J|A|C|S
ARTICLES

Published on Web 06/18/2008

Mapping of Two Networks of Residues That Exhibit Structural and Dynamical Changes upon Binding in a PDZ Domain Protein

Anne Dhulesia, Joerg Gsponer, and Michele Vendruscolo*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Received July 12, 2007; E-mail: mv245@cam.ac.uk

Abstract: We describe the changes in structure and dynamics that occur in the second PDZ domain of human tyrosine phosphatase 1E upon binding the small peptide RA-GEF2 by an analysis of NMR data based on their use as ensemble-averaged restraints in molecular dynamics simulations. This approach reveals the presence of two interconnected networks of residues, the first exhibiting structural changes and the second dynamical changes upon binding, and it provides a detailed mapping of the regions of increased and decreased mobility upon binding. Analysis of the dynamical properties of the residues in these networks reveals that conformational changes are transmitted through pathways of coupled side-chain reorientations. These results illustrate how the strategy we described, in which NMR data are used in combination with molecular dynamics simulations, can be used to characterize in detail the complex organization of the changes in structure and dynamics that take place in proteins upon binding.

Introduction

Proteins in solution exhibit a vast range of motions, spanning time scales from picoseconds to milliseconds and more and length scales from one to several angstroms. These motions play a central role in the ability of proteins to function as enzymes and to take part in regulatory processes.[1–8] Although the structural transitions associated with these activities may be prompted by localized events, such as the binding of small ligand molecules, the effects of these interactions are transmitted across the entire protein, and they often modulate the binding affinity at distant surfaces.[6,9] Such transitions take place by highly organized movements, and it is a long-standing goal in structural biology to describe this type of structural specificity with high accuracy.

A view is gaining ground according to which these large-scale motions are intrinsic properties of protein molecules.[1,6,10–13] According to this view, the structural and dynamical changes that take place during allosteric and enzymatic processes can be described as a redistribution of the statistical weights between pre-existing populations that dynamically interchange also in the absence of interaction partners. The existence of population shifts has been demonstrated by a series of recent experimental studies that have thus suggested that allostery and enzyme catalysis take place by exploiting the intrinsic dynamics of protein molecules.[1,10,13]

In this work we characterize the specific changes that take place in the second PDZ domain of human tyrosine phosphatase 1E (henceforth referred to as PDZ, Figure 1) upon binding the small peptide RA-GEF2 by following an approach that involves the determination of the dynamical changes on the nanosecond time scale associated with this process.[14] PDZ domains are involved in organizing the assembly and the cellular distribution of the protein complexes that mediate the process of synaptic communication.[15] These complexes are assembled and disassembled in a tightly regulated manner at specific cellular locations, most often near cell surfaces, to perform specialized tasks such as the regulation of cell junctions, signal transduction, protein trafficking, and maintenance of cell polarity.

The approach that we adopt in this study, to identify the conformational changes that take place in PDZ upon binding, is based on the use of NMR measurements as ensemble-averaged restraints in molecular dynamics simulations. This strategy enables the structure and dynamics of proteins to be

(1) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638–1642.
(2) Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman: New York, 1999.
(3) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.
(4) Hammes-Schiffer, S.; Benkovic, S. J. *Annu. Rev. Biochem.* **2006**, *75*, 519–541.
(5) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
(6) Kern, D.; Zuiderweg, E. R. P. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748–757.
(7) Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. *Methods* **2004**, *34*, 51–64.
(8) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224–228.
(9) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88–118.
(10) Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. *Nature* **2005**, *438*, 117–121.
(11) Kumar, S.; Ma, B. Y.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10–19.
(12) Vendruscolo, M.; Dobson, C. M. *Science* **2006**, *313*, 1586–1587.
(13) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. *Science* **2001**, *291*, 2429–2433.
(14) Fuentes, E. J.; Der, C. J.; Lee, A. L. *J. Mol. Biol.* **2004**, *335*, 1105–1115.
(15) Kim, E. J.; Sheng, M. *Nat. Rev. Neurosci.* **2004**, *5*, 771–781.

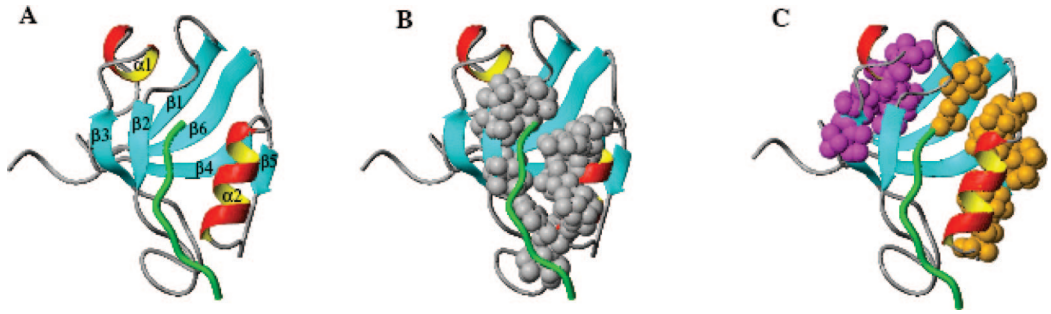

Figure 1. (A) Structure of PDZ in complex with the RA-GEF2 peptide (green). (B) Residues forming the binding site (gray): L18, G19, I20, V22, and V26 in strand β2 and H71, K72, V75, and L78 in helix α2. (C) Distal surfaces as determined in the present work: DS1 (V61, V64, L66, A69, T81, L87, and L88, shown in orange) and DS2 (A39, V40, A45, A46, and I52, shown in purple).

simultaneously characterized, resulting in high-resolution conformational ensembles.[16–18] In the present work we use NMR relaxation data (S^2 order parameters)[19–24] as restraints in molecular dynamics simulations[16–18] to determine both the free and bound states of PDZ. Since the transitions between the ligand-free and bound states take place on the millisecond time scale,[10] we study here only the inherent dynamics within the free and the bound states, not those during the transitions between these two free energy wells. We thus obtain a description of the fluctuations on the nanosecond time scale in both the free- and bound-state free energy wells, as well as information about the structural and dynamical differences between the two states by comparing the corresponding ensembles of structures resulting from the simulations. The results of this type of analysis reveal the existence of two intertwined networks of residues that change, respectively, their structure and their dynamics upon binding and that span across the whole protein structure.

Methods

Computational Strategy. In the dynamic ensemble refinement (DER) method, the structural information provided by NOE-derived distances is complemented by information about the dynamics contributed by the S^2 order parameters.[17] The simultaneous enforcement of both observables as ensemble-averaged restraints in molecular dynamics simulations enables ensembles of structures to be determined that provide an accurate representation of the thermal fluctuations of proteins on the nanosecond time scale.[16–18] The minimal underrestraining minimal-overrestraining (MUMO) method is a particular implementation of the DER approach in which different NMR observables are averaged over different number of replicas.[18] This procedure enables us to minimize simultaneously the effects of overrestraining and underrestraining for NMR observables that report on different time scales.[18] The MUMO approach has been extensively validated in the case of ubiquitin, where a large set of independent measurements is available for cross-validation.[18] We should note, however that particular care should be taken in validating the results for systems for which much less detailed experimental information is available.

