

Peptide binding to the PDZ3 domain by conformational selection

Sandra Steiner and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

ABSTRACT

The PDZ domains, a large family of peptide recognition proteins, bind to the C-terminal segment of membrane ion channels and receptors thereby mediating their localization. The peptide binding process is not known in detail and seems to differ among different PDZ domains. For the third PDZ domain of the synaptic protein PSD-95 (PDZ3), a lock-and-key mechanism was postulated on the basis of the almost perfect overlap of the crystal structures in the presence and absence of its peptide ligand. Here, peptide binding to PDZ3 is investigated by explicit solvent molecular dynamics (MD) simulations (for a total of 1.3 μ s) and the cut-based free energy profile method for determining free energy barriers and basins. The free energy landscape of apo PDZ3 indicates that there are multiple basins within the native state. These basins differ by the relative orientation of the α 2 helix and β 2 strand, the two secondary structure elements that make up the peptide binding site. Only the structure with the smallest aperture of the binding site is populated in the MD simulations of the complex whose analysis reveals that the peptide ligand binds to PDZ3 by selecting one of three conformations. Thus, the dynamical information obtained by the atomistic simulations increment the static, that is, partial, picture of the PDZ3 binding mechanism based on the X-ray crystallography data. Importantly, the simulation results show for the first time that conformational selection is a possible mechanism of peptide binding by PDZ domains in general.

Proteins 2012; 80:2562–2572.
© 2012 Wiley Periodicals, Inc.

Key words: free energy surface; PDZ domain; molecular dynamics; reaction coordinate optimization.

INTRODUCTION

PDZ domains are found in various proteins with diverse functions either alone or as arrays. They mediate interactions by binding to the C-terminal segment of target proteins and play an important role in signal transduction and scaffolding.^{1,2} PDZ domains are small (\approx 90 residues) and have a highly conserved fold consisting of five to six β -strands (β 1– β 6) and two α -helices (α 1, α 2). The C-terminus of the ligand protein binds in an extended conformation to an elongated pocket (termed hereafter binding site groove) between the β 2 strand and the α 2 helix thereby augmenting the antiparallel β 2– β 3 hairpin into a three-stranded β -sheet.

A large number of crystal and nuclear magnetic resonance (NMR) structures of PDZ domains in ligand-free (apo) and complexed (holo) state have been solved,^{3–6} and several studies have investigated the binding mechanisms of different PDZ domains and their peptide binding partners using a range of experimental techniques.^{5–10} Most of them have concerned three of the most thoroughly studied PDZ domains, the second PDZ domain of human tyrosine phosphatase PTPL1 or murine tyro-

sine phosphatase PTP-BL (human PDZ2, murine PDZ2) and the third PDZ domain of human synaptic protein PSD-95 (PDZ3). In particular, a lock-and-key binding mechanism has been suggested for PDZ3 due to a lack of visible conformational changes when overlaying crystal structures of apo and holo PDZ3.³ This finding has been corroborated by kinetic experiments suggesting a one-step binding mechanism without rate-limiting conformational changes.^{7,8} For murine PDZ2, on the other hand, NMR structures together with ultrafast kinetic experiments provide evidence for an induced-fit binding mechanism.⁸ Additionally, also for the human form of PDZ2 inspection of NMR structures suggests an induced-fit binding mechanism.⁵ Yet, more recent crystallographic experiments corroborated by residual dipolar couplings

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

Grant sponsor: Swiss National Science Foundation (to A.C.).

*Correspondence to: A. Caflisch Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. E-mail: caflisch@bioc.uzh.ch

Received 13 April 2012; Revised 4 June 2012; Accepted 14 June 2012

Published online 2 July 2012 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/prot.24137

were able to detect only slight conformational changes in the crystal structures.⁶ Complementing the crystal structures with Carr-Purcell-Meiboom-Gill relaxation dispersion data indicated a lock-and-key mechanism to be the most suitable binding mechanism for human PDZ2.⁶ An interesting structural feature regarding binding is the β 2– β 3 loop. It has been reported that for some PDZ domains this loop contributes to protein–ligand interactions⁵ and that it has often enhanced conformational flexibility.^{4,10} Therefore, it is speculated to assist in ligand-selectivity of PDZ domains.

Several experimental and computational studies have focused on the allosteric behavior of PDZ domains.^{11–15} Large conformational changes and rearrangements are rare among the resolved structures, thus particularly the intramolecular signaling properties have been of interest. Evolutionary conserved signaling pathways seem to connect the binding site residues to distal parts of the PDZ fold.¹⁶ Computational studies yielded similar pathways by investigating residue–residue interaction energy correlations.¹⁷ An NMR spectroscopy study has shown that the additional structural element of PDZ3, a C-terminal α -helix (α 3), influences ligand affinity by modifying the side chain dynamics of the whole domain.¹⁴ This study has provided evidence of dynamic allostery in the PDZ3 domain, as its C-terminal α -helix is distal to the ligand and dynamical fluctuations throughout the protein cause the changes in binding affinity.¹⁴

Binding mechanisms of several proteins have been investigated by atomistic simulations. It has been suggested that the distinction between the induced-fit and the conformational selection model is not absolute.¹⁸ A computational study regarding the binding mechanism of Lysine-, Arginine-, Ornithine-binding (LAO) protein corroborates this suggestion as it appears from the simulations that binding by conformational selection is followed by conformational readjustment,¹⁹ thus a combination of the two mechanisms “conformational selection” and “induced-fit” is at work. On the other hand, molecular dynamics (MD) simulations suggest that the flaviviral NS2B-NS3pro protease actually binds its peptide substrate by conformational selection,²⁰ although crystal structures displaying conformational differences between apo and holo conformations of NS2B-NS3pro had previously led to the speculation of an induced-fit binding mechanism. The MD simulations started from the apo structure showed that even without the substrate the catalytic competent conformation formed spontaneously and was kinetically stable on a 50-ns time scale.²⁰ Furthermore, a two-state model of the binding free energy surface of PDZ3 has recently emerged from atomistic Monte Carlo simulations with restraints on the protein.²¹

Here, the mechanism of peptide binding by the third PDZ domain (PDZ3) from the human synaptic protein PSD-95 is investigated by explicit solvent MD simulations

of its complexed (holo) and peptide-free (apo) state. This simulation study was motivated by the heterogeneity of binding mechanisms suggested by using a variety of experimental methods. In this context, there are several outstanding questions: How to reconcile the lock-and-key mechanism postulated for peptide binding to PDZ3^{3,7} with the evidence for a “delocalized conformational entropy mechanism” mediated by the α 3 distal helix¹⁴? Is it possible to shed light on the plasticity of the β 2– β 3 loop and helix α 2 and their influence on the binding mechanism? Is the functional role of α 3 only of dynamical nature or has α 3 also a structural impact? How can the equilibrium native state ensembles for apo and holo PDZ3 be characterized structurally? The simulation results indicate that PDZ3/peptide association follows most likely a conformational selection mechanism. While the complexed state is rather rigid the peptide-free state displays conformational flexibility. In particular, the β 2– β 3 loop and the tilting of the helix α 2 characterize structurally the different free energy basins of the apo state.

