilized by the ligand. In contrast to PDZ2, (black profile in Figure 4, center panel), the stabilization profile for PDZ3 does not change significantly in the unbound state, except for the region around 325 and 360, as well as, more intensely, at the C-terminal residues 400-403, at the C-terminal end of helix H3, whose peaks are lower. Structurally, the energy modulation is coupled to a partial unfolding of the terminal beta sheet, observed in both MD trajectories. If the explicit ligand energy is considered in the energy calculation (compare red and green profile in Figure 4), its main contribution to stabilization is located at the helical binding site region H2, around residue His 372. In PDZ3, the core element Phe337, corresponding to PDZ2 Tyr36 flipping, is kept stable by the interactions with the aromatic residues of helix H3, Tyr398 (Figure 1), both in the presence and in the absence of the ligand.

We summarize our findings on energetic as well as dynamic modulation upon binding in a residue-based representation in Table 1. There, we also compare our data with two reference sets of results, one investigating the effects of mutations identified from SCA coupling<sup>23</sup> and the other reporting NMR-based relaxation measurements where available, namely for PDZ2. {Fuentes, 2006 #13016}. The modulation of the energy eigenvector of PDZ2 involves regions including all residues identified by SCA, such as 20,24,25,34,61,62,75,79,85. Consistently, those regions are affected dynamically in our MD simulations in agreement with experimental relaxation measurements. Interestingly, other regions, such as distal surface 2, residues 40-45, are not



included in the SCA set, yet they show dynamical modulation both in experiments and in our simulations. Similarly, in PDZ3 the SCA residues fall into energy modulated regions, which however cover shorter stretches with respect to PDZ2. Interestingly, in PDZ3 the energy modulation is not accompanied by a significant modulation of fluctuations in the same regions of the domain.

We conclude that, the energy eigenvector calculated in the bound and in the unbound state is able to capture significant differences in the modulation of PDZ2 and PDZ3 upon ligand binding/unbinding in agreement with the SCA analysis. While the average structure in solution does not show global differences between bound and unbound state in the different cases, with the exception of the repositioning of helix H2 in one trajectory of PDZ2 (see Figure 1), the energy eigenvector and in particular the core residues can change upon binding. Also, in the unbound state of PDZ2 we see a complex conformational ensemble, characterized by different arrangements of the binding site involving helix H2 and loop L23. Therefore, according to our data, the impact of the ligand on the global structural fluctuations around the binding site and possibly affecting the overall fold, is more evident in PDZ2 than in PDZ3. The structural stability of the latter points to a structural preorganization for binding, which can be related to the higher thermodynamic stability of PDZ3 with respect to PDZ2<sup>62</sup>.

Given these findings, we hypothesize that the dynamic allostery experimentally observed in PDZ is associated to the modulation of core residues. The latter is more evident in PDZ2, where the unbound state is characterized by an alternative arrangement of the binding site, and correspondingly, of core structures; upon binding, the ligand stabilizes one binding site arrangement and correspondingly, one stabilization profile and one core structure. If this is the case, we should be able 1) to quantify a connection between the change in energy of core

residues and the fluctuations, which allows us to define the allosterically responding sites as the ones related to the core energy and 2) to isolate the differences responsible for the differential behavior of PDZ2 versus PDZ3. As to the second point, we address the question whether the main structural differences between PDZ3 and PDZ2, namely helix H3 and the C-terminal beta sheet, which apparently absorb the main destabilization provoked by removing the ligand, are also responsible for the reduced dynamical response of the PDZ3 domain upon unbinding and the enhanced preorganization of the binding site. To shed light on this aspect, we simulated the PDZ3-Delta construct, obtained by deleting helix H3 and C-terminal strand and compared the results in the following section.

**Comparison with PDZ3-Delta construct.** The PDZ3 construct lacking the terminal 25 residues was simulated for 400ns. The cluster analysis (Supplemental Information SI4) shows a stable bound state with no significant changes in the binding site and helix H2 with respect to the starting structure. In the unbound state, a set of representative conformations with higher variability of the loop elements and orientation of helix H2, and overall stable secondary structure elements, characterize the protein structural ensemble. In spite of the increased conformational variability resembling PDZ2, the fluctuation profile obtained for the PDZ3 construct on the 10ns timescale is overall conserved and rather similar to the longer PDZ3 sequence (Supplemental Information SI5). In particular, no ligand-dependent modulation of the fluctuations is observed beyond residue 346 in helix H1.

However, in contrast to the fluctuation data, even if the coupling energy between peptide and truncated PDZ3 molecule is conserved with respect to the non-truncated version (Supplemental

Information SI5), with the obvious exception of the interaction with residue 400, now missing, the stabilization energy profile of truncated PDZ3 (Figure 4 lower panel) is significantly modulated. It namely becomes more similar to the PDZ2 case, in that some energy modulation affects the N terminal strand, namely residues 310-320 (stabilized by the ligand). In contrast to PDZ2, however, the remaining hydrophobic core peaks 337, 355, 357, 359-360, are not destabilized in the unbound state. A high peak at the coordinating Histidine in helix H2, residue 372, due to interactions with the ligand (Figure 4) is present, and was already observed in the non-truncated construct, while it did not occur in PDZ2. According to our analysis, the truncated PDZ3 shows an intermediate behavior between PDZ2 and the full PDZ3 in terms of structural rearrangement of the binding site and average energy core modulation.

The residual discrepancy in the modulation of the stability core between the truncated PDZ3 construct and PDZ2 might be related to the different chemical and packing properties of the residues defining strand 2 and the shorter loop L23, which reduces the intrinsic motion accessible to this region. Indeed, the comparison of the first eigenvectors for the unbound states, taking the sequence alignment into account, shows that the sequence of strand 2 in PDZ3, FNIIGG, corresponds to ISVTGG in PDZ2. Phe326, Ile328 in PDZ3 provide better and more ordered packing than Ile20, Val22 in PDZ2. The critical role played by the substitution Ile-Phe at position 20 in enhancing the stability of the domain upon ligand binding was experimentally evidenced by mutational studies of PDZ2 in <sup>54</sup>. Moreover, although the energy eigenvector shows a clear modulation both in the bound and unbound state indicating a reasonable convergence of this measure (Figure 3, S3 and S11), we cannot exclude that some discrepancy between our fluctuation analysis and the experimental data in <sup>47</sup>, as well as some deviation in the fluctuation profile observed in Figure 3 for the unbound state, may arise from a residual non-equilibrium

behavior of the truncated system, due to the removal of a significant portion of the chain, which might require even longer simulations to be overcome. This should be taken as a critical caveat, although the general consistency of the results with the hypotheses and data from multiple runs supports the validity of the approach.

### **Cross correlation between energy and fluctuations.**

Overall, our results suggest a cross talk between the binding site (beta strand 2, loop L23 and helix H2), the stability core and the motions of the upper loops in PDZ2 and to a smaller extent in the truncated version of PDZ3. With the aim of exploring the relationships between fluctuations and energy modulation of the core, time dependent correlations of energy and fluctuations were investigated, both in the unbound- and in the bound-state simulations.

To this end, we built one residue-based time dependent representation of the stabilization energy contribution shared by each amino acid and one showing the progress of its spatial fluctuation (see Methods). Representative time dependent plots are shown in Supplemental Information Figure S6. We evaluated the cross correlation coefficient, at time lag=0, between the fluctuation time series of each residue, and the energy time series of any other residues. We expect a positive correlation value if the energetic destabilization of a residue is coupled to the increased motion of another residue, and a negative correlation if the energetic stabilization of a residue is coupled to the increased motion of another residue (and vice versa).