NMR Data. As structural restraints in the molecular dynamics simulations of the free and bound states we used 1192 and 1093 NOE-derived distances, respectively, determined by Ekiel and co-workers.[25,26] In addition, we used 70 backbone S^2 order parameters for the free state and 73 for the bound state, as well as 49 methyl S^2 order parameters for the free state and 47 for the bound state, which were determined by Lee and co-workers.[14] The availability of experimentally determined methyl S^2 order parameters provides invaluable information about the changes of structure and dynamics upon binding. However, due to the relatively small number of methyl groups in a protein, this information is intrinsically sparse in nature, and its incorporation in the molecular dynamics procedure that we adopted in this work enables significant additional information to be obtained about the motions of side chains in both the free and bound states. For this reason, the ensembles of structures that we determined provide a more complete description of the structural and dynamical changes in the side-chain rotameric distributions upon complex formation than that offered by comparison of individual NMR or X-ray structures of free and bound states, or that provided by considering only the side chains for which S^2 order parameters are available.

Initial Structures. We used the structures 3PDZ for the free state[25] and 1D5G for the bound state[26] as starting points for the molecular dynamics simulations described in this work.

Restrained Molecular Dynamics Simulations. We performed a 2 ns simulation for 16 replicas using NOE and S^2 data,[14,26,25] as restraints,[16–18] in addition to the CHARMM22 force field.[27] Simulations were carried out in a 4 Å shell of TIP3 water molecules[28] and a boundary potential was used to prevent water molecules from escaping.[29] All calculations used an atom-based truncation scheme with a list cutoff of 14 Å, a nonbond cutoff of 12 Å, and the Lennard-Jones smoothing function initiated at 10 Å. Electrostatic and Lennard-Jones interactions were force-switched. We used a 2 fs integration time step and covalent bonds involving hydrogen atoms were constrained with SHAKE.[30] All simulations were performed at 300 K; initial velocities were randomly assigned from a Maxwell−Boltzmann distribution at 300 K with a different random seed for each replica. Ensemble-averaged simulations were implemented by use of MPI, as described previously.[16]

The energy function used has the form

$$E_{\text{tot}} = E_{\text{CHARMM}} + E_{\text{NOE}} + E_{S^2} \quad (1)$$

in which E_{CHARMM} is the CHARMM22 force field[27] and E_{NOE} and E_{S^2} are the energies of the NOE and S^2 ensemble-averaged restraints, respectively. NOE distances and S^2 order parameters are calculated over the replicas at each time step (2 fs) of the dynamics and allow the determination of the energy penalty function:

(16) Best, R. B.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 8090–8091.
(17) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.
(18) Richter, B.; Gsponer, J.; Varnai, P.; Salvatella, X.; Vendruscolo, M. *J. Biomol. NMR* **2007**, *37*, 117–135.
(19) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959–8966.
(20) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624–1671.
(21) Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nat. Struct. Biol.* **2000**, *7*, 72–77.
(22) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
(23) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623–3640.
(24) Yang, D. W.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369–382.
(25) Kozlov, G.; Gehring, K.; Ekiel, I. *Biochemistry* **2000**, *39*, 2572–2580.
(26) Kozlov, G.; Banville, D.; Gehring, K.; Ekiel, I. *J. Mol. Biol.* **2002**, *320*, 813–820.

$$E_X[\rho(t)] = \begin{cases} (1/2)\alpha_X N_{\text{rep}}[\rho(t) - \rho_0(t)]^2 & \text{if } \rho(t) > \rho_0(t) \\ 0 & \text{if } \rho(t) \le \rho_0(t) \end{cases} \quad (2)$$

where N_{rep} is the number of replicas, X corresponds to either NOE or S^2 restraints, and α_X is the force constant associated with the restraint. α_X was fixed to 4×10^6 and 6.5×10^7 for NOE and S^2, respectively, and N_{rep} is set to 16.[17,18] $\rho_0(t)$ is defined as

$$\rho_0(t) = \min_{0 \le \tau \le t} \rho(\tau) \quad (3)$$

For NOE restraints

$$\rho_{\text{NOE}}(t) = \frac{1}{N_{\text{NOE}}} \sum_{i,j} \left(d_{ij}^{\text{exp}} - d_{ij}^{\text{sim}}\right)^2 \quad (4)$$

where the simulated distance is allowed to vary freely between the lower and upper bounds of the NMR restraints set and N_{NOE} is the number of NOE restraints. The simulated distance is calculated over the replicas with a $1/r^3$ average:[17,18]

$$d_{ij}^{\text{sim}} = \left[\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} r_{ij,k}^{-3}\right]^{-1/3} \quad (5)$$

For S^2 restraints:

$$\rho_{S^2}(t) = \frac{1}{N_{S^2}} \sum_{i,j} (S^{2,\text{exp}} - S^{2,\text{calc}})^2 \quad (6)$$

where the simulated S^2 values are calculated without assuming any motional model. For $S^{2,ij}$ relative to the motion of the bond between atoms i and j:

$$S^{2,ij} = \frac{3}{2}[\langle \hat{x}_{ij}^2 \rangle^2 + \langle \hat{y}_{ij}^2 \rangle^2 + \langle \hat{z}_{ij}^2 \rangle^2 + \langle \hat{x}_{ij}\hat{y}_{ij} \rangle^2 + \langle \hat{y}_{ij}\hat{z}_{ij} \rangle^2 + \langle \hat{z}_{ij}\hat{x}_{ij} \rangle^2] - \frac{1}{2} \quad (7)$$

where broken brackets describe an ensemble average over the 16 replicas and $\hat{x}$, $\hat{y}$, and $\hat{z}$ represent the normalized internuclear coordinates.

Sets of 16 structures were extracted every 200 ps and were then further refined for 50 ps with $^3J_{\text{HNH}\alpha}$ restraints[25] calculated with a Karplus relation,[31,32] with standard coefficients.[25,26] In the case of ^3J restraints we use the formula

$$\rho_{^3J}(t) = \frac{1}{N_{^3J}} \sum_{i,j} \left(^3J^{\text{exp}} - {}^3J^{\text{calc}}\right)^2 \quad (8)$$

At the end of this procedure, we obtained two ensembles of 160 structures for the free and bound states, respectively, which were used for the analysis.

Analysis of the S^2 Values. In order to estimate the errors in the $S^{2,\text{calc}}$ values, we considered 10 subensembles of 16 structures, calculated $S^{2,\text{calc}}$ values for these 10 subensembles, and then estimated the standard deviations of the $S^{2,\text{calc}}$ values. We found average values for the error of 0.017 and 0.019 for the free and bound states, respectively.

Structural Analysis. From the ensembles of structures that we determined, we calculated the average Cα−Cα distances and their associated standard deviations as an estimate of the distance fluctuations.

