J|A|C|S
ARTICLES
Published on Web 09/21/2005

A Flexible Docking Procedure for the Exploration of Peptide Binding Selectivity to Known Structures and Homology Models of PDZ Domains

Masha Y. Niv[†] and Harel Weinstein*[,†,‡]

Contribution from the Department of Physiology and Biophysics and Institute for Computational Biomedicine, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021

Received June 24, 2005; E-mail: haw2002@med.cornell.edu

Abstract: PDZ domains are important scaffolding modules that typically bind to the C-termini of their interaction partners. Several structures of such complexes have been solved, revealing a conserved binding site in the PDZ domain and an extended conformation of the bound peptide. A compendium of information regarding PDZ complexes demonstrates that dissimilar C-terminal peptides bind to the same PDZ domain, and different PDZ domains can bind the same peptides. A detailed understanding of the PDZ−peptide recognition is needed to elucidate this complexity. To this end, we have designed a family of docking protocols for PDZ domains (termed PDZ-DocScheme) that is based on simulated annealing molecular dynamics and rotamer optimization, and is applicable to the docking of long peptides (20−40 rotatable bonds) to both known PDZ structures and to the more complicated problem of homology models of these domains. The resulting protocol reproduces the structures of PDZ complexes with peptides 4−8 amino acids long within 1−2 Å from the experimental structure when the docking is performed to the original structure. If the structure of the target PDZ domain is an apo structure or a homology model, the docking protocol yields structures within 3 Å in 9 out of 12 test cases. The automated docking procedure PDZ-DocScheme can serve in the generation of a structural context for validation of PDZ domain specificity from mutagenesis and ligand binding data.

Introduction

PDZ (PSD-95, Discs-large, ZO-1) domains are modules 70−90 amino acids long that often form a tandem of multiple copies or occur in combination with other protein-binding motifs.[1−5] These modular protein-interaction domains are involved in the assembly of supramolecular complexes and play an important role in cellular signaling. Protein targeting and recruitment is achieved by sequence-specific binding between a PDZ domain in one protein and a PDZ-binding motif in another protein. Specific internal β-finger folded peptides are recognized by PDZ in some cases,[6] but the vast majority of PDZ ligands are C-terminal peptides.[5] Several PDZ domain structures, with and without their peptidic ligands, were solved with NMR or X-ray crystallography. The results show the PDZ domain to consist of two α helices and six β strands, compactly arranged in a globular structure. The binding pocket is located between the second β strand (βB) and the second α helix (αB), and the peptides bind in an extended β-strand conformation. A conserved Gly-Leu-Gly-Phe (GLGF) sequence of the PDZ domain is found within the βA−βB connecting loop and is important for hydrogen bond coordination of the C-terminal carboxylate (COO^-) group.[7] Peptide positions are numbered from P0 in the C terminal, with P(−1) representing the next residue N-terminal to it and so on. The important role of positions P0 and P(−2) was highlighted from initial structure−function studies, and the PDZ domains were classified as binding either a S/T−X−Φ consensus motif (type I) or a Φ−X−Φ consensus motif (type II) (where Φ is any hydrophobic residue).[8] However, this classification has been challenged by the growing amount of data showing (1) that some PDZ domains appear to have dual specificity and (2) the importance of additional ligand positions that are being revealed as interaction determinants.[2,9] Complicating the understanding of the determinants for specificity are the observations that different PDZ domains may bind the same ligand: for example, both GRIP and PICK1 (25% identity) bind GluR2 peptide (IESVKI), while only GRIP binds the phosphorylated (IEpSVKI) or the phosphomimicking

[†] Department of Physiology and Biophysics.
[‡] Institute for Computational Biomedicine.

(1) Sheng, M.; Sala, C. *Annu. Rev. Neurosci.* **2001**, *24*, 1−29.
(2) Nourry, C.; Grant, S. G.; Borg, J. P. *Sci. STKE* **2003**, *179*, RE7.
(3) Hung, A. Y.; Sheng, M. *J. Biol. Chem.* **2002**, *277*, 5699−5702.
(4) van Ham, M.; Hendriks, W. *Mol. Biol. Rep.* **2003**, *30*, 69−82.
(5) Beuming, T.; Skrabanek, L.; Niv, M. Y.; Mukherjee, P.; Weinstein, H. *Bioinformatics* **2005**, *21*, 827−828.
(6) Harris, B. Z.; Hillier, B. J.; Lim, W. A. *Biochemistry* **2001**, *40*, 5921−5930.
(7) Doyle, D. A.; Lee, A.; Lewis, J.; Kim, E.; Sheng, M.; MacKinnon, R. *Cell* **1996**, *85*, 1067−1076.
(8) 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*, 73−77.
(9) Bezprozvanny, I.; Maximov, A. *FEBS Lett.* **2001**, *509*, 457−462.

10.1021/ja054195s CCC: $30.25

(IEEVKI) peptide.[10] On the other hand, the same PDZ domain can bind various, dissimilar peptides: for example, the P(−1) position of MUPP1 PDZ10 ligands can be occupied by Ser, Asp, Leu, or Tyr; the P(−2) position can be occupied by Ser, Thr, or Asp (but Ser-to-Asp mutation in serotonin 5-HT2C receptor disrupts binding[11]), and P(−3) can be occupied by Ile, Val, Thr, His, Ala, or Lys (found by querying PDZbase[5]).

The structures of several PDZ domain complexes with peptides have helped researchers understand the basis for the various binding preferences, but there is overwhelming evidence in the literature that the manner in which the specificity of binding is achieved may require a broad investigation: There are about 500 different proteins comprising one or more PDZ domains in the human genome (http://smart.embl-heidelberg.de/), each interacting with one or several peptides, but only ∼15 PDZ/ligand complex structures have been solved so far. Consequently, there is a need for a computational procedure to obtain reliable models of the complexes for the cases where the PDZ domain was solved alone, with a different peptide, or not at all. The procedure for obtaining a model of the complex involves docking a peptide to a PDZ structure (in the latter case, a homology model), accounting for the structural rearrangement of both backbone and side chains of the peptides.

