JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

A Hybrid All-Atom Structure-Based Model for Protein Folding and Large Scale Conformational Transitions

Ludovico Sutto, Ilaria Mereu, and Francesco Luigi Gervasio*

Spanish National Cancer Research Center (CNIO), Structural Biology and Biocomputing Programme, Melchor Fernandez Almagro, 3, E-28029 Madrid, Spain

ABSTRACT: Structure-based models are successful at conjugating the essence of the energy landscape theory of protein folding with an easy and efficient implementation. Recently, their realm expanded beyond a single protein structure, and structure-based potentials have been used profitably to widely study complex conformational transitions. Still, when dealing with structural rearrangements between two, or more, well-defined structures, an unbiased and transferable description of the local backbone and side chain interactions could be advantageous. Here, we propose an all-atom model that merges a classical force field description of these local interactions with a structure-based long-range potential that takes into account the different conformations. We first validate the model simulating and characterizing the folding reaction and the transition state of two well-known proteins: the villin headpiece and the SH3 domain. Then, we characterize the activation mechanism of the catalytic domain of c-Src kinase. Such a process involves the conformational rearrangement of a large loop and the swing of an α helix. The appearance of a stable intermediate state in the free energy landscape between the two conformational end points suggests the mechanism of the loop opening. The low computational cost of the model together with the satisfactory accuracy of the results make it a promising approach to studying conformational transitions in large protein systems.

1. INTRODUCTION

Molecular simulations of biological systems are nowadays routinely exploited to investigate complex phenomena with a high resolution in time and space. However, a large number of relevant processes involving proteins take place on a time scale which is difficult to access with fully solvated all-atom force field simulations on modern computers. Protein folding, large scale structural rearrangement, and allosteric regulation are important examples where free energy barriers hinder classical molecular dynamics sampling. To overcome these difficulties, different routes have been followed, either reducing the number of degrees of freedom by using coarse-grained models[1−4] or by accelerating the dynamics with enhanced sampling methods.[5,6]

Pure structure-based models, or Gō-models,[7] with only native interactions are well suited to the study of protein folding since their Hamiltonians are by construction minimally frustrated.[2] The energy landscape is smoothly funneled toward the native structure, and folding occurs quickly and in a cooperative fashion. A variety of flavors has been proposed where the actual physicochemical interactions are all condensed in an effective native potential combined with different levels of geometric coarse-graining.[8−11] These models represent indeed the simplest implementation of the energy landscape theory of protein folding[12] and are able to reproduce many aspects of the protein folding process.[2] Their main tangible advantage is speed, as they are orders of magnitude faster than all-atom simulations, which translates into the ability to study larger systems and for longer times. But they are also extremely useful in testing our ability to catch the relevant descriptors of the system and to understand their physical meaning. However, they lack the transferability of all-atom and other coarse-grained force fields.

Besides protein folding, structure-based models have recently been used to elucidate peculiar fold switching behaviors,[13] to study large amplitude conformational transitions,[14−18] and to reproduce the B-DNA base flipping process.[19] In these cases, the contributions of two distinct structures must be merged in the Hamiltonian to reproduce the conformational transition. Different approaches have been proposed, ranging from simple summation of the two potential energies[13] to exponential averaging of two energy functions[14,17] or by solving a secular equation.[15,18,19] These models have by construction two energy minima corresponding to the input structures and lead to two free energy basins separated by an adjustable barrier that can be tuned on experimental data. However, it is not clear whether or not these mixing approaches correctly reproduce conformations outside the two energy basins (e.g., the transition state ensemble). For instance, no distinctions are made between the local angular and dihedral terms and the long-range interactions upon merging. A possible solution would be to mix structure-based approaches with some of the transferable terms of the atomistic force fields. A similar approach has been tried with DNA,[19] while in proteins Gō-like potentials have only been used as an extra biasing term on top of the original all-atom force field to predict the effects of mutations[20] or to accelerate folding.[21] Here, to provide a generally applicable and predictive structure-based approach which includes a transferable description of the local backbone and rotamer conformations, we merge the bonded, angular, and dihedral terms of an up-to-date all-atom semiempirical force field with a structure-based, long-range term. In so doing, we keep the advantages of a structure-based model where the long-range interactions can be set to include one or more structures while gaining at the same time a more correct description of the backbone and side chain rotamers and dynamics along the simulated trajectory.

Received: August 4, 2011
Published: October 24, 2011

To assess the validity of this hybrid model, we first use a single basin version to set the Gō term energy parameter and reproduce the folding of two well-studied globular proteins with two different secondary structures: the three helix bundle chicken villin headpiece subdomain (HP35) and the β-sheet rich SH3 domain of human c-Src. Both of these proteins are known to fold fast and cooperatively and have been extensively characterized experimentally.[22−29] Subsequently, we use the two-state model to simulate and characterize the large conformational transition taking place in the catalytic domain of c-Src tyrosine kinase upon activation. The role of this kinase in cancer[30] fostered an intense experimental characterization of its structure and dynamics.[17,31−35] Indeed several crystal structures exist with different poses of the catalytic domain in its active and inactive states.[34,36] The activation involves the crossing of a free energy barrier corresponding to the rearrangement of a 23-residue loop located near the catalytic site between the two lobes of the kinase. Due to its transient nature, such structural transition is difficult to characterize experimentally. At the same time, all-atom MD simulations require either conspicuous simulation time and resources or ad-hoc sampling techniques to accelerate the rate of barrier crossing. Indeed, the activations of c-Src and close Src family members have been studied only recently either using computational methods with a C_α multistate Gō model[17] or merging the information from multiple short molecular dynamics runs.[31,32] Here, we use the proposed two-state model to investigate the activation transition mechanism with an all-atom resolution.

2. COMPUTATIONAL METHODS

2.1. Model. The model takes into account all non-hydrogen atoms of the proteins while the solvent is treated implicitly by the Langevin dynamics. The Hamiltonian can be summarized by the sum of two distinct contributions, $H = H_{ff} + H_{sb}$, where H_{ff} contains only the following terms of the AMBER99SB-ILDN force field:[37] two-body bonded potential, three-body angle potential, four-body proper and improper dihedrals, and 1−4 short-range van der Waals potentials (two-body term between atoms separated by three bonds). These terms account for the correct backbone and side chain rotamers and geometries and are fully described elsewhere.[38] The H_{sb} term is the two-body structure-based potential and has the form:

$$H_{sb} = \sum_{i<j}^{\text{nativecontacts}} \varepsilon_0[(r_{ij}^0/r_{ij})^{12} - 2(r_{ij}^0/r_{ij})^6] + \sum_{i<j}^{\text{nonnativecontact}} \varepsilon_H(r_H/r_{ij})^{12}$$

where the first sum is over all of the native pairs i,j conferring to each native contact a Lennard-Jones potential with common depth ε_0 at the native distance r_{ij}^0. The second term is a generic hardcore repulsion for non-native atom pairs to avoid atom overlap with constants set to $\varepsilon_H = 1.0$ kJ/mol and $r_H = 0.2$ nm. The current conformation is specified by the set of pair distances r_{ij}.

