

Understanding the Structure and Apo Dynamics of the Functionally Active JIP1 Fragment

Neda Ojaghloo,[§] Justin Airas,[§] Lauren M. McRae, Cooper A. Taylor, Bill R. Miller, III, and Carol A. Parish*



Cite This: *J. Chem. Inf. Model.* 2021, 61, 324–334



Read Online

ACCESS |



Metrics & More



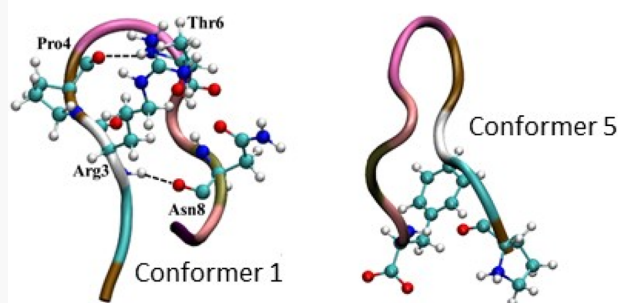
Article Recommendations



Supporting Information

ABSTRACT: Recent experiments indicate that the C-Jun amino-terminal kinase-interacting protein 1 (JIP1) binds to and activates the c-Jun N-terminal kinase (JNK) protein. JNK is an integral part of cell apoptosis, and misregulation of this process is a causative factor in diseases such as Alzheimer's disease (AD), obesity, and cancer. It has also been shown that JIP1 may increase the phosphorylation of tau by facilitating the interaction between the tau protein and JNK, which could also be a causative factor in AD. Very little is known about the structure and dynamics of JIP1; however, the amino acid composition of the first 350 residues suggests that it contains an intrinsically disordered region. Molecular dynamics (MD) simulations using AMBER 14 were used to study the structure and dynamics of a functionally active JIP1 10mer fragment to better understand the solution behavior of the fragment. Two microseconds of unbiased MD was performed on the JIP1 10mer fragment in 10 different seeds for a total of 20 μ s of simulation time, and from this, seven structurally stable conformations of the 10mer fragment were identified via classical clustering. The 10mer ensemble was also used to build a Markov state model (MSM) that identified four metastable states that encompassed six of the seven conformational families identified by classical dimensional reduction. Based on this MSM, conformational interconversions between the four states occur via two dominant pathways with probability fluxes of 55 and 44% for each individual pathway. Transitions between the initial and final states occur with mean first passage times of 31 (forward) and 16 (reverse) μ s.

Most Probable JIP1 10mer Conformations



INTRODUCTION

The c-Jun Amino-terminal kinase-interacting protein 1 (JIP1) is a scaffold protein that facilitates the activation of the c-Jun N-terminal kinase proteins (JNK) and enhances JNK signaling by creating a proximity effect between JNK and upstream kinases.^{1,2} The JNK proteins are serine/threonine protein kinases that are members of the MAPK family and they play an important role in cell apoptosis.^{3–5} There are three different JNK genes: JNK1, JNK2, and JNK3.⁶ Stress-inducing stimuli such as heat shock, UV irradiation, hyperosmolarity, and ischemic injury activate the JNK proteins, which leads to the phosphorylation of transcription factors and substrates involved in cell survival and proliferation, insulin receptor signaling, and mRNA stabilization. These are related to the pathogenesis of several diseases, such as diabetes, cancer, and Alzheimer's disease (AD).^{7–11} Because of this relationship, JNK is an important target for drug therapeutics.¹² JNK is thought to initiate cell death by activating c-Jun.¹³

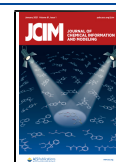
JIP1 is necessary for JNK activation in response to stress and its deficiency causes reduced apoptosis.¹⁴ Vaishnav et al. found that during the initiation of apoptosis, maximal JNK activation is observed when JIP1 is intact, whereas cleavage of JIP1

correlates with JNK inactivation and stops the progression of cell death.¹⁵ It has been shown that the JNK protein is atypically stimulated in AD patients.^{16–19} This may be due to the fact that JIP1 interacts directly with phosphorylated tau,²⁰ and phosphorylated tau proteins are present in the brains of AD patients and believed to be a factor in the proliferation of the disease.²¹ It has been proposed that JNK phosphorylates tau, and these interactions suggest that JIP1 could increase tau phosphorylation by facilitating the interaction between tau and JNK.^{22–24}

The 711-residue JIP1 protein is localized in the neuronal region of the brain, which suggests that it is critical in cell communication.²⁵ Barr et al. has shown an 11-residue JIP1 peptide fragment (residues 157–167) is the functionally active region within the protein.²⁶ The 11mer peptide activates the

Received: August 27, 2020

Published: December 30, 2020



ACS Publications

© 2020 American Chemical Society

324

<https://dx.doi.org/10.1021/acs.jcim.0c01008>
J. Chem. Inf. Model. 2021, 61, 324–334

JNK protein through competition with substrates and upstream kinases.^{27,28} Currently, there are two experimental structures containing 10 of the 11 residues (PDB IDs 2H96 and 1UKH). A root mean square comparison of the experimental structures shows that they are very similar structures (RMSD = 0.513 Å). The missing residue in both of these structures is Arg157.^{29,30} Arg160, Pro161, Leu164, and Leu166 have been shown to be the most critical residues in the retention of function, and removing either the C-terminus or the N-terminus of the 11-residue peptide has had no impact on function.²⁶

The sequence of the full JIP1 protein suggests that the first 350 residues are an intrinsically disordered region (IDR).³¹ IDRs lack a well-ordered, three-dimensional structure; i.e., they are disordered and conformationally dynamic yet biologically functional.³² They are characterized by sequences of low complexity and are typically enriched in charged and hydrophilic amino acids while lacking bulky hydrophobic residues.³³

Although there is a considerable amount of previous research that establishes the role of JIP1 in regulating JNK signaling, less is known about the structure or dynamics of apo-JIP1 or the functionally active apo-JIP1 10mer fragment.³⁰ The binding between JNK and JIP1 is imperative to many important cellular processes, most notably cell apoptosis, so understanding the solvent-accessible, atomistic behavior of the apo-JIP1 fragment will help guide future experimental and computational studies of binding. In this study, we focused on characterizing the structure and conformational flexibility of the apo-JIP1 10mer fragment using molecular dynamics (MD) simulations. A variety of classical clustering methods were used to analyze the resulting ensembles, and this allowed us to better understand the conformational flexibility of the JIP1 fragment as well as to compare the dependence of the results on the clustering methodology. In addition, a Markov state model was used to yield experimentally relevant timescales and atomistic representations that provide conformational insight.^{34–38}

METHODS

Structure Retrieval and Preparation. The initial structure of JIP1 was obtained from PDB 2H96 from the Protein Data Bank.^{29,39} This experimental structure contained a dimer of mitogen-activated protein kinase 8 complexed to two 10 residue fragments of JIP1. Both kinases were removed, along with one of the two JIP1 protein fragments (chain G). The sequence of the 10mer fragment is shown in Figure 1. Because Arg157 is at the N-terminus of the 11mer and its absence showed no major impact on function, it was not added

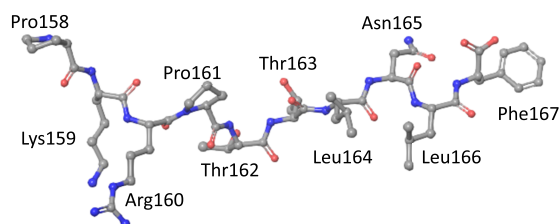


Figure 1. Representation of the 10mer fragment as it appears in the experimentally determined JNK-bound structure (PDB ID 2H96). The FASTA sequence of the full JIP1 protein can be found in Figure S1.

to the experimental 10mer structure we used to initiate this study.²⁶ The sequence was not ACE/NME-capped in the experimental structure²⁹ and, therefore, we did not add the end caps *in silico*. The sequence of the full JIP1 protein can be found in Figure S1.