The ability of docking methods to redock ligands into a known native structure was evaluated in several recent papers.[12−16] While docking of some ligands with up to 10 and sometimes up to 20 rotatable bonds is achievable with several docking algorithms, docking of highly flexible ligands remains nontrivial. In a comparative study performed by Bursulaya et al., ICM[17] docked 2 penicillopepsin ligands with 30 and 29 rotatable bonds within 2 Å; one of the ligands was also successfully docked by Autodock,[18] while the other had RMSD of 9.10 Å. DOCK,[19] FlexX,[20] and GOLD[21] misdocked these highly flexible ligands (with RMSD > 5.9 Å).[15] A new hierarchical method, Glide, was shown to perform very well for ligands with up to 20 rotatable bonds. For ligands with more than 20 rotatable bonds, 30 out of 55 ligands docked within 2 Å.[22] Kellenberger and co-workers' [16] test set included 5 ligands with more than 25 rotatable bonds: Surflex[23] docked 3 out of 5 with RMSD < 2 Å, FLexX[20] and FRED[24] docked 2 out of 5, GOLD[21] and GLIDE[22] docked 1 out of 5, and DOCK[19] docked none of these flexible ligands within 2 Å.[16] Liu and co-workers docked 13 out of 22 long peptides (4−16 amino acids) within dRMS of 3 Å, starting from the bound conformation of the peptide and using Monte Carlo annealing (Table 5 in ref 25) (dRMS is the root-mean-square deviation of the distances between the docked conformation and the native one). Desmet and co-workers have reproduced main-chain placement and many of the side-chain features for MHC I binding octapeptides.[26] The protocol used dead-end elimination (DEE) to find the optimal side-chain conformation for multiple main-chain orientations, similar to Leach[27] and Schaffer and Verkhivker.[28] Another work on MHC peptides employed the fact that N-terminal and C-terminal parts of the peptides dock to well-conserved sites. The edges were docked as rigid bodies using ICM, and the rest of the peptide was placed using a loop closure technique, achieving Cα RMSD < 1.5 Å for peptides 9−10 amino acids (aa) long (heavy atoms RMSD was not reported).[29] Taken together, these results show that docking of long peptides remains challenging and is achieved best when experimental information exists about binding anchors or about conformational properties. Such is the case for PDZ domains, where the position of P0 is conserved, and the peptides bind in an extended conformation. A more complex challenge is presented by docking to structures that are not known experimentally in their complexed forms (termed here "non-native structures"). The success of cross-docking is typically lower than docking back to the native structure (redocking), and the success is reduced with increasing number of rotatable bonds.[30] To the best of our knowledge, cross-docking of ligands with over 20 rotatable bonds has not been evaluated.

The problem is further complicated by the fact that for many real-life problems there are no experimentally determined structures available for the docking target, and one needs to resort to homology modeling. The use of homology models was successfully applied to pharmaceutically relevant questions, as reviewed recently by Hillisch et al.[31] Virtual screening against homology models was shown to enrich the hit factor compared to random screening (e.g., see refs 32−34). A direct comparison with a known complex structure was carried out by Schafferhans and Klebe.[35] Rigid inhibitors in their bound conformation were docked into models of thrombin generated from serine proteases with 28−40% identity, yielding binding modes with average RMS deviation of 1.4 Å. Three thrombin ligands with up to 5 rotatable bonds were rigidly docked to several homology models, starting from multiple conformations and producing first rank within 2.6 Å (and in some cases even below 1.5 Å).[35] Evers and co-workers used bound ligand models to introduce restraints

(10) Chung, H. J.; Xia, J.; Scannevin, R. H.; Zhang, X.; Huganir, R. L. *J. Neurosci.* **2000**, *20*, 7258−7267.
(11) Parker, L. L.; Backstrom, J. R.; Sanders-Bush, E.; Shieh, B. H. *J. Biol. Chem.* **2003**, *278*, 21576−21583.
(12) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. *J. Med. Chem.* **2004**, *47*, 558−565.
(13) Perola, E.; Walters, W. P.; Charifson, P. S. *Proteins* **2004**, *56*, 235−249.
(14) Schulz-Gasch, T.; Stahl, M. *J. Mol. Model. (Online)* **2003**, *9*, 47−57.
(15) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., III. *J. Comput. Aided Mol. Des.* **2003**, *17*, 755−763.
(16) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. *Proteins* **2004**, *57*, 225−242.
(17) Totrov, M.; Abagyan, R. *Proteins* **1997**; Suppl 1:215-220.
(18) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(19) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411−428.
(20) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470−489.
(21) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727−748.
(22) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(23) Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499−511.
(24) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. *Biopolymers* **2003**, *68*, 76−90.
(25) Liu, Z.; Dominy, B. N.; Shakhnovich, E. I. *J. Am. Chem. Soc.* **2004**, *126*, 8515−8528.
(26) Desmet, J.; Wilson, I. A.; Joniau, M.; De Maeyer, M.; Lasters, I. *Faseb J.* **1997**, *11*, 164−172.
(27) Leach, A. R. *J. Mol. Biol.* **1994**, *235*, 345−356.
(28) Schaffer, L.; Verkhivker, G. M. *Proteins* **1998**, *33*, 295−310.
(29) Tong, J. C.; Tan, T. W.; Ranganathan, S. *Protein Sci.* **2004**, *13*, 2523−2532.
(30) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. *J. Med. Chem.* **2004**, *47*, 45−55.
(31) Hillisch, A.; Pineda, L. F.; Hilgenfeld, R. *Drug Discov. Today* **2004**, *9*, 659−669.
(32) Oshiro, C.; Bradley, E. K.; Eksterowicz, J.; Evensen, E.; Lamb, M. L.; Lanctot, J. K.; Putta, S.; Stanton, R.; Grootenhuis, P. D. *J. Med. Chem.* **2004**, *47*, 764−767.
(33) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. *Proteins* **2003**, *50*, 5−25.
(34) McGovern, S. L.; Shoichet, B. K. *J. Med. Chem.* **2003**, *46*, 2895−2907.
(35) Schafferhans, A.; Klebe, G. *J. Mol. Biol.* **2001**, *307*, 407−427.

for further homology model refinements in the MOBILE application.[36]

PDZ domains share about 25% identity in sequence, but thus far the folding patterns observed experimentally were found to be very close. We demonstrate here successful docking to homology models of such PDZ domains. To the best of our knowledge no systematic docking studies were performed for PDZ domains until now. Peptides were modeled into X-ray structures of PDZ domains by building an extended peptide and minimizing it,[37] by homology with another complex,[38] and by molecular replacement followed by minimization and dynamics,[39] typically attempting to model four C-terminal residues of the peptide. While this is reasonable for most of the short peptides and some longer ones, the docking of long peptides calls for a more elaborate procedure. (For example, following the simpler procedures by building an extended conformation into the native PDZ domain from Dishevelled/Dapper complex and minimizing the complex yields an unsatisfactory RMSD of 6.8 Å for the 8-aa-long peptide.)

