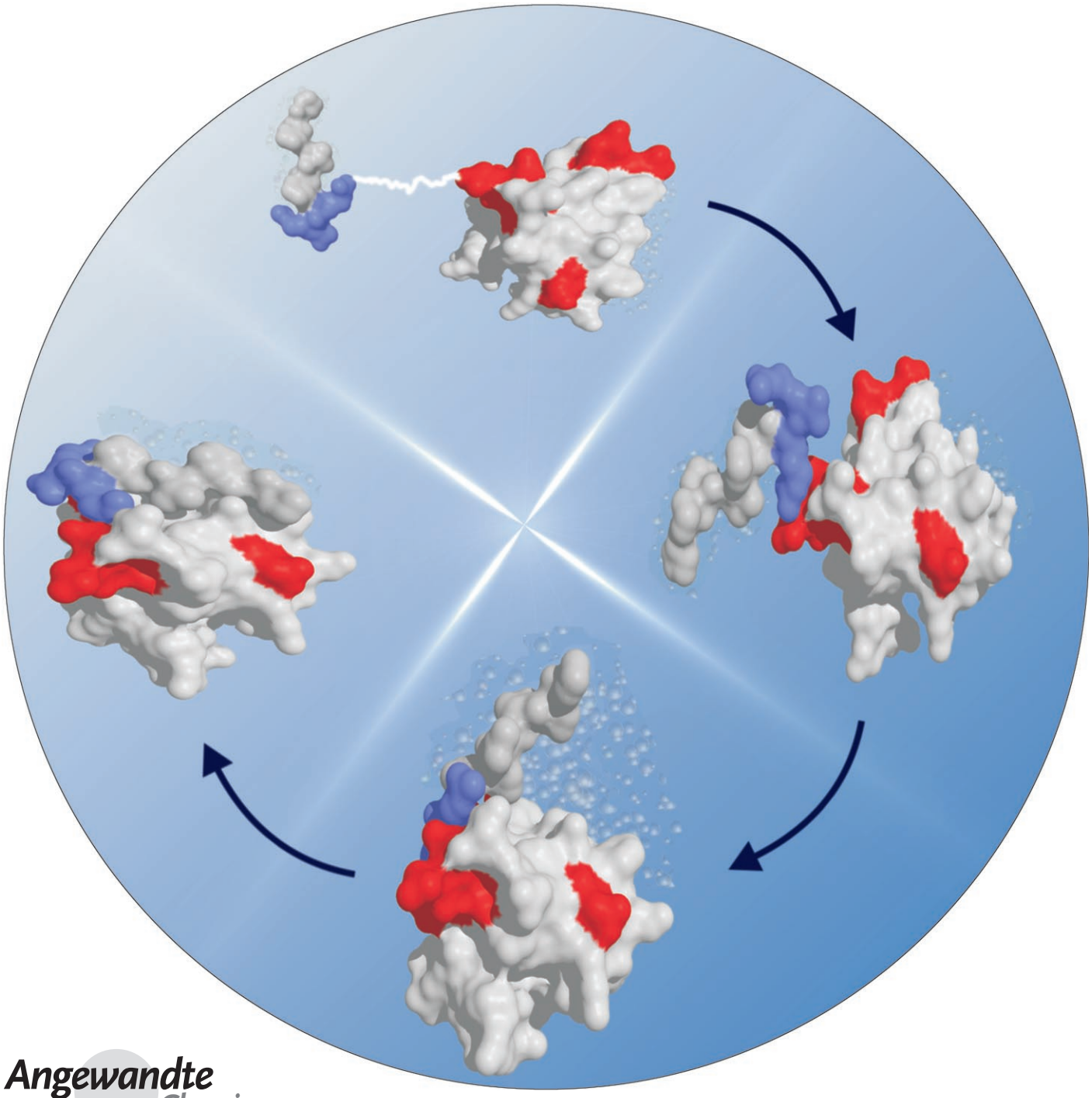
Peptide Recognition

DOI: 10.1002/anie.200801856

Mechanism of Fast Peptide Recognition by SH3 Domains**

*Mazen Ahmad, Wei Gu, and Volkhard Helms**



Protein–protein and protein–peptide recognition play central roles in the regulation of cellular processes, and much work has been devoted to unraveling the mechanistic details of these processes during the past decades.[1–4] However, our understanding of "the binding event" still awaits more detailed information from experiment and advances in the computational performance so that dynamic simulations may be extended to longer simulation times on which these events take place. It is commonly believed that formation of protein–protein complexes follows a pathway from a diffusive phase through one or more intermediate states to the final stereospecific complex.[5] Experimental studies by site-directed mutagenesis[1,2] and computational studies by Brownian dynamics simulations[3,4] have shown the important role of long-range electrostatic interactions in the diffusive phase. By using NMR paramagnetic relaxation enhancement (NMR-PRE), Clore and co-workers have recently demonstrated the existence and visualized the distribution of an ensemble of transient, nonspecific encounter complexes for a relatively weak protein–protein complex.[6] In spite of these advances, our mechanistic understanding of the transformation from nonspecific transient encounter complexes to the final stereospecific stable complex is still limited. Clearly, two of the main challenges for computational modeling of protein binding events are the role of solvent and the time scale of the binding events.

We present herein results on the mechanism of how proline-rich motifs (PRMs) are recognized by SH3 domains,[7,8] which are one of the most abundant protein interaction domains (at least 182 SH3 domains in the human genome sequence[9]), with many different biological roles. Protein recognition by SH3 domains is known to be mediated by a short PRM.[7,8] The canonical peptide-binding pocket of SH3 domains consists of a hydrophobic surface patch including two grooves that accommodate Px and xP residues of the peptide (see Figure 1 A). The flanking positively charged arginines of the peptide usually form contacts with the negatively charged residues in the RT and n-sCr loops of the SH3 domain. In the crystal structure of the complex the peptide adopts a polyproline type-II helix conformation (PPII), which is the known binding conformation for the proline-rich motifs bound to SH3 domains and other proline-recognition domains.[7,8]

