THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 31, pp. 27167–27175, August 5, 2011

Sequence-specific Long Range Networks in PSD-95/Discs Large/ZO-1 (PDZ) Domains Tune Their Binding Selectivity*[S]

Received for publication, March 16, 2011, and in revised form, May 30, 2011 Published, JBC Papers in Press, June 8, 2011, DOI 10.1074/jbc.M111.239541

Stefano Gianni[‡1], S. Raza Haq[§1,2], Linda C. Montemiglio[‡1], Maike C. Jürgens[§], Åke Engström[§], Celestine N. Chi[§3], Maurizio Brunori[‡], and Per Jemth[§4]

From the [‡]Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del Consiglio Nazionale delle Ricerche, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Sapienza Università di Roma, Piazzale A. Moro 5, 00185 Rome, Italy and the [§]Department of Medical Biochemistry and Microbiology, Uppsala University, Biomedical Center Box 582, SE-75123 Uppsala, Sweden

Protein-protein interactions mediated by modular protein domains are critical for cell scaffolding, differentiation, signaling, and ultimately, evolution. Given the vast number of ligands competing for binding to a limited number of domain families, it is often puzzling how specificity can be achieved. Selectivity may be modulated by intradomain allostery, whereby a remote residue is energetically connected to the functional binding site via side chain or backbone interactions. Whereas several energetic pathways, which could mediate intradomain allostery, have been predicted in modular protein domains, there is a paucity of experimental data to validate their existence and roles. Here, we have identified such functional energetic networks in one of the most common protein-protein interaction modules, the PDZ domain. We used double mutant cycles involving site-directed mutagenesis of both the PDZ domain and the peptide ligand, in conjunction with kinetics to capture the fine energetic details of the networks involved in peptide recognition. We performed the analysis on two homologous PDZ-ligand complexes and found that the energetically coupled residues differ for these two complexes. This result demonstrates that amino acid sequence rather than topology dictates the allosteric pathways. Furthermore, our data support a mechanism whereby the whole domain and not only the binding pocket is optimized for a specific ligand. Such cross-talk between binding sites and remote residues may be used to fine tune target selectivity.

Several studies on protein-protein interaction modules (1), such as SH2, SH3, and PDZ[5] domains, have highlighted the apparent problem of how selectivity is achieved in a cellular context (2–4). For example, the synapse contains several PDZ proteins (5) with apparently overlapping specificity, and yet they have distinct roles in scaffolding and signaling (6). One mechanism that could modulate specificity is intradomain allostery (7). Multisubunit proteins such as hemoglobin are well known for their allosteric behavior whereby the oxygen binding is cooperative because ligation of one subunit will modulate the affinity in the other subunits of the tetramer (8). But allostery is also present within monomeric proteins (9), and it may be invoked even in the absence of a detectable conformational change (10–15).

Previous studies on monomeric globular proteins have predicted allosteric pathways on an amino acid residue level based on analysis of co-evolution (16, 17), molecular dynamics simulations (18), or NMR-constrained molecular dynamics (19). Although the PDZ domain family has served as a prime model system for these studies, experimental data that detect intramolecular allosteric pathways are scarce and sometimes conflicting (16, 20, 21).

In protein folding studies, it proved illuminating to compare homologous proteins to unravel basic determinants of the underlying mechanism (22). Here, we employ the same strategy to assess the structural pattern of intradomain allostery by looking at the coupling free energy ($\Delta\Delta\Delta G_C$) upon binding of peptide ligands to two different members of the PDZ domain family. We used a double mutant cycle analysis (23) to quantify energetic coupling between different residues in the binding reaction (see Fig. 1).

The protein domains chosen for the study, protein tyrosine phosphatase Basophil-like PDZ2 (hereafter called PDZ2) and PSD-95 PDZ3 (PDZ3), are perfectly suited for this type of analysis on allosteric networks because they have been extensively investigated by experiments directed to folding and binding (21, 24–26). Importantly, computational studies on these PDZ domains predicted the presence of distinct allosteric networks (16, 18, 19, 27, 28). In this study, we present experimental evidence for such networks.

MATERIALS AND METHODS

Mutagenesis, Expression, and Purification of Recombinant PDZ2 and PDZ3 Domains—Site-directed mutants were prepared by inverted PCR using the cDNAs encoding pseudo-wild type PDZ2 and pseudo-wild type PDZ3, respectively, as templates. These cDNAs code for a Trp residue in homologous positions for the two domains, namely Y43W and F337W,

* This work was supported by the Swedish Research Council Grant 2009-5659 (to P. J.), the Human Frontiers Science Program (to P. J.), and the Italian Ministero dell'Istruzione dell'Università e della Ricerca FIRB RBRN07BMCT_007 (to M. B.).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1, Figs. S1 and S2, and an additional reference.

[1] These authors contributed equally to this work.

[2] Supported in part by a fellowship from the Higher Education Commission of Pakistan.

[3] To whom correspondence may be addressed. E-mail: Celestine.Chi@imbim.uu.se.

[4] To whom correspondence may be addressed. E-mail: Per.Jemth@imbim.uu.se.

[5] The abbreviations used are: PDZ, PSD-95/Discs large/ZO-1; Abu, aminobutyric acid; PSD-95, postsynaptic density protein 95.

respectively, as probe for binding (20, 24, 25, 29). PDZ2 and PDZ3 were expressed in *Escherichia coli* and purified as described previously (30, 31). Protein concentrations were determined by absorbance measurements using calculated extinction coefficients from the amino acid composition (PDZ3, $\epsilon_{280} = 8500\ \text{M}^{-1}\ \text{cm}^{-1}$; PDZ2, $\epsilon_{280} = 5500\ \text{M}^{-1}\ \text{cm}^{-1}$). The numbering of residues in this article is in accordance with NMR and crystal structures of PDZ2 and PDZ3: Protein Data Bank codes 1VJ6 (25) and 1BE9 (32), respectively.

Ligand Binding Kinetics—Single-mixing kinetic binding experiments were carried out on SX-18MV and SX-20MV stopped-flow instruments (Applied Photophysics, Leatherhead, UK) by following the change in fluorescence emission upon binding. Excitation was at 280 nm, and emission was measured by using either a 330 nm band-pass filter or a 435 nm cut-off filter. We thus used fluorescence resonance energy transfer between the tryptophan in the PDZ domain and a dansyl group covalently bound to the N terminus of the specific peptide. We have shown previously that this dansyl does not affect the binding kinetics of hexapeptides (20, 31, 33), and any small effects would cancel out in the calculation of $\Delta\Delta G$ for wild type and modified peptides.