MATERIALS AND METHODS

MD simulations

The explicit solvent MD simulations were carried out with CHARMM²² and GROMACS²³ using the CHARMM22 all-hydrogen force field²⁴ and the TIP3P water model.²⁵ The coordinates of the complex between PDZ3 and the 5-residue C-terminal motif Lys-Gln-Thr-Ser-Val were downloaded from the protein data bank (PDB code 1BE9) and used as starting structures. Note that residues 303–402 of the PDZ3 domain are termed here residues 1–100, in accordance with the work of Petit et al.¹⁴ To reproduce neutral pH conditions, the side chains of aspartates and glutamates as well as the C-terminus of the peptide were negatively charged, those of lysines and arginines were positively charged, and histidines were considered neutral. The protein was immersed in a truncated octahedral box of preequilibrated water molecules. The size of the box was chosen to have a minimal distance of 12 Å between the boundary and any atom of the protein. The simulation box contained five and four potassium ions to compensate for the net negative charge of apo PDZ3 and PDZ3/peptide complex, respectively.

Periodic boundary conditions were applied, long-range electrostatic interactions were treated with the Particle Mesh Ewald method²⁶ and the van der Waals interactions were truncated at a cutoff of 10 Å using a switch function starting from 8 Å. The MD simulations were performed at constant temperature (298 K) using the Berendsen thermostat²⁷ and constant pressure (1 atm) with an integration step of 2 fs and saving frequency of 2 ps. After an equilibration and heating phases the production runs were started using different seeds for the initial distribution of velocities.

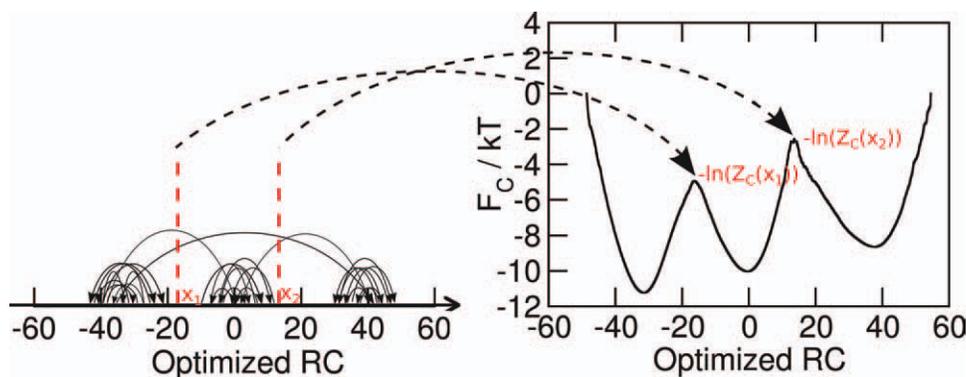


Figure 1

Schematic illustration of cFEP. By computing its optimized RC value (real number), every MD snapshot can be uniquely placed on the line representing the real numbers. The transitions are given naturally from snapshot i to snapshot $i + 1$ and are here illustrated by curved arrows (left). Then, for a fixed RC value, for example, x_1 , the cut-based partition function $Z_C(x_1)$ is the number of transitions through this point, that is, the number of curved arrows crossing the dashed line at x_1 . Computing this values for every x gives the cFEP as $F_C(x) = -kT \ln(Z_C(x))$ (right). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The analysis of the MD trajectories was carried out with CHARMM²² and the MD-analysis tool WORDOM.^{28,29}

Optimization of reaction coordinate

To construct a one-dimensional free energy profile, the MD trajectory has to be further processed. Here, the total of 655,000 snapshots was binned along a reaction coordinate (RC), another common method would be grouping the snapshots with a clustering algorithm (e.g., using a tree-based algorithm³⁰). The binning suggests the conventional way of projecting the free energy as $F_H(i) = -kT \ln Z_H(i)$ where $Z_H(i)$ denotes the partition function which is equal to the density of bin i , that is, $Z_H(i) = \frac{\text{\#frames in bin } i}{\text{bin size}}$ and F_H is short for histogram-based free energy.

A more kinetically motivated approach that is able to preserve the free energy barriers is to compute the cut-based free energy profile (cFEP).^{31–33} This approach emulates the cuts in flow-networks. In the case of conformational space sampled by MD, the nodes and links of the network are the mesostates (e.g., bins along a RC or clusters obtained by root mean square deviation [RMSD] clustering) and the sampled transitions between them, respectively.³⁴ The partition function of the free energy barrier separating two mesostates i and j is equal to the value of minimum cut between the mesostates in the network.³¹ Practically, this means that the network is partitioned in two distinct sets I and J , which include mesostates i and j , respectively, such that the number of transitions between the two sets are minimal. This minimal number of transitions is then assigned as partition function of the transition state between mesostates i and j as it represents the statistical weight of this transition. Similarly, one can estimate in continuous terms the statistical

weight for a RC and thereby the cut-based partition function $Z_C(x)$ at a point x , as the number of transitions through point x .³² Accordingly, the computation of the cut-based free energy is performed as $F_C(x) = -kT \ln(Z_C(x))$ where x is the RC and $Z_C(x)$ is the cut-based partition function at the point x defined as the number of transitions that pass through the point x . In terms of bins, the following formula applies $Z_C(i, i + 1) = \frac{1}{2}(\sum_{j \leq i < k} n_{jk} + \sum_{j > i \geq k} n_{jk})$ where n_{ij} is the number of transitions from bin j to bin i . In terms of continuous RC, the formula becomes $Z_C(x) = \frac{1}{2} \sum_t \theta((x(t) - x)(x - x(t + \Delta t)))$ where θ is the Heaviside step function and Δt is the sampling interval. A schematic illustration of this procedure is shown in Figure 1.

The aim of the optimization of the RC is to find a coordinate that shows diffusive dynamics (at least at the barrier regions). Diffusive dynamics is characterized by the mean square displacement growing linearly with time. The following equation defines the so-called subdiffusion exponent α : $\langle \Delta x^2(t) \rangle \sim t^{2\alpha}$. For diffusive dynamics, one has $\alpha = 0.5$, whereas for subdiffusive dynamics $\alpha < 0.5$. The coordinate dependent exponent $\alpha(x)$ can be computed from the distance between two profiles $F_C(x)$ at different sampling intervals.³⁵ Plotting $F_C(x)$ together with $\alpha(x)$ gives insight into the dynamics along a profile.