We conducted MD simulations from different unbound conformations for the C-CRK N-terminal SH3 domain with a PRM for which a crystal structure of their complex is available.[10] In all simulations, we observed a relatively fast diffusive phase (see Figure S3 in the Supplementary Information) leading to the formation of nonspecific encounter complexes (see Figure 2 A) stabilized by salt bridges between the oppositely charged residues in the domain and the peptide. In six out of thirteen simulations, the encounter complexes led to stable stereospecific complexes involving three different binding modes (1–3). We defined these complexes as stereospecific binding modes based on the comparison with experimentally known structures of complexes for SH3 domains. The determining step for these modes is the formation of the transient complex, where the peptide arginine forms salt bridges with the negatively charged residues in the RT or n-sCr loops leading to the binding mode of the crystal structure or to binding in the new pocket, or with the residue D142 leading to the binding with the peptide in opposite orientation:

1) Binding mode of the crystal structure (Figure 1 B): One simulation converged to a conformation similar to the crystal structure. Already during the first 50 ns of the simulation it formed a considerable number of native contacts. We extended this simulation to 150 ns, and it came very close to the crystal structure after 130 ns (Figure 1 B). The root-mean-square displacement (RMSD) for the whole complex backbone from the backbone of the crystal structure is (1.3 ± 0.2) Å averaged over the last 20 ns of the simulation. It contained the known native contacts in the crystal structure (Figure 2 B)

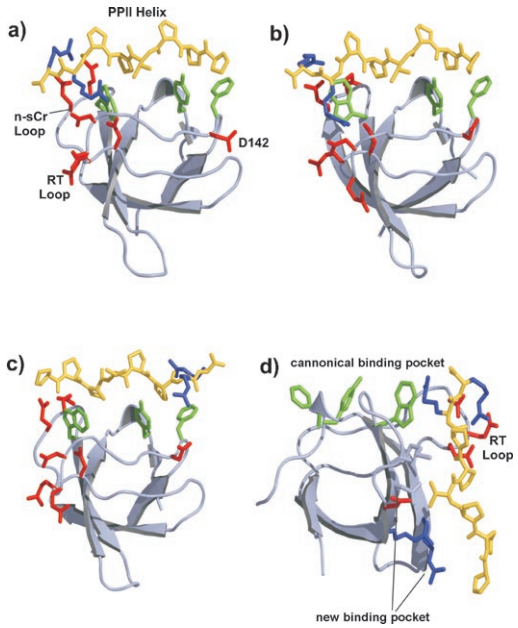


Figure 1. The SH3 domain complex studied here and the recognized binding modes. a) Crystal structure of the domain with the binding motif (PDB code 1ckb). The side chains of the negatively charged residues are colored red, the hydrophobic pocket in the domain is shown in green, the PPII helix is yellow, and the arginines are blue. b) MD snapshot at 130 ns in which the binding mode of the crystal structure is recovered. c) Observed peptide binding in the orientation opposite to that in the crystal structure. d) Observed binding mode in the new pocket. The graphics were generated using PYMOL.[29]