Usually, what sets the energy scale in pure Gō models is the depth of the native potential. Here, we have the force field contribution with its semiempirical parameters already correctly tuned to mimic the correct interactions around 300 K; hence, the strength of the native interaction is not a free parameter anymore. On the contrary, we have to set ε_0 to correctly reproduce the balance between the long-range interactions and the local interactions. Too large values of ε_0 would drive the model toward a pure Gō model, making pointless the force field contributions. On the other hand, ε_0 must be large enough to effectively account for the nonbonded interactions. For this reason, a screening on a range of ε_0 values was performed with the aim of identifying a coupling strength so to provide, on one side, the desired balance between folding speed and correct geometry, along with folding temperatures close to their experimental values. This survey led to the choice of $\varepsilon_0 = 3.8$ kJ/mol (see the Appendix for the technical details).

We chose, for the sake of simplicity, to use the same ε_0 value for all atom−atom native pair interactions regardless of the atom type, the residue, or the protein considered. Nevertheless, since it is an all-atom model, part of the residue characteristics such as its rotational degrees of freedom and its steric hindrance are automatically taken into account, relaxing the necessity for a pair-specific potential.

In the case of the conformational transition of the catalytic domain of c-Src, the structure-based term H_{sb} is the sum over the native contacts of both the open and closed conformations of the activation loop (A-loop). In order for such an approach to lead to the desired two free energy minima, the set of native contacts of each structure must be largely mutually exclusive. Otherwise, a structural arrangement of the backbone and side chains satisfying both sets of contacts would lead to a single global energy minimum structure, instead of the two distinct minima. In our case, the cutoff defining a native contact is small (5 Å) compared to the movement of the activation loop that spans a distance of 25 Å. Hence, the activation switch must proceed by breaking the contacts of the starting conformation to make those of the final structure, avoiding passing through conformations in which both sets of contacts are formed. Finally, the free energy minima are equally populated if the conformational entropy and the energetic contributions of the two structures are similar. This holds in our case since both conformations are compact and thus share a reasonably similar conformational entropy. Moreover, the number of native interactions specific to the open and to the closed structure is almost the same ($C_{open} \approx C_{closed}$, see below). As a result, the potential energies of the two conformations are indeed very close (-2438 ± 176 kJ/mol for the open and -2593 ± 116 kJ/mol for the closed conformation calculated averaging the long-range contributions coming from 2000 conformations picked in their respective free energy basins).

2.2. Native Contact Definition. Following the work of Wu et al.,[39] given a reference structure, a native contact between two atoms i and j belonging to two different, not adjacent, residues is defined when two requirements are met: (i) the atoms are within a cutoff distance $d_{ij} < 5$ Å and (ii) there is no other atom k such that $d_{ik} < d_{ij}$ and $\hat{ijk} < 35°$ or $d_{jk} < d_{ij}$ and $\hat{ijk} < 35°$. This definition approximates the screening effect occurring when the interaction between two atoms is shadowed by the presence of a third atom in between. With such a definition the dependence of the number of native contacts as a function of the cutoff distance increases more smoothly and slowly than it would without screening effects, thus relaxing the dependence of the model on the cutoff value.

We found 259 contacts for the chicken villin headpiece HP35 (PDB code 1YRF) and 382 for the SH3 domain of c-Src (PDB code 1FMK). In the case of the catalytic domain of c-Src (residues Gln250:Glu524, in human c-Src numbering), the closed loop conformation (PDB code 2SRC) has a total of 2146 contacts. A total of 723 of these contacts involve at least one atom in the region of major structural change upon activation.

To be sure not to limit excessively the protein flexibility, besides the activation loop (residues Ala403:Pro425) we include in the definition of this region also part of the N-lobe comprising the αC-helix (residues Ala256:Arg318), as shown in the red and green regions of Figure 6. The remaining C_{KD} = 1423 contacts are essentially the same in both open and closed conformation and will be referred to as "kinase domain contacts" (shown in cyan in Figure 6). C_{closed} = 438 contacts out of the 723 transition related contacts are specific to the closed conformation while the remaining C_{shared} = 285 are present in both open and closed conformation, and in that case the native distance is chosen as the smallest contact distance between the two conformations. Finally, C_{open} = 433 contacts of the open conformation (PDB code 1Y57) involve at least one atom in the transition region and are unique to the open structure. As a result, the total number of native contacts for the c-Src two basins model is $C_{tot} = C_{KD} + C_{closed} + C_{open} + C_{shared} = 2579$.

2.3. Simulation Details and Performance. Langevin dynamics have been performed with the molecular dynamics package GROMACS 4[40] using an inverse friction constant of 1 ps and an integration time step of 2 fs. A 500 ns-long simulation has been performed for both chicken villin headpiece HP35 and the SH3 domain. In the case of HP35, 15 replicas in the 270−350 K temperature range have been used. For the sharper transition of the SH3 domain, 40 replicas in the 350−370 K range have been used. In both cases, the exchange between adjacent replicas was attempted every 1 ps. The parallel tempering[41] technique and the long simulation time allow for a full convergence of the thermodynamic properties.

To simulate the activation transition of the catalytic domain of c-Src, we adopted the bias exchange metadynamics (BE)[42] approach in which replicas subject to different metadynamics biasing potentials are exchanged between Langevin dynamics simulations performed at the same temperature. In our case, we used T = 310 K and three replicas: one subject to the metadynamics bias with the contact map of the open form of the A-loop as a collective variable, one with the contact map of the closed form of the A-loop, and a third white replica with no metadynamics bias. In so doing, the sampling of the white replica conformations converges very quickly to the correct equilibrium distribution, and the free energy surface can be easily calculated.

To perform the metadynamics runs we used the PLUMED[43] plug-in for GROMACS.

As a reference, in the case of the 452 atoms of SH3, the model performs 102 ns/day per core per replica on a modern computer cluster. The same system simulated with an unbiased standard force field with implicit solvent performs 37 ns/day, while when it is solvated in 9120 water molecules in a 284 nm^3 cubic box it slows down to 2.2 ns/day. Obviously, since the Hamiltonians and the number of simulated particles are different, the direct comparison of these performances has to be taken with care. A more meaningful comparison could be made by measuring the overall simulation time needed to converge the observables of interest. In so doing, we can then appreciate that in a few days of calculations, with a standard computer cluster, our model is able to reproduce thousands of folding/unfolding events of a globular protein allowing for a validation of its thermodynamics and the structures of the transition state ensemble at atomic resolution. On the contrary, a few folding/unfolding events, with an unbiased all-atom force field approach, still requires nowadays months of calculations on specialized hardware.[44]

Similarly, the conformational transition between two known conformers separated by a high free energy barrier also requires months of computer time with a standard force field even with biased dynamics,[45] while we can sample the transition of c-Src in a few days of computer time. For these cases, the gain in time of our model is thus roughly 2 orders of magnitude over standard unbiased all-atom force fields.

2.4. Analysis. In order to characterize the folding reaction of HP35 and SH3, several parameters are used. The heat capacity at constant volume $C_V(T)$ as a function of temperature T is calculated as $C_V(T) = \sigma_E^2/(k_B T^2)$ where k_B is Boltzmann's constant and σ_E are the fluctuations of the energy. The latter are calculated from the most likely density of state of the system obtained collecting the energies of the 15 runs at different temperatures and optimally combining them with the multiple histogram method.[46,47] To measure the folding reaction cooperativity, use is made of Chan's parameter[48] κ_2, defined as the ratio of the van't Hoff enthalpy to calorimetric enthalpy: $\kappa_2 = 2T_{max}(k_B C_p(T_{max}))^{1/2}/\Delta H_{cal}$, where T_{max} is the temperature of the C_p peak and ΔH_{cal} is the calorimetric enthalpy of the reaction determined as the integral of the heat capacity across the transition region. We used the energy and the C_V instead of the enthalpy and the C_p to calculate κ_2 since for our purposes they can be considered equivalent.