To compare two different trajectories of the same system, we introduced a significance threshold for the (positive or negative) correlation coefficient between any two residues, namely we chose to retain only correlation coefficients for which the probability of occurring by chance is less than 1%,  $p < 0.01$ . Then we searched for residue pairs showing a significant correlation of

the same kind in both MD trajectories of the same system. We notice here that the correlation analysis can also be carried out on two single energy and fluctuation time series from the trajectory obtained by concatenating the two independent runs. By filtering out correlation coefficients below the significance threshold  $p < 0.001$ , the same protein regions showing anti correlation to energy are identified (see Supplemental Information figure SI10, blue regions).

In the unbound state of PDZ2, (Figure 5) two groups of residues show fluctuations significantly correlated to stabilization energy. These are: residues 15-25 on loop L12 and strand 2 and residues 46-47 and 58. They are positively correlated to their own stabilization energy and negatively correlated to the energy of core residues. In particular, residues 16,17,21 in beta strand 2 are the ones most coupled to the previously discussed energy peaks 36,56,57,89-90. The same holds for residues 26 to 28 on loop L23, while 46-47 (H1 and loop L34) and 58 are coupled to the N terminus. The correlation between helix H1 and the N terminus confirm the structural observation shown in Supplemental Information figure SI9, where the unfolding of helix H1 has been connected to the interaction with strand 1. Intriguingly, the interaction between helix H1 and strand 1 was hypothesized to be modulated by PDZ1 domain, thereby allosterically affecting the binding specificity of PDZ2<sup>48</sup>.

**Overall**, the stabilization of the core appears associated with more intense fluctuations, which we actually observed in the bound state in the upper loops L12 and L34.

In the PDZ2 bound state, less correlation signals are present, which is due to a lower modulation of the core energy as well as of the binding site motions. As an example, the correlation profile of motions of residue 21 is reported both in the unbound and bound simulations (Supplemental Information SI8). Interestingly, in the bound state, energy-anti-correlated fluctuation emerges

around distal surface 1 residues 61, and residues 76, 78, 85 (coupled to the energy changes around 80). Overall, the modulation of residue fluctuations that are coordinated to the energy changes upon ligand binding offers a better agreement with the NMR relaxation data of <sup>49</sup> than the fluctuation profiles alone (Figure 3 and Table I).

Also, these findings confirm, in agreement with mutational studies, <sup>54</sup> that in PDZ2 the proximity of the ligand to the stability core residues affects its stability through the modulation of residues on beta strand 2, around residue 20, and on loop L23.

A connection between core residues and distal motions is also found in PDZ3 (Figure 6), but the energy anti-correlated fluctuations are here mainly localized at the C-terminal helix and beta sheet. In the unbound state, they are limited to the beta sheet. In the bound state, on the other hand, a stronger signal emerges for helix H3, while the terminal beta sheet has lower fluctuations. Similarly to the case of PDZ2, we notice that ligand binding turns off the energy peaks at the binding site (Figure 5,6).

As expected, in the PDZ3 construct lacking helix H3, fluctuations anti correlated to energy are higher than in the full molecule. In the unbound state a significant fluctuation peak is present on the edge of loop L23 (residue 334) rather close to core residue 336, and other peaks are found at residues 370-375 at helix H2. In the bound state, on the other hand, energy anti correlated fluctuations are found especially at the N terminal residues 305-310, the C-terminal end at 390, with respect to the energy of the binding site residues of beta 2, loop L23 and helix H2 (Figure 7). The increase in energy modulation at the binding site in the presence of the ligand might indicate a less favorable binding and is coupled to remarkable fluctuations on the N terminal side.

We conclude that in PDZ3 the dynamic modulation induced by the ligand is enhanced in the absence of helix H3 and affects regions in the domain that are not modulated in the full protein. In the latter, H3 primarily responds to the ligand by modulating its dynamics, while the binding site and the remaining core are barely perturbed.

This picture agrees with the experimental findings of Petit et al.<sup>47</sup>, who studied by NMR the dynamic modulation of the truncated protein, showing a stronger ligand induced modulation of relaxation parameters compared to the full chain, and also measured a significant decrease in binding affinity in the absence of helix H3. According to our results on the full chain, helix H3 is responsible of the preorganization for binding of beta strand 2 and loop L23. The mediating role played by helix H3 in absorbing the ligand perturbation upon binding was also recently discussed in both the theoretical<sup>30</sup> and experimental contexts<sup>63</sup>. Both papers highlight, in accordance with our findings and observations, the role of the helix module in enhancing the preorganization for binding in different PDZ domains.

## **Conclusions**

The present study analyzes the relationship between non-covalent interactions and internal dynamics of proteins, with the aim of describing the mechanism that underlies the long range modulation observed for PDZ2 and PDZ3 upon ligand binding. From the methodological point of view, we attempted a combination of the current approaches on the study of allostery, namely on one hand the analysis of structural fluctuations and on the other hand the pairwise energy interactions. The main element underlying our approach is the use of the Energy Decomposition Method, which allows us to identify the most stabilizing energy contributions and from there characterize the modulation of the protein stability along the MD trajectory and induced by

binding. We find that, by evaluating the time evolution of the energy eigenvector along our simulations, we can correlate it to the spatial fluctuations. Then, by focusing on the fluctuations that are anti-correlated to the energy of the stabilizing residues, we are able to isolate the allosterically responsive regions. This approach allows us to highlight and measure the relevance of the high energy fluctuations for specific residues as principal effectors of the allosteric change in the protein motions.

In PDZ, intramolecular interactions of core residues can be modulated by the presence or absence of the ligand, even in the absence of stable conformational changes involving the backbone, and their energy fluctuations are then reflected in increased motions at less restrained regions, such as peripheral loops. In particular, the energy core is more significantly rearranged by the ligand in PDZ2 than it is in PDZ3, by means of the tunable region between beta strand 2 and loop L23. This strand, directly sensing the ligand, turns out to be the most responsive region and to be related to the overall dynamic modulation induced by binding, as highlighted by our correlation data as well as by the structural cluster analysis pointing out the interaction between L23 and H2. In PDZ3 this region is less plastic with a higher level of preorganization, due both to an intrinsically more stable core, to a shorter L23, and finally to the presence of helix H3, which shares the most part of stabilization energy of the domain. In fact, removing helix H3 increases the ligand-based modulation of the PDZ3 domain within its stability core, although differently from the PDZ2 case.

As an important caveat, it is worth underlining once more that we aim at studying the coordination between the variations in the energy distribution in the core determined by peptide ligand and the modulation of fluctuations in distal regions. It is clear that introducing mutations in the core can and would most probably result in the modification of the folding-unfolding



profile of the proteins making the present analysis of correlations not immediately applicable. This however would hold true not only for simulative but also for experimental biochemical approaches, in which the relative stabilities of folded states is modified.

Overall, our observations and the agreement with experimental data suggest that calculating **the modulation of the energy eigenvector and** the correlation of fluctuations to stabilization energy might constitute an effective criterion for selecting dynamically responsive sites and therefore identifying allosteric effects from MD simulations.

## **Methods**

### **Molecular Dynamics Simulations**

All MD simulations and standard structural analyses were performed with the GROMACS suite <sup>64</sup>, using the GROMOS force field <sup>65, 66</sup> and the SPC water model <sup>67</sup>. PDZ2 from protein tyrosine phosphatase PTP-1E, in its free state (PDB structure: 3LNX) and in complex with the RAGEF2 C-terminal peptide, (sequence: ENEQVSAV) (PDB structure:3LNY) <sup>49</sup> and of PDZ3 from the synaptic protein PSD-95, both in complex with the pentapeptide CRIPT (sequence: KQTSV), and in the unbound state (PDB structures: 1BE9 and 1BFE) <sup>50</sup> were considered as the starting structure for the simulations, adopting a protonation state consistent with the pH 7.