The optimization procedure starts from a seed RC which is able to capture the configuration space sufficiently well, that is, partially distinguish between free energy basins, but is suboptimal, meaning that projection of the free energy onto this coordinate results in possible subdiffusive dynamics at barrier and other regions and in underestimating barrier heights due to overlapping of the configuration space. The seed RC chosen here is the RMSD of the C_α atoms of residues 6–97 from the first

frame of the MD trajectory. As projecting to a suboptimal RC results in faster kinetics by overlapping of different regions of the configuration space, the optimal RC is defined by having the longest mean first passage times (MFPT) from one free energy basin to another. The optimal RC is then constructed by maximizing the MFPT that is computed from the trajectory using Kramer's equation in the following form

$$\begin{aligned} \langle t_{A,B} \rangle &= \int_a^b dx \frac{e^{\beta F_H(x)}}{D(x)} \int_{-\infty}^x dy e^{-\beta F_H(y)} \\ &= \frac{\Delta t}{\pi} \int_a^b dx \frac{Z_H(x)}{Z_C^2(x)} \int_{-\infty}^x dy Z_H(y), \end{aligned}$$

where $\beta = \frac{1}{kT}$, $D(x)$ is the coordinate-dependent diffusion coefficient, and a and b denote two representative coordinate values for two basins A and B , respectively.^{32,35} The functional form of the RC is the “smoothed number of native contacts” $R(x, \alpha) = \sum_{i,j} \alpha_{ij} \theta(\Delta_{ij} - r_{ij})$ where r_{ij} is the distance between atoms i and j while α_{ij} and Δ_{ij} are the parameters to be optimized (the former is either 1 or -1 while the latter is a threshold for a formed contact $\in \{0\text{\AA}, 0.5\text{\AA}, \dots, 30\text{\AA}\}$) and θ is a smoothed step function, that is, $\theta(x) = \min(1, x)$ if $x > 0$ and zero otherwise.³⁵ The optimization is carried out iteratively by randomly picking a pair ij and selecting α_{ij} and Δ_{ij} according to the given ranges above, such that MFPT is maximized. In addition, pairs of coordinates ij that do not bring substantial improvement are removed automatically (see Ref. 35 for further details).

There are significant differences between the cFEP method and other frequently used techniques for projecting the free energy surface, for example, histogram-based profiles and principal component analysis. The main advantage of the cFEP method is that the dynamical information present in the trajectory is fully taken into account. Moreover, the cFEP with RC optimization does not require any geometrical clustering. By considering the actual transition matrix, the approach used in this work performs dimensionality reduction while preserving the dynamic information.³⁵ In contrast, other frequently used methods reduce the dimensionality by trying to preserve the proximity in configuration space. The vicinity in configuration space does not necessarily imply dynamical closeness. In other words, geometrically similar conformations can be separated by high free energy barriers.

Tilting angle

The tilting angle between $\alpha 2$ helix and $\beta 2$ strand is a measure of scissor-like partial widening of the binding site groove. It is computed as the angle between two directional vectors given by $\beta 2$ (from C_α of Ile25 to carbonyl C of Gly22) and $\alpha 2$ (from carbonyl C of Glu71 to carbonyl C of Lys78).

RESULTS

The following analysis is based on four explicit solvent MD runs for each of the apo and holo states. The total simulation time is 660 ns for apo PDZ3 and 650 ns for holo PDZ3 during which the native fold is preserved as expected at the temperature of 298 K. The residues 303–402 of PDZ3 according to the PDB file 1BE9 are hereafter termed residues 1–100 while the five C-terminal residues of the peptide ligand are referred to as Val0, Ser-1, Thr-2, Gln-3, and Lys-4 (from C-terminus to N-terminus).

Differences in the free energy surface and plasticity of apo and holo

The cFEP,³¹ projected onto an optimized RC³⁵ (see “Materials and Methods” section), displays three free energy basins [called A, B, and C in Fig. 2(A)] within the native state. From the time series of the value of optimized RC [Fig. 2B (top)] along the simulation trajectories, it is evident that in the apo simulations all the three free energy basins are sampled, whereas in the holo runs the system never leaves basin A. Defining the ranges of optimized RC according to basin A, B, and C as -38 to -25 , -8 to 6 , and 25 to 45 , respectively, every snapshot of the simulation can be allocated either to an individual basin or to the transition region according to its RC value. From this, we can derive the relative populations of the three basins in the apo simulations as approximately 36%, 45%, and 13% for basins A, B, and C, respectively, whereas the remaining 6% of the snapshots populate the barrier regions. The optimization procedure is started from an appropriate seed RC which should be able to capture the configuration space well enough. Although such a conventionally chosen RC often leads to subdiffusive dynamics and underestimating of barrier heights due to overlapping of the configuration space, the optimization procedure yields a RC which is optimal in the sense that at the barrier regions the projected dynamics are diffusive. Furthermore, possible configuration space overlap is removed. We used the RMSD of the C_α atoms of residues 6–97 from the first frame of the MD trajectory as a seed RC for the optimization. The resulting optimized RC is more appropriate than the RMSD as it separates the three free energy basins better [Fig. 2(B)] and the dynamics at the barriers on projection is diffusive [Fig. 2(A)]. The time series of the RMSD of the C_α atoms of residues 6–97 never exceeds 2 Å and is below 1.5 Å most of the time which shows that the folded state is preserved in all MD runs. Carrying out the cFEP analysis using the apo and holo sampling separately confirms the finding that apo PDZ3 samples three different free energy basins whereas holo PDZ3 samples only one. Projection of the free energy for the apo sampling onto the seed RC and later on onto the optimized RC reveals

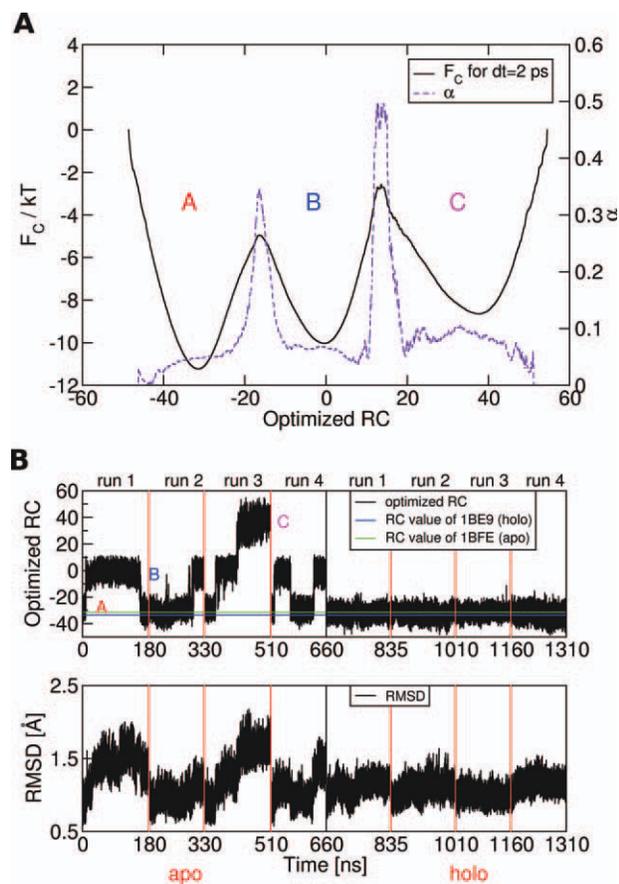


Figure 2

Only one of the three free energy basins of apo PDZ3 is populated by the PDZ3/peptide complex. (A) Cut-based free energy F_C along the optimized RC for the accumulated apo and holo MD runs. The three free energy basins are labeled A, B, and C. The F_C (solid line) is shown together with the subdiffusion coefficient α (dashed line). The value of the subdiffusion coefficient $\alpha = 0.5$ in the barrier regions indicates that the optimization of the RC yields diffusive dynamics³⁵ at the transition regions between basins in the native state. Note that the subdiffusion $\alpha < 0.5$ at the bottom of the basins does not influence significantly the overall kinetics which are dominated by the barrier regions. B, (top) Time series of value of optimized RC along the simulation trajectories. Vertical lines indicate the beginning/end of each MD run and the black vertical line separates apo from holo runs. The time series of optimized RC shows that only the basin A is occupied in the absence of the peptide ligand while all of the three basins [labeled as in (A)] are populated in the complexed state. The value of the optimized RC for the X-ray structures of apo (PDB code 1BFE) and holo (PDB code 1BE9) are shown with green and blue horizontal lines, respectively. B (bottom) Time series of RMSD of the C_α atoms of residues 6–97 from the first frame. Vertical lines are shown as in the top panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

three free energy basins, whereas for holo PDZ3 projection of the free energy onto the seed RC results in a single free energy basin (see Supporting Information Figs. S1 and S2).

