THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

αC Helix as a Switch in the Conformational Transition of Src/CDK-like Kinase Domains

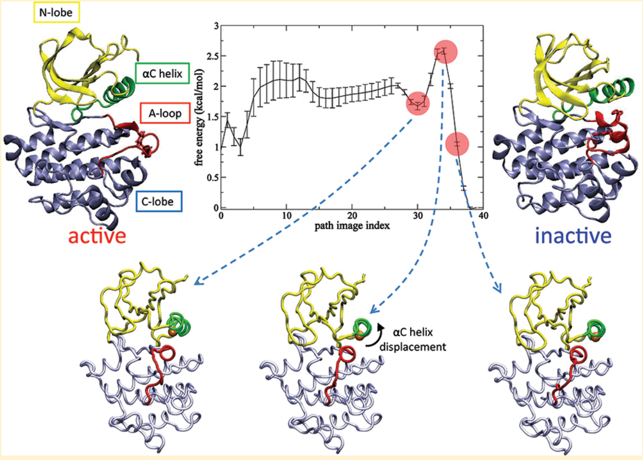
He Huang,[†] Ruijun Zhao,[‡,§] Bradley M. Dickson,[†] Robert D. Skeel,[‡] and Carol Beth Post*[,†]

[†]Department of Medicinal Chemistry and Molecular Pharmacology, Markey Center for Structural Biology and Purdue Cancer Center, Purdue University, West Lafayette, Indiana 47907, United States

[‡]Department of Computer Science, Purdue University, West Lafayette, Indiana 47907, United States

S *Supporting Information*

ABSTRACT: One mechanism of regulating the catalytic activity of protein kinases is through conformational transitions. Despite great diversity in the structural changes involved in the transitions, a certain set of changes within the kinase domain (KD) has been observed for many kinases including Src and CDK2. We investigated this conformational transition computationally to identify the topological features that are energetically critical to the transition. Results from both molecular dynamics sampling and transition path optimization highlight the displacement of the αC helix as the major energy barrier, mediating the switch of the KD between the active and down-regulated states. The critical role of the αC helix is noteworthy by providing a rationale for a number of activation and deactivation mechanisms known to occur in cells. We find that kinases with the αC helix displacement exist throughout the kinome, suggesting that this feature may have emerged early in evolution.



■ INTRODUCTION

Protein kinases constitute a large family of proteins that plays essential roles in regulating cellular pathways. To maintain normal cellular functionality, their activities are subject to tight regulation. Kinases share a highly conserved kinase domain (KD), which undergoes a conformational change that is central to this regulation.

As previously noted,[22,25,27,42] the KDs of different kinases share a similar active structure, whereas the inactive conformations often differ. Despite the variability associated with the inactive forms, one inactive conformation has been observed for a number of different kinases, including the cyclin-dependent kinases (CDKs),[7,12] the Src family kinases (SFKs),[51,56] the epidermal growth factor receptor (EGFR),[59] the zeta-chain-associated protein kinase 70 (ZAP70),[13] and Bruton's tyrosine kinase (BTK).[41] The conformational activation/deactivation and the regulatory mechanisms for some of these kinases have been reviewed.[27]

This inactive conformation is compared to the active structure for a Src family kinase in Figure 1. Three major structural features are involved in the active to inactive conformational transition. First, in the inactive form, the catalytic cleft between the N- and C-terminal lobes (N-lobe and C-lobe) is more closed due to a hinge motion of the two lobes. Second, the αC helix is rotated outward relative to the β strands in the N-lobe. This outward displacement of the αC helix breaks a catalytically important salt bridge between a conserved glutamate and a conserved lysine residue (E310 and K295 in chicken c-Src numbering), and is believed to contribute to rendering the KD catalytically inactive. Finally, the activation loop (A-loop) rearranges from an extended conformation to a more compact one, with the N-terminal part forming a short helix packed against the αC helix without a flip of the conserved DFG motif as observed in another class of inactive structures.[27] The C-terminal part of the A-loop of different kinases may take different conformations, and in the case of Src, another short α helix is formed.

Although a good number of structures for the two end states have been solved experimentally, the transition process of this conformational change remains to be elucidated. In the present work, we investigate this conformational transition computationally, taking an SFK member Lyn and a CDK member CDK2 as examples. We aim to define the important topological features that control the transition. A Gō-type coarse-grained protein model[29,30] enabled us to obtain well converged computational results despite the large-scale and multifaceted nature of the conformational transition. Instead of providing complete atomistic details of the molecule, a Gō potential

Received: February 18, 2012
Revised: March 23, 2012
Published: March 26, 2012

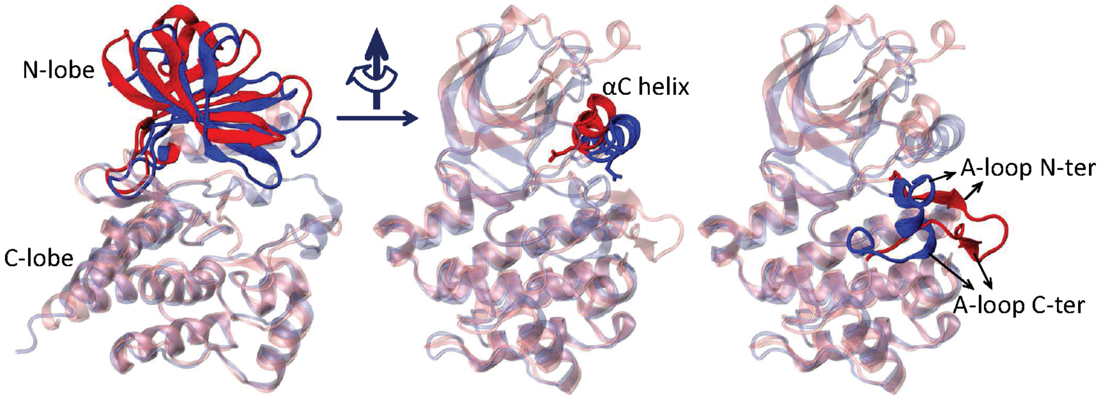


Figure 1. A comparison of the active (red) and inactive (blue) conformations of the kinase domain of the Src family kinase Lyn with the C-lobe superimposed. The conformational change mainly involves three structural elements as highlighted with solid colors in each panel: the N-lobe (left), the αC helix (middle), and the A-loop (right). The molecule is rotated by 90° along the vertical axis in the middle and right panels relative to the orientation in the left panel. The side chain of E310 is highlighted in the middle panel to show the rotation of the αC helix. The structures shown are homology models of Lyn based on crystal structures of the Src family kinases Lck (PDB id: 3LCK)[57] and Hck (PDB id: 1QCF)[51] as reported previously.[46]

models a protein at the residue level and provides descriptions of the chain connectivity and the contact pattern, in other words, of the topological features of a protein. In this study, we hope to capture the part of the mechanism underlying the transition process that is encoded in the topology of the molecules. Considering that a similar conformational transition is apparently broadly distributed over remotely related kinases, it is especially reasonable that a common topology-encoding mechanism might exist.

We characterize the molecular system in two ways. First, we use enhanced molecular dynamics sampling to explore the free energy landscape of the system. Second, we use the recently developed maximum flux transition path (MFTP) method[60] to locate an optimal path for the transition process by maximizing the "traffic" along the path connecting the two reference states. Results from both independent methodologies highlight the αC helix as a structural element critical to the energetics of the transition process. This finding provides a physical understanding of the previous observation that, repeatedly, interactions to the αC helix are involved in the biological mechanisms for regulation of different kinases. A systematic analysis of the KD structures in the Protein Data Bank (PDB) is carried out to examine the potential relevance of the current finding over the human kinome.

■ METHODS

The Double-Basin Gō Model. The kinase domains (KDs) of Lyn and CDK2 were both modeled using a double-basin Gō potential. Gō-type potentials realize a single-basin energy surface with the minimum energy at a known three-dimensional structure.[16] Whereas these single-basin potentials are widely used to study protein folding mechanisms,[9,11,44,54] Gō models featuring multiple basins have been developed to study conformational transitions by combining the single-basin Gō potentials built on different conformations.[6,23,39,43,61] For each of the two KDs, two single-basin Gō potentials V_{act} and V_{ina} were built based on the active and inactive structures, respectively, as described by Karanicolas and Brooks.[29] Detailed energy terms of this Gō model are described in the Supporting Information. The Gō models have a melting temperature of about 250 K. On the basis of this observation, we took 200 K as an estimate of "room temperature" in this study.

Following the procedure of ref 6, we then constructed a double-basin Gō potential V_{double} for each KD by merging the two single-basin potentials V_{act} and V_{ina}. A similar procedure applied to Src kinase has been reported before.[58] The merging procedure takes the form of an exponential average:

$$V_{\text{double}} = -\frac{1}{\beta}\ln\{\exp[-\beta(V_{\text{act}} + \alpha)] + \exp(-\beta V_{\text{ina}})\} \quad (1)$$

where α and β are parameters that modulate the relative stability of the two states and the barrier height between them, respectively. For Lyn and CDK2, these parameters were chosen on the basis of trial simulations. The α values were chosen to roughly balance the active and inactive forms, yet slightly favoring a different side in the two systems. Specifically, α was set to −4.0 kcal/mol for Lyn, which slightly favors the inactive form, and to −12.5 kcal/mol for CDK2, which slightly favors the active form. The β value for Lyn was set to 0.02 mol/kcal, which resulted in a few spontaneous transitions in a trial MD simulation of 400 ns at 200 K. Such a transition rate allowed us to converge the 2D free energy landscapes in 1 μs of replica exchange molecular dynamics simulation but is likely to be faster than the physical transition rate given the complexity of the conformational transition. To also examine the system at a slower transition rate, a greater β of 0.03 mol/kcal was used in the maximum flux transition path calculations of CDK2, which resulted in a significantly higher barrier compared to the Lyn system.