All mutants were subjected to kinetic binding measurements using three different peptide ligands: i) the wild type peptides, EQVSAV for PDZ2 as described in Gianni *et al.* (25) and YKQTSV derived from the C terminus of the protein CRIPT (34) for PDZ3; ii) the wild type peptides containing a $\text{Val}^0 \rightarrow$ Abu mutation for both PDZ domains and iii) a $\text{Ser}^{-2} \rightarrow$ Thr or $\text{Thr}^{-2} \rightarrow$ Ser mutation in the peptides for PDZ2 and PDZ3, respectively. Abu is an α-aminobutyric acid, *i.e.* a Val with one γ-methyl group deleted. Peptides were purchased from GL Biochem (Shanghai, China) and JPT Peptide Technologies (Berlin, Germany). The concentration of peptide was determined by measuring the dansyl absorbance at 330 nm using $\epsilon_{330} = 4{,}300\ \text{M}^{-1}\ \text{cm}^{-1}$.

All measurements on PDZ2 were performed using 2 μM protein concentration in 50 mM phosphate buffer (pH 7.2) and 0.4 M Na_2SO_4 at 4 °C. Experiments with PDZ3 were performed using 3 μM protein concentration in 50 mM phosphate (pH 7.45) at 10 °C. In all cases, the time course for binding was satisfactorily fitted by a single exponential function. The observed reaction was thus consistent with the simplest kinetic mechanism involving a bimolecular association rate constant and a monomolecular dissociation rate constant. Kinetic traces (supplemental Figs. S1*A* and S2*A*) were fitted to a single exponential equation using the software provided by Applied Photophysics to obtain the observed rate constant, k_{obs} (Equation 1),

$$A = \Delta A_{EQ}(1 - e^{-k_{obs}t}) + C \qquad \text{(Eq. 1)}$$

where A is the fluorescence signal at time t, ΔA_{EQ} is the amplitude of the kinetic trace, and C is the fluorescence at the beginning of the reaction.

The k_{obs} values were then plotted *versus* the peptide concentration and analyzed by either Equation 2, which is valid under pseudo-first-order conditions (PDZ2) or the general equation for a second-order bimolecular association (35) for PDZ3 (Equation 3).

$$k_{obs} = k_{on} \times n + k_{off} \qquad \text{(Eq. 2)}$$

$$k_{obs} = ((k_{on}{}^2(n - [A]_0)^2 + k_{off}{}^2 + 2 \times k_{on}k_{off}(n + [A]_0))^{0.5} \qquad \text{(Eq. 3)}$$

In these equations, k_{on} is the association (on-rate) constant, k_{off} is the dissociation (off-rate) constant, and $[A]_0$ and n are the initial concentrations of PDZ and peptide, respectively.

Displacement Kinetics—The off-rate constant k_{off} for PDZ3 was in many cases $<5\ \text{s}^{-1}$ and could therefore not be estimated with good enough accuracy from Equation 3. All k_{off} values for PDZ3 were therefore determined in displacement reactions (33). PDZ3 and dansylated peptide were first preincubated (0.5–3 μM PDZ and 0.25–3 μM dansylated peptide, depending on the respective affinity between the particular pair of PDZ and peptide). The complex was mixed rapidly with an excess of unlabeled peptide. At sufficiently high concentration of unlabeled peptide the dissociation of dansylated peptide from the protein is the rate-limiting step (k_{off}) (supplemental Fig. S2, *C* and *D*). Kinetic traces were single exponential and fitted to Equation 1 to obtain observed rate constants (supplemental Fig. S2*C*). These k_{obs} values were plotted *versus* the concentration of the unlabeled peptide used to displace the complex (supplemental Fig. S2*D*). The data were analyzed by Equation 4 to estimate a value for k_{off}. This equation is valid under pseudo-first-order conditions, *i.e.* at high concentrations of unlabeled peptide. But, it also serves as a good model-free equation at lower concentrations of unlabeled peptide,

$$k_{obs} = k_{off} + k_{on}' \times K_D/(K_D + [\text{unlabeled peptide}]) \qquad \text{(Eq. 4)}$$

where K_D is the equilibrium dissociation constant for the PDZ and the unlabeled peptide and k_{on}' the apparent (first-order) on-rate constant for the labeled peptide and the PDZ domain at the chosen concentration of labeled peptide. Note that k_{off} can be accurately estimated from this experiment as the asymptotic value of k_{obs} at high concentration of unlabeled peptide. All data were fitted to Equations 2–4 using KaleidaGraph software (Synergy Software).

Calculation of Coupling Energies—Coupling energies were calculated as described (23). In brief, for a hypothetical wild type protein P-AB, the changes in free energy for single mutants at positions A and B can be calculated assuming the following (Equations 5–7).

$$\begin{array}{ccc} \text{P-AB} & \rightarrow & \text{P-A} \\ \downarrow & & \downarrow \\ \text{P-B} & \rightarrow & \text{P} \end{array} \qquad \text{(Eq. 5)}$$

$$\Delta\Delta G_{\text{P-AB}\rightarrow\text{P-B}} = \Delta G_{\text{P-AB}} - \Delta G_{\text{P-B}} \qquad \text{(Eq. 6)}$$

$$\Delta\Delta G_{\text{P-A}\rightarrow\text{P}} = \Delta G_{\text{P-A}} - \Delta G_{\text{P}} \qquad \text{(Eq. 7)}$$

P-A is the variant where B is deleted, P-B is the variant where A is deleted, and P is the variant where both A and B are deleted. By applying a thermodynamic cycle (Equation 5) to the four species P-AB, P-A, P-B, and P, it may be noted that if A and B do not interact energetically, the change in energy upon deleting A

in P-AB ($\Delta\Delta G_{\text{P-AB}\rightarrow\text{P-B}}$) should equal the change in energy upon deleting A in P-A ($\Delta\Delta G_{\text{P-A}\rightarrow\text{P}}$). Thus, a coupling energy $\Delta\Delta\Delta G_C$ may be calculated as the difference in $\Delta\Delta G$ for deleting A in P-AB and for deleting A in P-A.

$$\Delta\Delta\Delta G_C = \Delta\Delta G_{\text{P-AB}\rightarrow\text{P-B}} - \Delta\Delta G_{\text{P-A}\rightarrow\text{P}} \quad \text{(Eq. 8)}$$

A detectable $\Delta\Delta\Delta G_C \neq 0$ implies that the probed positions A and B interact energetically, and its value is a quantitative measure of the strength of this interaction.

Fersht and co-workers (23, 36, 37) initially developed the double mutant cycle methodology for probing intramolecular interactions. But, as shown by Schreiber and Fersht (38, 39), it works equally well for intermolecular interactions, where the energetic interaction between pairs of residues on different molecules can be monitored directly. This first and most detailed analysis of a protein-protein interaction by double mutant cycles was on that between barnase and barstar.

