JCTC Journal of Chemical Theory and Computation

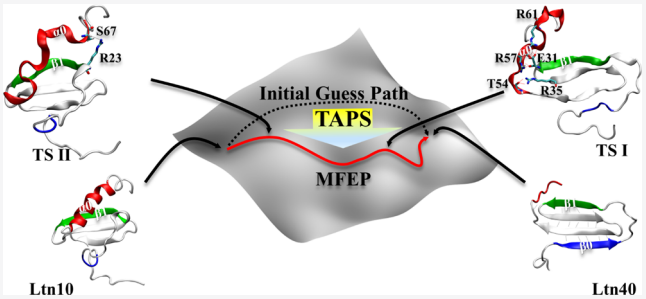
pubs.acs.org/JCTC Article

Assessing the Performance of Traveling-salesman based Automated Path Searching (TAPS) on Complex Biomolecular Systems

Kun Xi, Zhenquan Hu, Qiang Wu, Meihan Wei, Runtong Qian, and Lizhe Zhu*

Cite This: J. Chem. Theory Comput. 2021, 17, 5301−5311 Read Online

ABSTRACT: Though crucial for understanding the function of large biomolecular systems, locating the minimum free energy paths (MFEPs) between their key conformational states is far from trivial due to their high-dimensional nature. Most existing path-searching methods require a static collective variable space as input, encoding intuition or prior knowledge of the transition mechanism. Such information is, however, hardly available *a priori* and expensive to validate. To alleviate this issue, we have previously introduced a Traveling-salesman based Automated Path Searching method (TAPS) and demonstrated its efficiency on simple peptide systems. Having implemented a parallel version of this method, here we assess the performance of TAPS on three realistic systems (tens to hundreds of residues) in explicit solvents. We show that TAPS successfully located the MFEP for the ground/excited state transition of the T4 lysozyme L99A variant, consistent with previous findings. TAPS also helped identifying the important role of the two polar contacts in directing the loop-in/loop-out transition of the mitogen-activated protein kinase kinase (MEK1), which explained previous mutant experiments. Remarkably, at a minimal cost of 126 ns sampling, TAPS revealed that the Ltn40/Ltn10 transition of lymphotactin needs no complete unfolding/refolding of its β-sheets and that five polar contacts are sufficient to stabilize the various partially unfolded intermediates along the MFEP. These results present TAPS as a general and promising tool for studying the functional dynamics of complex biomolecular systems.



1. INTRODUCTION

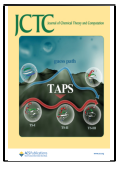
As one of the basic tasks of biomolecular simulations, revealing the transition pathways among the stable conformational states is essential for understanding the functional dynamics of biomolecular systems.[1] When sufficient amount of sampling is available, the pathways can be derived by visual inspection or from kinetic network models, such as the Markov State Models.[2−10] However, due to the ubiquitous high barriers along the paths and the numerous coordinates involved, the sampling process itself remains far from trivial.

One class of techniques used to tackle this challenge is path sampling, which aims to obtain an ensemble of transition paths directly defined in the phase space and therefore requires no prior knowledge about the coordinates involved in the transition. For example, as the most established method in this class, transition path sampling (TPS) and its derivatives[11−20] conduct a Monte Carlo sampling scheme in the path space, on top of unbiased molecular dynamics simulations, to obtain this path ensemble and then compute the corresponding reaction rate. The degrees of freedom most relevant to the transitions are identified from the path ensemble through post-analysis.[17−19]

When the kinetics is not of central interest, one may instead apply path-searching methods[21−28] that use biased sampling to locate the minimum free energy path (MFEP) closest to a guess path generated by various enhanced sampling techniques.[29−36] Path-searching methods are typically less resource-demanding than path sampling and can in fact be used as an initial sampling technique for constructing Markov State Models.[37−41] However, the overall efficiency of most path-searching methods is restricted by two factors: (a) the path is defined in a low-dimensional collective variable (CV) space chosen *a priori*, heavily relying on prior knowledge about the system, which is difficult to obtain, particularly for large systems; (b) the sampling around the path node is typically local, leading to slow evolution of the path toward the MFEP.

To alleviate these issues, we have recently developed Traveling-salesman based Automated Path Searching (TAPS), a highly parallelizable and efficient path-searching method that adopts nonlocal sampling for a rapid search perpendicular to the reference path and uses a traveling-salesman scheme to reorder the newly generated path nodes.

Received: February 19, 2021
Published: July 16, 2021



https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

Importantly, TAPS avoids the definition of the preselected static CV space but is based on a distance metric (a way to measure the distance between any pair of high-dimensional images of the system) used to compute the path collective variable (PCV)[42] and therefore minimizes the prior knowledge required for the search. Such improvements have led to a 5−8 times speedup of TAPS over the string method with swarms of trajectories (SMwSTs) for peptide systems in vacuum and explicit solvent.[43]

To assess the performance of TAPS on complex biomolecules with hundreds of residues or more, we implemented a fully parallel version of TAPS in Python 3.7 with GPU support and standardized the tools for convergence check via the multidimensional scaling method (MDS).[44] This new implementation was applied to three representative systems with explicit solvent: the L99A variant of T4 lysozyme (T4L L99A, ∼30 000 atoms), the mitogen-activated protein kinase kinase (MEK1, ∼50 000 atoms), and the human chemokine lymphotactin (Ltn, ∼80 000 atoms). The corresponding intermediates and transition states (IS/TS) were identified from the free energy surface along the located MFEPs obtained via umbrella sampling.[45]

For T4L L99A, we elucidated the mechanism of the interconversion between its ground (G) and excited (E) states that dictates the hydrolysis activity in peptidoglycan.[46−48] For MEK1, TAPS identified the MFEP for its loop-in/loop-out transition (an important component in the activation of mitogen-activated protein kinase[49]) and revealed the role of the polar contacts R227:L235 and Y229:E255 in this transition.

For Ltn, a critical chemokine, we studied the transition between its two distinct functional folds Ltn10 (a functional GPCR XCR1 agonist, but fails to bind heparin[50,51]) and Ltn40 (agonist for glycosaminoglycans, yet cannot activate XCR1[52,53]). Remarkably, TAPS revealed unprecedented details about this transition. Contrary to a previous hypothesis that complete unfolding/refolding of the β−sheet must occur, our MFEPs pointed to a three-step process involving only partial unfolding. The partially unfolded structures are stabilized by some of the five polar contacts E31:R57, E31:R61,R35:T54, T16:R43 and R23:S67 at different stages of the transition.

Collectively, these results well present TAPS as a promising approach for studying the functional dynamics of complex biomolecular systems.

2. METHODS

To enable the application of TAPS to realistic biomolecular systems, we implemented an MPI version of TAPS in Python 3.7 with GPU support, based on the code published previously.[43] This new implementation follows the protocol shown in Figure 1, which mainly comprises four stages: (a) initial path generation; (b) path optimization; (c) convergence check and (d) free energy calculations. In this section, we describe in detail the motivation for each step and provide the general guidance for setting the corresponding input parameters. Special attention is given to the newly added module for convergence check.

2.1. Generation of the Initial Path. An initial path must be generated before the optimization by TAPS, typically via coarse-graining[54] or various biased sampling methods,[29−31,34−36] e.g., steered MD.[33] The overall efficiency for this step remains an open question and is worth further exploration. For the three systems in this work, we chose

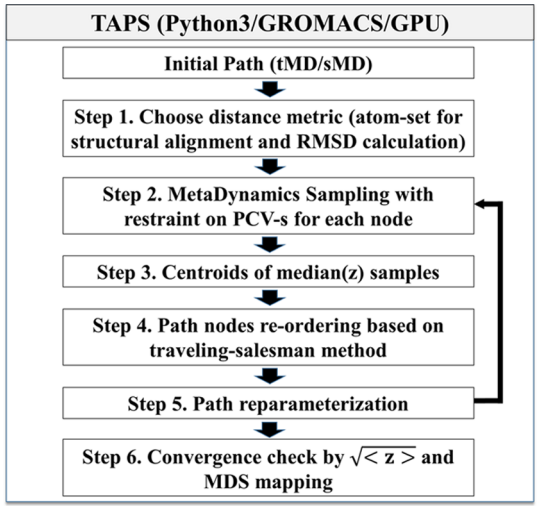


Figure 1. Flowchart of the Traveling-salesman based Automated Path Searching (TAPS). The initial/guess path generated by tMD,[32] sMD,[33] or other enhanced sampling methods.[34−36] Well-tempered MetaD simulation for each node is performed parallelly on the "progress component" s of the path collective variable (PCV)[55−57] with a restraint potential on s. The backbone nodes of the new path correspond to the centroids of the sampled conformations of each node with the median value of the "deviation component" z of PCV. Re-ordering of these new nodes with a traveling-salesman scheme. Reparameterization is finally performed to maintain the resolution of the new path (detailed in ref 42). Convergence is checked by $\sqrt{\langle z \rangle}$ and the visualization of the path nodes on a low-dimensional space is obtained by multidimensional scaling (MDS).[44]

targeted MD (tMD)[32] for the initial path generation, which drives the system toward a target structure via the root mean square distance (RMSD) from it. For simplicity, all C_α atoms for residues with well-defined secondary structures (α-helix/β-sheet) were used for structural alignment, and all heavy atoms were used to calculate the RMSD. Note that these two atom sets, which determine a distance metric only for the initial path generation, need not be the same as the ones in subsequent TAPS optimization.