The reference structures for Lyn KD were obtained from homology modeling[46] based on the crystal structures of active Lck (PDB id: 3LCK)[57] and inactive Hck (PDB id: 1QCF).[51] For CDK2, the crystal structures of its active (PDB id: 1FIN)[24] and inactive (PDB id: 1HCK)[53] forms were used as the reference structures. A short loop from residues 37−40 missing in the inactive CDK2 structure was modeled with the program MODELLER.[50] For all structures, missing hydrogen atoms were added, and a mild energy minimization was carried out with the CHARMM program.[8] All molecular dynamics simulations carried out for these models were propagated with the Langevin integrator of CHARMM[8] using a time step of 10 fs and a friction coefficient of 5.0 ps^{-1}, and without a cutoff distance for the nonbonded energy terms.

Replica Exchange Simulation and Free Energy Landscape Construction. We first explored the conformational space of Lyn KD using replica exchange (Rex) enhanced molecular dynamics sampling.[19] The Rex runs consisted of five replicas spanning a temperature ladder of 200−240 K with a uniform step of 10 K. For each replica, 0.8 μs of dynamics were generated following 0.2 μs of equilibration. During the simulation, each trajectory made more than 1300 trips between the two ends of the temperature ladder. Results from simulations initiated with the active or inactive structure are highly similar, showing that 0.8 μs is long enough to converge the sampling.

2D free energy landscapes for different pairs of order parameters were constructed on the basis of the sampling data to visualize the conformational space. The examined order parameters are all difference quantities, each characterizing the progress of a certain property of the system. These order parameters include the energy difference $\Delta V \equiv V_{\text{act}} + \alpha - V_{\text{ina}}$, the rmsd difference $\Delta\text{rmsd}(X) = \text{rmsd}(X, X_{\text{act}}) - \text{rmsd}(X, X_{\text{ina}})$, and a contact number difference $\Delta Q = Q_{\text{act}} - Q_{\text{ina}}$. The physical meaning of the ΔV parameter is given at the beginning of the Results section. In the definition of Δrmsd, X is the examined configuration, whereas X_{act} and X_{ina} are the active and inactive structures, respectively. The contact difference ΔQ depends on the contacts unique to either the active or inactive states, namely, the pairs that are defined as Gō contacts in the one single-basin model but not in the other. For a given configuration, Q_{act} or Q_{ina} counts the number of distances from the active or inactive list of unique pairs, respectively, that are within 1.45 times their values in the corresponding reference structure.

The weighted histogram analysis method[14,32] was used to combine the sampling data obtained at each temperature while assuming each temperature was sampled independently. Such an assumption is justified by the large number of travels across the temperature ladder made by each replica, as discussed in ref 10. Specifically, a histogram H was first constructed by dividing the configurations sampled at all L different temperatures into M bins according to their potential energy V. The density of states Ω_i associated with each potential energy bin i was calculated by solving the following equation iteratively:

$$\Omega_i = \frac{H_i \sum_{m=1}^{M} \Omega_m \sum_{l=1}^{L} e^{-\beta_l V_m}}{N \sum_{l=1}^{L} e^{-\beta_l V_i}} \tag{2}$$

where N is the total number of configurations, H_i the number of those in bin i, V_i the center potential energy of bin i, and β_l the lth inverse temperature. A 2D histogram $H^{P,Q}$ for the pair of order parameters P and Q at the inverse temperature β_0 corresponding to the reference temperature of 200 K was then constructed by binning all sampled configurations according to their P and Q values but reweighting each count with a weighting factor

$$\omega_m = \frac{\Omega_m e^{-\beta_0 V_m}}{H_m} \tag{3}$$

where m is the index of the bin in the 1D histogram H that the potential energy of the configuration falls into. A 2D free energy map was obtained by calculating the free energy as $F_{ij} = -\beta_0^{-1} \ln[H_{ij}^{P,Q}/\max(H_{ij}^{P,Q})]$. Equations 2 and 3 are equivalent to eqs 72 and 68 of ref 10.

The Maximum Flux Transition Path Calculation. We further characterized the transition process of the KDs of both Lyn and CDK2 using the maximum flux transition path (MFTP) method.[60] The MFTP method is a generalization of a path optimization method for the Cartesian space due to refs 5 and 21. The algorithm of the MFTP method is described in ref 60. In brief, it resolves the path that carries maximum flow between two metastable states as a series of images in a collective variable space through iterative updates starting from an initial guess. In each iteration, an update vector is estimated on the basis of restrained molecular dynamics sampling at each of the path images. In all calculations reported here, the path was represented by 40 images. In each iteration, 50 ps of restrained equilibration and 500 ps of restrained sampling at 200 K were carried out for each image. The force constant was 10 kcal/(mol·Å^2) for all restraints. The virtual time step used for updating the path is 0.001 in CHARMM time units squared, or (1.55 fs)^2.

The MFTP calculations were carried out in a space consisting of seven collective variables for both KDs. Each of the seven variables characterizes the progress of the transition in terms of the relative orientation between a pair of structural elements, a full list of which is described in the Results section. Between two structural elements, a collective variable is defined as a combination of individual inter-residue distances connecting them in the following form:

$$z_i = \sum_{j=1}^{n_i} \frac{r_{ij}^{\text{act}} - r_{ij}^{\text{ina}}}{|r_{ij}^{\text{act}} - r_{ij}^{\text{ina}}|} r_{ij} \tag{4}$$

where r_{ij} is the jth distance for defining collective variable z_i and r_{ij}^{s} is the reference value of this distance in state s. The distances included in the summation are those differing by at least 1 Å in the two end states and at the same time contributing to the contact energy in at least one of the two single-basin Gō potentials (see Table S1 in the Supporting Information for a detailed list).

For each KD, the MFTP calculation was started from three different initial paths, one generated by linear interpolation (*linear*) and two defined from targeted molecular dynamics (TMD)[52] runs (α*C-Aloop*, *Aloop-*α*C*). To generate the starting configurations for the *linear* path, the inactive structure was gradually pulled toward the active structure with harmonic restraints applied to each inter-residue distance used to define the collective variables. The pulling process involves changing the reference values of the harmonic restraints from the inactive distances to the active distances in 2000 consecutive steps. In each step, the structure was subject to 100 steps of minimization and 20 ps of restrained dynamics. The α*C-Aloop* and *Aloop-*α*C* paths were generated each by two TMD runs starting from the active structure. In the first run, the rmsd to the inactive structure of the N-lobe (including αC) or C-lobe (including A-loop) was minimized for the α*C-Aloop* or *Aloop-*α*C* path, respectively. In the second run, the rmsd of the whole molecule to the inactive structure was minimized. In each of the three initial paths, two critical structural changes happen in a different order: the α*C-Aloop* initial path has the αC helix displacement preceding the A-loop rearrangement, whereas the *Aloop-*α*C* initial path has the opposite sequence. These two events happen concurrently in the *linear* initial path.

We monitored the convergence of the path optimization by calculating the distance between the path at the last iteration to the paths at the previous iterations. The distance between two

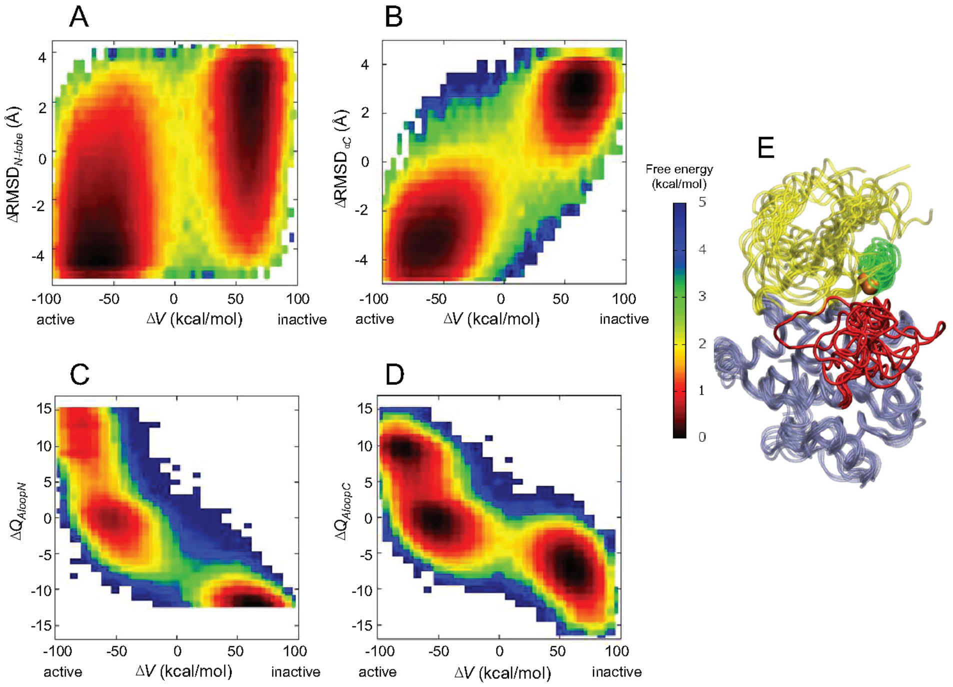


Figure 2. (A−D) Free energy landscapes of Lyn at the reference temperature for different structural parameters against the transition progress variable ΔV. The relative motion of the two lobes (A) and the movement of the αC helix (B) are examined via the corresponding Δrmsd parameters. Δrmsd gives the difference between the specific rmsd to the active structure and that to the inactive structure. The C-lobe was superimposed when calculating the rmsd for N-lobe, and the N-lobe excluding the αC helix was superimposed when calculating the rmsd for the αC helix. The structural changes of the N-terminal (C) and C-terminal (D) segment of the A-loop are examined via the corresponding ΔQ parameters. ΔQ gives the difference between the numbers of formed active-unique contacts and inactive-unique contacts. (E) Representative structures from the intermediate basin where the A-loop is disordered. The N-lobe, αC helix, A-loop, and C-lobe are colored in yellow, green, red, and light blue, respectively. The orange sphere highlights the inward position of E310, indicating the rotational orientation of the αC helix. For a temperature dependence of the stability of the intermediate basin, see also Figure S1 in the Supporting Information.

paths P and Q was defined as the distance between each corresponding image pair averaged over all pairs along the paths:

$$d(P, Q) = \frac{1}{40}\sum_{i}^{40}\sqrt{\sum_{j}^{7}(P_{ij} - Q_{ij})^2} \tag{5}$$

where P_{ij} is the value of the jth collective variable of the ith image in path P.