Molecular Dynamics Analysis of 10mer and Full Apo-JIP1. Unrestrained MD was used to investigate the structure and dynamics of the apo-JIP1 10mer. All simulations were performed using AMBER ff14SB force fields^{40,41} using *tleap* in AmberTools⁴² to neutralize the system with Na⁺ and Cl[−] ions and solvate it using a truncated octahedron periodic box with TIP3P water molecules.⁴³ The box had a minimum 8.0 Å of solvent between the solute and the edge of the unit cell. The AMBER series of force fields has long been used for the accurate modeling of protein secondary structure and dynamics.⁴⁴ AMBER ff14SB provides improved secondary structure prediction and reproduces NMR X1 scalar coupling constants for proteins in solution.⁴⁵ The GPU-accelerated *pmemd* code of AMBER 14 was used to perform minimization, heating, equilibration, and MD steps for all simulations.^{41,46,47} Minimization was performed in seven stages with 1000 steps of the steepest descent minimization followed by 4000 steps of conjugate gradient minimization for a total of 5000 steps. In the first stage, a restraint of 10.0 kcal/mol/Å² was applied to all nonhydrogen atoms, and this restraint was systematically lowered to 0.0 kcal/mol/Å² by stage seven. The protein was then restrained using a 2.0 kcal/mol/Å² force during heating from 10 to 300 K. Equilibration was also performed over seven steps in which the restraint on all 10 residues was lowered from 10.0 to 0.0 kcal/mol/Å² by stage seven. Following this equilibration, MD was performed at constant pressure, temperature, and pH. The SHAKE algorithm was used to restrain all bonds to hydrogen atoms and a 2.0 fs time step was used in these simulations.⁴⁸ For the 10mer, 10 different random seeds, initiated from the 2H96 experimental structure and subjected to minimization and equilibration as described above, were run for 2 μs each to ensure that the potential energy surface was sufficiently sampled; energies and coordinates were saved every 0.1 ns (100 ps).

MD Analysis. For each trajectory, AmberTools 14 *cpptraj*⁴² was used to calculate root-mean-square deviations (RMSD), RMS residue fluctuations (RMSF), the occurrence of secondary structure, cluster analysis, and hydrogen bond formation.^{41,42} Structures and trajectories were visualized using VMD, UCSF Chimera, and PyMOL.^{49–51} For *cpptraj* clustering, we analyzed every 10th structure in the ensemble reducing our 20 μs dataset to 20,000 structures. Using the default threshold value of 3.0 Å, *cpptraj* clustered 18,988 (out of 20,000) of the structures into the same family (Table S1). Because default *cpptraj* clustering did not identify statistically significant families, we initially analyzed the ensembles using RMSD analysis and trajectory visualization. We calculated RMSD values relative to the solvated and equilibrated initial structure every 0.1 ns.⁵² Trajectory visualization indicated that regions of unchanging or slowly changing RMSD values corresponded to similar structures. We calculated average structures over trajectory regions corresponding to flat or unchanging RMSD and used these structures to identify the most prevalent families within the ensemble. Following this manual clustering, we analyzed the ensemble using the Multiscale Modeling Tools for Structural Biology (MMTSB) clustering toolkit.⁵³ We also revisited *cpptraj* clustering and determined that clustering using pairwise distances resulted in

reasonably good clustering (*vide infra*). MMTSB is a cluster analysis program that clusters using RMSD mutual similarity, meaning that it compares the RMSD of each structure, and clusters those that are similar according to an RMSD cutoff.⁴² The MMTSB tool set has K-means⁵⁴ and hierarchical clustering algorithms; we used the K-means algorithm to cluster based on atomic RMSD with a cut-off value of 2 Å. We performed this analysis using the *enscluster.pl* script provided in the MMTSB toolset.⁵³ We utilized 49,990 structures corresponding to every fourth frame of the full 20 μ s ensemble.

Clustering by manual RMSD with visualization, *cpptraj* pairwise RMSD, and MMTSB K-means clustering grouped the ensemble into seven conformational families.

(In what follows below, we are using an upper case “Conformer” to refer to these seven families.) The average structures used for this analysis were obtained either from MMTSB⁵³ (Conformers 1, 2, 4, and 5), RMSD analysis (Conformers 1–6), or *cpptraj* (Conformers 1–7). Each structure in the overall ensemble was associated with a Conformer if the RMSD with the average structure representing that Conformer was within 2.0 Å. This allowed us to report fraction occurrence and a persistence lifetime for each Conformer family. In this fashion, the fraction occurrence is simply the number of structures assigned to each Conformer divided by the total number of structures. Average lifetime analysis was performed on all trajectories to quantify the persistence of a specific Conformer within the entire simulation. Using RMSD-based binned structures, we calculated the persistence length of consecutive occurrences of each Conformer over each trajectory and summed those values to obtain an average lifetime for each of the seven Conformers. For example, if we measure a particular conformation occurring with a persistence length of f over N consecutive frames (N_{frames}) in our simulation, then the lifetime τ is defined as

$$\tau = \frac{\sum (N_{\text{frames}} \times f)}{\sum f} \times \frac{0.1 \text{ ns}}{\text{frame}} \quad (1)$$

where N_{frames} is the number of consecutive frames for which the Conformer occurred, $\sum f$ is the total number of frames that contain one of the seven Conformers, and f is the persistence length of such consecutive occurrences. Using this metric, a conformation that persists for many consecutive frames will have a longer lifetime value.

Markov State Model Construction. Markov state modeling (MSM) provides an alternative method for understanding conformational behavior from molecular dynamics time series data.^{34–38,55–59} Rather than relying on geometrically based clustering methods to group ensemble members into structurally similar families, Markov state models group ensembles into long-lived (metastable) states with rapid interconversion within states and slow conversion between states.^{34,35} To build the MSM model, we transformed the MD ensemble into a series of microstates. The dimensionality of the ensemble was reduced by the featurization method^{55,60} and the dimensional reduction algorithm.^{55,56,61,62} Transition probabilities were defined by counting how many transitions happen between each state at a specific time (lag time). The resulting transition probability matrix $P(\tau) = [p_{ij}(\tau)]$ ⁵⁵ was used to describe the probability of transitioning from state i at time t to state j at time $t + \tau$. From this matrix, we can

determine pathways between states and the probabilities of transitions between them.⁵⁵

The MSM model developed in this work was carried out using the open-source python package PyEMMA 2.4^{55,63} using the following steps:

- (1) Load the coordinates and select appropriate features. The VAMP-2⁶⁰ score was used to select the best features for discretizing the data.
- (2) Apply dimensional reduction and clustering algorithms to discretize the data.
- (3) Confirm the Markovianity of the system by computing implied timescales and performing the Chapman–Kolmogorov (CK) test.^{38,55,64}
- (4) Compute the free energy landscape and the probability fluxes^{35,65} from the Markov state model as a function of the two slowest internal coordinates (ICs) to visualize the number of metastable states and the conformational transitions between them.

RESULTS AND DISCUSSION

The purpose of this study was to understand the time evolution and dynamical behavior of the functionally active fragment of the JIP1 protein. Toward this end, we performed a collective 20 μ s of unbiased molecular dynamics on the apo-JIP1 10mer, characterized the conformations of this fragment using classical dimensional reduction tools (RMSD, K-means, and pairwise distance clustering), and utilized the ensemble to develop a Markov state model to describe the dynamics of the 10mer fragment.