Motivated by the key role of PDZ domains in cell signaling, by the puzzling observations in attempts to classify PDZ−peptide recognition selectivity, and by the sheer abundance of PDZ−ligand complexes without known structure, we set to dock peptides to PDZ structures and homology models. To this end we developed CHARMM[40]-based simulated annealing protocols and tested them against several types of PDZ targets. Simulated annealing molecular dynamics (SA/MD) as an effective and efficient search strategy for flexible docking has been highlighted by Vieth and co-workers.[41−43] Our docking protocol PDZ-DocScheme is based on SA/MD with soft core potential[42] or flexible binding site side chains; this is followed by rotamer optimization using SCRWL,[44] and CHARMM minimization and scoring. Docking performance is tested against a set of native complexes, non-native crystal structures, and homology models.

Methods

The PDZ-DockScheme consists of four consecutive steps:

Step A: Generation of Initial Conformations. The optimal position of the backbone of the C-terminal residue (P0) is identified as follows: in redocking to a known PDZ structure from the complex, or to a PDZ structure from a complex with another peptide, the position of the P0 backbone is taken from the crystal structure of the same PDZ domain. In docking to apo-structures, the P0 backbone position is determined by superimposing the apo-structure on the known structure of a complex. In homology models, the P0 position is modeled on the same templates as the protein (only peptide-bound proteins are used as templates). The rest of the peptide is built in an extended conformation and hydrogen atoms are placed. The Cα atom of P0 is then tethered by a 10 kcal/mol· Å^2 harmonic force to its position. The complex is minimized for 800 steps using the adopted basis Newton−Raphson (ABNR) method. Next the system is heated during 10 ps using a 1-fs time step, a leapfrog Verlet integrator, and a distance-dependent dielectric of 2r. The system is propagated for 600 ps, with snapshots saved every picosecond. The procedure termed **A_rigid** is the above procedure with the protein kept frozen and the elevated temperature set to 700 K; **A_flex** is the above procedure with the peptide and the protein side chains within a 6 Å radius from the peptide allowed to move while the rest of the protein is frozen and the elevated temperature is 1000 K. The CHARMM trajectory is unpacked using SIMULAID software (http://fulcrum.physbio.mssm.edu/~mezei/simulaid/simulaid.html) to provide the initial geometries for step B.

Step B: Simulated Annealing. Two different approaches were employed, termed **B_soft** and **B_flex**.

For **B_soft**, the soft-core simulated annealing routine follows that of Wu and co-workers[42] and is summarized here briefly. During the docking process, nonbonded interactions are softened during three cooling cycles. [The first run has the softest potential (E_{max}(vDW) = 0.6 kcal/mol, E_{max}(att) = −0.4, E_{max}(rep) = 8.0) and starts from 700 K, next cycle has a harder potential (E_{max}(vDW) = 3.0, E_{max}(att) = −20.0, E_{max}(rep) = 40.0) and starts from 500 K, and the last cycle (E_{max}(vDW) = 30.0, E_{max}(att) = −200.0, E_{max}(rep) = 400) starts at 400 K and cools to 50 K; altogether the cooling is achieved in 35 ps ("att" stands for attractive electrostatic and "rep" stands for repulsive electrostatic).] Finally the system is minimized for 300 ABNR steps with all degrees of freedom flexible. The soft-core potential is given by $E_{ij}(r_{ij}) = E_{max} - ar_{ij}^b$ if $|E_{ij}^*| > 0.5E_{max}$, where E_{ij}^* is the regular potential and a and b are extracted from the equations that express equality of regular and soft potential and forces at the switching distance; a distance-dependent dielectric of 3r [30] is used. The procedure is performed for each snapshot obtained in step **A**.

In contrast to the soft-core potential in **B_soft**, cooling is performed in **B_flex** using the unmodified par22 potential[45] of CHARMM.[40] The peptide and the protein side chains within 6 Å from it are flexible, and the rest of the protein is frozen. Cooling is performed in 9 ps, so despite many more degrees of freedom being treated as flexible, **B_flex** is only 2.5 times slower than **B_soft**. (Typically **B_flex** takes 2.3 min per snapshot on an SGI ALTIX 1300-MHz processor. Use of a tabulated look-up potential would accelerate the calculations[42] but was not implemented in this work.) The system is minimized for 300 ABNR steps with all degrees of freedom flexible.

Step C: Side Chain Optimization. For each conformation obtained from step **B**, optimal side-chain rotamers are chosen using SCWRL3.0.[44] SCWRL uses a backbone-dependent rotamer library and an energy function based on log probabilities for these rotamers and a simple repulsive term. Dead-end elimination (DEE) of rotamers that are excluded from the global minimum energy configuration is employed: rotamer s_i of residue i, r_i, always has a lower interaction energy with all other side chains and the backbone, regardless of which rotamer is chosen for the other side chains. The SCRWL3.0 version employs results from graph theory to solve the combinatorial problem that remains after the DEE step. The resulting structure is fed back into CHARMM, the hydrogens are placed and minimized for 100 steps with the steepest descent method, and then all degrees of freedom in the complex are minimized for 1500 steps with ABNR. The minimization enables protein backbone adjustment to the docked peptide, as well as adjustment of the P0 position, since the tether is relieved at this stage.

Scoring. The poses are scored by the sum of interaction energies between the PDZ domains and the peptide, and the internal energy of the peptide. The performance of the algorithm is evaluated by the RMSD of all heavy atoms of the peptide after superposition of the PDZ backbone with the X-ray structure. The pose with the best score is considered as the result produced by the docking procedure.