After the path evolution had converged, a long iteration with 1.5 μs of restrained sampling was carried out to estimate the free energy of the collective variables along the optimal paths by numerically integrating the free energy gradient ∇F evaluated at each path image as described in ref 60. Specifically, $F_{39} = 0$ and $F_i = {}^1/_2\sum_{k=i}^{38}(\nabla F_k + \nabla F_{k+1})\cdot(Z_k - Z_{k+1})$ for $i < 39$, where Z_k is the coordinate of the kth path image in the collective variable space.

Analysis of Kinase Domain Structures in PDB. A systematic analysis of the kinase structures in the Protein Data Bank (PDB) was carried out to examine the prevalence of the displacement of the αC helix. To identify structures of kinase domains in the Protein Data Bank (PDB), we first ran a structural alignment with the PDBeFold server[31] using the inactive model structure of Lyn as the query structure against the whole PDB archive with 40% query lowest acceptable match and 20% target lowest acceptable match. The sequences of the 2781 chains reported as hits in the structural alignment were then aligned against 478 human eukaryotic protein kinase (ePK) sequences[37] using BLASTP.[1] 2202 of the 2781 chains match at least one ePK sequence with more than 95% identities plus gaps, and are identified as structures of human ePK kinase domains. For each of the remaining chains, the Cα−Cα distance between the two residues that align with E310 and K295 of c-Src was calculated if applicable as an indicator of the position of the αC helix with respect to the N-lobe β strands.

RESULTS

2D Free Energy Landscapes for Lyn Kinase Domain. In this section, we characterize the conformational space of the kinase domain (KD) of Lyn by examining free energy landscapes for multiple structural parameters constructed from replica exchange (Rex) enhanced molecular dynamics sampling. Each free energy landscape examines one structural parameter against a common coordinate, $\Delta V \equiv V_{\text{act}} + \alpha - V_{\text{ina}}$, which is a natural progress variable to characterize the relative position of a configuration with respect to the two basins. In the merged double-basin model, the force $\mathbf{f}$ that a configuration feels is related to the forces it would feel in the two original single-basin models $\mathbf{f}_{\text{act}}$ and $\mathbf{f}_{\text{ina}}$ by a linear combination $\mathbf{f} = c_{\text{act}}\mathbf{f}_{\text{act}} + c_{\text{ina}}\mathbf{f}_{\text{ina}}$, where the ratio of the coefficients c_{act} and c_{ina} is directly related to ΔV by

$$\frac{c_{\text{act}}}{c_{\text{ina}}} = \exp(-\beta\Delta V) \tag{6}$$

Therefore, a large negative/positive value of ΔV indicates that the force applied to the molecule is mostly contributed by the active/inactive component of the merged potential, whereas values close to zero indicate a configuration is close to the barrier region.

As described in the Introduction, the transition involves mainly the movement of three structural elements, namely, the hinge between the N-lobe and C-lobe, the αC helix and the A-loop. These features are natural choices for structural parameters to characterize and such parameters are used to construct the free energy surfaces reported below. It should be noted that the relative stability of the two states and the barrier height between them reflected in the free energy landscapes are related to the choice of the α and β parameters. Accordingly, the results identify qualitative topological features, such as the shapes and positions of the free energy basins, while quantitative conclusions on the relative stability and transition rate cannot be assessed.

Position of the αC Helix Correlates with the Reaction More Strongly than Does Lobe–Lobe Orientation. Parts A and B of Figure 2 show the free energy landscapes for the structural parameters $\Delta\text{rmsd}_{\text{Nlobe}}$ or $\Delta\text{rmsd}_{\alpha\text{C}}$ versus ΔV. For $\Delta\text{rmsd}_{\text{Nlobe}}$, the rmsd's to the two reference structures are calculated over the N-lobe residues excluding the αC helix with the C-lobe superimposed, and for $\Delta\text{rmsd}_{\alpha\text{C}}$, they are calculated for the αC helix residues with the rest of the N-lobe superimposed. These two coordinates characterize the closing/opening hinge motion of the N- and C-lobes and the swinging/rotating motion of the αC helix within the N-lobe, respectively.

On both landscapes, the progress of the transition, ΔV, clearly divides the space into two metastable regions with negative ΔV corresponding to the active state and positive ΔV to the inactive state. The broad basins with extensive overlap along the vertical axis in Figure 2A imply that the lobe–lobe motion occurs with little cost in free energy. This facileness of the lobe–lobe motion has been observed also in all-atom simulations of Src kinases.[3,46] In contrast, the two basins in Figure 2B for the αC helix motion show significantly less overlap in the $\Delta\text{rmsd}_{\alpha\text{C}}$ coordinate, suggesting the position of the αC helix correlates more strongly with the progress of the transition than does the relative position of the two lobes.

An A-Loop Melted Intermediate on the Active Side ($\Delta V < 0$). The rearrangement of the A-loop involves partial unfolding and refolding, and is more complex than the lobe–lobe and αC helix motion. In the inactive conformation, the A-loop folds into two short helical segments of which the N-terminal one packs against the αC helix (Figure 1C). In the active conformation, both segments adopt extended conformations and form contacts to different parts of the C-terminal lobe. Parts C and D of Figure 2 examine the two segments of the A-loop separately by plotting the free energy landscapes for the contact number difference ΔQ of these two segments. The definition of ΔQ was given in the Methods section. When calculating ΔQ_{AloopN} or ΔQ_{AloopC}, only residue pairs that involve at least one residue in the N-terminal or C-terminal segment of the A-loop, respectively, are used to build the pair lists.

In both parts C and D of Figure 2, the active and inactive states appear as minima near the upper-left and lower-right corners. Interestingly, an additional metastable region exists, which is energetically more similar to the active state as suggested by the negative ΔV. Figure 2E shows representative structures for this intermediate with both ΔQ_{AloopC} and ΔQ_{AloopN} within $[-4, 1]$ and ΔV within $[-70$ kcal/mol, -30 kcal/mol$]$. Consistent with the strong correlation between the position of the αC helix and ΔV, configurations in this metastable state mainly have the αC helix in the inward position. When the αC helix is not "out", the A-loop can adopt a variety of conformations, consistent with the fact that this intermediate gains stability as the temperature rises (see Figure S1, Supporting Information).

A Floppy C-Terminal Segment of A-Loop on the Inactive Side ($\Delta V > 0$). A striking difference between Figure 2C and 2D is in the shape of the basin on the inactive side, which indicates that the N- and C-terminal segments of the A-loop behave differently near the inactive state. The inactive basin in Figure 2C has a flat shape. Its small vertical span suggests that the N-terminal segment has very well-defined structure in the inactive state. In contrast, the inactive basin of the C-terminal segment of the A-loop (Figure 2D) spans a wide range of ΔQ values, suggesting a flexible conformation.

The difference of the two parts of the A-loop discussed above implies that the ordered structure of the N-terminal segment plays an important role in maintaining the overall inactive shape. This role of the N-terminal segment is likely due to its close proximity to the αC helix in the inactive structure. Its helical form is required for making correct contacts with the αC helix so as to stabilize the αC helix in the displaced conformation. In contrast, the C-terminal segment can take a variety of conformations on both the active and inactive sides, suggesting that this part of the A-loop is less relevant to the energetics controlling the transition process. The flexibility of the C-terminal region of the A-loop revealed by our coarse-grained simulation agrees with all-atom simulations of Hck[4] and suggests that such flexibility arises from the contact topology of the structures.

That the N-terminal region of the A-loop is important for the structural integrity of the inactive state, while the C-terminal region is not, gives a physical understanding of previous mutagenesis studies of the Src family kinases in which it was shown that substitutions in the N-terminal rather than the C-terminal region of the A-loop increase the kinase activity regardless of the down-regulation by C-terminal tail phosphorylation.[17,34] In a broader sense, the computational finding provides a rationale for the fact that the helical structure formed by the N- but not the C-terminal segment is often conserved in different kinases that have the αC helix displaced outward in the inactive structures.