From the atomic coordinates we also determined the covariances of the displacements of atom pairs. We performed this covariance analysis with CHARMM for each pair of Cα atoms, as well as for side chains; in the latter case, the covariance is averaged over individual side chain atom pairs. The covariance is defined as

$$C_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle = \langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle \quad (9)$$

where r_i and r_j are the position vectors of atoms i and j, respectively, and averages are taken over the generated ensembles.

Structural Similarity. We determined the RMSD between all pairs of structures, in both free and bound states and between structures belonging to the two different ensembles. The RMSD were computed on structures already overlaid on the lowest energy conformer of the NMR bound ensemble. Highly mobile residues of the C and N termini (residues P1 and K2, as well as P95 and T96) were removed in the RMSD calculations.

Clustering of Structures. We performed a Cα cluster analysis of the free- and bound-state ensembles based on a structural similarity algorithm of Daura et al.[33] and of Ferrara and Caflisch[34] The Cα RMSD was computed for all pairs of structures and allowed the determination of the number of neighbors for each structure: two structures are defined as neighbors when their Cα RMSD is less than a given threshold. The center of the first cluster is identified as the structure having the highest number of neighbors, and its associated cluster is composed of its neighbors. A similar analysis is done on the nonneighboring structures until each structure is assigned to one cluster. This clustering procedure allows the main conformations and their relative populations to be determined. RMSDs have been computed based on Cα coordinates, with different thresholds of 1.0, 1.5 and 2 Å.

Comparison of Dihedral Angle Distributions. We computed dihedral angle distributions with bins of 5° and then quantified the differences between the distributions of the free and bound states with the χ^2 factor.[22] A value $\chi^2 = 0$ indicates identical distributions, while χ^2 larger than unity indicates significantly different distributions; the maximum value of a χ^2 factor is 2. For dihedral angles where the last atom is a hydrogen of a methyl group, the 3-fold degeneracy was taken into account by considering the smallest χ^2 factor among the three that were computed with offsets of 0°, 120°, and 240°. A similar procedure was used when the last atom had a 2-fold degeneracy (with offsets of 0° and 180°); this is typically the case for oxygen atoms at carboxylic ends.

We tested the convergence of the results by computing the χ^2 factor between angle distributions made out of 11 200 and 160 structures (this corresponds to different time intervals between the extraction of structures during the dynamics). An average of $\chi^2 = 0.04$ suggests that the sampling over our ensembles of 160 structures is both efficient and sufficient.

Computation of the Structural Network. We determined the 20 residues with the largest structural change by considering the sum of the χ^2 factors summed for the dihedral angles ϕ, ψ, $\chi 1$, and $\chi 2$. We then considered the energetics of these residues. The distribution of the interaction energies of residue pairs exhibits a clear tail for negative energies smaller than −0.1 kcal/mol, corresponding to 12% of the residue pairs. The 20 residues exhibiting the largest structural changes interact with at least one partner with an energy lower than −0.1 kcal/mol, and 23% of all possible interactions are smaller than −0.1 kcal/mol, in comparison with 12% for the entire protein. These results show that the residues of the network interact more tightly with one another than randomly selected ones.

Determination of Structural Pathways. Structural pathways were determined by requiring that consecutive residues in the pathways interact with an energy lower than −0.1 kcal/mol. We determined 290 and 185 pathways containing four and five residues, respectively. The five pathways found to contain six residues were not considered in the following analysis. We determined how often

(27) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(29) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050–9063.
(30) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(31) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11–15.
(32) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *J. Biomol. NMR* **2005**, *32*, 273–280.
(33) Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 269–280.
(34) Ferrara, P.; Caflisch, A. *J. Mol. Biol.* **2001**, *306*, 837–850.

Table 1. Comparison between Experimental and Calculated NOE, S^2, and ^3J Values[a]

	free state	bound state
NOE (angstroms)	0.091 ± 0.002	0.069 ± 0.006
^3J (Hz)	0.11 ± 0.01	0.12 ± 0.020
backbone S^2	0.042 ± 0.00	0.035 ± 0.00
side-chain S^2	0.035 ± 0.00	0.035 ± 0.00

[a] Values correspond to root mean square deviations (RMSDs) between experimental and calculated values; standard deviations were estimated from the analysis of the 10 subensembles of 16 structures determined in this work.

a given residue occurred at any given position in the pathway, in the center or at the end points. We then focused on the residues in the central regions of the pathways and determined how often they occurred in different types of pathways: (i) for all possible pathways, (ii) for pathways beginning from residues L18 and L78 of the binding site, and (iii) for pathways beginning and ending at a solvent-exposed residue (with a solvent-accessible surface area larger than 50 Å^2).

Results and Discussion

Determination of Ensembles of Conformations. We used molecular dynamics simulations restrained through the use of NOE, S^2, and ^3J data[14,25,26] (see Methods) to determine two structural ensembles representing the free and bound states of PDZ. Interproton NOE distances, S^2 order parameters of backbone amide groups and side-chain methyl groups, and ^3J coupling constants calculated over ensembles of 160 conformations are reported in Table 1 and Figure 2A,B; statistical errors were estimated by considering the 10 ensembles of 16 members each (see Methods).

Characterization of Distal Regions. The changes (ΔS^2) in the S^2 order parameters upon binding allow a comparison between the dynamics on the nanosecond time scale of the free and bound states to be made. In this work, in addition to the experimentally measured S^2 order parameters, $S^{2,\mathrm{exp}}$, we also consider the S^2 order parameters, $S^{2,\mathrm{calc}}$, calculated from the ensembles that we determined. Significant changes $\Delta S^{2,\mathrm{exp}}$ in the side-chain $S^{2,\mathrm{exp}}$ experimental order parameters are observed upon binding.[14,35] As the experimental errors $\delta S^{2,\mathrm{exp}}$ on the $S^{2,\mathrm{exp}}$ values in both the free and bound states were comparable to $\Delta S^{2,\mathrm{exp}}$, a change $\Delta S^{2,\mathrm{exp}}$ was considered to be significant if it was larger than $\delta S^{2,\mathrm{exp}}_{\mathrm{free}} + \delta S^{2,\mathrm{exp}}_{\mathrm{bound}}$. A similar criterion was adopted for correlation times.[14,35] These criteria provided a list of 14 residues undergoing significant changes in side-chain dynamics upon binding.[14] The peptide binding site region was found to be connected dynamically to two distant regions in the structure (Figure 1). These regions were identified as distal surface 1 (DS1; residues V61, V64, L66, A69, T81, and V85) and distal surface 2 (DS2; residues A39 and I40).[14]