(36) Evers, A.; Gohlke, H.; Klebe, G. *J. Mol. Biol.* **2003**, *334*, 327−345.
(37) Au, Y.; Atkinson, R. A.; Guerrini, R.; Kelly, G.; Joseph, C.; Martin, S. R.; Muskett, F. W.; Pallavicini, A.; Faulkner, G.; Pastore, A. *Structure (Cambridge)* **2004**, *12*, 611−622.
(38) Perez, J. L.; Khatri, L.; Chang, C.; Srivastava, S.; Osten, P.; Ziff, E. B. *J. Neurosci.* **2001**, *21*, 5417−5428.
(39) Zhou, H.; Xu, Y.; Yang, Y.; Huang, A.; Wu, J.; Shi, Y. *J. Biol. Chem.* **2005**, *280*, 13841−13847.
(40) Brooks, B. R.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(41) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L. *J. Comput. Chem.* **1998**, *19*, 1623−1631.
(42) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. *J. Comput. Chem.* **2003**, *24*, 1549−1562.
(43) Wu, G.; Vieth, M. *J. Med. Chem.* **2004**, *47*, 3142−3148.
(44) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. *Protein Sci.* **2003**, *12*, 2001−2014.
(45) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

Table 1. Summary of Docking Schemes

step	heating	cooling	side-chains optimization
PDZ-DocScheme 1	**A_rigid**	**B_soft**	not performed
PDZ-DocScheme 2	**A_rigid**	**B_soft**	performed
PDZ-DocScheme 3	**A_flex**	**B_flex**	not performed
PDZ-DocScheme 4	**A_flex**	**B_flex**	performed

[a]Details of heating and cooling steps are described in the Methods section.

The simulated annealing schemes we have tested are summarized in Table 1. All four variations of PDZ-DocScheme treat all degrees of freedom of the peptide as flexible and do not use information about bound conformation. The only input is the backbone position of the C-terminal residue (P0). This constraint is justified by the observation that the P0 position is highly similar in known PDZ complexes. The initial conditions are obtained from a high-temperature trajectory of a peptide tethered to the protein by its P0 Cα. Different degrees of flexibility of the PDZ domain are incorporated in these simulated annealing schemes: PDZ-DocScheme 1 is similar to the CDOCKER routine, where the protein flexibility is accounted for by softening of the potential.[42] However, since PDZ-DocScheme 1 is geared toward docking long peptides, we produce 300−600 conformations from a high-temperature trajectory of the peptide tethered to the protein, while CDOCKER uses 50 conformations produced with CORINA.[46] PDZ-DocScheme 2 adds optimization of protein side chains for multiple backbone conformations obtained in PDZ-DocScheme 1. PDZ-DocScheme 3 treats the protein binding site side chains' flexibly throughout heating and cooling (no soft-core potential is used), and PDZ-DocScheme 4 adds the rotamer optimization step using SCRWL on top of PDZ-DocScheme 3.

PDZ Targets. To enable the comparative studies, X-ray structures of PDZ domains that are noncovalently bound to C-terminal peptides of 5−8 amino acids were chosen if additional structures exist (to enable cross-docking) or can be obtained from reliable homology modeling. This test set includes: GRIP1 PDZ6 (1N7F), postsynaptic density protein PSD95 (1BE9), Dishevelled (1L60), Syntenin (1OBY), and Na^+/H^+ exchanger regulatory factor, NHERF (1GQ4). We have also used one NMR structure, 1N7T for Erbin.

Homology Modeling. Homology models were constructed with Modeller 6V2.0 (refs 47 and 48) using at least two templates of bound PDZ domains, with a minimal identity to the target of 25%. The model with the best objective score out of 20 was selected. Problematic regions (highlighted by Procheck[49]) and the βB/βC loop were submitted to the MODLOOP server,[50] and then the side chains were optimized with SCWRL3.0.[44] (The model for Syntenin, which does not have a βB/βC insertion, was built using multiple templates, without the optimization steps with MODLOOP and SCWRL.)

Results and Discussion

Redocking. Redocking (docking to the PDZ domain of the original complex) with PDZ-DocScheme 1 produced best-scoring poses with heavy atoms RMSD of the peptide below 2 Å, as summarized in Table 2. The only case where (despite good sampling as shown in the "best RMSD" column) satisfactory docking was not achieved is Syntenin (row 7 of Table 2). This was due to the insufficiently defined N-terminal residues of the peptide that interact only loosely with the protein, and disregarding the N-terminal residue gives high-quality docking results for the C-terminal 5-aa-long peptide (row 6 of Table 2). (Consequently, the same strategy was used in subsequent docking experiments with Syntenin structures; i.e., only the last 5 aa were used for scoring and RMSD calculation, although all 6 residues were present in the docking run.)

Table 2. Redocking of Peptides to Native Structures Using PDZ-DocScheme 1

structure	PDZ	peptide sequence	rotatable bonds	RMSD of best score	best RMSD
1N7F	GRIP PDZ6	ATVRTYSC	38	1.9	1.2
1BE9	PSD95	KQTSV	26	2.0	1.4
1L6O	Dishevelled	SLKLMTTV	44	1.7	1.8
1GQ4	NHERF	NDSLL[a]	20	1.4	1.4
1N7T	Erbin	WETWV[a]	23	1.7	1.7
1OBY	Syntenin	NEFYA[a]	21	1.7[b]	1.2
1OBY	Syntenin	TNEFYA	26	3.7	1.4

[a] Docking was performed for a longer sequence, but the RMSD and score are calculated for the last five residues only. [b] An additional pose with a nearly identical scoring energy had an RMSD of 6.3 Å. It was discarded after inspection showed an intrapeptide hydrogen bond between the asparagine (P(−5)) side chain and the phenylalanine (P(−2)) backbone.

In the NMR structures of Erbin, the Thr P(−6) and Gly P(−5) residues are very divergent, and therefore only the P(−4) through P0 part of the peptide (WETWV) is used for scoring and RMSD calculation. Analysis of preliminary results for Erbin showed that an arginine on the βB/βC loop presented a very high barrier to sampling. Once the temperature at step A was elevated to 1200 K, this barrier was overcome and the favorable results are shown in Table 2.

In the crystalline state of NHERF, the carboxy-terminal sequence of PDZ1 of NHERF occupies the peptide-binding pocket of a neighboring PDZ1 molecule related by two-fold crystallographic symmetry.[51] This was used by the authors to solve PDZ1 with cognate peptides: peptides of interest were concatenated to the C-terminal tail of the NHERF PDZ1 domain and the chimeric sequences were crystallized.[52] Since the N-capped NDSLL peptidic sequence corresponds poorly to the physicochemical properties of the chimera (where the negatively charged sequence DPE, not a positively charged H_3^+N, precedes the terminal NDSLL stretch), we used the DPENDSLL peptide in the docking experiments. RMSD and scoring were calculated for the actual NDSLL ligand only.