Maximum Flux Transition Paths of Lyn Kinase Domain. While the free energy landscape for two coordinates can be easily visualized as a 2D map constructed from sampling data, a more detailed characterization of the transition would require the simultaneous examination of more than two variables, so that constructing the full landscape quickly becomes an infeasible task as the dimensionality goes beyond a very small number. Path optimization methods provide a way to explore higher dimensional spaces. In this section, we report the use of the maximum flux transition path (MFTP) method[60] for locating optimal transition paths in a space consisting of seven collective variables. An MFTP is a path along which the flow rate between two stable states is maximized. It approaches the minimum free energy path at zero temperature but is closer to an ideal path at finite temperatures. In the optimization, the path is represented as a series of points or images, each with a specified set of values for the collective variables.

As described in the Methods section, the collective variables used in this study are defined as linear combinations of individual inter-residue distances connecting different structural elements of the KD. Figure 3A shows a list of the seven pairs of

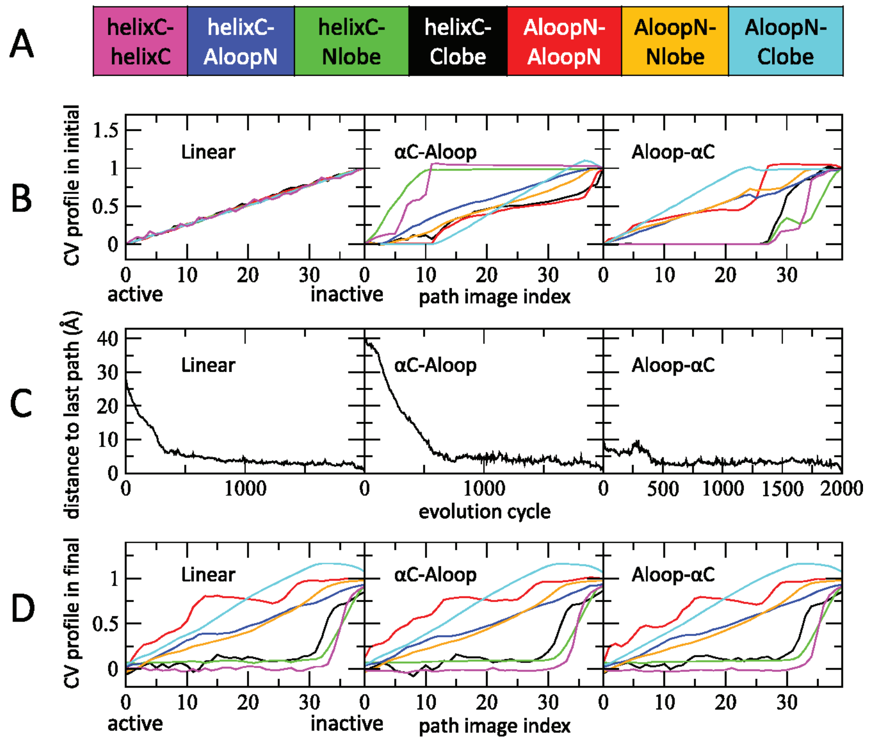


Figure 3. MFTP path optimization of Lyn KD using seven collective variables converges to the same path starting from three different initial paths. (A) Pairs of structural elements used to define the collective variables. For a full list of individual distances used to define each collective variable, refer to Table S1 in the Supporting Information. (B) Profile of each collective variable (CV) in the initial path shown by plotting the normalized values of each variable along the path. Each collective variable z is normalized to $\overline{z} = (z - z_{\text{act}})/(z_{\text{ina}} - z_{\text{act}})$ so that $\overline{z}$ has a value of 0 in the active structure and 1 in the inactive structure. The collective variable curve is colored as in part A. (C) Distance between the last path and the paths from previous iterations showing the convergence of the optimization. (D) The profiles of the collective variables in the final paths are nearly identical for all three initial paths.

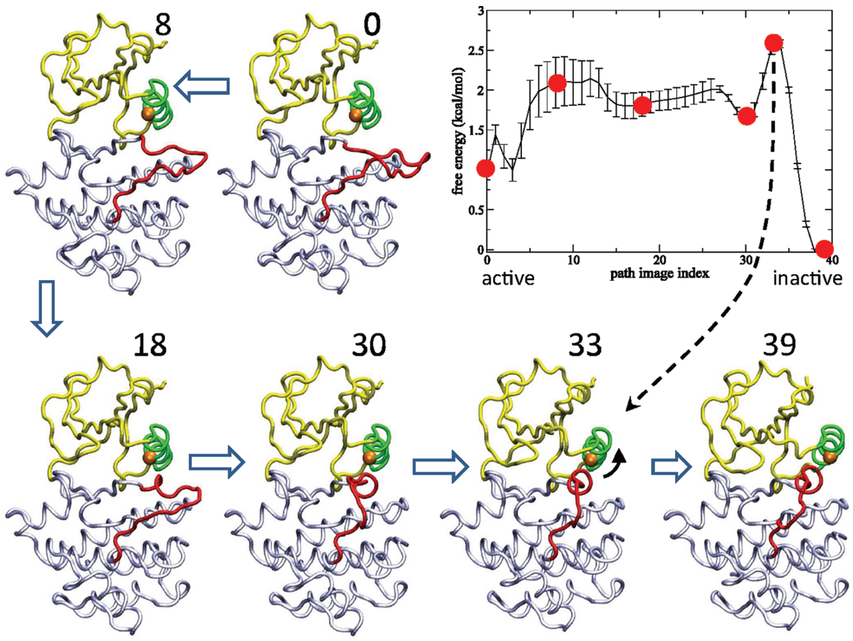


Figure 4. The free energy profile along the optimal transition path of Lyn KD and representative structures at different path images. The error bars show the standard error for the free energies calculated from the first and second halves of the 1.5 μs of sampling time. The average structure over the entire sampling time is shown for the six labeled path images in tube representation with the same color scheme as that used in Figure 2E. For an evolution history of the free energy of the *αC-Aloop* path, see also Figure S2 in the Supporting Information.

structural elements used to define the seven collective variables. Collectively, they characterize the configuration of the αC helix and the N-terminal segment of the A-loop both internally and with respect to the relatively rigid N-lobe β strands and the C-lobe. The C-terminal segment of the A-loop was not included because it is flexible in both active and inactive states in the Rex simulations. Similarly, the relative orientation between the N-lobe and C-lobe was excluded from the list because of the broad basins in Figure 2A.

Like other transition path optimization methods such as the nudged elastic band[26] method and the string method,[40] the MFTP method evolves an initially guessed path into a locally optimized path, and thus potentially depends on the initial guess. To get a more global idea of the path space, we started the evolution from three initial paths, each generated to have a different sequence of events (see Methods). These three initial paths are shown in Figure 3B by plotting each normalized collective variable as a function of the path image index with curves colored corresponding to Figure 3A. The normalization shifts and scales each collective variable so that they all take a value of 0 in the active structure and 1 in the inactive structure. The differences in the three initial paths are apparent from the comparison of the three panels.

Three Initial Paths Converge to the Same Final Path. Each initial path was optimized using the MFTP method for 2000 iterations. The three panels of Figure 3C plot the distance from the path at each iteration to the final path to visualize the convergence of the optimization (see Methods for the definition of the distance between two paths). For all three calculations, the distance-to-final curves plateau to a small value over the last 1000 iterations, indicating that each path has reached a stable position, about which it fluctuates due to sampling error in evaluating the free energy gradient in individual iterations.

The three panels of Figure 3D show the profile of each collective variable along the three final paths. In contrast to the initial paths, the final paths are strikingly similar, indicating that the three calculations have converged to one consensus path. In this common final path, the movement of the αC helix as indicated by a concerted jump in three collective variables (magenta, black, and green) occurs near the inactive state. In other words, it is the last step to accomplish when going from active to inactive. Prior to this event, the N-terminal part of the A-loop has arranged itself in an inactive-like conformation. Specifically, at path image 30, the value of the red collective variable has reached 1, indicating a helix-like shape of the N-terminal segment of the A-loop. Moreover, the blue, orange, and cyan collective variables all have values close to 1, indicating that this segment is located close to its inactive position relative to the αC helix, the N-lobe, and the C-lobe. A value greater than 1 for the cyan collective variable is caused by a movement of the A-loop further into the center of the catalytic cleft and away from the C-lobe residues. This overshooting allows the A-loop to make contacts with the αC helix which is still close to its active position.

Free Energy along the Path Highlights the Critical Role of the αC Helix. The free energy profile and representative average structures along the consensus optimal path are shown in Figure 4. Starting from the active side, the A-loop first breaks its contacts with the C-lobe (image 0 to 8), causing a rise at the beginning of the free energy curve. It then moves a long distance toward the catalytic cleft center in a diffusive fashion, as indicated by the long and relatively flat shoulder of the free energy curve from image 8 to 30. The N-terminal part of the A-loop starts to form the helical structure when it gradually moves into the space below the αC helix. The minor dip in free energy at image 30 corresponds to a conformation in which this short helix has almost formed while the αC helix is still at its active position and where the cyan collective variable value is greater than 1. The displacement of the αC helix occurring around image 33 corresponds to the highest barrier along the path, after which the free energy decreases by about 2.6 kcal/mol.

The evolution history of the path and its free energy profile during the optimization further establishes the critical role of the αC helix. As the initial path evolves toward the optimal path, the height of the major free energy barrier decreases in all three optimizations. The most significant decrease occurred in the optimization of the *αC-Aloop* path, the evolution of which is shown in Figure S2 in the Supporting Information. Regardless of the change in barrier height and position along the path, the barrier position always closely follows the motion of the αC helix (Figure S2, Supporting Information). Such coincidence shows that the largest energy penalty in the transition process comes from the movement of the αC helix, which is physically reasonable considering that its displacement involves simultaneous breaking of a number of contacts. To minimize this energy penalty, the molecule chooses the path in which the A-loop prearranges itself before the αC helix moves so that its N-terminal part will be ready to make contacts with the relocated αC.

