Optimal Parameter Selection in Markov State Models for Biomolecular Conformational Dynamics

Robert T. McGibbon, Christian R. Schwantes, and Vijay S. Pande

Markov state models (MSMs) are a powerful tool for the analysis of molecular dynamics simulations, but have been hampered by the need for manual selection of the number of states. We report a new method for the optimal selection of the number of states in an MSM based on the Bayesian information criterion. We demonstrate the approach on three systems of increasing complexity...

I. INTRODUCTION

Protein are highly complex molecular machines, and their dynamics are an essential aspect of biomolecular function. These dynamics span a wide range of length scales, timescales and complexity, including folding and aggregation, conformational change between functional native substates, ligand binding, and allostery [1–4]. Whereas classical experimental probes have often been interpreted in two-state frameworks, ensemble measurements with increasingly high temporal resolution as well as sensitive single molecule probes have uncovered a vast array of complex multi-state kinetics [5, 6]. But the atomic-resolution characterization of these dynamics is often an enormous challenge – as molecular probes like Förster resonance energy transfer, small-angle x-ray scattering, and nuclear magnetic resonance techniques measure complex projections of the intrinsic structure, generally reporting simultaneously on many degrees of freedom[7, 8].

Computer simulations can complement experiments by providing atomic insight into conformational dynamics. With advances at the algorithmic, hardware, and software levels, modern molecular simulation paradigms, incorporating specialized or accelerated hardware, often in combination with highly parallel distributed computing frameworks, are capable of generating extensive simulation data sets[9–13]. In fact, the minimally-biased kinetic analysis of such simulations is often a central bottleneck and presents a major challenge to the field. The analysis paradigms often entail the construction of lower resolution models parametrized from the high resolution simulation data set which capture the essential features in an interpretable framework[14, 15]. For example, by projecting the data down onto one or two degrees of freedom we create a simpler model for the system, such as one characterized by diffusion along a single reaction coordinate[16].

Markov state models (MSMs) are one approach for analyzing MD data sets and driving further MD simulations that are able to smoothly move between high and low-resolution models[17–20]. Such detailed models maintain quantitative agreement with the underlying simulation data, while low-resolution models capture the salient features of the potential energy landscape, sacrificing some degree of model complexity. In an MSM, the dynamics are modeled as a memory-less jump process between a discrete set of conformational states. The two key quantities which define the MSM are thus the state definitions, an indicator function basis over phase space, and the pairwise transition probabilities or transition rates, which parameterize the kinetics. The matrix of transition probabilities can be used to locate the systems transition paths[21], and its dominant eigenvectors to identify the metastable states[22] and long-timescale dynamical modes[].

A significant challenge in the automated construction of Markov state models is the choice of the number of states[23]. Although classical Hamiltonian dynamics form a continuous-time Markov chain in $\mathbb{R}^{6N}$, the Markov property does not hold after the projecting the dynamics onto a basis of discrete indicator functions. In particular, when states contain within them free energy barriers of substantial magnitude, the validity of the Markov assumption begins to suffer considerably. While this source of modeling error can be addressed by increasing the number of microstates, the reduction in one error comes at the expense of the increase in another. This second source of error is statistical in origin. As the number of states in the model grows, so does the number of parameters required to completely specify the kinetic model between all pairs of states. Because the amount of data is constant, each additional parameter leads to a decrease in the amount of data available per model parameter, which makes the approach susceptible to over-fitting.

Here, we seek to build models that are *suitably* complex, given the data, yielding complex descriptions of the system only to the extent that their additional parameters are implied by the observed dynamics. To that end, we introduce a procedure for scoring the likelihood of an MSM, which, together with standard statistical model selection techniques, enables the optimal selection of the state space, which we express both in terms of the number of states and the clustering algorithm employed to group sampled conformations into states. This approach complements validation procedures performed primarily based on human intuition, such as Chapman-Kolmogorov tests, and enables the treatment of model selection as an optimization problem amenable to automated methods.

II. PRIOR WORK

In the context of MSMs, model selection has traditionally been performed by analyzing the self-consistency of

the models with respect to changes in the Markov lag time.