No information of the original geometry of the peptide was used except for the backbone position of P0 residue, as discussed in Methods. Table 2 shows that the conformational space was sampled very well with the soft representation of the rigid protein (lowest heavy atoms RMSD of the peptides are between 0.95 and 1.8 Å), and the simple scoring used (interaction + peptide energy in distance-dependent dielectric) identified poses with heavy atoms RMSD of the peptide lower than 2 Å. Figure 1 shows the scoring energy vs RMSD calculated for the heavy atoms of the peptide (after superposition of the PDZ Cα atoms) for all poses in test cases 1−6 detailed in Table 2. The results indicate satisfactory distribution of energies with RMSD, in that the best energy scoring poses have low RMSD and can be readily identified. This is in agreement with results reported by Vieth and co-workers[41] and further shows that soft-core molecular dynamics simulated annealing can be successfully

(46) Sadowski, J.; Gasteiger, J.; Klebe, G. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(47) Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. *Proteins* **1995**, *23*, 318−326.
(48) Sanchez, R.; Sali, A. *Methods Mol. Biol.* **2000**, *143*, 97−129.
(49) Laskowski, R. A., M. M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(50) Fiser, A.; Sali, A. *Bioinformatics* **2003**, *19*, 2500−2501.
(51) Karthikeyan, S.; Leung, T.; Birrane, G.; Webster, G.; Ladias, J. A. *J. Mol. Biol.* **2001**, *308*, 963−973.
(52) Karthikeyan, S.; Leung, T.; Ladias, J. A. *J. Biol. Chem.* **2002**, *277*, 18973−18978.

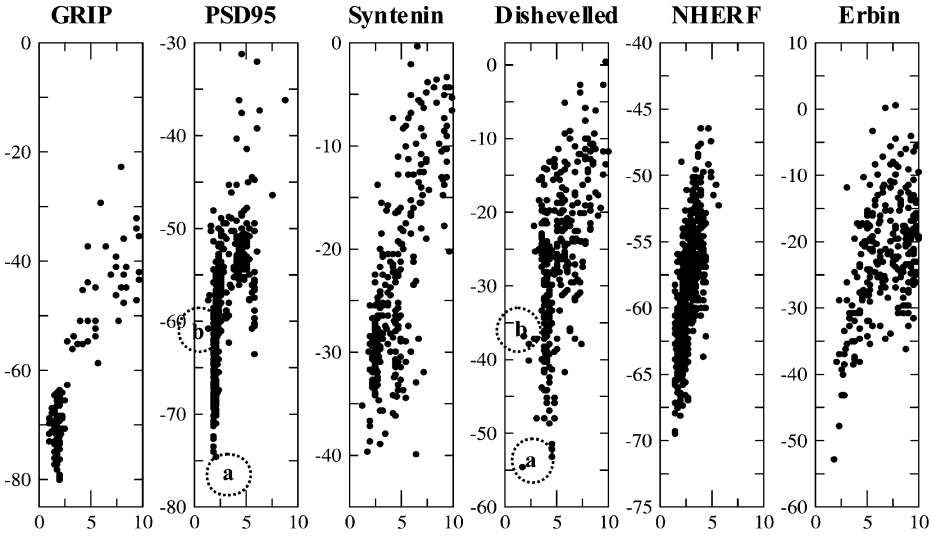

Figure 1. Redocking (docking to native structures). Scoring energy (sum of peptide−protein interaction and internal peptide energy calculated with CHARMM22, in kcal/mol) is plotted vs heavy atoms RMSD (in angstroms) between the redocked and the native peptides. For PSD95 and Dishevelled, poses with the best-scoring energy are circled and labeled "a", while poses with similar RMSD and higher energy are circled and labeled "b".

extended to 20−40 flexible degrees of freedom in the ligand when redocking to the native structure.

A general trend of increasing scoring energy with increasing RMSD is evident in Figure 1, though the RMSD-to-scoring-energy relation is not linear. For example, in the PSD95 case, pose "b" (RMSD of 1.6 Å from the X-ray structure, 2.2 Å from the best-scoring pose "a"), is 25 kcal/mol higher in scoring energy than pose "a". The difference in energy is due to the reorientation of a single side chain that causes a loss of most of the electrostatic interaction between the PDZ and Lys in position P(−4) of the peptide. In Dishevelled, pose "b" (RMSD 1.8 Å) scores 18 kcal/mol higher than the best-scoring pose "a", mainly due to smaller electrostatic contribution from the interaction of Lys P(−5) and Ser P(−7) with the PDZ domain, but the RMSD difference is not significant (1.7 Å from the X-ray structure and 2.4 Å from pose "a"). It is clear, therefore, that poses similar to the best-scoring one may have significantly higher scoring energy, mostly due to slight conformational changes in charged residues. Nevertheless, best-scoring poses in each case have RMSD values of less than 2 Å relative to the known structures, and this makes it possible to pick a correct pose within 2 Å RMSD solely on the basis of the scoring energy.

The best-scoring pose for the GRIP PDZ6 is superimposed on the X-ray complex in Figure 2, illustrating the excellent agreement achieved.

Cross-Docking. Once the ability to redock peptides to their original structures was established, we attempted to dock the same peptides to other existing experimental structures of their cognate PDZ domains (except for Dishevelled, for which an additional experimental structure is not available). The backbone RMSD values of the PDZ structures in this test set, from the PDZ structures in the native complexes, are below 1 Å for all cases but the Erbin (Table 3). Erbin has a long twisted βB/βC loop and other regions that vary between the NMR complex structure (1N7T) and the X-ray structure (1MFG). For the cross-docking and for docking to homology models, we consider RMSD > 3 Å as a misdocked pose following Kontoyianni and co-workers.[12] Even applying this looser criterion, PDZ-DocScheme 1 (that performed remarkably well in redocking experiment), fails in three cases of cross-docking (PSD95, GRIP, and Erbin; see Table 3). We therefore continued to develop other schemes to obtain a better agreement in cross-docking. Comparative results from these alternative protocols are summarized in Table 3.

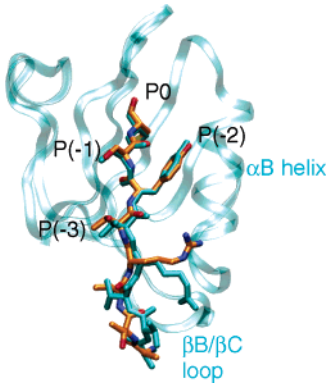

Figure 2. Grip/liprinα peptide complex (1n7f) and redocked liprinα peptide. GRIP PDZ domain is depicted in ribbon. The native peptide is shown in cyan, and the redocked peptide is colored by atom types (C, orange; N, blue; O, red).

