

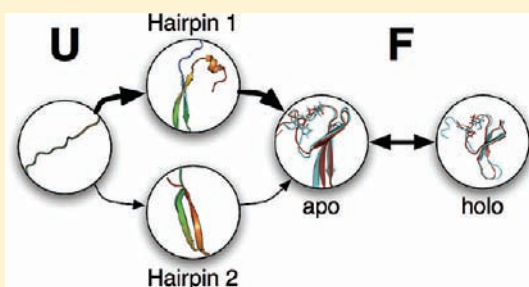
Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories

Thomas J. Lane,[†] Gregory R. Bowman,[†] Kyle Beauchamp,[‡] Vincent A. Voelz,^{†,§} and Vijay S. Pande^{*,†}

[†]Department of Chemistry and [‡]Biophysics Program, Stanford University, Stanford, California 94305, United States

S Supporting Information

ABSTRACT: Two strategies have been recently employed to push molecular simulation to long, biologically relevant time scales: projection-based analysis of results from specialized hardware producing a small number of ultralong trajectories and the statistical interpretation of massive parallel sampling performed with Markov state models (MSMs). Here, we assess the MSM as an analysis method by constructing a Markov model from ultralong trajectories, specifically two previously reported 100 μ s trajectories of the F1P35 WW domain (Shaw, D. E. et al. *Science* 2010, 330, 341–346). We find that the MSM approach yields novel insights. It discovers new statistically significant folding pathways, in which either beta-hairpin of the WW domain can form first. The rates of this process approach experimental values in a direct quantitative comparison (time scales of 5.0 μ s and 100 ns), within a factor of ~ 2 . Finally, the hub-like topology of the MSM and identification of a *holo* conformation predicts how WW domains may function through a conformational selection mechanism.



INTRODUCTION

A detailed understanding of protein dynamics is necessary for a complete picture of biology at a molecular scale. Atomistic molecular dynamics (MD) simulations have the potential to yield a detailed kinetic description of these dynamics but remain extremely computationally challenging. Moreover, even if the computational challenges can be surmounted, a new obstacle emerges: the efficient and accurate interpretation of massive MD data sets. Ideally, one could employ a method allowing for facile interpretation of the data, including comparison to experiment for theoretical validation.

Recently, advances have been made in both generation and analysis of MD simulations. To overcome the issue of computational intractability, our group and others have turned to the statistical approach of massively parallel simulation.¹ To unify the information acquired from independent parallel simulations, Markov state models (MSMs) have been employed for analysis. MSMs are discrete-time models based on kinetic exchange between states, which represent a partitioning of phase space into metastable regions.^{2–4} This approach has been largely successful, documented in recent reports of the kinetics of moderately sized proteins (80+ residues) at long time scales (microseconds to milliseconds).^{5–7}

Another approach is to design specialized hardware capable of producing single, long MD trajectories. Recently, this concept has been implemented by D. E. Shaw et al. in the form of Anton, a powerful and innovative supercomputer fine-tuned for the calculation of MD simulations.⁸ Anton is unique in its ability to calculate single, long trajectories, the first of which have been recently reported. During these simulations, multiple folding events of a WW domain, F1P35, were observed, the analysis of

which led to the conclusion of a single dominant folding pathway.⁹ This conclusion was particularly noteworthy because it was in disagreement with previous simulation work on WW domains demonstrating multiple distinct significant folding mechanisms,^{10,11} including ref 10, which was a pioneering study in the application of MSM theory that showed heterogeneous pathways in the folding of a WW domain. This conclusion also contradicted the general view that protein folding occurs through heterogeneous pathways, which has prevailed for over a decade.^{12–14}

Here, we reanalyze these ultralong F1P35 trajectories and demonstrate that MSMs can accurately capture long-time scale dynamics and provide a powerful foundation for analysis. A Markov model with short memory (100 ns Markov lag time) built from two 100 μ s trajectories of the F1P35 WW domain recapitulates the dynamics of the original trajectories, up to the ~ 10 μ s-scale folding time, the slowest process observed. The MSM, which identifies transitions through statistical inference, reveals many parallel folding pathways with heterogeneous molecular mechanisms. While a majority of the folding flux flows through the path reported by Shaw et al., we find other, distinct routes to the global free energy minimum, suggesting that the folding mechanism of WW domains is heterogeneous, a result previously found by Noé et al., who employed MSM theory.¹⁰ The time scales of this process match experiment to within a factor of 2, recapturing the rates and amplitudes of the biexponential process seen in T-jump fluorescence experiments.^{15,16}