In the case of intermolecular interactions, binding rate constants are measured for the four different species PA (wild type protein), P (single mutant protein where A is deleted), LB (wild type ligand), and L (single mutant ligand where B is deleted). The experimental observables are the association and dissociation rate constants for PA with LB, *i.e.* $k_{\text{on(PA-LB)}}$ and $k_{\text{off(PA-LB)}}$, and to L, $k_{\text{on(PA-L)}}$ and $k_{\text{off(PA-L)}}$, and the association and dissociation rate constants for the mutant protein P with LB, *i.e.* $k_{\text{on(P-LB)}}$ and $k_{\text{off(P-LB)}}$, and to L, $k_{\text{on(P-L)}}$ and $k_{\text{off(P-L)}}$.

By applying the double mutant cycle method (see Fig. 1), a coupling energy $\Delta\Delta\Delta G_C$ may be calculated as follows.

$$\Delta\Delta G_{\text{PA-LB}\rightarrow\text{P-LB}} = \text{RTln}\left(\frac{k_{\text{on(PA-LB)}}}{k_{\text{off(PA-LB)}}}\cdot\frac{k_{\text{off(P-LB)}}}{k_{\text{on(P-LB)}}}\right) \quad \text{(Eq. 9)}$$

$$\Delta\Delta G_{\text{PA-L}\rightarrow\text{P-L}} = \text{RTln}\left(\frac{k_{\text{on(PA-L)}}}{k_{\text{off(PA-L)}}}\cdot\frac{k_{\text{off(P-L)}}}{k_{\text{on(P-L)}}}\right) \quad \text{(Eq. 10)}$$

$$\Delta\Delta\Delta G_C = \Delta\Delta G_{\text{PA-LB}\rightarrow\text{P-LB}} - \Delta\Delta G_{\text{PA-L}\rightarrow\text{P-L}} \quad \text{(Eq. 11)}$$

Two examples of calculation of $\Delta\Delta\Delta G_C$ are given in supplemental Figs. S1 and S2 for PDZ2 and PDZ3, respectively.

RESULTS

Allosteric systems are characterized by an energetic coupling between remote sites in a protein. Indeed, if allostery is at play, the existence of functional energetic networks between residues located in the ligand binding pocket or even the ligand itself, and those in other parts of the protein should be detectable. Pinpointing such networks is the focus of this study.

Double Mutant Cycles—A powerful approach to address quantitatively the energetic connectivity between remote residues in a protein is the double mutant cycle analysis (Fig. 1). This methodology has been used to capture the strength and the location of both intra- and intermolecular interactions in proteins (23, 38).

The basic experimental parameters to be determined are the rate constants for the association and dissociation of a specific ligand to the known binding site on the protein. In a double mutant cycle analysis the kinetics of the wild type protein are compared with those of two single mutants at positions A and B

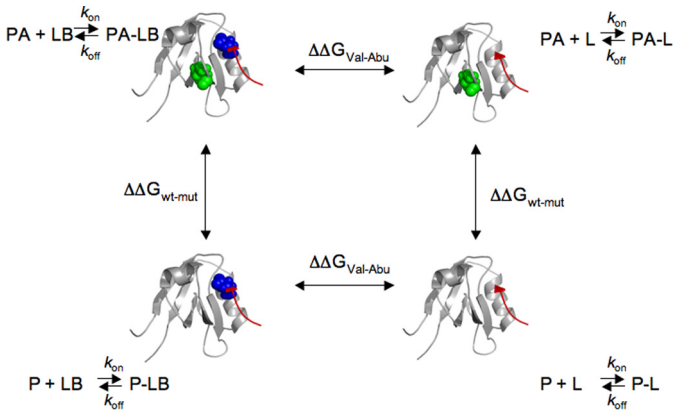

FIGURE 1. **Thermodynamic double mutant cycle for the PDZ-peptide interaction.** As described in the text, we measured the association and dissociation rate constants (k_{on} and k_{off}) for the binding of peptide ligands to PDZ domains. Each thermodynamic cycle involves four different species: PA (wild type protein with residue A in *green*), P (single mutant protein at position A), LB (wild type peptide ligand with residue B in *blue*), and L (single mutant peptide ligand at position B) and their four complexes: PA-LB, PA-L, P-LB, and PL. The free energy of interaction between the two probed positions A on the protein and B on the peptide ligand is calculated from the determined rate constants. In the scheme, positions A and B are exemplified as Val^{44} of the PDZ2 domain (highlighted in *green*) and Val^{0} of the peptide ligand (highlighted in *blue*).

and of the corresponding double mutant (both A and B being mutated). In our study, residue A belongs to the protein, the PDZ domain, and residue B to its specific peptide ligand.

We have studied 31 different mutants for PDZ2 and 31 mutants for PDZ3. The binding kinetics for each one of these mutants have been characterized for three different peptide ligands. The change in binding free energy of the wild type peptide LB, upon mutation of residue A, is quantified by $\Delta\Delta G_{\text{PA-LB}\rightarrow\text{P-LB}}$ (Equation 9). Likewise, the change in binding free energy of the mutant peptide L, upon mutation of residue A, is quantified by $\Delta\Delta G_{\text{PA-L}\rightarrow\text{P-L}}$ (Equation 10). If the two residues A and B are functionally independent in the binding reaction, the effect of mutating residue A in the protein would be identical in the binding reactions with wild type peptide LB and mutant peptide L, respectively. On the other hand, if the residues A and B are energetically coupled, $\Delta\Delta G_{\text{PA-LB}\rightarrow\text{P-LB}}$ and $\Delta\Delta G_{\text{PA-L}\rightarrow\text{P-L}}$ will not be identical and therefore a detectable coupling energy $\Delta\Delta\Delta G_C$, according to Equation 11, will be observed.

Binding Kinetics of PDZ Mutants with Their Peptide Ligands—Side chains of core residues are likely candidates for transmitting allosteric signals through a protein domain (40). The two PDZ domains were extensively mutated such that all regions of the hydrophobic core and a few surface patches were probed by conservative mutations. All 62 mutants, although typically destabilized relative to the wild type, were all folded under the experimental conditions (24, 29).