Another useful parameter to characterize folding is the native contact fraction Q, calculated as the ratio of the number of formed native contacts over the total number of native contacts. A native contact is considered formed if the atoms are within 1 Å of their native distance.

To characterize the folding transition state (TS) of HP35 and SH3, the Φ values per residue can be approximated by $\Phi_i = \langle Q_i \rangle_{TS}$ where Q_i is the fraction of formed native contacts of residue i and where the average $\langle ... \rangle_{TS}$ is done over the set of conformations constituting the transition state.[49,50]

Two structural parameters are also used to measure the similarity between a current structure Γ and a reference structure Γ^0: the distance root-mean-square deviation $dRMSD = (1/(N(N-1)\Sigma_{i<j}(d_{ij}-d_{ij}^0)^2))^{1/2}$ and the root-mean-square deviation $RMSD = (1/N\Sigma_{i=1}^{N}(r_i-r_i^0)^2)^{1/2}$. Both quantities are calculated using N Cα atoms and involve either their pairwise distance d_{ij}^0 or their position r_i^0 in the reference structure. In the latter case, r_i indicates the position of the Cα atom i of the structure Γ aligned on the reference Γ^0.

The free energy as a function of a generic reaction coordinate q is calculated from the equilibrium probability distribution $p(q)$ obtained during the simulation as $F(q) = -k_B T \ln(p(q))$.

3. RESULTS AND DISCUSSION

3.1. Folding of Villin Headpiece HP35 and SH3 Domain. In order to test the ability of our model to correctly fold both α-helical and β-sheet rich proteins, we chose two different domains: HP35 and SH3. HP35 consists of three α-helices packed together and, with its 35 residues, is one of the smallest proteins showing a cooperative folding.[51] Several computational[25,52−55] and experimental[25−29,51,56] works thoroughly characterized its folding reaction describing the transition state as well as its secondary structure rich unfolded state. The chosen value of the native potential energy ε_0 results in a cooperative folding transition for HP35 with a folding temperature T_f = 316 K. Indeed, at $T = T_f$ the heat capacity C_v features a peak, and the average fraction of native contacts $\langle Q \rangle$ shows its steepest variation (Figure 1).

The fraction of native contacts Q is not sufficient alone to distinguish the role of each helix during the folding reaction. For this reason, the average fraction of native contact $\langle Q \rangle$ is decomposed according to four subsets of native contacts (Figure 1, lower panel), and the free energy is projected along both Q and the C_α distance root-mean-square deviation of the first and second helices ($dRMSD_{I-II}$) and the second and third helices ($dRMSD_{II-III}$) in Figure 2. The folding barrier $\Delta G_{U\ddagger} = 2$ kJ/mol at $T = 1.03T_f$ is relatively small because of the high content of residual secondary structure in the unfolded state. The conformations populating this unfolded basin are characterized by the loss of the majority of the tertiary native structure ($\langle Q_{inter} \rangle \approx 0.2$, Figure 1) and most of the helical content of helices I and II ($dRmsd_{I-II} \approx 0.4$ nm, Figure 2), while helix III maintains roughly half of its internal native interactions ($dRmsd_{II-III} \approx 0.2$ nm, Figure 2 and $\langle Q_{h3} \rangle \approx 0.5$, Figure 1). This residual structure, which involves the N-terminal residues of helix III, vanishes at $T = 1.05T_f$ leading to an extended unfolded basin (Figure 2, right-hand panels).

Despite the absence of any appreciable intermediate state with native secondary structure and poor tertiary interactions as found by solid-state NMR experiments,[27] the description of the unfolded state rich in helical content, in particular in the helix III region, is in good agreement with all-atom molecular dynamics calculations in explicit solvent[52,55] as well as with recent triplet–triplet energy transfer experiments.[25,26] A number of independent

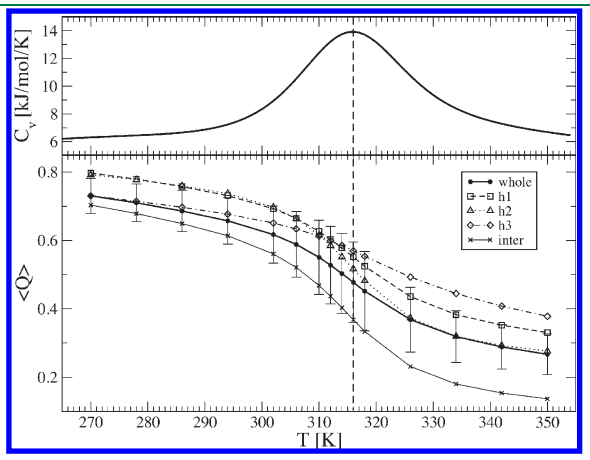

Figure 1. Thermodynamics of HP35 folding. In the upper panel, the heat capacity C_V is reported as a function of temperature T. Its peak at $T_f = 316$ K defines the folding temperature. In the lower panel, the average fraction of native contact as a function of the temperature of the whole protein (solid thick line, filled dots) is shown together with its standard deviation. In the same plot, the average fraction of native contacts divided into different subsets is also displayed: the internal contacts of helix 1 (dashed line, square), helix 2 (dotted line, triangles), and helix 3 (dash-dotted line, diamonds) and the tertiary native contacts between them (solid line, crosses). These curves highlight that the folding reaction mainly depends on the assembly of the helices while they keep a certain degree of nativeness even at high temperatures. As expected by its small size, helix 2 is the least stable and the one that loses most structure at high temperatures. On the contrary, helix 3 is the most stable and retains almost 40% of its structure at the highest sampled temperature.

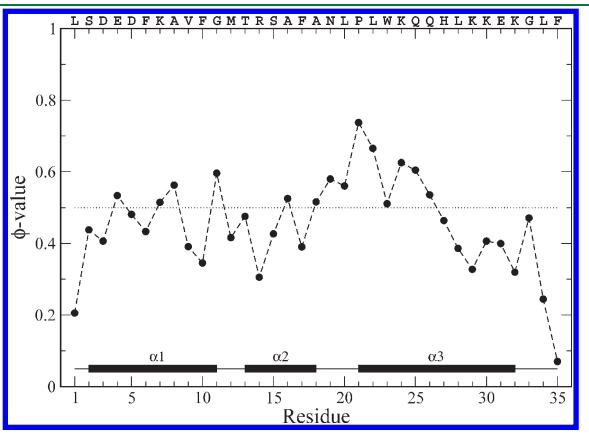

Figure 3. Characterization of HP35 folding transition state. The Φ value is calculated for each residue using a set of 5178 structures picked from the top of the folding free energy barrier at $T = T_f$. The highest values correspond to residues Pro21 and Leu22 belonging to the first turn of helix III. The helices are represented along the sequence as solid bars, and a dotted line is shown at $\Phi = 0.5$.