Results for PDZ-DocScheme 2 in Table 3 show that inclusion of rotamer optimization step **C** improves the RMSD only marginally. However, rendering the binding site side chains flexible (PDZ-DocScheme 3) results in successfully docked poses for PSD95, GRIP, Syntenin (1R6J), and NHERF. Inclusion of SCRWL side-chain optimization (PDZ-DocScheme 4, shown in bold in Table 3) improves the results for GRIP, PSD95, Syntenin (1OBX), and Erbin. Overall, PDZ-DocScheme 4 is able to dock 5 out 7 peptides of 5−8-aa length to non-native PDZ structures with RMSD < 2.8 Å, and in all of the 7 cases rank 9 or below docks with RMSD < 2.4 Å. Plots of scoring energy vs RMSD obtained with PDZ-DocScheme 4 are shown in Figure 3.

Inspection of the best RMSD results for PDZ-DocScheme 4 (Table 3) reveals that conformations very close to the known structure are sampled in the GRIP, PSD95, and NHERF docking tests, but even the closest poses in the Syntenin and Erbin test cases are slightly above 2 Å. This may be due to the special properties of the 1OBY structure of Syntenin, which is a dimer

Table 3. Docking of Peptides to Non-native Crystal Structures Using Four PDZ-Doc Schemes

structure	PDZ	bb RMSD (Å)[a]	peptide sequence	PDZ-DocScheme 1	PDZ-DocScheme 2	PDZ-DocScheme 3	**PDZ-DocScheme 4**	best RMSD[b]
1N7E	GRIP PDZ6 (apo)	0.8	ATVRTYSC	3.2	3.1	2.5	**1.8**	1.5
1BFE	PSD95 (apo)	0.7	KQTSV	7.0	6.3	2.2	**1.5**	1.3
1R6J	Syntenin (apo)	0.7	NEFYA[c]	1.8	2.5	2.6	**2.8**	2.2
1OBX	Syntenin (cross)	0.7	NEFYA[c]	3.0	2.9	4.0	**2.4**	2.2
1GQ5	NHERF (cross)	0.7	NDSLL[c]	1.4	3.5	1.8	**2.0**	1.2
1I92	NHERF cross)	0.9	NDSLL[c]	1.7	1.7	1.6	**3.7**	1.0
1MFG	Erbin (cross)	1.7	WETWV[c]	3.3	3.5	4.4	**3.5**	2.3

[a] Backbone RMSD between the native PDZ domain and the one used for docking. [b] Obtained from PDZ-DocScheme 4. [c] Docking was performed for a longer sequence, but the RMSD and score are calculated for the last five residues only.

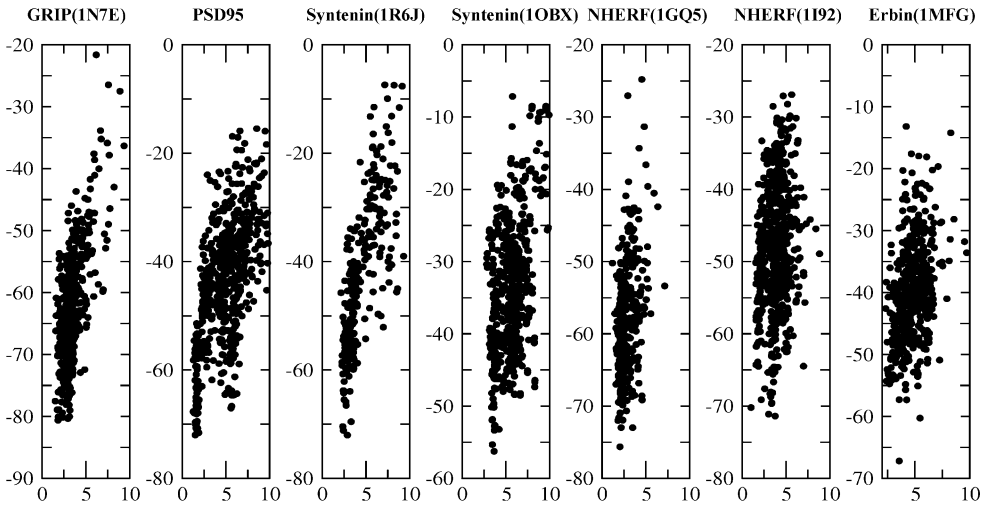

Figure 3. Cross-docking (docking to experimental non-native structures) using PDZ-DocScheme 4. Scoring energy (sum of peptide−protein interaction and internal peptide energy calculated in CHARMM22) is plotted vs heavy atoms RMSD between the peptides docked to non-native structures with PDZ-DocScheme 4 and the native peptides.

of PDZ−peptide complexes with contacts between the peptide and the other complex. Clearly, this situation is not reproduced in the docking experiment. Erbin presents an interesting case as well: the results from PDZ-DocScheme 2 and from PDZ-DocScheme 4 have RMSD of 0.2 Å from each other, effectively representing an identical pose. For this pose, the RMSD of four extreme residues (P(−3) through P0) from the native structure is 1.3 Å. The next residue, P(−4), drastically deviates from the native structure, in which Trp points toward the βB strand, by pointing toward the αB helix (Figure 4B). This P(−4) orientation is found in many other complexes (including GRIP, shown in Figure 4A, and in the Erbin/ErbB2 complex (1MFG). We thus propose that the best-ranking pose may represent an alternative feasible conformation of the peptide.

Docking to Homology Models. The ability to dock ligands to homology models is essential for the study of the vast majority of proteins for which the structure was not solved. Because the structural conservation of PDZ domains is high despite the relatively low sequence identity (≤30%), we were able to obtain good-quality homology models (as assessed from the backbone RMSD relative to the X-ray structures, shown in Table 4).