To generate an ensemble for characterizing the dynamics of the 10mer fragment, 2 μ s simulations were performed, initiated from 10 different random seeds each initiated from the experimental structure (2H96; Figure 2). Time series data

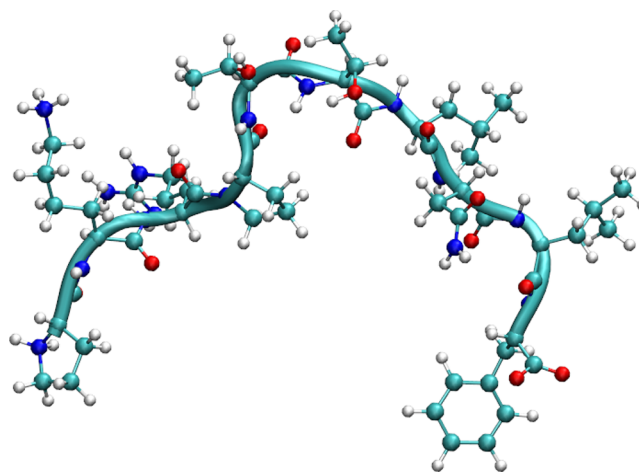


Figure 2. Solvated and equilibrated JIP1 10mer experimental structure used as the reference for RMSD calculations for each trajectory. The sequence of this fragment is Pro1-Lys2-Arg3-Pro4-Thr5-Thr6-Leu7-Asn8-Leu9-Phe10.

using secondary structure characterization for each of these simulations can be found in the Supporting Information (Figure S2). Over the course of each trajectory, the 10mer fragment adopts a wide variety of secondary structures: turns, bends, and partial helices. These secondary structures were identified using *cpptraj*, which uses the DSSP algorithm.⁴² Using classical clustering methods, we identified seven

structurally stable conformations of the apo-JIP1 10mer. Throughout each simulation, hundreds of different conformations were sampled, but we searched for conformations that were persistent and stable over a period of time, suggesting that they will be significantly populated in experimental ensembles because they are located in broad and deep energy minima. A Markov state model applied to the ensemble suggests that six of the seven conformations are part of four important and long-lived metastable states. With 20 μ s of simulation data and multiple methods of conformational analysis, we are confident that we are describing the conformational flexibility of the JIP1 10mer fragment.

These seven conformations (described in more detail below) were obtained from the 20 μ s ensemble via three different clustering methods. We used different clustering methods so as to avoid a methodological dependence in our results and to provide a comparison of the various methods. For instance, manual clustering of the 200,000 structure ensemble by RMSD relative to the initial structure identified six Conformers. However, because multiple structures may give the same RMSD to a single structure,⁶⁶ we also utilized MMTSB⁵³ K-means and *cpptraj*⁴² pairwise distance clustering, which found 10 and 19 Conformers, respectively (details below).

For our manual analysis, if the RMSD relative to the solvated and equilibrated structure shown in Figure 2 was within 1 Å, then we considered the molecule conformationally stable. Using this semiquantitative criterion, we then visualized and overlaid the structures in every flat region that extended more than 250 ns in the RMSD plots (Figure 1, black rectangle). We performed an RMSD analysis of these structures within VMD to confirm that the same structure was found throughout the region of structural stability. A typical trajectory is shown in Figure 3 and a complete set of trajectories can be found in

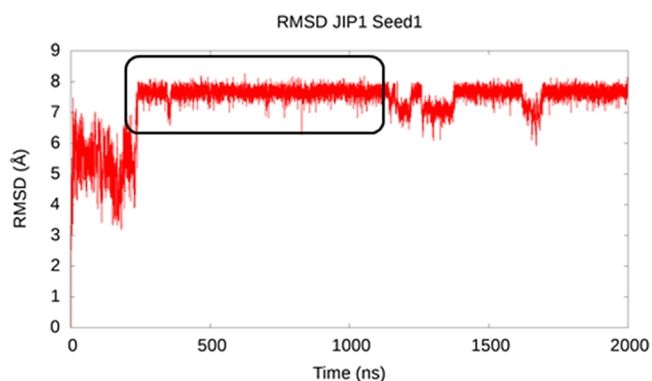


Figure 3. Trajectory initiated from Seed 1 where Conformer 1 was identified. RMSD trajectory values indicate that the simulation was stable, and visualization of the ensemble after 250 ns reveals a persistent conformation (Conformer 1). This Conformer persisted for 1750 of 2000 ns. This period of simulation stability is highlighted with a black box.

Figure S3. Using this approach, we identified six of the seven conformationally unique, persistent Conformers reported in this study.

K-Means and Pairwise Distance Clustering. To further understand the conformational behavior and to compare clustering methods, we also subjected every fourth structure in the apo 10mer ensemble ($\sim$ 50,000 structures) to K-means clustering using the MMTSB toolkit.⁵³ This clustering produced 115 structural families. The 10 most populated

families contained more than 1000 structures within each family. Of those 10 families, 8 families corresponded to Conformers 1, 2, 4, and 5 identified via RMSD analysis. Conformers 3 and 6 were not identified by K-means clustering. We also utilized clustering methods available in *cpptraj* to group every fifth structure in the 10mer ensemble ($\sim$ 40,000 structures).^{41,42} We analyzed a variety of *cpptraj* clustering criteria and found that using pairwise distances resulted in clusters with a good distribution of structures within conformational families. Pairwise distance clustering using all atoms grouped the ensemble into 19 conformational families with fractional occurrences ranging from 1 to 17% (Table S1). The six Conformers identified via RMSD analysis were identified in *cpptraj* clustering, and a seventh cluster was identified (details below).

We performed an unequal variance (Welch's) two-sample *t* test⁶⁷ to ascertain the significance of the clustering of ensemble structures into each of seven Conformers. In this test, we used the RMSD values relative to each Conformer to compare the deviation of all structures to the deviation of those structures binned into each Conformer (Table S2). In all cases, the *p* values suggest that the clustering is significant. We also compared the RMSD values between all Conformers, which ranged from 0.8 Å (Conformers 1 and 5) to 7.8 Å (Conformers 1 and 4; Table S3). We describe below the structural features of the Conformers found via classical clustering before constructing a Markov state model using the 200,000 JIP1 10mer dataset.

Conformer 1. Based on our classical clustering (non-Markovian), Conformer 1 was present for the longest period. This conformation persisted for 1760 ns of 2000 ns of the first trajectory, representing a 78% occurrence in trajectory #1 (Figure 3) and an overall 1.2–17.4% occurrence in the 20 μ s ensemble (Table 1). Conformer 1 is stabilized throughout the simulation by a pair of hydrogen bonds between residues Arg3-Asn8 and Pro4-Thr6, as well as, less frequently, by interactions between Pro1-Phe10 and Arg3-Leu7. Figure 4 depicts Conformer 1 in two different orientations, highlighting the two different hydrogen bonding networks. For Conformer 1, the individual H-bonds shown in Figure 4A,B were present for 76.2% (Asn8-Arg3), 73.4% (Pro4-Thr6), 27.1% (Arg3-Leu7), and 23.4% (Pro1-Phe10) of the 20 μ s ensemble. A visualization of the ensemble reveals that hydrogen bonding interactions shown in Figure 4A,B are driving the loop structure that we see throughout the 20 μ s simulation.

Conformers 2–7. Average structures from the other six Conformers found by classical clustering are shown in Figure 5. Conformers 2, 3, and 7 were the only Conformers that contained a persistent secondary structure with alpha helical content in their middle and outer residues (residues 5–7 and 7–9). All of the Conformers adopt a similar loop structure in the middle of the 10mer, except for Conformer 4, which has a loop structure at the C-terminal end of the fragment, due to different patterns of hydrogen bonding. Conformer 2 and 7 adopt a 3–10 helical motif formed between residues Thr5-Thr6-Leu7, while Conformer 3 adopts a helical motif between residues Leu7-Asn8-Leu9, respectively.

To further gauge Conformer importance, we evaluated the percent occurrence and the persistence, or lifetime, of each Conformer (Table 1). To calculate percent occurrence and lifetimes, we binned each structure in each trajectory of the ensemble into one or more Conformer by calculating the RMSD of each structure relative to the representative average

Table 1. Ensemble Analysis Using Classical Clustering^a

conformer	1	2	3	4	5	6	7
RMSD % occurrence	17.4	6.7	15.7	1.4	17.1	2.5	22.6
K-means % occurrence	10.2	2.2		4.5	4.9		
pairwise (cpptraj) % occurrence	1.24	4.89	15.60	2.94	16.86	2.24	12.53
bootstrapped pairwise % occurrence	1.23	4.90	15.44	2.97	16.89	2.23	12.56
% error in % occurrence	0.645	0.051	1.016	1.106	0.215	0.446	0.271
lifetime (ns)	4.48	0.86	0.83	0.34	6.41	1.51	0.74

^aFor each Conformer identified by MMTSB (Conformers 1, 2, 4, and 5), RMSD (Conformers 1–6), or *cpptraj* (Conformers 1–7), we report the percent occurrence and lifetime based on RMSD comparison to average structures (*vide infra*). Lifetimes are based on *cpptraj* clustering as that method identified all seven families.