2.2. Path Optimization by TAPS. As described previously,[43] the initial path is optimized toward its closest MFEP via the following iterative process (steps 2−5 in Figure 1): (a) choose a distance metric (two atom sets for computing RMSD), based on which the path collective variable (PCV)[42,55] is defined; (b) perform well-tempered PCV-metadynamics[56] (PCV-MetaD) simulations with a restraint potential on the PCV-s component for each path node; (c) extract new path nodes from the PCV-MetaD simulation results; (d) reorder the path nodes via a Traveling-salesman scheme; and (e) re-parameterize the path to ensure the straightness of the path and the proximity of the neighbor nodes.

2.2.1. Choosing a Distance Metric. Most existing path-searching methods rely on a static CV space defined *a priori*. This requirement, however, contradicts the most common situation where they are used—sampling a complex system without sufficient prior knowledge. The lack of prior knowledge necessitates a tedious trial−error process of finding a proper set of CVs,[57] limiting their overall efficiency. In TAPS, we attempt to avoid this static CV space, but use the PCV[55,56] as a temporary coordinate system in each iteration, which

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

projects the original high-dimensional system onto the high-dimensional path obtained in one previous iteration.

To define the PCV, a distance metric $d_{i,j}$ discriminating any two high-dimensional conformations i and j (typically RMSD) must be specified, which is the only major assumption of the TAPS method (step 1). Since one only needs to specify two atom sets to define $d_{i,j}$ (for structural alignment and RMSD computation respectively), the overall information required to perform the optimization is less than those relying on CVs typically with physical meaning. Note that improper choice of the metric may result in incomplete description of the transition mechanism (missing important atoms) or problematic potential of mean force along the apparently converged MFEP in the subsequent free energy calculations. To alleviate this issue, the current TAPS implementation generates paths whose nodes are all full system images, allowing adjustments of $d_{i,j}$ in any iteration.

2.2.2. Nonlocal Perpendicular Sampling. With $d_{i,j}$, one can compute the two-component PCV[55,56]

$$s = \frac{\sum_{i=1}^{N} i e^{-\lambda d_{x,i}^2}}{\sum_{i=1}^{N} e^{-\lambda d_{x,i}^2}} z = -\frac{1}{\lambda} \ln\left(\sum_{i=0}^{N} e^{-\lambda d_{x,i}^2} \right) \quad (1)$$

where $d_{x,i}$ is the distance between any high-dimensional conformation $\boldsymbol{x}$ and the ith node of the path α containing N nodes. λ is a constant, typically computed by $\lambda = 2.3(N-1)/\sum_{i=1}^{N-1} d_{i,i+1}^2$, where $d_{i,i+1}$ is the distance between the ith node and its neighbor node $i+1$. Here s is a unit-less quantity, meaning the location of such projection on path α. z is expressed in the unit of distance square, meaning the average squared distance of this conformation from path α.

As stated in ref 43, paths other than the MFEP are unstable in directions orthogonal to themselves. This fact means that the sampling for each node can be performed within the perpendicular plane $P_{ij}(\alpha)$ of path α to accelerate the search for segments of its closest MFEP (step 2). We approximate this nonlocal perpendicular sampling by setting a restraint potential on PCV-s. Since PCV-s is a unit-less quantity, the restraint potential is easy to design and needs no modifications even when the distance metric $d_{i,j}$ is changed. Thanks to the newly introduced MPI implementation, such sampling can be performed for all path nodes in parallel.

In addition to the restraint potential, during the perpendicular sampling on PCV-z, we also added the MetaD bias potential on PCV-s. This is because reparameterization of the path at a later stage requires a pool of conformations containing a sufficient number of different structures. Introducing the MetaD bias reduces the amount of extra sampling required (see the following section).

2.2.3. Path Node Extraction, Re-ordering and Path Reparameterization. Subsequently, the centroids of the conformations with the median value of the "deviation component" z of PCV can be extracted as the backbone nodes of the new path (step 3). Because each path node is allowed to move arbitrarily far from its initial position along its own preferred directions in the perpendicular hyperplane of $N-1$ dimensions, the nodes are generally unlikely to maintain their original order in the reference path. Thus, TAPS uses a Traveling-salesman scheme to regenerate an order based on the geometric distances among the new nodes (step 4).

Due to the tendency of the sampling to drop into the known free energy minima, the nodes of the new path are often unevenly distributed, necessitating the final reparameterization of the new path. This process includes three simple steps: terminal truncation, node insertion, and path straightening (step 5). Here, two parameters are essential for defining the geometrical resolution of the path: (a) the largest distance allowed for neighbor nodes δ and (b) the straightening factor σ (details in ref 43). While σ can easily be set to a number from 1.1 to 1.5, the choice of δ needs more careful thought. When $d_{i,i+1} > \delta$ for any two neighbor nodes, we insert a conformation in-between them, which is chosen from the samples of perpendicular sampling. When no such conformation is available, extra targeted MD is performed. Clearly, smaller δ yields a larger number of nodes, more detailed description of the transition mechanism, yet higher computational costs. It is up to the users to specify such balance between efficiency and path resolution.

After the reparameterization, the iteration of path searching starts again until the path converges to (fluctuates around) the closest MFEP, where convergence is measured as described below.

2.3. Convergence Check: $\sqrt{\langle z \rangle}$ and MDS analysis. Convergence of the optimization is checked in two manners: (a) the distance between the path of current iteration and the initial guess path, as measured by $\sqrt{\langle z \rangle}$—square root of the z component of PCV averaged for all current path nodes[53−55] and (b) the projection of the path nodes obtained in all iterations on a low-dimensional space generated by multi-dimensional scaling (MDS).[44]

A simple measure of the distance of the new path α from the reference path is the square root of average PCV-z for all nodes

$$\sqrt{\langle z \rangle} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} z_{i,\text{ref.path}}} \quad (2)$$

where $z_{i,\text{ref path}}$ denotes the PCV-z of node i of path α with respect to the reference path. The convergence for the optimization process can be checked according to the $\sqrt{\langle z \rangle}$ of the path in each iteration with respect to the initial path. When converged, further TAPS iterations shall not alter the relative position of the path nodes significantly ($\sqrt{\langle z \rangle}$ fluctuating mildly).

Besides, with the distances among all nodes from all paths available, one can map all the paths from a high-dimensional space to a low-dimensional (typically two-dimensional (2D)) space via MDS.[44] In this low-dimensional space, each point represents a node with the distance between points nearly identical to the RMSD between the corresponding high-dimensional nodes. With this projection, the optimization process is conveniently visualized.

2.4. Free energy calculations. After convergence, one may use various approaches, e.g., MetaD and umbrella sampling, to evaluate the free energy surface along the PCV-s of the found MFEP. The transition states and intermediates can be further identified from the free energy surface accordingly. Here we used umbrella sampling for all three systems because of its parallelizability. During the umbrella sampling on PCV-s, wall potentials were put on PCV-z to ensure a local sampling surrounding the found MFEP.