Received: August 16, 2011

Published: October 11, 2011

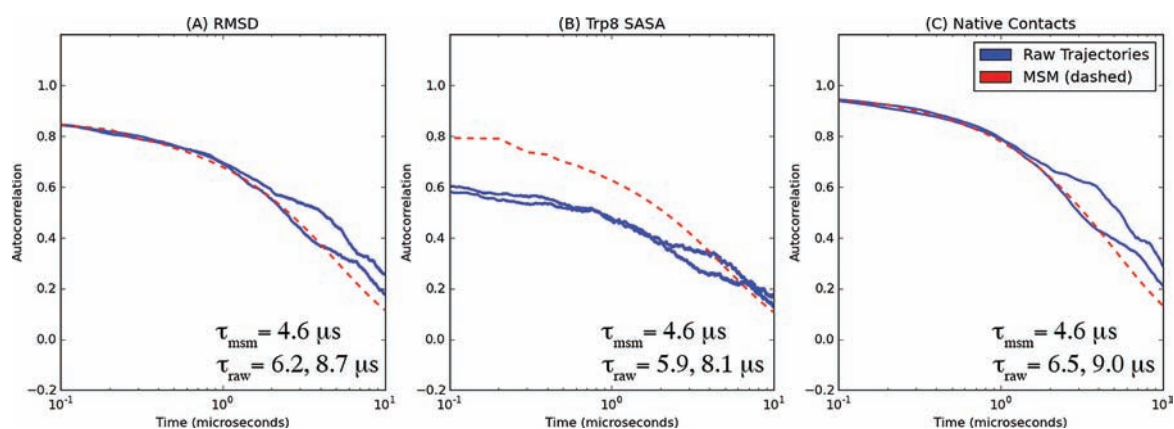


Figure 1. Autocorrelation functions of (A) rmsd to the native state, (B) Trp8 SASA, and (C) number of native contacts. Traces are overlaid for the two raw trajectories (blue, solid) and the Markov Chain relaxation (red, dashed). Each of these trajectories was initiated in the same state and propagated for 100 μs . Each autocorrelation was fit to a single exponential—the relaxation constants of this fit are indicated on the appropriate panel.

In addition to these folding dynamics, we report on the *functional dynamics* of the folded protein. That is, once a protein reaches the native state, how do we expect it to carry out its biological function? As in other systems, we have found that the free energy landscape near the native state is a kinetic hub,¹⁷ with many rapidly interconverting native-like states. One native-like state is especially interesting, as it corresponds to *holo* conformers of the Pin1 WW domain, the wild-type parent of Fip35 that binds proline-rich peptides.

Finally, we explicitly show how the MSM approach, a powerful sampling technique,⁵ is also a powerful analysis tool, representing an advance over traditional techniques in three major ways. First, states are defined based on physical criteria, rather than intuition. Therefore, one can objectively assess how many relevant folding pathways are present rather than assuming a single dominant mechanistic pathway *a priori*. Second, the MSM allows one to simulate experiments and calculate interesting properties such as committor values without running additional costly simulations. Finally, the MSM provides an intuitive framework for understanding of dynamics; for instance, key features, such as the native and *holo* states, are easy to identify. Each of these aspects has been leveraged to yield new insight into the dynamics of Fip35.

METHODS

Two-hundred microseconds of atomistic MD simulation (Amber ff99SB-ILDN force field,¹⁸ TIP3P solvent¹⁹) was previously performed by Shaw et al.⁹ using the Anton supercomputer. A Markov State Model (MSM) was constructed by clustering all saved 10^6 snapshots of this simulation (200 ps intervals). A 26 104 state model was generated with a *k*-centers algorithm, ensuring that no cluster had a larger radius than 4.5 Å. This radius cutoff ensured that equilibrium properties were preserved but maximized the resolution of the model (Figure S1, Supporting Information). Ten rounds of local *k*-medoids was then used to improve this model. The model was assessed for Markovian behavior based on the criterion of implied time scale invariance (Figure S1, Supporting Information). A 200-state macrostate model was constructed from a PCCA-lumping²⁰ of a second 10 000 microstate *k*-medoids model. This second model was used to generate the macrostate model because *k*-medoids provides a good partitioning of native dynamics, allowing us to distinguish the detailed native kinetics and identify the *holo* conformer. All clustering and data analysis was performed using the MSMBuild2 software package (<http://simtk.org/home/msmbuilder>), documented

in ref 21. This reference and the references within contain detailed information on the model construction practices employed here.

RESULTS

Models. Markov state models consist of a decomposition of phase space into discrete metastable states, and a master equation describing the kinetics between those states represented by a transition matrix estimated from simulation. From the raw trajectories, two models were created, a microstate model (26 104 states) and macrostate model (200 states).²² The microstate model is used for quantitative calculation, while the macrostate model has been employed in visualization and qualitative analysis. See Methods for details on model construction.

Recapitulation of the Raw Data. It is essential that a Markov Model parametrized from MD simulation recovers the kinetics of the simulation. Ultralong trajectories give a unique opportunity to test if MSMs can recapitulate long-time scale dynamics. Autocorrelation functions provide a direct way to determine dynamics, and have been used here to compare the raw data and MSM. The autocorrelations of rmsd to the native state, Trp8 solvent-exposed surface area (SASA), and native contacts show that the MSM recapitulates the raw data (Figure 1). Interestingly, there is still significant deviation between the two trajectories themselves, indicating that additional sampling (beyond two 100 μs trajectories) is necessary to draw statistically robust estimates of the kinetic properties of the Fip35 system.