We extended this analysis by exploiting the possibility of estimating the S^2 order parameters for the backbone amide and the side-chain methyl groups in the protein from the structural ensembles that we calculated (Figure 2C,D,F). In the calculations, the force constant associated with the S^2 order parameter restraints (α_{S^2}) was chosen to restrict the $S^{2,\mathrm{exp}}$ values within the interval $S^{2,\mathrm{exp}} \pm \delta S^{2,\mathrm{exp}}$ (see Methods). A statistical error $\delta S^{2,\mathrm{calc}}$ was calculated for each methyl-bearing group (see Methods). Consistent with the approach used by Fuentes et al.[14] for the analysis of experimental changes $\Delta S^{2,\mathrm{exp}}$, we determined a list of residues exhibiting significant changes in side-chain $S^{2,\mathrm{calc}}$ order parameter upon binding; that is, $\Delta S^{2,\mathrm{calc}} > \delta S^{2,\mathrm{calc}}_{\mathrm{free}} + \delta S^{2,\mathrm{calc}}_{\mathrm{bound}}$. This criterion provided six residues with significant $\Delta S^{2,\mathrm{calc}}$ values (T23, A45, A46, I52, V58, and L87) (Figure 2F) in addition to those identified experimentally (Figure 2G). According to their location in the structure, residues were added to distal surfaces 1 (L87) and 2 (A45, A46, and I52) (Figure 1). In particular, our simulations show that residues V58 and L87, which are completely buried both in the free and in the bound states, change in side-chain $S^{2,\mathrm{calc}}$, suggesting that the hydrophobic core of the protein is significantly affected by the binding of the RA-GEF2 peptide.

Description of Structural and Dynamical Changes upon Binding. The ensembles of structures that we determined enable the changes in structure and dynamics in the second PDZ domain of human tyrosine phosphatase 1E upon binding the small peptide RA-GEF2 to be characterized in detail. We determined the structural fluctuations by computation of the standard deviations of the Cα−Cα distances within the ensembles of conformations (Figure 3C) (see Methods). Residues showing large changes in fluctuations are mostly localized in specific regions, including the binding site (residues in the region between L18 and V26), strand β2, DS1 (residues in the region between V61 and A69), DS2 (residues in the regions of V40, A45, and I52), and the region of residue V30, which is in the loop between strands β2 and β3. These results indicate that the binding site and DS2 become more rigid, while DS1 and the region of residue V30 become more flexible upon binding (Figure 3C).

We also calculated the side-chain covariance for atomic displacements for the free and the bound states, respectively (Figure 3A,B; see Methods). The covariance maps reveal complex patterns of correlated displacements that extend across the protein structure. A comparison between the two states indicates the presence of a loss or gain in correlated atomic displacements. In the free state, the protein exhibits two regions in anticorrelated motion (yellow boxes in Figure 3A); the first region is composed of β2 (belonging to the binding site) and DS2, whereas the second region comprises helix α2 (the other part of the binding site) and DS1. This anticorrelation is lost upon binding, as indicated by the covariance patterns that are more spread throughout the protein (Figure 3B). Within the second region itself, helix α2 and DS1 also clearly show correlated motions (yellow and orange) in the free state, whereas they exhibit anticorrelated motions (blue) in the bound state. Together with the results on the fluctuations (Figure 3C), this analysis indicates that upon ligand binding the motion of DS1 increases and becomes decoupled from that of helix α2 in the binding site. In contrast, DS2 becomes more rigid and more tightly linked with strand β2 in the binding site, a result consistent with the findings of Milev et al.[36] and Piserchio et al.[37] This analysis also shows that the two parts of the binding site, helix α2 and strand β2, change in the way their motions are correlated. Interestingly, in a recent paper De Los Rios et al.[38] reported that the residues dynamically affected by the binding have high mobility in a single normal mode.

A Cα cluster analysis was carried out to probe the structural heterogeneity in the backbone in the free and bound ensembles (see Methods). The larger number of clusters found in the free

(35) Fuentes, E. J.; Gilmore, S. A.; Mauldin, R. V.; Lee, A. L. *J. Mol. Biol.* **2006**, *364*, 337–351.
(36) Milev, S.; Bjelic, S.; Georgiev, O.; Jelesarov, I. *Biochemistry* **2007**, *46*, 1064–1078.
(37) Piserchio, A.; Fellows, A.; Madden, D. R.; Mierke, D. F. *Biochemistry* **2005**, *44*, 16158–16166.
(38) De Los Rios, P.; Cecconi, F.; Pretre, A.; Dietler, G.; Michielin, O.; Piazza, F.; Juanico, B. *Biophys. J.* **2005**, *89*, 14–21.