[*] M. Ahmad, Dr. W. Gu, Prof. Dr. V. Helms
Zentrum für Bioinformatik
Universität des Saarlandes
66041 Saarbrücken (Germany)
Fax: (+49) 681-302-64180
E-mail: volkhard.helms@bioinformatik.uni-saarland.de
Homepage: http://gepard.bioinformatik.uni-saarland.de

[**] M.A. acknowledges Tishreen University for a predoctoral fellowship. This project was funded in part by Volkswagenstifung (Project I/80469).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200801856.

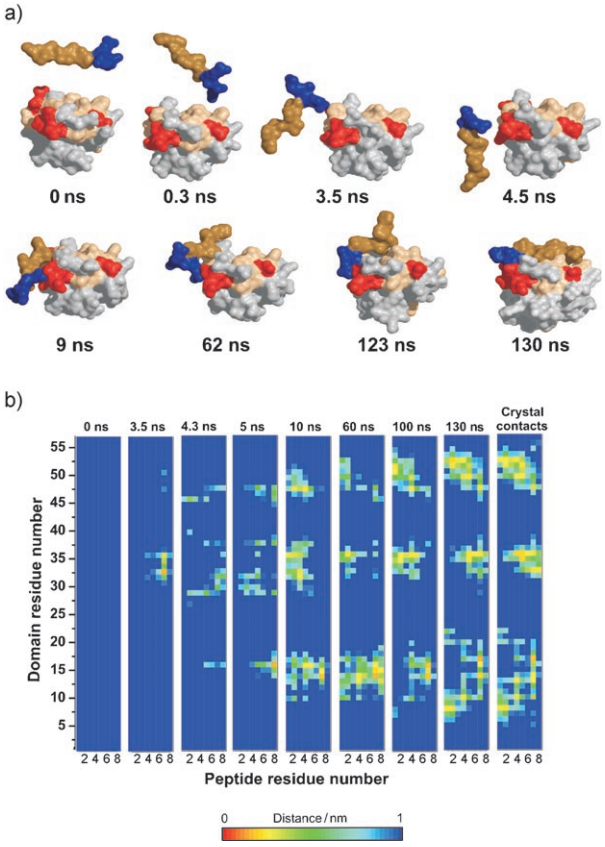


Figure 2. Conformational snapshots along the binding pathway that leads to the complex with the orientation found in the crystal structure. a) Snapshots at different simulation times showing the transformation from the starting structure to the final complex through transient encounter complexes. Solvent molecules are not shown. b) Maps of the contacts between the domain and the peptide for different snapshots in comparison to those in the crystal structure. A movie is available in the Supporting Information.

and a polyproline type-II helix conformation for the peptide.

2) Binding mode with opposite orientation: Three simulations converged within 20–30 ns of simulation time to a conformation with the peptide bound in the same canonical pocket as in the crystal structure but in the opposite orientation (Figure 1C). The peptide adopts a PPII helix conformation as well and is symmetric (opposite orientation) to the one in the crystal structure. One of the peptide arginines forms a salt bridge with residue D142 in the domain during the transient encounter stage, but this contact was not permanently established in the final complex. Interestingly, aspartic acid residues are frequently found in this position among SH3 domains. These complexes were stable during the remaining simulation time (up to 50 ns). The possibility of SH3 domains binding the PRMs in two opposite orientations is well characterized[11] and has been found in many other proline-recognition domains as well.[7,8] The structural basis for this is the symmetry of the PPII helix which enables packing into two different orientations in the same binding pocket. This novel binding mode for the C-CRK SH3 domain described here suggests the possibility of the same peptide binding in two different orientations to the same SH3 domain. To our knowledge, this has not yet been described experimentally. Estimating the populations of the two orientations for the same peptide with the same domain is beyond the aim of this paper.