Folding Time. Given that the model recapitulates the simulation, we now turn to the question of whether the simulation is an accurate representation of reality. Based on their definitions of the unfolded and folded states, Shaw et al. reported a folding time of $10 \pm 3 \mu s$. However, the rate calculation is highly dependent on human-defined states (Figure S2, Supporting Information), a common problem for prediction of rates directly from MD simulation with traditional (non-MSM) methods.¹³ The MSM, based on states defined from a *kinetic* perspective in an algorithmic manner, provides an objective solution to this problem. In particular, the MSM allows one to move away from relying on intuition for state boundaries and determine a physical state decomposition.

We take advantage of this to a mimic the temperature-jump experiment used to determine the folding time of Fip35^{15,16} without performing additional MD simulation. This is accomplished

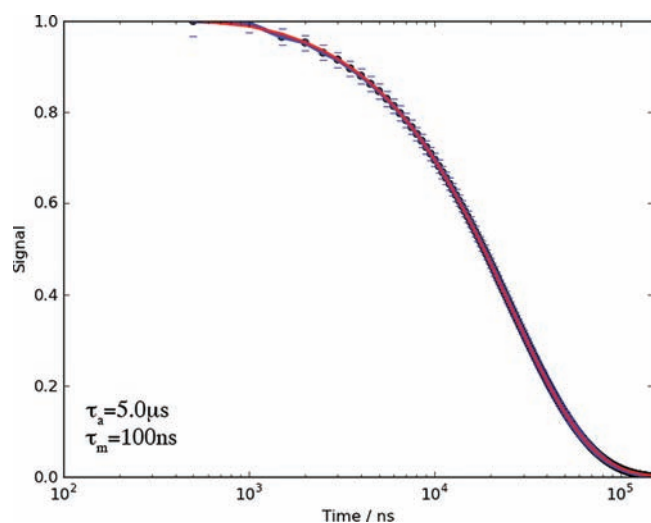


Figure 2. Simulation of a T-jump experiment. Ensemble averages of the Trp8 SASA are shown in blue, with a double-exponential fit to the data shown in red. The fit results in two time scales of 5.0 μ s and 100 ns, respectively. Error bars show one standard deviation around the mean, calculated from an averaging of 100 stochastically generated initial population vectors. The signal intensity has been normalized to span [0,1].

by stochastically perturbing the equilibrium populations of the microstates and watching the Markov chain relax back to detailed balance over time, mimicking a relaxation to equilibrium after perturbation by a T-jump. The T-Jump perturbation was mimicked by choosing two states at random, and shifting 25% of the population of one state to the other, repeating this process a number of times equal to the number of states in the model. 100 separate T-Jumps were performed and averaged to obtain the final result. The time scales observed in the T-Jump were robust to the precise perturbation used—specifically, the two time scales were observed regardless of the perturbation.

Projecting the population at each time onto an observable, here the Trp8 solvent-accessible surface area (SASA, a proxy for fluorescence intensity), we generate a time trace approximately proportional to measured fluorescence over time. This projection procedure amounts to calculating the average Trp8 SASA for the entire ensemble of states as the system relaxes from some random distribution of microstate populations to equilibrium. A double exponential with time scales of 5.0 μ s and 100 ns accounts for all variation in the data (Figure 2).

Experimentally, two time scales were found and fit in T-jump experiments for FiP35, and reported as activated and molecular rates.¹⁶ Our multiexponential fit can be interpreted as corresponding to these rates, which were reported to be about 11 μ s and 150 ns at 381K, the highest temperature measured.^{15,16} Additionally, the relative amplitudes of the two processes match in both theory and experiment, at ~ 0.8 and ~ 0.2 , for the slow and fast processes, respectively. This correspondence is evidence that the underlying dynamics observed in the model are the same as those in experiment, though definite proof of this cannot be obtained from a single observable.

By analyzing the eigenvectors and eigenvalues of the MSM transition matrix, one can obtain information about the dynamics of the system. The eigenvectors correspond to transitions between the states, and the eigenvalues give the time scales of those processes.^{3,4} By analyzing the four slowest eigenprocesses,

we find that the 5 μ s transition corresponds to folding (Figure 3). The next eigenvectors show quick dynamics within the non-native state, including dynamics between states where Trp8 is either highly buried or very exposed, but no productive folded structure is present. We found no one eigenprocess directly accounts for the second time scale observed in the T-Jump, which has contributions from non-native state dynamical processes that occur over an order of magnitude in time scales (Figure S3, Supporting Information). We conclude that the faster time scale measured in the T-Jump is in fact due to rearrangements between unfolded states, rather than some fast process occurring along the folding pathway.