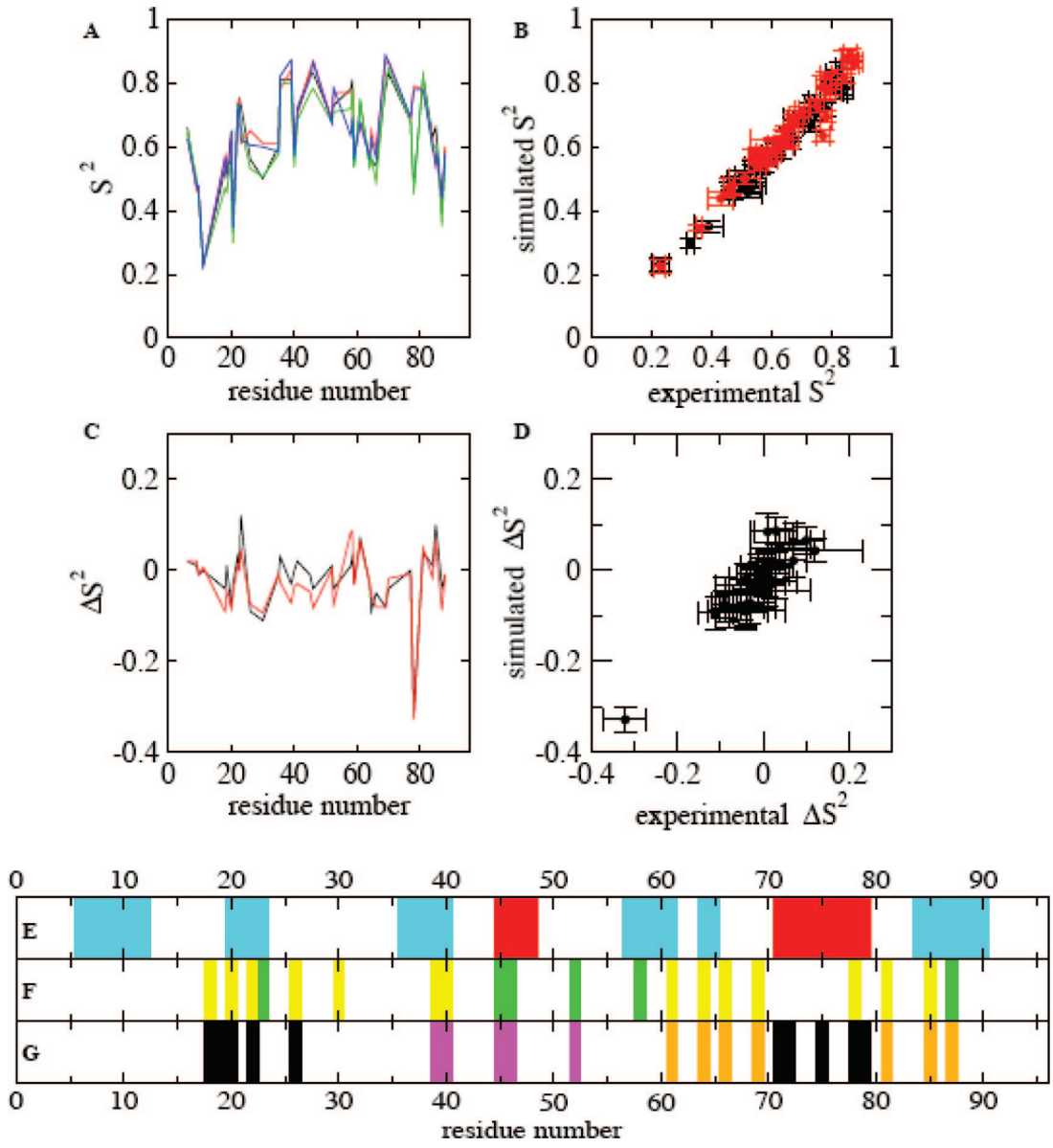

Figure 2. Comparison between calculated and experimental side-chain S^2 order parameters and schematic representation of the significant changes in the side-chain S^2 order parameters upon binding. (A) S^2 order parameters as a function of the residue number: (black) experimental values in the free state, (red) experimental values in the bound state, (green) calculated values in the free state, (blue) calculated values in the bound state. (B) Comparison of S^2 order parameters: (black) experimental and calculated values in the free state, (red) experimental and calculated values in the bound state. (C) Changes (ΔS^2) in the S^2 order parameters upon binding as a function of the residue number: (black) experimental values, (red) simulated values. (D) Correlation between experimental and calculated ΔS^2 values. (E) Schematic representation of the secondary structure elements: β-strands (blue) and α-helices (red). (F) Residues with a large change in side-chain S^2 order parameter upon binding: residues that change according to both experiments and simulations (yellow) and residues that change according to simulations only (green) are indicated. (G) Important regions resulting from the analysis of Fuentes et al.[14] of the S^2 order parameters. The binding site is shown in black; the extended DS1 and DS2 are shown in orange and purple, respectively.

state compared to the bound state suggests that the backbone on the whole becomes more rigid upon binding (Table 3). However, while the binding site and DS2 exhibit this type of behavior, DS1 is more heterogeneous in the bound than in the free state.

To probe the degree of structural heterogeneity in the side-chain conformations, pairwise side-chain RMSDs were calculated for both ensembles (Figure 3H and Table 2). The average pairwise side-chain RMSD was found to be 3.2 Å in the free ensemble and 3.0 Å in the bound ensemble, indicating the presence of significant side-chain variability in both cases. The side-chain RMSD analysis for different regions in the free and bound states reveals the variations of the heterogeneity of the side chains upon binding. The protein as a whole (Figure 3E), the binding site (Figure 3F), and DS2 (Figure 3H) are conformationally more heterogeneous in the free state than in the bound state. Upon peptide binding, a significant decrease in the structural diversity is found in DS2 (Figure 3H), while the opposite is found for DS1 (Figure 3G).

Pairwise side-chain RMSD distributions allow the characterization of structural similarity between free- and bound-state ensembles. The free and bound states are found to be more variable with respect to each other than to themselves (Figure 3E−H), consistent with the conclusion that a change in conformation takes place upon binding.

Overall, these results characterize in detail the changes in structure and the significant redistribution of dynamics that happen upon ligand binding in the PDZ domain, and they show that different regions of the protein, in particular DS1 and DS2,

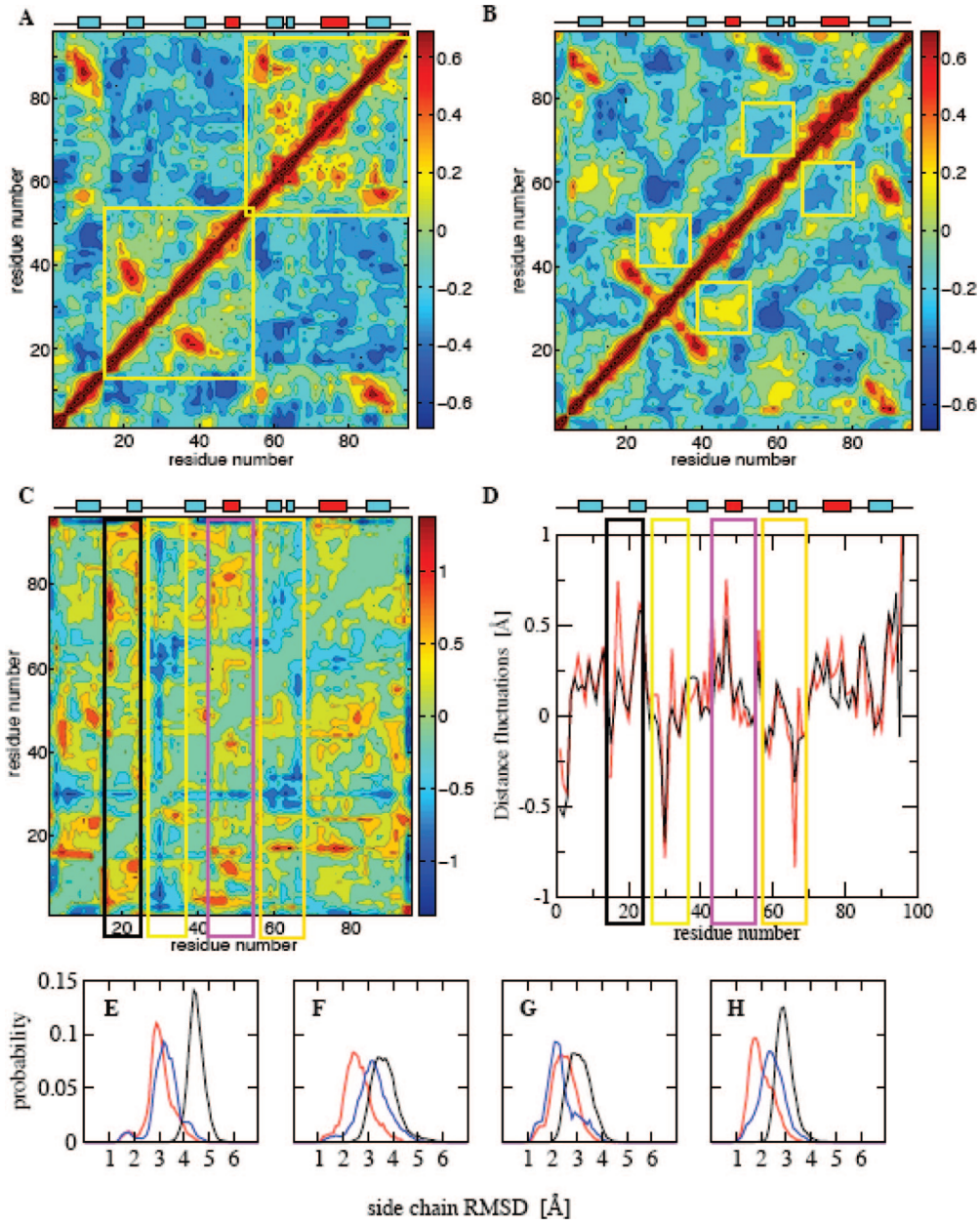

Figure 3. Analysis of structural changes of PDZ upon binding as determined from the ensembles of structures described in this work. (A) Side-chain covariance (see Methods) in the free state. (B) Side-chain covariance in the bound state. (C) Differences in Cα−Cα distance fluctuations in free and bound states. (D) Residue-specific distance fluctuations in free (black) and bound (red) states. In panels C and D, regions of increased mobility are boxed in black (binding site) and purple (distal region 2), and regions of decreased mobility are boxed in yellow (distal region 1 and region of residue V30). (E−H) Side-chain RMSD distributions for different regions of the protein within the ensembles that we have determined; the RMSDs in the free state are indicated in blue, those in the bound state in red, and those between the free and the bound states in black. (E) Entire protein; (F) binding site; (G) distal region 1; (H) distal region 2.

