J|A|C|S
ARTICLES
Published on Web 09/09/2006

# Thermodynamic Basis for Promiscuity and Selectivity in Protein−Protein Interactions: PDZ Domains, a Case Study

Nathalie Basdevant,[‡] Harel Weinstein,[‡,§] and Marco Ceruso*[,†]

*Contribution from the Department of Chemistry, CUNY College of Staten Island, 2800 Victory Boulevard, Staten Island, New York 10314, Department of Physiology and Biophysics, and HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021*

Received February 3, 2006; E-mail: ceruso@mail.csi.cuny.edu

***Abstract:*** Like other protein−protein interaction domains, PDZ domains are involved in many key cellular processes. These processes often require that specific multiprotein complexes be assembled, a task that PDZ domains accomplish by binding to specific peptide motifs in target proteins. However, a growing number of experimental studies show that PDZ domains (like other protein−protein interaction domains) can engage in a variety of interactions and bind distinct peptide motifs. Such promiscuity in ligand recognition raises intriguing questions about the molecular and thermodynamic mechanisms that can sustain it. To identify possible sources of promiscuity and selectivity underlying PDZ domain interactions, we performed molecular dynamics simulations of 20 to 25 ns on a set of 12 different PDZ domain complexes (for the proteins PSD-95, Syntenin, Erbin, GRIP, NHERF, Inad, Dishevelled, and Shank). The electrostatic, nonpolar, and configurational entropy binding contributions were evaluated using the MM/PBSA method combined with a quasi-harmonic analysis. The results revealed that PDZ domain interactions are characterized by overwhelmingly favorable nonpolar contributions and almost negligible electrostatic components, a mix that may readily sustain promiscuity. In addition, despite the structural similarity in fold and in recognition modes, the entropic and other dynamical aspects of binding were remarkably variable not only across PDZ domains but also for the same PDZ domain bound to distinct ligands. This variability suggests that entropic and dynamical components can play a role in determining selectivity either of PDZ domain interactions with peptide ligands or of PDZ domain complexes with downstream effectors.

## Introduction

Protein−protein interaction domains are one of the most remarkable features in protein recognition. Grouped in families of structurally homologous proteins, protein−protein interaction domains such as FYVE, PH, PB, SH3, or PDZ participate in a wide variety of biological functions.[1] They do so by binding specific peptide sequence motifs within target proteins and assembling these proteins into supramolecular complexes generally endowed with a new biological function. However, one characteristic of protein−protein recognition is its degeneracy, i.e., the ability that certain protein−protein interaction domains have to bind more than one target sequence motif.[2] This property is also referred to in the literature as degenerate specificity, multivalent specificity, or simply promiscuity.

To investigate the thermodynamic aspects underlying the specific example of promiscuity in the protein−protein recognition function of such modules, we have chosen to focus on PDZ domains.[3] Named after the combination of Post-synaptic density PSD-95, Discs large Dlg, and Zona occludens-1 ZO-1—the first proteins in which PDZ domains have been identified—these structurally conserved domains are among the most common protein−protein interaction domains.[4−8] They consist of 80 to 100 amino acids and have a tertiary structure formed by six $\beta$-strands, $\beta$A through $\beta$F, and two $\alpha$-helices, $\alpha$A and $\alpha$B (Figure 1a).

PDZ domains generally recognize peptide motifs located at the C-terminus of other proteins. These bind to PDZ domains by what appears to be a $\beta$-sheet augmentation mechanism observed for other protein−protein interaction domains as well.[9−11] The canonical mode of peptide recognition by PDZ domains is shown in Figure 1a. One of the hallmarks of PDZ domain recognition is the carboxylate group of the last residue in the ligand protein (the $P_0$ position) interacting through several

† CUNY College of Staten Island.
‡ Department of Physiology and Biophysics, Weill Medical College of Cornell University.
§ HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Medical College of Cornell University.

(1) Cesareni, G.; Gimona, M.; Sudol, M.; Yaffe, M. *Modular Protein Domains*; Wiley: New York, 2004.
(2) Sudol, M. *Oncogene* **1998**, *17*, 1469−74.
(3) Doyle, D. A.; Lee, A.; Lewis, J.; Kim, E.; Sheng, M.; MacKinnon, R. *Cell* **1996**, *85*, 1067−76.
(4) Jelen, F.; Oleksy, A.; Smietana, K.; Otlewski, J. *Acta Biochim. Pol.* **2003**, *50*, 985−1017.
(5) Nourry, C.; Grant, S. G.; Borg, J. P. *Sci. STKE* **2003**, 2003, RE7.
(6) van Ham, M.; Hendriks, W. *Mol. Biol. Rep.* **2003**, *30*, 69−82.
(7) Sheng, M.; Sala, C. *Annu. Rev. Neurosci.* **2001**, *24*, 1−29.
(8) Kim, E.; Sheng, M. *Nat. Rev. Neurosci.* **2004**, *5*, 771−81.
(9) Cowburn, D. *Structure* **1996**, *4*, 1005−8.
(10) Cowburn, D. *Curr. Opin. Struct. Biol.* **1997**, *7*, 835−8.
(11) Harrison, S. C. *Cell* **1996**, *86*, 341−3.


**10.1021/ja060830y CCC: $33.50** 

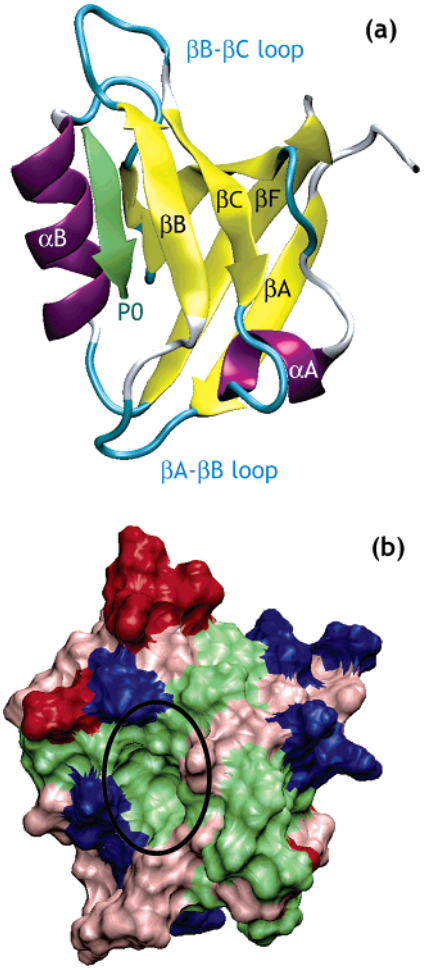


***Figure 1.*** Representative structure of a PDZ domain in complex with its target peptide. (a) Ribbon representation of the crystal structure of PDZ3 of PSD-95 in complex with a peptide in its binding site.[3] The peptide is represented in green, and the C-terminus of the peptide is indicated by $P_0$. (b) Molecular surface representation of PDZ3 of PSD-95 with basic residues in blue, acidic residues in red, polar residues in pink, and hydrophobic residues in green. The binding pocket for the peptide is indicated by a black circle. Figures created using VMD.[92]

hydrogen bonds with the main-chain amide protons of the $\beta$A−$\beta$B loop, also called the carboxylate-binding loop, that contains the conserved GLGF motif.[3] Two main classes of PDZ domains have been defined:[12−14] class I PDZ domains recognizing C-terminus peptides with a serine or threonine at $P_{-2}$, within a peptide sequence motif defined as −[X-S/T-X-Φ], where Φ is any hydrophobic amino acid and X is any amino acid; and class II PDZ domains that can bind peptides with any hydrophobic residue at $P_{-2}$ within a peptide sequence motif defined as −[X-Φ-X-Φ]. In addition, at least two more classes have been defined, which correspond to different sequence motifs in the C-terminal peptides,[4,5,15] but there is no consensus on this classification.[16,17]

The difficulties encountered in the various attempts at classifying PDZ domains are but an illustration of the remarkable promiscuity that seems to characterize PDZ domain interactions. Indeed, some PDZ domains have more than one biological target, such as the first NHERF PDZ domain which binds both the NDSLL and EDSFL peptide motifs located at the C-terminus of the $\beta$2-adrenergic and Platelet-derived Growth Factor receptors, respectively.[18] Other PDZ domains exhibit promiscuity by binding to peptide sequence motifs that belong to more than one class. For example, the PDZ domain of Erbin binds the class I C-terminus of p0071/delta-catenin protein[19] and the class II C-terminus of ErbB2.[20] Similarly, the second PDZ domain of Syntenin recognizes both the class I motif at the C-terminus of the IL5 receptor ($\alpha$ chain) and the class II motif at the C-terminus of syndecan.[21] Other examples of this dual-class promiscuity include PDZ3 and PDZ5 of CIPP1,[16,22] the PDZ domain of PICK1,[23] PDZ3 of hINADL,[17] and PDZ1 of MINT1 and Par6 PDZ domains.[16] Finally, it has been shown that the PDZ domain of syntrophin can recognize internal peptide sequences that lack a free carboxylate,[24,25] and a binding site distinct from the canonical binding site in the seventh PDZ domain of GRIP has also been proposed.[26]

Much information is available from experiments concerning the binding of PDZ domains. But there is considerably less data from measurements of binding affinities for PDZ domains than for many of the other modular domains, such as the SH2 domains. Moreover, measured binding affinities for PDZ domains are often not comparable due to differences in experimental conditions. The study of the thermodynamics of recognition of C-terminal peptides by PDZ domains utilizing computational methods can therefore offer an important complementation to the sparse experimental data.