Folding Pathways. We find the folding of FiP35 to be complex, exhibiting a diverse set of folding mechanisms (Figure 4). The macrostate MSM reveals many parallel pathways to the native state, which is characterized by two beta-hairpins and a hydrophobic Pro5-Pro33 core.

Looking for a pathway directly in MD simulations is difficult. Since systems typically exhibit rapid fluctuations and spend a great deal of time exploring off-pathway (orthogonal) motions, reductionist analysis methods are necessary. Additionally, pathways are only traveled a finite number of times during the simulation, making general extrapolation difficult. The MSM addresses these problems by projecting out fast degrees of freedom, keeping relevant motions we care about, and providing a statistical basis of observation. To elucidate the most important transitions, we employ transition path theory (TPT),^{23,24} which represents a kinetic process as reactive flux along pathways of interest.

TPT analysis of FiP35 reveals many parallel, heterogeneous folding pathways traveled with high probability. Despite this heterogeneity, some generalizations can be made. During folding, backbone secondary structure forms first—either of the two beta hairpin collapse, followed by the other—forming a molten globule. Analysis of the folding flux in the microstate MSM gives the probabilities that either hairpin forms first. By calculating the total flux along all folding pathways, defined as routes from states with no hairpins formed to those with both formed (by DSSP²⁵), we determined the relative probabilities for either hairpin to form first. Formation of the N-terminal (first) beta hairpin first was most common, occurring in 39% of folding events, while the second hairpin formed first in 16% of pathways. Due to the discrete nature of MSMs, and the long lag time employed in this particular model, in 49% of transitions both hairpins form concurrently in a single state-to-state transition. These transitions are therefore concerted, with both hairpins forming within one model lag time (100 ns).

Corroborating the heterogeneous paths found by the MSM, direct inspection of the trajectories shows instances of both pathways. A direct analysis predicts approximately a 4:1 ratio in hairpin formation, based on secondary structure assignments of folding events. Additionally, anecdotes of each hairpin's formation preceding folding have been found in the raw simulation (Figure S4, Supporting Information). Our results show the folding of FiP35 does not always occur along a single pathway, but is heterogeneous, a finding previously reported by this group and others.^{10,11,26}

Hub-like Nature of Native Structures. Our analysis of the folding pathway reveals that folding flux is channeled through many parallel hubs on the way to a native conformation. In the macrostate MSM, all the states with native structure are highly connected and interconvert rapidly, on the order of 100s of ns.