In general, both soft docking and flexible side-chain docking to these models perform well, misdocking only 1 out of 5 peptides. Even for the Dishevelled PDZ domain, which could not be modeled with high accuracy due to a long βB/βC loop, the four C-terminal residues of the peptide were docked with an RMSD of 2 Å. The inclusion of rotamer optimization step

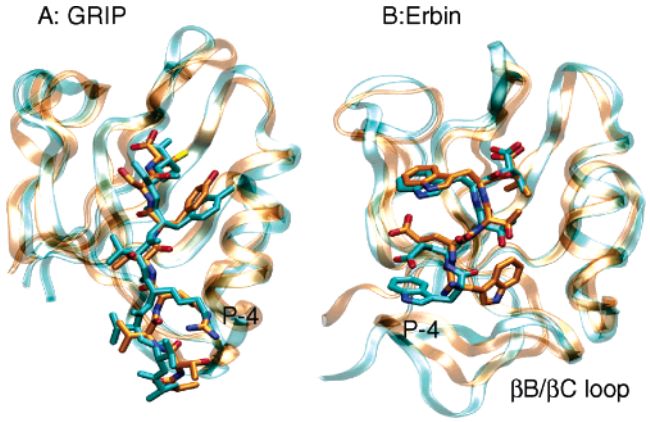

Figure 4. Cross-docked complexes. (A) GRIP PDZ domain is depicted in cyan ribbon, and the superimposed apo-structure is in orange ribbon. The native peptide is shown in cyan, and the best-scoring pose of the peptide docked to the apo-structure is colored by atom types (C, orange; N, blue; O, red). (B) The native Erbin PDZ domain (1N7T) is depicted in cyan ribbon, and the superimposed domain (from Erbin/ErbB2 complex, 1MFG) is in orange ribbon. The native peptide is shown in cyan, and the best-scoring pose of the peptide docked to the non-native structure is colored by atom types (C, orange; N, blue; O, red).

C generally improves the results both for soft-core procedures (PDZ-DocScheme 2 vs PDZ-DocScheme 1, Table 4) and for flexible binding site side-chain procedures (PDZ-DocScheme 4 (shown in bold) vs PDZ-DocScheme 3, Table 4). Since PDZ-DocScheme 4 outperformed PDZ-DocScheme 2 in the cross-docking test (Table 3), it has the best overall performance

Table 4. Docking of Peptides to Homology Models

PDZ	bb RMSD[a] (Å)	peptide sequence	PDZ-DocScheme 1	PDZ-DocScheme 2	PDZ-DocScheme 3	**PDZ-DocScheme 4**	best RMSD[b]
GRIP (model)	1.6	ATVRTYSC	3.4	2.5	5.1	**2.5**	2.5
PSD95 (model)	2.1	KQTSV	2.9	2.4	2.5	**2.6**	1.7
Syntenin (model)	1.6	NEFYA[c]	3.8	2.8	5.7	**4.1**	1.9
Dishevelled (model)	4.2	MTTV[c]	2.3	2.1	2.2	**2**	1.9
NHERF (model)	1.6	NDSLL[c]	3.5	3.5	4.1	**2.3**	1.6

[a] Backbone RMSD between the model and the native PDZ domain. [b] Obtained using PDZ-DocScheme 4. [c] Docking was performed for a longer sequence, but the RMSD and score are calculated for the last five residues for Syntenin and NHERF and the last four residues for Dishevelled.

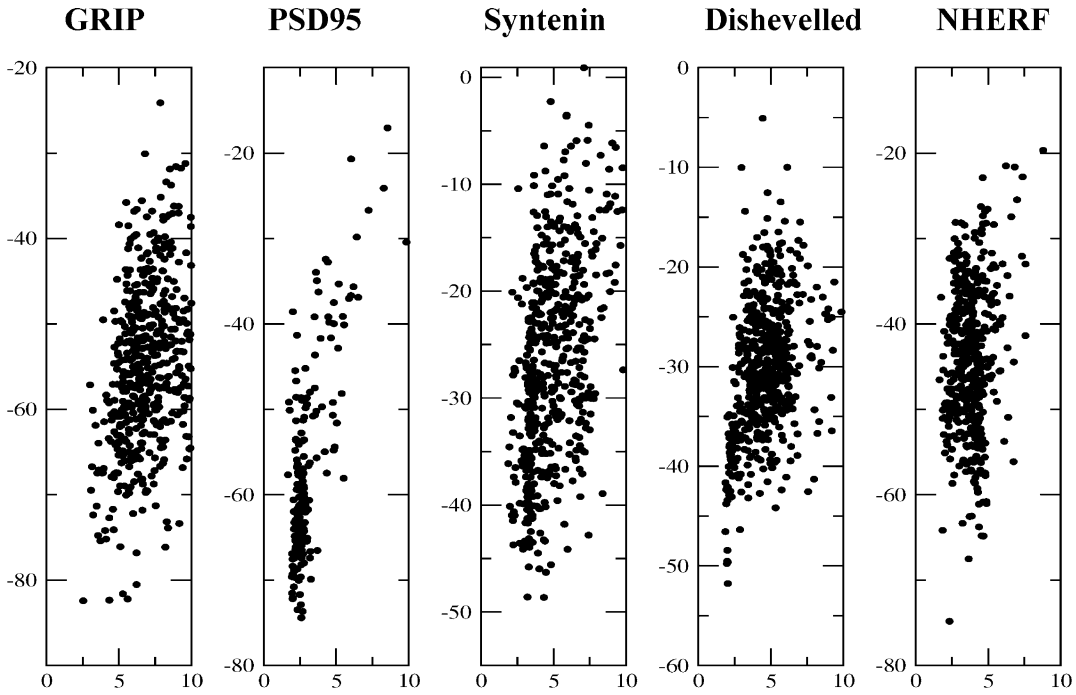

Figure 5. Docking to homology models using PDZ-DocScheme 4. Scoring energy (sum of peptide−protein interaction and internal peptide energy calculated in CHARMM22) is plotted vs heavy atoms RMSD between the peptides docked to homology models and the native peptides.

for docking to non-native structures, and results of docking peptides to homology models using PDZ-DocScheme 4 are presented in Figure 5.

As an example of this successful application, we show in Figure 6 the native GRIP complex and the best-scoring complex of peptide docked to the homology model. The superimposition was performed using the PDZ part of the complexes and results in good overlay of the modeled and the native peptides.