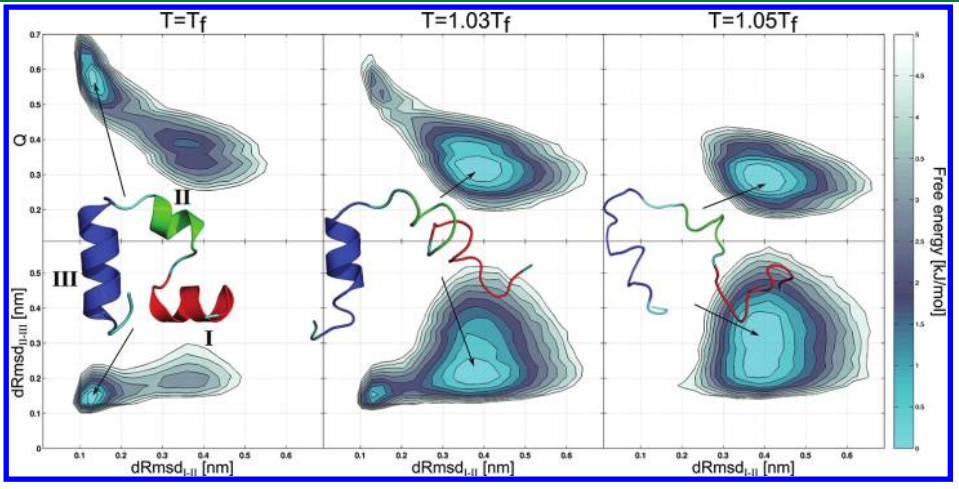

Figure 2. Free energy surfaces of HP35. The free energy surface at three increasing temperatures is projected along either the fraction of native contacts Q and the distance root-mean-square deviation with respect to the first and second helices ($dRMSD_{I-II}$) in the upper row or the distance root-mean-square deviation with respect to the second and third helices ($dRMSD_{II-III}$) and $dRMSD_{I-II}$ in the lower row. The contour lines are drawn every 0.5 kJ/mol. Three representative conformations of each basin are also shown.

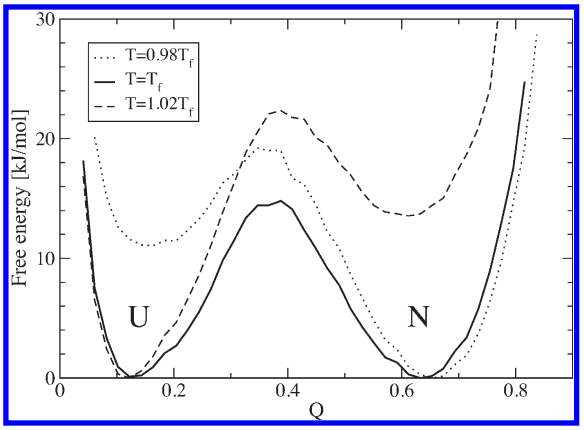

Figure 4. Free energy of the SH3 domain as a function of the fraction of formed native contacts Q at three different temperatures. At $T = T_f$, the native (N) and unfolded (U) states are equally populated (solid curve) and separated by a free energy barrier of about 14 kJ/mol located at $Q = 0.4$. At lower ($T = 0.98T_f$, dotted curve) or higher temperatures ($T = 1.02T_f$, dashed curve), the native or unfolded state is respectively predominant.

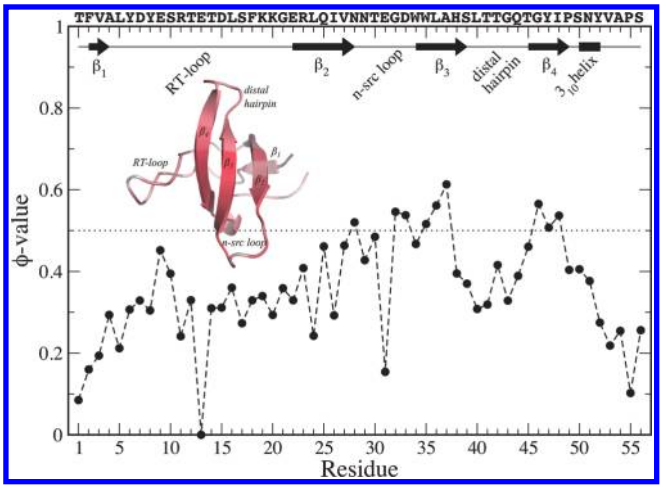

Figure 5. Characterization of SH3 folding transition state. Φ values are calculated for each residue using a set of 1350 structures picked from the top of the folding free energy barrier at $T = T_f$. Along the upper x axis is shown the sequence and a schematic representation of the secondary structure. A native conformation with color coded Φ values (darker color being 1, lighter color being 0) is also shown. A dotted line is plotted at $\Phi = 0.5$.

computational works also found an early formation of helices II and III, and a slower formation of helix I.[53−55] Moreover, recent molecular dynamics simulations of the HP35 folding found the intermediate state caused by an incorrect docking of helices I and III where non-native interactions play an important stabilization role.[25,52] This could explain the absence of the intermediate state in our calculations given the lack of any non-native tertiary interactions in our Gō-like potential.

The putative transition state, corresponding to the top of the free energy barrier, features the formation of most of the native contacts of Pro21 and neighboring residues as visible from the highest Φ value (Figure 3). This corresponds to the formation of the loop between helices II and III and the first turn of helix III. This residue is indeed known to play a crucial role for the protein stability since if mutated to alanine the villin headpiece is unable to fold to a native structure anymore.[57] The N-terminal half of helix III as well as residues 4, 7, and 8 belonging to helix I have more than 50% of the native interactions formed in the transition state ($\Phi > 0.5$). The last forming interactions, in the free energy descent toward the native state, involve the tail of helix III with both internal and tertiary contacts with the head of helix I that eventually lock the protein in its characteristic bundle.

To check the ability of our model to correctly fold also β-sheet structures, we simulated the 60-residue globular domain SH3 and repeated the thermodynamic analysis and folding characterization. As expected, the folding transition of the β domain SH3 is much more cooperative than for the α-helical HP35 as measured by Chan's parameter[48] κ_2 resulting in $\kappa_2^{SH3} = 0.80$ as compared to $\kappa_2^{HP35} = 0.53$ (see the Computational Methods). Intuitively, an α helix can form independently of the rest of the protein since it involves only local interactions, while a β sheet requires a wider structural organization. The three α helices of HP35 form gradually, losing entropy in favor of the energetic contributions of the local contacts. On the contrary, the formation of the β sheets of SH3 involves longer-range interactions with a larger entropic cost compensated by the formation of several native interactions, resulting in an emphasized all-or-none character.

To further support this picture, we calculated the difference in internal energy between the native and the unfolded state for HP35 and SH3 at their respective folding temperatures. At $T = T_f$, we have by definition $\Delta G_{UN} = 0$ thus $T_f \Delta S_{UN} = \Delta E_{UN}$ where ΔG_{UN}, ΔS_{UN}, and ΔE_{UN} are respectively the difference in free energy, entropy, and internal energy between the unfolded (U) and native (N) states. Picking 30 000 structures in both basins for both proteins, the average difference in internal energy is $\Delta E_{UN}^{HP35} = 134 \pm 94$ kJ/mol and $\Delta E_{UN}^{SH3} = 647 \pm 129$ kJ/mol, accounting for the 6% and 25% of their respective average native state energy. As a result, the chicken villin shows a much lower difference in internal energy between structures of the unfolded and the native state due to the large amount of residual helical structure in the U state. On the contrary, the unfolded basin of SH3 is a minimum in free energy mainly because of its large entropy, which compensates the poor native content responsible for the large difference in internal energy between the U and N state conformations.