The role of the αC helix revealed in the path calculations confirms the conclusions drawn from the 2D free energy surfaces in a higher dimensional space. In this space, the disordered A-loop conformations from the intermediate state on the 2D landscapes are better resolved and correspond to the long segment from image 8 to 30 of the optimal path with relatively flat free energies. We note that the αC helix is also the origin of the transition barrier reported by Yang and Roux;[58] however, the sequence of events in the optimal path described above is apparently inconsistent with theirs, in which a double-basin Gō potential was used to characterize the active and inactive states of the KD of a different Src family kinase Hck. The free energy surface reported in ref 58 showed that, when going from active to inactive, the αC helix moves first and the A-loop moves later. Part of this discrepancy can be explained by the fact that the A-loop was treated as a whole piece[58] but separated into two parts in the present study. With respect to the C-terminal part, our results agree with those of ref 58 in that the C-terminus of the A-loop can still be unfolded after the αC helix moves outward. Nevertheless, the N-terminal part of the A-loop makes the active to inactive transition before the αC helix moves in our calculation. This discrepancy is likely due to differences in the Gō models. In the model used in the present study (see the Supporting Information for details), the contact strengths are specific to each contact pair instead of uniform as in ref 58, potentially preserving more details from the all-atom representation. Moreover, sequence instead of structure derived dihedral references are used to remove the dominating effect of the dihedral terms in determining the conformation of the system.[29] Finally, the force constant and radius of the repulsive terms in the Gō model used in ref 58 result in a barrier as small as 0.25 kcal/mol for the protein chain to cross over itself. The parameters used in the present study give barriers that are 2 orders of magnitude greater and thus prevent the chain crossing in the dynamics.

Maximum Flux Transition Paths of CDK2 Kinase Domain. As mentioned in the Introduction, inactive conformations similar to those of Src kinases have been observed for a number of other kinases, among which CDK2 is a famous example.[7] Since the optimal path calculated for Lyn can be understood on the basis of the contact topology of the molecule, the same mechanism is expected to also apply to the KDs of these kinases. To test this prediction, MFTP optimization was carried out for CDK2. The cyclin-dependent kinases are serine/threonine kinases belonging to the CMGC group of protein kinases. They are relatively remotely related in sequence to SFKs among the different kinases for which the similar inactive conformation has been observed. Figure 5A compares the inactive structure of CDK2 and that of Lyn KD. The two structures differ mainly near helix αG in the C-lobe and at the C-terminal part of the A-loop. The αC helix of the inactive CDK2 structure is outwardly displaced, similarly to Lyn, but with its N-terminal end further from the N-lobe β strands.

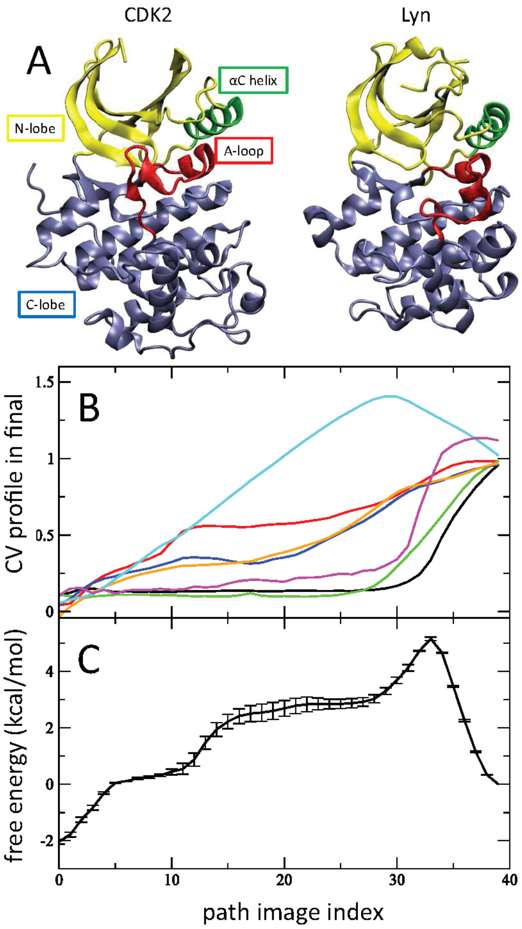


Figure 5. MFTP path optimization of CDK2 using seven collective variables converges to a path similar to that of Lyn KD. (A) A comparison of the inactive structure of CDK2 and Lyn KD. Both structures are shown in cartoon representation with the same color scheme used in Figure 2E. The C-terminal part of the A-loop of CDK2 takes a different conformation compared to Lyn. The C-lobes of the two structures also differ substantially. (B) Profiles of the collective variables along the optimal transition path of CDK2, colored as in Figure 3A. (C) Free energy profile along the optimal path. The calculation of the free energy profile and error bars are the same as described in the caption of Figure 4 for the Lyn system.

The activation/deactivation path of CDK2 was determined with the MFTP method in the same way as for Lyn, using the same collective variables and starting from three similar initial paths. As in the Lyn case, all three optimizations arrived at the same final path. Parts B and C of Figure 5 show the profiles of the collective variables and the free energy along this common final path. The prediction of the same mechanism as observed for Lyn is well supported. First, the final path defined by the collective variable profiles agrees well with that of Lyn for the overall progress pattern, in which the active to inactive transition initiates with the A-loop leaving the C-lobe and concludes with the αC helix moving away from the β strands in the N-lobe. Second, the free energy profile shows a prominent peak close to the inactive end, coincident with the movement of the αC helix as indicated by the rise of the magenta, black, and green coordinates. As in the transition of Lyn, the movement of the αC helix is the energetically costly step. We note that the α and β parameters chosen for the double-basin potential of CDK2 differ from those used for Lyn KD (see Methods). Notably, the CDK2 value for α switches the favored form to the active state, and the value for β results in a higher barrier for CDK2 compared to Lyn. As such, the above-noted qualitative characteristics of the transition behavior and free energy profile are not sensitive to the relative stability of the two states as determined by α and the height of the barrier as determined by β at least within the variation range observed for the two systems. Unlike the Lyn path, the free energy along the optimal path of CDK2 increases monotonically from the active state to the major barrier. This difference is related to the different α and β parameters rather than a physically relevant feature.

■ DISCUSSION

The Key Role of the αC Helix in SFK Regulation. In the present study, we examined the KD of a Src family kinase, Lyn, to determine which of those features observed from crystal structures of active and inactive kinase states are energetically critical to the transition. Our results obtained from both conformational sampling and path calculation show a strong correlation of the αC helix motion with the major barrier of the transition, suggesting a picture in which the αC helix acts as an energetic switch between the active and inactive conformations.

Whereas the present study focuses on the KD, full length Src kinases contain two regulatory domains SH2 and SH3 in addition to the KD. In vivo, the activity of the KD is down-regulated when the three domains form an assembly. The highlighted critical role of the αC helix suggests the potential of this structural element as a target for interdomain allosteric regulation. Indeed, several previous studies on SFKs have pointed out that, in the formation of the down-regulating SH3-SH2-KD assembly, the allosteric signal from the SH2 and SH3 domains is transmitted to the KD through the N-terminal segment of KD via its interaction with the αC helix.[2,18,33,47] The critical role of the αC helix found in the present study is consistent with such a picture. On the other hand, our results are less supportive of an alternative mechanism that has been proposed to explain the allosteric regulation by SH2−SH3, in which the formation or dissociation of the assembly affects the relative orientation of the two lobes of the KD which in turn affects the flexibility of the C-terminal segment of the A-loop and hides or exposes the phosphorylation site Tyr416.[56] Our dynamics studies show the C-terminal segment of the A-loop is intrinsically flexible. Moreover, its flexibility is not affected by restraints on the relative orientation of the two KD lobes.

Interaction with the αC Helix as a Common Regulatory Mechanism in the Kinome. The key role of the αC helix in the transition of Lyn KD characterized by the Gō potential is related to the fact that, as a relatively rigid structural element, the motion of the αC helix has the potential

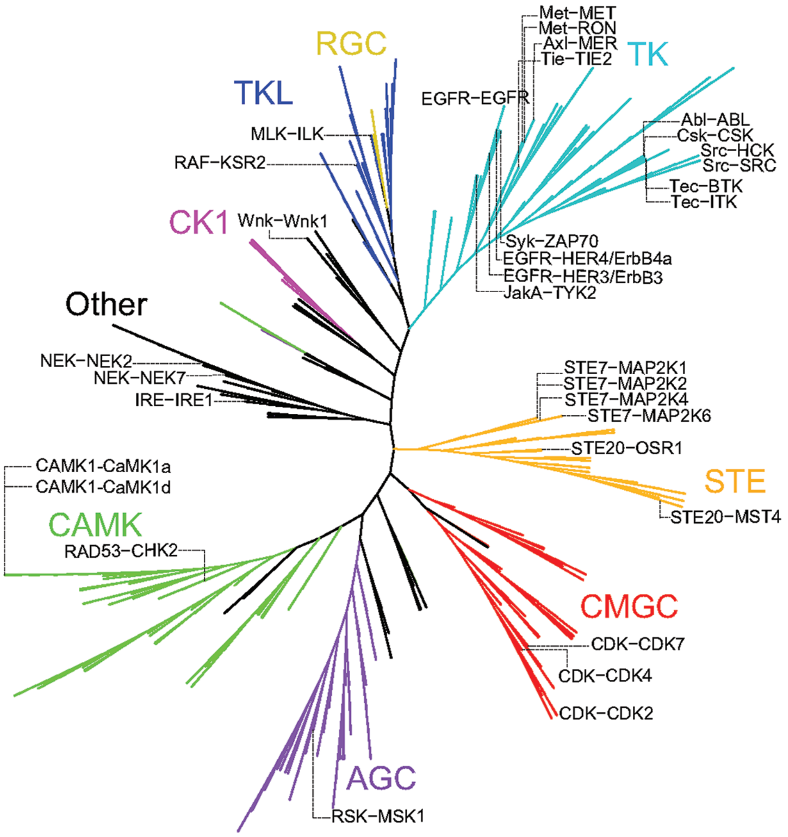


Figure 6. The distribution in the human kinome of kinases with the αC helix displaced outward from the N-lobe. Each kinase with a structure in the PDB with a distance greater than 14.5 Å between the two residues aligned with E310 and K295 of c-Src is labeled. Major groups are colored and labeled. The aligned sequences and phylogeny data of the human kinome are taken from www.kinase.com/human/kinome.[37] The dendrogram of the phylogenetic tree is drawn with the web-based program Interactive Tree Of Life.[36]