From the computational studies presented in this work, we aim to understand the origins of promiscuity and selectivity in peptide/PDZ domain recognition by examining various contributions to the energy in the binding interactions of known complexes of PDZ domains. In order to evaluate these energy contributions, namely the electrostatic, nonpolar, and configurational entropy components, a set of distinct PDZ domains complexes was subjected to molecular dynamics (MD) simulations and the resulting trajectories were analyzed using the MM/PBSA method (Molecular Mechanics/Poisson−Boltzmann Surface Area). This set includes a variety of same-class and interclass pairs: 8 different PDZ domains in complex with different target peptides, for a total of 12 different complexes for which 3D-structures were available. The traditional MM/PBSA approach was combined with a quasi-harmonic analysis to estimate the configurational entropies of molecules. In

(12) Songyang, Z.; Fanning, A. S.; Fu, C.; Xu, J.; Marfatia, S. M.; Chishti, A. H.; Crompton, A.; Chan, A. C.; Anderson, J. M.; Cantley, L. C. *Science* **1997**, *275*, 73−7.
(13) Stricker, N. L.; Christopherson, K. S.; Yi, B. A.; Schatz, P. J.; Raab, R. W.; Dawes, G.; Bassett, D. E., Jr.; Bredt, D. S.; Li, M. *Nat. Biotechnol.* **1997**, *15*, 336−42.
(14) Tochio, H.; Zhang, Q.; Mandal, P.; Li, M.; Zhang, M. *Nat. Struct. Biol.* **1999**, *6*, 417−21.
(15) Harris, B. Z.; Lim, W. A. *J. Cell. Sci.* **2001**, *114*, 3219−31.
(16) Bezprozvanny, I.; Maximov, A. *FEBS Lett.* **2001**, *509*, 457−62.
(17) Vaccaro, P.; Dente, L. *FEBS Lett.* **2002**, *512*, 345−9.
(18) Karthikeyan, S.; Leung, T.; Ladias, J. A. *J. Biol. Chem.* **2002**, *277*, 18973−8.
(19) Laura, R. P.; Witt, A. S.; Held, H. A.; Gerstner, R.; Deshayes, K.; Koehler, M. F.; Kosik, K. S.; Sidhu, S. S.; Lasky, L. A. *J. Biol. Chem.* **2002**, *277*, 12906−14.
(20) Birrane, G.; Chung, J.; Ladias, J. A. *J. Biol. Chem.* **2003**, *278*, 1399−402.
(21) Kang, B. S.; Cooper, D. R.; Devedjiev, Y.; Derewenda, U.; Derewenda, Z. S. *Structure (Camb)* **2003**, *11*, 845−53.
(22) Kurschner, C.; Mermelstein, P. G.; Holden, W. T.; Surmeier, D. J. *Mol. Cell. NeuroSci.* **1998**, *11*, 161−72.
(23) Madsen, K. L.; Beuming, T.; Niv, M. Y.; Chang, V.; Dev, K. K.; Weinstein, H.; Gether, U. *J. Biol. Chem.* **2005**.
(24) Hillier, B. J.; Christopherson, K. S.; Prehoda, K. E.; Bredt, D. S.; Lim, W. A. *Science* **1999**, *284*, 812−5.
(25) Oschkinat, H. *Nat. Struct. Biol.* **1999**, *6*, 408−10.
(26) Feng, W.; Fan, J. S.; Jiang, M.; Shi, Y. W.; Zhang, M. *J. Biol. Chem.* **2002**, *277*, 41140−6.

***Table 1.*** PDZ Domains Used for MD Simulations

| protein | PDZ | class | PDB code | residue IDs | ligand length | ligand sequence[a] |
|---|---|---|---|---|---|---|
| NHERF | PDZ1 | Apo | 1G9O[93] | A:10−94 | | |
| | | I | 1GQ5[18] | A:10−94/A:95−99 | 5 | -EDSFL (“a” pep.) |
| | | I | 1GQ4[18] | A:10−94/A:95−99 | 5 | -NDSLL (“b” pep.) |
| | | I | 1I92[33] | A:10−94/A:95−99 | 5 | -QDTRL (“c” pep.) |
| PSD-95 | PDZ3 | Apo | 1BFE[3] | A:306−415 | | |
| | | I | 1BE9[3] | A:306−415/B:5−9 | 5 | -KQTSV |
| Dishevelled | | Apo | 1L6O[32] | A:254−340 | | |
| | | I | 1L6O[32] | A:254−340/D:1−8 | 8 | -SLKLMTTV |
| Shank[b] | | Apo | 1Q3O[34] | A:584−686 | | |
| | | I | 1Q3P[34] | A:584−686/C:1−6 | 6 | -EAQTRL |
| Syntenin | PDZ2 | Apo | 1NTE[21] | A:197−270 | | |
| | | I | 1OBX[21] | A:197−270/B:5−8 | 4 | -(ETLE)DSVF |
| | | II | 1OBY[21] | B:197−270/Q:1−6 | 6 | -TNEFYA |
| Erbin | | Apo | 1MFG[20] | A:1280−1367 | | |
| | | I | 1N7T[30] | model 2 A:12−99/B:301−307 | 7 | -TGWETWV |
| | | II | 1MFG[20] | A:1280−1367/B:1247−1255 | 9 | -EYLGLDVPV |
| GRIP | PDZ6 | Apo | 1N7E[94] | A:668−753 | | |
| | | II | 1N7F[94] | A:668−753/B:1−8 | 8 | -ATVRTYSC |
| Inad | PDZ1 | Apo | 1IHJ[31] | A:12−105 | | |
| | | II | 1IHJ[31] | A:12−105/D:3−7 | 5 | -(GK)TEFCA |

[a] Residues in parentheses are disordered in the crystal structure and were not included in the MD simulations. [b] The coordinates of residues 602 to 621 in the holo structure were taken from the crystal structure of the apo form of Shank (PDB code: 1Q3O); see “Molecular Systems” subsection in Materials and Methods.

addition, some of the dynamic effects associated with binding that could not be captured in the MM/PBSA calculations because they are based exclusively on the MD trajectories of the PDZ/peptide complex were estimated through the calculation of an “adaptation free energy” defined separately for each PDZ domain and each ligand as the difference in free energies between the bound and free state of each molecule.[27,28] Computing the adaptation free energy required separate MD simulations for each PDZ domain and each peptide ligand.

The results of these computational experiments show that nonpolar contributions dominate the thermodynamics of the interaction between PDZ domains and their ligands. This was in marked contrast with the contribution of the electrostatic component that was almost consistently insignificant. The configurational entropy contribution, which reflects the degree to which the dynamics of the PDZ domain and the peptide ligand are coupled in the complex, was always unfavorable to binding. Moreover, it was found to be very variable across PDZ domains, even for complexes involving the same PDZ domain but different peptides. But when comparing complexes involving the same PDZ domain, the configurational entropy contribution did not usually alter the binding preferences set by the non-entropic free energy component (corresponding to the sum of nonpolar and electrostatic contributions). Finally, the evaluation of the various configurational entropy contributions and adaptation free energies revealed several remarkable features about the events that make up the binding reaction. Notably, we found that upon binding all peptide ligands underwent significant ordering (loss of configurational entropy), but while an increase in configurational order was also observed for the majority of the PDZ domains, in 5 out of the 12 complexes, the PDZ domains were as ordered or less ordered in the bound state than in the free (apo) state.

Based on these findings we argue that the combined trends observed for the nonpolar, electrostatic, and configurational entropy contributions are poised to sustain the promiscuous interactions in which PDZ domains engage. Furthermore, it is proposed here that the remarkable variability in the degree to which the dynamics of the peptide and the cognate PDZ domain are coupled in the complex, as well as the varied ordering response of PDZ domains, suggests that entropic contributions could play a role in determining the specificity of the interactions between the PDZ domain and its target ligand or of the bound PDZ/peptide ligand toward a downstream effector.

## Materials and Methods

**Molecular Systems.** The PDZ domains selected for this study are listed in Table 1. For each domain, independent molecular dynamics (MD) simulations were performed for the apo (without a bound peptide) and holo (in complex with a peptide) forms of the PDZ domain and for the isolated peptide ligand. Thus, a total of 12 holo systems, 12 isolated peptide ligands, and 8 apo systems were prepared for MD simulations. The starting coordinates for all systems were taken from known 3D structures of PDZ domains deposited in the Protein Databank[29] and subjected to a few modifications as detailed below. All 3D-structures were obtained from X-ray diffraction experiments except for Erbin 1N7T,[30] which was an NMR structure (model 2 was used, as it is considered by the authors to be the “best representative structure” in the ensemble).

When the sequence length of the apo and holo forms differed in the crystal structures, a few residues at the N- and/or C-termini of the PDZ domains were deleted in order to maintain the same sequence in both forms. The simulated chain segments are defined in Table 1 for each system.

(27) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. *Acc. Chem. Res.* **2000**, *33*, 889−97.
(28) Reyes, C. M.; Kollman, P. A. *J. Mol. Biol.* **2000**, *297*, 1145−58.
(29) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−42.
(30) Skelton, N. J.; Koehler, M. F.; Zobel, K.; Wong, W. L.; Yeh, S.; Pisabarro, M. T.; Yin, J. P.; Lasky, L. A.; Sidhu, S. S. *J. Biol. Chem.* **2003**, *278*, 7645−54.

For the Erbin,[20] Inad,[31] and Dishevelled[32] PDZ domains, no apo structure was known; therefore, the starting structure for the apo simulations was built from the corresponding holo structure by deletion of the peptide. The three crystal structures of the NHERF PDZ domain used in this study contain a non-natural C-terminus which corresponds to the C-terminus of a known interacting protein.[18,33] This C-terminus is bound to the neighboring PDZ molecule in the crystal, thereby mimicking a peptide/PDZ domain interaction. Thus, the holo starting structures for the NHERF PDZ domains were reconstructed from the crystal symmetry by manually excising the peptide ligand from the C-terminus of a neighboring PDZ domain. In the Shank PDZ domain,[34] the coordinates of residues 610 to 614, located in the $\beta$B-$\beta$C loop, are missing in the crystal structure of the holo form. In order to build the coordinates for these missing residues, the apo and holo forms of Shank were superimposed and residues 602 to 621 from the apo form were incorporated in the initial holo structure. Since the structure of the $\beta$B−$\beta$C loop differs in the apo and holo forms, residues flanking both sides of the missing segment were incorporated to include at least one structurally overlapping residue. Minimization and equilibration (see next section) were performed with positional restraints on the rest of the molecule in order to relax the resulting holo structure.