The thermodynamic stability of SH3 is also higher than for HP35 with a folding temperature of $T_f = 357$ K. The free energy projected along the fraction of native contacts Q shows two basins of equal depth at $T = T_f$ corresponding to the native (N) and unfolded (U) states separated by a high barrier of 14 kJ/mol (Figure 4).

The ensemble of structures on the top of this free energy barrier, containing about 40% of formed native contacts (Q in $[0.31-0.44]$), corresponds to the transition state of our model. These native contacts are not distributed homogeneously along the sequence but are rather concentrated in specific regions. Indeed, from the calculated Φ values, the most natively structured residues involve the n-src loop and the β_3 and β_4 strands (see Figure 5). Consequently, the less formed regions involve the RT-loop and the N- and C-terminal regions. These observations are in good agreement with the experimental characterization of the transition state done by the thorough Φ-value analysis of Riddle et al.[23] where they find the central three-stranded β sheet with the most structured residues. The linear correlation coefficient between the calculated Φ values Φ_{calcd} and the 37 available experimental ones Φ_{exptl} is $r = 0.48$ with a root-mean-square

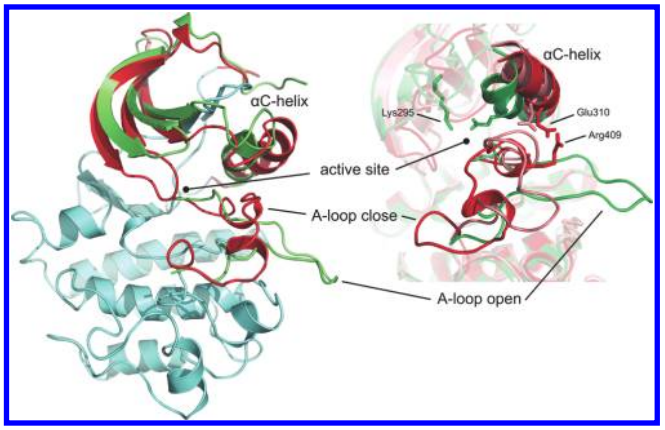

Figure 6. (Left) The catalytic domain of c-Src in its open (green) and closed (red) conformations. The cyan part, common to both conformations, defines the kinase domain native contacts while the open-specific contacts are taken from the green region, and the closed-specific native contacts from the red one. (Right) Close-up of the activation loop and the αC-helix from another viewpoint. Besides the open and closed conformations, also a representative conformation of the intermediate C* basin is shown in pink. The catalytically critical Glu310 is also shown with sticks, together with the Lys295 and Tyr382 residues that stabilize the αC-in and αC-out conformations, respectively.

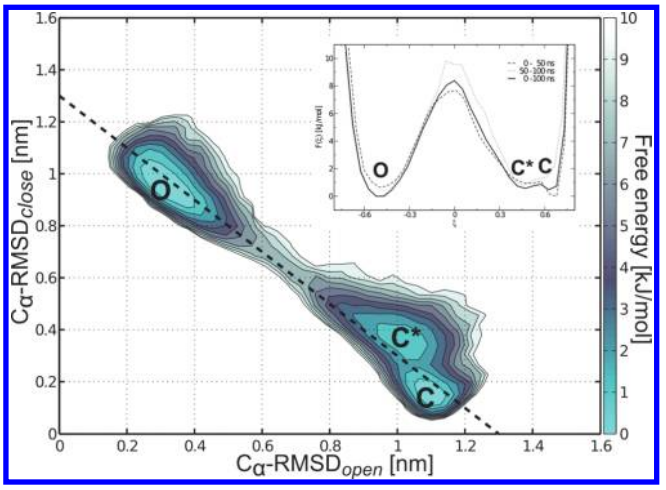

Figure 7. Transition free energy of c-Src at T = 310 K. The two main minima corresponding to the open (O) A-loop and to the closed (C) A-loop conformations are separated by a free energy barrier of 8 kJ/mol. A secondary shallow minimum in the closed basin corresponds to an intermediate state of conformations (C*) where the activation loop adopts a particular hybrid structure. The contour lines are drawn every 1 kJ/mol. In the inset is shown the free energy projected along the dashed line of the bidimensional plot, corresponding to the linear combination of $RMSD_{open}$ and $RMSD_{closed}$ that best distinguishes the minima (solid line). In this monodimensional projection are also shown the free energies calculated for the first half of the simulation (dashed line) and for the second half (dotted line).

deviation of $\sigma = ((\Sigma (\Phi_{exptl} - \Phi_{calcd})^2/N))^{1/2} = 0.32$. This result shows a semiquantitative agreement with the actual transition state similarly to other all-atom structure-based models[58,59] while performing better than a C_α heterogeneous G$\bar{o}$ model.[9] In particular, we overestimate the native content of Gly46 of strand β_4 while underestimating that of Gly43 belonging to the distal hairpin. In fact, when these two outliers are removed, the correlation coefficient of the remaining 35 Φ values increases to r = 0.62 with a smaller RMSD σ = 0.29.

These results validate the model as a valuable tool to reproduce the folding of both α and β protein domains. Indeed, the transition state ensembles are accessible with atomic detail and correspond with an acceptable confidence to the experimental ones, as shown by Φ-value analysis. This suggests that the model could be equally fit to study the large-scale conformational transitions in larger proteins. The main difference between folding and conformational transitions stems from the fact that, in the latter case, the conformations of the two extremes of the reaction are precise structures, while folding proceeds from a large ensemble of structures belonging to the unfolded state. Exploiting this difference through a two-basin energy landscape, we are able to characterize the activation of the catalytic domain of c-Src kinase.

3.2. The Activation Loop Dynamics of c-Src Catalytic Domain. c-Src is a nonreceptor tyrosine kinase critically involved in the regulation of fundamental cellular processes including cell growth, differentiation, and migration.[60] These functions are carried out through phosphorylation of the tyrosine residue of a protein substrate once the kinase is in its active conformation.[34,60−62] An elevated and uncontrolled activation of c-Src is associated with many tumor types,[30,62,63] highlighting the importance of the structural and dynamical characterization of its activation mechanism.

The catalytic domain is constituted by 274 residues forming a characteristic two-lobe structure with the catalytic active site located in a deep cleft between them (see Figure 6). The activation mechanism involves the rearrangement of a centrally located loop, called the activation loop or A-loop (residues Ala403 to Pro425), that switches from a folded and closed conformation obstructing the active site to a more extended one which serves as a platform for substrate binding.[61] At the same time, the αC-helix (residues Pro304 to Lys315) in the N-terminal lobe also swings from an αC-out to an αC-in conformation. In so doing, the catalytically critical Glu310 points inward to the active site pocket where, together with the DFG motif (Asp404, Phe405, Gly406) at the beginning of the A-loop, it coordinates the transfer of the γ-phosphoryl group of ATP to the tyrosine residue of a protein substrate.[64] The open A-loop conformation is stabilized by the phosphorylation of a tyrosine residue (Tyr416), located in the middle of the loop, albeit the open conformation has been crystallized also with the unphosphorylated tyrosine.[36] This tyrosine is in fact accessible for phosphorylation only when exposed to the solvent and thus when the loop is not in its fully closed conformation. This suggests that the open and closed conformations are both accessible to the unphosphorylated catalytic domain, in agreement with the significant catalytic activity retained by the unphosphorylated form.[65,66] As a consequence, we did not consider any phosphorylation effect in our model, and we neglected the two regulatory domains SH2 and SH3, focusing our attention on the sole ability of the catalytic domain to undergo the conformational rearrangement.