The binding rate constants were determined by fluorescence spectroscopy using the stopped-flow technique (supplemental Figs. S1 and S2). Association (k_{on}) and dissociation (k_{off}) rate constants with three different hexapeptide ligands (Fig. 2) were determined for each PDZ mutant (supplemental Table S1). These peptide ligands were as follows: (i) the wild type peptide corresponding to a previously identified natural ligand specific for each PDZ domain and (ii) peptides containing a $Val^{0} \rightarrow$ Abu

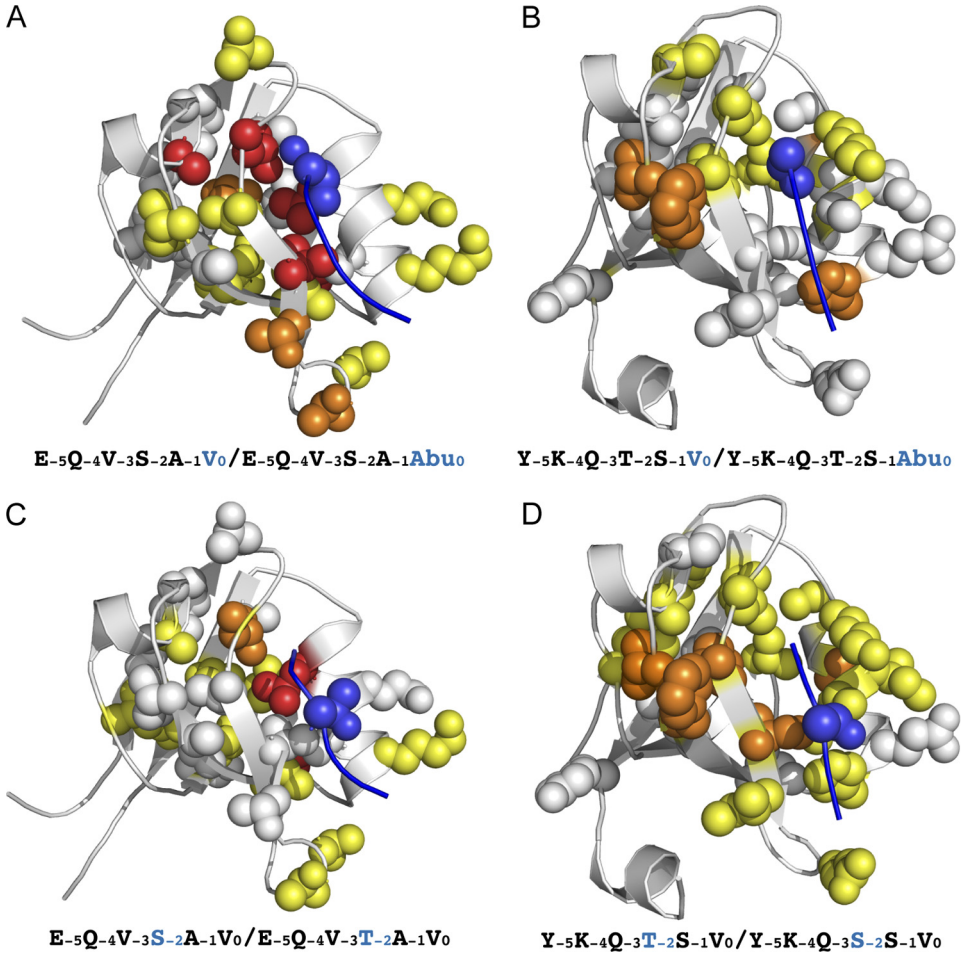

FIGURE 2. **Coupling free energies for the interaction between PDZ domains and their ligands.** All of the mutated residues in protein tyrosine phosphatase Basophil-like (*PTP-BL*) PDZ2 (*A* and *C*) and PSD-95 PDZ3 (*B* and *D*) are represented as *spheres* on the structures of the respective PDZ domain-ligand complex (*white ribbon* representation, Protein Data Bank codes 1VJ6 and 1BE9). The coupling free energies (absolute values) are divided into four groups: $\Delta\Delta\Delta G_C < 0.2$ kcal mol^{-1} (white), $0.2 < \Delta\Delta\Delta G_C < 0.4$ kcal mol^{-1} (*yellow*), $0.4 < \Delta\Delta\Delta G_C < 0.7$ kcal mol^{-1} (*orange*), and $\Delta\Delta\Delta G_C > 0.7$ kcal mol^{-1} (*red*). Coupling energies were calculated from microscopic rate constants as described under "Materials and Methods" and in supplemental Figs. S1 and S2. The peptides used in the different experiments are shown *below* each panel (Val^0 panels, *A* and *B*; or Ser/Thr^{-2} panels, *C* and *D*). The figure was drawn in PyMOL (44).

mutation (deletion of a γ-methyl group from Val^0) or (iii) a Ser^{-2} → Thr or Thr^{-2} → Ser mutation for PDZ2 and PDZ3, respectively (see Fig. 2 for numbering of the peptide).

We chose these peptide mutations because they target residues that interact with the binding pocket of PDZ domains. Furthermore, deletion of a methyl group is a relatively conservative mutation (41) that perturbs the affinity between PDZ and peptide to a perfect degree for analysis (usually 0.5–1 kcal mol^{-1}) and for our stopped flow measurements (k_{off} values between 0.1–300 s^{-1}).

The wild type peptide for PDZ2 contains at the −2 position a Ser (Fig. 2) that could not be mutated to Ala because the ensuing very low affinity between PDZ and the ligand would jeopardize analysis (20); thus, the Ser^{-2} → Thr mutation was chosen to probe this position. Importantly, PDZ2 is expected to display a poor selectivity for Ser *versus* Thr at position −2 of the ligand (42). Thus, in this case, the dissociation equilibrium constant K_D for the PDZ2-peptide complex is essentially unchanged (supplemental Table S1).

Coupling Free Energies between PDZ and Peptide Side Chains—The coupling free energies between each mutated residue in the protein and the side chain of the two peptide side chains were calculated using the determined rate constants k_{on} and k_{off} (Table 1). We found that both PDZ domains display long range communication between the peptide and residues in the protein that are spatially distant from the binding site (up to 15 Å C_α–C_α distance), thereby supporting an allosteric mechanism.

The structural distribution of the coupling energies for the different residues can be seen in Figs. 2 and 3 for each PDZ domain. By comparing the two, it is immediately apparent that the wiring of the long range networks of coupling energies are different for the two PDZ domains, indicating distinct allosteric pathways. The overall coupling pattern for PDZ3 suggests a segment through the core with higher coupling free energies. This segment extends through the core at an angle compared with the two β-sheets of the PDZ-peptide complex. Furthermore, it appears to include the Thr^{-2} of the peptide (Fig. 2, *B* and *D*). For PDZ2, on the other hand, the observed segment of higher coupling free energies goes from the loop between β2 and β3 (*bottom* of Fig. 2, *A* and *C*) and along the bound peptide toward Val^0.

This interesting feature of specific coupling patterns is further demonstrated by the lack of correlation when the coupling

TABLE 1

Coupling free energies $\Delta\Delta\Delta G_C$ for the interaction between residue sidechains in PDZ2 and PDZ3 and sidechains in their respective peptide ligands

Corresponding topological positions are aligned.