It is useful to supplement the cFEP analysis with the differences in the root mean square fluctuations (RMSF)

calculated from the MD runs. The RMSF of the C_α atoms in the apo state are higher than in the complexed state (Fig. 3) which is consistent with the cFEP analysis. The RMSF differ most at the loop regions, that is, the $\beta 1$ – $\beta 2$ and $\beta 2$ – $\beta 3$ loops, as well as at the α -helix ($\alpha 2$) which lines the peptide binding groove. Overall, the RMSF calculated from the MD trajectories are in good agreement with the crystallographic temperature factors for 1BFE (apo) and 1BE9 (holo) (Fig. 3). Moreover, the X-ray structure of the apo state has higher B-factors than the one of the holo state, and the difference is larger in the same segments of the sequence as for the RMSF difference, that is, the loop regions and $\alpha 2$ helix.

It has to be mentioned that all of the seven X-ray structures of PDZ3 (apo: 1BFE, 1TQ3, 3I4W, and 3K82; holo: 1BE9, 1TP3, and 1TP5) are located in basin A with values of the optimized RC ranging from -23.8 for 3K82 to -34.6 for 1TP3. There are two possible reasons for the fact that according to their value of the optimized RC the four apo X-ray structures belong to basin A: it is either the most populated and/or the one favored by the crystal packing. Given such structural similarity among the available X-ray structures, the plasticity of the native state of apo PDZ3, as it emerges from the MD simulations, is rather unexpected. Thus, it is important to investigate the structural features and transitions that characterize the individual free energy basins within the native state of (apo) PDZ3, which are presented next.

Structural analysis

As indicated by the RMSF of the C_α atoms there are mainly two regions that display significant structural plasticity: the $\beta 2$ – $\beta 3$ loop and the $\alpha 2$ helix which make up the binding site groove. To analyze the relative displacement of these secondary structure element, it is useful to determine the structural ensembles corresponding to each of the three basins identified by the optimized RC. The three basins can be characterized by the position of the $\beta 2$ – $\beta 3$ loop and the relative orientation of $\alpha 2$ and $\beta 2$ (Fig. 4). The tilting angle between $\alpha 2$ and $\beta 2$ (for definition see Materials and Methods section) widens by about 8° and 10° in going from basin A to basin B and C, respectively (see Supporting Information Figs. S3 and S4). Similarly, NMR structures of apo and holo LARG (leukemia-associated Rho guanine nucleotide exchange factor) PDZ show that peptide binding tilts $\alpha 2$ and $\beta 2$ apart.¹⁰

The distribution of selected pairs of distances are useful to further characterize the motion of the binding site [Fig. 4(C)] by complementing the observed widening of the tilting angle. The histograms of three distances between pairs of C_α atoms in $\alpha 2$ and $\beta 2$ (Gly22-Lys78, Asn24-Ala74, Ile26-His70) indicate that the widening of the tilting angle corresponds to a scissor-like motion of $\alpha 2$ with respect to $\beta 2$ with the pivot somewhere between the middle and the N-terminal turn of helix $\alpha 2$. Thus,

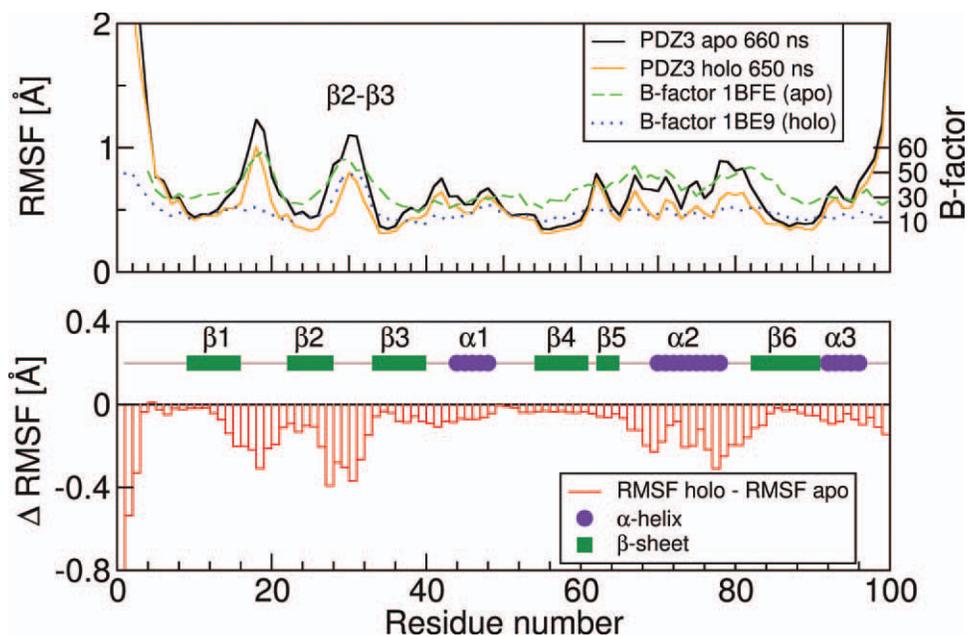


Figure 3

Differences in backbone flexibility between apo and holo PDZ3. (Top) Values of the RMSF of the C_{α} atoms for apo and holo PDZ3 (black and orange lines, y -axis on the left) and the crystallographic B-factors (blue dotted and green dashed lines, y -axis on the right) as a function of residue number. The RMSF values were averaged over simulation intervals of 5 ns. The first 5 ns of each MD run were discarded. (Bottom) The differences between the RMSF of holo PDZ3 and apo PDZ3 are shown by red bars. Secondary structure assignment was done with WORDOM^{28,29} using the X-ray structure of holo PDZ3 (1BE9). The individual elements of secondary structure are: β 1 strand, 10–15; β 2 strand, 23–27; β 3 strand, 34–39; β 4 strand, 55–60; β 5 strand, 63–64; β 6 strand, 83–89; α 1 helix, 44–48; α 2 helix, 70–78; α 3 helix, 91–97. The nomenclature used for individual α -helices and β -strands corresponds to the one introduced in Ref. 3, where βA to βF and αA to αC , are replaced by $\beta 1$ to $\beta 6$ and $\alpha 1$ – $\alpha 3$, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the largest displacement when going from basin A to basin B and C (>2 Å) takes place at the C-terminal turn (Gly22–Lys78 distance), whereas for the N-terminal part of helix $\alpha 2$ (Gly22–Lys78 distance, opposite of pivot) a smaller (≈ 1 Å) displacement in the opposite direction is observed. This scissor-like motion, therefore, widens or narrows a hydrophobic cavity of the binding site groove located between the C-terminal part of helix $\alpha 2$ and the N-terminal part of strand $\beta 2$.