To characterize the conformational switch, a bidimensional free energy has been calculated from the simulation. The parameter that better distinguishes the structural change is the root-mean-square deviation (RMSD) of carbon α atoms of the A-loop with respect to either the open or closed conformation after proper alignment of the whole catalytic domain to the respective reference structure. In such space, the projection of the free energy shows two major basins, C and O, separated by a barrier of 8−9 kJ/mol: one corresponding to the folded, closed (C), A-loop, while the other corresponds to the extended, open (O),

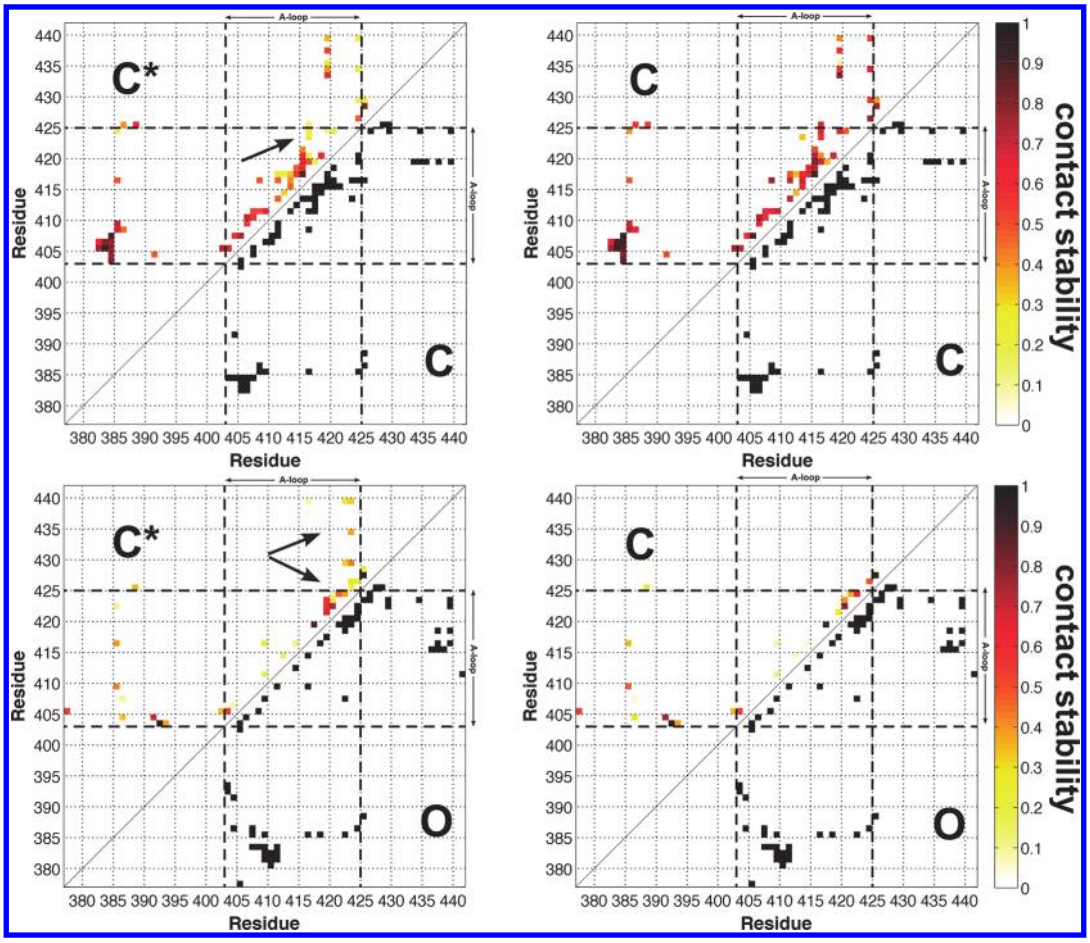

Figure 8. Stability contact maps of the A-loop of the intermediate state (C*) compared to the closed state (C). The represented quantity is the probability of the native contact formation per residue calculated out of a set of 2000 conformations picked from either the intermediate state (C*, left-hand side) or the closed state (C, right-hand side). For an easier reading, the native contacts are divided into closed-specific native contacts (top row) and open-specific native contacts (bottom row). The intermediate state contains hybrid conformations in which the C-terminal region of the activation loop partly disrupts its closed-like contacts in favor of open-like ones, as highlighted by the arrows.

conformation (see Figure 7). In the 100 ns simulation run, the catalytic domain undergoes multiple barrier crossings helped by the metadynamics bias on two of the three replicas, guaranteeing the convergence of the free energy profile (see the Computational Methods). To assess such convergence, the free energies of the first and second halves of the trajectory have been projected on a linear combination of $RMSD_{open}$ and $RMSD_{closed}$ for a more convenient comparison (see inset of Figure 7). The variable ξ, defined as $\xi = RMSD_{open}/(2)^{1/2} - RMSD_{closed}/(2)^{1/2}$, shows a difference of 2.5 kJ/mol $\approx 1k_BT$ in the top of the barrier as the largest deviation between the first and second half of the simulation, providing an estimate of the error in the free energy determination.

Interestingly, adjacent to the closed free energy basin C, a shallower second basin emerges in the bidimensional FES. It is separated by a small barrier to state C but nonetheless populated by well characterized structures. Such a free energy minimum, hereby denoted as C*, corresponds to an intermediate state between the closed and open conformations that shares specific native contacts of the open loop and the closed loop. More precisely, C* is populated by an ensemble of conformations where the centrally located residues of the A-loop (residues 411−419 which comprise most of a short 1.5 turn helix) form less stable interactions with their neighbors than in the closed basin C (see a representative conformation in Figure 6 and the contact maps in the top row of Figure 8). This is even more pronounced in the region near the C-terminal end of the A-loop where the probability of closed-specific native contact formation between residues 419−425 and the following residues of the C-lobe, residues 426−439, falls below 0.2. At the same time, a cluster of open-like native interactions appears at the C-terminal end of the A-loop (see Figure 8, bottom row), while they are totally absent in the C state.

These observations suggest that the switching mechanism of the activation loop passes through an intermediate state (C*) characterized by loose short-range interactions of the C-terminal half of the loop where a hybrid conformation of the activation loop is stably populated. This intermediate state has more open-like content than the C state, mostly located in the C-terminal end of the loop, indicating that this region is most prone to starting the transition to the full open structure (see Figure 6).