PDZ2			PDZ3		
Mutant	$\Delta\Delta\Delta G_c$ (kcal mol^{-1})		Mutant	$\Delta\Delta\Delta G_c$ (kcal mol^{-1})	
	$Val_0 \rightarrow Abu$	$Ser_{-2} \rightarrow Thr$		$Val_0 \rightarrow Abu$	$Thr_{-2} \rightarrow Ser$
V16A	0.14 ± 0.18	-0.02 ± 0.20	I314V	0.15 ± 0.07	0.24 ± 0.06
L18A	0.05 ± 0.18	-0.04 ± 0.23	I316A	0.21 ± 0.07	0.30 ± 0.06
T21S	0.31 ± 0.16	-0.004 ± 0.15			
L25A	1.7 ± 0.4	-0.47 ± 0.13	L323A	0.37 ± 0.08	0.38 ± 0.06
I27V	0.22 ± 0.16	-0.06 ± 0.15	F325A	-0.29 ± 0.09	0.41 ± 0.08
V29A	1.02 ± 0.18	-0.07 ± 0.16	I327V	0.19 ± 0.18	0.48 ± 0.06
T30S	0.45 ± 0.20	0.06 ± 0.14	V328A	0.10 ± 0.10	0.22 ± 0.09
V33A	0.06 ± 0.21	0.28 ± 0.15			
V33G	0.22 ± 0.16	0.26 ± 0.13			
T35G	0.40 ± 016	0.16 ± 0.13			
T35S	0.52 ± 0.19	0.32 ± 0.14			
			E331A	0.14 ± 0.27	0.23 ± 0.29
I42V	0.33 ± 0.20	0.22 ± 0.13	I336V	0.06 ± 0.32	0.03 ± 0.32
V44A	0.30 ± 0.19	0.01 ± 0.16	I338A	-0.15 ± 0.11	0.09 ± 0.11
A46G	0.10 ± 0.16	0.15 ± 0.19	F340A	0.46 ± 0.10	0.45 ± 0.08
I47V	0.35 ± 0.16	0.06 ± 0.16	I341V	0.23 ± 0.07	0.32 ± 0.07
			I341A	0.39 ± 0.08	0.57 ± 0.14
A52G	0.12 ± 0.16	0.13 ± 0.15	P346G	0.35 ± 0.13	0.19 ± 0.12
A53G	0.80 ± 0.17	0.28 ± 0.22	A347G	0.19 ± 0.09	0.25 ± 0.08
I59V	0.17 ± 0.16	-0.30 ± 0.13	L353A	0.08 ± 0.08	0.36 ± 0.07
V65A	0.36 ± 0.19	-0.36 ± 0.20			
L66A	0.22 ± 0.19	-0.05 ± 0.20			
V68A	0.18 ± 0.19	0.40 ± 0.43	V362A	0.29 ± 0.07	0.20 ± 0.06
			H372A	0.59 ± 0.16	0.24 ± 0.09
L73A	1.64 ± 0.17	1.08 ± 0.19			
K79G	0.17 ± 0.15	0.26 ± 0.18	E373G	0.18 ± 0.07	0.15 ± 0.06
K79A	0.24 ± 0.16	0.09 ± 0.23	E373A	0.16 ± 0.08	0.14 ± 0.07
A81G	0.15 ± 0.16	-0.002 ± 0.16	A375G	0.08 ± 0.13	0.22 ± 0.11
			A376G	0.27 ± 0.08	0.32 ± 0.07
E83G	0.18 ± 0.18	-0.11 ± 0.17	I377G	0.14 ± 0.08	0.30 ± 0.07
E83A	0.31 ± 0.16	-0.04 ± 0.17	I377A	-0.23 ± 0.16	-0.09 ± 0.16
			A378G	0.22 ± 0.09	0.40 ± 0.09
L85A	0.80 ± 0.16	1.16 ± 0.54			
			K380G	0.27 ± 0.16	0.14 ± 0.09
			K380A	0.45 ± 0.08	0.34 ± 0.09
			A382G	-0.04 ± 0.08	0.23 ± 0.07
V92A	0.18 ± 0.15	0.09 ± 0.14	V386A	0.16 ± 0.08	0.32 ± 0.07
L94A	0.46 ± 0.22	-0.38 ± 0.18	I388V	0.06 ± 0.09	0.09 ± 0.09
L96A	0.06 ± 0.24	-0.21 ± 0.21	A390G	0.10 ± 0.08	-0.10 ± 0.17
			Y392A	0.29 ± 0.09	0.16 ± 0.09

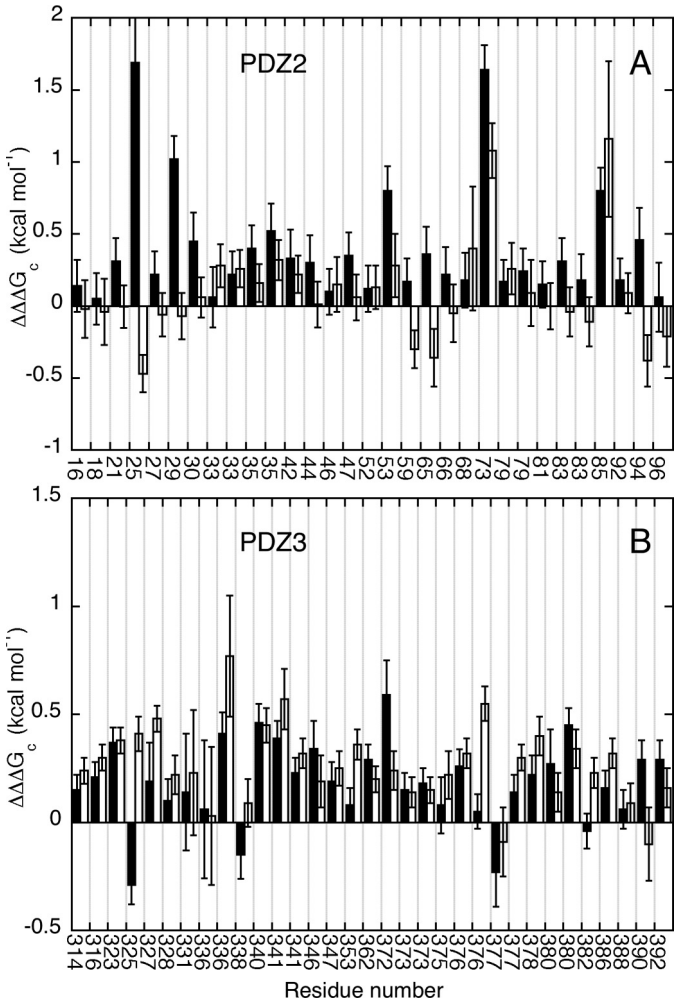

FIGURE 3. **Histograms of coupling energies for the PDZ-peptide interactions.** *A*, PDZ2; *B*, PDZ3. *Black bars* in both panels represent the $\Delta\Delta\Delta G_C$ values calculated for the interaction between the two PDZ domains and peptides containing a $Val^0 \rightarrow$ Abu mutation, whereas the *white bars* show the $\Delta\Delta\Delta G_C$ values determined for the interaction between PDZ domains and the peptides containing a $Ser^{-2} \rightarrow Thr$ or $Thr^{-2} \rightarrow Ser$ mutation for PDZ2 and PDZ3, respectively. *Error bars* show S.E. Note that coupling free energies are mainly positive except for the nonselective Ser^{-2} position in the PDZ2 peptide (see "Discussion").