exhibit opposite responses. Increased or decreased mobility upon binding has been observed in a variety of cases,[39–42] and entropic contributions have been recently suggested to play an important role in determining the free energy changes associated with protein−protein or protein−ligand association.[43–47] The data that we present indicate that, for the PDZ complex studied here, entropy changes upon binding can have different signs in different regions of the structure. The conformational changes in PDZ occurring upon binding allow the distal surface 1 to gain in flexibility and may thus complement, at least in part, the rigidification of the binding site and distal surface 2.

Networks of Residues Changing in Structure and Dynamics. In order to probe differences between the conformational ensembles representing the free and bound states, we monitored all the backbone ϕ and ψ as well as all the side-chain $\chi1$ and $\chi2$ dihedral angles (Figure 4) and analyzed the differences in their distributions between both states (see Methods). The largest changes in $\chi1$ and $\chi2$ dihedral angles are observed in specific

(39) Atkinson, R. A.; Kieffer, B. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *44*, 141–187.
(40) Finerty, P. J.; Muhandiram, R.; Forman-Kay, J. D. *J. Mol. Biol.* **2002**, *322*, 605–620.
(41) MacRaild, C. A.; Daranas, A. H.; Bronowska, A.; Homans, S. W. *J. Mol. Biol.* **2007**, *368*, 822–832.
(42) Zhuravleva, A.; Korzhnev, D. M.; Nolde, S. B.; Kay, L. E.; Arseniev, A. S.; Billeter, M.; Orekhov, V. Y. *J. Mol. Biol.* **2007**, *367*, 1079–1092.

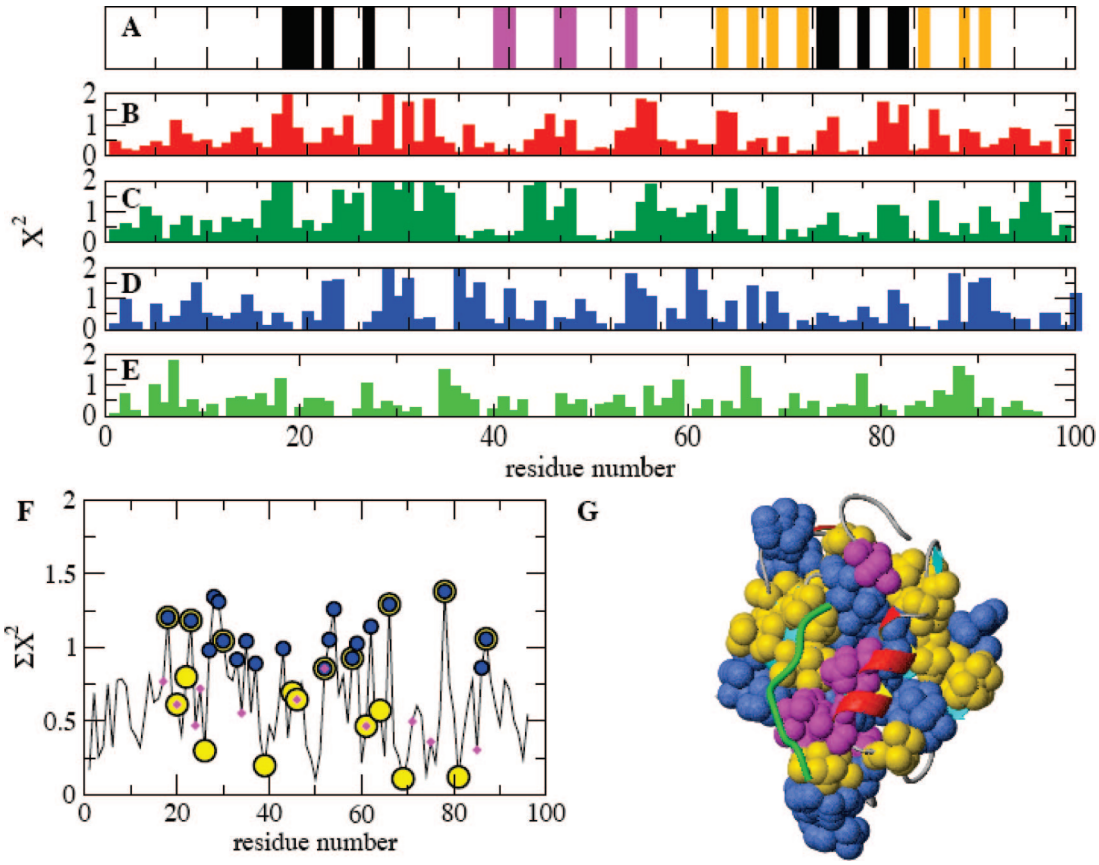

Figure 4. Schematic representation of changes upon binding in the dihedral angles through the use of distances, measured by the χ^2 factors, between free- and bound-state distributions. (A) Location of the binding site (black) and extended distal regions 1 (orange) and 2 (purple). (B) ϕ angle; (C) ψ angle; (D) $\chi1$ angle; (E) $\chi2$ angle. (F) Sum of χ^2 factors for ϕ, ψ, $\chi1$, and $\chi2$ angles along protein sequence. The evolutionary network (purple diamonds) the structural network (blue circles), and the dynamical network (yellow circles) are also indicated; eight residues belong to both the structural and dynamical networks (see Supporting Information Table 1). (G) Representation of evolutionary, structural, and dynamical networks on the PDZ structure; the color code is the same as in panel F.

Table 2. Ensemble-Averaged Side-Chain RMSDs and Corresponding Standard Deviations for the Free and Bound States[a]

	all	BS	DS1	DS2
free state RMSD (Å)	3.22 ± 0.57	3.16 ± 0.65	2.30 ± 0.59	2.36 ± 0.50
bound state RMSD (Å)	2.98 ± 0.52	2.55 ± 0.55	2.41 ± 0.47	1.97 ± 0.46

[a] "All" indicates residues P3−S94, BS is the binding site, and DS1,2 distal surfaces 1 and 2.