Besides the rearrangement of the A-loop, the activation of the catalytic domain also involves the rotation of the αC-helix where the catalytically critical Glu310 residue breaks its contacts with Tyr382, Arg409, and Leu410 and points toward the active site. In the active site pocket, favorable interactions between Glu310 and Lys295, Ile336, and the DFG-motif stabilize the αC-in conformation. Analyzing the fraction of native contacts made both by

Table 1. Inspection of the A-loop Closed (C), Intermediate (C*), and Open (O) State in Terms of the Conformation of the αC-helix[a]

	closed (C)	intermediate (C*)	open (O)
$\langle Q_{\alpha C\text{-in}} \rangle$	0.51 (0.37)	0.52 (0.39)	0.57 (0.50)
$\langle Q_{\alpha C\text{-out}} \rangle$	0.53 (0.43)	0.53 (0.42)	0.43 (0.29)

[a] The fraction of formed native contacts specific to the αC-helix in (out) conformation $Q_{\alpha C\text{-in}}$ ($Q_{\alpha C\text{-out}}$) is averaged over a set of 2000, 2000, and 6000 structures from the C, C*, and O ensembles, respectively. In parentheses, only a subset of the native contacts of the αC-helix is considered, namely, those involving the Glu310 residue. The statistical standard deviation for all of the calculated quantities is around 0.07.

the αC-helix and also specifically by Glu310, we do not find a sharp correlation between the rotation of the helix and the conformation of the A-loop. In other words, in the ensemble of both open and closed conformations, we can find structures with either an inward rotation of the αC-helix or an outward rotation albeit with slightly different propensities (see Table 1). More in general, the helix rearrangement occurs with small cooperativity through a set of conformations stabilized by a blend of favorable interactions coming from both the αC-in and αC-out reference structures. In fact, in the closed and intermediate basins (C and C*), the αC-helix forms roughly half of the native contacts specific to the αC-in and half of those specific to the αC-out conformation, while the structures found in the open basin (O) are mainly characterized by the expected αC-in conformation (see Table 1).

Taken together these results are suggestive of the existence of an intermediate step in the activation transition of c-Src and the ability of the αC-helix to swing between two distinct conformations. These features have been previously predicted by short unbiased all-atom runs along an optimal free energy activation pathway.[31] In that case, the intermediate state was structurally different from the present C* ensemble of conformations showing a clearer open-like structure of the A-loop and an outward rotation of the αC-helix. Interestingly, targeted molecular dynamics (TMD) simulations of the Src family member Lyn suggested a first barrier to the activation to be the helix−coil transition of the 1.5 turn helix at the C-terminal end of the A-loop.[67] This short helix involves residues that we find to be indeed much less structured in the proposed C* intermediate state. Further evidence of the stepwise mechanism of activation comes from the Src family member Hck where two distinct intermediate states have been observed in the activation transition together with an alternative activation pathway involving the partial unfolding of the N-lobe.[32] An intermediate state has also been found in CDK5, Abl, and EGFR tyrosine kinases using either metadynamics molecular dynamics[45] or TMD.[68]

While most features predicted by our hybrid model are in accordance with several computational and experimental observations, the predicted peculiar behavior of the αC-helix was not reported previously either in c-Src or in the close relatives Hck and Lyn. On the contrary, coarse-grained computations and a combination of short all-atom simulations predicted a rigid αC-helix leading to a two-step activation mechanism requiring the loop opening to facilitate the helix rotation to the inward conformation.[17,32,68,69] As a consequence, a closed loop with an inward pointing Glu310 residue has never been reported for c-Src, while it is consistently present in both our C and C* free energy basins. Our model shows that the combination of an αC-helix and Glu310 pointing inward with a folded A-loop is accessible and realistic from a steric point of view. However, without further experimental validation, we cannot rule out that this peculiar combination is due to a slight unbalance of the interactions between the two conformations in the model used.

4. CONCLUSIONS

We describe a novel hybrid all-atom structure-based model able to reproduce the folding transition of both helical and β-sheet globular proteins in good agreement with the available experimental and computational data. The inclusion of multiple reference structures makes it useful in characterizing large-scale conformational transitions as the activation transition of the A-loop of the catalytic domain of c-Src. The major approximation of structure-based models, namely, the absence of any non-native favorable interactions, has been partially addressed in our model by including transferable local interactions that are independent of the reference structures. Albeit the model is limited by the need of an a priori knowledge of one or more reference structures, it can still be applied to study a plethora of biological phenomena of interest for which the crystal structures are available. The approach is much faster than all-atom MD with an explicit solvent; still it has a predictive value that goes well beyond the reference structures that are used to define it. For instance, an accurate reproduction of the transition states of the villin and SH3 folding are indicative of the correct description of a complex and realistic conformational ensemble of metastable structures. Moreover, the emergence of a third minimum in the free energy landscape of the conformational transition of c-Src, not included a priori in the model but in accordance with the proposed stepwise activation mechanism, further shows the versatility of the model. Altogether, these results suggest that the proposed hybrid model can be a viable approach to studying large-scale conformational transitions in proteins.

APPENDIX

To determine the value of the ε_0 interaction parameter, we first estimated its scale knowing that the van der Waals interaction energy between two carbon atoms in the AMBER force field is about 0.6 kJ/mol (this value turned out to be very small for our model). We then proceeded to the identification of a reasonable range of values through the visual inspection of short molecular dynamics trajectories at different values of ε_0. Mainly, two behaviors have been observed: either a fast unfolding took place resulting from an ε_0 value too small or, on the contrary, the atoms' fluctuations were very small around the starting native conformation, which suggested a too large value of ε_0. Once a reasonable range has been identified (3.0−4.0), we proceeded to a more quantitative analysis.

For each of the values of ε_0 (3.0, 3.2, 3.5, 3.8, 4.0), we simulated the SH3 protein using the parallel tempering scheme to sample a large range of temperatures. We compared both the temperature corresponding to the peak in the specific heat to the experimental folding temperature and the correlation of the atom fluctuations with a classical, explicit solvent, AMBER99SB-ILDN force field[37] simulation of the same system at T = 350 K.

As a result, even though a value of ε_0 = 3.5 would have given a lower folding temperature for SH3, closer to the experimental one $T_f^{exp} \approx 350$ K,[22] we eventually chose ε_0 = 3.8 because it better reproduced the atom fluctuations as well as resulted in a faster folding protein (see Table 2). We therefore think that in the small

Table 2. Comparison of Different Value of the Interaction Parameter ε_0.[a]

ε_0	T_f	$r_{C\alpha-\mathrm{RMSF}}$
3.0	335 K	0.78
3.2	336 K	0.65
3.5	348 K	0.82
3.8	357 K	0.84
4.0	n/a	0.84

[a] For each value of ε_0 are indicated the calculated SH3 folding temperature T_f as the peak temperature of the C_v and the Pearson's correlation coefficient $r_{C\alpha-\mathrm{RMSF}}$ of the C_α root mean square fluctuation (RMSF) between a short standard molecular dynamics simulation in explicit solvent and our model. For the case of $\varepsilon_0 = 4.0$, our simulation time did not allow for a sufficient sampling of the folding/unfolding transition.

range (3.5−3.8) the model would give quantitatively similar, and qualitatively identical, results.

AUTHOR INFORMATION

Corresponding Author
*E-mail: flgervasio@cnio.es.

ACKNOWLEDGMENT

We acknowledge support by the Spanish Ministry of Science and Innovation (MICINN) grant (BIO2010-20166, "AlteredDynamics") and the Barcelona Supercomputing Center for computer time. LS acknowledges support by a Juan de la Cierva Research Fellowship from MICINN. IM acknowledges support by a La Caixa Fellowship. The authors acknowledge G. Tiana for the implementation of the multiple histogram method.