to induce the greatest perturbation to the system. The similar optimal transition path from MFTP calculation for CDK2 confirms that such a role is a consequence of the overall topology of the system rather than specific interactions. We therefore propose that the αC helix may play a general and critical role in the activation/deactivation transitions of different kinases that possess the Src/CDK-like inactive conformation, in which the αC helix is displaced outward and the N-terminal segment of the A-loop takes a helical form. These kinases include but are not limited to EGFR,[59] Zap70,[13] Mer,[20] BTK,[38] NEK7,[49] OSR1,[35] and CDPK.[55]

In fact, regulatory mechanisms involving interactions with the αC helix have been observed/proposed repeatedly among well studied kinases with a Src/CDK-like inactive conformation. For example, the activation of CDKs and EGFR requires the binding of cyclin or another kinase domain directly to a hydrophobic patch close to the αC helix[12,59] region. A similar mechanism has been proposed for the activation of Nek7 by its binding partner Nek9.[49] The alternative regulatory mechanism of SFKs by assembly of the regulatory domains induces specific interactions to the interface or hinge between the αC helix and the β4 strand in the N-lobe and thus activates or deactivates the kinase domain. Besides Src, examples include Zap70[13,28] and CDPK.[55]

Given the key role of the αC helix suggested from the present study, we considered the potential relevance of αC in regulation for an even broader range of kinases in the kinome, namely, those kinases with an αC helix displaced outwardly from a catalytically active conformation. We therefore surveyed all structures in the Protein Data Bank (PDB) based on a sequence alignment against human eukaryotic protein kinase sequences to identify all available structures of kinase domains (see Methods). Of these structures, those with a $C\alpha{-}C\alpha$ distance greater than 14.5 Å between the two residues aligned with E310 and K295 of c-Src were determined. As shown in Figure 6, such conformations are present in all major branches of the human kinome and are especially concentrated in the group of tyrosine kinase (TK). In addition to the TK group and groups CMGC and CAMK with kinase members noted above, the groups TKL, STE, and AGC also have known structures with the αC displaced outward and have potential to be regulated by intra- or intermolecular interactions to stabilize the aC orientation, and thus control catalytic activity. The extensive coverage of the kinome with an αC helix displaced conformation suggests that a regulatory role of the αC helix as an activation switch might have emerged at an early stage of the evolution and is a preserved topological feature in very different branches of the kinome.

Methodological Aspects of the Path Calculation. The collective variables we used in the path calculation are linear combinations of individual distances. They are more collective and of a smaller number compared to tens or hundreds of individual distances or Cartesian coordinates, which are the usual choices of similar path calculations.[15,45,48] Use of a large set of collective variables has the advantage that more degrees of freedom are defined by the collective variables so that the space consisting of the rest of the degrees of freedom is usually small and requires less sampling, which is essential for applying the method to systems that are very expensive to sample. However, while reducing the dimensionality of the orthogonal space, such an approach results in a high-dimensional and potentially rugged collective variable space, which poses two problems for path optimization methods. First, very densely spaced images would be needed to characterize a path. More importantly, there would be a large number of locally optimal paths connecting the two end states, which would result in a different final path from nearly every different initial path, making it very hard to interpret the result from a single or a few optimizations. In fact, we observed both problems during early stages of our study when we used a set of more than 100 inter-residue distances as the collective variables. In that case, the free energies along the paths are rugged and the sequence of events of the optimized path is always the same as that of the initial path. To our knowledge, the present work is the first one in which a consensus optimal path is reached starting from alternative initial paths with disparate sequence of events for a biological system. We note that such results are only possible when the collective variable space is not rugged. At the same time, it is reasonable to interpret a globally relevant mechanism based on a few optimizations only when the collective variable space is smooth so that the total number of locally optimal paths is also no more than a few.

■ CONCLUSIONS

We investigated the conformational transition between the active and inactive states of the KDs of Src and CDK2 with conformational sampling and path optimization approaches. A robust result from both methodologies and both molecular systems is the identification of the displacement of the αC helix as a topological feature that dominates the major energy barrier for activation. The calculation also defines the sequence of events along the optimal path to be that the A-loop folds into the active site followed by the αC helix movement when going from the active to the inactive state. The key finding that displacement of the αC helix is the origin of the major energy barrier and thus controls the switch of the KD between the active and down-regulated states provides an explanation for a number of known activation and deactivation mechanisms of many kinases. Finally, we find that kinases with the αC helix displacement exist throughout the kinome, suggesting that this feature may have emerged early in evolution.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Detailed description of the single-basin Gō potential, a 2D free energy landscape of Lyn KD at three different temperatures, lists of individual distances used to define the seven collective variables, the evolution history of one path of Lyn KD. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: cbp@purdue.edu. Phone: 765-4945980. Fax: 765-494141.

Present Address

[§]Department of Mathematics and Statistics, Minnesota State University, Mankato, MN, 56001, USA.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) grants GM039478 and GM083605 (C.B.P.). H.H. was supported by Purdue Research Foundation Fellowships.

■ ABBREVIATIONS

KD, kinase domain; CDK, cyclin-dependent kinase; SFK, Src family kinase; SH2, Src homology domain 2; SH3, Src homology domain 3; N-lobe, N-terminal lobe; C-lobe, C-terminal lobe; αC, α helix C; A-loop, activation loop; AloopN, N-terminal segment of A-loop; AloopC, C-terminal segment of A-loop; MFTP, maximum flux transition path; rmsd, root mean squared deviation; CV, collective variable

■ REFERENCES

(1) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403−410.

(2) Banavali, N. K.; Roux, B. *Structure* **2005**, *13*, 1715−1723.

(3) Banavali, N. K.; Roux, B. *Proteins* **2007**, *67*, 1096−1112.

(4) Banavali, N. K.; Roux, B. *Proteins* **2009**, *74*, 378−389.

(5) Berkowitz, M.; Morgan, J. D.; McCammon, J. A.; Northrup, S. H. *J. Chem. Phys.* **1983**, *79*, 5563−5565.

(6) Best, R. B.; Chen, Y. G.; Hummer, G. *Structure* **2005**, *13*, 1755−1763.

(7) Bondt, H. L. D.; Rosenblatt, J.; Jancarik, J.; Jones, H. D.; Morgant, D. O.; Kim, S. H. *Nature* **1993**, *363*, 595−602.

(8) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(9) Cho, S. S.; Weinkam, P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 118−123.

(10) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3*, 26−41.

(11) Das, P.; Wilson, C. J.; Fossati, G.; Wittung-Stafshede, P.; Matthews, K. S.; Clementi, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14569−14574.

(12) Davies, T. G.; Tunnah, P.; Meijer, L.; Marko, D.; Eisenbrand, G.; Endicott, J. A.; Noble, M. E. *Structure* **2001**, *9*, 389−397.

(13) Deindl, S.; Kadlecek, T. A.; Brdicka, T.; Cao, X.; Weiss, A.; Kuriyan, J. *Cell* **2007**, *129*, 735−746.

(14) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

(15) Gan, W.; Yang, S.; Roux, B. *Biophys. J.* **2009**, *97*, L8−L10.

(16) Gō, N. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.

(17) Gonfloni, S.; Weijland, A.; Kretzschmar, J.; Superti-Furga, G. *Nat. Struct. Biol.* **2000**, *7*, 281−286.

(18) Gonfloni, S.; Williams, J. C.; Hattula, K.; Weijland, A.; Wierenga, R. K.; Superti-Furga, G. *EMBO J.* **1997**, *16*, 7261−7271.

(19) Hansmann, U. H. *Chem. Phys. Lett.* **1997**, *281*, 140−150.

(20) Huang, X.; Finerty, P.; Walker, J. R.; Butler-Cole, C.; Vedadi, M.; Schapira, M.; Parker, S. A.; Turk, B. E.; Thompson, D. A.; Dhe-Paganon, S. *J. Struct. Biol.* **2009**, *165*, 88−96.

(21) Huo, S.; Straub, J. E. *J. Chem. Phys.* **1997**, *107*, 5000−5006.

(22) Huse, M.; Kuriyan, J. *Cell* **2002**, *109*, 275−282.

(23) Hyeon, C.; Lorimer, G. H.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18939−18944.

(24) Jeffrey, P. D.; Russo, A. A.; Polyak, K.; Gibbs, E.; Hurwitz, J.; Massagué, J.; Pavletich, N. P. *Nature* **1995**, *376*, 313−320.

(25) Johnson, L. N. *Biochem. Soc. Trans.* **2009**, *37*, 627−641.

