JCTC Journal of Chemical Theory and Computation

Letter

pubs.acs.org/JCTC

Visualizing the Induced Binding of SH2-Phosphopeptide

T. Giorgino,[†,‡,§] I. Buch,[†,§] and G. De Fabritiis*[,†]

[†]Computational Biochemistry and Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park, C/Dr. Aiguader 88, 08003 Barcelona, Spain

[‡]Institute of Biomedical Engineering, CNR - National Research Council of Italy, Corso Stati Uniti 4, 35127 Padova, Italy

Supporting Information

ABSTRACT: Approximately 100 proteins in the human genome contain an SH2 domain recognizing small flexible phosphopeptides. It is therefore important to understand in atomistic detail the way these peptides bind and the conformational changes that take place upon binding. Here, we obtained several spontaneous binding events between the p56 lck SH2 domain and the pYEEI peptide within 2 Å RMSD from the crystal structure and with kinetic rates compatible with experiments using high-throughput molecular dynamics simulations. Binding is achieved in two phases, fast contacts of the charged phosphotyrosine and then rearrangement of the ligand involving the stabilization of two important loops in the SH2 domain. These observations provide insights into the binding pathways and induced conformations of the SH2−phosphopeptide complex which, due to the characteristics of SH2 domains, should be relevant for other SH2 recognition peptides.

Src homology 2 (SH2) domains are small (∼100 amino acids) and well-conserved protein domains[1] which recognize with high affinity and specificity short amino acid sequences (three to six residues) containing a phosphorylated tyrosine. The recognized phosphopeptide with sequence pYEEI binds the p56lck SH2 domain, with nanomolar affinity,[2,3] approximately perpendicular to the central rigid β sheet formed by three antiparallel strands. The structure, common to all SH2 domains,[4] defines a hydrophilic pocket below the BC loop face and hydrophobic cavity between the EF and BG loops (Figure 1A). Experimental studies have elucidated the structural determinants[5] and the contributions of the ligand's sequence[6] to the binding specificity but do not directly provide the sequence of dynamical molecular events that lead to the final bound pose. In particular, flexibility in the BC loop and EF-BG regions has been previously observed, but its functional role in the pathways of the association is still to be specified.[7]

The recovery of binding and unbinding trajectories has been possible, so far, through biased molecular dynamics (MD).[8−10] More recently, unbiased MD has been used to directly simulate binding and folding processes, incurring a larger computational cost, but without making a priori assumptions on the process, thus reproducing the pathways and binding kinetics of the process up to the accuracy allowed by the atomistic force field.[11−14] In this work, we performed high-throughput unbiased MD simulations on initially unbound pYEEI to elucidate the pathway leading to association with SH2. Using a protocol similar to ref 14, the protein was modeled with the CHARMM27 all-atom force-field in explicit TIP3P water and simulated in the NVT ensemble at a temperature of 295 K. An ensemble of 772 free diffusion trajectories, each 200 ns long, was started with the ligand placed 40 Å away from the binding site in a periodic box of 60 × 66 × 98 Å^3; a slightly smaller flat-bottomed restraint box was created around the protein to increase the ligand concentration (Figure 2; see also Supporting Information). The simulations were performed with the ACEMD software[15] on the GPUGRID distributed computing network.[10]

After 200 ns, five out of the 772 trajectories reached a bound configuration, defined as the ligand's backbone RMSD < 2 Å with respect to the 1LKK crystal structure (Figure 1B; Supporting Figures S1 and S2). From the number of events observed, it is possible to determine an order-of-magnitude estimate for the *on* binding kinetics, assuming first-order kinetics,[11] which yields $k_{on} = (1.6 \pm 0.8) \times 10^6$ M^{-1} s^{-1} (95% confidence interval: 0.6−3.7 × 10^6 M^{-1} s^{-1}; see Supporting Information). The standard binding free energy $\Delta G^\circ = -9.0 \pm 0.5$ kcal/mol was computed using a protocol based on umbrella sampling with 51 windows along the reaction coordinate and 10 independent replicas per window[16] in good agreement with the experimental value of −8.0 kcal/mol. The kinetic constant k_{off} can be therefore estimated indirectly using the relation $K_D = k_{off}/k_{on}$, yielding a result on the order of 10^{-1} s^{-1}. The k_{on} value is also within the range of those obtained in vitro on various ligands,[2,17,18] showing that the number of events recovered is therefore consistent with its experimental kinetics.