3. SIMULATION DETAILS

3.1. TAPS. All TAPS simulations were performed via GROMACS-2019.4,[58] PLUMED-2.6.0,[59] and Concorde,[60]

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

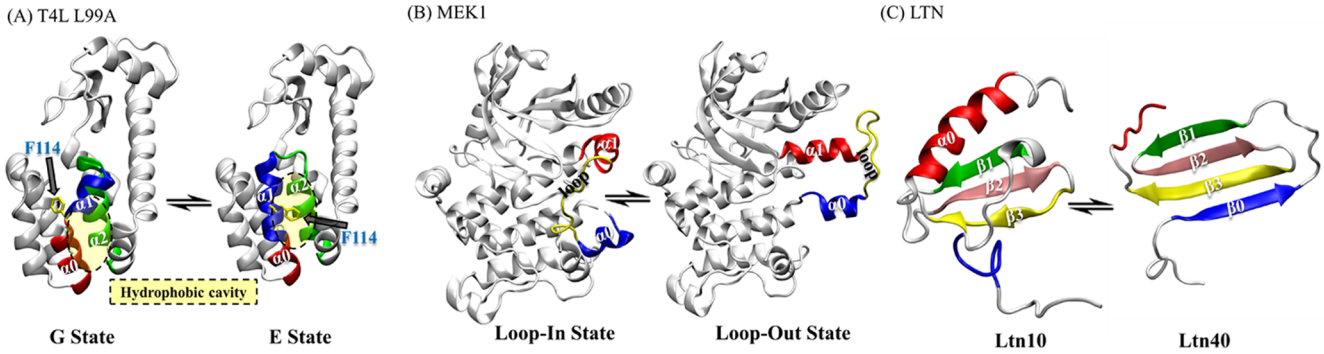


Figure 2. Structures of (A) L99A variant of the T4 Lysozyme (T4L L99A),[46−48] (B) MEK1,[49] and (C) lymphotactin (Ltn).[50,51] (A) Interconversion between the G state[46] and E state[47] of T4L L99A accompanied by a shift of residue F114 (yellow) to the center of the hydrophobic cavity, surrounded by three α helices: α0 (red), α1 (blue), and α2 (green). (B) Interconversion between the loop-out state and loop-in state for the activation segment of MEK1; the two α-helices (α0: blue, α1: red) and loop (yellow) are highlighted. (C) Interconversion between Ltn40[50] and Ltn10;[51] the disordered C-termini (residues 61−76) are omitted for clarity;[51] the four important β-sheets (β0: blue, β1: green, β2: pink, β3: yellow) and α-helix (α0: red) are highlighted.

wrapped within a python script publicly available at https://github.com/xikun2020/TAPS. All simulations were performed in the NVT ensemble at T = 300 K, using the velocity-rescale thermostat.[61]

3.2. System Studied. *3.2.1. L99A variant of the T4 Lysozyme.* T4 lysozyme (T4L) is an enzyme that catalyzes the hydrolysis of peptidoglycan, which is the major component of the gram-positive bacterial cell wall.[46−48,62−67] Such catalysis, however, is only possible if T4L is in an excited state (E), typically activated from a ground state (G). Thus, elucidating the mechanism of this G/E transition is essential for fine-tuning the activation and therefore catalysis of T4L. So far, the L99A variant of T4L (T4L L99A) and the structures of G state (PDB ID: 3DMV) and E state (PDB ID: 2LCB) have been resolved[46,47] (Figure 2A). The G/E state interconversion of T4L L99A only involves the residue F114 turning into the hydrophobic cavity (yellow in Figure 2A) and local rearrangement of three α helices: α0 (red), α1 (blue), and α2 (green). Details can be found in Figure S1A.

3.2.2. Mitogen-Activated Protein Kinase 1. The mitogen-activated protein kinase pathway (MAPK) has important roles in tumor activity. As one of the most critical components in this pathway, activation of the MAP kinase 1 (MEK1) warrants exploration. During activation, the loop of the activation segment (residue 211−236) transforms from the loop-out state (PDB ID: 5YT3) to the loop-in state (PDB ID: 3QED)[49] (Figures 2B and S1B). The main features of this transition are the transformation of the loop (yellow) and the rotation of two α helices: α0 (red) and α1 (blue).

3.2.3. Lymphotactin. Lymphotactin (Ltn) is a human chemokine that directs the migration of leukocytes toward areas of inflammation. Recent studies have revealed the folds of Ltn:[50] a chemokine fold preferred at 10 °C (Ltn10, PDB ID: 1J9O)[50] and a four-stranded antiparallel β-sheet preferred at 40 °C (Ltn40, PDB ID: 2JP1)[51] (Figure 2C). Here, Ltn10 is a functional XCR1 agonist, but fails to bind heparin. In contrast, Ltn40 can bind glycosaminoglycans with high affinity, yet is unable to activate XCR1.[52,53] Thus, dissecting this Ltn40/Ltn10 transition is crucial for understanding the function of Ltn and other temperature-sensitive chemokines in general. As this transition must involve unfolding (loss of a β-strand β0) and refolding, (formation of an α-helix α0)[50,51] (Figure S1C), it shall pose a proper challenge to TAPS.

3.3. System Preparation. T4L L99A, MEK1, and Ltn were solvated in a dodecahedron box with 10331, 14227, and 25714 TIP3P waters and 10 Na^+/18 Cl^-, 29 Na^+/27 Cl^-, and 8 Cl^- ions,[68] respectively (details in Table S1). The CHARMM36 forcefield[69] was used to describe the interactions in the three systems. Energy minimization was performed for 50 000 steps by the steepest descent algorithm and then by a conjugate gradient algorithm. Then, a 100 ps NVT simulation was performed at 300 K for solvent equilibration, followed by a 2 ns NpT equilibration to 1 atm using the Berendsen barostat.[70] Long-range electrostatic interactions were treated by the particle-mesh Ewald method.[71,72] The short-range electrostatic and van der Waals interactions both used a cutoff of 10 Å. All bonds were constrained by the LINCS algorithm.[73]

3.4. Generation of the Initial Path and TAPS Optimization. For the three systems, the initial paths were generated by tMD.[32] tMD was performed for 400 ps with the frame recorded every 0.2 ps. In each iteration of TAPS, the perpendicular sampling of all path nodes amounted to 1 ns in total (300 ps for T4L L99A). The cutoff distance δ between neighbor nodes was set as 0.8 Å (T4L L99A), 1.0 Å (MEK1), and 1.6 Å (Ltn). The straightening factor was chosen as σ = 1.25 for all three systems. Gaussians of height 0.25 kJ/mol and width 0.5 were deposited every 0.1 ps, with the frames recorded at the same frequency. The complete parameters used in TAPS are listed in Table S2.

3.5. Umbrella Sampling. After the final path got converged, the free energy landscape was demonstrated by umbrella sampling.[45] To escape the possible non-overlapping between neighbor nodes, extra nodes with an average gap of 0.25 were added and finally, 88, 124, and 1340 nodes were obtained for T4L L99A, MEK1, and Ltn, respectively (see the details in Table S3). For the three systems, 40 ns, 40 ns and 2 ns of umbrella sampling was performed for each node of T4L L99A, MEK1, and Ltn, respectively (see the sampling result in Figures S2−S4). Finally, the free energy landscapes of the three systems were revealed by WHAM.

4. RESULTS

4.1. Choice of Distance Metric. The choice of the atom sets for structural alignment and RMSD calculation is vital for TAPS. Structural alignment can be performed using the rigid parts of the molecule without significant conformational changes. Yet, note that a mere comparison between the two