The distance between the C_{α} atoms of residues 29 and 78 reflects the relative displacement of the $\beta 2$ – $\beta 3$ loop with respect to $\alpha 2$. The histograms of this distance show significant differences for the three free energy basins [Figure 4(C)]. The two-peak distribution for basin B indicates that there are two main orientations of this loop in basin B which differ by about 2 Å. Since the histogram of the Gly22–Lys78 distance for basin B is single-peaked, the two peaks arise from two different positions of Glu29, that is, two different conformations of the loop. The structural overlap of representative MD snapshots of basins A, B, and C visualizes the described structural differences among basins manifested in helix $\alpha 2$ and the $\beta 2$ – $\beta 3$ loop [Fig. 4(A)]. Furthermore, overlap of the available crystal structures of apo and holo PDZ3

with the representative structures of the individual basins reveals, as suggested by analysis of RC values, that apo and holo X-ray structures match best with the representative snapshot of basin A. Importantly, a tilting angle of helix $\alpha 2$ comparable to the one of the representative snapshot of basin B is not observed among the crystal structures [Fig. 4(D)].

The $\alpha 3$ helix is a structural element of PDZ3 that is additional to the usual PDZ fold, and has been previously suggested to modulate binding by a dynamic mechanism.¹⁴ Intriguingly, $\alpha 3$ includes four charged residues (Lys91, Glu93, Glu94, and Arg97) that form a complex network of salt bridges. It is, therefore, useful to relate individual basins of the cFEP to the presence/absence of salt bridges that “connect” the helix $\alpha 3$ to the rest of the fold. Interestingly, there are differences between apo and holo state in this network of salt bridges. In particular, Arg97 is involved in a salt bridge with Glu29 or Glu32 in the apo state and almost only with the latter in the complexed state (Fig. 5). Moreover, the salt bridge between Arg7 and Glu93 is almost always formed in the holo state while often disrupted in apo PDZ3. On the other hand, the Glu94–Arg97 salt bridge and the Lys91–Glu94 water-bridged interaction, both within helix $\alpha 3$, are almost

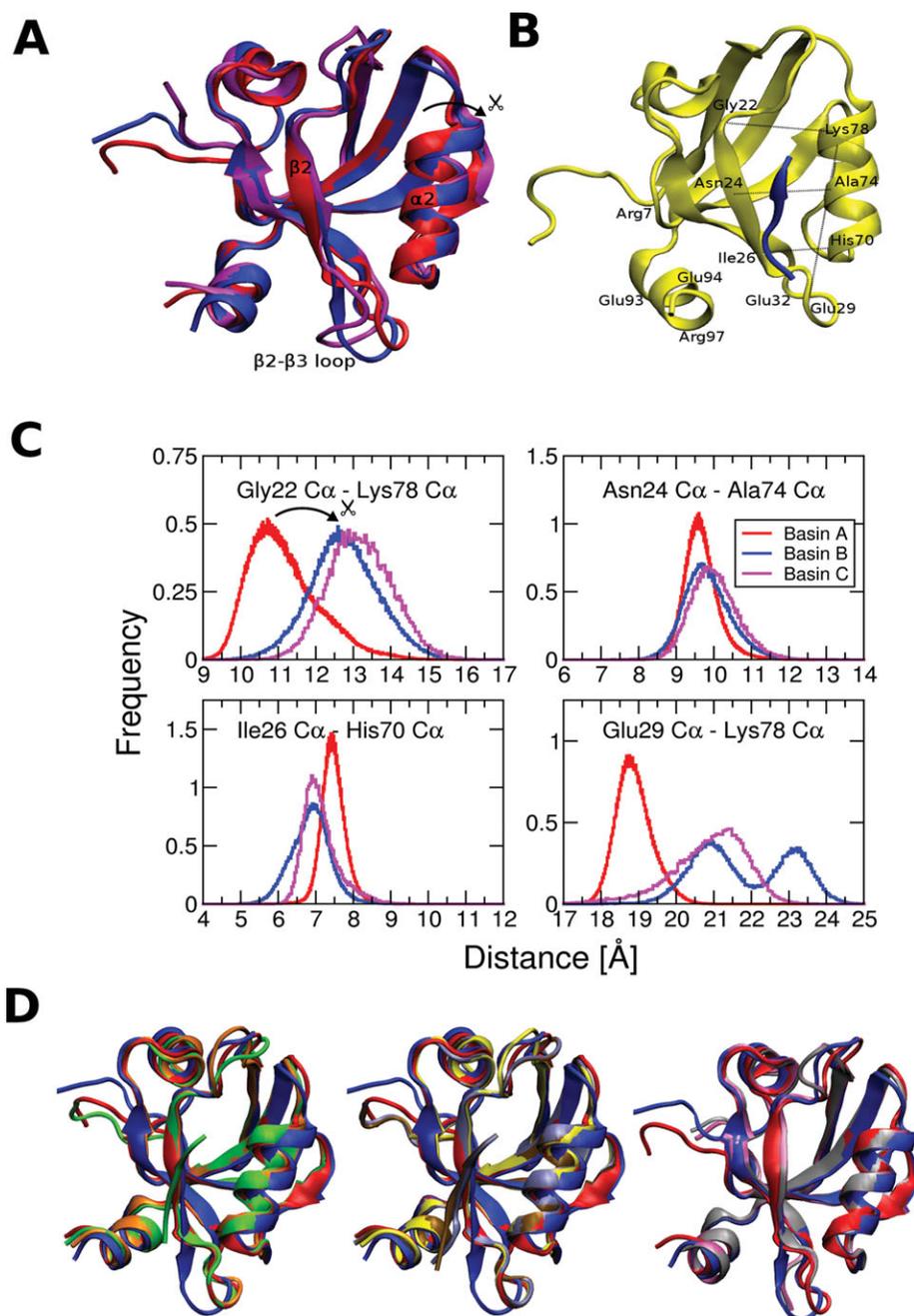


Figure 4

Structural characteristics of free energy basins of apo PDZ3. (A) Structural overlap of the representative structures of basins A (red), B (blue), and C (magenta). The representative structure is the MD snapshot within the smallest RMSD of the C_{α} atoms of residues 6–97 from the average structure of the basin. The representatives of basins A, B, and C differ from their respective average structure by RMSD values of 0.36 Å, 0.56 Å, and 0.47 Å, respectively (for distribution of RMSD values within basins see Supporting Information Fig. S5). (B) Snapshot of holo MD simulations showing the binding site groove with the bound peptide and indicating the locations of key residues used for the analysis of the tilting of helix $\alpha 2$ and the network of salt bridges. (C) Histograms of distances between pairs of C_{α} atoms in the binding site. The MD snapshots from all runs with value of the optimized RC in the range -38 to -25 , -8 to 6 , and 25 to 45 were used to calculate the histograms of basins A, B, and C, respectively. The colors used for the individual histograms are consistent with those used for the structures shown in the top, left. (D) Overlap of X-ray structures and MD snapshots representative of basins A and B. In all three figures, the representative snapshot of basin A is shown in red and the one of basin B in blue. The following X-ray structures are shown: (Left) 1BFE (apo) in orange and 1BE9 (holo) in green; (Middle) 1TQ3 (apo) in yellow, 1TP3 (holo) in ochre, and 1TP5 (holo) in iceblue; (Right) 3I4W (apo) in mauve, and 3K82 (apo) in silver. All structures were overlapped using the C_{α} atoms of residues 6–97.