- Swope-Pitera http://pubs.acs.org/doi/abs/10.1021/jp037421y
- Noe http://www.pnas.org/content/106/45/19011
- Park and Pande
- Sergio
- Kellogg

After diagonalization of the transition matrix, the relaxation of any initial probability distribution can be described by a superposition of exponential relaxation processes, with a characteristic time given by $-\frac{1}{\ln \lambda_i}$.

III. LIKELIHOOD OF A MARKOV STATE MODEL

With the kinetic model expressed as a set of pairwise state to state transition probabilities at a given lag time, the likelihood of an ensemble of trajectories after projection into the indicator function basis is given simply by the product of the transition matrix elements along the observed trajectories. However, as we vary the number of states, it is not permissible to simply compare these likelihoods as part of an optimization of the state definitions. In doing so, the optimal model would always be the trivial one state model, whose computed likelihood is unity regardless of the data.

The appropriate likelihood is instead a path action in phase space, on which the discrete states are merely an indicator function basis. With $s(X)$ as the function mapping conformations into the indicator function basis set, $s : \mathbb{R}^{3N} \to \{1, 2, \ldots, K\}$, the likelihood can be written as

$$P[x_{0\ldots T-1}]dx^T = \prod_{i=0}^{T-1} T(s(x_i) \to s(x_{i+1})) \cdot \prod_{i=0}^{T} p_{s(x_i)}(x_i) \tag{1}$$

With a discrete, non-overlapping state space, the likelihood of a sampled trajectory can be decomposed into a product of terms of two types: the state to state transition probabilities, $T(s_i \to s_j)$, and so-called emission distributions of each state, the conditional probability of observing a conformation at a given location in phase space given that the conformation, x_t is within a certain state, $s(x_t)$.

For example, consider two Markov state models sharing the same transition matrix, T. In one model, the state emission distributions are highly peaked at specific locations in phase space, whereas in the other model the emission distributions are uniform over the volume of the states. If an observed trajectory does go through the first models' regions of high likelihood, it is appropriately termed a more likely model given the data.

This emission distribution can be potentially modeled in multiple ways, each of which, in combination with a transition matrix, parameterizes a different statistical model. We propose two possible distributions: first, one may model the conditional distribution $P(x_t|s(x_t))$ as a uniform distribution over the (hyper)volume of $s(x_t)$, as shown in eq. 2. This model has the advantage of appearing agnostic with respect to intra-state dynamics but is intractable in high dimensions when the states are defined by general polytopes, such as those produced by a data-driven Voronoi tessellation. A more tractable alternative is a Gaussian emission model where $P(x_t|s(x_t))$ is modeled as multivariate normal, where the pairwise overlap between the emission distributions is treated as a truncation error, as shown in eq. 2 and used in the popular X-means algorithm.[24] The truncation error steps from the fact that the emission probability distributions from every state assign nonzero probability to each conformation in a dataset, while the parameterization procedure for the Markov transition probabilities assumes each conformation to be uniquely assigned to a single state. Relaxing this approximation is possible with the forward-backward algorithm,[25] but is beyond the scope of this work.

$$P[x_{0\ldots T-1}]dx^T = \prod_{i=0}^{T-1} T(s(x_i) \to s(x_{i+1})) \cdot \prod_{i=0}^{T} \frac{1}{V_{s(x_i)}} \tag{2}$$

$$P[x_{0\ldots T-1}]dx^T = \prod_{i=0}^{T-1} T(s(x_i) \to s(x_{i+1})) \cdot \prod_{i=0}^{T} (2\pi)^{\frac{d}{2}} |\Sigma_{s(x_i)}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_i - \mu_{s(x_i)})' \Sigma^{-1}_{s(x_i)} (x_i - \mu_{s(x_i)})\right) \tag{3}$$

where $\mu_{s(x_i)}$ is the cluster center assigned to x_i and $\Sigma_{s(x_i)}$ is the corresponding covariance matrix. For simplicity, we follow Pelleg *et al.* [24] and employ a spherical Gaussian model with a single shared variance parameter across the states estimated by maximum likelihood, under which conditions eq. 3 reduces to