The Delta construct of PDZ3 was modelled by removing the C-terminal 21 residues from the original sequence of length 110 residues, resulting in a chain of 89 amino acids.

Each system was solvated with around 6000 explicit water molecules filling an octahedral box. After minimization using steepest descent and 1ns of restrained MD simulation, production MD simulations of length 400 ns were run in the NPT ensemble with a time step of 2 fs. Two

independent trajectories were generated with different random seeds. Pressure was controlled with Berendsen barostat <sup>68</sup> and at 1.0 bar with a time constant of 1 ps. The pressure coupling was used isotropically. Bond lengths were constrained by the LINCS algorithm <sup>69</sup>, while water molecules were constrained with the SETTLE algorithm <sup>70</sup>. Temperature was controlled with Berendsen thermostat <sup>68</sup> and at 300 K with a time constant of 0.1 ps. Lennard–Jones interactions were calculated with a 0.9/1.4 nm twin-range cutoff, the short-range part of the electrostatic interactions with a 0.9 nm cutoff, and the particle-mesh-Ewald algorithm was used for the long-range part <sup>71</sup>.

### Energy Decomposition Method

The Energy Decomposition Method is based on the calculation of the interaction matrix  $M_{ij}$ , which is determined by evaluating average, interresidue, nonbonded (van der Waals and electrostatics) interaction energies between residue pairs, calculated over all structures visited during an MD trajectory. For a protein of  $N$  residues, this calculation yields an  $N \times N$  matrix of pair couplings  $m_{ij}$  such that the total average non bonded energy of the protein is given by the sum over the matrix entries. We showed that, after diagonalizing the matrix  $M_{ij}$ , one can approximate pair couplings using the first eigenvalue  $\lambda_1$  and eigenvector  $w$ :

$$E_{MD} = \sum_{i,j=1}^N m_{ij} = \sum_{i,j=1}^N \sum_{k=1}^N \lambda_k w_i^k w_j^k \approx \sum_{i,j=1}^N \lambda_1 w_i^1 w_j^1 \quad (1)$$

The eigenvector profile reports on the single residue contributions to the essential stabilization energy <sup>59,72</sup>.

## Binding energy approximation

Under a continuum representation, in order to implicitly describe the ligand effect within the energy decomposition formalism, we can consider the protein, described by its energy function  $H_0(\mathbf{r})$ , and a perturbation  $\epsilon(\mathbf{r})$ , which is function of the protein's degrees of freedom  $\{\mathbf{r}\}$  and represents the effect of the ligand. The latter is obtained after integrating on the ligand degrees of freedom.

Then, the binding reaction can be thought of as a conformational transition undergone by the protein between state U (unbound) and state B (bound). The protein structural change is due to the “bias” perturbation  $\epsilon(\mathbf{r})$ , which stabilizes state B with respect to state U. The free energy difference is given by:

$$\begin{aligned}\Delta G &= -RT \ln \frac{Z_B}{Z_U} = -RT \ln \frac{\int e^{-\beta(H_0+\epsilon)}}{\int e^{-\beta(H_0)}} \approx -RT \ln \frac{\int_B e^{-\beta(H_0+\epsilon)}}{\int_U e^{-\beta(H_0)}} = \\ &= -RT \langle e^{-\beta\epsilon} \rangle_B - RT \frac{\int_B e^{-\beta(H_0)}}{\int_U e^{-\beta(H_0)}} = -RT \langle e^{-\beta\epsilon} \rangle_B + (G_B^0 - G_U^0)\end{aligned}\quad (2)$$

According to this model, the change in the protein free energy due to binding is separated in two parts: on one hand, the energetic expense paid by the protein to leave the unbound state in order to reach the bound structure (second term), which is unfavorable in the absence of the ligand. On the other hand, the energy gain coming from the ligand (first term) and represented by the ligand energy, which compensates that expense and leads to a globally favorable (negative) free energy change.

In the framework of the Energy Decomposition Approach, we can define, in analogy with the previous free energy description, two energetic terms to describe the binding process. The first one is:

$$-RT \langle e^{-\beta\epsilon} \rangle_B \rightarrow E_{B+l} - E_B \quad (3)$$

where  $E_{B+l}$  and  $E_B$  represent the average non bonded internal energies calculated for the protein in the bound state, taking into account the ligand ( $E_{B+l}$ ) or not ( $E_B$ ). This term reports on the “net” energy contribution provided by the ligand to the bound state. The second term, on the other hand, can be associated with the internal energy difference between bound and unbound state of the protein:

$$(G_B^0 - G_U^0) \rightarrow E_B - E_U \quad (4)$$

The three terms  $E_{B+l}$ ,  $E_B$  and  $E_U$  are then reexpressed in terms of pair couplings in an  $N \times N$  matrix, without considering the explicit degrees of freedom of the ligand.  $E_B$  and  $E_U$  are calculated as in (Equation 1) by considering the bound and unbound state respectively, but in the absence of any ligand. On the other hand  $E_{B+l}$  can be defined as follows:

$$E_{B+l} = \sum_{i,j=1}^N m_{i,j} + \sum_{i=1}^N e_{lig,i} \quad (5)$$

where  $m_{ij}$  are the off diagonal protein-protein non bonded energy terms calculated in the bound state as in  $E_B$ , while  $e_{lig,i}$  are on-diagonal terms recapitulating for each residue  $i$  the non

bonded couplings with the ligand. By diagonalizing this matrix, one hence implicitly accounts for the ligand contribution, which will affect eigenvectors and eigenvalues.

### **Time series and correlation between energy eigenvector and fluctuations**

We divided the 400ns trajectories into sequential 10ns intervals and calculated within each interval the fluctuation profile, by means of the `g_rmsf` tool of GROMACS<sup>64</sup> after fitting the structure on the average for the 10ns window. The option `-res` was used in order to obtain the average RMSF value per residue  $RMSF(i,t)$ .

On the same 10ns window, the energy matrix was also evaluated. Then, by means of the Energy Decomposition Approach, the protein energy eigenvector and its corresponding eigenvalue were calculated. The degrees of freedom of the ligand were implicitly considered on the diagonal (see previous section). Within each time window  $t$ , the stabilization energy of residue  $j$  results as:

$$E(j, t) = \sum_{i=1}^N \lambda_i(t) w_i^1(t) w_j^1(t) \quad (6)$$

In this way, we end up having a fluctuation time series  $RMSF(i,t)$  and an energy time series  $E(j,t)$  for every  $i,j=1,\dots,N$  and for  $t=1,\dots,35$ . The first 50ns of simulation were neglected to cut out possible off-equilibrium fluctuations. The choice of 10ns windows is compatible with the relaxation of self-correlations both of energy components and of fluctuations.