3) Binding in the new pocket: In two of our simulations the peptide bound in a different pocket of the SH3 domain within 30 ns of simulation time (Figure 1D). These complexes were also stable during the remaining 20 ns of the simulation time. The role of this pocket in peptide binding by SH3 domains has been discovered in a recent crystal structure[12] (PDB code 2p4r) where the SH3 domain showed the possibility of binding a peptide with two PRMs. The first motif bound to the canonical pocket and the other motif bound to this face of the domain. Moreover, this pocket has recently been observed in another new structure (PDB code 2drm) as well.

Concerning the mechanism of binding, the pathway of complex formation was found to consist of three phases. A fast diffusive phase leads to the formation of various electrostatically stabilized intermediate complexes, of which some can bind into the stereospecific stable bound complex. Here we define the diffusive phase as the time during which both reacting chains diffuse before they form short-range contacts. Although the diffusion phase was quite short in our simulations, the simulations showed little dependence on the starting structures of the peptide. In many of the thirteen simulations the peptide completely changed its orientation during the diffusive phase. As expected, the electrostatic interactions between the oppositely charged residues in both proteins play an essential role. They guide and accelerate the diffusion to terminate by forming a nonspecific complex stabilized by salt bridges between the side chains of arginine in the peptide and the negatively charged residues in the domain (RT and n-sCr loops or residue D142). In control simulations a mutant peptide carrying two R→A mutations did not form stable contacts within 20 ns (see the Supporting Information).

Experiments show that the electrostatically accelerated association of proteins is very rapid.[1] In particular, electrostatic acceleration increases the affinity by increasing the association rate (k_{on}), without affecting the disassociation rate (k_{off}).[2] Such acceleration of diffusion can be critical for protein–protein association if the association is diffusion controlled or influenced.[13] The presence of many negatively charged residues on the surface of the domain therefore results in an ensemble of transient complexes and a binding process with multiple pathways. These observations agree with the picture of encounter complexes emerging from experimental data.[6,14] The electrostatic nature of the intermediate states in protein–protein association has also been characterized using double mutant cycles.[15] Moreover, NMR-PRE experiments revealed a correlation between the spatial distribution of nonspecific encounter complexes and the electrostatic potential isosurface.[6] The population of nonspecific encounter complexes was shown to be significantly more affected by ionic strength than that of the

stereospecific complex, demonstrating the importance of electrostatic interactions in the formation of the ensemble of nonspecific encounter complexes.[16] Those salt bridges in the transient encounter complexes that led to the stereospecific complex are close to or part of the possible contacts in the last stable complex. This enables these encounter complexes to run a search process with fewer degrees of freedom, resulting in a "reduction-in-dimensionality"[3] to find the final stereospecific complex. The main role of electrostatic interactions in diffusion and stabilization of the transient encounter complexes explains the importance of charged residues in the binding motifs for SH3 domains. The positively charged residues in the binding motifs (R, K) are essential for binding, where the consensus (PxxPxR/K) is essential for class II motifs and (R/KxxPxxP) for class I.[7,8]

One of the current unknowns in the area of protein–protein interactions is the transition from the transient complexes to the specific complex which involves depletion of solvent and docking of the interfaces. To investigate the possible role of dewetting at hydrophobic interfaces prior to binding, we calculated the water density in the intermolecular gaps for six snapshots out of the transient stage of the simulation that led close to the crystal structure. Interestingly, the intermolecular gaps showed a significant decrease in water density (see the Supporting Information), which indicates clear partial dewetting of the interfaces before binding. To characterize the distance dependence of this effect, we ran ten further similar simulations starting from the crystal structure after displacing the peptide by distances of 2.5–9.0 Å along the connection of the two centers of geometry for the peptide and the domain (see Figure 3), and keeping it there by harmonic restraints. For comparison we also ran two simulations for the peptide alone (and for the domain alone) and computed the water density inside a virtual gap with the same volume and shape as in the complexes. Even the hydrophobic pockets in the separate domain and around the PPII helix in the separate peptide showed a reduced water density at the hydrophobic surfaces. However, this effect is more pronounced in the complexes, where a significant degree of dewetting is found for all complexes with interfacial distances of < 5 Å. Recent work has added to the understanding of the hydrophobic effect as a major component of the forces that fold and stabilize the structure of biomolecules.[17–22] For example, Lum et al.[23] argued that a vaporlike layer forms around large hydrophobic surfaces and showed this effect in simulations. Zhou et al.[20] also showed a distance dependence of dewetting effects during MD simulations for protein folding. Control simulations (see the Supporting Information) showed that the partial dewetting accelerates the collapse of the peptide into its binding pocket. This suggests that dewetting can guide the search process for the final specific complex after the reduction in dimensionality in the transient state. In the light of these new findings we can now explain why the transient complexes close to the final complex will converge so quickly that they have short life times and low occupancy.[6]