REFERENCES

(1) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.
(2) Clementi, C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 10–15.
(3) Hills, R. D.; Lu, L.; Voth, G. A. *PLoS Comput. Biol.* **2010**, *6*, e1000827.
(4) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
(5) Laio, A.; Gervasio, F. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(6) Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer: New York, 2007.
(7) Ueda, Y.; Taketomi, H.; Gō, N. *Biopolymers* **1978**, *17*, 1531–1548.
(8) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins* **2009**, *75*, 430–441.
(9) Sutto, L.; Tiana, G.; Broglia, R. A. *Protein Sci.* **2006**, *15*, 1638–1652.
(10) Derreumaux, P.; Mousseau, N. *J. Chem. Phys.* **2007**, *126*, 025101.
(11) Hills, R. D.; Brooks, C. L. *IJMS* **2009**, *10*, 889–905.
(12) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.
(13) Camilloni, C.; Sutto, L. *J. Chem. Phys.* **2009**, *131*, 245105.
(14) Best, R. B.; Chen, Y.-G.; Hummer, G. *Structure* **2005**, *13*, 1755–1763.
(15) Okazaki, K.-i.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl Acad. Sci. U.S.A.* **2006**, *103*, 11844–11849.
(16) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. *Proc. Natl Acad. Sci. U.S.A.* **2007**, *104*, 18043–18048.
(17) Yang, S.; Roux, B. *PLoS Comput. Biol.* **2008**, *4*, e1000047.
(18) Li, W.; Wolynes, P. G.; Takada, S. *Proc. Natl Acad. Sci. U.S.A.* **2011**, *108*, 3504–3509.
(19) de Marco, G.; Várnai, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10694.
(20) Pogorelov, T. V.; Luthey-Schulten, Z. *Biophys. J.* **2004**, *87*, 207–214.
(21) Meinke, J.; Hansmann, U. *J. Phys.: Conden. Matter* **2007**, *19*, 285215.
(22) Grantcharova, V. P.; Baker, D. *Biochemistry* **1997**, *36*, 15685–15692.
(23) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E. J.; Ruczinski, I.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 1016–1024.
(24) Grantcharova, V. P.; Riddle, D. S.; Baker, D. *Proc. Natl Acad. Sci. U.S.A.* **2000**, *97*, 7084–7089.
(25) Beauchamp, K. A.; Ensign, D. L.; Das, R.; Pande, V. S. *Proc. Natl Acad. Sci. U.S.A.* **2011**, *108*, 12734–12739.
(26) Reiner, A.; Henklein, P.; Kiefhaber, T. *Proc. Natl Acad. Sci. U.S.A.* **2010**, *107*, 4955–4960.
(27) Hu, K.-N.; Yau, W.-M.; Tycko, R. *J. Am. Chem. Soc.* **2010**, *132*, 24–25.
(28) Havlin, R.; Tycko, R. *Proc. Natl Acad. Sci. U.S.A.* **2005**, *102*, 3284.
(29) Kubelka, J.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2003**, *329*, 625–630.
(30) Ishizawar, R.; Parsons, S. J. *Cancer Cell* **2004**, *6*, 209–214.
(31) Gan, W.; Yang, S.; Roux, B. *Biophys. J.* **2009**, *97*, L8–L10.
(32) Yang, S.; Banavali, N. K.; Roux, B. *Proc. Natl Acad. Sci. U.S.A.* **2009**, *106*, 3776–3781.
(33) Young, M.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. *Cell* **2001**, *105*, 115–126.
(34) Xu, W.; Doshi, A.; Lei, M.; Eck, M.; Harrison, S. *Mol. Cell* **1999**, *3*, 629–638.
(35) Gonfloni, S.; Williams, J. C.; Hattula, K.; Weijland, A.; Wierenga, R. K.; Superti-Furga, G. *EMBO J.* **1997**, *16*, 7261–7271.
(36) Cowan-Jacob, S. W.; Fendrich, G.; Manley, P. W.; Jahnke, W.; Fabbro, D.; Liebetanz, J.; Meyer, T. *Structure* **2005**, *13*, 861–871.
(37) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.
(38) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(39) Wu, L.; Zhang, J.; Qin, M.; Liu, F.; Wang, W. *J. Chem. Phys.* **2008**, *128*, 235103.
(40) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
(41) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
(42) Piana, S.; Laio, A. *J Phys Chem B* **2007**, *111*, 4553–4559.
(43) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
(44) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.
(45) Berteotti, A.; Cavalli, A.; Branduardi, D.; Gervasio, F. L.; Recanatini, M.; Parrinello, M. *J. Am. Chem. Soc.* **2009**, *131*, 244–250.
(46) Ferrenberg, A.; Swendsen, R. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
(47) Ferkinghoff-Borg, J. *Eur. Phys. J. B* **2002**, *29*, 481–484.
(48) Kaya, H.; Chan, H. S. *Proteins* **2000**, *40*, 637–661.
(49) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928–1931.
(50) Li, A.; Daggett, V. *Proc. Natl Acad. Sci. U.S.A.* **1994**, *91*, 10430–10434.
(51) McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S. *J. Mol. Biol.* **1996**, *260*, 126–134.
(52) Freddolino, P. L.; Schulten, K. *Biophys. J.* **2009**, *97*, 2338–2347.
(53) Yang, J. S.; Wallin, S.; Shakhnovich, E. I. *Proc. Natl Acad. Sci. U.S.A.* **2008**, *105*, 895–900.
(54) Lei, H.; Duan, Y. *J. Mol. Biol.* **2007**, *370*, 196–206.
(55) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl Acad. Sci. U.S.A.* **2007**, *104*, 4925.
(56) Bunagan, M. R.; Gao, J.; Kelly, J. W.; Gai, F. *J. Am. Chem. Soc.* **2009**, *131*, 7470–7476.

(57) Vermeulen, W.; van Troys, M.; Bourry, D.; Dewitte, D.; Rossenu, S.; Goethals, M.; Borremans, F. A. M.; Vandekerckhove, J.; Martins, J. C.; Ampe, C. *J. Mol. Biol.* **2006**, *359*, 1277–1292.
(58) Lam, A. R.; Borreguero, J. M.; Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. *J. Mol. Biol.* **2007**, *373*, 1348–1360.
(59) Mitomo, D.; Nakamura, H. K.; Ikeda, K.; Yamagishi, A.; Higo, J. *Proteins* **2006**, *64*, 883–894.
(60) Parsons, S. J.; Parsons, J. T. *Oncogene* **2004**, *23*, 7906–7909.
(61) Xu, W.; Harrison, S.; Eck, M. *Nature* **1997**, *385*, 595–602.
(62) Rosen, N.; Bolen, J. B.; Schwartz, A. M.; Cohen, P.; DeSeau, V.; Israel, M. A. *J. Biol. Chem.* **1986**, *261*, 13754–13759.
(63) Verbeek, B. S.; Vroom, T. M.; Adriaansen-Slot, S. S.; Ottenhoff-Kalff, A. E.; Geertzema, J. G.; Hennipman, A.; Rijksen, G. *J. Pathol.* **1996**, *180*, 383–388.
(64) Bose, R.; Holbert, M. A.; Pickin, K. A.; Cole, P. A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 668–675.
(65) Porter, M.; Schindler, T.; Kuriyan, J.; Miller, W. T. *J. Biol. Chem.* **2000**, *275*, 2721–2726.
(66) Adams, J. A. *Biochemistry* **2003**, *42*, 601–607.
(67) Ozkirimli, E.; Post, C. B. *Protein Sci.* **2006**, *15*, 1051–1062.
(68) Dixit, A.; Verkhivker, G. M. *PLoS Comput Biol* **2009**, *5*, e1000487.
(69) Banavali, N. K.; Roux, B. *Proteins* **2007**, *67*, 1096–1112.