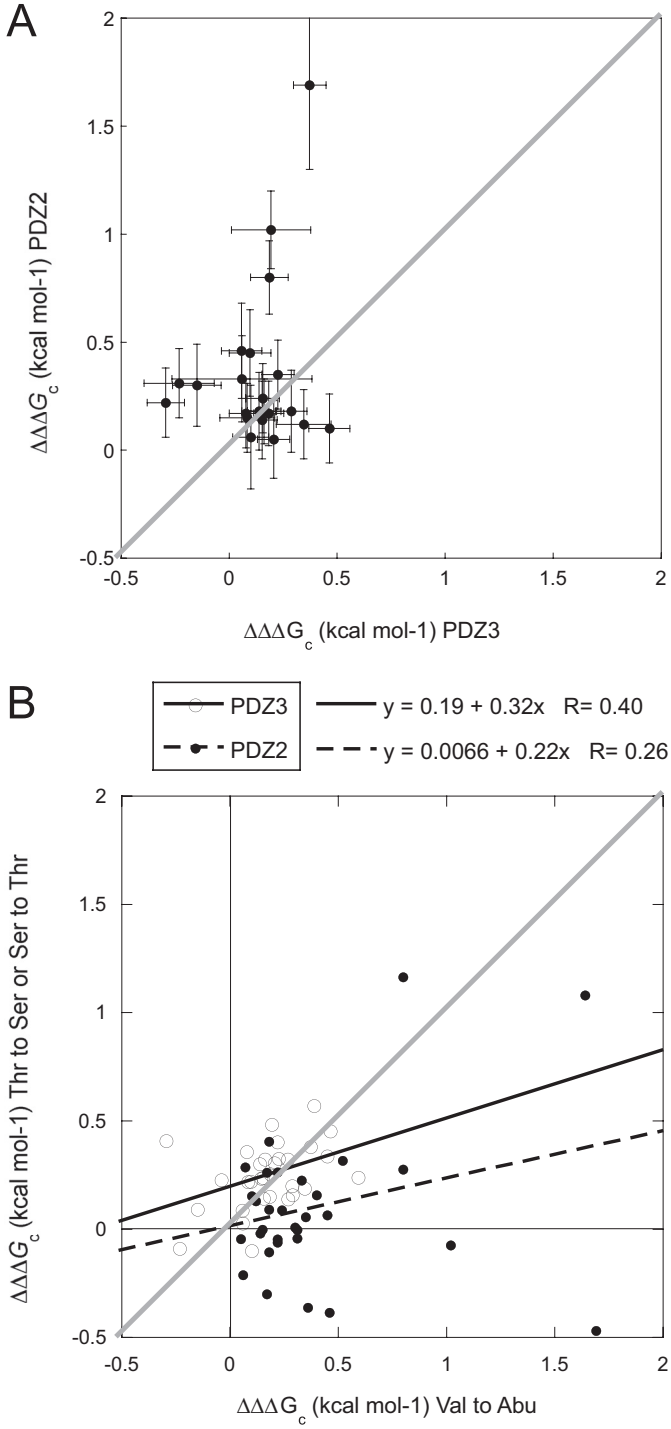

FIGURE 4. *A*, coupling energies for PDZ2 *versus* those for the corresponding topological position in PDZ3, for the $Val^0 \rightarrow$ Abu mutation in the peptide. *B*, coupling energies for the C-terminal Val *versus* coupling energies for the Ser/Thr at position −2 of the peptide. The *horizontal* and *vertical lines* divide positive coupling energies from negative. The *thick gray lines* correspond to a slope of 1, which would be a perfect correlation.

energies for corresponding topological positions in the two native structures are plotted against each other (Fig. 4*A*). This is in stark contrast to folding Φ values, which correlate strongly for these two protein domains, suggesting a common folding mechanism (24). It thus appears that whereas folding pathways are conserved within the PDZ domain family, their functional energetic networks are more variable. In other words, although native topology sculpts the free energy landscape for folding in a rather common way, the amino acid composition plays a key role in allostery.

The sequence dependence of the allosteric pathways is illustrated by some of the positions with the highest coupling energy. For example, Val^{29} in PDZ2 couples strongly to Val^0 in the peptide, whereas the corresponding Ile^{327} in PDZ3 has a $\Delta\Delta\Delta G_C < 0.2$ kcal mol^{-1}. However, the same Ile^{327} in PDZ3 has one of the highest $\Delta\Delta\Delta G_C$ for Thr^{-2} (0.48 kcal mol^{-1}), whereas $\Delta\Delta\Delta G_C$ between the corresponding positions in PDZ2 is close to zero.

Nevertheless, a few positions display high coupling for both PDZ domains, for example position Leu^{25} in PDZ2 and Leu^{323} in PDZ3. This position was predicted as an allosteric network residue (18, 19, 28), and indeed, it couples with the C-terminal Val of the peptide (Table 1 and supplemental Fig. S1).

Similarly, a subset of residues couples to both the Val^0 and Ser/Thr^{-2} positions (*e.g.* Leu^{25}, Thr^{35}, Val^{65}, Leu^{73}, Leu^{85}, and