Table 3. Cα Cluster Analysis of Different Regions of the Structure of PDZ, for the Bound and the Free Ensembles, Showing the Number of Clusters Obtained with Different Thresholds[a]

threshold	state	all residues	binding site	distal region 1	distal region 2
0.5 Å	free	160	100	31	22
0.5 Å	bound	160	61	71	16
1.0 Å	free	30	17	3	6
1.0 Å	bound	30	8	9	2
1.5 Å	free	14	7	2	2
1.5 Å	bound	8	2	2	1
2.0 Å	free	7	2	1	1
2.0 Å	bound	2	1	1	1

[a] The use of different RMSD thresholds shows the convergence of the results.

regions that differ in part from the corresponding ones for the ϕ and ψ angles. These results suggest that backbone and side chains exhibit different behaviors upon peptide binding, confirming the importance of probing the motion of both. Interestingly, DS1 and DS2, which are defined on the basis of their differences in S^2 order parameter between free and bound states, do not emerge clearly as the regions that include residues with the largest differences in dihedral angles between the two ensembles. The same remark holds for residues belonging to the binding site. Therefore, we defined two sets of residues that change significantly upon binding either in structure or in dynamics, respectively.

The first set (the structural network) is formed by residues that exhibit a large change in dihedral angle distributions upon binding (see Methods). The 20 residues belonging to this set are found to be highly interconnected and strongly interacting with each other (see Methods), suggesting that they form a tightly knit network (see Methods and Supporting Information Table 1 for the complete list). Among the residues of the binding pocket, only residues L18 and L78 belong to this set. They therefore emerge from our analysis as the key residues that are associated with the large-scale conformational changes that span the structure from the binding site to the distal regions.

The second set (the dynamical network) consists of those residues that show significant changes in $S^{2,\mathrm{calc}}$ order parameter (see above). The residues in this second set are also strongly interacting with each other, but are notably less solvent-exposed than the first set (Supporting Information Figure 1).

Eight residues are in common between the two networks (Supporting Information Table 1), and the remaining ones are

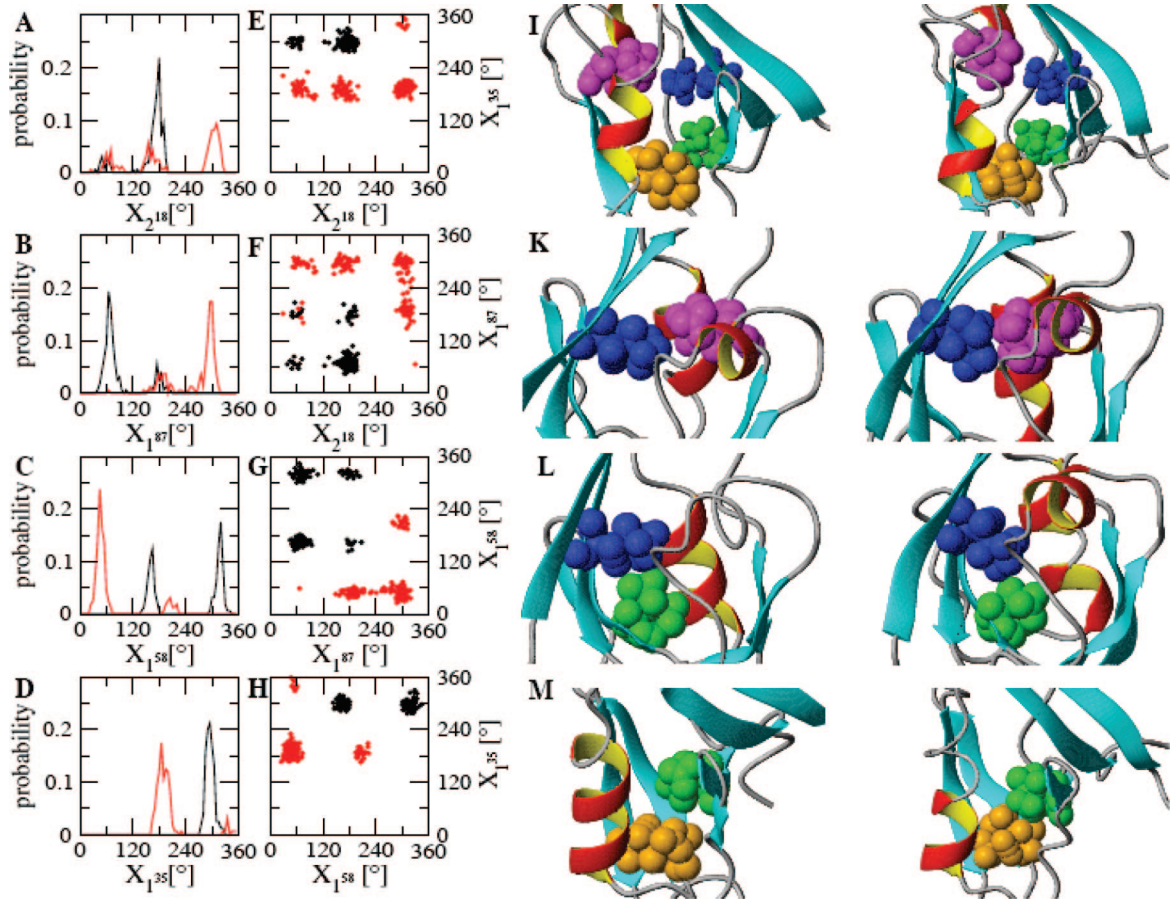

Figure 5. Representation of the correlations in the changes of dihedral angle distributions upon binding for neighboring residues. We illustrate here the pathway of coupled side-chain reorientations from residue L18 to residue I35 through residues L87 and V58. (A−D) Distributions of rotameric states in the free (red) and bound (black) conformations for residues (A) L18 (χ2), (B) L87 (χ1), (C) V58 (χ1), and (D) I35 (χ1). (E−H) Joint distributions of the side-chain dihedral angles for residues (E) L18 and I35 (χ2 and χ1), (F) L18 and L87 (χ2 and χ1), (G) L87 and V58 (χ1 and χ1), and (H) V58 and I35 (χ1 and χ1), in the free (red) and bound (black) states. (I, K−M) Illustration of the changes presented in panels E−H from residue L18 (purple) to I35 (orange) through residues V58 (blue) and L87 (green). The first column refers to the free state, the second to the bound state.

intertwined. The eight residues undergoing both dynamical and structural changes appear to be the key ones associated with conformational changes and they are located in the region of the binding site (L18, T23 and L78), in the distal surfaces (L66 and L87 in DS1 and I52 in DS2), or in the hydrophobic core of the protein (I35, V58).

Network Analysis of Cooperative Structural Changes within Proteins. The free and bound states of the PDZ domain that we studied here are characterized by different conformations and different dynamics on the nanosecond time scale. As we have shown here, two networks of residues are associated with the changes in structure and dynamics observed in the two states. These networks connect distant regions in the structure of the protein and are associated with the large-scale conformational changes across the protein from the binding site to the opposite surface.