Finally, the starting structures for the simulations of the peptide ligands were extracted from the 3D structure of the complexes. A charged ammonium group ($NH_3^+$) was added at the N-terminus of all peptide ligands in simulations of complexes and peptides alone. Note that, although the N-termini segments of some peptide ligands were disordered in the crystal structure, the coordinates for these missing residues were not reconstituted. This was the case for Syntenin's 1OBX[21] and Inad's 1IHJ,[31] in which the coordinates of four and two residues, respectively, are missing. In order to understand whether the presence of the ammonium group could affect our analyses, we performed a control simulation on the Syntenin's 1OBX complexed with a peptide capped by a neutral acetyl group instead of the charged ammonium group. We found that the resulting trajectory and the calculated energies were nearly identical to the simulation with the ammonium group (data not shown).

**Molecular Dynamics Simulations.** All MD simulations were performed with the GROMACS simulation package[35,36] using the OPLS all-atom force-field.[37] Each protein system was immersed in a pre-equilibrated cubic box of simple point-charge (SPC) water molecules.[38] The distance between the edge of the box and any protein atom was at least 0.8 nm. When present, crystallographic waters were also included in the initial protein system. Sodium or chlorine counterions were added to neutralize the systems when necessary. All MD simulations were performed using periodic boundary conditions in the isobaric−isotherm ensemble, at 300 K and 1 bar. The temperature and the pressure were maintained using the Berendsen algorithm[39] with coupling constants of $\tau_T = 0.1$ ps and $\tau_P = 1$ ps for the temperature and pressure, respectively. Twin-range nonbonded cutoffs of 0.9 and 1.2 nm were used for the Lennard−Jones potentials. Electrostatic interactions were calculated using the Particle−Mesh Ewald (PME) summation method[40,41] with a cutoff of 0.9 nm in the direct space. The pair lists were updated every 10 steps. The Lincs algorithm[42] was used to constrain the bond lengths and bond angles between atoms, thus allowing an integration time step of 2 fs.

Initially, all heavy atoms were restrained to their crystallographic positions using a harmonic potential (force constant equal to 1000 kJ/mol/nm$^2$), while surrounding water molecules were first minimized and then subjected to 50 ps of MD simulations (NPT, 300 K, 1 bar). Then, the resulting system was energy minimized, without any restraints, and the MD simulations were carried out for 20 or 25 ns depending on the equilibration time (see Results).

In the case of Shank (holo structure only), in order to relax and equilibrate the coordinates for residues 602 to 621 that had been manually incorporated (see previous section), the 50 ps restrained MD run was decomposed into five 10-ps-long restrained MD runs in which the restraining force constant for the heavy atoms in residues 602 to 621 was reduced by 250 kJ/mol/nm$^2$ in each restrained MD run (initially set at 1000 kJ/mol/nm$^2$ in the first run) while keeping the restraining force constant at 1000 kJ/mol/nm$^2$ on the remaining heavy atoms.

**Binding Free Energy.** Binding free energies were approximated using the MM/PBSA approach.[27,43−47] The MM/PBSA method is an efficient way to evaluate the binding free energy of a complex from a classic MD simulation in explicit water. It has been applied to many different systems, including other protein recognition domains (e.g., the SH3 domains[48]), and the affinities derived from the method have been compared successfully to experimental values.[43,49] A variant of this method, termed MM/GBSA, in which the implicit Generalized Born (GB) solvent model replaces the PB calculation, has been introduced recently.[50]

In the MM/PBSA approach, the binding free energy, $\Delta G_b$, associated with the binding of a PDZ domain (P) to its cognate peptide ligand (L) to form a protein/ligand complex (P::L) is written

$$\Delta G_b = \langle G^{P::L}_{MMPBSA}\rangle - \langle G^{P}_{MMPBSA}\rangle - \langle G^{L}_{MMPBSA}\rangle \qquad (1)$$

where $\langle G_{MMPBSA}\rangle$ is an approximation of the free energy of a species in solution given by

$$\langle G_{MMPBSA}\rangle = \langle E_{MM}\rangle + \langle \Delta G^{np}_{sol}\rangle + \langle \Delta G^{ele}_{sol}\rangle - TS \qquad (2)$$

with $\langle E_{MM}\rangle$, the molecular mechanical energy of the species in vacuum averaged over snapshot structures extracted from a molecular dynamics trajectory; $\langle \Delta G^{np}_{sol}\rangle$ and $\langle \Delta G^{ele}_{sol}\rangle$, the nonpolar and electrostatic contributions to the solvation free energy of the species also computed as averages over snapshot structures extracted from a molecular dynamics trajectory; and $S$, the entropy of the molecular species in a vacuum (see "Computational Details" for estimate of the entropy).

**Adaptation Free Energy.** The adaptation free energy of a given molecule, either a PDZ domain or a peptide ligand, was defined as the difference in free energy between the bound (holo) and free (apo) states of that molecule.[27,28] Thus, for a given PDZ domain, the adaptation free energy associated with its binding reaction in an aqueous solution was defined as


(31) Kimple, M. E.; Siderovski, D. P.; Sondek, J. *EMBO J.* **2001**, *20*, 4414−22.
(32) Cheyette, B. N.; Waxman, J. S.; Miller, J. R.; Takemaru, K.; Sheldahl, L. C.; Khlebtsova, N.; Fox, E. P.; Earnest, T.; Moon, R. T. *Dev. Cell* **2002**, *2*, 449−61.
(33) Karthikeyan, S.; Leung, T.; Ladias, J. A. *J. Biol. Chem.* **2001**, *276*, 19683−6.
(34) Im, Y. J.; Lee, J. H.; Park, S. H.; Park, S. J.; Rho, S. H.; Kang, G. B.; Kim, E.; Eom, S. H. *J. Biol. Chem.* **2003**, *278*, 48099−104.
(35) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comp. Phys. Commun.* **1995**, *91*, 43−56.
(36) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306−317.
(37) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(38) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, 1981; pp 331−342.
(39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(40) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(41) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(42) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(43) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786−3791.
(44) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(45) Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D. *J. Med. Chem.* **2003**, *46*, 5628−37.
(46) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(47) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 671−82.
(48) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Luo, R.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 3986−94.
(49) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221−30.
(50) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238−50.

$$\Delta G_{adapt} = \langle G^{P\_holo}_{MMPBSA} \rangle - \langle G^{P\_apo}_{MMPBSA} \rangle \quad (3)$$

and similarly for the peptide ligand

$$\Delta G_{adapt} = \langle G^{L\_bound}_{MMPBSA} \rangle - \langle G^{L\_free}_{MMPBSA} \rangle \quad (4)$$

**Computational Details.** Analysis of the results from the MD simulations was carried out with various algorithms implemented in the program METAPHORE (in-house software). All analyses were performed using the last 15 ns of the equilibrated trajectories. Equilibration was evaluated from the time evolution of the radius of gyration and the root-mean-square deviation (rmsd) of the structures from their initial configuration.

**A. Binding and Adaptation Free Energy**. All contributions to the free energy of binding and adaptation, with the exception of the corresponding entropy contributions, were calculated as averages of instantaneous values computed for snapshot structures extracted from the MD trajectories. The snapshot structures were taken every 10 ps over the last 15 ns of the trajectory (1500 structures for each molecular system: protein, ligand, or protein/ligand complex).

Contributions to $\langle E_{MM} \rangle$ were calculated using the potential energy terms of the OPLS force field as implemented in the GROMACS package. The van der Waals, $E_{vdw}$, and Coulomb, $E_{coul}$, energies were calculated without periodic boundary conditions or a cutoff for long-range interactions.

The solvation free energy terms were computed following the MM/PBSA approach. For each structure, the nonpolar component of the solvation free energy $\Delta G^{np}_{sol}$ was taken to be linearly dependent on the solvent accessible surface area, $A$, of the solute[51]

$$\Delta G^{np}_{sol} \approx \gamma A + b \quad (5)$$

with $\gamma = 0.0054$ kcal/mol and $b = 0.92$ kcal/mol.[27,51] The solvent accessible surface area of the solute molecules was calculated using the MSMS program.[52] The PARSE radii were used to define the atomic radii of the solutes, and the radius of the probe sphere was taken as 1.4 Å to be consistent with the electrostatic part of the solvation free energies (see below).

The electrostatic component of the solvation free energy, $\Delta G^{ele}_{sol}$, was calculated with the program Delphi,[53,54] which solves the Poisson−Boltzmann equation numerically for a set of fixed charges in a cavity. We used the PARSE parameter set[27,51] for the radii of the solute atoms used to define the solute cavity; the atomic charges were taken from the OPLS force field (to be consistent with the calculation of $E_{coul}$). The interior dielectric constant for the protein was set to $\epsilon_i = 4$. The exterior dielectric constant was set to $\epsilon_{sol} = 80$ for water, and the dielectric boundary was calculated using a spherical probe with a 1.4 Å radius. The grid was chosen with 3 points per Å, and the solute molecule occupied 80% of the box.

**B. Configurational Entropy**. Only the configurational part of the entropy was included in the entropic contribution to the free energy. Thus, the entropic changes associated with binding and adaptation were calculated as $-T\Delta S_b \approx -T(S^{P::L}_{conf} - S^{P}_{conf} - S^{L}_{conf})$ for the binding and $-T\Delta S^{P}_{adapt} \approx -T(S^{P\_holo}_{conf} - S^{P\_apo}_{conf})$ and $-T\Delta S^{L}_{adapt} \approx -T(S^{L\_bound}_{conf} - S^{L\_free}_{conf})$ for the adaptation of the protein (P) and the ligand (L), respectively.

The configurational entropy of each system (protein, ligand, and protein/ligand complex) was estimated using the quasi-harmonic analysis.[55,56] This method has been used previously to evaluate the entropy of peptides and proteins.[57] The quasi-harmonic analysis yields an upper bound approximation, $S_{ho}$, to the real configurational entropy, $S$, of a molecule based on the all-atom covariance matrix that can be calculated from an MD trajectory. For a molecule of $N$ atoms, $S_{ho}$ is given by the following equation:

$$S \le S_{ho} = k_B \sum_{i=1}^{3N-6} \left[ \frac{\gamma}{e^{\gamma} - 1} - \ln(1 - e^{-\gamma}) \right] \quad (6)$$

where $\gamma = h/2\pi\sqrt{1/k_B T\lambda_i}$, $h$ is the Planck constant, $k_B$ is the Boltzmann constant, $T$ is the temperature, and $\lambda_i$ are the eigenvalues of the all-atom mass-weighted covariance matrix of fluctuations $(\sigma_{ij})_{ij}$, where $\sigma_{ij} = \sqrt{m_i m_j}\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$.