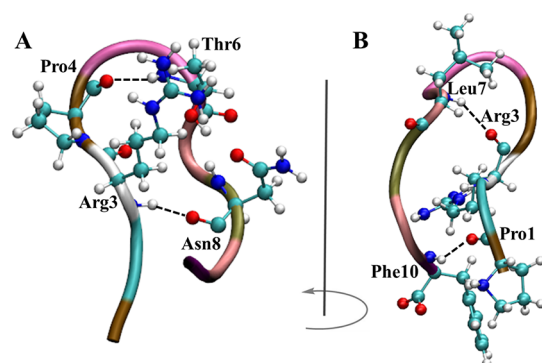


Figure 4. Hydrogen bonding in Conformer 1. (A) The hydrogen bonds highlighted in the left image occurred for 76.2% (Asn8–Arg3) and 73.4% (Pro4–Thr6) of the simulation. (B) Hydrogen bonds shown in the right image occurred for 27.1% (Arg3–Leu7) and 23.4% (Pro1–Phe10) of the simulation. The image in (B) is generated by rotating the image in (A) 180° from front to back. Percentages refer to the presence of each hydrogen bonding interaction throughout the entire 20 μ s ensemble.

structure of each Conformer. If the RMSD values fell within a user-defined cut-off window, then that structure was assigned to that Conformer family.

We evaluated a range of RMSD cut-off windows (Table S4) to find the optimal RMSD tolerance that would minimize

structures being assigned to multiple Conformers while binning as many structures into families as was physically meaningful. In other words, we wanted RMSD tolerance that was flexible enough to account for the bond and angle movements that are possible within one conformation without including dihedral angle changes that interconverted different conformations. Based on the number of duplicate groupings versus the number of structures binned, we selected a cut-off window of 2.0 Å. This binned 83% of the structures into at least one conformational family while only binning 17 and 0.07% into more than one or two families, respectively. From this analysis (Table 1), we see that based on our classical clustering, Conformers 1, 5, and 6 have the longest lifetimes and so are the most persistent Conformers in the overall ensemble. Conformers 1, 3, 5, and 7 display the largest percent occurrence and therefore dominate the ensemble. Conformer 1 forms often and, once formed, is quite persistent, whereas apo-JIP1 10mer does not adopt Conformer 6 very often, but once formed, it persists for, on average, 1.51 ns. Conformer 7 has a relatively high occurrence (12.5–22.6%) along with a relatively short lifetime (0.74 ns), suggesting that this Conformer forms often but does not persist for long. A graphical representation of the persistence frequencies for all Conformers is shown in Figures S4–S9.

To estimate the error associated with the percent occurrences reported in Table 1, we performed a random sampling with replacement (bootstrapping) analysis; i.e., we

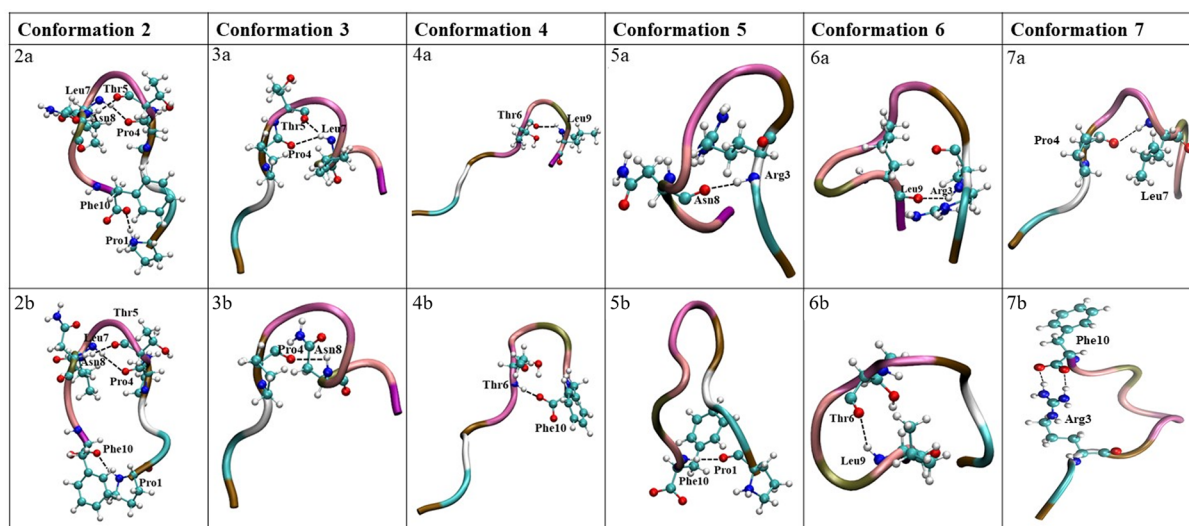


Figure 5. Representative structures of the seven Conformers found for the apo-JIP1 10mer. For each Conformer, hydrogen bonds and their associated residues are highlighted. Residues Thr5–Thr6–Leu7 of Conformer 2 contain 3–10 helical character. (a) and (b) denote different orientations of the same structural snapshot and illustrate different hydrogen bonding patterns.

randomly sampled our ensemble structure-to-Conformer assignments.^{68,69} In both cases, the randomly generated sample size was constrained by the original sample size. The population averages from our resampled percent occurrences are in good agreement with our original data (Table 1) with percent occurrence errors less than 2%.

To ascertain the occurrence of various Conformers in the overall ensemble, as well as within each of the 10 2 μ s trajectories, we used *cpptraj* to perform pairwise distance clustering to bin structures into Conformers 1–7 (Table S5 and Figure S10). From this analysis, we can see that 78% of the structures align with one of the seven Conformers, and all trajectories, except those initiated from Seeds 7 and 8, sample at least two Conformers with percentages greater than 10%. There is some interconversion between various conformations within a trajectory (especially for trajectories originating from Seeds 1–6); however, particular Conformers are more likely to be found in some trajectories; i.e., of all frames sampling Conformer 1, ~50% of those frames are found in the trajectory originating from Seed 1 (Table S6, bottom). A graphical analysis of the Conformer sampling throughout the entire ensemble, organized by trajectory, can be seen in Figure S10. This analysis also suggests that our manual clustering based on RMSD visualization was qualitatively useful, at least for this particular ensemble of a relatively small, disordered protein fragment.

We also performed a percent occurrence and error analysis using the *cpptraj* clustering results on each trajectory initiated using a different random seed (Table S5a). From this, we can see that sampling within a trajectory is not sufficient to find all seven Conformers; however, within the pooled ensemble, we do find all Conformers except for Conformer 1. Given the structural similarity between Conformers 1 and 5 (RMSD = 0.86 Å), it is not surprising that the *cpptraj* pairwise clustering method cannot distinguish these Conformers.

The representative structures for each Conformer (Figure 5) reveal that the JIP1 10mer spends a considerable amount of time in a folded loop conformation with some amount of helical character and that this conformational dynamic is likely driven by hydrogen bonding. To quantify this and identify residues involved in these stabilizing interactions, we analyzed the 20 μ s ensemble for the occurrence of hydrogen bonding associated with each Conformer (Table 2) as well as the full ensemble (Table S6).

For those structures that are identified as belonging to Conformer 1, hydrogen bonding occurs between Asn8 and Arg3, Pro4 and Thr6, and Asn8 and Arg3 for 76.22, 73.45, and 52.77% of the ensemble, respectively. In addition to these most prevalent H-bonding patterns, the hydrogen bond between Pro1 and Phe10 occurred in three Conformer families, while H-bonds between Arg3 and Asp8, Pro4 and Leu7, and Thr6 and Leu9 all occurred in three Conformer families. These interactions contributed to the stability of each of the conformations in which they were involved and therefore are highly important in understanding the overall stability. It is notable that there are no pairwise hydrogen bonds that occur in all conformations. JIP1 takes on a similar shape in each of the conformations but reaches those shapes due to different hydrogen bonding patterns. Examining the hydrogen bonding behavior for the full ensemble (Table S6) suggests that interactions between Asn8 and Arg3, Phe10 and Arg3, and Pro4 and Leu7 dominate with 13.06, 10.06, and 10.06% of the ensemble, respectively.