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

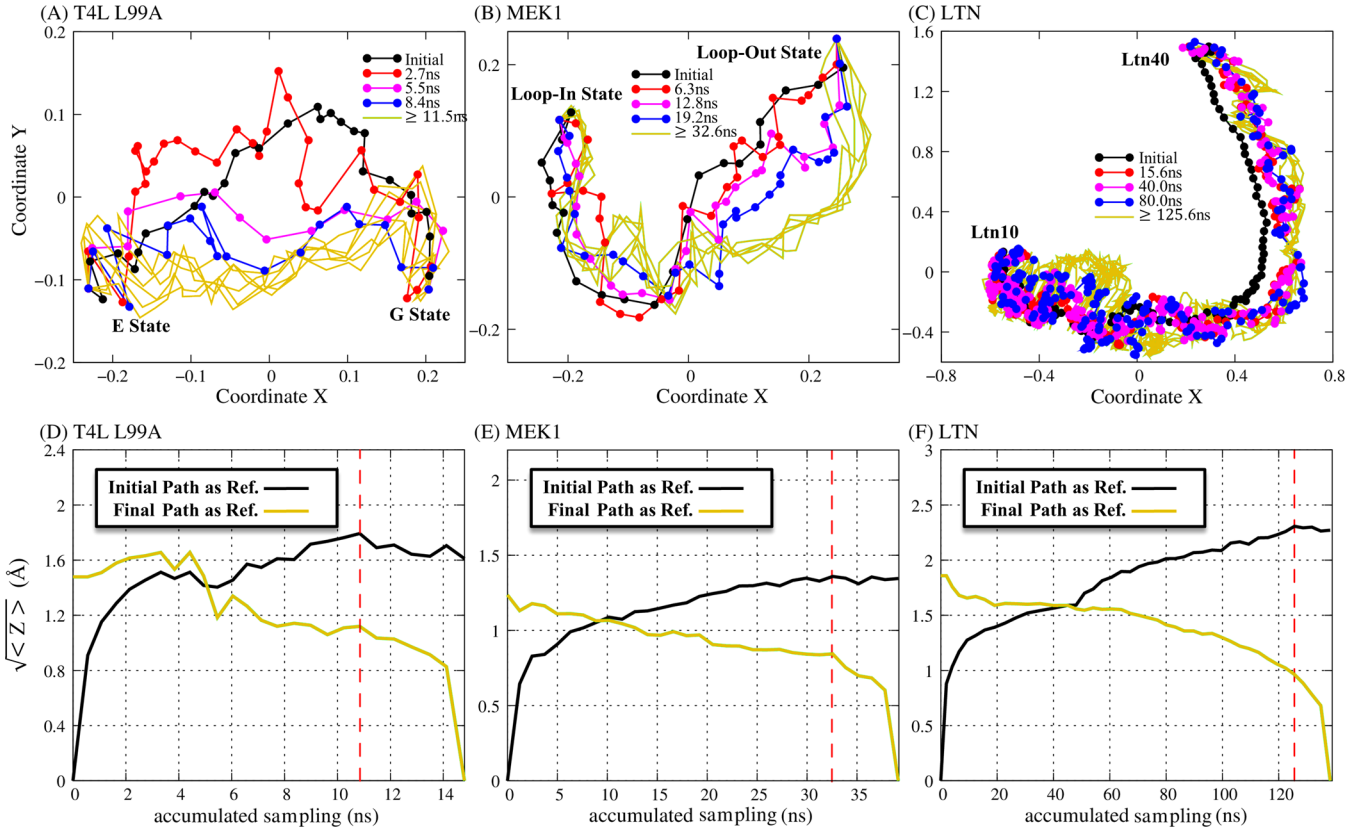


Figure 3. Convergence check. The paths at different TAPS iterations for T4L L99A (A) MEK1 (B) and Ltn (C) are mapped to a two-dimensional space obtained by multidimensional scaling (MDS);[44] The accumulated sampling time for each iteration (perpendicular sampling and targeted MD) is highlighted by a different color. Convergence of the optimization processes for T4L L99A (D), MEK1 (E), and Ltn (F) is measured by the progress of $\sqrt{\langle z \rangle}$ along the accumulated sampling time. Here, the black and orange lines denote two different reference paths: initial path (black) and final path (orange), respectively. Convergence times are highlighted by red dashed lines.

known stable states is not always sufficient to define this atom set. The path revealed by TAPS might indicate additional mobile parts that move back and forth during the transition, as shown by our example of Ltn. In this case, one may need to redefine this atom set to obtain a reasonable potential of mean force along the final MFEP. For RMSD calculations, if important atoms are excluded from the metric, one may observe unphysical high barriers in the final free energy surface obtained. Redundancy in this atom set is less of a problem, yet adds more burden to the PCV calculations.

Here for T4L L99A, we chose all Cα atoms of the residues from α-helices and β-sheets for structural alignment, only excluding residues 92−124. As residues 92−124 are sufficient to describe the G/E transition, we used all heavy atoms of residues 92−124 for RMSD calculations. For MEK1, a similar strategy was used: Cα atoms of α-helices and β-sheets were used for structural alignment except the activation segment (residues 211−236); all heavy atoms of this segment were used for RMSD calculation.

For Ltn, even though all three β-sheets $\beta1/\beta2/\beta3$ barely differ between the known stable states Ltn10 and Ltn40, we chose only $\beta2/\beta3$ for the alignment, since we observed that $\beta1$ tilted mildly with respect to $\beta2/\beta3$ during the transition. The position of the intermediate state would be distorted if all three β-sheets were used for structural alignment. In addition, as this Ltn10/Ltn40 transition involves complicated folding/refolding, all heavy atoms of the molecule were used for RMSD calculation.

4.2. Convergence Check. As we mentioned before, the convergence of TAPS can be conveniently analyzed by MDS[44] and $\sqrt{\langle z \rangle}$ (Figure 3). For all three systems, representative paths have been chosen and mapped into a two-dimensional space to characterize the optimization process (Figure 3A−C). It is obvious that a big difference appeared between the initial path (black color) and the path only after five iterations (red) and the optimization converged to the final path (orange) within 20−30 iterations. In fact, the path (blue) started to fluctuate around the final MFEP only within 10−15 iterations for all three systems, demonstrating the efficiency of TAPS.

As shown in Figure 3D−F, $\sqrt{\langle z \rangle}$ was also calculated using both the initial path (black), labeled with $\sqrt{\langle z_i \rangle}$, and the final MFEP (last iteration, orange, labeled $\sqrt{\langle z_f \rangle}$) as the reference path. Undoubtedly, similar trends of convergence were observed for the three systems: $\sqrt{\langle z_i \rangle}$ always quickly increased ($\sqrt{\langle z_f \rangle}$ quickly decreased) and reached the maximum at the accumulated sampling times (perpendicular and tMD sampling) of ∼11.5, ∼32.6, and ∼125.6 ns for T4L L99A, MEK1, and Ltn, respectively. The results show that the time needed for optimization was influenced by the complexity of the transition. Remarkably, for the complicated Ltn10/40 transition involving unfolding−refolding, TAPS is able to converge within only 40 ns.

4.3. Resolution and Straightness of the Final Path. After convergence, we performed umbrella sampling[45] to

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

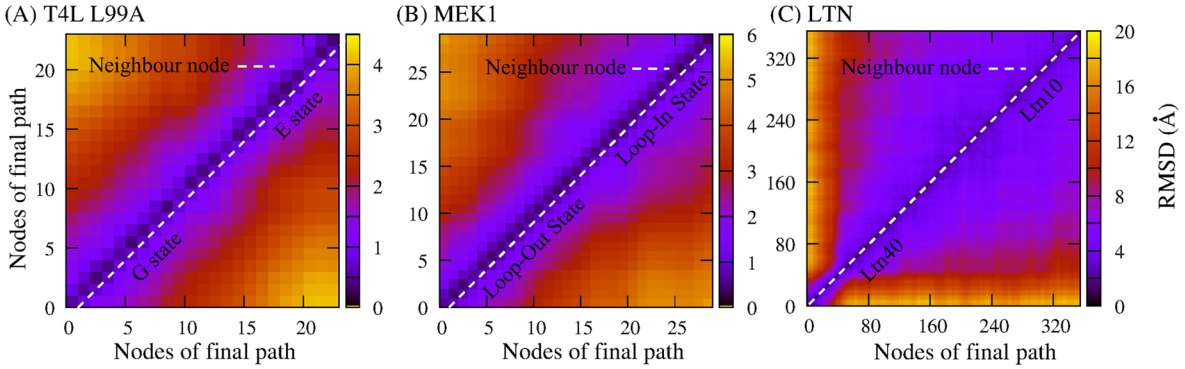


Figure 4. Pairwise RMSD matrix for the nodes on the found minimum free energy path for T4L L99A (A), MEK1 (B), and Ltn (C). RMSD values for neighbor nodes are highlighted by white dashed lines.