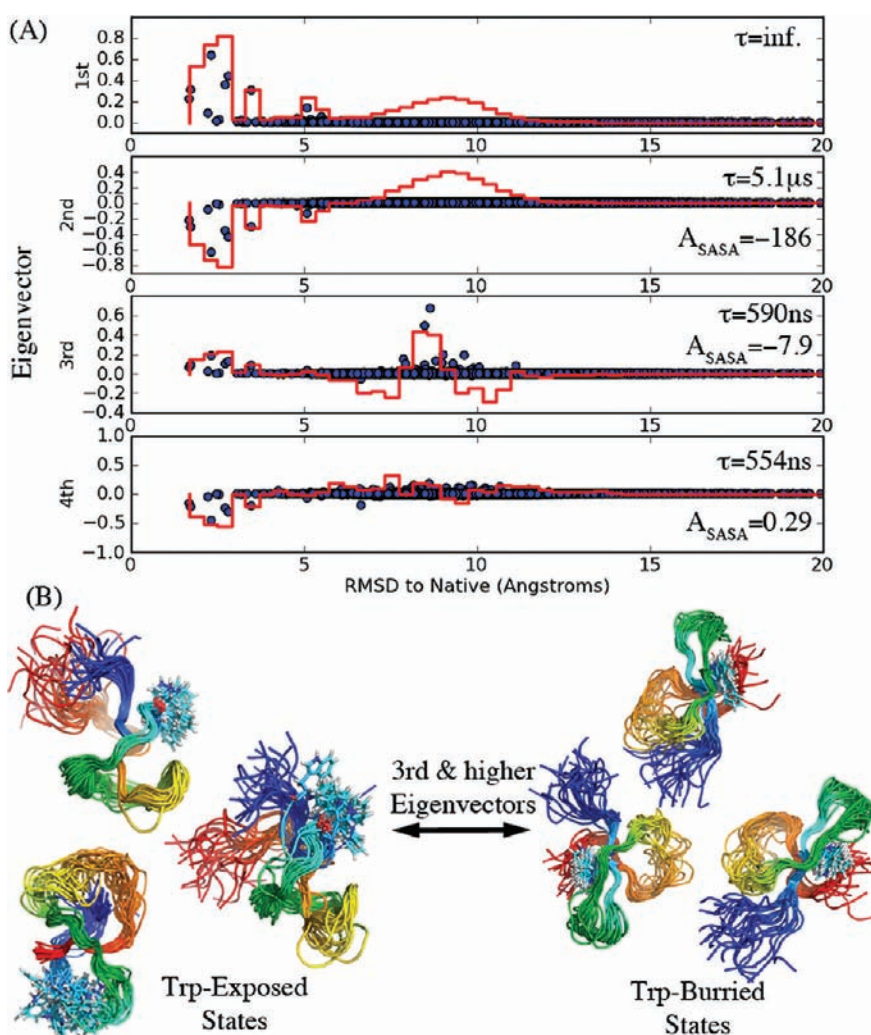


Figure 3. Eigenprocesses occurring in the transition matrix. The eigenvectors of the MSM transition matrix give the dynamical processes occurring in that matrix, while the eigenvalues give the time scales on which those processes occur. (A) Plot of the eigenvectors projected onto rmsd to native. Blue dots are individual states, red lines are weighted histograms showing the binned sum of the individual state's contributions to the eigenvector. (B) Illustration of six example states from the 3rd and 4th eigenprocesses, which are dominated by interconversions between unfolded states. Many of these unfolded states have large deviations in their Trp8 exposed surface area. Trp8 shown in sticks.

These internal transitions occur 2 orders of magnitude faster than those between non-native states (10s of μs), as measured by mean-first-passage times (MFPT), which provide a qualitative measure of the distance between states. Additionally, the average MFPTs from non-native to native macrostates show transitions to the native state are rapid (on the order of $1 \mu s$, not weighted by equilibrium population). The disparity between the $1 \mu s$ MFPT and the $11 \mu s$ folding rate found in the T-Jump (Figure 2, Figure 3) is a result the macrostate model's artificially fast dynamics. This is a result of macrostates being too diverse to ensure kinetic similarity at a macroscopic level and is an intrinsic drawback of current methods to generate macrostate models. The qualitative features described here are valid, however, and are retained in the qualitatively accurate microstate model.

Quick transitions to the native conformation lead us to characterize our MSM as a network with the native states as a hub at the center (Figure 5). Any state within the model is close to a native-like state, and transitions from any one non-native state to any other non-native state are likely to pass through one of the native macrostates. This hub-like nature has been previously

noted in MSMs, and appears to be a recurrent theme.¹⁷ Any state can reach any other state in 4 moves or less, and 94% of states are connected by one state or less. This high degree of connectivity, along with the centrality of the native states are indicative of a hub-like kinetic network.

Identification of the Functionally Relevant Holo State. The rapid interconversions within the native network have been found to be important for function,²⁷ allowing a protein to reach a functionally relevant *holo* state. We have identified macrostates that correspond to the *apo* and *holo* conformers of a related mutant, the human Pin1 WW domain (hPin1 WW, Figure 5).²⁸

NMR solution structures^{28,29} of Pin1 WW correspond to the highlighted states in Figure 4 and Figure 5. The *apo* structure is structurally similar to all structures in the native state, within 3.0 \AA rmsd of each state. We have highlighted the macrostate with the lowest rmsd to the NMR structure as an *apo* exemplar for comparison purposes (Figure 4, Figure 5). Direct analysis of the key interactions in the *apo* state indicates that these features are common to all of the states in the native state besides the *holo* state, with slight differences due to mutation.