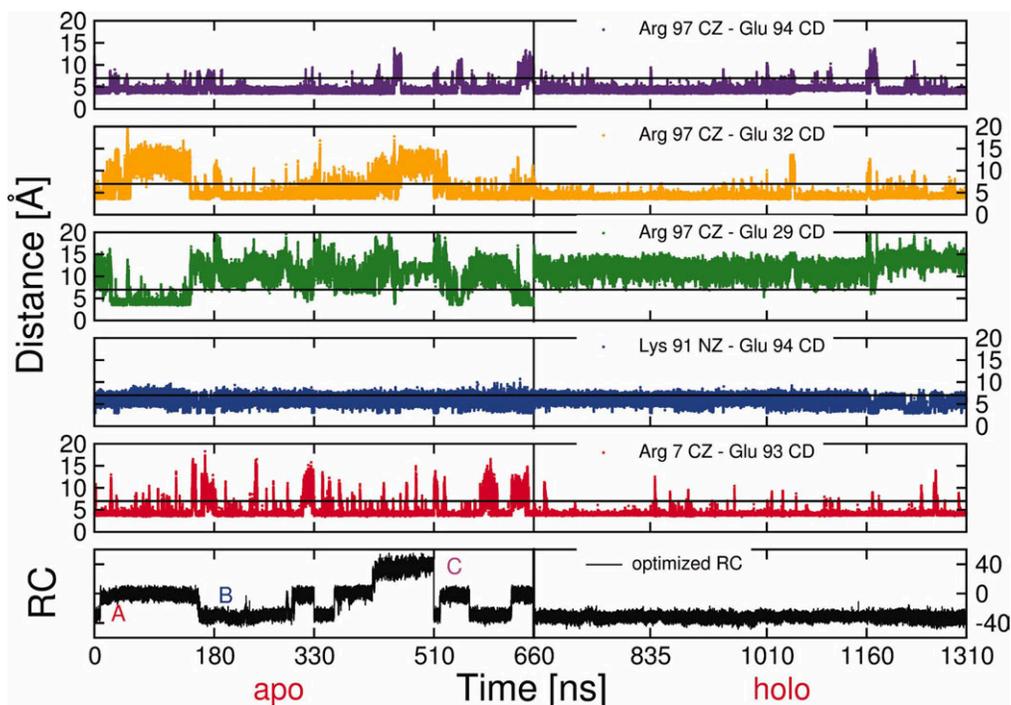


Figure 5

Salt bridges involving residues in the $\alpha 3$ helix. Time series of salt bridge distances between residues in helix $\alpha 3$ and the rest of the domain (color lines) are shown together with the time series of the optimized RC (black line, bottom panel). The black vertical lines at 660 ns indicate the concatenation point of apo and holo trajectories and the tick marks along the time axis indicate the individual apo and holo MD runs. The black horizontal lines at distance = 7 Å is drawn to discriminate between formation and rupture of salt bridges. This rather large threshold takes also into account single water-bridged polar interactions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

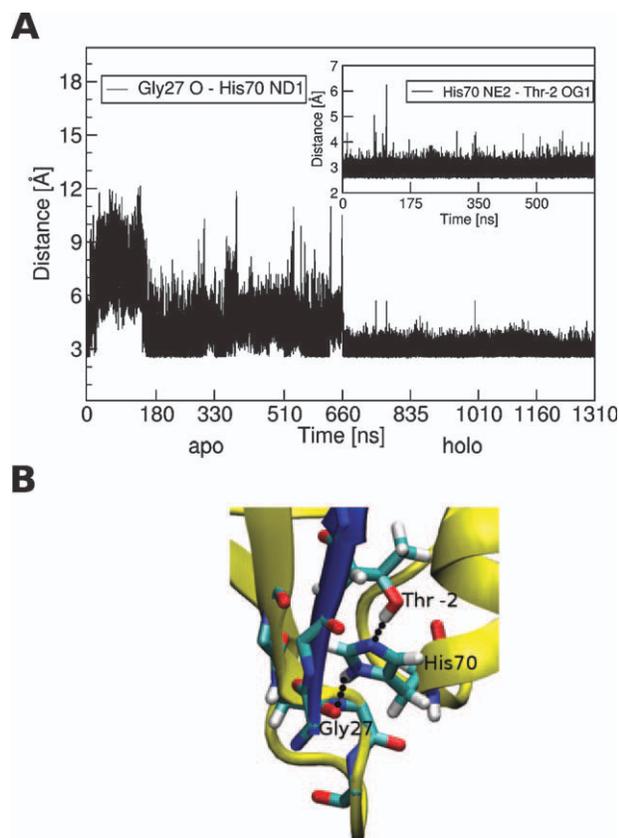
always present and their behavior is indistinguishable between ligand-free and complexed state. It is clear from Figure 5 that the formation/rupture of these salt bridges is sometimes but not always concomitant with changes in the optimized RC, that is, transitions between free energy basins.

As noted above, one of the main structural differences between the three basins is the position and conformation of the $\beta 2$ – $\beta 3$ loop. In this context, the salt bridge Arg97-Glu29 plays an important role, as its formation or rupture has a direct impact on the conformation of this loop. While this salt bridge is almost never formed in the complexed state, it spontaneously forms several times during the apo simulations, particularly in basin B (Fig. 5), thereby directly modifying the conformation of the $\beta 2$ – $\beta 3$ loop. This simulation result is interesting because experimental studies have provided evidence of the involvement of the $\beta 2$ – $\beta 3$ loop in binding. Using 2D and 3D heteronuclear NMR spectroscopy, Kozlov et al.⁵ have reported that this loop is directly involved in the interactions between the peptide ligand and human PDZ2. In addition, Liu et al.¹⁰ have observed by NMR experiments a high conformational flexibility for the $\beta 2$ – $\beta 3$ loop of LARG PDZ in the ligand-free state in contrast to relative rigidity in the complexed state, and have

speculated that conformational plasticity might facilitate binding of different peptide ligands. Our MD simulations indicate that the chain of hydrogen bonds involving the carbonyl oxygen of Gly27, the imidazole ring of His70, and the hydroxyl oxygen of Thr-2 in the ligand, originally observed in the X-ray structure of the PDZ3/peptide complex,³ restricts the $\beta 2$ – $\beta 3$ loop's flexibility in the complexed state and keeps it fixed in a single conformation. This chain of hydrogen bonds is present throughout the holo MD simulations (Fig. 6). Furthermore, it has been shown experimentally, that the mutation of Gly27 into an alanine decreases the peptide affinity by almost 100-fold, which is most probably due to the bulkier side chain of alanine preventing the $\beta 2$ – $\beta 3$ loop from adopting the right conformation to form the hydrogen bond with His70.³⁶ Whether the flexibility of the $\beta 2$ – $\beta 3$ loop facilitates or hinders binding by providing conformers to which the peptide can not bind is not evident from the simulations.