Table 2. Hydrogen Bond Analysis for Each Conformer Identified in the JIP1 10mer Ensemble^a

conformer	hydrogen bond	percent occurrence	average RMSD (Å)	helical character?
1	Arg3-H:Asn8	76.22	7.58	no
	Thr6-H:Pro4	73.45		
	Leu7-H:Arg3	22.71		
	Phe10-H:Pro1	21.91		
2	Asn8-H:Thr5	20.64	5.57	yes (aa 5–7)
	Leu7-H:Pro4	14.19		
	Pro1-H:Phe10	13.38		
3	Leu7-H:Thr5	21.90	5.91	yes (aa 7–9)
	Asn8-H:Pro4	16.20		
	Leu7-H:Pro4	10.08		
4	Leu9-H:Thr6	9.49	6.31	no
	Thr6-H:Phe10	7.99		
5	Arg3-H:Asn8	52.77	7.74	no
	Phe10-H:Pro1	29.67		
6	Leu9-H:Thr6	11.68	5.78	no
	Arg3-H:Leu9	11.09		
7	Leu7-H:Pro4	59.68	8.58	yes (aa 5–7)
	Arg3-H:Phe10	0.85		

^aBinding interactions within 3.0 Å and 135° between specific residues are indicated along with their percent occurrence in the 20 μ s ensemble. Also highlighted are the average RMSD values for each average structure relative to the solvated and equilibrated initial structure and whether each conformation contained helical character. The helical character was determined using *cpptraj* default angles.⁴²

Taken together, the data suggests that JIP1 is capable of adopting structurally diverse conformations. The presence of seven different conformations suggests that the functionally active 10mer is conformationally quite flexible. RMSD comparison to the JNK-bound 10mer structure (Table 2) suggests that the fragment is not likely to be preorganized for binding to JNK, but the enthalpies of each Conformer do not bias the ensemble into adopting or avoiding particular conformations.

Comparing Classical Methods of Conformational Analysis. Clustering by RMSD comparison to the solvated and equilibrated structure (Figure 2) identified six Conformers. Three of the MMTSB families adopted a structure similar to Conformer 1 (MMTSB family 1 RMSD = 0.748 Å; MMTSB family 2 RMSD = 0.776 Å; MMTSB family 3 RMSD = 0.880 Å), suggesting that Conformer 1-like structures were present for 10.2% of the full ensemble (5107 structures). The fourth most populated MMTSB family corresponded to Conformer 2 (MMTSB family 4 RMSD = 0.418 Å) with a 2.2% occurrence (1094 structures). There were two families with structures similar to Conformer 4 (MMTSB family 5 RMSD = 2.781 Å; MMTSB family 6 RMSD = 1.291 Å), with a 4.5% occurrence (2267 structures). These families were identified as having the same structure as Conformer 2 because they took on the same hydrogen bonds; the larger RMSD is a result of the tails not being aligned. Conformer 5 was present in two families (MMTSB family 7 RMSD = 0.813 Å; MMTSB family 8 RMSD = 0.725 Å) or 4.9% of the ensemble and was represented by 2439 structures. Conformers 3 and 6 were not found within the 10 largest families. A representative structure of each MMTSB cluster is shown in Figures S11–S16 where RMSD superpositions with Conformers 1, 2, 4, and 5 are also shown. Clustering by pairwise distances using *cpptraj* resulted in 19 relatively well-populated

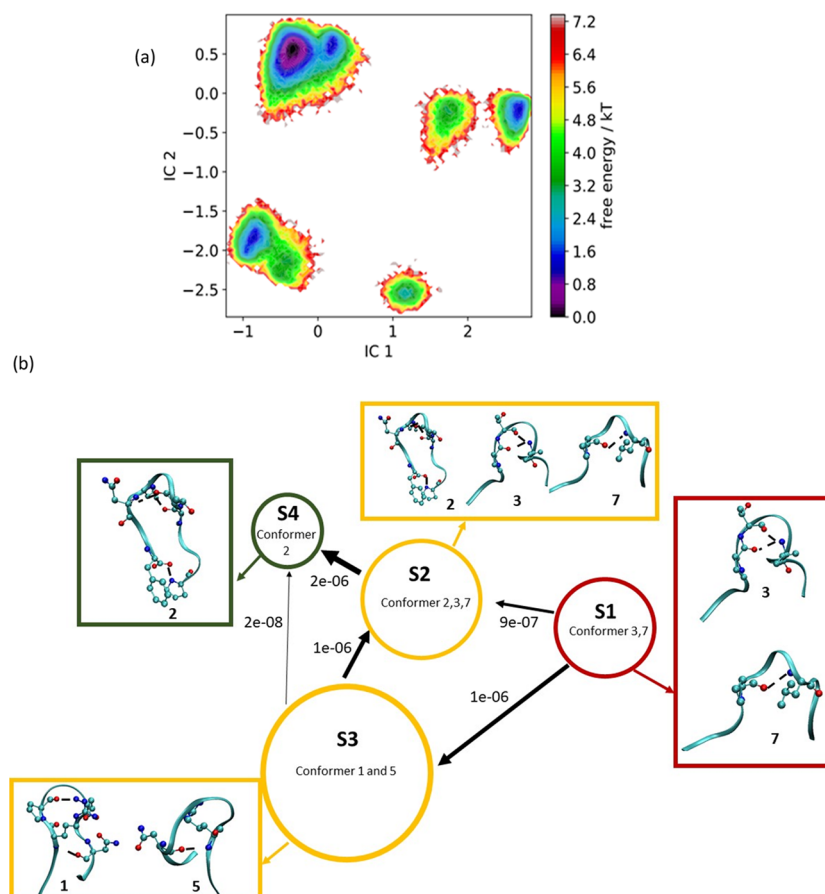


Figure 6. (a) Free energy as a function of the two slowest ICs. The minima in the free energy plot represent the metastable states in our system. (b) TPT analysis for the four metastable-state MSM model with a lag time of 10 ns. Only the dominant coarse-grained fluxes are shown. A cut-off of 0.95 was used to generate the MSM probability flux plots as this preserved 95% of the kinetic variance. The visual size of each state is proportional to its free energy and the arrow size is proportional to the main flux. The red, yellow, and green colors indicate the initial, intermediate, and final states of TPT analysis, respectively. The numbers associated with the arrows are the transition fluxes.

conformational families (Table S7) that confirmed Conformers 1–6 and identified a seventh cluster. Structural agreement between Conformers identified with different clustering methods was assessed using RMSD comparison (Tables S1 and S7) and structural visualization. A total of 78% of the 200,000 frames in the 20 μ s ensemble superimposed within 2 Å to at least one of the seven Conformers (Table S5).

Markov State Modeling. To complement the clustering described above, we also used time-based Markov state modeling to locate long-lived (metastable) states within the 20 μ s ensemble and to identify discrete transitions between those states.^{36–38,70–72} We began the process by selecting molecular features that maximized the total kinetic variance.^{55,63} We analyzed features such as backbone torsions, backbone atom distances, side-chain torsion angles (χ_1), XYZ coordinates, and backbone atom coordinates (Figure S17).⁵⁵ While the VAMP-2^{60,73} score is slightly lower for backbone torsions relative to backbone atom distances, backbone torsions reduce the dimensions most significantly relative to other features (Table S8). Figure S18 shows the VAMP-2^{60,73} score as a function of the number of clusters for the different clustering algorithms. This suggests that using the K-means algorithm, the number of clusters becomes constant after $k = 100$ (Figure S18). Therefore, we choose a k value larger than 100. We also performed a comparison between the clustering algorithms and the number of microstates (Figures S19–S21)

and these results suggest that both uniform time⁶³ and K-means clustering are well distributed in the first two TICA and that the $k = 250$ clustering level is sufficient to discretize the 20 μ s JIP1 dynamics. In this paper, we are presenting only the results from the K-means clustering algorithm.