(26) Jónsson,H.; Mills, G.; Jacobsen, K.W. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific: Singapore, 1998; p 385.

(27) Jura, N.; Zhang, X.; Endres, N. F.; Seeliger, M. A.; Schindler, T.; Kuriyan, J. *Mol. Cell* **2011**, *42*, 9−22.

(28) Kannan, N.; Neuwald, A. F.; Taylor, S. S. *Biochim. Biophys. Acta* **2008**, *1784*, 27−32.

(29) Karanicolas, J.; Brooks, C. L. *Protein Sci.* **2002**, *11*, 2351−2361.

(30) Karanicolas, J.; Brooks, C. L. *J. Mol. Biol.* **2003**, *334*, 309−325.

(31) Krissinel, E.; Henrick, K. *Acta Crystallogr., Sect. D* **2004**, *60*, 2256−2268.

(32) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(33) LaFevre-Bernt, M.; Sicheri, F.; Pico, A.; Porter, M.; Kuriyan, J.; Miller, W. T. *J. Biol. Chem.* **1998**, *273*, 32129−32134.

(34) Laham, L. E.; Mukhopadhyay, N.; Roberts, T. M. *Oncogene* **2000**, *19*, 3961−3970.

(35) Lee, S. J.; Cobb, M. H.; Goldsmith, E. J. *Protein Sci.* **2009**, *18*, 304−313.

(36) Letunic, I.; Bork, P. *Nucleic Acids Res.* **2011**, *39*, W475−W478.

(37) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912−1934.

(38) Mao, C.; Zhou, M.; Uckun, F. M. *J. Biol. Chem.* **2001**, *276*, 41435−41443.

(39) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807−822.

(40) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 24106.

(41) Marcotte, D. J.; Liu, Y. T.; Arduini, R. M.; Hession, C. A.; Miatkowski, K.; Wildes, C. P.; Cullen, P. F.; Hong, V.; Hopkins, B. T.; Mertsching, E.; Jenkins, T. J.; Romanowski, M. J.; Baker, D. P.; Silvian, L. F. *Protein Sci.* **2010**, *19*, 429−439.

(42) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. *Science* **2004**, *303*, 1800−1805.

(43) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844−11849.

(44) Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7129−7131.

(45) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. *J. Chem. Phys.* **2011**, *134*, 085103.

(46) Ozkirimli, E.; Post, C. B. *Protein Sci.* **2006**, *15*, 1051−1062.

(47) Ozkirimli, E.; Yadav, S. S.; Miller, W. T.; Post, C. B. *Protein Sci.* **2008**, *17*, 1871−1880.

(48) Pan, A. C.; Sezer, D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3432−3440.

(49) Richards, M. W.; O'Regan, L.; Mas-Droux, C.; Blot, J. M. Y.; Cheung, J.; Hoelder, S.; Fry, A. M.; Bayliss, R. *Mol. Cell* **2009**, *36*, 560−570.

(50) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779−815.

(51) Schindler, T.; Sicheri, F.; Pico, A.; Gazit, A.; Levitzki, A.; Kuriyan, J. *Mol. Cell* **1999**, *3*, 639−648.

(52) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291−308.

(53) Schulze-Gahmen, U.; Bondt, H. L. D.; Kim, S. H. *J. Med. Chem.* **1996**, *39*, 4540−4546.

(54) Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11698−11700.

(55) Wernimont, A. K.; Artz, J. D.; Finerty, P.; Lin, Y. H.; Amani, M.; Allali-Hassani, A.; Senisterra, G.; Vedadi, M.; Tempel, W.; Mackenzie, F.; Chau, I.; Lourido, S.; Sibley, L. D.; Hui, R. *Nat. Struct. Mol. Biol.* **2010**, *17*, 596−601.

(56) Xu, W. Q.; Doshi, A.; Lei, M.; Eck, M. J.; Harrison, S. C. *Mol. Cell* **1999**, *3*, 629−638.

(57) Yamaguchi, H.; Hendrickson, W. A. *Nature* **1996**, *384*, 484−489.

(58) Yang, S.; Roux, B. *PLoS Comput. Biol.* **2008**, *4*, e1000047.

(59) Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. *Cell* **2006**, *125*, 1137−1149.

(60) Zhao, R.; Shen, J.; Skeel, R. D. *J. Chem. Theory Comput.* **2010**, *6*, 2411−2423.

(61) Zuckerman, D. M. *J. Phys. Chem. B* **2004**, *108*, 5127−5137.

Supplemental Information

Supplementary Tables and Figures

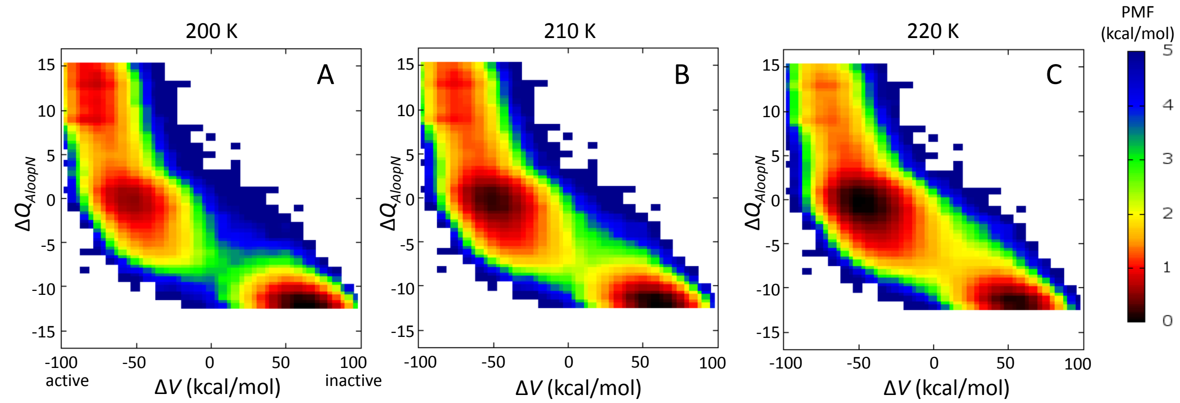


Figure S1: Free energy landscape of Lyn KD along the structural coordinate ΔQ_{AloopN} and the progress variable ΔV at 200K (A), 210K (B) and 220K (C). The intermediate state at $\Delta Q_{AloopN} \approx 0$ and $\Delta V \approx -50$ kcal/mol becomes more favorable as temperature increases, suggesting that this state is associated with greater entropy compared to the other states.

Table S1: List of residue pairs used to define the collective variables for Lyn and CDK2. Each row of the table lists the pairs used to define one collective variable as indicated by the first column. Distances that increase from the active to the inactive state and thus contribute positively to the collective variable are listed directly. Distances that decrease from active to inactive and thus contribute negatively to the collective variable are listed in parenthesis.

	Lyn	CDK2
αC-αC	(A286-E290) (A286-A291) (F287-E290) (F287-A291)	R50-L55 E51-L55 E51-K56
αC-AloopN	(S283-V391) (S283-E393) (A286-V391) (F287-V391) E290-D385 E290-F386 (E290-V391) L293-F386 L293-A389 L293-V391 (M294-L388)	(A48-F152) R50-R150 E51-F146 (E51-A151) (I52-L148) (I52-A151) (I52-F152) L54-A149 L54-R150 L54-A151 L55-F146 (L55-L148)
αC-Nlobe	K275-F287 K275-E290 L277-V284 L277-F287 (V284-E312) (F287-V308) F287-I317 L288-I315 E290-I317 A291-L305 A291-I317 L293-Q298 M294-V303 M294-L305 M294-T319	Y15-T47 Y15-E51 K33-E51 I35-A48 A48-L76 E51-L78 E51-F80 I52-L66 I52-L78 (S53-V69) L55-I63 (K56-V69) L58-I63
αC-Clobe	(E290-Y363) (L293-Y363) (M294-Y363)	L54-V123 E57-R122
AloopN-AloopN	(D385-L388) F386-A389 (G387-R390) (L388-V391) (L388-I392)	(D145-L148) F146-A149 (G147-A151) (G147-F152) (L148-A151) (L148-F152) (L148-G153) (A149-F152) (A149-G153)
AloopN-Nlobe	(F258-I392) (K275-D385) (T281-E393) (M282-V391) (M282-I392) (I317-L388)	(Y15-L148) (Y15-A149) (Y15-F152) (K33-L148) (I35-F152) (I63-F146) (L76-F152) (L78-L148) (F80-D145)
AloopN-Clobe	K361-I392 N362-V391 N362-I392 N362-E393 (Y363-F386) Y363-R390 Y363-V391 Y363-I392 I364-A389 I364-R390 I364-V391 I364-I392 (H365-F386) (H365-G387) H365-R390 R366-L388 (R366-R390) D367-L388 (D367-A389) (I383-F386)	(L115-F146) (C118-F146) H121-F152 R122-A151 R122-F152 R122-G153 (V123-F146) V123-A149 V123-R150 V123-F152 L124-A149 L124-R150 L124-A151 L124-F152 (H125-F146) H125-R150 R126-L148 D127-L148 (L143-F146) (A144-L148) (F146-D185) F152-T182