Finally, when calculating the entropic contribution to the free energy of binding, $-T\Delta S_b$, the structures used to compute the covariance matrix of fluctuations for each individual system (protein, ligand, protein/ligand complex) were all extracted from the MD trajectory of the corresponding protein/ligand complex (similarly as the MM/PBSA binding free energy calculations). This calculation does not take into account the loss or gain of self-entropy associated with the rotational, translational, and conformational changes of each component upon binding. However, it yields the configurational entropy due to the motions of the peptide and the PDZ domain with respect to each other and to the cross-terms from correlated motions between the PDZ domain and the peptide. Thus, a high entropic cost of binding indicates a high degree of coupling between the dynamics of the PDZ domain and the bound peptide.

## Results

**MM/PBSA.** The components of the binding free energies for the 12 complexes were calculated with the MM/PBSA method associated with a quasi-harmonic approximation for the configurational entropy. The snapshot structures used for these calculations were extracted from the equilibrated portions of Molecular Dynamics (MD) trajectories carried out in the presence of explicit solvent molecules. MD simulations were computed for 12 holo PDZ domains, 12 unbound peptide ligands, and 8 apo PDZ domains, listed in Table 1. All simulations were performed using the same conditions and with the same protocol (see Materials and Methods).

The equilibration of the MD trajectories was monitored from the convergence of the plots of the radius of gyration as a function of time and the time dependence of the root-mean-square deviation of Cα carbon atoms (Cα-rmsd) from their initial configuration. The equilibration time varied among systems, but it was particularly slow, ranging from 3 to 10 ns, for the apo PDZ domains of Inad and Dishevelled for which the starting structures had been extracted from the crystal structure of the corresponding holo PDZ domain. These two apo systems were therefore simulated for 25 ns instead of 20 ns for all the other systems, in order to ensure an equilibrated portion of the MD trajectory of at least 15 ns long. All analyses were thus performed on the last 15 ns of the trajectories.

It should be noted that long simulations are rarely used to evaluate the binding free energy using surface area continuum solvent methods. To our knowledge, only one case has been reported of an MM/GBSA (with a solvation term calculated using Generalized Born and Surface Area) analysis over a 10 ns window of a 12 ns MD simulation of a protein−protein complex.[50] That study showed that averaging over such a long time window was necessary because of the large fluctuations

(51) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(52) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−20.
(53) Gilson, M. K.; Honig, B. *Proteins* **1988**, *4*, 7−18.
(54) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−9.
(55) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289−6292.
(56) Tidor, B.; Karplus, M. *J. Mol. Biol.* **1994**, *238*, 405−414.
(57) Pohlmann, T.; Bockmann, R. A.; Grubmuller, H.; Uchanska-Ziegler, B.; Ziegler, A.; Alexiev, U. *J. Biol. Chem.* **2004**, *279*, 28197−201.

**Table 2.** Binding Free Energies of PDZ Domains with Their Target Peptides in kcal/mol Calculated with the MM/PBSA Method[a]

| protein | $\Delta E_{vdw}$ | $\Delta E_{coul}$ | $\Delta\Delta$ | $G_{sol}^{ele}\Delta\Delta G_{sol}^{np}$ | $\Delta G_b^{ele\,b}$ | $\Delta G_b^{np\,c}$ | $-T\Delta S_b{}^c$ | $\Delta G_b{}^e$ |
|---|---|---|---|---|---|---|---|---|
| NHERF[a] | −43.2 (*4.3*) | −56.6 (*7.4*) | 56.6 (*6.5*) | −5.9 (*0.4*) | 0.0 (*1.7*) | −49.1 (*4.3*) | 9 (*2*) | **−40.1** (*4.9*) |
| NHERF[b] | −42.2 (*4.0*) | −36.4 (*4.3*) | 38.1 (*3.3*) | −5.7 (*0.9*) | 1.7 (*1.7*) | −47.9 (*4.2*) | 7 (*3*) | **−39.1** (*6.4*) |
| NHERF[c] | −43.8 (*4.3*) | −40.2 (*7.0*) | 42.7 (*6.1*) | −6.2 (*0.4*) | 2.5 (*1.7*) | −50.0 (*4.4*) | 14 (*2*) | **−33.6** (*5.6*) |
| PSD-95 | −42.1 (*4.1*) | −87.1 (*14.9*) | 87.3 (*13.1*) | −5.8 (*0.2*) | 0.2 (*2.5*) | −47.9 (*4.1*) | 9 (*3*) | **−38.8** (*6.2*) |
| Dishevelled | −61.7 (*4.3*) | −65.5 (*9.2*) | 68.2 (*9.0*) | −7.5 (*0.3*) | 2.7 (*1.5*) | −69.2 (*4.3*) | 22 (*3*) | **−44.4** (*7.0*) |
| Shank | −36.4 (*4.8*) | −38.4 (*9.3*) | 39.5 (*8.1*) | −5.6 (*1.0*) | 1.1 (*1.9*) | −42.0 (*4.9*) | 45 (*2*) | **+4.1** (*5.8*) |
| Syntenin I | −35.5 (*4.5*) | −37.3 (*8.1*) | 38.6 (*6.5*) | −5.0 (*0.9*) | 1.3 (*2.2*) | −40.4 (*4.5*) | 8 (*1*) | **−31.1** (*3.9*) |
| Syntenin II | −38.7 (*3.8*) | −25.4 (*3.8*) | 27.4 (*3.0*) | −5.0 (*0.1*) | 2.0 (*1.4*) | −43.7 (*3.8*) | 5 (*3*) | **−36.7** (*5.5*) |
| Erbin I | −62.4 (*5.1*) | −39.8 (*11.5*) | 43.6 (*9.1*) | −7.4 (*0.3*) | 3.9 (*3.1*) | −69.8 (*5.1*) | 12 (*2*) | **−53.9** (*6.9*) |
| Erbin II | −50.9 (*6.4*) | −44.3 (*5.5*) | 48.2 (*4.9*) | −7.0 (*1.0*) | 3.9 (*2.4*) | −57.9 (*6.6*) | 23 (*2*) | **−31.0** (*7.4*) |
| GRIP | −46.2 (*4.6*) | −60.0 (*6.1*) | 60.7 (*5.3*) | −6.9 (*0.5*) | 0.8 (*1.7*) | −53.1 (*4.6*) | 24 (*4*) | **−28.3** (*7.9*) |
| Inad | −33.0 (*3.7*) | −29.6 (*3.8*) | 32.2 (*3.2*) | −4.5 (*1.0*) | 2.5 (*1.6*) | −37.5 (*3.8*) | 5 (*2*) | **−30.0** (*5.2*) |
| Average | **−44.7** (*9.5*) | **−46.7** (*17.5*) | **48.6** (*17.0*) | **−6.0** (*1.0*) | **1.9** (*1.3*) | **−50.7** (*10.4*) | **15** (*12*) | **−33.6** (*13.9*) |

[a] Standard deviations of averages are shown in parentheses. The last line is the average of each column and its corresponding standard deviation. [b] $\Delta G_b{}^{ele} = \Delta E_{coul} + \Delta\Delta G_{sol}{}^{ele}$. [c] $\Delta G_b{}^{np} = \Delta E_{vdw} + \Delta\Delta G_{sol}{}^{np}$. [d] $-T\Delta S_b$ calculated with the quasi-harmonic approximation on the complex trajectory. [e] $\Delta G_b = \Delta G_b{}^{elec} + \Delta G_b{}^{np} - T\Delta S$.

observed in the computed free energies. We have performed similar convergence tests and found in agreement with this previous study that long simulations were necessary to ensure better convergence of the calculated free energy components and that averages obtained over short time windows are unreliable (data not shown). We also found that long simulations were necessary to obtain sufficient convergence of the entropic part of the binding free energies.

For each system, in order to improve the precision of the computed contributions to the binding free energy, all the snapshot structures for the protein, the ligand, and the protein/ligand complex were extracted from a single MD trajectory of the corresponding protein/ligand complex. This procedure provides only an approximation to the binding free energy because it does not take into account the possible conformational changes of the PDZ domain or the peptide in the process of binding. But we found that when snapshot structures were taken from separate trajectories, the standard deviation and the average values for the various free energy contributions had the same order of magnitude, thereby hampering the interpretation of the results. This is consistent with a previous study[58] which showed that taking separate trajectories of the components to evaluate binding free energies led to very large standard deviations on the averages.

Binding free energies of the twelve complexes calculated with MM/PBSA are presented in Table 2. While it is well-known that the binding free energies calculated with MM/PBSA do not reproduce experimental values accurately, MM/PBSA binding free energies have been shown to correlate with experiments well.[43,49] Although our purpose here is to use MM/PBSA calculations to reveal the forces governing binding through a decomposition of binding free energies, rather than to rank the different complexes, the ranking of binding free energies was tested for the Erbin and Syntenin systems for which the affinities were known experimentally.

In agreement with the experimental finding that the Erbin PDZ domain binds class I peptides with higher affinity than class II ones,[20] our MM/PBSA calculations showed that the binding free energy of the Erbin PDZ domain with the class I peptide was 23 kcal/mol more favorable than that with the class II peptide. The exaggerated magnitude of this difference is obviously due to the approximations in the MM/PBSA method, but the trend is reassuringly correct. We also found that the Syntenin PDZ domain had a binding free energy that was more favorable by 5.5 kcal/mol for the class II peptide than for the class I peptide. This is also consistent with experiments[59] that show that the PDZ2 domain of Syntenin binds slightly better to the class I than to the class II peptide. Together, these results show that the MM/PBSA method is capable of reproducing the important preference trends in the comparison of class I/II binding.