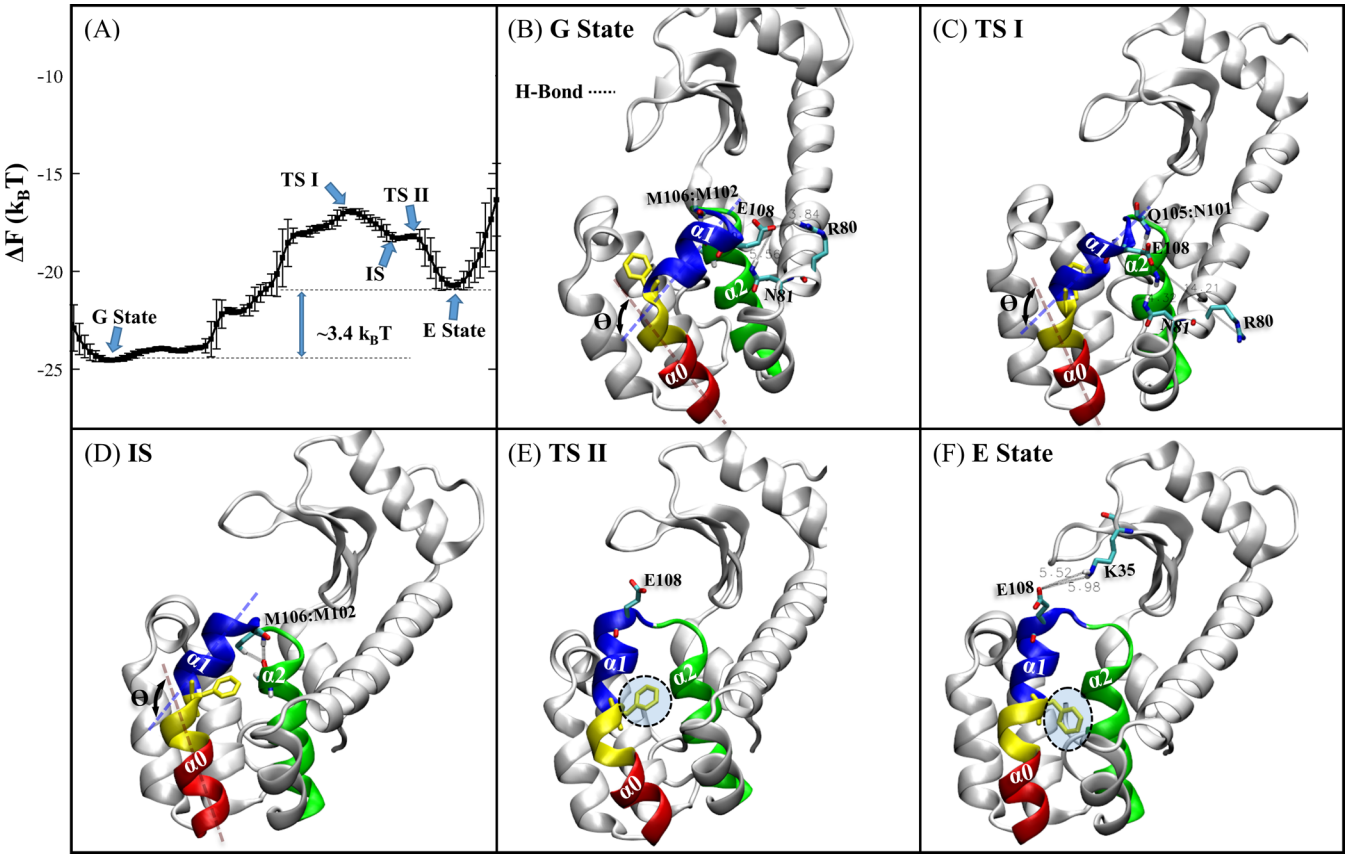


Figure 5. (A) Free energy landscapes along the final path for T4L L99A calculated by umbrella sampling.[45] The important intermediate state (IS) and transition state (TS) are defined based on the free energy landscape. (B–F) Structures of G state/E state and IS/TSs are displayed by VMD[74] with highlighted α0(red)/α1(blue)/α2(green), H-bonds, and polar contacts. The residues 115–118 that formed α0 in G state and α1 in E state are highlighted in yellow. The angle θ between the central axis of α0 helix (red dashed line) and α1 helix (blue dashed line) is labeled (B–D), as well as the key residue F114 (yellow).

obtain the free energy surface along the PCV-*s* of the final path and locate thereafter the intermediate and transition states. However, proper definition of PCV requires the nodes to be evenly distributed (equal distance between neighbor nodes; the path is straight enough without loops of high curvature). To check this, we computed the pairwise RMSD matrix for all nodes of the final path and highlighted the neighbor nodes list by the white dashed lines in Figure 4. The neighbor node distances $d_{i,i+1}$ were satisfactory and close to the chosen cutoff distance δ (0.8–1.6 Å). Straightness was also optimal for all systems, as the $d_{i,i+j}$ values for j = 2–5 were 3–5 Å, significantly larger than the neighbor distance $d_{i,i+1}$.

4.4. Revealed Transition Mechanism. The free energy landscapes along the PCV-*s* of the final MFEP were obtained by umbrella sampling (details in Table S3, SI Figures S2–S4). We then identified the state (IS) and the transition state (TS) (Figures 5–7). Important polar contacts/H-bonds during the transition are particularly highlighted (details in SI Figure S5 and Table S2). All the three-dimensional (3D) structures were displayed by VMD.[74]

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301–5311

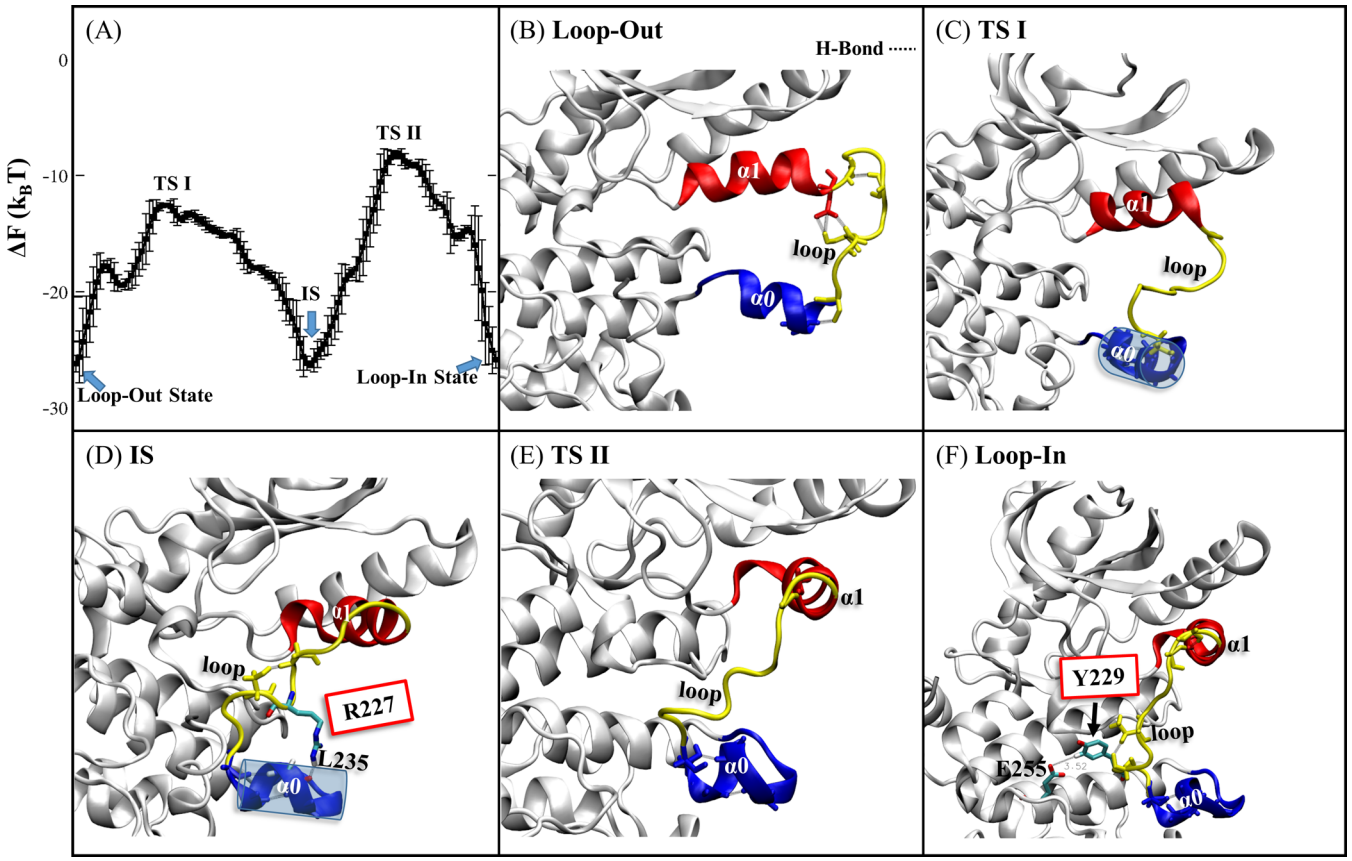


Figure 6. (A) Free energy landscapes along the final path of MEK1 revealed by umbrella sampling.[45] The important intermediate state (IS) and transition state (TS) are defined based on the free energy landscape. (B–F) Structures of loop-out state/loop-in state and ISs/TSs are displayed by VMD[74] combined with the highlighted α0(blue)/α1(red)/loop(yellow), H-bonds, and polar contacts.