The binding trajectories (T1−T5) were examined in order to gain some insight into the mechanism of binding and verify the existence of principal binding pathways (Supporting Video S1, the binding trajectories are available on request). The first step in binding is diffusion of the ligand in the bulk. Phosphopeptide association is determined[19−21] by the double negative charge on the phosphotyrosine and the positive charge in the corresponding binding pocket. The combined effect of diffusion and electrostatic forces drives the ligand from 40 Å away to the vicinity of the binding surface. In at least four of the five trajectories, the charged pY(+0) residue forms clear first contacts with the protruding side chains of Lys182(βD6) and Arg134(αA2), which thus constitute a "capture set" that drives the association through an encounter complex (Figure 3A,B;

Received: January 4, 2012
Published: March 10, 2012

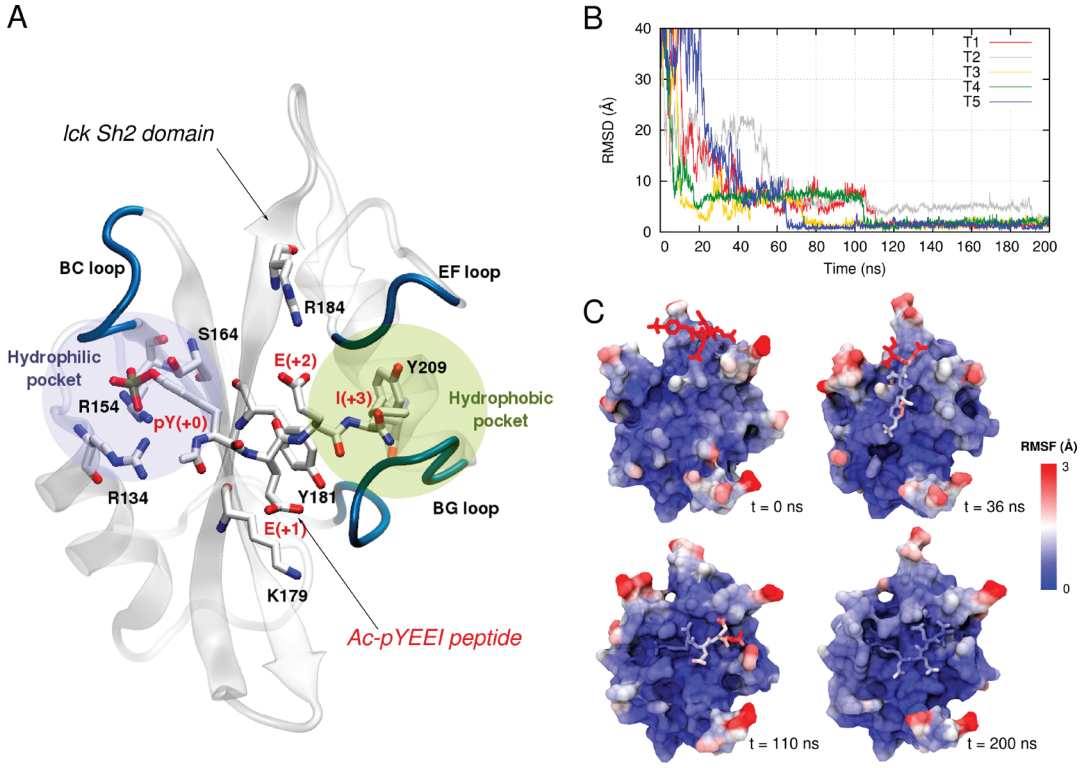

Figure 1. (A) The bound pose in the 1LKK PDB crystal structure, with the protein's secondary structure (transparent) and relevant loops highlighted (blue). The Ac-pY-E-E-I peptide (sticks) plugs into two pockets: a hydrophobic one shown on the left, "proximal", which buries phosphotyrosine pY(+0), and a hydrophobic one shown on the right, "distal", accommodating I(+3). Significant residues forming native contacts between the protein and the peptide are labeled in black and red, respectively. Secondary structure elements are named according to Eck et al.[4] (B) RMSD of the ligand's backbone in all five selected trajectories T1−T5 with respect to time. At 200 ns, all trajectories reached the imposed cutoff for binding, i.e., RMSD < 2 Å. (C) Snapshots of trajectory T1 at different times, shown with the same orientation as in subfigure A. The protein surface and the ligand's atoms are color-coded to highlight the local RMSF (1 ns window), on a scale from blue (lowest fluctuation) to red (highest fluctuation). A "freezing" of the ligand and BC loop can be observed as binding proceeds.