Direct comparison of the other results with experiment is not straightforward, because, as mentioned in the introduction, few experimental binding affinities are currently available for PDZ domain complexes. Although some experimental binding affinities are in the literature, it is often not possible to compare them among themselves, because the assays are done under very different conditions. Moreover, comparison with our calculations would require the peptide sequences used in the binding assays to match those in our simulations which were imposed by the available crystal structures. Nevertheless, experimental data are available for some cognate systems. For example, for NHERF there are three different binding experiments showing that the NHERF PDZ domain binds strongly to the three peptides (a, b, and c) but slightly better to the b-peptide ($K_d$ = 18 nM, ref 60) than to the a-peptide ($K_d$ = 26 nM, ref 61), and that both a- and b-peptides bind better than the c-peptide ($K_d$ = 48 nM, ref 62). In slight disagreement with the experiments, we found that the NHERF PDZ domain binds equally well to the a-peptide and to the b-peptide, but we agree with the experiments in finding the c-peptide to have the lowest affinity (6−7 kcal/mol less than the a- and b-peptides) for the NHERF PDZ domain. For PSD-

(58) Swanson, J. M.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.
(59) Kang, B. S.; Cooper, D. R.; Jelen, F.; Devedjiev, Y.; Derewenda, U.; Dauter, Z.; Otlewski, J.; Derewenda, Z. S. *Structure* **2003**, *11*, 459−68.
(60) Hall, R. A.; Ostedgaard, L. S.; Premont, R. T.; Blitzer, J. T.; Rahman, N.; Welsh, M. J.; Lefkowitz, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8496−501.
(61) Maudsley, S.; Zamah, A. M.; Rahman, N.; Blitzer, J. T.; Luttrell, L. M.; Lefkowitz, R. J.; Hall, R. A. *Mol. Cell. Biol.* **2000**, *20*, 8352−63.
(62) Wang, S.; Raab, R. W.; Schatz, P. J.; Guggino, W. B.; Li, M. *FEBS Lett.* **1998**, *427*, 103−8.

95, binding assays were conducted[63] with the CRIPT peptide, but with a shorter or a longer sequence than the one we simulated ($K_d = 2\ \mu$M for -YKQTSV and $K_d = 300\ \mu$M for -QTSV). Experiments on the Dishevelled PDZ domain showed that its complex with the Dapper peptide, with a longer sequence (-SGSLKLMTTV) than the one in our simulations, has a dissociation constant $K_d = 16\ \mu$M.[64] Unfortunately, to our knowledge, no binding data are available for the Shank PDZ domain in complex with the GKAP peptide, for the GRIP PDZ domain with liprin alpha, or for the Inad PDZ in complex with NorpA. Our predictions of relative PDZ domain binding affinities from the computational study with MM/PBSA are currently the only quantitative information available for these systems.

The case of the Shank PDZ domain is noteworthy, because the total calculated binding free energy of the complex is slightly positive (+4.1 kcal/mol), but the complex exists in the crystal structure. It becomes clear from the energy decomposition (cf. Table 2) that the calculated unfavorable binding free energy is entirely due to a very high entropic cost (+45 kcal/mol), which is by far the largest entropic cost of all the studied systems. One possible explanation is that this effect is due to the modeling procedure required to prepare the starting structure for this particular system: in the crystal structure of the Shank complex residues are missing in the $\beta$B−$\beta$C loop, and these residues were modeled based on the $\beta$B−$\beta$C loop of the apo form (see Methods). This procedure may have been responsible for an uncommon dynamic behavior of the complex, resulting in an exceptionally high entropic cost. However, another more intriguing explanation could be that this unfavorable or very small computed binding affinity reflects a *true* lack of affinity of the isolated/monomeric Shank PDZ domain for its cognate peptide−ligand. This would be consistent with the recent suggestion[34] that the functional unit of the Shank PDZ domain is a dimer rather than a monomer (the crystallographic unit does contain a dimer of Shank PDZ domains, and we only simulated an isolated monomer).

**PDZ Domain Binding Interactions.** One of the advantages of the MM/PBSA approach is that it enables a decomposition of the free energy into identifiable contributions. Thus, electrostatic ($\Delta G_b^{ele} = \Delta E_{coul} + \Delta\Delta G_{sol}^{ele}$), nonpolar or hydrophobic ($\Delta G_b^{np} = \Delta E_{vdw} + \Delta\Delta G_{sol}^{np}$), and entropic ($-T\Delta S_b$) contributions were analyzed separately (Table 2).

**A. Nonpolar** Contributions. The decomposition (Table 2) of the binding free energies into electrostatic, nonpolar, and entropic components showed that nonpolar contributions, and more specifically intermolecular van der Waals forces, dominate peptide/PDZ domain interactions. The absolute value of the nonpolar component represented on average 77% of the sum of the absolute values of the three components (electrostatic, nonpolar, and entropy). The average value of $\Delta G_b^{np}$ was −50.7 kcal/mol with a standard deviation of 10.4 kcal/mol. The highly favorable nonpolar binding free energy probably reflects the great number of favorable interactions from the various hydrophobic pockets that cover the surface of the binding site in the PDZ domains (cf. Figure 1a).

**B. Electrostatic Contributions**. In contrast to the nonpolar components, the electrostatic interactions were found to make a very small contribution to the binding free energy. Values of $\Delta G_b^{ele}$ were on average +1.9 kcal/mol with a standard error of 1.3 kcal/mol. They ranged from 0 kcal/mol for NHERF to about +4 kcal/mol for class I and II Erbin. The weakness of electrostatic interactions was surprising because the interaction between the conserved carboxylate binding loop and the C-terminal carboxyl group in the peptide ligand is one of the hallmarks of peptide/PDZ domain recognition, and also because the peptide ligands also comprise charged and polar residues that in some cases make distinct hydrogen-bonding interactions with residues in the PDZ domain receptor. However, decomposition of the free electrostatic contribution into its Coulombic $\Delta E_{coul}$ and solvation $\Delta G_{sol}^{ele}$ components showed that indeed the direct intermolecular electrostatic interactions were always favorable to the binding but their contributions could not compensate the large desolvation penalties associated with the binding event, thereby always leading to an unfavorable $\Delta G_b^{ele}$. This compensation phenomenon has been previously observed in several studies of protein/ligand interactions in solution.[45,65−67]

It should be noted that the electrostatic contributions were computed assuming a value of $\epsilon_i = 4$ for the dielectric constant of the protein interior. This value is commonly used for proteins in this kind of study.[45,49,68,69] However, there is no consensus on what the value of the dielectric constant for proteins should be: $\epsilon_i$ values greater than 1 are typically used to include polarization effects ($\epsilon_i = 2$) and/or structural reorganization of the protein ($\epsilon_i = 4$) implicitly.[70] To establish the effect of the value for the interior dielectric constant on the relative contribution of the electrostatic component to the total binding free energy, different ($\epsilon_i = 1$, 2, 4, 10, and 20) interior dielectric constant values were tested. These tests showed that the absolute value of the electrostatic binding free energy is roughly inversely proportional to the value of the dielectric constant (data not shown) in agreement with a previous report.[71] However, we found that using a different dielectric constant did not affect significantly our result of a relative small contribution of electrostatic interactions to binding free energies.

**C. Entropic Contributions**. There is no explicit method to calculate the entropy contribution to the binding free energy in the MM/PBSA approach. This contribution is sometimes neglected,[72] calculated with a normal-mode analysis on a few minimized structures extracted from the simulations,[43,48−50] or computed using a quasi-harmonic approximation.[50,73] Because normal-mode analysis is computationally very costly when used on a large number of configurations, we chose to use a quasi-harmonic analysis[55,56] to calculate the configurational part of the entropy contribution to the free energy of binding, as described in Materials and Methods. It should be mentioned

(63) Niethammer, M.; Valtschanoff, J. G.; Kapoor, T. M.; Allison, D. W.; Weinberg, T. M.; Craig, A. M.; Sheng, M. *Neuron* **1998**, *20*, 693−707.
(64) Wong, H. C.; Bourdelas, A.; Krauss, A.; Lee, H. J.; Shao, Y.; Wu, D.; Mlodzik, M.; Shi, D. L.; Zheng, J. *Mol. Cell.* **2003**, *12*, 1251−60.
(65) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−6.
(66) Smith, B. J.; Colman, P. M.; Von Itzstein, M.; Danylec, B.; Varghese, J. N. *Protein Sci.* **2001**, *10*, 689−96.
(67) Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567−82.
(68) Lee, M. R.; Tsai, J.; Baker, D.; Kollman, P. A. *J. Mol. Biol.* **2001**, *313*, 417−30.
(69) Suenaga, A.; Hatakeyama, M.; Ichikawa, M.; Yu, X.; Futatsugi, N.; Narumi, T.; Fukui, K.; Terada, T.; Taiji, M.; Shirouzu, M.; Yokoyama, S.; Konagaya, A. *Biochemistry* **2003**, *42*, 5195−200.
(70) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−32.
(71) Archontis, G.; Simonson, T. *J. Am. Chem. Soc.* **2001**, *123*, 11047−56.
(72) Laitinen, T.; Kankare, J. A.; Perakyla, M. *Proteins* **2004**, *55*, 34−43.
(73) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658−63.

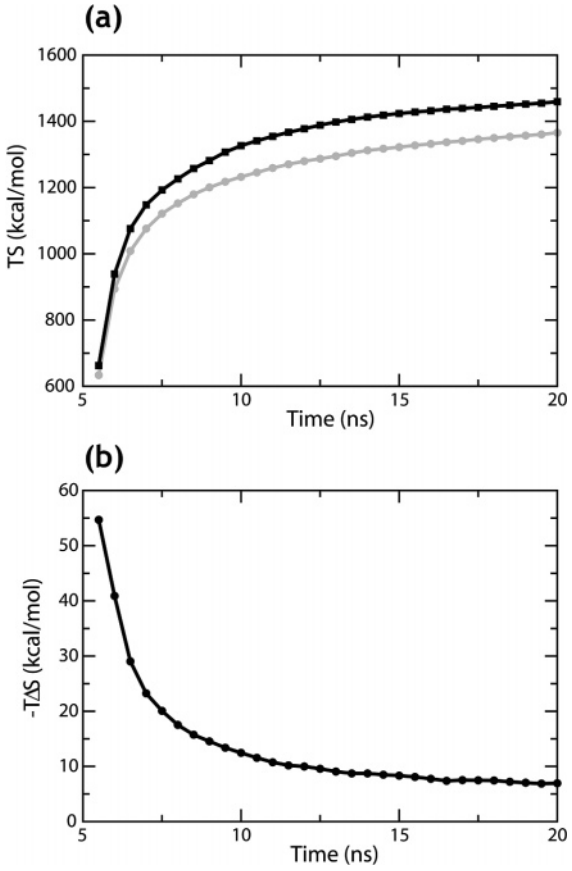


***Figure 2.*** Representation of the convergence of entropy calculations using the quasi-harmonic approximation. (a) Entropy (*TS*) of the NHERF PDZ domain apo (black line) and holo in complex with the "b" peptide (gray line) calculated on cumulative blocks, starting at 5 ns of MD simulations in intervals increasing by 0.5 ns. (b) Entropy part of the binding free energy (see Methods) calculated on cumulative blocks for the NHERF PDZ domain in complex with the "b" peptide.