$$P[x_{0\ldots T-1}]dx^T = \prod_{i=0}^{T-1} T(s(x_i) \to s(x_{i+1})) \cdot \prod_{i=0}^{T} \left(2\pi\hat{\sigma}^2\right)^{-\frac{d}{2}} \exp\left(-\frac{||x_i - \mu_{s(x_i)}||^2}{2\hat{\sigma}^2}\right) \tag{4}$$

where

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{i=0}^{T} ||x_i - \mu_{s(x_i)}||^2 \tag{5}$$

where k is the number of states.

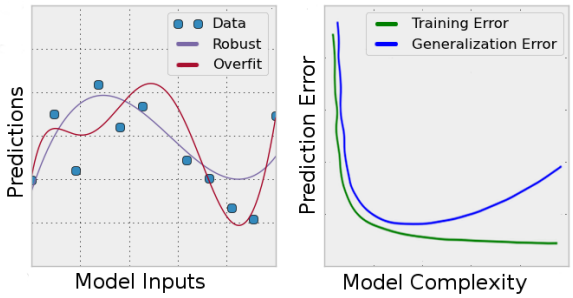

FIG. 1. An example of overfitting. Left: in the presence of noise, a more complex model is able to fit the observed training data better, with lower residues and a higher empirical likelihood, but fails to distinguish the signal from the noise. Right: more complex model classes exhibit lower training errors, but generalize poorly to unobserved data.

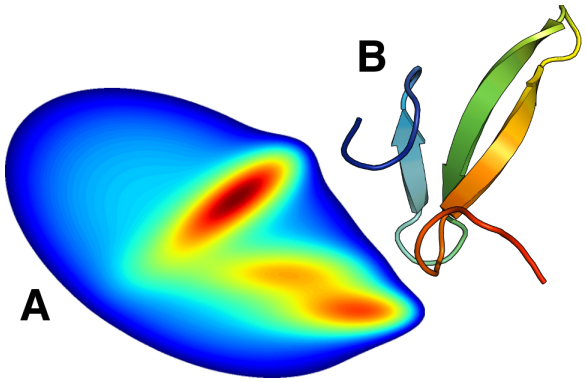

FIG. 2. Systems studied in this work. (A) Langevin dynamics on the two dimensional Müller potential. (b) 200 μs of dynamics of the Fip35 WW domain[31], courtesy of D.E. Shaw research [32].

IV. CROSS VALIDATION AND THE BAYES INFORMATION CRITERION

Likelihood maximization is insufficient for model selection when the number of parameters varies between proposed models, as more complex models generally exhibit higher empirical likelihoods, often at the cost of larger generalization errors due to overfitting[26, 27]. Statistical learning theory provides a number of alternative approaches for this problem. Conceptually, the most straightforward is a full Bayesian treatment in which all unknown model parameters are represented by probability distributions. The evidence for a model is computed by formally integrating over the model parameters and the evidence ratio, or Bayes factor[28], then provides a rigorous basis of model selection that appropriately punishes overly complex models as they become poorly constrained in parameter space. Unfortunately such approaches are intractable for problems of this size because of the need to integrate over all possible Markov models of a given size.

Instead, we explore three alternative procedures: cross validation, Schwarz's Bayesian information criterion (BIC)[29] and the Akaike information criterion (AIC)[30] for choosing the number of states in a Markov state model. In cross validation, we attempt to directly measure the Markov model's generalization error. We parameterize the model, building both the state space and transition matrix on a subset of the available molecular dynamics trajectories, but evaluate the likelihood on the left-out portion. The AIC and BIC are augmented likelihood function approaches which do not require leaving out portions of the available data during fitting, and instead employ asymptotic approximations to the Bayesian evidence to directly penalize models with additional free parameters.

$$\mathrm{BIC} \equiv -2 \cdot \ln L + k \cdot \ln N \tag{6}$$

$$\mathrm{AIC} \equiv -2 \cdot \ln L + 2 \cdot k \tag{7}$$

where L is the likelihood, k is the number of free parameters, and N is the number of data points, assumed to be independent and identically distributed.