The correlations between each fluctuation time series  $RMSF(i,t)$  and  $E(j,t)$  with  $i,j=1,\dots,N$  and  $t=1,\dots,35$ , were calculated as cross-correlation coefficient:

$$c(i, j) = \frac{\langle (RMSF(i,t) - \langle RMSF(i,t) \rangle) (E(j,t) - \langle E(j,t) \rangle) \rangle}{\langle RMSF(i,t) - \langle RMSF(i,t) \rangle \rangle \langle E(j,t) - \langle E(j,t) \rangle \rangle} \quad (7)$$

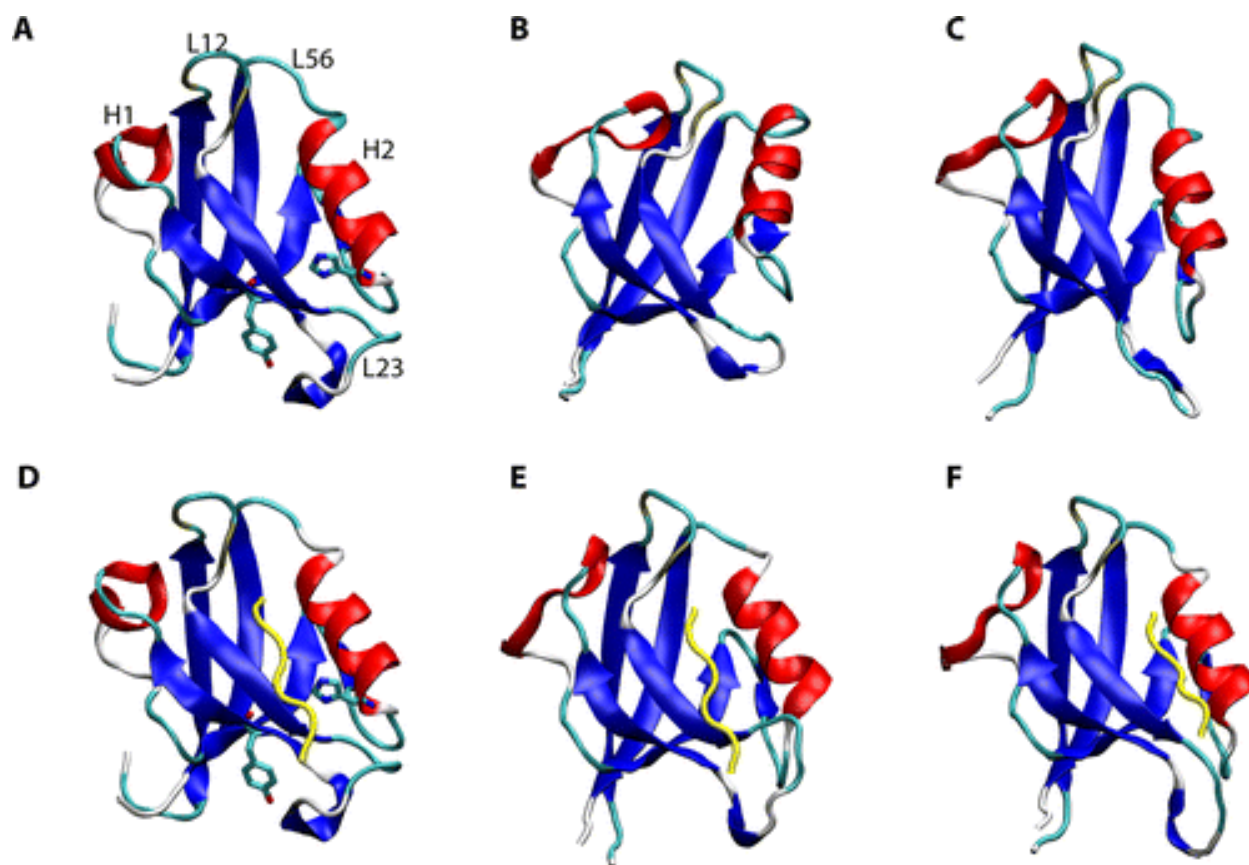
where the  $\langle \dots \rangle$  brackets indicate average over the time series.

## TABLES

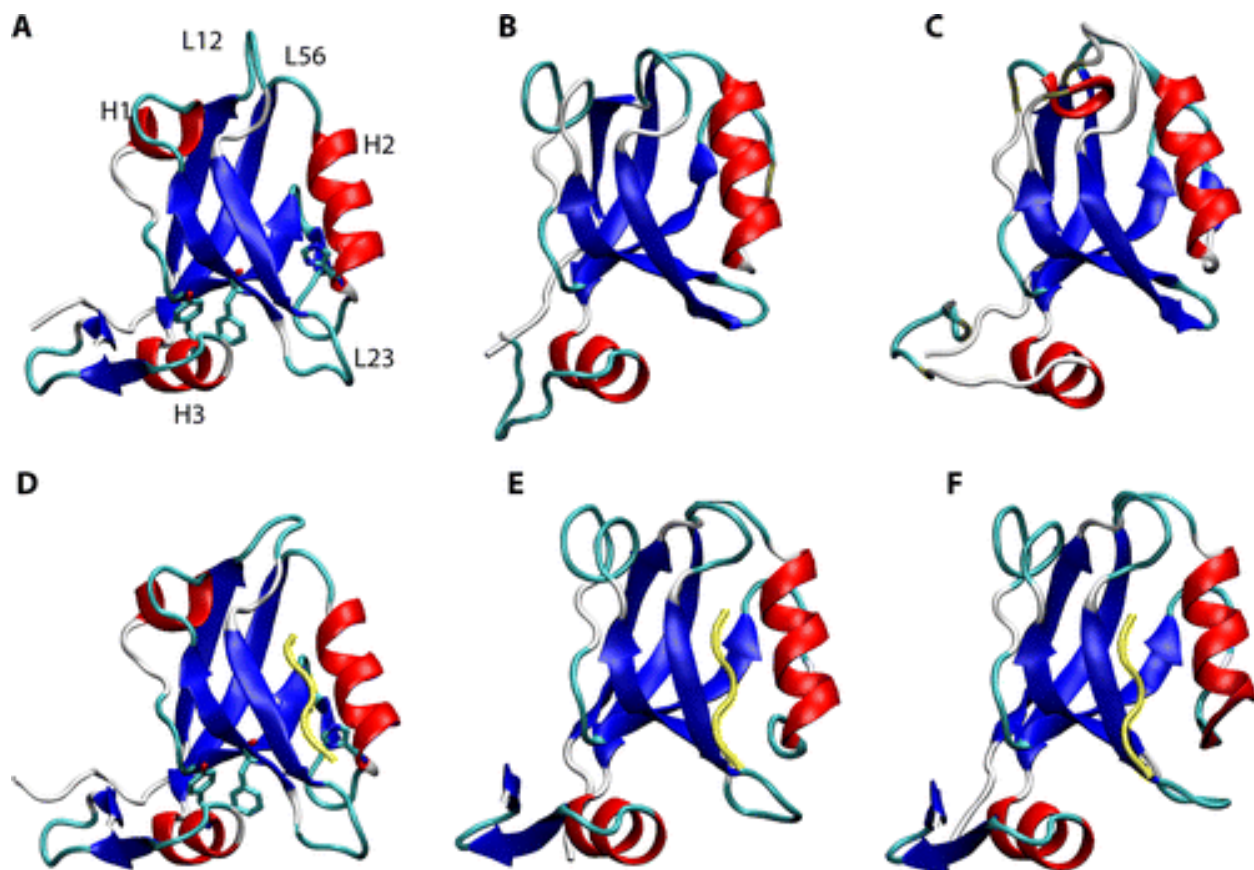
PDZ2	NMR	fluct	ener	SCA	corr	PDZ3	fluct	ener	SCA	corr	Delta	fluct	ener	SCA	corr
15			ener			306					306	ener			
16						307					307				
17						308					308				
18						309					309				
19						318					318				
20						319	fluct				319				
21						320					320				
22						321					321				
23						322					322				
24						324		ener			324				
25						325					325				
26						326					326				
27						327					327				
28						328					328				
29						329					329				
30						330					330				
31						331					331				
32						332					332				
33						333					333				
34						334					334				
35						335					335				
36						336					336				
37						340					340				
38						341					341				
39						342					342				
40						343					343				
41						344					344				
42						345					345				
43						355					355				
44						356					356				
45						357					357				
46						358					358				
47						359					359				
48						360					360				
49						361					361				
50						362					362				
51						363					363				
52						364					364				
53						365					365				
54						366					366				
55						367					367				
56						368					368				
57						369					369				
58						370					370				
59						371					371				
60						372					372				
61						373					373				
62						374					374				
63						375					375				
64						376					376				
65						377					377				
68						378					378				
69						379					379				
70						380					380				
71						383					383				
72						384					384				
73						385					385				
74						386					386				
75						387					387				
76						388					388				
77						396					396				
78						397					397				
79						398					398				
80						399					399				
81						400					400				
85						401					401				
89						402					402				

**Table1.** Residue-based summary of ligand-induced modulation emerging from the current study, compared to experimental data. Left, PDZ2. Center, PDZ3. Right, Delta construct. The fluctuation column reports less (blue) and more (red) flexible regions upon binding resulting from our RMSF analysis. The *ener* column reports (green) modulation hotspots in the energy eigenvector induced by binding. (see also Supp Fig 11) The *corr* column highlights regions for which a significant correlation between fluctuations and core energy is found. Residues identified by Lockless et al.<sup>23</sup> are reported in the *SCA* column. NMR-based measurements of dynamical modulation<sup>54</sup> are reported for PDZ2 in the *NMR* column.