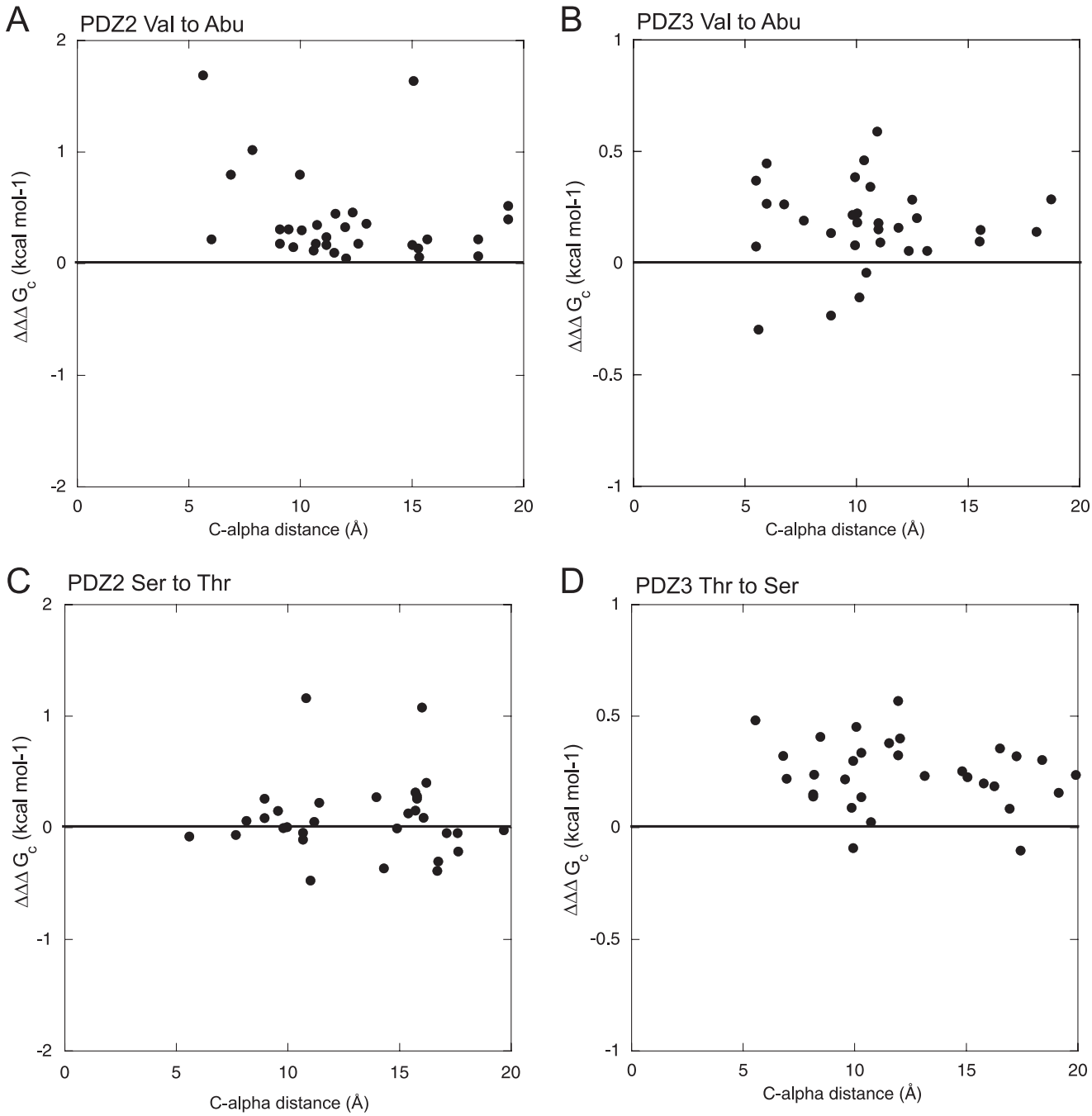

FIGURE 5. **Coupling energies for PDZ2 (*A* and *C*) and PDZ3 (*B* and *D*) plotted *versus* C_α–C_α distance between the two sites of mutation.** The *horizontal lines* divide positive coupling energies from negative ones.

Leu^{94} in PDZ2 and Leu^{323}, Phe^{340}, Ile^{341}, and Lys^{380} in PDZ3; Table 1). Yet, no general correlation was observed between residues that couple to position Val^{0} of the peptide and residues that couple to position Ser/Thr^{-2} (Fig. 4*B*). Interestingly, for PDZ2, Leu^{25}, Val^{65}, and Leu^{94} have positive coupling energies with position Val^{0} but negative coupling energies with Ser^{-2}.

DISCUSSION

The crowded and heterogeneous cellular environment demands the interactions between its numerous components to be highly specific, to avoid potentially harmful side reactions. In fact, it is somewhat surprising that the cell employs effectively a limited number of distinct protein-protein interaction domains (2), such as the PDZ family, which is involved in the control of quite different cellular events. Despite the fact that this protein family displays a simple topology and a rather conserved binding pocket, the distinct roles of the various PDZ domains in cellular signaling require selectivity. Until recently, PDZ domains were grouped into three classes, depending on their consensus target sequence. A detailed analysis of the PDZ domains in the mouse proteome, however, suggested that separation into discrete classes might not be justified; instead, PDZ domains are evenly distributed throughout selectivity space,

suggesting they have been optimized to increase the repertoire of specificities but maintaining minimal cross-reactivity (4). In this perspective, it is of interest to discuss the general significance of the coupling free energies that we obtained for the different peptides, which seem to imply a mechanism of optimization involving the whole domain rather than the binding site only.

The experimental data reported above show that all the mutations of PDZ2 that we have explored lead to a positive coupling energy for the Val^0 to Abu substitution (Fig. 3), although the absolute values are quite variable. That is, mutation of any core residue in the protein will affect the binding of the peptide such that the effect of deleting a methyl group from the native C-terminal Val^0 of the peptide will be smaller in the mutant as compared with the wild type PDZ2. This intriguing finding suggests that the entire PDZ scaffold has been under selective pressure to optimize the binding of the ligand by intradomain allosteric coupling. Sequence variation in the protein domain is therefore not neutral to peptide binding, strongly indicating that selectivity by the domain is not solely determined by the subset of residues directly involved in ligand binding. Additional support for this conclusion comes from the effect of the Ser → Thr mutation in the peptide for the binding with PDZ2. In this case, the rather even distribution of negative coupling energies (Figs. 3 and Fig. 5*C*) suggests that this position is not optimized for Ser relative to Thr (42), as also indicated by the nearly identical K_D for wild type PDZ2 and the two peptide variants (supplemental Table S1).

Likewise, PDZ3 also displays mainly positive coupling energies. However, in this case, deletion of the methyl group of the Thr is more detrimental to the binding than the Val → Abu substitution (supplemental Table S1). Indeed, the Thr^{-2} → Ser mutation displayed an even higher number of positive coupling energies than Val^0 → Abu (Figs. 3 and 5). In conclusion, the high number of positive $\Delta\Delta\Delta G_C$ values suggests that PDZ3 is optimized for both Thr at position −2 and Val at position 0.

The coupling free energy between pairs of residues was found previously to depend strongly on the distance between the two residues (38, 43). A similar conclusion was made for PDZ3 when coupling to the evolutionarily conserved residue His^{372} was assessed, although with a limited data set (20). This behavior is not surprising because neighboring residues are expected to interact more effectively than distant ones. However, in the case of both PDZ2 and PDZ3, we did not observe a clear trend in the dependence of the coupling energies on distance, for neither of the peptide residues Val^0 and Ser/Thr^{-2} (Fig. 5). This result corroborates the hypothesis that sequence rather than topology governs the observed inter-residue communication. It is also in line with the observed distinct segments of higher coupling free energies for PDZ2 and PDZ3, respectively (Fig. 2).