V. COMPUTATIONAL METHODS

The uniform distribution emission model presents a computational challenge: its use requires the calculation of the (hyper)volume of the MSM's states, which, when defined by clustering are high-dimensional Voronoi cells. While trivial in two or three dimensions, this computational geometry task becomes challenging in high-dimensional settings. The computation of such volumes has occupied significant attention in recent years in the computational geometry literature, especially via randomized algorithms[33–35]. We opt to approximate the volumes using naive Monte Carlo rejection sampling, which we find tractable for large systems only when the molecular dynamics dataset is first projected into a suitable small vector space of up to perhaps ten dimensions.

A further challenge is the procedure by which to model the volume of states which are at the edge of the MSM – whose Voronoi cells extend to infinity in some direction. In these cases the Voronoi cells are of unbounded volume. Because the convex hull of our simulation datasets are computationally inaccessible, we define the outer extent of our data sets to be the set of all trial points such that the nearest sampled configuration to the trial point is closer than a certain cutoff, R.

For cross validation, we find it essential to use a dense prior during transition matrix estimation. Because the maximum likelihood transition matrix fit on a subset of the data assigns probability zero to unobserved transitions, we observe cross validation log likelihoods of negative infinity in all but the most data-rich regimes with maximum likelihood methods. For simplicity, we use the Jeffreys prior on each row of the transition matrix – a

Dirichlet prior with $\alpha = \frac{1}{2}$ – which can be interpreted as a psuedocount of one half of an observed transition between each pair of states.

VI. RESULTS AND DISCUSSION

A. Müller Potential

We simulated two trajectories with 10^6 steps of Langevin dynamics on the Müller potential[]. Simulations were performed with a timestep $t = 0.1$, friction coefficient of $\gamma = 10^3$, and $kT = 15$. The first trajectory was clustered using the k-centers clustering algorithm, and state volumes were computed for the uniform emission model using 10^5 rounds of Monte Carlo rejection sampling with a padding distance of $R = 0.2$ around the state centers from the 500 state model to define the dataset's envelope. The second trajectory was used as a test set.

The volumes did not change drastically upon doing ten times more Monte Carlo rejection sampling, and the likelihood remained essentially constant, which indicates that the use of approximation volumes is sufficient to obtain robust estimates of the likelihood.

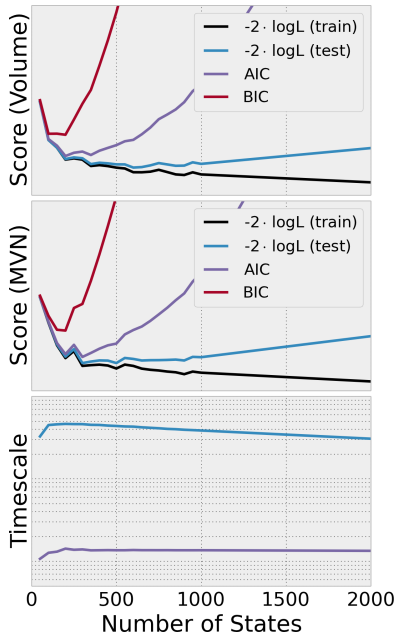

FIG. 3. In models built using between 50 and 1,000 states with the k-centers algorithm, the log likelihood function increased quickly and plateaued at approximately 300 states. There are many methods that can be used to test the transferability of an MSM. We've plotted the AIC, BIC, and cross-validation scores for the Müller potential simulations. These results indicate that, at least for this dataset, the BIC and AIC may penalize the number of states too heavily.

As shown in Fig. 3, models built with too few states achieve a drastically reduced likelihood, but above a threshold region the likelihood increases relatively slowly. Here, the AIC and BIC's penalty on the number of parameters, which scales with the square of the number of states, begins to dominate. The optimal models according to the three methods are between 200 and 1000 states for this system. This observation is consistent with the convergence of the relaxation timescales of the models.