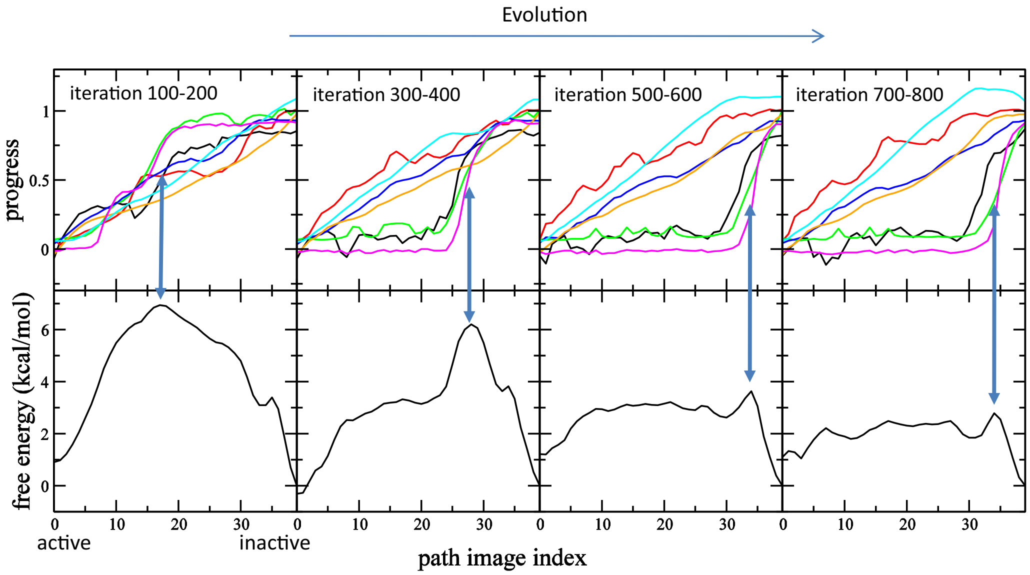


Figure S2: The evolution of the αC-$Aloop$ path of Lyn KD during the MFTP optimization. The profiles of the collective variable (top) and the free energy (bottom) are shown for the path at different stages of optimization. Each column displays the collective variables and the free energy averaged over a 100-iteration period. The blue double-headed arrows show where the αC helix moves and where the free energy peaks along the path.

Supplementary Methods

The single-basin Gō model

We used the Gō potential developed by Karanicolas and Brooks [15,16] as the single-basin model. The model represents each residue of the protein with a single particle at the Cα position and the potential function consists of both bonded and non-bonded terms describing interactions between the particles:

$$V = \sum_{bonds} k_b(b_i - \bar{b}_i)^2 + \sum_{angles} k_\theta(\theta_i - \bar{\theta}_i)^2 + \sum_{dihedrals} \sum_{n=1}^{4} k_{\phi,i}^n (1 - \cos n(\phi_i - \bar{\phi}_i^n)) + \sum_{nbond} V_{ij}^{nbond}, \tag{1}$$

where b_i, θ_i, ϕ_i are the individual bond length, bond angle and dihedral angle, $\bar{b}_i$, $\bar{\theta}_i$, $\bar{\phi}_i^n$ are the corresponding reference values and k_b, k_θ and k_ϕ's are the force constants. The reference values for the dihedral angles in this specific Gō model are defined solely based on the protein sequence and hence have no dependence on the reference structure. Reference values for all other terms are derived from the reference structure. The model defines native contacts between a pair of residues if their side-chain heavy atoms are within 4.5 Å or if they are directly hydrogen-bonded. For those pairs that are in contact, a 12-10-6-order energy term with an attractive well and a dissociation penalty is used:

$$V_{12-10-6}(\epsilon, \sigma, r) = \epsilon \left[13 \left(\frac{\sigma}{r}\right)^{12} - 18 \left(\frac{\sigma}{r}\right)^{10} + 4 \left(\frac{\sigma}{r}\right)^{6} \right], \tag{2}$$

where r is the distance between the two particles and ϵ and σ are parameters determining the strength and characteristic distance of the contact, respectively. For all the other pairs, the non-bonded interaction is described by a simple 12-order repulsive term:

$$V_{12}(\epsilon, \sigma, r) = \epsilon \left(\frac{\sigma}{r}\right)^{12}, \tag{3}$$

where ϵ is the force constant and σ is the repulsive diameter.

Consolidation of two single-basin models

As described in *Methods*, an exponential averaging procedure is used to merge two single-basin Gō potentials into one double-basin potential. For the exponential averaging scheme to work properly, it is necessary to make modifications to the two single-basin Gō models to eliminate factors that may cause an artificially large ΔV. These factors involve all three structure-derived terms. We made modifications to each of them to make the two single-basin models compatible with each other.

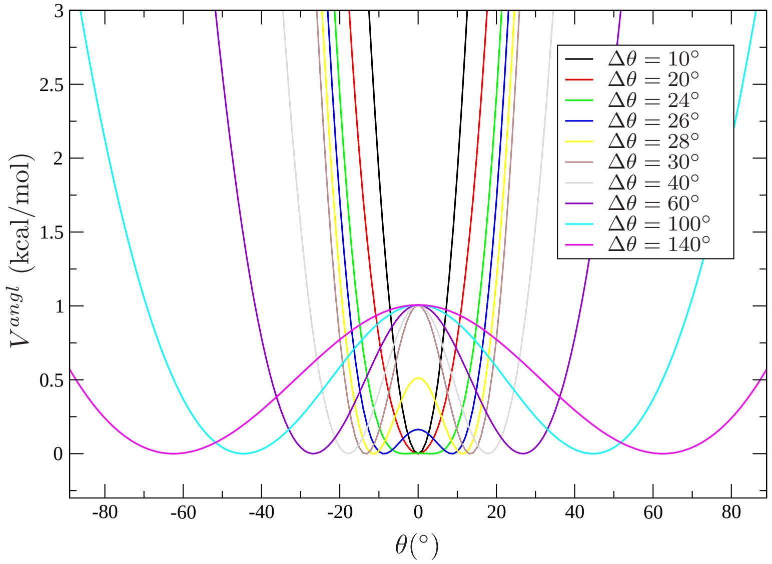


Figure S3: Consolidated angle potential for angles with reference values differing by different degrees in the two single-basin models.

Bonds

The bond terms were consolidated by taking the average of the reference values in the two single-basin models for each bond. This modification does not affect either model significantly because the length of a C_α-C_α pseudo-bond varies little as the configuration changes.

Angles

Different from the bond terms, an angle term may have very different reference values in the two single-basin models. Taking the average does not work very well for the angle terms because it would distort the two reference structures. Instead of averaging, we changed the form of the angle term from a single-basin harmonic well to the following double-basin potential:

$$V^{angl}(\theta) = -\beta_1 \ln(\exp(-\beta_2(k(\theta-\theta_0)^2)) + \exp(-\beta_2(k(\theta-\theta_0')^2))). \qquad (4)$$

where θ_0 and θ_0' are reference values in the two models, $k = 75.6$ kcal/mol/rad^2 is the original harmonic force constant, $\beta_1 = 0.155$ kcal/mol, and $\beta_2 = \beta_1(\frac{\pi/6}{\max(|\theta_0-\theta_0'|,\pi/6)})^2$. The definition of β_2 ensures that any angle with $|\theta_0 - \theta_0'| > \pi/6$ has a constant barrier of 1 kcal/mol between its two minima, whereas for angles with $|\theta_0 - \theta_0'| \leq \pi/6$ the barrier height shrinks as the difference between the two minima decreases.

Non-bonded terms

Both types of terms characterize the excluded volume effect as a steep repulsive wall. Because the original single-basin models use structure derived non-bonded parameters, the same pair of residues may experience the repulsive wall at different distances in the two single-basin models, which would result in a huge energy gap between the two models. In the consolidated models, we modified the non-bonded terms so that the same pair of residues always experience the same repulsive wall. The specific forms of the consolidated non-bonded terms are summarized below. In the summary q is a flag the value of which is true if the pair is in contact and false otherwise. q_{ij}, ϵ_{ij} and σ_{ij} denote corresponding attributes of pair $\{ij\}$ in the present (active/inactive) model before consolidation, whereas q'_{ij}, ϵ'_{ij} and σ'_{ij} denote the same attributes in the other (inactive/active) single-basin model. $\epsilon_0 = 0.18$ kcal/mol and $\sigma_0 = 4.0$ Å.

- **if $\mathbf{q_{ij}}$**:
  - **if $\mathbf{q'_{ij}}$ and $\sigma_{\mathbf{ij}} > \sigma'_{\mathbf{ij}}$**:
    $$V_{ij}^{nbond} = \begin{cases} \min[V_{12-10-6}(\epsilon_{ij}, \sigma_{ij}, r_{ij}), \max(V_{12-10-6}(\epsilon'_{ij}, \sigma'_{ij}, r_{ij}), 0)] & \text{for } r_{ij} < \sigma'_{ij} \\ \min(V_{12-10-6}(\epsilon_{ij}, \sigma_{ij}, r_{ij}), 0) & \text{for } \sigma'_{ij} \le r_{ij} < \sigma_{ij} \\ V_{12-10-6}(\epsilon_{ij}, \sigma_{ij}, r_{ij}) & \text{for } r_{ij} \ge \sigma_{ij} \end{cases}$$
  - **else**: $V_{ij}^{nbond} = V_{12-10-6}(\epsilon_{ij}, \sigma_{ij}, r_{ij})$
- **else**:
  - **if $\mathbf{q'_{ij}}$**: $V_{ij}^{nbond} = \begin{cases} \max(V_{12-10-6}(\epsilon'_{ij}, \sigma'_{ij}, r_{ij}), 0) & \text{for } r_{ij} < \sigma'_{ij} \\ 0 & \text{for } r_{ij} \ge \sigma'_{ij} \end{cases}$
  - **else**: $V_{ij}^{nbond} = V_{12}(\epsilon_0, \sigma_0, r_{ij})$