## FIGURES

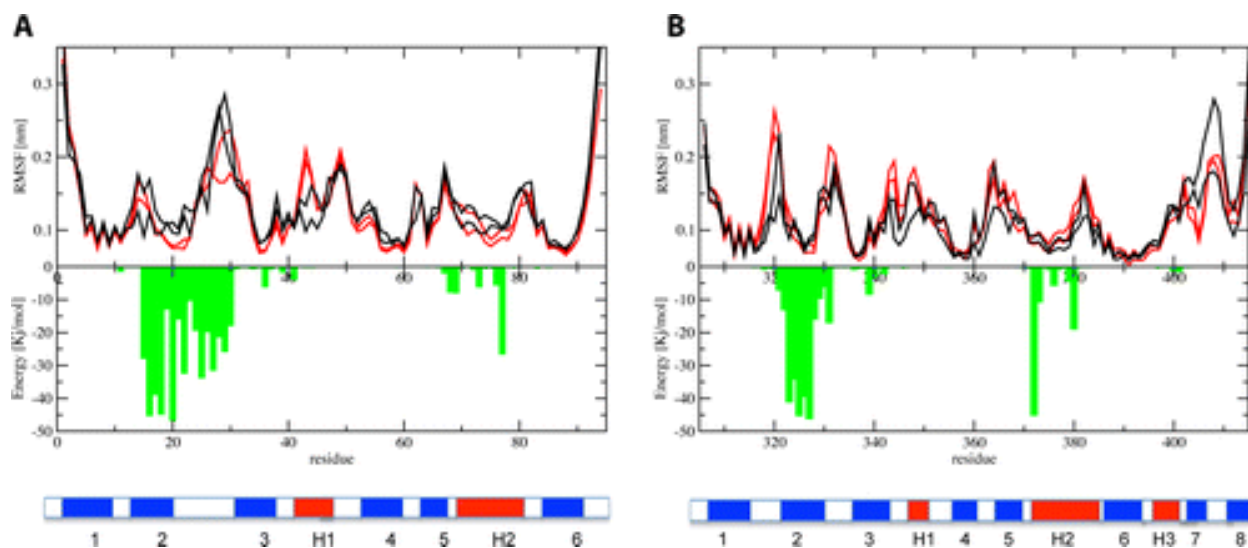


**Figure 1.** Illustrations of PDZ2. (Left) Starting protein structures used for production MD of PDZ2 ((A) unbound and (D) ligand bound), showing the main structural elements discussed in the present work, secondary structure elements, core residue Tyr36 and binding His71. (Middle and Right) Average structures resulting from the two independent 400-ns MD trajectories of (B,C) the unbound state and (E,F) the bound state.

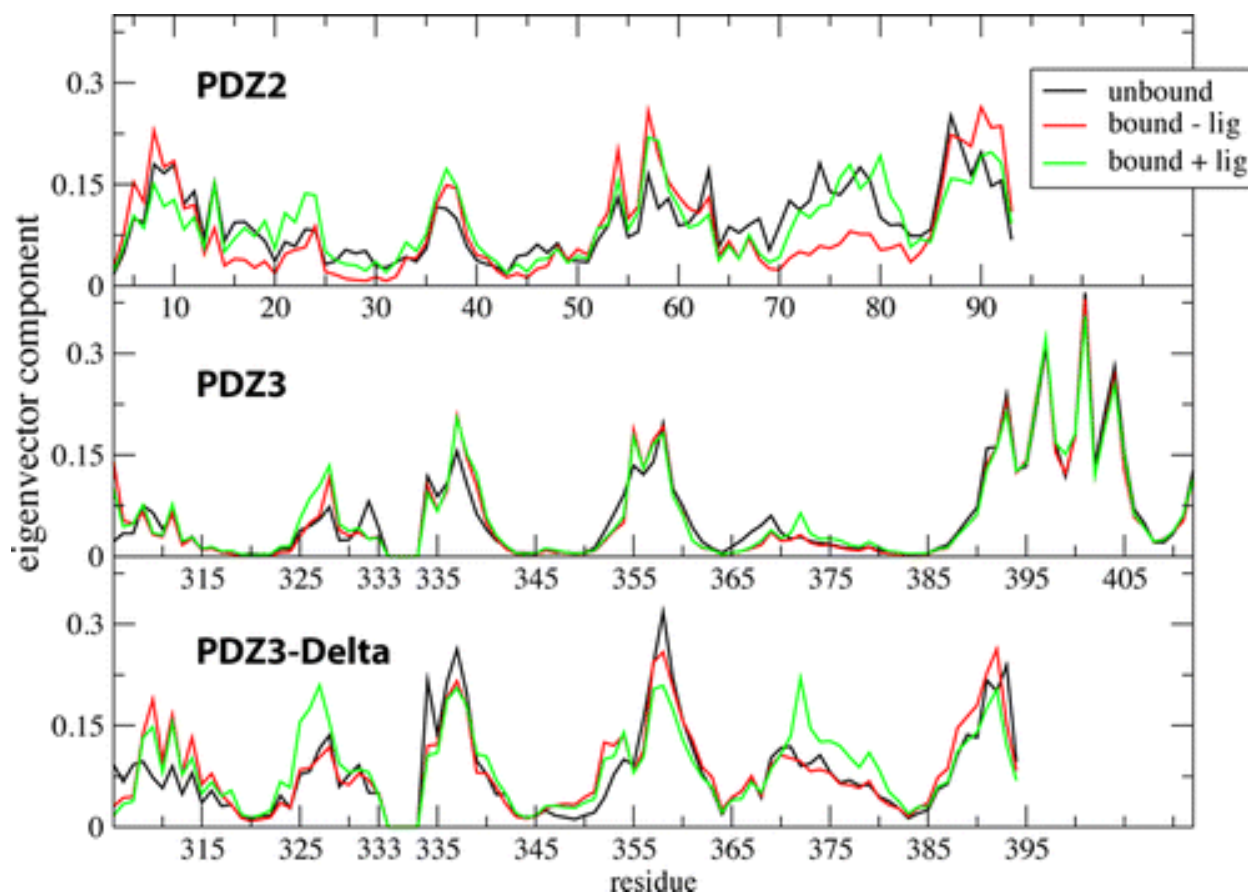


**Figure 2.** Illustrations of PDZ3. (Left) Starting protein structures used for production MD of PDZ3 ((A) unbound and (D) ligand bound), showing the main structural elements discussed in the present work, secondary structure elements, core residue Phe337, binding His376, and Tyr397 on H3. (Middle and Right) Average structures resulting from the two independent 400-ns MD trajectories of (B,C) the unbound state and (E,F) the bound state.