In summary, we identified by double mutant cycles and binding kinetics, energetically coupled residues in two different homologous protein-ligand complexes. Our experimental results confirm the existence of long range networks suggested to exist in different proteins (17) and in particular predicted in the PDZ domain family (*e.g.* Refs. 16, 18, 19, 27, 28). Several predictions capture overall features and even details (18, 19) of our present results. Our data may thus be used to benchmark the extensive available and forthcoming theoretical work on the PDZ system.

We suggest that distinct networks will be found in most proteins, but the extent of functional selectivity by a particular allosteric network will have to be assessed in each case. Nevertheless, these coupling networks exist and may be of advantage for a specific function and fixed by natural selection. Overall, our data suggest an attractive scenario whereby a subtle allosteric mechanism reduces cross-reactivity by optimizing ligand recognition using the whole domain. The mechanism involves distinct pathways that modulate specificities in different PDZ domains.

REFERENCES

1. Bhattacharyya, R. P., Reményi, A., Yeh, B. J., and Lim, W. A. (2006) *Annu. Rev. Biochem.* **75,** 655–680
2. Kaneko, T., Sidhu, S. S., and Li, S. S. (2011) *Trends Biochem. Sci.* **36,** 183–190
3. te Velthuis, A. J., Sakalis, P. A., Fowler, D. A., and Bagowski, C. P. (2011) *PLoS ONE* **6,** e16047
4. Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A., and MacBeath, G. (2007) *Science* **317,** 364–369
5. Harris, B. Z., and Lim, W. A. (2001) *J. Cell Sci.* **114,** 3219–3231
6. Feng, W., and Zhang, M. (2009) *Nat. Rev. Neurosci.* **10,** 87–99
7. Goodey, N. M., and Benkovic, S. J. (2008) *Nat. Chem. Biol.* **4,** 474–482
8. Monod, J., Wyman, J., and Changeux, J. P. (1965) *J. Mol. Biol.* **12,** 88–118
9. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) *Nature* **438,** 117–121
10. Cooper, A., and Dryden, D. T. (1984) *Eur. Biophys. J.* **11,** 103–109
11. Akke, M., Bruschweiler, R., and Palmer, A. G., 3rd (1993) *J. Am. Chem. Soc.* **115,** 9832–9833
12. Pan, H., Lee, J. C., and Hilser, V. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97,** 12020–12025
13. Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) *Nat. Struct. Mol. Biol.* **13,** 831–838
14. Tsai, C. J., del Sol, A., and Nussinov, R. (2008) *J. Mol. Biol.* **378,** 1–11
15. Tzeng, S. R., and Kalodimos, C. G. (2009) *Nature* **462,** 368–372
16. Lockless, S. W., and Ranganathan, R. (1999) *Science* **286,** 295–299
17. Süel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003) *Nat. Struct. Biol.* **10,** 59–69
18. Kong, Y., and Karplus, M. (2009) *Proteins* **74,** 145–154
19. Dhulesia, A., Gsponer, J., and Vendruscolo, M. (2008) *J. Am. Chem. Soc.* **130,** 8931–8939
20. Chi, C. N., Elfström, L., Shi, Y., Snäll, T., Engström, A., and Jemth, P. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4679–4684
21. Fuentes, E. J., Gilmore, S. A., Mauldin, R. V., and Lee, A. L. (2006) *J. Mol. Biol.* **364,** 337–351
22. McCallister, E. L., Alm, E., and Baker, D. (2000) *Nat. Struct. Biol.* **7,** 669–673
23. Horovitz, A., and Fersht, A. R. (1990) *J. Mol. Biol.* **214,** 613–617
24. Calosci, N., Chi, C. N., Richter, B., Camilloni, C., Engström, A., Eklund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M., and Jemth, P. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 19241–19246
25. Gianni, S., Walma, T., Arcovito, A., Calosci, N., Bellelli, A., Engström, A., Travaglini-Allocatelli, C., Brunori, M., Jemth, P., and Vuister, G. W. (2006) *Structure* **14,** 1801–1809
26. Jemth, P., and Gianni, S. (2007) *Biochemistry* **46,** 8701–8708
27. Ho, B. K., and Agard, D. A. (2010) *Protein Sci.* **19,** 398–411
28. Liu, Z., Chen, J., and Thirumalai, D. (2009) *Proteins* **77,** 823–831
29. Gianni, S., Geierhaas, C. D., Calosci, N., Jemth, P., Vuister, G. W., Travaglini-Allocatelli, C., Vendruscolo, M., and Brunori, M. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 128–133
30. Gianni, S., Calosci, N., Aelen, J. M., Vuister, G. W., Brunori, M., and Travaglini-Allocatelli, C. (2005) *Prot. Eng. Des. Sel.* **18,** 389–395
31. Gianni, S., Engström, A., Larsson, M., Calosci, N., Malatesta, F., Eklund, L.,

Ngang, C. C., Travaglini-Allocatelli, C., and Jemth, P. (2005) *J. Biol. Chem.* **280,** 34805–34812
32. Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996) *Cell* **85,** 1067–1076
33. Chi, C. N., Bach, A., Engström, A., Wang, H., Strømgaard, K., Gianni, S., and Jemth, P. (2009) *Biochemistry* **48,** 7089–7097
34. Niethammer, M., Valtschanoff, J. G., Kapoor, T. M., Allison, D. W., Weinberg, R. J., Craig, A. M., and Sheng, M. (1998) *Neuron* **20,** 693–707
35. Malatesta, F. (2005) *Biophys. Chem.* **116,** 251–256
36. Carter, P. J., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1984) *Cell* **38,** 835–840
37. Horovitz, A., and Fersht, A. R. (1992) *J. Mol. Biol.* **224,** 733–740
38. Schreiber, G., and Fersht, A. R. (1995) *J. Mol. Biol.* **248,** 478–486
39. Schreiber, G., and Fersht, A. R. (1996) *Nat. Struct. Biol.* **3,** 427–431
40. Fuentes, E. J., Der, C. J., and Lee, A. L. (2004) *J. Mol. Biol.* **335,** 1105–1115
41. Fersht, A. R., and Sato, S. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101,** 7976–7981
42. Kozlov, G., Banville, D., Gehring, K., and Ekiel, I. (2002) *J. Mol. Biol.* **320,** 813–820
43. Green, S. M., and Shortle, D. (1993) *Biochemistry* **32,** 10131–10139
44. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, Schrodinger, LLC, Portland, OR