Side chain flexibility

Some but not all residues of the PDZ3 binding site show enhanced side chain flexibility in the ligand-free state compared to the complexed state (Fig. 7). The con-

**Figure 6**

His70 hydrogen bonds. (A) Time series of distances between the atoms involved in the two hydrogen bonds formed by His70 in the crystal structure of the complexed state.³ The inset shows the intermolecular hydrogen bond. (B) β 2– β 3 loop and part of the peptide binding pocket taken from an arbitrary snapshot of holo MD simulations. The backbone of PDZ3 (yellow) and peptide ligand (blue) are shown in a ribbon representation while residues involved in hydrogen bonds with His70 are shown by sticks colored by atom type. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

trary is not observed. Three different χ_1 values are accessible to His70 in the ligand-free state while only the trans rotameric state is populated in the complex because, as aforementioned, the His70 side chain is directly involved in peptide binding (Fig. 6). Furthermore, the side chains of Asn24 and Lys78, which are partially exposed to solvent, show a decreased flexibility in the complexed state. While Asn24 forms a hydrogen bond with Gln-3 of the peptide ligand in the crystal structure of the complex,³ for Lys78 no such interactions have been suggested. Interestingly, the MD simulations reveal that the side chain amino group of Lys78 can act as donor in a hydrogen bond with the carbonyl oxygen of Ser-1 (see Supporting Information Fig. S6 for time series of PDZ3/peptide hydrogen bonds). The χ_1 -angle flexibility of Leu77 decreases upon peptide binding. Leu77 is part of the hydrophobic cavity of the binding site, as is Phe23, that

is filled by the side chain of Val0 of the peptide ligand. The crystal structure of the complexed state displays a hydrogen bond between the carbonyl oxygen of Leu77 and a guanidinium nitrogen of Arg16. However, the latter interaction is equally probable for ligand-free and complexed PDZ3 (see Supporting Information Fig. S6). Therefore, the decreased flexibility of Leu77 in holo state is most likely due to the further burial of Leu77 by the side chain of Val0. There are also residues in β 2 strand and α 2 helix that show similar flexibility in the holo and apo state. As an example, the χ_1 of Phe23 populates a single rotameric state because its phenyl ring is tightly packed in the hydrophobic part of the binding pocket.

DISCUSSION

We have investigated the differences between the complexed and peptide-free states of the third PDZ domain of the synaptic protein PSD-95 by multiple explicit solvent MD simulations and analysis of free energy surface. Two main results emerge from our simulation study. First, the free energy landscape of the apo native state is more complex than the one of the peptide-bound state. For apo PDZ3 there are three main basins separated by barriers of about 3–4 kcal/mol. The presence of multiple free energy basins in the native state of apo PDZ3 is consistent with the folding intermediate observed under native conditions by native-state hydrogen exchange experiments.³⁷ In contrast, the holo state consists of a single basin which implies that there is an entropic penalty in binding the peptide ligand, in agreement with the conformational entropy mechanism and dynamic allostery unveiled by NMR experiments of wild type and a C-terminal truncated form of PDZ3.¹⁴ Interestingly, the additional C-terminal helix α 3 (residues 91–97), which has been shown to modify side chain dynamics, plays also a structural role in binding by influencing the conformation of the β 2– β 3 loop (residues 28–33) via a network of salt bridges involving Lys91, Glu94, and Arg97 in the former, and Glu29 and Glu32 in the latter.

Second, the cFEP analysis shows that the holo state samples only one conformation, which is also one of the three conformations sampled by the apo state. Thus, the peptide ligand binds to PDZ3 by selection of one of the three conformations sampled by the apo state. As the present simulation study does not sample the binding process explicitly, binding to one of the two apo-only conformations followed by spatial rearrangement (to the conformation of basin A) cannot be conclusively excluded. Nevertheless, experimental evidence supports a one-step binding mechanism without rate-limiting conformational change.^{7,8} Therefore, the theoretical framework of the conformational selection binding mechanism is the one most in line with our observations. The different types of analysis of the MD trajectories provide con-

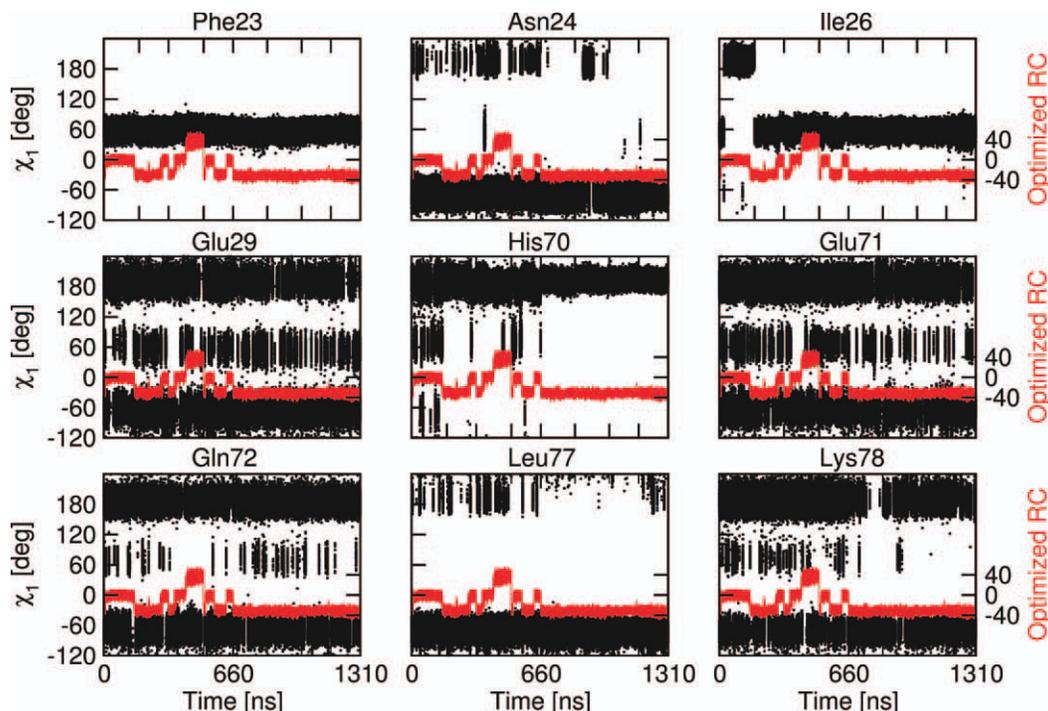


Figure 7

Flexibility of side chains. Time series of χ_1 -angles of binding site residues (black, y -axis on the left) together with the time series of the optimized RC (red, y -axis on the right). The tick mark at time = 660 ns indicates the separation of apo and holo trajectories, additional tick marks on the x -axis indicate the independent apo and holo MD runs. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

sistent evidence that the native conformation of PDZ3 is more heterogeneous in the apo state than the holo state. Moreover, the increased plasticity in the apo state is not restricted to side chain fluctuations, as suggested by NMR spectroscopy data,¹⁴ but involves also the backbone and in particular the relative orientation of the regular elements of secondary structure lining up the binding site ($\beta 2$ strand and $\alpha 2$ helix) as well as the $\beta 2$ – $\beta 3$ loop. The scissor-like motion of helix $\alpha 2$ and the thereby induced partial widening of the binding site groove takes place on a time scale of 10–100 ns and is reversible. In contrast, the single conformer of the peptide-bound state of PDZ3 is the one with the smallest tilting angle of the $\alpha 2$ helix with respect to the $\beta 2$ strand.