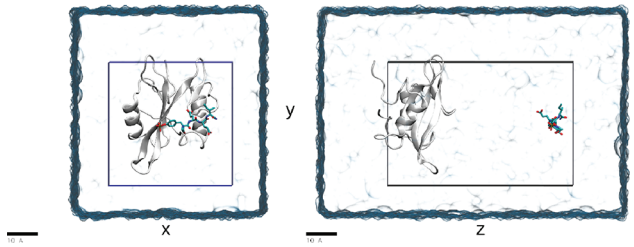

Figure 2. Simulation cell (rough surface, 60 × 66 × 98 Å^3) and flat bottom box (smooth surface, 40 × 40 × 60 Å^3) used for the production simulations, shown as projections in the *xy* (left) and *zy* (right) planes. The pYEEI ligand is shown in its starting position, 40 Å away from the bound pose. An ensemble of 772 production simulations of 200 ns each was performed in the NVT ensemble at 295 K and explicit TIP3P[29] water.

Supporting Figures S3−S7). Once formed, a minimum of one of these contacts is maintained until the ligand achieves its final binding pose. The stabilizing role of the capture residues is exhibited by looking at the per atom root-mean-squared fluctuation (RMSF) in the encounter complex (Figure 1C; Supporting Video S2): as the phosphotyrosine points toward the protein, its RMSF is markedly reduced with respect to the rest of the ligand. Thus, in the encounter complex, the C-terminal residues of the peptide undergo a search in conformational space. After the capture set is established, the first step toward the establishment of complete binding occurs when pY(+0) slides into the charged pocket formed by the regions αA, βB, and βD and the BC loop, as seen in four of the five trajectories (Figure 3A-B; Supporting Figures S3−S7).

Charged residues E(+1) and E(+2) have preferential contacts in the native state: the side chain of E(+1) faces the bulk and makes intermittent, water-mediated contacts with Lys179(βD3) and Tyr181(βD5); E(+2) has a well-defined partner in Arg184(βD′1). In the native pose, I(+3) is buried in the hydrophobic cavity between the EF and BG loops (Figure 1A). Consistent with the findings by Bradshaw and Waksman,[22] the binding of residues +1 to +3 appears to be cooperative: these native contacts are established concurrently in three of the trajectories. Once formed, E(+1) and E(+2) are relatively stable; I(+3) is still relatively flexible (also seen in the control simulation, Supporting Figure S8).