Time-lagged independent component analysis (TICA)^{56,61,82,74} with a lag time of 10 ns was used to reduce the dimensions of our molecular simulation data to the 11 dimensions that preserve 95% of the kinetic variance and to capture the slowest transition in the system. TICA is one of the best approaches for reducing system dimensions into the slowest subspace by keeping the highest kinetic variance.⁶⁰ During TICA, our kinetic_map variable was set to “FALSE”. Finally, the K-means⁵⁴ algorithm was applied to cluster the MD trajectory and assigned the clusters to separate metastable states.

Model Validation. For building the MSM, we need to identify a lag time that is long enough to capture the dynamics within each metastable state while also being short enough to properly characterize interstate pathways.⁵⁵ The implied timescale t_i can be used to determine the appropriate lag time τ . Equation 2 shows the relationship between the implied timescale and lag time

$$t_i(\tau) = \frac{-\tau}{\ln|\lambda_i(\tau)|} \quad (2)$$

where τ is the lag time and $t_i(\tau)$ is the i th slowest implied timescale determined from the i th largest eigenvalue $\lambda_i(\tau)$ of the transition probability matrix.⁵⁵ If the implied timescale becomes constant with lag time, then the Markov assumption is satisfied.⁶³ Figure S22a shows that the implied timescale curves reach a constant value when the lag time increases beyond 10 ns. Therefore, the MSM estimation of the JIP1 10mer ensemble becomes Markovian at a lag time of 10 ns. Error estimates are computed using Bayesian sampling and are shown in Figure S22a with colored shading.^{55,75}

Performing the Chapman–Kolmogorov (CK) test allows us to validate our MSM.^{38,55,64} For this CK analysis, we reduced the ensemble to four metastable states. This number of metastable states was defined based on the large differences between the implied timescales of the fourth and fifth timescales as shown in Figure S22b.³⁵ We used PCCA^{76–78} to define the membership distribution and choose the state with the highest and lowest memberships as the initial and final states, respectively. We choose these endpoints as our goal was to build an MSM that identified pathways between all states and best described the conformational diversity of the ensemble.

Figure S22c shows that there is good agreement between the CK analysis of the change in probability as a function of time for the chosen metastable states predicted from the MSM (blue dashed lines) and the probability of MSMs estimated for the same metastable states with longer lag times (black solid lines). This suggests that our model is indeed Markovian. The CK^{38,55,64} analysis also confirms that our system contains four metastable states. In Figure S22c, the Bayesian method was used to determine the 95% confidence intervals (the blue shaded area).^{55,75} Both the CK test and implied timescale analysis show that our system is Markovian. Because the MSM is Markovian and the transition matrix has one large eigenvalue $\lambda = 1$,^{55,79} it will allow transitions from one state to all other states and should provide a good description of the equilibrium dynamics of the apo-JIP1 protein fragment. This is a feature benefit of performing an MSM analysis.

We have plotted the free energy of the system over the two independent coordinates (ICs) that resolves our ensemble into four metastable states. These states are shown as minima in the free energy plot (Figure 6a). In Figure S23, the time series of the first three slowest ICs illustrates the transition between the minima.

To relate our MSM to the classical clustering described above, we also built an MSM using the cpptraj-assigned clusters (RMSD of each cluster) as an input to a new MSM analysis. We observed that the implied timescales and the number of metastable states of this MSM model are in good agreement with the MSM results that used the 20 μ s MD trajectory (Figures S24 and S25).

We used a coarse-grained transition path theory (TPT) method^{57,64,80,81} to illustrate the network of transition pathways obtained from our MSM results (Figure 6b). TPT computes the fluxes between states and extracts information about the kinetics of transitions between initial and final states. As can be seen in Figure 6b, there are two pathways emanating from S1 with transition probabilities of 1×10^{-6} and 9×10^{-7} . These pathways lead to S2 and S3. S3 transitions to S2 as an intermediate state before reaching the final state S4. The percentages of flux for pathways $[S1 \rightarrow S3 \rightarrow S2 \rightarrow S4]$, $[S1 \rightarrow S2 \rightarrow S4]$, and $[S1 \rightarrow S3 \rightarrow S4]$ from S1 to S4 are 55, 44, and 0.01%, respectively. This suggests that two pathways $[S1 \rightarrow$

$S3 \rightarrow S2 \rightarrow S4]$ and $[S1 \rightarrow S2 \rightarrow S4]$ dominate the flux between S1 and S4. The mean first passage time (MFPT)⁷⁵ from S1 to S4 is 31 μ s, while the MFPT to return from S4 to S1 is 16 μ s.

An RMSD comparison between the average structures of each of the seven Conformers, with 2000 sample structures from each of the four metastable states, allows us to relate the results from classical clustering to the results from the MSM. Using the sample by state method,^{55,63} we find that Conformers identified by classical clustering are also found in the MSM, except for Conformer 6 (Table S9). There is relatively good agreement between the methods used in this study as Conformer 6 is one of the least populated structures in the classical clustering analysis and not found via MMTSB K-means clustering (Table 1). This suggests that Conformer 6 is more likely a rare-event state. Our MSM-based percent occurrence results (Table S9), and RMSD values relative to the representative structure of the seven Conformers (Table S10), show that S1 contains mostly Conformers 3 and 7, S2 contains Conformer 7, S3 is dominated by Conformers 1 and 5, and S4 contains Conformer 2. This is in good agreement with our results from classical clustering.

In addition, Conformers 1 and 5 in S3 are the conformers that have the highest population in our MSM results, which is again in agreement with our classical clustering. It is also notable that Conformers 1 and 5 have the same occurrence in state S3, which is in good agreement with our previous findings (Table 1) showing that Conformers 1 and 5 have similar structures and percent occurrences in both models. Both classical clustering and MSM analysis indicate that Conformers 1, 3, 5, and 7 are highly populated and dominate the ensemble, existing in S1–S3.

CONCLUSIONS

Understanding the apo dynamics of the JIP1 protein is important because of the crucial role that this protein plays in a variety of disease states ranging from obesity, cancer, and AD. Conformations in which JIP1 are stable can have important implications for binding to other proteins. Past studies have indicated that JIP1 activates cell apoptosis via the JNK protein. JIP1 also may enable the binding between JNK and phosphorylated tau. Our studies suggest that the functionally active JIP1 10mer adopts six or seven energetically and structurally stable conformations. The fragment takes on a similar shape in all the Conformers, and the Conformers are stabilized by different hydrogen bonding patterns. The ensemble is dominated by structures that adopt a folded loop conformation with some amount of helical character in the middle and outer residues (residues 5–7 and 7–9). Clustering via RMSD, pairwise distance comparison, and K-means clustering resulted in differing numbers of conformers, but seven Conformers produced by at least two of the methods were unique and persistent. A Markov state model built using the ensemble indicates that the JIP1 10mer exists in four long-lived, metastable states that encompass six of the seven Conformers identified via classical clustering.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c01008>.

Trajectories and hydrogen bonding occurring in all 10 JIP1 10mer seeds, cpptraj⁴² and MMTSB clustering

details, and lifetime analysis and homology model details (PDF)

AUTHOR INFORMATION

Corresponding Author

Carol A. Parish – Department of Chemistry, University of Richmond, Richmond, Virginia 23173, United States;
orcid.org/0000-0003-2878-3070; Email: cparish@richmond.edu

Authors

Neda Ojaghlo – Department of Chemistry, University of Richmond, Richmond, Virginia 23173, United States
Justin Airas – Department of Chemistry, University of Richmond, Richmond, Virginia 23173, United States
Lauren M. McRae – Department of Chemistry, University of Richmond, Richmond, Virginia 23173, United States
Cooper A. Taylor – Department of Chemistry, University of Richmond, Richmond, Virginia 23173, United States
Bill R. Miller, III – Department of Chemistry, Truman State University, Kirksville, Missouri 63501, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.0c01008>

Author Contributions

[§]N.O. and J.A. contributed equally to this work.