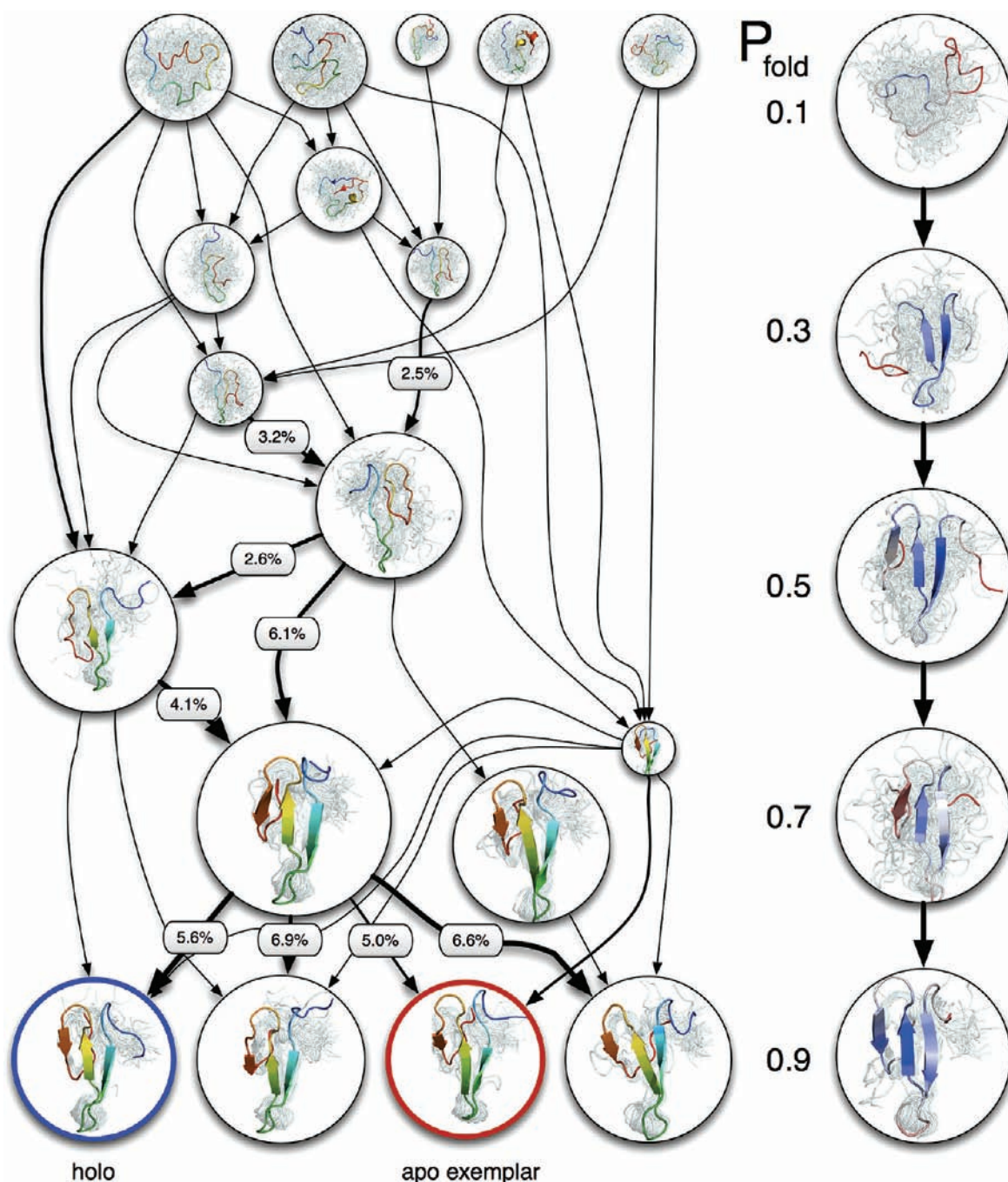


Figure 4. Folding of FiP35, with committor (P_{fold}) values determined from transition path theory. (Left) Top 12 pathways of the folding probability flux, showing heterogeneous, parallel folding pathways for FiP35. Arrow width is proportional to flux, node size is proportional to the free energy of the state, and the *apo* and *holo* states are highlighted in red and blue, respectively. (Right) Overlaid structures with the indicated P_{fold} , showing an alternate view of folding. The exemplar structure is highlighted with a temperature factor proportional to the structure of each residue, as measured by rmsd. Red indicates an unstructured residue, blue indicates a structured residue, and white is intermediate between the two.

More interesting is the *holo* state itself. Two available ligand-bound NMR structures of hPin1 WW correspond closely to a single macrostate (avg. $C\alpha$ rmsd of 1.8 and 1.9 Å, respectively). This *holo* macrostate looks qualitatively different than the rest of the observed native states, with Arg11 rotated inward, in a presentation very similar to that of the corresponding residue in ligand-bound hPin1 WW. The *holo* state is connected to all of the states in the N state, and has a high equilibrium population of 5.5%. This indicates the *holo* state is part of the functional

dynamics of the native basin, and is populated even in the absence of the ligand.

WW Domains bind proline-rich peptide sequences, and in the wild-type hPin1 WW, the dominant interaction occurs when Arg12, in the loop between sheets 1 and 2, rotates outward, binding the peptide. Simultaneously, Trp29 stabilizes this interaction (Figure 6).²⁹ The correspondence of hPin1 WW to FiP35 is not perfect, due to mutations made in FiP35 to speed folding. The *holo* topology is the same in both mutants, however, with the