Interestingly, a transient closure of the EF-BG hydrophobic pocket was observed in two of the simulations before binding (Figure 4A). The closed state is not compatible with I(+3) binding due to steric constraints. The transition of the EF-BG pocket to the open state is therefore a prerequisite for binding (Figure 4B). In trajectory T1, the distal pocket closed and reopened once before accommodating I(+3) (Figure 4C). Closure in the EF-BG pocket loop had also been observed in NMR measurements on the *v*-Src SH2.[23]

The charged pocket accommodating the phosphotyrosine is relatively buried; thus, plugging into it requires the BC loop to become transiently open (Figure 4A). The BC loop opened during the equilibration in absence of the ligand. In all of the

dx.doi.org/10.1021/ct300003f | J. Chem. Theory Comput. 2012, 8, 1171−1175

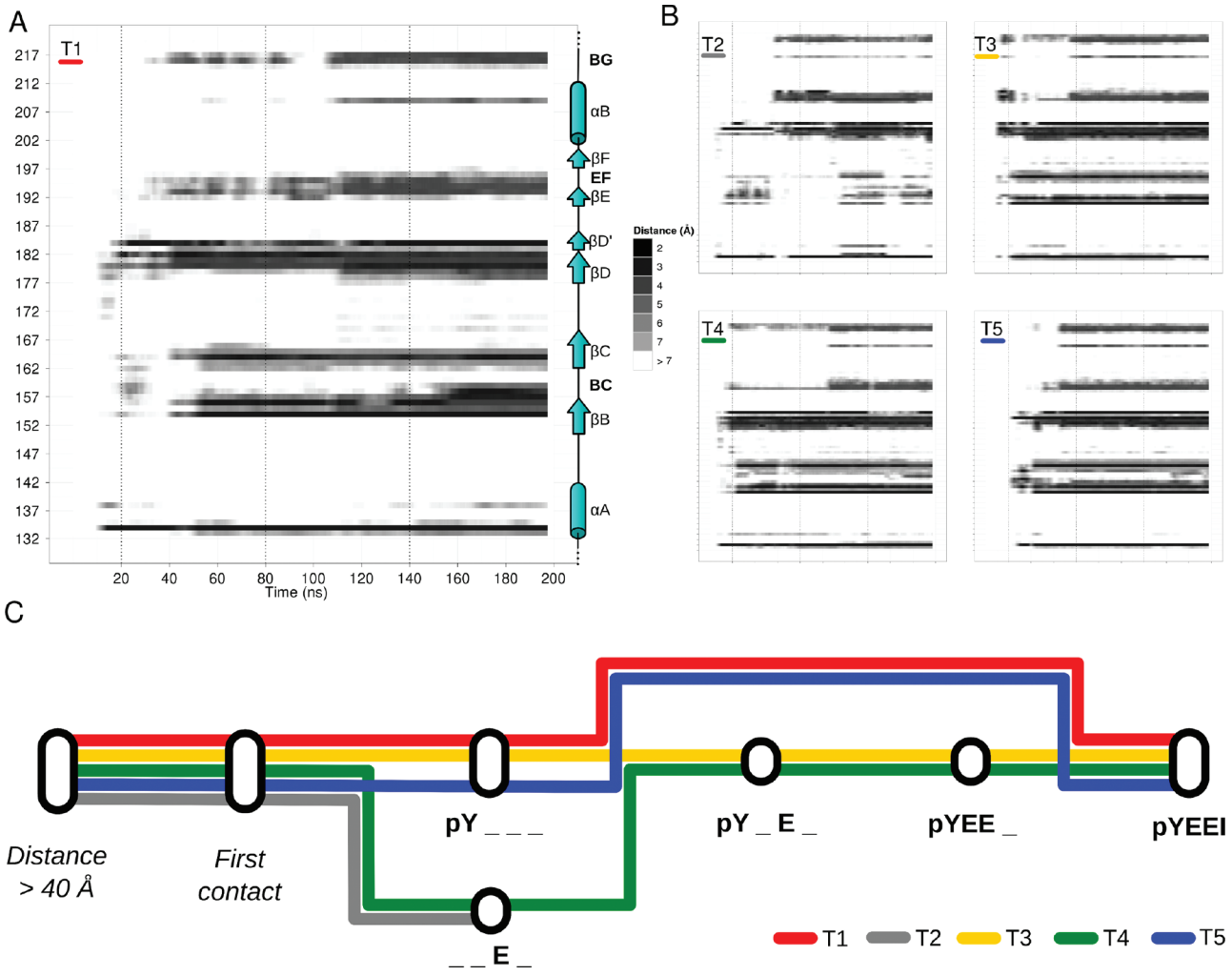

Figure 3. (A) Trajectory T1 residue-ligand distance timeline, highlighting the evolution of the ligand's contacts with each of the protein's residues. The charged pY(+0) residue forms first contacts with the protruding side chain of Arg134(αA2), part of "capture set" with Lys182(βD6). pY(+0) slides into an encounter complex at the charged pocket formed by the regions αA, βB, and βD and the BC loop. A swift transition to the native pose is visible around 110 ns, as is the BC loop (residues 157−161) locking at 160 ns. (B) Residue−ligand distance timelines for trajectories T2−T5. As expected, the ligand's fingerprint on the protein sequence is equivalent across all trajectories. (C) Network of the states, defined as combinations of native contacts, visited by four binding trajectories (T2 did not form native contacts after 200 ns). In T1 (red), T3 (yellow), and T5 (blue), the first residue of the ligand to form native contacts is the pY(+0), on average after 36 ns. In T4 (green), on the contrary, the first residue to form native contacts is E(+2) in 18 ns, almost immediately followed by pY(+0) at 22 ns. After the binding of pY, in T1 and T5 the ligand establishes simultaneous native contacts with the rest of the residues after 110 and 65 ns, respectively. In T3, the ligand forms the native contacts gradually, first with E(+2) in 45 ns, followed by E(+1) in 50 ns, and finally I(+3) in 75 ns. Similarly, T4 progressively binds E(+1) and I(+3) in 40 and 110 ns, respectively. The four trajectories reach the native bound pose by 110 ns. Additional data are available as Supporting Information.