Notes

The authors declare no competing financial interest.
Authors will release the atomic coordinates and experimental data upon article publication.

ACKNOWLEDGMENTS

We would like to thank Dr. Chris Stevenson of the University of Richmond for assistance with the lifetime analysis. We would also like to thank Dr. Julie Pollock of the University of Richmond for fruitful discussions regarding peptide synthesis and end-capping. We would like to thank Zachary Stewart for assistance with some of the RMSD measurements. This research was supported by funding from the NSF RUI program (grants CHE-1213271 and CHE-1800014), the Donors of the American Chemical Society Petroleum Research Fund, the University of Richmond, and the Floyd D. and Elisabeth S. Gottwald Endowment. J.A., L.M.M., and C.A.T. were the recipients of a summer fellowship from the University of Richmond, School of Arts & Sciences, as well as from the University of Richmond Chemistry Department. Computational resources were provided, in part, by the MERCURY supercomputer consortium under NSF grant CHE-1626238.

REFERENCES

- (1) Nihalani, D.; Wong, H. N.; Holzman, L. B. Recruitment of JNK to JIP1 and JNK-dependent JIP1 Phosphorylation Regulates JNK Module Dynamics and Activation. *J. Biol. Chem.* **2003**, *278*, 28694–28702.
- (2) Whitmarsh, A. J.; Cavanagh, J.; Tournier, C.; Yasuda, J.; Davis, R. J. A mammalian scaffold complex that selectively mediates MAP kinase activation. *Science* **1998**, *281*, 1671–1674.
- (3) Kyriakis, J. M.; Avruch, J. pp54 microtubule-associated protein 2 kinase. A novel serine/threonine protein kinase regulated by phosphorylation and stimulated by poly-L-lysine. *J. Biol. Chem.* **1990**, *265*, 17355–17363.
- (4) Dérijard, B.; Hibi, M.; Wu, I.-H.; Barrett, T.; Su, B.; Deng, T.; Karin, M.; Davis, R. J. JNK1: a protein kinase stimulated by UV light

and Ha-Ras that binds and phosphorylates the c-Jun activation domain. *Cell* **1994**, *76*, 1025–1037.

(5) Kallunki, T.; Su, B.; Tsigelny, I.; Sluss, H. K.; Derijard, B.; Moore, G.; Davis, R.; Karin, M. JNK2 contains a specificity-determining region responsible for efficient c-Jun binding and phosphorylation. *Genes Dev.* **1994**, *8*, 2996–3007.

(6) Gupta, S.; Barrett, T.; Whitmarsh, A. J.; Cavanagh, J.; Sluss, H. K.; Dérijard, B.; Davis, R. J. Selective interaction of JNK protein kinase isoforms with transcription factors. *EMBO J.* **1996**, *15*, 2760–2770.

(7) Stebbins, J. L.; De, S. K.; Machleidt, T.; Becattini, B.; Vazquez, J.; Kuntzen, C.; Chen, L.-H.; Cellitti, J. F.; Riel-Mehan, M.; Emdadi, A.; Solinas, G.; Karin, M.; Pellecchia, M. Identification of a new JNK inhibitor targeting the JNK-JIP interaction site. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16809–16813.

(8) Ip, Y. T.; Davis, R. J. Signal transduction by the c-Jun N-terminal kinase (JNK)—from inflammation to development. *Curr. Opin. Cell Biol.* **1998**, *10*, 205–219.

(9) Leppä, S.; Bohmann, D. Diverse functions of JNK signaling and c-Jun in stress response and apoptosis. *Oncogene* **1999**, *18*, 6158–6162.

(10) Minden, A.; Karin, M. Regulation and function of the JNK subgroup of MAP kinases. *Biochim. Biophys. Acta* **1997**, *1333*, F85–F104.

(11) Whitmarsh, A. J.; Davis, R. J. Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways. *J. Mol. Med.* **1996**, *74*, 589–607.

(12) Manning, A. M.; Davis, R. J. Targeting JNK for therapeutic benefit: From junk to gold? *Nat. Rev. Drug Discovery* **2003**, *2*, 554–565.

(13) Harding, T. C.; Xue, L.; Bienemann, A.; Haywood, D.; Dickens, M.; Tolkovsky, A. M.; Uney, J. B. Inhibition of JNK by Overexpression of the JNK Binding Domain of JIP-1 Prevents Apoptosis in Sympathetic Neurons. *J. Biol. Chem.* **2001**, *276*, 4531–4534.

(14) Whitmarsh, A. J.; Kuan, C.-Y.; Kennedy, N. J.; Kelkar, N.; Haydar, T. F.; Mordes, J. P.; Appel, M.; Rossini, A. A.; Jones, S. N.; Flavell, R. A.; Rakic, P.; Davis, R. J. Requirement of the JIP1 scaffold protein for stress-induced JNK activation. *Genes Dev.* **2001**, *15*, 2421–2432.

(15) Vaishnav, M.; MacFarlane, M.; Dickens, M. Disassembly of the JIP1/JNK molecular scaffold by caspase-3-mediated cleavage of JIP1 during apoptosis. *Exp. Cell Res.* **2011**, *317*, 1028–1039.

(16) Pei, J.; Braak, E.; Braak, H.; Grundke-Iqbal, I.; Iqbal, K.; Winblad, B.; Cowburn, R. Localization of active forms of C-jun kinase (JNK) and p38 kinase in Alzheimer's disease brains at different stages of neurofibrillary degeneration. *J. Alzheimer's Dis.* **2001**, *3*, 41–48.

(17) Shoji, M.; Iwakami, N.; Takeuchi, S.; Waragai, M.; Suzuki, M.; Kanazawa, I.; Lippa, C.; Ono, S.; Okazawa, H. JNK activation is associated with intracellular beta-amyloid accumulation. *Mol. Brain Res.* **2000**, *85*, 221–233.

(18) Yoon, S. O.; Park, D. J.; Ryu, J. C.; Ozer, H. G.; Tep, C.; Shin, Y. J.; Lim, T. H.; Pastorino, L.; Kunwar, A. J.; Walton, J. C.; Nagahara, A. H.; Lu, K. P.; Nelson, R. J.; Tuszynski, M. H.; Huang, K. JNK3 perpetuates metabolic stress induced by A β peptides. *Neuron* **2012**, *75*, 824–837.

(19) Zhu, X.; Raina, A. K.; Rottkamp, C. A.; Aliev, G.; Perry, G.; Bux, H.; Smith, M. A. Activation and redistribution of c-jun N-terminal kinase/stress activated protein kinase in degenerating neurons in Alzheimer's disease. *J. Neurochem.* **2001**, *76*, 435–441.

(20) Ittner, L. M.; Ke, Y. D.; Götz, J. Phosphorylated Tau interacts with C-Jun N-terminal kinase-interacting protein 1 (JIP1) in Alzheimer disease. *J. Biol. Chem.* **2009**, *284*, 20909–20916.

(21) Ballatore, C.; Lee, V. M.-Y.; Trojanowski, J. Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat. Rev. Neurosci.* **2007**, *8*, 663–672.

(22) Orejana, L.; Barros-Miñones, L.; Aguirre, N.; Puerta, E. Implication of JNK pathway on tau pathology and cognitive decline in

a senescence-accelerated mouse model. *Exp. Gerontol.* **2013**, *48*, 565–571.

(23) Reynolds, C. H.; Utton, M. A.; Gibb, G. M.; Yates, A.; Anderton, B. H. Stress-activated protein kinase/c-jun N-terminal kinase phosphorylates tau protein. *J. Neurochem.* **1997**, *68*, 1736–1744.

(24) Margevicius, D.; Bastian, C.; Fan, Q.; Davis, R.; Pimplikar, S. JNK-interacting protein 1 mediates Alzheimer's-like pathological features in AICD-transgenic mice. *Neurobiol. Aging* **2015**, *36*, 2370–2379.