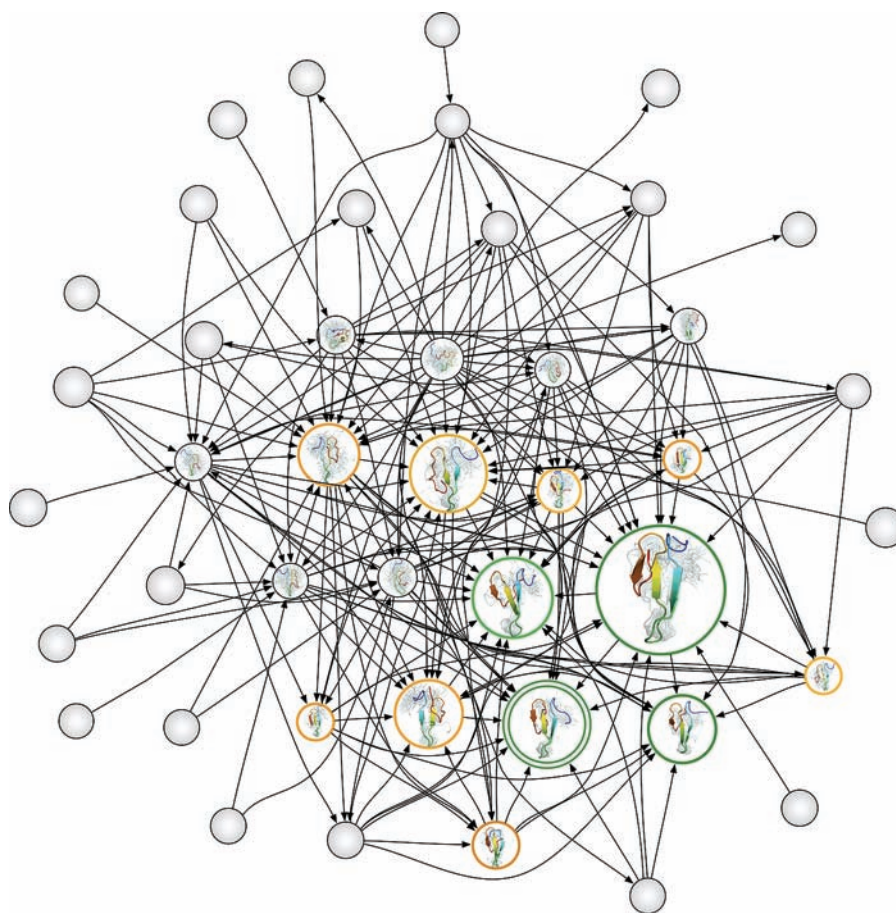


Figure 5. Hub-like nature of the FiP35 MSM. A subset of the macrostate MSM was represented as a graph, with centrality automatically determined by number of connections (Omnigraph software). We see that the native macrostates, in green, are centrally located, surrounded by molten globule states, in orange. The non-native states (uncolored) sit on the periphery of the graph, but can reach native states in few transitions, through hubs. The *holo* state is highlighted (double ring) and is centrally located in the hub-like network. Transitions to this state are relevant for biological function.

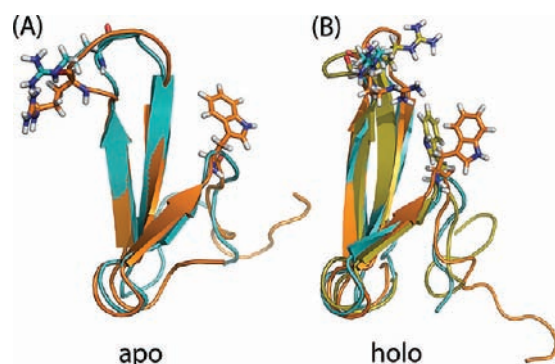


Figure 6. Structural overlays of the (A) *apo* (PDB: 1I6C) and (B) *holo* states (PDB: 1I8G, 1I8H). The NMR *holo* states here have ligands bound in the structure, a distinct ligand for each PDB entry, but ligands are not shown for clarity. NMR structures are shown in orange, MSM state exemplars are shown in teal. Key residues are highlighted: Arg12 (Arg11 in FiP35), along with Trp29 are responsible for ligand binding.

appropriate key twist in the hairpin between sheets 1 and 2, rotating, in FiP35, Arg11 outward. FiP35 lacks a second Trp residue due to mutation; such mutations have been demonstrated to destroy any ligand-binding activity.³⁰ The native state dynamics of Pin1 WW have been previously studied using MSMs,

which showed native state dynamics similar to what is seen here, with some differences due to mutation.³¹ These studies suggest that WW domains may bind ligands with a conformational selection mechanism, though definite conclusions cannot be made since these studies have spanned only two mutants and have not explicitly simulated ligand binding.

DISCUSSION

Previously, MSMs combined with parallel simulation have effectively overcome the intractability of simulating phenomena involving millisecond or longer time scales in systems with tens of thousands of atoms. With Anton, however, a new opportunity to directly produce long, serial trajectories has arisen. Here, we show an MSM built from serial data can yield insight not possible with traditional methods. Our results indicate that there appears to be little difference between Markovian models parametrized from few long or many short trajectories.

Moreover, the MSM's probabilistic interpretation is shown to significantly aid analysis of long trajectories, statistically leveraging the data to find transitions missed by other analysis methods. In addition to extracting all possible information from existing data, the MSM provides a powerful, physical foundation for comprehending simulation. In the study of FiP35, an MSM allows us to analyze folding as a probability flux through multiple

pathways, illuminating the hub-like structure of the equilibrium distribution.

Finally, in addition to capturing long time scale folding dynamics, the MSM allows for facile analysis of the protein's functional dynamics. FiP35 is shown to present itself in a *holo* conformation, even in the absence of any ligand. This presentation is of the sort predicted by the conformational selection model of ligand binding, as opposed to an induced fit model,²⁷ though this simulation represents only one anecdotal observation, and should not be overgeneralized.