simulations, it remained open (more than 4 Å RMSD from the crystal structure) and exhibited large fluctuations while in the *apo* form (RMSF ∼ 2.5 Å), consistent both with crystallographic experimental data[24] and with the variability between the structures' binding pockets (Figure 4C; Supporting Figures S3−S7).[25] The control simulation of the bound pose also showed fluctuation in the BC loop, being generally at 2.5 ± 1 Å RMSD from the crystal but occasionally fluctuating to about 5.0 Å (Supporting Figure S8). Such fluctuations are indicative that, despite the ligand-bound BC-closed conformation being energetically favored, the open state is occasionally visited as well. The BC loop is therefore expected to lock on the phosphotyrosine upon binding, thus increasing the buried surface and complex stability. BC loop locking was observed to occur within the simulated time in two out of the five reacting trajectories (Supporting Table S1).

Once the distinctive features of the binding trajectories have been identified, it is interesting to verify how consistently they are represented in the binding events collected so far. We therefore defined a set of quasi-bound states based on which combination of residues of the ligand was forming native contacts at a given time. Although, in principle, the native pose could be reached forming native contacts in any arbitrary order, again we see (Figure 3C; Supporting Video S1) a nonexclusive preferentiality for pathways in which pY(+0) binds first, and EEI follows cooperatively. Although a "single pathway" description that can be obtained from a single trajectory is an appealing simplification, the binding events are likely to occur through sequences of steps which exhibit a degree of variability. This observation is consistent to the variability recently described in different systems binding small molecules.[11,14] The network shown in Figure 3C can be formalized