Importantly, conformational selection as peptide-binding mechanism for a PDZ domain is a novel suggestion. It remains to be investigated if it applies also to other PDZ domains. Up to now, experimental data on binding kinetics is available for some but not all PDZ domains, for example for the murine form of PDZ2⁸ and the second PDZ domain of SAP97.⁹ Nevertheless, in several cases where such data is not available, hypotheses for the binding mechanisms were made by analysis of crystal structures and/or NMR spectroscopy data of PDZ domains. The present simulation study shows that information on the free-energy landscape obtained by MD,

even if restricted to the native basin, can significantly alter the picture of the binding mechanism suggested on the basis of X-ray crystallography and/or NMR spectroscopy data.

It is interesting to compare the behavior of PDZ3 with other PDZ domains. The present MD simulations reveal that the conformation accessible to the complexed form of PDZ3 is the one with the smallest aperture of the hydrophobic cavity of the binding site groove, whereas the ensembles of NMR structures of apo and holo murine PDZ2 of PTP-BL⁸ show the exact contrary, that is, an opening of the binding site on peptide binding by PDZ2. Note also that the accessibility of different tilting angles/orientation of the $\alpha 2$ helix with respect to the $\beta 2$ strand has also been shown experimentally for other PDZ domains as LARG PDZ¹⁰ and the PDZ domain of Par-6.³⁸ However, the way this tilt is modified on peptide binding, and thereby widens or narrows the binding site groove, might depend not only on the specific PDZ domain itself but also on the bulkiness of the side chains of the peptide ligand.

ACKNOWLEDGMENTS

The authors thank Sergei Krivov for very interesting discussions. The simulations were carried out on the Schrödinger compute cluster of the University of Zurich.

REFERENCES

- Kim E, Sheng M. PDZ domain proteins of synapses. *Nat Rev Neurosci* 2004;5:771–781.
- Fan J-S, Zhang M. Signaling complex organization by PDZ domain proteins. *Neurosignals* 2002;11:315–321.
- Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* 1996;85:1067–1076.
- Walma T, Spronk CA, Tessari M, Aelen J, Schepens J, Hendriks W, Vuister GW. Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. *J Mol Biol* 2002;316:1101–1110.
- Kozlov G, Banville D, Gehring K, Ekiel I. Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the β 2- β 3 loop to PDZ domain-ligand interactions. *J Mol Biol* 2002;320:813–820.
- Zhang J, Sapienza PJ, Ke H, Chang A, Hengel SR, Wang H, Phillips GN, Lee AL. 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.
- Gianni S, Engström A, Larsson M, Calosci N, Malatesta F, Eklund L, Ngang CC, Travaglini-Allocatelli C, Jemth P. The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. *J Biol Chem* 2005;280:34805–34812.
- Gianni S, Walma T, Arcovito A, Calosci N, Belli A, Engström A, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure* 2006;14:1801–1809.
- Chi CN, Bach A, Engström A, Wang H, Strømgaard K, Gianni S, Jemth P. A sequential binding mechanism in a PDZ domain. *Biochemistry* 2009;48:7089–7097.
- Liu J, Zhang J, Yang Y, Huang H, Shen W, Hu Q, Wang X, Wu J, Shi Y. Conformational change upon ligand binding and dynamics of the PDZ domain from leukemia-associated Rho guanine nucleotide exchange factor. *Protein Sci* 2008;17:1003–1014.
- van den Berk LCJ, Landi E, Walma T, Vuister GW, Dente L, Hendriks JA. An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry* 2007;46:13629–13637.
- Zhang M. Scaffold proteins as dynamic switches. *Nat Chem Biol* 2007;3:756–757.
- Sohn J, Grant RA, Sauer RT. Allostery is an intrinsic property of the protease domain of DegS: implications for enzyme function and evolution. *J Biol Chem* 2010;285:34039–34047.
- Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci USA* 2009;106:18249–18254.
- Gerek ZN, Ozkan SB. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Comput Biol* 2011;7:e1002154.
- Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 1999;286:295–299.
- Kong Y, Karplus M. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins* 2009;74:145–154.
- Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* 2010;35:539–546.
- Silva D-A, Bowman GR, Sosa-Peinado A, Huang X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput Biol* 2011;7:e1002054.
- Ekonomiuk D, Caflisch A. Activation of the West Nile virus NS3 protease: molecular dynamics evidence for a conformational selection mechanism. *Protein Sci* 2009;18:1003–1011.
- Staneva I, Wallin S. Binding free energy landscape of domain-peptide interactions. *PLoS Comput Biol* 2011;7:e1002131.
- Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. *J Comput Chem* 2009;30:1545–1614.
- van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem* 2005;26:1701–1718.
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Jr., Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Więrkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998;102:3586–3616.
- Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
- Darden T, York D, Pedersen LG. Particle mesh Ewald: an $n \log(n)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089.
- Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
- Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 2007;23:2625–2627.
- Seeber M, Felling A, Raimondi F, Muff S, Friedman R, Rao F, Caflisch A, Fanelli F. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* 2011;32:1183–1194.
- Vitalis A, Caflisch A. Efficient construction of mesostate networks from molecular dynamics trajectories. *J Chem Theor Comput* 2012;8:1108–1120.
- Krivov SV, Karplus M. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J Phys Chem B* 2006;110:12689–12698.
- Krivov SV, Karplus M. Diffusive reaction dynamics on invariant free energy profiles. *Proc Natl Acad Sci USA* 2008;105:13841–13846.
- Krivov SV, Muff S, Caflisch A, Karplus M. One-dimensional barrier preserving free-energy projections of a beta-sheet miniprotein: new insights into the folding process. *J Phys Chem B* 2008;112:8701–8714.
- Rao F, Caflisch A. The protein folding network. *J Mol Biol* 2004;342:299–306.
- Krivov SV. Is protein folding sub-diffusive? *PLoS Comput Biol* 2010;6:e1000921.
- Chi CN, Elfström L, Shi Y, Snäll T, Engström A, Jemth P. Reassessing a sparse energetic network within a single protein domain. *Proc Natl Acad Sci USA* 2008;105:4679–4684.
- Feng H, Vu N-D, Bai Y. Detection of a hidden folding intermediate of the third domain of PDZ. *J Mol Biol* 2005;346:345–353.
- Peterson FC, Penkert RR, Volkman Brian F, Prehoda KE. Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. *Mol Cell* 2004;13:665–676.