An important question to ask at the end is how relevant this model is for understanding protein–protein recognition in general. First of all, the existence of a large hydrophobic aromatic pocket is a common feature in many proline-recognition domains. Yet, the hydrophobic nature of protein–protein interfaces is even well known as a general principle. Typically, between 30–50 % of the protein interface area is taken up by hydrophobic amino acids.[24,25] On the other hand, the presence of salt bridges at protein–protein interfaces is a general feature, too, because on average two ion pairs per interface were found.[26] Therefore, we suggest that the mechanism found here to guide the association of the C-CRK N-terminal SH3 domain and its peptide-binding motif applies to many other protein pairs too.

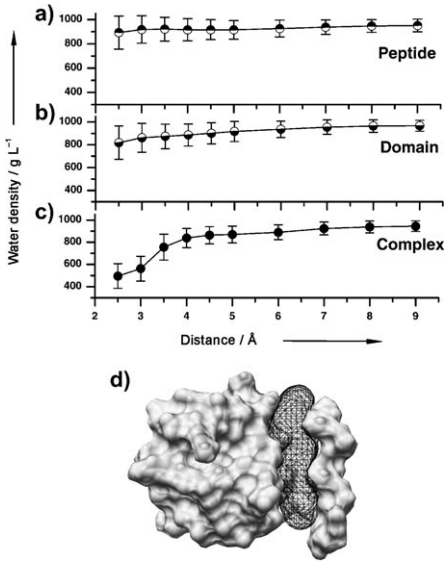


Figure 3. Water density in the hydrophobic interfacial gap with the peptide at fixed distance. a) Water density in the interfacial gap for the peptide alone, b) for the domain alone, and c) for the complex. The water density was averaged over the last 1.7 ns for each 2 ns position-restraint MD simulation. The gap volumes for the free peptide and the free domain were defined by superposition of the corresponding part from the complex simulation to all snapshots of the protein–water system. d) Representation of the defined interfacial gap between the PPII helix and the hydrophobic pocket in the domain at an interfacial distance of 4 Å.

One lesson we have learned from the simulations is the synergistic nature of the driving forces for binding (Figure 4). The long-range electrostatic effects play the main role during diffusion and stabilize the transient complexes formed by the electrostatic parts in the interface. At short distances, this then enables the partial-dewetting effect to increase the probability for the collapse of the hydrophobic part of the interface and the convergence to the final specific complex. This model is one example for the simplicity of protein recognition in spite of the apparent complexity of this process. As in many other cases, it appears that nature does not just throw dice here.

Experimental Section

The starting structure of the simulation was extracted from a crystal structure of a complex for the C-CRK N-terminal SH3 domain (PDB code 1ckb) with a PRM.[10] Thirteen independent unbiased atomistic

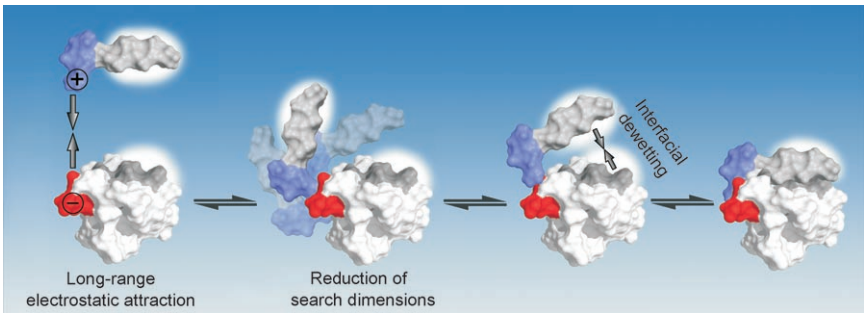


Figure 4. Mechanism of the binding process. The positively charged and negatively charged residues are colored blue and red, respectively. The hydrophobic interfaces are colored gray. Dewetting of the hydrophobic interfaces of the peptide and the SH3 domain is indicated by white clouds around the interfaces.