dx.doi.org/10.1021/ct300003f | J. Chem. Theory Comput. 2012, 8, 1171−1175

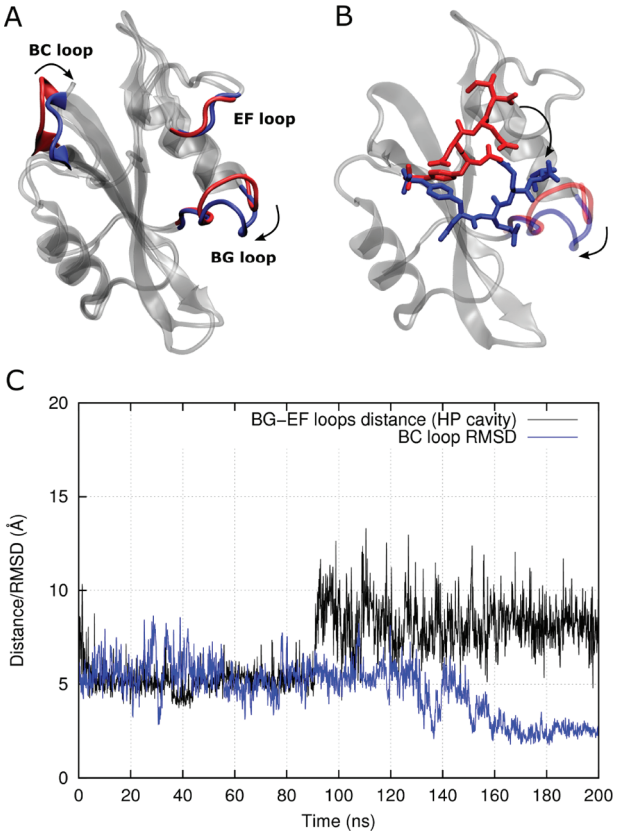

Figure 4. (A) Protein conformational changes required for binding include an increase in the EF−BG distance by opening of the BG loop, and BC loop closing and locking over pY(+0) once it binds to the proximal pocket. In red, loops are shown respectively in their open, flexible conformation (BC loop) and in their closed non-native conformation (BG loop). In blue, the native conformation of the BC loop and the binding-competent conformation for the BG loop are shown. (B) Opening of the BG loop increases the EF−BG loops gap and exposes the distal pocket, which is required for the accommodation of the ligands' hydrophobic residue I(+3), hence the native bound conformation. (C) In trajectory T1, a transient closure of the EF−BG (black) hydrophobic pocket is observed, opening at ∼90 ns, which allows the I(+3) residue to achieve its bound pose in the distal pocket (20 ns later); finally, closure of the BC loop on the ligand occurs at 160 ns (blue: RMSD with respect to the crystal).

quantitatively by constructing Markov state models from simulation data,[26] but sufficient sampling is required for all the transitions involved.

Both the locking of the BC loop and the opening of the hydrophobic pocket gate indicate a role for local flexibility. Although most of the protein acts as a rigid scaffold, the residual mobility of the loops protruding in the solvent (Supporting Video S2) gives rise to dock-lock (BC loop) or guard (the hydrophobic gap between EF and BG loops) mechanisms (Figure 4A). Our observations therefore provide a dynamical interpretation to both the induced-fit response proposed on the basis of the *apo* and *holo* forms of the Src SH2 domain[14] and the NMR observation of the EF−BG loop closure.[23]

The flexible nature of proteins gives rise to nontrivial kinetic pathways of binding that need to be elucidated in order to fully understand molecular recognition. We have shown how, in a system with moderate structural complexity and a flexible partner, the molecular determinants and their temporal dependencies could be revealed. We recovered the binding pose within 2 Å RMSD of the crystal structure and obtained binding kinetics compatible with experimental values. In retrospect, structural determinants of the binding kinetics can be considered in the drug design process; together with trajectory-ensemble methodologies[13,14,27] they may enable, for example, the rational optimization of drug kinetic parameters.[28]

■ ASSOCIATED CONTENT

Supporting Information

Data for binding trajectories T1−T5; detailed computational methods; video of the five binding trajectories superimposed; video of the RMSF of domain and ligand during the binding event. The data files containing the binding trajectories are available on request. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gianni.defabritiis@upf.edu.

Author Contributions

§These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank all the volunteers of GPUGRID who donated GPU computing time to the project. T.G. acknowledges support from the Beatriu de Pinós programme of the Generalitat de Catalunya. I.B. acknowledges support from the Fundació La Marató de TV3. G.D.F. acknowledges support from the Ramón y Cajal scheme and support by the Spanish Ministry of Science and Innovation (ref. BIO2011-27450).