As a consequence of the presence of networks, residue pairs in different regions of the structure are often linked through pathways.[48–57] We illustrate the correlation of structural changes of residues in different regions of the protein in the specific case of the pathway between residues L18 and I35 (Figure 5 and Supporting Information Table 2). The approach that we describe in this work reveals that while these residues are not directly in contact, their rotameric states are highly correlated through the existence of networks within the structure of the PDZ domain. The change in the rotameric state of L18 due to ligand binding is associated with a series of changes in the statistical weights of the rotameric states of the side chains strongly coupled to it (Figure 5). These changes are correlated

(43) Akke, M.; Bruschweiler, R.; Palmer, A. G. *J. Am. Chem. Soc.* **1993**, *115*, 9832–9833.
(44) Maler, L.; Blankenship, J.; Rance, M.; Chazin, W. J. *Nat. Struct. Biol.* **2000**, *7*, 245–250.
(45) Popovych, N.; Sun, S. J.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831–838.
(46) Niv, M. Y.; Weinstein, H. *J. Am. Chem. Soc.* **2005**, *127*, 14072–14079.
(47) Basdevant, N.; Weinstein, H.; Ceruso, M. *J. Am. Chem. Soc.* **2006**, *128*, 12766–12777.
(48) Clarkson, M. W.; Gilmore, S. A.; Edgell, M. H.; Lee, A. L. *Biochemistry* **2006**, *45*, 7693–7699.
(49) del Sol, A.; Fujihashi, H.; Amoros, D.; Nussinov, R. *Protein Sci.* **2006**, *15*, 2120–2128.
(50) Dima, R. I.; Thirumalai, D. *Protein Sci.* **2006**, *15*, 258–268.
(51) Fleishman, S. J.; Yifrach, O.; Ben-Tal, N. *J. Mol. Biol.* **2004**, *340*, 307–318.
(52) Formaneck, M. S.; Ma, L.; Cui, Q. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 846–867.
(53) Kass, I.; Horovitz, A. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 611–617.
(54) Lockless, S. W.; Ranganathan, R. *Science* **1999**, *286*, 295–299.
(55) Rousseau, F.; Schymkowitz, J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 23–30.
(56) Swain, J. F.; Gierasch, L. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 102–108.
(57) Ota, N.; Agard, D. A. *J. Mol. Biol.* **2005**, *351*, 345–354.

from residue L18 to residue I35 through residues L87 and V58. This example thus highlights the importance of residues V58 and L87, which are common to the structural and dynamical networks, in the conformational transition upon binding.

We carried out a systematic network analysis that identified 258 similar structural pathways involving four residues and 142 pathways involving five residues. We then analyzed how often the residues that take part in the structural network appear in these 475 pathways and where they are located, that is, whether they are found at the end points of the pathways or in their centers (Supporting Information Figure 2). Six residues emerge from this analysis as the most connected ones: I35, V37, I52, V58, L78, and L87. As far as binding residues are concerned, residue L18 mainly acts as a first point in pathways, whereas residue L78 acts both as first and as central residue. This type of analysis helps identify residues that are mainly end points of the structural pathways, such as residues N27−S29, situated in the loop between β2 and β3, residue Q43, situated close to DS2, and residues N62 and L66, close to or in DS1. These three regions were also identified by the fluctuation and covariance analysis presented above as being the part of the protein undergoing the more dramatic changes upon peptide binding.

Residues I35, V37, I52, V58, and L87 appear as the most frequently present central residues for all types of pathways, including those connecting the binding site (L18 and L78) to other regions, and those connecting a solvent-exposed part of the PDZ domain to another (Supporting Information Figure 3). These five residues form in part the hydrophobic core of the protein and they exhibit significant changes in their rotameric states. Three of them (I52, V58, L87) also undergo significant changes in S^2. In addition, most of the pathways involving one of the binding residues L18 (located in helix α2) and L78 (located in strand β2) also involve the other one, suggesting that there is a concerted motion of the two parts of the binding site upon binding. This observation complements the results of the covariance analysis that showed a change in correlated motion between the two parts of the binding site upon binding.

Link with an Evolutionary Network of PDZ Domains. In addition to the two distinct networks that we described, the PDZ domain that we analyzed contains another type of network that is formed by coevolving residues.[55] Evolutionary analysis represents a possible approach for understanding the molecular basis of conformational changes by taking advantage of the fact that residues important for function are likely to be preserved through evolution and leave a trace in the pattern of sequence conservation that is characteristic of members of the protein family.[55,58,59] Evolutionary networks may, at least in principle, contain residues that are important for different reasons, including for stability or for function. Lee and co-workers[14] previously observed the overlap between the evolutionary and dynamical networks. We also find that this evolutionary network corresponds more closely to the dynamical network in our analysis rather than to the structural network. We also increased the overlap between the evolutionary and dynamical networks by the incorporation of residues A46 and I52 into the dynamical network; we included these two residues because they exhibited significant changes upon binding as identified by the restrained molecular dynamics simulations that we used.

Residues in the evolutionary network are strongly interacting but do not exhibit large changes in the distributions of dihedral angles upon binding, and therefore they do not take part in the structural network (Figure 5 and Supporting Information Table 1). For example, residue H71 in the evolutionary network, which is important for binding in PDZ domains, does not change significantly in dihedral angles according to our analysis. This residue thus appears to act as an anchor residue by having its side chains already prepared to bind the peptide.

Conclusions

We have used molecular dynamics simulations with ensemble-averaged NMR restraints to study the free and bound states of the second PDZ domain of human tyrosine phosphatase 1E. Analysis of the resulting ensembles of conformations has enabled a map of the changes in structure and dynamics to be obtained, which has revealed the presence of two interconnected networks of residues associated with the response to ligand binding. The first network is formed by a set of strongly interacting residues that undergo changes in the distribution of their rotameric states. The second is composed by a set of residues that experience a variation in their picosecond to nanosecond dynamics.

The conformational ensembles that we have presented enable a description of a mechanism of conformational changes upon binding that involves coupled side-chain rotamer reorientations. We thus provide a detailed characterization of the relationship between structural and dynamical changes that take place in this PDZ domain protein upon binding and obtain a mapping of the pathways connecting residues that exhibit correlated changes in rotameric states. The methodology that we have described is general and should enable the dynamics on the nanosecond time scale to be obtained for a range of proteins, particularly those involved in enzymatic catalysis and allosteric communication.

Acknowledgment. We are grateful to I. Ekiel and G. Kozlov for sending us their unpublished data on the *J* couplings of the bound state and to M. Akke, P. De Los Rios, and A. Lee for many interesting discussions. This work was supported by the Swiss National Science Foundation (J.G.), the Medical Research Council (J.G.), the EU (M.V.), the Leverhulme Trust (M.V.), and the Royal Society (M.V.).

Supporting Information Available: Supporting Tables 1 and 2 and Supporting Figures 1−3. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0752080

(58) Socolich, M.; Lockless, S. W.; Russ, W. P.; Lee, H.; Gardner, K. H.; Ranganathan, R. *Nature* **2005**, *437*, 512–518.
(59) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. *Nat. Struct. Biol.* **2003**, *10*, 232–232.