molecular dynamics simulations in explicit solvent and 50–150 ns in length were started from different unbound conformations at minimum distances of 13–20 Å using the GROMACS simulation package.[27] The total length of simulation times is 0.85 μs (details in the Supporting Information). The intermolecular gaps were defined by using the program SURFNET.[28]

Received: April 21, 2008
Revised: July 3, 2008
Published online: August 27, 2008

Keywords: adaptor domains · hydrophobic dewetting · molecular dynamics simulation · proline-rich motifs · proteins

[1] G. Schreiber, A. R. Fersht, *Nat. Struct. Biol.* **1996**, *3*, 427.
[2] T. Selzer, S. Albeck, G. Schreiber, *Nat. Struct. Biol.* **2000**, *7*, 537.
[3] S. H. Northrup, J. O. Boles, J. C. Reynolds, *Science* **1988**, *241*, 67.
[4] A. Spaar, C. Dammer, R. R. Gabdoulline, R. C. Wade, V. Helms, *Biophys. J.* **2005**, *90*, 1913.
[5] G. Schreiber, *Curr. Opin. Struct. Biol.* **2002**, *12*, 41.
[6] C. Tang, J. Iwahara, G. M. Clore, *Nature* **2006**, *444*, 383.
[7] L. J. Ball, R. Kuhne, J. Schneider-Mergener, H. Oschkinat, *Angew. Chem.* **2005**, *117*, 2912; *Angew. Chem. Int. Ed.* **2005**, *44*, 2852.
[8] A. Zarrinpar, R. P. Bhattacharyya, W. A. Lim, *Sci. STKE* **2003**, *2003*, re8.
[9] J. C. Venter et al., *Science* **2001**, *291*, 1304.
[10] X. Wu, B. Knudsen, S. M. Feller, J. Zheng, A. Sali, D. Cowburn, H. Hanafusa, J. Kuriyan, *Structure* **1995**, *3*, 215.
[11] M. Saraste, A. Musacchio, *Nat. Struct. Biol.* **1994**, *1*, 835.
[12] J. M. Janz, T. P. Sakmar, K. C. Min, *J. Biol. Chem.* **2007**, *282*, 28893.
[13] R. R. Gabdoulline, R. C. Wade, *J. Mol. Biol.* **2001**, *306*, 1139.
[14] T. L. Blundell, J. Fernandez-Recio, *Nature* **2006**, *444*, 279.
[15] G. Schreiber, A. R. Fersht, *J. Mol. Biol.* **1995**, *248*, 478.
[16] J. Y. Suh, C. Tang, G. M. Clore, *J. Am. Chem. Soc.* **2007**, *129*, 12954.
[17] D. Chandler, *Nature* **2005**, *437*, 640.
[18] P. Ball, *Nature* **2003**, *423*, 25.
[19] P. Liu, X. Huang, R. Zhou, B. J. Berne, *Nature* **2005**, *437*, 159.
[20] R. Zhou, X. Huang, C. J. Margulis, B. J. Berne, *Science* **2004**, *305*, 1605.
[21] N. Choudhury, B. M. Pettitt, *J. Am. Chem. Soc.* **2007**, *129*, 4847.
[22] J. Dzubiella, J. M. J. Swanson, J. A. McCammon, *J. Chem. Phys.* **2006**, *124*, 084905.
[23] K. Lum, D. Chandler, J. D. Weeks, *J. Phys. Chem. B* **1999**, *103*, 4570.
[24] L. Lo Conte, C. Chothia, J. Janin, *J. Mol. Biol.* **1999**, *285*, 2177.
[25] S. Ansari, V. Helms, *Proteins Struct. Funct. Genet.* **2005**, *61*, 344.
[26] D. Xu, C. J. Tsai, R. Nussinov, *Protein Eng.* **1997**, *10*, 999.
[27] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.
[28] R. A. Laskowski, *J. Mol. Graph.* **1995**, *13*, 323.
[29] W. L. DeLano, DeLano Scientific, San Carlos, CA, **2002**.