4.4.1. T4L L99A. Comparison between the G and E states of T4L L99A reveals that the G- > E transition must involve the rotation of residue F114 into the hydrophobic cavity, the merging of residues 115–118 (yellow in Figure 5) into the helix α1 (blue in Figure 5), which can be measured by the increase of angle θ between the central axis of α0 and α1 (definition in Figure S6). As shown by the free energy surface in Figure 5A, the free energy difference between the G and E states is ~3.4 k_BT, consistent with the experimental measurements, where the fractional populations of G and E states were ~97 and ~3%, respectively.[47]

This transition occurred in two steps. Starting from the G state, the system must climb an ~8 k_BT high barrier to reach TS I, where the polar contact E108:N81 and the H-bond M106:M102 stabilizing the loop connecting α1 and α2 are broken. The loosened loop enables α1 to rotate around its central axis with F114 approaching the hydrophobic cavity as θ is increased from ~101 to ~113° accordingly. To compensate such energy loss, the polar contact E108:R80 and a newly formed H-bond Q105:N101 are maintained in this first transition state (Figure 5B,C). T4L L99A reaches a barely stable intermediate IS (Figure 5D), whereas θ increases further to ~134° with the M106–M102 H-bond being reformed and F114 moving to the entrance of the hydrophobic cavity. The second barrier is fairly easy to overcome (~1.5 k_BT). TS II is featured by re-breaking of the M106:M102 H-bond and the formation of a longer α1 (merging the residues 115–118) and further movement of F114 to the edge of the hydrophobic cavity (Figure 5D,E), which results in the suspension of the elongated α1 in solution without contact with other domains. Finally, the elongated α1 finds a polar contact with K35 (E108:K35, Figure 5E,F), which is described in ref 48, with F114 rotated 180° around the bond between C_α and C_β and fully buried in the cavity.

4.4.2. MEK1. For the loop-out/loop-in transition of MEK1, two steps were found, with the second step being rate-limiting (Figure 6A). For clarity, we only visualized the residues close to the active segment in Figure 6B–F (details in Figure S7). Starting from the loop-out state, where α0 (blue) is facing outward with α1 (red) parallel to it,[49] the system has to first reach TS I with a mild inward rotation of α0 (Figure 6B,C).[75]

To overcome this ~14 k_BT barrier, several H-bonds stabilizing the loop between α0 and α1 (yellow) had to be broken. Then, the system goes downhill to an intermediate state IS with similar stability to the loop-out state with a newly formed polar contact R227:L235 to stabilize the loop and further inward rotation of α0 (Figure 6D). This polar contact is broken again in subsequent climbing over the second ~18 k_BT barrier (Figure 6E), enabling a loosen loop segment and therefore the rotation of α1 to become orthogonal to α0. Finally, MEK1 goes downhill to the loop-in state with newly formed H-bonds within the loop and an additional polar contact Y229(**loop**):E255 (Figure 6F), where E255 belongs to the core kinase domain with a critical role in the function of the enzyme (Figure S1C). Note that the stability of α1 in the loop-in state is partially provided by the stretching force from the stabilized loop. Interestingly, our findings on the two residues R227 and L235 provide an insightful interpretation for

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301–5311

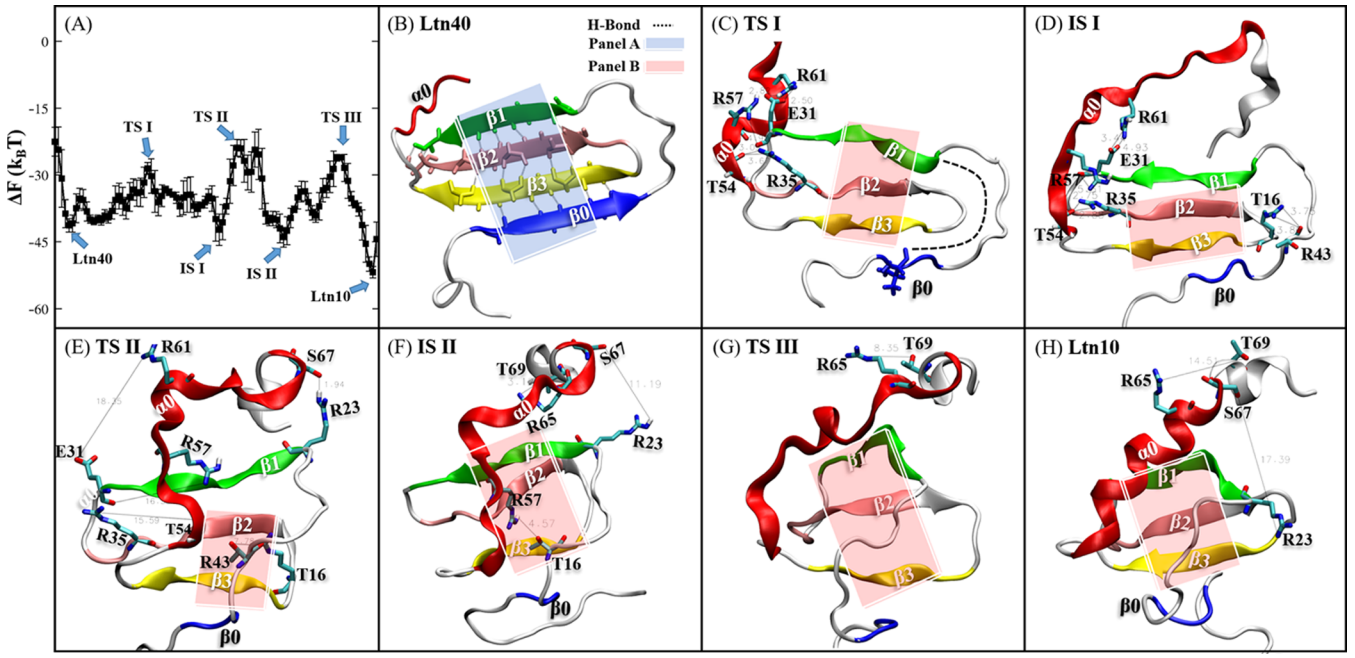


Figure 7. (A) Free energy landscape along the minimum free energy path for lymphotactin (Ltn), calculated by umbrella sampling.[45] The important intermediate state (IS) and transition state (TS) are labeled. (B−H) Structures of Ltn40/Ltn10 and IS/TS are displayed by VMD.[74] Key structural features β0(blue)/β1(green)/β2(pink)/β3(yellow)/α0(red), H-bonds, and polar contacts are highlighted. The two panels formed by the β-sheets are defined in square sheets: panel A (blue) and panel B (red).

a recent experiment where point mutation on them caused deactivation of MEK1.[76]

4.4.3. Lymphotactin. A structural comparison between Ltn40 and Ltn10 (Figure 7B,H) indicates that this transition must at least include (a) destruction of one β strand (β0: blue); (b) formation of a new α-helix (α0: red); and (c) rearrangement of the other three β-sheets (β1: green, β2: pink, and β3: yellow). Previous experimental efforts suggest that such change is so dramatic that it is unlikely to occur without complete unfolding/refolding of the corresponding β-sheets.[50] Remarkably, our results revealed no complete unfolding/refolding of Ltn at all. Instead, a few polar contacts are sufficient to stabilize the partially unfolded β-sheets at all stages of the transition.

The free energy surface along the MFEP found by TAPS revealed a three-step process (two ISs and three TSs, Figure 7A), with the height of barriers ranging from 15 to 19 k_BT. To reach the first transition state TS I (Figure 7C), Ltn40 (Figure 7B) must first destabilize and unfold β0 from the original four β-sheets β0−β3 (panel A, blue), resulting in a plane comprising β1−β3 only (panel B, red). Meanwhile, the flexible loop that finally turned into a long helix α0 formed a compact structure partly due to three polar contacts: E31($\boldsymbol{\beta 1}$):R57($\boldsymbol{\alpha 0}$), E31($\boldsymbol{\beta 1}$):R61($\boldsymbol{\alpha 0}$), and R35($\boldsymbol{\beta 2}$):T54($\boldsymbol{\alpha 0}$). Additional H-bonds are also found between the loop that connects β0/β1 and the loop that connects β2/β3 in TS I. Subsequently, these loop−loop H-bonds break again with β0 unfolded, leading to increased flexibility of β1 but also a newly formed polar contact T16:R43 that appears to stabilize the terminal of β1. Such collective motions led the system to reach the first intermediate IS I (Figure 7D).