Interestingly, the AIC and BIC penalize the larger state models much more than the cross-validation approach. This could be due to the simplicity of this system. Since the Müller potential is only two-dimensional, it is difficult to actually produce a drastically over-fit model. We suspect that the cross-validation score on a real protein system would be closer to the BIC and AIC. However, we note, that cross-validation requires estimating a model without some subset of the data, which is not feasible for large systems in the data-sparse regime.

We also tested the Gaussian likelihood on MSMs built using k-means clustering on the Müller potential. The likelihood (Fig. 4) peaks at a five-state model and decreases with the addition of further states. This maximum can be interpreted by inspecting the k-means cluster centers on the Müller potential. The five state model is the first model to have a cluster center located in the least stable well and as more states are added, the most stable well ends up being heavily split. We hypothesize that this over-splitting in the most stable minima gives rise to the decrease in MSM likelihood, and indicates a deficiency in clustering-based procedures which determine the state space without directly maximizing equations 2 or 3.

B. Fip35 WW Domain

To test the procedure on a larger protein system, we reanalyzed two ultra-long 100 μs molecular dynamics trajectories of the Fip35 WW domain[31], provided courtesy of D.E. Shaw Research [32]. Because the likelihood calculations involving the uniform distribution emission model have exponential complexity in the dimensionality of the state space we first preprocess the trajectories with time-structure based independent components analysis (tICA) [36, 37], extracting only the the four slowest uncorrelated linear combinations of residue-residue distances. Using the Gaussian emission model, we compare both models built in the 3 dimensional tICA subspace and those built using k-centers clustering with the cartesian root-mean squared deviation (RMSD) over the Fip35 WW domain's 252 non-redundant heavy atoms, computed using the quaternion characteristic polynomial method[38].

How many states are required in a Markov state model? The two 100 μs molecular dynamics trajectories of the Fip35 WW domain simulated by Shaw et. al.[32] have been used to parameterize multiple Markov state models. Lane *et al.* [39] proposed a 26,000 state MSM after clustering the data by RMSD, whereas Kellogg *et al.* [40] clustering on contact maps, present a 1000

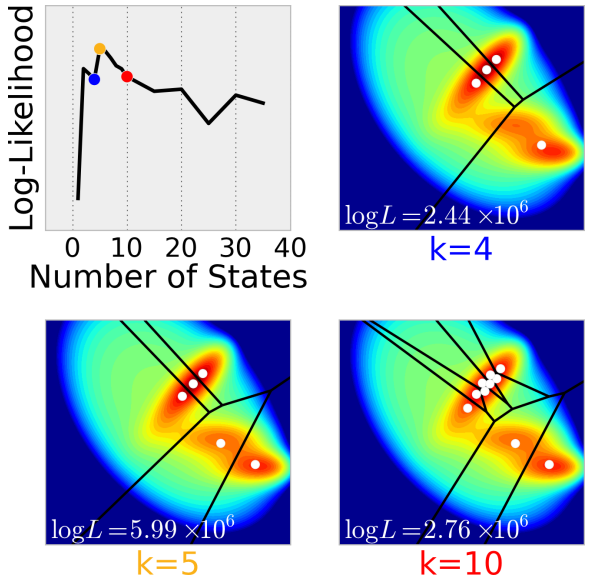

FIG. 4. MSMs were built using K-Means clustering on the Müller potential simulations. Since K-Means optimizes an objective function that is related to the emission probability, our likelihood picks only a few states as the most likely model. Interestingly, the likelihood decreases as you introduce additional states. This behavior is due to the state-to-state likelihood decreased rapidly when adding additional states, which may be a result of the large number of states in the most stable well.