The PDZ-DocScheme 2 can serve to eliminate some structural biases introduced by artifacts from the flexible side-chain procedure, as illustrated by docking to the model of Syntenin. This is the worst result of docking using PDZ-DocScheme 4 (the best-scoring pose has RMSD 4.1 Å). This pose is artificial, however, because the structure is distorted by an intramolecular H-bond connecting P(−1) to P(−5). (A similar problem was encountered in redocking to the native Syntenin structure, with internal H-bonding producing a peptide docking pose with an RMSD of 6.3 Å from the native one.) The use of PDZ-DocScheme 2 yields a more realistic pose in this case (RMSD = 2.8 Å). (Discarding poses on the basis of divergence from a known binding mode has been successfully and systematically used by Wu and Vieth.[43]) Notably, however, the overall success of our flexible protein side-chain PDZ-DocScheme 4 is in some contrast with the findings of Taylor et al., that upon inclusion of protein side-chain flexibility the correct binding mode becomes energetically indistinguishable from other modes in most cases.[53]

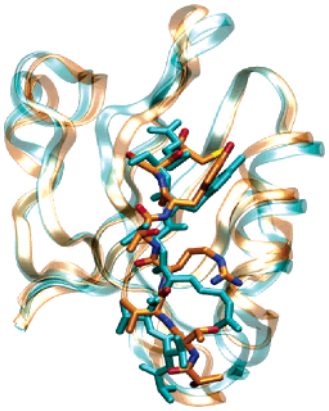

Figure 6. GRIP/liprinα peptide complex (1n7f) and liprinα peptide docked to GRIP homology model. GRIP PDZ domain is depicted in cyan ribbon, and the superimposed homology model is in orange ribbon. The native peptide is shown in cyan, and the best-scoring peptide (which, in this case, has also the best RMSD) is docked to the model structure and colored by atom types (C, orange; N, blue; O, red).

(53) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *J. Comput. Chem.* **2003**, *24*, 1637−1656.

Conclusions

In the most recent docking studies, ligands are typically represented as partially or fully flexible, while protein flexibility remains a challenge. Attempts have been made to introduce protein flexibility via soft docking, multiple copies, unbound dynamics of the protein, and side-chain flexibility (reviewed in refs 54 and 55). Here we have developed and tested a family of docking protocols to PDZ domains, PDZ-DocScheme. PDZ-DocScheme employs soft-core potential or protein side-chain flexibility in a simulated annealing molecular dynamics protocol followed by rotamer optimization. PDZ-DocScheme was applied to peptide docking cases that are challenging in two respects: first, the ligands are very flexible and contain 20−44 rotatable bonds; second, docking was tested not only to original known structures of the complexes, but also to other crystal structures of the same target PDZ domains and, most importantly, to homology models. Redocking to native structures using soft-core potential (PDZ-DocScheme 1) yielded excellent results for the highly flexible ligands tested here (RMSD < 2 Å). In docking to other structures and to homology models, we found that the best performance for all 12 test cases studied was achieved by PDZ-DocScheme 4. PDZ-DocScheme 4 starts from a peptide in extended conformation with the backbone of the P0 residue tethered to the canonical position. The peptide and the binding site side chains are heated with the rest of the protein frozen, and snapshots are taken from the hot trajectory and cooled. Optimal rotamers for every cooled pose are found using SCRWL.[44] The poses are minimized and scored by the sum of interaction and internal peptide energy. In 9 out of 12 cases in which PDZ-DocScheme 4 was applied to structures other than native, the best-scoring poses had RMSD < 2.8 Å for heavy atoms of peptides 4−8 amino acids long. The rotamer optimization step in the procedure was found to improve the RMSD in most of the tested cases. PDZ-DocScheme 2 (soft-core potential followed by rotamer optimization and minimization) was also shown to perform very well for docking to homology models. Its successful application was recently demonstrated in characterizing the complex binding specificity of the PICK1 PDZ domain to the C-terminus of dopamine transporter protein (DAT).[56] In that study, the binding results for a variety of derivative peptides measured with a novel fluorescence polarization-based binding assay were interpreted using structures obtained for peptides docked into a homology-based model of PICK1.[56] These models were used to suggest validating mutations. For example, a model of PICK1 binding a peptide corresponding to eight C-terminal residues of the DAT (-TLHRWLKV) highlighted the interaction of the peptide with the αB1 position, which is occupied by a lysine (Lys^{83}). In the model, the aliphatic chain of Lys^{83} is part of a hydrophobic pocket that also includes Val^{84}, Val^{86}, and Ala^{87} and accommodates the leucine at the P(−2) position of the ligand. According to the model, there are no interactions between residues in the peptide and the charged headgroup of Lys^{83}. This led to the inference that the aliphatic chain of Lys^{83} acts as a hydrophobic residue present in regular type II PDZ domains and, accordingly, that its charge would not contribute much to affinity. To test this hypothesis and thus mimic a canonical type II interaction, the atypical lysine in the αB1 position was substituted into valine, a hydrophobic residue commonly seen at this position. The substitution (K83V) was predicted to fully preserve the hydrophobic pocket. Moreover, the presence of the additional hydrophobic residues in the P(−2) pocket (Val^{84}, Val^{86}, and Ala^{87}) was predicted to enable a number of favorable hydrophobic interactions of the DAT peptide in the K83V mutants as well. In agreement with these predictions, the experimental data showed that the affinity for the DAT peptide increased slightly in the uncharged mutant construct.

Structure-based information provided from valid docking approaches should offer useful insights into complex selectivity patterns observed for PDZ domains. The importance of PDZ domains as anchoring and adaptor modules in cellular signaling pathways are major reasons for the great attention they have received. In addition, PDZ domains were recently highlighted as promising drug design target candidates.[57] Docking methods that could be used with low-resolution structures[58−60] and homology models as demonstrated in this work and in ref 35 should be very useful both in modeling protein interactions in signaling and in drug design, possibly in conjunction with peptide library techniques.[61] Further development and refinement of such methods seems well worth pursuing.

Acknowledgment. We thank Dr. Ernest Mehler and Thijs Beuming for helpful discussions and Dr. Michal Vieth for providing the CDOCKER input file. 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.

Supporting Information Available: Complete ref 45. This material is available free of charge via the Internet at http://pubs.acs.org.

JA054195S

(54) Carlson, H. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(55) Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. *Curr. Pharm. Des.* **2005**, *11*, 323−333.
(56) Madsen, K. L.; Beuming, T.; Niv, M. Y.; Chang, C. W.; Dev, K. K.; Weinstein, H.; Gether, U. *J. Biol. Chem.* **2005**, *280*, 20539−20548.
(57) Dev, K. K. *Nat. Rev. Drug Discov.* **2004**, *3*, 1047−1056.
(58) Wojciechowski, M.; Skolnick, J. *J. Comput. Chem.* **2002**, *23*, 189−197.
(59) Bindewald, E.; Skolnick, J. *J. Comput. Chem.* **2005**, *26*, 374−383.
(60) Vakser, I. A. *Biopolymers* **1996**, *39*, 455−464.
(61) Kay, B. K.; Kehoe, J. W. *Chem. Biol.* **2004**, *11*, 423−425.