To overcome the second ∼19 k_BT barrier, the polar contacts among E31/R57/R61/R35/T54 have to break. To compensate for such energy loss, an additional polar contact R23($\boldsymbol{\beta 1}$):S67($\boldsymbol{\alpha 0}$) is newly formed in TS II (Figure 7E), which never appeared in either Ltn40 or Ltn10 (Figure 7H). Though this contact did not last long, it assisted the formation of the polar contact T16:R57($\boldsymbol{\alpha 0}$), which promoted the formation of a complete panel B in the second intermediate IS II (Figure 7F). Besides, in IS II, α0 started to fold from its terminal due to the polar contact R65:T69 (not present in Ltn10).

The last barrier is of similar height to the second. To overcome this barrier, the R65:T69 contact disappeared and panel B transformed to a highly stretched structure where H-bonds among β1−3 were almost broken (Figure 7G). Finally, panel B turned into a standard β-sheet with α0 completely folded (Figure 7H).

5. DISCUSSION

Similar to other path-searching/optimization methods, TAPS is a sampling technique built upon the assumption that the geometric distance among neighbor nodes (conformations) of the found MFEP can well approximate their kinetic distance. This class of methods is well suited for initial sampling of the system of interest, where no kinetic information is yet available. This distinguishes TAPS from analysis techniques such as the time-structure based Independent Component Analysis (tICA)[77−79] that attempts to identify the slowest mode of motions from abundant trajectories with unbiased kinetics information. As the MFEP found by TAPS is by definition statistically significant, the conformations on the MFEP and the free energy surface along the MFEP can guide the subsequent computation of kinetics via methods such as Markov State Models,[40,41] milestoning,[80] markovian milestoning[81] and weighted ensemble milestoning.[82]

Different from other path-searching methods, TAPS aims to minimize the overall information provided as input to the path optimization process. However, unlike path-sampling techniques where the paths are defined directly in the phase space

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

with no prior assumption, the execution of TAPS still relies on a preset distance metric, which is its only major assumption. To alleviate the difficulty in choosing this metric, we provided guidelines in this manuscript and used the full system images as the path nodes so that adjustments in the metric can be performed in any iteration.

For all three systems in the present work, targeted MD has been used to generate an initial path. However, complex transitions of bio-macromolecules often occur through multiple parallel pathways. Therefore, a systematic yet affordable protocol able to provide different parallel pathways is crucial for further development of the TAPS approach.

6. CONCLUSIONS

We have implemented a parallel version of the automated path-searching method TAPS with GPU support and additional module for convergence check. Using this new implementation, we successfully located the MFEP connecting the ground and excited states of the L99A variant of the T4 lysozyme, which is in good agreement with previous findings. TAPS also revealed the critical role of polar contacts R227:L235 and Y229:E255 in the loop-in/loop-out transition of MEK1. In particular, TAPS provided unprecedented insights into the Ltn40/Ltn10 transition of lymphotactin involving complicated unfolding/refolding: five polar contacts (E31:R57/E31:R61/R35:T54/T16:R43/R23:S67) are sufficient to stabilize various partially unfolded structures and thus no complete unfolding/refolding of its β-sheets is necessary. Based on these results, we conclude that TAPS is a general and promising approach for studying the functional dynamics of complex biomolecular systems. Guidelines for its input parameters and discussion on possible further improvements are also provided.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00182.

Relation between sequences and secondary structures for protein; details of the MD simulation of the three systems; parameters used in TAPS simulation and analysis methods; details and results of umbrella sampling for the three systems; definition of polar contact and h-bond; structural template used for define the central axis of α−helix; complete structures of MEK1 during the loop-in and loop-out transition (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Lizhe Zhu − *Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, P. R. China;* orcid.org/0000-0001-8252-7807; Email: zhulizhe@cuhk.edu.cn

Authors

Kun Xi − *Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, P. R. China; School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China*

Zhenquan Hu − *Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, P. R. China; School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China*

Qiang Wu − *School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, P. R. China*

Meihan Wei − *Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, P. R. China*

Runtong Qian − *Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, P. R. China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00182

Author Contributions

L.Z. and K.X. designed the research; K.X., Z.H., Q.W., M.W., and R.Q. performed the calculations; L.Z. and K.X. analyzed the data; K.X., Z.H., and L.Z. wrote the manuscript. All authors discussed the results and reviewed the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Warshel Institute for Computational Biology Funding from Shenzhen City and Longgang District, the National Science Foundation of China (Grant 31971179), and the Science, Technology and Innovation Commission of Shenzhen Municipality (Grant JCYJ20200109150003938, and RCYX20200714114645019). We appreciate the National Supercomputer Center in Guangzhou for high-performance computing services. We thank Prof. Ulrich H. E. Hansmann, University of Oklahoma for introducing the Ltn system and Dr. Huihui Liu and Dr. Wenping Lyu for fruitful discussions.

■ REFERENCES

(1) Rohrdanz, M. A.; Zheng, W.; Clementi, C. Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 295−316.

(2) Malmstrom, R. D.; Lee, C. T.; Van Wart, A. T.; Amaro, R. E. Application of molecular-dynamics based markov state models to functional proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2648−2657.

(3) Bowman, G. R.; Huang, X.; Pande, V. S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **2009**, *49*, 197−201.

(4) Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19765−19769.

(5) Zhang, B. W.; Dai, W.; Gallicchio, E.; He, P.; Xia, J. C.; Tan, Z. Q.; Levy, R. M. Simulating replica exchange: Markov state models, proposal schemes, and the infinite swapping limit. *J. Phys. Chem. B* **2016**, *120*, 8289−8301.

(6) Morcos, F.; Chatterjee, S.; McClendon, C. L.; Brenner, P. R.; López-Rendón, R.; Zintsmaster, J.; Ercsey-Ravasz, M.; Sweet, C. R.; Jacobson, M. P.; Peng, J. W.; Izaguirre, J. A. Modeling conformational ensembles of slow functional motions in Pin1-WW. *PLoS Comput. Biol.* **2010**, *6*, No. e1001015.