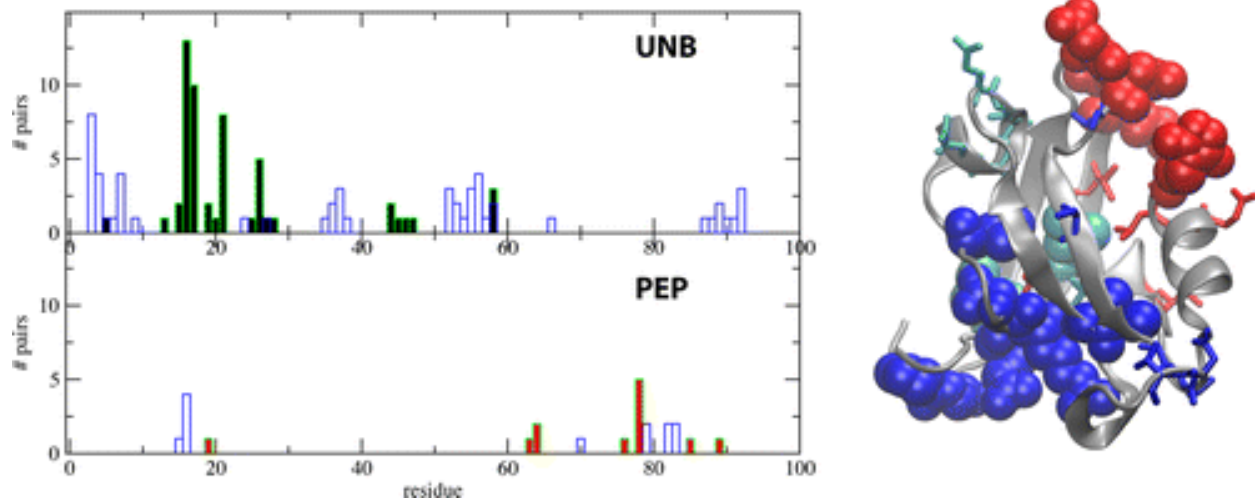




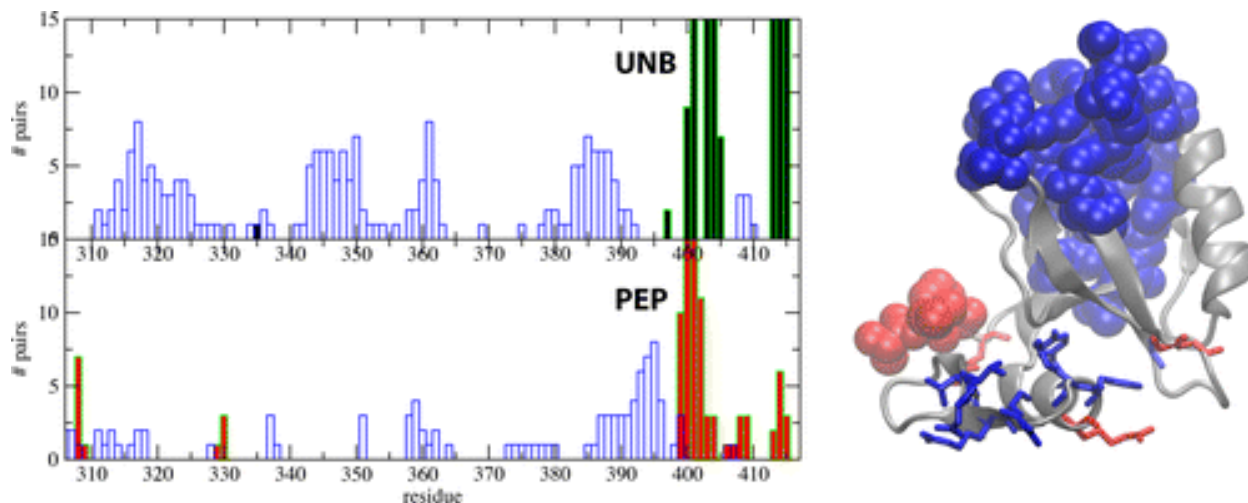
**Figure 3.** RMSF modulation and ligand interactions. (A) PDZ2: (top) residue-based RMSF profile of PDZ2, resulting from the two unbound MD trajectories (black lines) and the two bound MD trajectories (red lines); (bottom) MD average residue based interactions with the ligand (green). (B) PDZ3: (top) residue-based RMSF profile of PDZ3, resulting from the two unbound MD trajectories (black lines) and the two bound MD trajectories (red lines); (bottom) MD average residue-based interactions with the ligand (green).



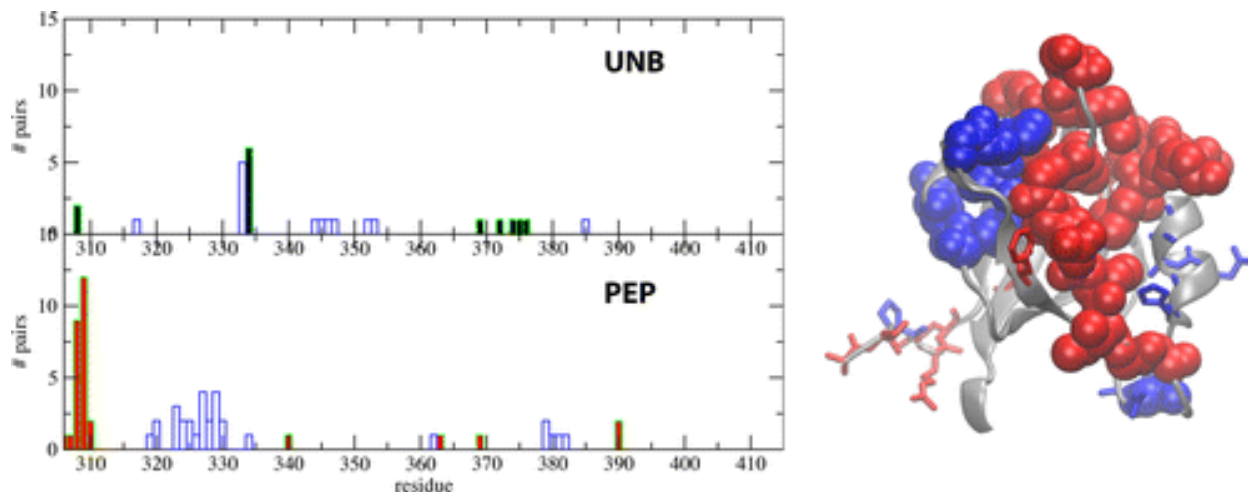
**Figure 4.** Energy profiles. Energy eigenvector profiles obtained by applying the energy decomposition method to the unbound-state MD trajectory (black trace), to the bound-state MD trajectory without considering the ligand (red trace), and to the bound state taking into account the ligand–protein interaction terms (green trace) for PDZ2 (top panel), full length PDZ3 (middle panel), and delta construct PDZ3 (bottom panel). The analysis considers one trajectory per system. PDZ3 and PDZ3-delta are aligned to PDZ2. The energy profile for the second trajectory is given in the Supporting Information.



**Figure 5.** Cross-correlation between RMSF and energy components in PDZ2. (Left) Histogram (top, unbound state, bottom, bound state) showing solid bars for the residues whose fluctuations are negatively correlated to the stabilization energy of residues shown as empty bars; the height of the bars represents the number of counts. (Right) Depiction of the protein structure; sticks correspond to the residues shown as solid bars in the histogram on the left, and spheres correspond to the residues shown as empty bars in the histogram on the left. Blue and cyan colors indicate pairs identified in the unbound state, and red color indicates pairs identified in the bound state.