that another entropy approximation, the Schlitter entropy,[74] which is also based on the assumption of quasi-harmonic behavior, has been extensively tested and used to evaluate the configurational entropy of peptides and proteins.[75−77] When the quasi-harmonic approximation holds, the Schlitter entropy is an upper bound approximation to $S_{ho}$ and is considered a less accurate approximation to the real configurational entropy than the quasi-harmonic entropy.[55]

The principal difficulty when evaluating the configurational entropy based on the quasi-harmonic approximation is that the covariance matrix of fluctuations needed for computing the entropy converges very slowly.[78] A previous MM/GBSA study reported that convergence was not achieved even when using 10-ns-long simulations.[50] In testing the approach for convergence by computing the entropy on cumulative blocks starting at 5 ns of MD, we found that the entropy of each molecule was not yet converged even after 20 ns of simulation. This is illustrated in Figure 2a for the apo and holo PDZ domain simulations of NHERF. However, we found that, despite the lack of convergence of the configurational entropy calculated for the individual molecules, differences taken between individual configurational entropies were sufficiently converged. The convergence of the entropic contribution to the binding free energy, $-T\Delta S_b = -T(S^{P::L} - S^P - S^L)$, is illustrated in Figure 2b, which represents a plot of $-T\Delta S_b$ on cumulative blocks as a function of the length of the time interval. We chose, therefore, to use the value of such entropy differences calculated from the last 15 ns of the trajectories (the last point on the graphs) as the *configurational entropy difference* and evaluated the error on this value as the difference between the value calculated from the last 15 ns (the time window from 5 to 20 ns of the trajectory or 10−25 ns for the longer simulations) and the value calculated in a time window of 10 ns (from 5 to 15 ns of the trajectory). This error is shown in parentheses in the entropy column of Tables 2−4.

Entropy convergence was tested further by performing multiple simulations on two systems: the PSD-95 complex and the Syntenin class I complex. These systems are the largest (PSD-95) and the smallest (Syntenin class I) complexes in total number of atoms. Four additional 20 ns MD simulations (one for PSD-95 and three for Syntenin) were performed with different seeds. For the PSD-95 complex the configurational binding entropy $-T\Delta S_b$ was $11 \pm 3$ kcal/mol for the additional simulation and $9 \pm 3$ kcal/mol for the simulation presented in Table 2. For the Syntenin class I complex the configurational binding entropies were $8 \pm 1$, $10 \pm 2$, and $9 \pm 1$ kcal/mol for the additional simulations and $8 \pm 1$ kcal/mol for the simulation presented in Table 2. These values are comparable within error and thus indicate that the entropy calculation is not simulation dependent. We conclude that the calculated differences in the values of configurational binding entropies for the various PDZ systems we study are not sampling artifacts.

It is important to note however that the quasi-harmonic approximation holds only in the case of a multivariate Gaussian probability distribution. Thus, when a system transits from a stable conformation (free energy well) to another stable conformation (free energy well), the quasi-harmonic approximation will overestimate the configurational entropy. These transitions can also lead to an overestimation of the configurational binding entropy. For example, we observed in the case of GRIP that a sudden change of conformation of the peptide interfered with the convergence of the configurational entropy difference. But, if the transition phases are removed from the trajectory (by cutting the trajectory into intervals during which the molecules are globally stable), the quasi-harmonic approximation holds and the convergence of the configurational entropy difference is restored.

The results in Table 2 show that the configurational entropy part is always unfavorable to binding. This is not unexpected because the entropies of the PDZ domain and the ligand, as well as the entropy of the complex, were calculated from a single simulation of the complex, and thus the entropy of the protein/ligand complex is by definition smaller than the sum of the entropies of its components, leading to an unfavorable entropy contribution to binding. In fact, as stated in the Methods section, the configurational entropy contribution to binding calculated with this protocol represents only the magnitude of correlated motions between the PDZ domain and the peptide ligand. And

(74) , Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617−621.
(75) Schafer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **2001**, *43*, 45−56.
(76) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809−7817.
(77) Schafer, H.; Smith, L. J.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **2002**, *46*, 215−24.
(78) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.

***Table 3.*** Adaptation Free Energy of the Peptides Upon Binding: $\Delta G^{L}_{adapt}$ in kcal/mol[a]

| protein | $\Delta E_{int}$ [b] | $\Delta G^{ele}$ [c] | $\Delta G^{np}$ [d] | $-T\Delta S_{adapt}$ [e] | $\Delta G^{*}_{adapt}$ [f] | $\Delta G_{adapt}$ [g] |
|---|---|---|---|---|---|---|
| NHERF[a] | 0.1 (*11.6*) | 0.5 (*3.0*) | 1.0 (*7.3*) | 57 (*3*) | 1.5 (*11.7*) | 58.5 (*14.7*) |
| NHERF[b] | −3.2 (*11.1*) | 1.2 (*1.9*) | −0.6 (*5.8*) | 59 (*2*) | −2.6 (*10.8*) | 56.4 (*12.8*) |
| NHERF[c] | −5.8 (*12.5*) | 4.2 (*2.1*) | −2.3 (*7.0*) | 32 (*4*) | −3.9 (*12.4*) | 28.1 (*16.4*) |
| PSD-95 | −3.3 (*12.1*) | 1.3 (*2.3*) | 1.8 (*6.0*) | 57 (*2*) | −0.2 (*11.7*) | 56.8 (*13.7*) |
| Dishevelled | 0.3 (*14.7*) | 0.8 (*2.9*) | 0.0 (*8.0*) | 70 (*18*) | 1.2 (*14.6*) | 71.2 (*32.6*) |
| Shank | −2.2 (*12.8*) | 1.8 (*2.7*) | −0.4 (*7.7*) | 25 (*10*) | −0.8 (*12.9*) | 24.2 (*22.9*) |
| Syntenin I | 3.0 (*10.6*) | −0.3 (*2.0*) | 1.8 (*5.5*) | 30 (*3*) | 4.5 (*10.1*) | 34.5 (*13.1*) |
| Syntenin II | 3.7 (*12.6*) | −2.6 (*2.5*) | 4.4 (*6.9*) | 63 (*3*) | 5.5 (*12.5*) | 68.5 (*15.5*) |
| Erbin I | 9.5 (*12.4*) | 0.5 (*2.3*) | 1.9 (*7.4*) | 73 (*5*) | 11.8 (*12.7*) | 84.8 (*17.7*) |
| Erbin II | 0.0 (*13.8*) | 0.9 (*2.1*) | 1.4 (*9.2*) | 55 (*15*) | 2.3 (*14.4*) | 57.3 (*29.4*) |
| GRIP | −3.9 (*13.8*) | −0.1 (*2.3*) | 1.6 (*7.2*) | 72 (*1*) | −2.4 (*13.5*) | 69.6 (*14.5*) |
| Inad | −1.1 (*10.5*) | 0.2 (*1.5*) | −0.2 (*5.6*) | 22 (*2*) | −1.1 (*10.3*) | 20.9 (*12.3*) |
| Average | **−0.2** (*4.1*) | **0.7** (*1.6*) | **0.9** (*1.7*) | **51.3** (*18.8*) | **1.3** (*4.3*) | **52.6** (*20.8*) |

[a] Standard deviations of averages are shown in parentheses. The last line is the average of each column and its corresponding standard deviation. [b] Difference in internal energy (sum of the bond, angle, and dihedral energies). [c] Difference in electrostatic free energy (sum of the intramolecular electrostatic energy and the electrostatic solvation free energy). [d] Difference in nonpolar free energy (sum of the van der Waals energy and the nonpolar part of the solvation free energy). [e] $-T\Delta S^{L}_{adapt} \approx -T(S^{L\_bound}_{conf} - S^{L\_free}_{conf})$ where $S^{L\_bound}_{conf}$ and $S^{L\_free}_{conf}$ are calculated with the quasi-harmonic approximation, respectively, from the complex and the free simulations. The error is calculated as $|-T\Delta S[5-20\ \text{ns}] + T\Delta S[5-15\ \text{ns}]|$. [f] $\Delta G^{*}_{adapt} = \Delta E_{int} + \Delta G^{ele} + \Delta G^{np}$. [g] $\Delta G_{adapt} = \Delta G^{*} - T\Delta S_{adapt}$.

thus, our results indicate that, in the bound state, the dynamics of the peptide and PDZ domain are generally tightly coupled.

However the magnitude of this dynamical coupling is very variable, as indicated by the variation in the configurational entropy contributions, $-T\Delta S_b$, to the binding free energies in Table 2. Configurational entropy contributions ranged from +5 or +10 kcal/mol for Inad, class I and II Syntenin, PSD-95 and NHERF (with a- and b-peptides), to +45 kcal/mol for Shank. Overall, the standard deviation for $-T\Delta S_b$ was of 12 kcal/mol. This standard deviation is as large as that of the nonpolar contribution (10.4 kcal/mol) and could suggest that the configurational entropy contributions could counterbalance nonpolar contributions in determining the specificity peptide/PDZ domain interactions. But this counterbalancing effect was only observed when analyzing the energetics of NHERF bound to peptide "c" vs peptide "b". In these complexes, the entropic contribution $-T\Delta S_b$ was twice as large (and unfavorable) for the "c" peptide (+14 kcal/mol) vs the "b" peptide (+7 kcal/mol). This lead to a weaker affinity between NHERF and peptide "c" (−33.6 kcal/mol) than that between NHERF and peptide "b" (−39.1 kcal/mol) although the nonentropic contribution ($\Delta G^{ele}_b + \Delta G^{np}_b$) to the binding free energy was slightly more favorable in the complex of NHERF with peptide "c" (−47.5 kcal/mol) than with peptide "b" (−46.2 kcal/mol). Nevertheless, the entropic contributions were remarkably variable. Notably, it is noteworthy that the entropic contributions, $-T\Delta S_b$, for peptides binding to the same PDZ domain were twice or almost twice as large in several complexes, including NHERF/peptide "c" vs NHERF/peptide "a" or "b", Syntenin/class I vs Syntenin/class II, and Erbin/class II vs Erbin/class I. This variability in the degree of coupling between the dynamics of the peptide and PDZ domain, while not influencing the affinity of the interaction (see above), is a property that could be exploited in interactions with downstream effectors.