■ REFERENCES

(1) Liu, B. A.; Jablonowski, K.; Raina, M.; Arcé, M.; Pawson, T.; Nash, P. D. *Mol. Cell* **2006**, *22*, 851−868.
(2) Payne, G.; Shoelson, S. E.; Gish, G. D.; Pawson, T.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4902−4906.
(3) Lee, T. R.; Lawrence, D. S. *J. Med. Chem.* **2000**, *43*, 1173−1179.
(4) Eck, M. J.; Shoelson, S. E.; Harrison, S. C. *Nature* **1993**, *362*, 87−91.
(5) Lubman, O. Y.; Waksman, G. *J. Mol. Biol.* **2003**, *328*, 655−668.
(6) Songyang, Z.; Shoelson, S. E.; Chaudhuri, M.; Gish, G.; Pawson, T.; Haser, W. G.; King, F.; Roberts, T.; Ratnofsky, S.; Lechleider, R. J.; Neel, B. G.; Birge, R. B.; Fajardo, J. E.; Chou, M. M.; Hanafusa, H.; Schaffhausen, B.; Cantley, L. C. *Cell* **1993**, *72*, 767−778.
(7) Bradshaw, J. M.; Grucza, R. A.; Ladbury, J. E.; Waksman, G. *Biochemistry* **1998**, *37*, 9083−9090.
(8) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. *J. Am. Chem. Soc.* **2010**, *132*, 7361−7371.
(9) Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2005**, *127*, 2600−2607.
(10) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. *J. Chem. Inf. Model.* **2010**, *50*, 397−403.
(11) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.
(12) Ahmad, M.; Gu, W.; Helms, V. *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 7626−7630.
(13) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526−1528.
(14) Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc. Natl. Acad. Sci. U.S.A.* **2011**.

(15) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. *J. Chem. Inf. Model.* **2009**, *5*, 1632−1639.
(16) Buch, I.; Sadiq, S. K.; De Fabritiis, G. *J. Chem. Theory Comput.* **2011**, *7*, 1765−1772.
(17) Panayotou, G.; Gish, G.; End, P.; Truong, O.; Gout, I.; Dhand, R.; Fry, M. J.; Hiles, I.; Pawson, T.; Waterfield, M. D. *Mol. Cell. Biol.* **1993**, *13*, 3567−3576.
(18) Morelock, M. M.; Ingraham, R. H.; Betageri, R.; Jakes, S. *J. Med. Chem.* **1995**, *38*, 1309−1318.
(19) De Fabritiis, G.; Geroult, S.; Coveney, P. V.; Waksman, G. *Proteins* **2008**, *72*, 1290−1297.
(20) Gan, W.; Roux, B. *Proteins* **2009**, *74*, 996−1007.
(21) Bradshaw, J. M.; Waksman, G. *Biochemistry* **1998**, *37*, 15400−15407.
(22) Bradshaw, J. M.; Waksman, G. *Biochemistry* **1999**, *38*, 5147−5154.
(23) Taylor, J. D.; Ababou, A.; Fawaz, R. R.; Hobbs, C. J.; Williams, M. A.; Ladbury, J. E. *Proteins* **2008**, *73*, 929−940.
(24) Waksman, G.; Shoelson, S. E.; Pant, N.; Cowburn, D.; Kuriyan, J. *Cell* **1993**, *72*, 779−790.
(25) Tong, L.; Warren, T. C.; Lukas, S.; Schembri-King, J.; Betageri, R.; Proudfoot, J. R.; Jakes, S. *J. Biol. Chem.* **1998**, *273*, 20238−20242.
(26) Noé, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154−162.
(27) Held, M.; Metzner, P.; Prinz, J.-H.; Noé, F. *Biophys. J.* **2011**, *100*, 701−710.
(28) Swinney, D. C. *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 31−39.
(29) Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.