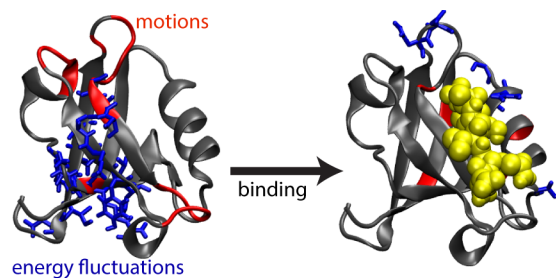


**Figure 6.** Cross-correlation between RMSF and energy components in PDZ3. (Left) Histogram (top, unbound state, bottom, bound state) showing solid bars for the residues whose fluctuations are negatively correlated to the stabilization energy of residues shown as empty bars; the height of the bars represents the number of counts. (Right) Depiction of the protein structure; sticks correspond to the residues shown as solid bars in the histogram on the left, and spheres correspond to the residues shown as empty bars in the histogram on the left. Blue color indicated pairs identified in the unbound state, and red color indicates pairs identified in the bound state.



**Figure 7.** Cross-correlation between RMSF and energy components in the PDZ3-delta construct. (Left) Histogram (top, unbound state, bottom, bound state) showing solid bars for the residues whose fluctuations are negatively correlated to the stabilization energy of residues shown as empty bars; the height of the bars represents the number of counts. (Right) Depiction of the protein structure; sticks correspond to the residues shown as solid bars in the histogram on the left, and spheres correspond to the residues shown as empty bars in the histogram on the left. Blue color indicates the unbound state, and red color indicates the bound state.

**For Table of Contents only:**



## REFERENCES

1. Smock, R. G.; Gierasch, L. M., Sending signals dynamically. *Science* **2009**, 324, 198-203.
2. del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R., The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* **2009**, 17, 1042-1050.
3. Gunasekaran, K.; Ma, B. Y.; Nussinov, R., Is allostery an intrinsic property of all proteins? *Proteins: Struct. Funct. Genet.* **2004**, 57, 433-443.
4. Tsai, C. J.; del Sol, A.; Nussinov, R., Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* **2008**, 378, 1-11.
5. Damm, K. L.; Carlson, H. A., Exploring experimental sources of multiple protein conformations in structure-based drug design. *Journal of the American Chemical Society* **2007**, 129, 8225-8235.
6. Henzler-Wildman, K.; Kern, D., Dynamic Personalities of proteins. *Nature* **2007**, 450, 964-972.
7. Fuentes, E. J.; Der, C. J.; Lee, A. L., Ligand dependent dynamics of intramolecular signaling in a PDZ domain. *J. Mol. Biol.* **2004**, 335, 1105-1115.
8. Schrank, T. P.; Bolen, D. W.; Hilser, V. J., Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. *Proc. Natl. Acad. Sci. USA* **2009**, 106, 16984-16989.
9. Morra, G.; Verkhivker, G. M.; Colombo, G., Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full length dimer. *PLOS Comp. Biol.* **2009**, 5, e1000323.
10. Mayer, M. P., Gymnastics of molecular chaperones. *Mol. Cell* **2010**, 39, 321-331.
11. Cooper, A.; Dryden, D. T., Allostery without conformational change. A plausible model. *Eur. Biophys. J.* **1984**, 11, 103-109.
12. Morra, G.; Neves, M. A. C.; Plescia, C. J.; Tsutsumi, S.; Neckers, L.; Verkhivker, G.; Altieri, D. C.; Colombo, G., Dynamics-Based Discovery of Allosteric Inhibitors: Selection of New Ligands for the C-terminal Domain of Hsp90. *J. Chem. Theory and Computation* **2010**, 6, 2978-2989.
13. Hardy, J. A.; Wells, J. A., Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.* **2004**, 14, 706-715.
14. Genoni, A.; Pennati, M.; Morra, G.; Zaffaroni, N.; Colombo, G., Ligand selection from the analysis of protein conformational substates: new leads targeting the N-terminal domain of Hsp90. *RSC Advances* **2012**, 2, 4268-4283.
15. Jain, M. K.; Berg, O. G., Coupling of the i-face and the active site of phospholipase A2 for interfacial activation. *Current Opinion in Chemical Biology* **2006**, 10, 473-479.
16. Sadowsky, J. D.; Burlingame, M. A.; Wolan, D. W.; McClendon, C. L.; Jacobson, M. P.; Wells, J. A., Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proceedings of the National Academy of Sciences* **2011**, 108, 6056-6061.
17. Kern, D.; Zuiderweg, E. R. P., The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **2003**, 13, 748-757.
18. Swain, J. F.; Dinler, G.; Sivendran, R.; Montgomery, D. L.; Stotz, M.; Gierasch, L. M., Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Molecular Cell* **2007**, 26, 27-39.

19. Oldham, W. M.; Van Eps, N.; Preininger, A. M.; Hubbell, W. L.; Hamm, H. E., Mapping allosteric connections from the receptor to the nucleotide-binding pocket of heterotrimeric G proteins. *Proceedings of the National Academy of Sciences* **2007**, 104, 7927-7932.
20. Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A., X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **2007**, 40, 191-285.
21. Taraska, J. W.; Zagotta, W. N., Structural dynamics in the gating ring of cyclic nucleotide-gated ion channels. *Nat Struct Mol Biol* **2007**, 14, 854-860.
22. Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R., Evolutionary conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **2003**, 10, 59-69.
23. Lockless, S. W.; Ranganathan, R., Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **1999**, 286, 295-299.
24. Dima, R. I., Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Science* **2006**, 15, 258-268.
25. 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. *Biophysical journal* **2005**, 89, 14-21.
26. Hawkins, R. J.; McLeish, T. C. B., Coarse-Grained Model Of Entropic Allostery. *Physical Review Letters* **2004**, 93, 098104.
27. Ming, D.; Wall, M. E., Allostery in a Coarse-Grained Model of Protein Dynamics. *Physical Review Letters* **2005**, 95, 198103.
28. Chennubhotla, C.; Yang, Z.; Bahar, I., Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Molecular BioSystems* **2008**, 4, 287.
29. Raimondi, F.; Felling, A.; Seeber, M.; Mariani, S.; Fanelli, F., A Mixed Protein Structure Network and Elastic Network Model Approach to Predict the Structural Communication in Biomolecular Systems: The PDZ2 Domain from Tyrosine Phosphatase 1E As a Case Study. *Journal of Chemical Theory and Computation* **2013**, 9, 2504-2518.
30. Gerek, Z. N.; Ozkan, S. B., Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS computational biology* **2011**, 7, e1002154.
31. Francesco, P.; Yves-Henri, S., Long-range energy transfer in proteins. *Physical Biology* **2009**, 6, 046014.
32. Nussinov, R.; Tsai, C.-J., Allostery in Disease and in Drug Discovery. *Cell* **2013**, 153, 293-305.
33. Cilia, E.; Vuister, G. W.; Lenaerts, T., Accurate Prediction of the Dynamical Changes within the Second PDZ Domain of PTP1e. *PLoS computational biology* **2012**, 8, e1002794.
34. McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P., Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *Journal of Chemical Theory and Computation* **2009**, 5, 2486-2502.
35. Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z., Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences* **2009**.
36. VanWart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E., Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *Journal of Chemical Theory and Computation* **2012**, 8, 2949-2961.

37. Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N., Structural and Energetic Basis of Allostery. *Annual Review of Biophysics* **2012**, 41, 585-609.
38. Motlagh, H. N.; Li, J.; Thompson, E. B.; Hilser, V. J., Interplay between allostery and intrinsic disorder in an ensemble. *Biochemical Society transactions* **2012**, 40, 975-80.
39. Vijayabaskar, M. S.; Vishveshwara, S., Interaction Energy Based Protein Structure Networks. *Biophysj* **2010**, 99, 3704-3715.
40. Kong, Y.; Karplus, M., Signaling pathways of PDZ2 domain: A molecular dynamics interaction correlation analysis. *Proteins* **2009**, 74, 145-154.
41. Ota, N.; Agard, D. A., Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion  
. *J. Mol. Biol.* **2005**, 351, 345-354.
42. Ho, B. K.; Agard, D. A., Conserved tertiary couplings stabilize elements in the PDZ fold, leading to characteristic patterns of domain conformational flexibility. *Protein Science* **2010**, NA-NA.
43. Erdmann, K. S.; Kuhlmann, J.; Lessmann, V.; Herrmann, L.; Eulenburg, V.; Muller, O.; Heumann, R., The Adenomatous Polyposis Coli-protein (APC) interacts with the protein tyrosine phosphatase PTP-BL via an alternatively spliced PDZ domain. *Oncogene* **2000**, 19, 3894-3901.
44. Kim, E. J.; Sheng, M., PDZ domain proteins of synapses. *Nature Reviews Neuroscience* **2004**, 5, 771-781.
45. Gianni, S.; Haq, S. R.; Montemiglio, L. C.; Jurgens, M. C.; Engstrom, A.; Chi, C. N.; Brunori, M.; Jemth, P., Sequence-specific Long Range Networks in PSD-95/Discs Large/ZO-1 (PDZ) Domains Tune Their Binding Selectivity. *Journal of Biological Chemistry* **2011**, 286, 27167-27175.
46. Chi, C. N. E., Lisa; Shi, Yao; Snall, Tord; Engstrom, Ake; Jemth, Per, Reassessing a sparse energetic network within a single protein domain. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, 4679-4684.
47. Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L., Hidden dynamic allostery in a PDZ domain. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, 106, 18249-18254.
48. van den Berk, L. C. J.; Landi, E.; Walma, T.; Vuister, G. W.; Dente, L.; Hendriks, W. J. A. J., An Allosteric Intramolecular PDZ-PDZ Interaction Modulates PTP-BL PDZ2 Binding Specificity †. *Biochemistry* **2007**, 46, 13629-13637.
49. Zhang, J.; Sapienza, P. J.; Ke, H.; Chang, A.; Hengel, S. R.; Wang, H.; Phillips, J., George N; Lee, A. L., Crystallographic and Nuclear Magnetic Resonance Evaluation of the Impact of Peptide Binding to the Second PDZ Domain of Protein Tyrosine Phosphatase 1E. *Biochemistry* **2010**, 49, 9280-9291.
50. Doyle, D. A. L., Alice; Lewis, John; Kim, Eunjoon; Sheng, Morgan; MacKinnon, Roderick, Crystal Structures of a Complexed and Peptide-Free Membrane Protein-Binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell* **1996**, 85, 1067-1076.
51. Buchli, B.; Waldauer, S. A.; Walser, R.; Donten, M. L.; Pfister, R.; Blöchliger, N.; Steiner, S.; Cafilisch, A.; Zerbe, O.; Hamm, P., Kinetic response of a photoperurbed allosteric protein. *Proceedings of the National Academy of Sciences* **2013**, 110, 11725-11730.
52. Steiner, S.; Cafilisch, A., Peptide binding to the PDZ3 domain by conformational selection. *Proteins* **2012**, 80, 2562-2572.
53. Munz, M.; Hein, J.; Biggin, P. C., The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. *PLoS computational biology* **2012**, 8, e1002749.



54. Fuentes, E. J.; Gilmore, S. A.; Mauldin, R. V.; Lee, A. L., Evaluation of Energetic and Dynamic Coupling Networks in a PDZ Domain Protein. *J. Mol. Biol.* **2006**, 364, 337-351.
55. 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. *Journal of the American Chemical Society* **2008**, 130, 8931-8939.
56. Gianni, S.; Walma, T.; Arcovito, A.; Calosci, N.; Bellelli, A.; Engström, Å.; Travaglini-Allocatelli, C.; Brunori, M.; Jemth, P.; Vuister, G. W., Demonstration of Long-Range Interactions in a PDZ Domain by NMR, Kinetics, and Protein Engineering. *Structure* **2006**, 14, 1801-1809.
57. McLaughlin, J., Richard N; Poelwijk, F. J.; Raman, A.; Gosal, W. S.; Ranganathan, R., The spatial architecture of protein function and adaptation. *Nature* **2012**, 491, 138-142.
58. Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglio, R. A.; Colombo, G., Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Science* **2004**, 13, 113-124.
59. Morra, G.; Colombo, G., Relationship between energy distribution and fold stability: Insights from molecular dynamics simulations of native and mutant proteins. *Proteins: Struct. Funct. and Bioinf.* **2008**, 72, 660-672.
60. Torella, R.; Moroni, E.; Caselle, M.; Morra, G.; Colombo, G., Investigating dynamic and energetic determinants of protein nucleic acid recognition: analysis of the zinc finger zif268-DNA complexes. *BMC Struct. Biol.* **2010**, 10, 42.
61. Morra, G.; Baragli, C.; Colombo, G., Selecting sequences that fold into a defined 3D structure: A new approach for protein design based on molecular dynamics and energetics. *Biophysical chemistry* **2010**, 146, 76-84.
62. Chi, C. N.; Gianni, S.; Calosci, N.; Travaglini-Allocatelli, C.; Engström, Å.; Jemth, P., A conserved folding mechanism for PDZ domains. *FEBS Letters* **2007**, 581, 1109-1113.
63. Bhattacharya, S.; Ju, J. H.; Orlova, N.; Khajeh, J. A.; Cowburn, D.; Bu, Z., Ligand-Induced Dynamic Changes in Extended PDZ Domains from NHERF1. *J. Mol. Biol.* **2013**, 425, 2509-2528.
64. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory and Computation* **2008**, 4, 435-447.
65. van Gunsteren, W. F.; Daura, X.; Mark, A. E., GROMOS Force Field. *Encyclopedia of Computational Chemistry* **1998**, 2, 1211-1216.
66. Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; Gunsteren, W. F. V., The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, 103, 3596-3607.
67. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, P. R., The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, 91, 6269-6271.
68. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 81, 3684-3690.
69. Hess, B.; Bekker, H.; Fraaije, J. G. E. M.; Berendsen, H. J. C., A linear constraint solver for molecular simulations. *J. Comp. Chem.* **1997**, 18, 1463-1472.
70. Miyamoto, S.; Kollman, P. A., SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.* **1992**, 13, 952-962.
71. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, 98.

72. Tiana, G.; Simona, S.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G., Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Science* **2004**, 13, 113-124.

## ASSOCIATED CONTENT

**Supporting Information.** This material is available free of charge via the Internet at

<http://pubs.acs.org>.

## Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Funding Sources

AIRC (Associazione Italiana Ricerca sul Cancro) through the grant: IG 11775; from Fondazione Cariplo through grant 2011.1800 for the RST call –“Premio fondazione cariplo per la ricerca di frontiera”; and from the Italian Ministry of Education and Research through the Flagship (PB05) “InterOmics”.

## ACKNOWLEDGMENT

GC kindly acknowledges funding from AIRC (Associazione Italiana Ricerca sul Cancro) through the grant: IG 11775; from Fondazione Cariplo through grant 2011.1800 for the RST call –“Premio fondazione cariplo per la ricerca di frontiera”; and from the Italian Ministry of Education and Research through the Flagship (PB05) “InterOmics”. GM thanks Dr. Harel Weinstein and Michael LeVine for many interesting discussions.