In summary, the decomposition of binding free energies for the 12 different PDZ domains in complex with their target peptides indicated that nonpolar interactions provided the largest unopposed contribution to the binding affinity. In contrast, electrostatic interactions, always unfavorable, did not contribute significantly to the binding affinity, and entropic effects, always unfavorable to binding as well, did not counterbalance (with one exception) the magnitude of the affinity set by the nonpolar interactions. However, the entropic effects were remarkable in their variability, notably when comparing distinct peptides bound to the same PDZ domain.

**Adaptation Free Energies.** In order to compare the free energy of the bound and unbound/free states of each molecule (peptide or PDZ domain), the free energies of the "free" state of the PDZ domain and of the "free" peptide were computed from MD trajectories obtained for each molecule individually. The free energies of the free states thus obtained were then subtracted from the already computed free energies of the bound state of each molecule (based on the MD trajectories of the complexes), and these differences were considered to represent "adaptation free energies" as presented previously elsewhere.[27,28] This sort of adaptation free energy incorporates some of the effects that are missing from the previous binding free energy calculations, which used structures extracted only from the protein/ligand complex trajectories.

The adaptation free energies for the PDZ domains, $\Delta G^{P}_{adapt} = \langle G^{P}_{holo} \rangle - \langle G^{P}_{apo} \rangle$, and for the peptide ligands, $\Delta G^{L}_{adapt} = \langle G^{L}_{bound} \rangle - \langle G^{L}_{free} \rangle$, are reported in Tables 3 and 4, respectively. It should be noted that since the standard deviations associated with the adaptation free energies are very large, about 40−50 kcal/mol (the standard deviation of the free energy average for each simulation is about 20−25 kcal/mol), it was not possible to interpret these results precisely, but we could still identify some general features and compare the different PDZ domains. Notably, in order to distinguish between structural and dynamical features linked to the variations in adaptation free energy, the nonentropic $\Delta G^{*}_{adapt}$ and entropic $-T\Delta S_{adapt}$ contributions to the free energy of adaptation were analyzed separately.

For the peptide ligands, the adaptation free energy was very unfavorable in all cases, ranging from approximately +21 kcal/mol in Inad to +85 kcal/mol in Erbin/class I (Table 3). Interestingly, the nonentropic part of the adaptation free energy was negligible (+1.3 kcal/mol on average) in comparison to the

***Table 4.*** Adaptation Free Energy of the PDZ Domains Upon Binding: $\Delta G^{P}_{adapt}$ in kcal/mol[a]

| protein | $\Delta E_{int}$ [b] | $\Delta G^{ele}$ [c] | $\Delta G^{np}$ [d] | $-T\Delta S_{adapt}$ [e] | $\Delta G^{*}_{adapt}$ [f] | $\Delta G_{adapt}$ [g] |
|---|---|---|---|---|---|---|
| NHERF[a] | −6.0 (*41.5*) | 4.0 (*9.1*) | 1.4 (*32.2*) | 95 (*8*) | −0.6 (*42.3*) | 94.4 (*50.3*) |
| NHERF[b] | −4.2 (*40.5*) | 3.3 (*9.4*) | −2.1 (*34.0*) | 85(*3*) | −3.0 (*41.9*) | 82.0 (*44.9*) |
| NHERF[c] | −14.8 (*41.3*) | 3.1 (*10.2*) | 10.2 (*31.8*) | 95 (*3*) | −1.5 (*42.1*) | 93.5 (*45.1*) |
| PSD-95 | 30.0 (*47.1*) | −1.2 (*10.4*) | 7.5 (*35.7*) | 20 (*10*) | 36.3 (*47.2*) | 56.3 (*57.2*) |
| Dishevelled | −15.8 (*41.2*) | −1.2 (*8.2*) | 44.3 (*30.7*) | −17 (*5*) | 27.3 (*41.1*) | 10.3 (*46.1*) |
| Shank | 8.2 (*44.6*) | 0.1 (*8.9*) | 5.8 (*32.3*) | −1 (*3*) | 14.1 (*44.7*) | 13.1 (*47.7*) |
| Syntenin I | 7.9 (*38.0*) | −0.6 (*7.2*) | 11.1 (*26.7*) | −15 (*7*) | 18.4 (*36.8*) | 3.4 (*43.8*) |
| Syntenin II | −2.4 (*37.2*) | 3.6 (*7.0*) | 4.2 (*26.8*) | 15 (*2*) | 5.3 (*36.8*) | 20.3 (*38.8*) |
| Erbin I | 17.2 (*42.0*) | 6.2 (*16.4*) | 4.7 (*31.3*) | 4 (*2*) | 28.0 (*45.2*) | 32.0 (*47.2*) |
| Erbin II | 2.0 (*41.4*) | 3.3 (*9.8*) | 6.7 (*32.2*) | 0 (*11*) | 12.0 (*41.9*) | 12.0 (*52.9*) |
| GRIP | 10.0 (*40.8*) | −3.1 (*9.2*) | 17.7 (*32.4*) | 10 (*10*) | 24.7 (*43.2*) | 34.7 (*53.2*) |
| Inad | −0.8 (*42.6*) | −0.2 (*10.7*) | 8.5 (*34.2*) | 15 (*8*) | 7.6 (*43.6*) | 22.6 (*51.6*) |
| Average | **2.6** (*13.1*) | **1.4** (*2.8*) | **10.0** (*11.9*) | **26** (*42*) | **14.1** (*13.0*) | **39.6** (*33.5*) |

[a] Standard deviations of averages are shown in parentheses. The last line is the average of each column and its corresponding standard deviation. [b] Difference in internal energy (sum of the bond, angle, and dihedral energies). [c] Difference in electrostatic free energy (sum of the intramolecular electrostatic energy and the electrostatic solvation free energy). [d] Difference in nonpolar free energy (sum of the van der Waals energy and the nonpolar part of the solvation free energy). [e] $-T\Delta S^{P}_{adapt} \approx -T(S^{P\_holo}_{conf} - S^{P\_apo}_{conf})$ where $S^{P\_holo}_{conf}$ and $S^{P\_apo}_{conf}$ are calculated with the quasi-harmonic approximation, respectively, from the holo and the apo simulations. The error is calculated as $|-T\Delta S(5-20\text{ ns}) + T\Delta S(5-15\text{ ns})|$. [f] $\Delta G^{*}_{adapt} = \Delta E_{int} + \Delta G^{ele} + \Delta G^{np}$. [g] $\Delta G_{adapt} = \Delta G^{*}_{adapt} - T\Delta S_{adapt}$.

entropic contribution (+51.3 kcal/mol on average). Thus, the cost of adaptation came almost exclusively from entropic contributions, indicating that ordering of the peptide within the binding site was a critical aspect of the binding event (from the peptide point of view).

For the PDZ domains, the total adaptation free energy was also always unfavorable (Table 4) and very variable among PDZ domains, with an average value of about +39.6 kcal/mol and a standard deviation of 33.5 kcal/mol over all systems. But the picture emerging from the decomposition of the adaptation free energy cost into entropic ($-T\Delta S_{adapt}$ in Table 4) and nonentropic ($\Delta G^{*}_{adapt}$ in Table 4) contributions was not as clear as that for the peptide ligands: no general trend concerning the relative contribution of these two quantities ($\Delta G^{*}_{adapt}$ vs $-T\Delta S_{adapt}$) was readily apparent.

The values of the nonentropic part of the adaptation free energy for the PDZ domains, ($\Delta G^{*}_{adapt}$ in Table 4) ranged from about 0 kcal/mol for the three NHERF PDZ domains, Inad and class II Syntenin, to +37 kcal/mol for the PDZ domain of PSD-95. These values indicate that some of the polar and nonpolar internal interactions that stabilize the PDZ domain in the free state are lost in the bound state, leading to an unfavorable nonentropic free energy of adaptation, but for several PDZ domain complexes (5 out 12) these differences are only minor. Note that the nonentropic adaptation free energy cost did not correlate with the rmsd between the average structures (for the backbone atoms only) from the apo and holo simulations, a quantity that should reflect the degree of conformational change upon binding (data not shown). This lack of correlation suggests that, when present, the cost associated with the nonentropic part of the free energy of adaptation is probably related to subtle side-chain rearrangements, with concomitant packing and hydrogen-bonding or salt-bridge disruptions.

The values for the entropic part of the adaptation free energy for the PDZ domains ranged from −17 kcal/mol for Dishevelled to ∼90 kcal/mol for NHERF complexes. The limits of this wide range of values indicated not only that the degree of PDZ domain ordering upon binding was very variable but also that unlike what was observed for the peptides, the PDZ domains were not necessarily more ordered when bound than when free. Systems such as the Erbin class I and class II and the Shank PDZ domains showed almost no entropic adaptation. But even more remarkable were systems such as the Syntenin class I and Dishevelled PDZ domains for which the binding event was associated with an entropic gain ($-T\Delta S_{adapt}$ values of −15 and −17 kcal/mol, respectively) rather than an entropic cost. Overall, in 5 out of the 12 systems the holo state of the PDZ domain became less or equally ordered than the corresponding free/apo state. It is worth mentioning that this form of vibrational entropic gain represents a way to recover partially from the loss of rotational and translational degrees of freedom associated with a binding event. Steinberg and Scheraga[7979] first proposed to evaluate the magnitude of this entropic gain that arises from the creation of new internal/vibrational degrees of freedom in the complex. More recently, Tidor and Karplus[56] showed that these effects contribute −7.2 kcal/mol to the dimerization of insulin.