(7) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, No. 174105.
(8) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386−2396.
(9) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, No. 155101.
(10) Pan, A. C.; Roux, B. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **2008**, *129*, No. 064107.
(11) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(12) Moroni, D.; van Erp, T. S.; Bolhuis, P. G. Investigating rare events by transition interface sampling. *Physica A* **2004**, *340*, 395−401.
(13) Bolhuis, P. G. Rare events via multiple reaction channels sampled by path replica exchange. *J. Chem. Phys.* **2008**, *129*, No. 114108.
(14) Swenson, D. W. H.; Bolhuis, P. G. A replica exchange transition interface sampling method with multiple interface sets for investigating networks of rare events. *J. Chem. Phys.* **2014**, *141*, No. 07B617_1.
(15) Du, W.; Bolhuis, P. G. Equilibrium kinetic network of the villin headpiece in implicit solvent. *Biophys. J.* **2015**, *108*, 368−378.
(16) Cabriolu, R.; Skjelbred Refsnes, K. M.; Bolhuis, P. G.; van Erp, T. S. Foundations and latest advances in replica exchange transition interface sampling. *J. Chem. Phys.* **2017**, *147*, No. 152722.
(17) Best, R. B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.
(18) Weinan, W.; Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391−420.
(19) Peters, B. Reaction coordinates and mechanistic hypothesis tests. *Annu. Rev. Phys. Chem.* **2016**, *67*, 669−690.
(20) Faidon, B. Z.; Bolhuis, P. G. Approximating free energy and committor landscapes in standard transition path sampling using virtual interface exchange. *J. Chem. Phys.* **2019**, *151*, No. 174111.
(21) Ee, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, No. 052301.
(22) Ren, W.; Vanden-Eijnden, E.; Maragakis, P.; Weinan, E. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.* **2005**, *123*, No. 134109.
(23) Ee, W.; Ren, W.; Vanden-Eijnden, E. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.
(24) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, No. 024106.
(25) Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446*, 182−190.
(26) Vanden-Eijnden, E.; Maddalena, V. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **2009**, *130*, No. 194103.
(27) Leines, G. D.; Ensing, B. Path finding on high-dimensional free energy landscape. *Phys. Rev. Lett.* **2012**, *109*, No. 020601.
(28) Chen, C.; Huang, Y.; Jiang, X.; Xiao, Y. A fast tomographic method for searching the minimum free energy path. *J. Chem. Phys.* **2014**, *15*, No. 154109.
(29) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(30) Pan, A. C.; Sezer, D.; Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432−3440.
(31) Weiss, D. R.; Levitt, M. Can morphing methods predict intermediate structures? *J. Mol. Biol.* **2009**, *385*, 665−674.
(32) Schlitter, J.; Engels, M.; Krüger, P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graphics* **1994**, *12*, 84−89.
(33) Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224−230.
(34) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, *60*, 96−123.
(35) Vreede, J.; Juraszek, J.; Bolhuis, P. G. Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2397−2402.
(36) Lee, J.; Lee, I. H.; Joung, I. S.; Lee, J.; Brooks, B. R. Finding multiple reaction pathways via global optimization of action. *Nat. Commun.* **2017**, *8*, No. 15443.
(37) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* **2014**, *5*, No. 3397.
(38) Gan, W.; Yang, S.; Roux, B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophys. J.* **2009**, *97*, L8−L10.
(39) Pontiggia, F.; Pachov, D. V.; Clarkson, M. W.; Villali, J.; Hagan, M. F.; Pande, V. S.; Kern, D. Free energy landscape of activation in a signaling protein at atomic resolution. *Nat. Commun.* **2015**, *6*, No. 7284.
(40) Wang, X.; Unarta, I. C.; Cheung, P. P.-H.; Huang, X. Elucidating molecular mechanisms of functional conformational changes of proteins via Markov state models. *Curr. Opin. Struct. Biol.* **2021**, *67*, 69−77.
(41) Wang, W.; Cao, S. Q.; Zhu, L. Z.; Huang, X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *WIREs Comput. Mol. Sci.* **2018**, *8*, No. e1343.
(42) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, No. 054103.
(43) Zhu, L. Z.; Sheong, F. K.; Cao, S. Q.; Liu, S.; Unarta, I. C.; Huang, X. TAPS: A traveling-salesman based automated path searching method for functional conformational changes of biological macromolecules. *J. Chem. Phys.* **2019**, *150*, No. 124105.
(44) Cox, M. A. A.; Cox, T. F. Multidimensional Scaling. In *Handbook of Data Visualization*; Chen, C.-h.; Härde, W. K.; Unwin, A., Eds.; Springer: Berlin, 2008; pp 315−347.
(45) Hénin, J.; Chipot, C. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* **2004**, *121*, 2904−2914.
(46) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **1992**, *355*, 371−373.
(47) Bouvignies, G.; Vallurupalli, P.; Hansen, D. F.; Correia, B. E.; Lange, O.; Bah, A.; Vernon, R. M.; Dahlquist, F. W.; Baker, D.; Kay, L. E. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **2011**, *477*, 111−114.
(48) Brotzakis, Z. F.; Parrinello, M. Enhanced sampling of protein conformational transitions via dynamically optimized collective variables. *J. Chem. Theory. Comput.* **2019**, *15*, 1393−1398.
(49) Fischmann, T. O.; Smith, C. K.; Mayhood, T. W.; Myers, J. E.; Reichert, J. P.; Mannarino, A.; Carr, D.; Zhu, H.; Wong, J.; Yang, R. S.; Le, H. V.; Madison, V. S. Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors. *Biochemistry* **2009**, *48*, 2661−2674.
(50) Tyler, R. C.; Mrray, N. J.; Peterson, F. C.; Volkman, B. F. Native-state interconversion of a metamorphic protein requires global unfolding. *Biochemistry* **2011**, *50*, 7077−7079.
(51) Harvey, S. R.; Porrini, M.; Konijnenberg, A.; Clarke, D. J.; Tyler, R. C.; Langridge-Smith, P. R. R.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E. Dissecting the dynamic conformations of the

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311

metamorphic protein lymphotactin. *J. Phys. Chem. B* **2014**, *118*, 12348−12359.
(52) Tuinstra, R. L.; Peterson, F. C.; Elgin, E. S.; Pelzek, A. J.; Volkman, B. F. An engineered second disulfide bond restricts lymphotactin/XCL1 to a chemokine-like conformation with XCR1 agonist activity. *Biochemistry* **2007**, *46*, 2564−2573.
(53) Fox, J. C.; Thomas, M. A.; Dishman, A. F.; Larsen, O.; Nakayama, T.; Yoshie, O.; Rosenkilde, M. M.; Volkman, B. F. Structure-function guided modeling of chemokine-GPCR specificity for the chemokine XCL1 and its receptor XCR1. *Sci. Signal.* **2019**, *12*, No. eaat4128.
(54) Saunders, M. G.; Voth, G. A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.
(55) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(56) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603.
(57) Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate: phosphate antiporter. *Nat. Commun.* **2015**, *6*, No. 8393.
(58) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(59) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(60) Applegate, D. L.; Bixby, R. E.; Chvátal, V.; Cook, W. J. *The Traveling Salesman Problem: A Computational Study*; Princeton University Press, 2006; pp 1−58.
(61) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(62) Vallurupalli, P.; Chakrabarti, N.; Pomès, R.; Kay, L. E. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. *Chem. Sci.* **2016**, *7*, 3602−3613.
(63) Brotzakis, Z. F.; Parrinello, M. Enhanced sampling of protein conformational transitions via dynamically optimized collective variables. *J. Chem. Theory Comput.* **2019**, *15*, 1393−1398.
(64) Tzeng, S. R.; Kalodimos, C. G. Allosteric inhibition through suppression of transient conformational states. *Nat. Chem. Biol.* **2013**, *9*, 462−465.
(65) López, C. J.; Yang, Z. Y.; Altenbach, C.; Hubbell, W. L. Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E4306−E4315.
(66) Mondal, J.; Ahalawat, N.; Pandit, S.; Kay, L. E.; Vallurupalli, P. Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLoS Comput. Biol.* **2018**, *14*, No. e1006180.
(67) Feher, V. A.; Schiffer, J. M.; Mermelstein, D. J.; Mih, N.; Pierce, L. C. T.; et al. Mechanisms for benzene dissociation through the excited state of t4 lysozyme l99a mutant. *Biophys. J.* **2019**, *116*, 205−214.
(68) Li, P. F.; Song, L. F.; Merz, K. M., Jr. Systematic parameterization of monovalent ions employing the nonbonded model. *J. Chem. Theory Comput.* **2015**, *11*, 1645−1657.
(69) Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(70) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(72) Darden, T.; York, D.; Pedersen, L. J. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *Chem. Phys.* **1993**, *98*, 10089−10092.
(73) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(74) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(75) Liu, Y.; Zhu, J. X.; Guo, X. Q.; Huang, T. C.; Han, J. R.; Gao, J. J.; Xu, D.; Han, W. W. How oncogenic mutations activate human MAP kinase 1 (MEK1): a molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* **2019**, *38*, 3942−3958.
(76) Hanrahan, A. J.; Sylvester, B. E.; Chang, M. T.; Elzein, A.; Gao, J.; Han, W.; Liu, Y.; Xu, D.; Gao, S. P.; Gorelick, A. N.; Jones, A. M.; Kiliti, A. J.; Nissan, M. H.; Nimura, C. A.; Poteshman, A. N.; Yao, Z.; Gao, Y.; Hu, W.; Wise, H. C.; Gavrila, E. I.; Shoushtari, A. N.; Tiwari, S.; Viale, A.; Abdel-Wahab, O.; Merghoub, T.; Berger, M. F.; Rosen, N.; Taylor, B. S.; Solit, D. B. Leveraging systematic functional analysis to benchmark an in silico framework distinguishes driver from passenger MEK mutants in cancer. *Cancer Res.* **2020**, *80*, 4233−4243.
(77) Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(78) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, No. 15102.
(79) Naritomi, Y.; Fuchigami, S. Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* **2013**, *139*, No. 215102.
(80) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, *120*, 10880−10889.
(81) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, No. 194101.
(82) Ray, D.; Andricioaei, I. Weighted ensemble milestoning (WEM): A combined approach for rare event simulations. *J. Chem. Phys.* **2020**, *152*, No. 234114.

https://doi.org/10.1021/acs.jctc.1c00182
J. Chem. Theory Comput. 2021, 17, 5301−5311