This work indicates that MSMs provide a powerful analysis tool not only for many short simulations, but also for long trajectories. Given this, future work will focus on the most efficient manner of generating MD simulations to construct Markov models. We suspect that clever combinations of long and short trajectories will be synergistic; shorter simulations are considerably more amenable to parallelization, but due to their short nature individually only explore limited regions of phase space and require methods such as adaptive sampling⁵ to collectively sample the required space. Long trajectories, while expensive to generate, do not suffer from this limitation. Therefore, it is possible to imagine that a combination of the two would lead to models that are both well sampled and cover phase space completely. Future work is needed to refine and test this hypothesis.

■ ASSOCIATED CONTENT

S Supporting Information. Complete ref 8. Four supplemental figures are included, depicting (1) the implied time scales and model error as a function of lag time, (2) the lack of robustness in the folding time as a function of state definition, (3) the projection of the transition matrix eigenvectors onto the Trp8 SASA observable, and (4) raw data depicting heterogeneous pathways. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author
pande@stanford.edu

Present Addresses

⁵Department of Chemistry, Temple University, Philadelphia, Pennsylvania, 19121, United States.

■ ACKNOWLEDGMENT

We kindly thank Stefano Piana-Agostinetti, Kresten Lindorff-Larsen and their co-workers at DE Shaw Research for providing the two trajectories and a critical reading of the manuscript. Thanks also to Jesus Izaguirre and Christian Schwantes for a many comments and to Liz Kellogg for spotting an error prior to publication. Computational resources were funded by the NSF (MRI-R2-0960306). V.S.P. acknowledges support from NIH R01-GM06-2868, NSF-DMS-0900700, NSF-MCB-0954714. V.S.P. and V.A.V. acknowledge support from NSF EF-0623664. T.J.L. was supported by an NSF GRF. G.R.B. was funded by the Berry Foundation.

■ REFERENCES

(1) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. *Biopolymers* **2003**, *68*, 91–109.

- (2) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *J. Chem. Phys.* **2009**, *131*, 124101.
- (3) Schütte, C.; Fischer, A.; Huisinga, W.; Deuffhard, P. *J. Comput. Phys.* **1999**, *151*, 146–168.
- (4) (a) Noé, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162. (b) Prinz, J. H.; Keller, B.; Noé, F. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912–16927.
- (5) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *133*, 664–667.
- (6) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526–1528.
- (7) Bowman, G. R.; Ensign, D. L.; Pande, V. S. *J. Chem. Theory Comput.* **2010**, *6*, 787–794.
- (8) Shaw, D. E.; et al. *ACM SIGARCH Comput. Archit. News* **2007**, *35*, 1.
- (9) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wrighers, W. *Science* **2010**, *330*, 341–346.
- (10) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011–19016.
- (11) Xu, J.; Huang, L.; Shakhnovich, E. I. *Proteins* **2011**, *79*, 1704–1714.
- (12) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70–75.
- (13) Ensign, D. L.; Kasson, P. M.; Pande, V. S. *J. Mol. Biol.* **2007**, *374*, 806–816.
- (14) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Ann. Rev. Biophys.* **2008**, *9*, 289–316.
- (15) Liu, F.; Du, S.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2369–2374.
- (16) Liu, F.; Nakaema, M.; Gruebele, M. *J. Chem. Phys.* **2009**, *131*, 195101.
- (17) Bowman, G. R.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10890–10895.
- (18) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78* (8), 1950–1958.
- (19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (20) Deuffhard, P.; Huisinga, W.; Fischer, A.; Schütte, C. *Lin. Alg. Appl.* **2000**, *315*, 39–59.
- (21) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I.; Pande, V. S. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.
- (22) Bowman, G. R.; Huang, X.; Pande, V. S. *Methods* **2009**, *49*, 197–201.
- (23) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. *Multiscale Model. Simul.* **2009**, *7*, 1192.
- (24) Berezhkovskii, A.; Hummer, G.; Szabo, A. *J. Chem. Phys.* **2009**, *130*, 205102.
- (25) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- (26) Ensign, D. L.; Pande, V. S. *Biophys. J.* **2009**, *96*, 53–55.
- (27) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789–796.
- (28) Wintjens, R.; Wieruszkeski, J.-M.; Drobecq, H.; Rousselot-Pailley, P.; Buée, L.; Lippens, G.; Landrieu, I. *J. Biol. Chem.* **2001**, *276*, 25150–25156.
- (29) Peng, T.; Zintsmaster, J. S.; Namanja, A. T.; Peng, J. W. *Nat. Struct. Biol.* **2007**, *14*, 325–331.
- (30) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10648–10653.
- (31) 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. *PLoS Comp. Biol.* **2010**, *6*, e1001015.