Finally, the contrasting behavior of the Syntenin PDZ domain when bound to class I and class II peptides is worth noting, since the computed values of $-T\Delta S_{adapt}$ indicate that upon binding the class II peptide the Syntenin PDZ domain becomes more ordered relative to its free state, but that upon binding the class I peptide the same Syntenin PDZ domain becomes less ordered. This behavior along with the remarkable variation in $-T\Delta S_{adapt}$ values is reminiscent of that observed in the previous section for the configurational entropy contribution and the class I/class II pair in the Erbin PDZ domain, and it is again tempting to speculate that these variations in the entropic response to binding could play a biological role in the differential recognition by downstream effectors.

## Discussion

The set of complexes of PDZ domains with their target peptides used in this study includes a variety of same-class and interclass pairs, so that sources of promiscuity or specificity in recognition could be expected to emerge from the comparisons

(79) Steinberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.* **1963**, *238*, 172−81.

of the various binding free energy contributions. We used a decomposition of the binding free energies and calculations of adaptation free energies from molecular dynamics (MD) simulations to evaluate the relative contribution of various free energy components, namely the electrostatic, nonpolar, and configurational entropy components, in peptide/PDZ domain recognition.

The free energies of all molecules were calculated using the MM/PBSA method, combined with a quasi-harmonic approximation for the configurational entropy part. Statistically reliable averages for free energies were shown to require long simulations. Binding free energies were evaluated from a single trajectory of the peptide/PDZ domain complex, and the resulting absolute values often overestimated experimental affinities. This was done although using separate trajectories (from separate simulations of the PDZ domain and the peptide ligand alone, and of the complex) may seem a more appropriate way to evaluate binding free energies, as they would take into account the structural changes involved in binding. However, such a protocol led to very large standard errors.

The results we obtained from the single trajectory calculations reproduced correctly the experimental ranking of binding free energies, including the preference of the Erbin PDZ domain for the class I peptide, the preference of the Syntenin PDZ domain for the class II peptide, and the preference of the NHERF PDZ domain for the a- and b-peptides over the c-peptide. The paucity of experimental data available on affinities of PDZ domain complexes (see Results), and the need for comparisons of results from assays done under the same conditions, emphasize the importance of our computational study on binding free energies and the results presented here.

**Nonpolar Interactions Dominate PDZ/Peptide Interactions.** The decomposition of binding free energies into a nonpolar part (sum of the van der Waals interactions and the nonpolar part of the solvation binding free energy), an electrostatic part (sum of direct Coulomb interactions and electrostatic solvation binding free energy), and a configurational entropy part showed that the nonpolar contribution was the largest one, representing, on average, 77% of the total binding free energy. The dominance of nonpolar interactions in PDZ/ligand binding is not surprising. PDZ domains interact mostly with hydrophobic ligands, and the typical PDZ domain binding pocket comprises many hydrophobic residues (Figure 1b). Nonpolar interactions have been shown to dominate the thermodynamics of protein−ligand recognition for many ligand classes including small ligands,[65,80] peptides interacting with immune response proteins such as MHC class I,[81] or other protein−protein interaction domains such as the SH3 domain,[49] and even highly charged ligands such as RNA[28] or DNA molecules.[82]

**Electrostatic Energies Do Not Contribute Significantly to PDZ/Peptide Interactions.** In contrast, the results indicated a very small and unfavorable contribution of the electrostatic interactions (0 to +3 kcal/mol) to the binding free energy. This was somewhat surprising because each PDZ domain complex studied here contains a bound peptide that carries a charged carboxylate ion, and this carboxylate ion is specifically recognized by the conserved GLGF sequence motif present on each PDZ domain. Nevertheless, a number of studies have shown that electrostatic contributions to the binding free energy are often small or unfavorable even when the ligands contain charged groups.[28,82,83] Generally, this relatively small or unfavorable contribution of electrostatic effects occurs because the direct intermolecular electrostatic interactions that are usually favorable cannot always compensate for the large desolvation penalties associated with ligand binding.[45,49,65,66] It should be noted, however, that in some cases electrostatic interactions drive complex formation.[84,85]

Our results are consistent with experimental observations about the role of electrostatic effects in PDZ domain interactions. Notably, a recent experimental study on PDZ domain ligand recognition that examined the salt dependence of terminal and internal ligand recognition in the mouse alpha1-syntrophin PDZ domain concluded that the charge on the C-terminus does not play a significant role in determining binding affinity.[86]

**Nonpolar and Electrostatic Contributions Enable Promiscuity in PDZ−Ligand Interactions.** The relatively small thermodynamic importance of the electrostatic contribution together with the dominance of nonpolar contributions indicates the thermodynamic basis for promiscuity in the PDZ domain interactions. This idea has been previously advanced by Harris and co-workers.[15,86] Together with their findings, the specifics of our proposition suggest that a given PDZ domain could bind with similar affinity a variety of peptides as long as these peptides can provide a certain threshold of nonpolar interactions, i.e., by projecting individual residues into specific patches on the surface of a PDZ domain binding pocket. This is not to say that any peptide could be recognized but that steric constraints,[15,86] rather than electrostatic ones, seem to define the domain-specific interactions[87] between the PDZ receptor and its putative cognate ligands. The suggestion that in contrast to their traditional role[88−91] electrostatic interactions do not seem to determine the specificity of peptide/PDZ interactions is probably a direct consequence of the recognition mechanism by $\beta$-sheet augmentation,[10,11] since in this form of recognition it is the backbone atoms of the ligand and not the side chains that form most of the interactions with the PDZ domain. A similar MM/PBSA study on an SH3 domain also found that electrostatic contributions to binding were small and unfavorable (+1.6 to +2 kcal/mol) and that nonpolar contributions played a primary role in determining affinity.[48] It seems important, therefore, to determine whether this unequal balance between nonpolar and electrostatic contributions may be a general thermodynamic

(80) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997−10002.
(81) Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293−301.
(82) Jayaram, B.; McConnell, K.; Dixit, S. B.; Das, A.; Beveridge, D. L. *J. Comput. Chem.* **2002**, *23*, 1−14.
(83) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14330−5.
(84) Lee, L. P.; Tidor, B. *Nat. Struct. Biol.* **2001**, *8*, 73−6.
(85) Sheinerman, F. B.; Honig, B. *J. Mol. Biol.* **2002**, *318*, 161−77.
(86) Harris, B. Z.; Lau, F. W.; Fujii, N.; Guy, R. K.; Lim, W. A. *Biochemistry* **2003**, *42*, 2797−805.
(87) Wiedemann, U.; Boisguerin, P.; Leben, R.; Leitner, D.; Krause, G.; Moelling, K.; Volkmer-Engert, R.; Oschkinat, H. *J. Mol. Biol.* **2004**, *343*, 703−18.
(88) Hendsch, Z. S.; Nohaile, M. J.; Sauer, R. T.; Tidor, B. *J. Am. Chem. Soc.* **2001**, *123*, 1264−5.
(89) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211−26.
(90) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1999**, *8*, 1381−92.
(91) Sindelar, C. V.; Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1998**, *7*, 1898−914.
(92) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, 14, 33−8, 27−8.
(93) Karthikeyan, S.; Leung, T.; Birrane, G.; Webster, G.; Ladias, J. A. *J. Mol. Biol.* **2001**, *308*, 963−73.
(94) Im, Y. J.; Park, S. H.; Rho, S. H.; Lee, J. H.; Kang, G. B.; Sheng, M.; Kim, E.; Eom, S. H. *J. Biol. Chem.* **2003**, *278*, 8501−7.

mechanism for sustaining promiscuity in molecular recognition by protein−protein interaction domains.

**Entropic Contributions Are Poised for Selectivity.** We have tried to capture several dynamic and entropic effects by computing both configurational entropy contributions (Table 2) and adaptation free energy changes (Tables 3 and 4) associated with the binding process. We found that the configurational entropy contribution to the binding free energy was sufficiently converged after 20 ns of MD simulation. This convergence was confirmed by doing multiple simulations on the PSD-95 complex and the Syntenin class I complex.

The changes in these quantities sketch a dynamic picture of PDZ domain binding interactions. For the peptide ligand, the binding event involves a large loss of vibrational entropy ($-T\Delta S_{adapt}$ in Table 3). This free energy loss relates to the significant loss of conformational freedom when the peptide is bound to the PDZ domain. This is a common observation in the binding of flexible ligands to protein receptors and suggests that preordering of the ligand, as in a constrained peptide or peptide mimetic, could significantly improve binding affinity.

In marked contrast, contributions from the PDZ domains do not share the uniformity seen in the behavior of the peptides. PDZ domains lost but also gained configurational entropy upon binding the ligands ($-T\Delta S_{adapt}$ in Table 4). This variety of responses was surprising because PDZ domains share a common fold and the complexes studied here all follow the canonical mode of PDZ domain/peptide recognition. And thus, it was also surprising to see that the degree to which the dynamics of the peptide ligands were coupled to those of the PDZ domains varied greatly ($-T\Delta S_b$ in Table 2). For comparison, in SH3 domains,[48] the entropic contribution was consistently evaluated at about 30 kcal/mol for all complexes. Moreover, the variability was observed not only when comparing complexes of distinct PDZ domains but also when contrasting distinct complexes of the same PDZ domain (i.e., Erbin class I vs Erbin class II in Table 2 and Syntenin class I vs class II in Table 4).

Together, our results show that both dynamic and entropic responses are complex-specific, in spite of the commonality of fold and mode of recognition in PDZ domains. Such complex-specific dynamical or entropic responses may form the basis for selective interactions either of PDZ domains with specific peptide ligands or of PDZ domain complexes with downstream effectors. In support of this hypothesis a recent combined experimental and computational study has shown that the recognition of a peptide loaded with MHC I molecule (Major Histocompatibility Complex class I) by the cognate T-cell receptor depends on the dynamics properties of the MHC I/peptide complex.[57] It is thus tempting to speculate that such a role for the dynamic properties of the PDZ domains in the recognition of their cognate ligands may indicate an involvement of the dynamics in ensuing mechanisms triggered by the complexation. Such inferences can be probed with simulation approaches such as those used in this study.

**Acknowledgment.** The work was supported in part by NIH grants from the National Institute on Drug Abuse (K05-DA00060, P01-DA12923, and P01-DA12408) and by the Institute for Computational Biomedicine (ICB) at Weill Medical College of Cornell University. The computations were carried out with the resources of the ICB, which are gratefully acknowledged.

JA060830Y