(25) Pellet, J.-B.; Haefliger, J.-A.; Staple, J. K.; Widmann, C.; Welker, E.; Hirling, H.; Bonny, C.; Nicod, P.; Catsicas, S.; Waeber, G.; Riederer, B. M. Spatial, temporal and subcellular localization of islet-brain 1 (IB1), a homologue of JIP-1, in mouse brain. *Eur. J. Neurosci.* **2000**, *12*, 621–632.

(26) Barr, R. K.; Kendrick, T. S.; Bogoyevitch, M. A. Identification of the critical features of a small peptide inhibitor of jnk activity. *J. Biol. Chem.* **2002**, *277*, 10987–10997.

(27) Dickens, M.; Rogers, J. S.; Cavanagh, J.; Raitano, A.; Xia, Z.; Halpern, J. R.; Greenberg, M. E.; Sawyers, C. L.; Davis, R. J. A cytoplasmic inhibitor of the JNK signal transduction pathway. *Science* **1997**, *277*, 693–696.

(28) Ho, D. T.; Bardwell, A. J.; Abdollahi, M.; Bardwell, L. A docking site in MKK4 mediates high affinity binding to JNK MAPKs and competes with similar docking sites in JNK substrates. *J. Biol. Chem.* **2003**, *278*, 32662–32672.

(29) Zhao, H.; Serby, M. D.; Xin, Z.; Szczepankiewicz, B. G.; Liu, M.; Kosogof, C.; Liu, B.; Nelson, L. T.; Johnson, E. F.; Wang, S.; Pederson, T.; Gum, R. J.; Clampitt, J. E.; Haasch, D. L.; Abad-Zapatero, C.; Fry, E. H.; Rondinone, C.; Trevillyan, J. M.; Sham, H. L.; Liu, G. Discovery of Potent, Highly Selective, and Orally Bioavailable Pyridine Carboxamide c-Jun NH₂-Terminal Kinase Inhibitors. *J. Med. Chem.* **2006**, *49*, 4455–4458.

(30) Heo, Y.-S.; Kim, S.-K.; Seo, C. I.; Kim, Y. K.; Sung, B.-J.; Lee, H. S.; Lee, J. I.; Park, S.-Y.; Kim, J. H.; Hwang, K. Y.; Hyun, Y.-L.; Jeon, Y. H.; Ro, S.; Cho, J. M.; Lee, T. G.; Yang, C.-H. Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125. *EMBO J.* **2004**, *23*, 2185–2195.

(31) Oates, M. E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M. J.; Xue, B.; Dosztányi, Z.; Uversky, V. N.; Obradovic, Z.; Kurgan, L.; Dunker, A. K.; Gough, J. D²P²: database of disordered protein predictions. *Nucleic Acids Res.* **2012**, *41*, D508–D516.

(32) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta) Physics. *Front. Phys.* **2019**, *7*, 10.

(33) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29.

(34) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*, 99–105.

(35) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.

(36) Schütte, C.; Fischer, A.; Huisinga, W.; Deuffhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146–168.

(37) Swope, W. C.; Pitera, J. W.; Suits, F. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.

(38) 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*, 174105.

(39) Berman, H. B.; Westbrook, J.; Feng, Z.; Gilliland, H.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank Nucleic Acids Research. **2000**, *28*, 235–242, DOI: 10.1093/nar/28.1.235.

(40) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(41) Case, D. A.; Babin, V.; Berryman, J.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *Amber* 14; 2014.

(42) Roe, D. R.; Cheatham, T. E., III PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(43) Jorgensen, W. L.; Madura, J. D. Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water. *J. Am. Chem. Soc.* **1983**, *105*, 1407–1413.

(44) Hayre, N. R.; Singh, R. R. P.; Cox, D. L. Evaluating force field accuracy with long-time simulations of a β -hairpin tryptophan zipper peptide. *J. Chem. Phys.* **2011**, *134*, No. 035103.

(45) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(46) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878.

(47) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

(48) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(49) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(50) Schrödinger, L. L. C. *The PyMOL Molecular Graphics System*; Version 1.8. 2015.

(51) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(52) Miller, B. R., III; McGee, T. D., Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(53) Feig, M.; Karanicolas, J.; Brooks, C. L., III MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377–395.

(54) Lloyd, S. P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137.

(55) Wehmeyer, C.; Scherer, M. K.; Hempel, T.; Husic, B. E.; Olsson, S.; Noé, F. Introduction to Markov state modeling with the PyEMMA software [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1*, 5965.

(56) Perez-Hernandez, 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*, 015102.

(57) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* **2009**, *7*, 1192–1219.

(58) Prakashchand, D. D.; Ahalawat, N.; Bandyopadhyay, S.; Sengupta, S.; Mondal, J. Nonaffine Displacements Encode Collective Conformational Fluctuations in Proteins. *J. Chem. Theory Comput.* **2020**, *16*, 2508–2516.

(59) Wolfe, D. K.; Persichetti, J. R.; Sharma, A. K.; Hudson, P. S.; Woodcock, H. L.; O'Brien, E. P. Hierarchical Markov State Model Building to Describe Molecular Processes. *J. Chem. Theory Comput.* **2020**, *16*, 1816–1826.

- (60) Noé, F.; Nüske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **2013**, *11*, 635–655.
- (61) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **1994**, *72*, 3634–3637.
- (62) Noé, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2015**, *11*, 5002–5011.
- (63) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542.
- (64) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci.* **2009**, *106*, 19011–19016.
- (65) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestoning. *J. Chem. Phys.* **2008**, *129*, 174102.
- (66) Grossfield, A.; Patrone, P. N.; Roe, D. R.; Schultz, A. J.; Siderius, D.; Zuckerman, D. M. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2018**, *1*, 5067.
- (67) Ruxton, G. D. The unequal variance *t*-test is an underused alternative to Student's *t*-test and the Mann–Whitney *U* test. *Behav. Ecol.* **2006**, *17*, 688–690.
- (68) Karain, W. I. Detecting transitions in protein dynamics using a recurrence quantification analysis based bootstrap method. *BMC Bioinform.* **2017**, *18*, 525.
- (69) Schinkel, S.; Marwan, N.; Dimigen, O.; Kurths, J. Confidence bounds of recurrence-based complexity measures. *Phys. Lett. A* **2009**, *373*, 2245–2250.
- (70) Noé, F.; Rosta, E. Markov Models of Molecular Kinetics. *J. Chem. Phys.* **2019**, *151*, 190401.
- (71) Harada, R.; Shigeta, Y. Hybrid Cascade-Type Molecular Dynamics with a Markov State Model for Efficient Free Energy Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 680–687.
- (72) Sidky, H.; Chen, W.; Ferguson, A. L. High-Resolution Markov State Models for the Dynamics of Trp-Cage Miniprotein Constructed Over Slow Folding Modes Identified by State-Free Reversible VAMPnets. *J. Phys. Chem. B* **2019**, *123*, 7999–8009.
- (73) Wu, H.; Noé, F. Variational Approach for Learning Markov Processes from Time Series Data. *J. Nonlinear Sci.* **2020**, *30*, 23–66.
- (74) 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.
- (75) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. Estimation and uncertainty of reversible Markov models. *J. Chem. Phys.* **2015**, *143*, 174101.
- (76) Deuffhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* **2005**, *398*, 161–184.
- (77) Kube, S.; Weber, M. A coarse graining method for the identification of transition rates between molecular conformations. *J. Chem. Phys.* **2007**, *126*, 024103.
- (78) Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data. Anal. Classif.* **2013**, *7*, 147–179.
- (79) Chu, B. K.; Tse, M. J.; Sato, R. R.; Read, E. L. Markov State Models of gene regulatory networks. *BMC Syst. Biol.* **2017**, *11*, 14.
- (80) Berezhkovskii, A.; Hummer, G.; Szabo, A. Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J. Chem. Phys.* **2009**, *130*, 205102.
- (81) Weinan, E.; Vanden-Eijnden, E. Towards a theory of transition paths. *J. Stat. Phys.* **2006**, *123*, 503–523.