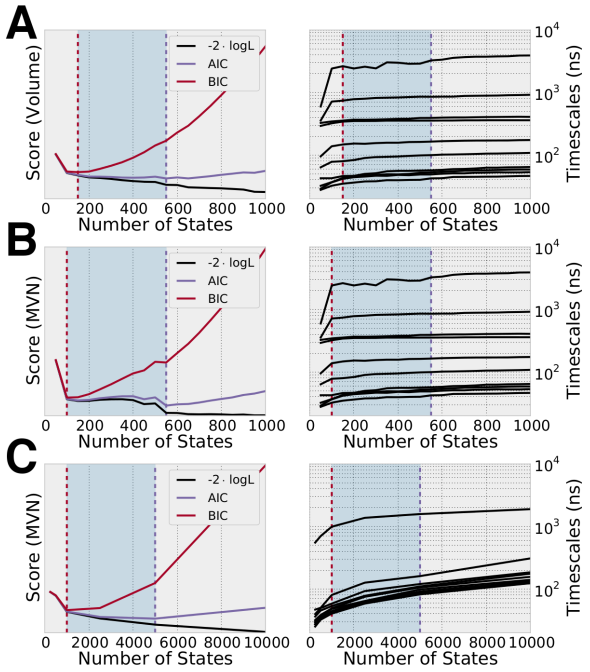

FIG. 5. For models built using tICA, we were able to compute volumes as the dimensionality was low (A). These likelihoods were qualitatively similar to the MVN likelihoods (B). Because the MVN likelihood is easy to compute, we can use it to calculate the likelihood of a model clustered using the RMSD distance metric (C). Our likelihood indicates we should construct a model with 5000 states in the RMSD metric, whereas previous studies used a model with 26,000 states [39].

state model. We find the optimal number of states to be highly dependent on space in which the data is clustered. With RMSD, the Gaussian likelihood indicates that Markov models built with on the order of 5,000 states optimally balance model quality and generalization error. On the other hand, the tICA method's identification of slow order parameters in the simulations has the effect of reducing the number of states necessary when constructing Markov models: here we find high-quality models can be constructed using only on the order of 500 states. Even with 200-600 states, Markov models constructed using tICA are able to accurately identify the simulations folding process on the 1-10 μs timescale, as well as the faster near-native dynamics identified in this dataset by McGibbon and Pande [23].

C. Limitations and Future Work

The likelihood functions described herein permits the comparison of Markov state models with varying number of states, however it requires the definition of a vector space over which the emission model is defined. As such, models built on different projections of an underlying space cannot be directly compared. Furthermore, this approach is not amenable to optimizing the Markov lag time, the discretization interval for the Markov jump process. Nonetheless, we anticipate that automatic selection of the number of states in a Markov state model will remove much of the guess work in MSM construction, as the selection of the lag time is more amenable to existing Chapman-Kolmogorov tests.

In the future, we plan to extend this work to the consideration of models without discrete states, where the requirement that states strictly partition phase space into a set of discrete indicator functions is relaxed.

VII. CONCLUSIONS

Markov State Models are powerful and popular tool for the analysis of molecular simulations, with a growing literature and two open source source software implementations[41, 42]. There are, however a number of steps in the construction process that require human intervention, which limits the use of MSMs to experts and introduces significant biases into the model building process. Additionally, the ability to automatically construct Markov state models on the fly while simulations are in progress, an important point for so-called adaptive sampling procedures[43], is hampered when manual model selection is required. In the finite-data regime, additional complexity in the Markov model, in the form of increasing number of states, does come at a statistical

cost. These results take a step toward fully automating the MSM construction process and controlling this statistical error by providing a quantitative method for by which the most suitable discretization procedure can be selected.

[1] C. M. Dobson, Nature **426**, 884 (2003).
[2] S. Kim, B. Born, M. Havenith, and M. Gruebele, Angew. Chem. Int. Ed. **47**, 6486 (2008).
[3] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, Biochemistry **14**, 5355 (1975).
[4] I. Bahar, C. Chennubhotla, and D. Tobi, Curr. Opin. Struct. Biol. **17**, 633 (2007).
[5] G. Cosa, Y. Zeng, H.-W. Liu, C. F. Landes, D. E. Makarov, K. Musier-Forsyth, and P. F. Barbara, J. Phys. Chem. B **110**, 2419 (2006).
[6] X. Zhang, V. Q. Lam, Y. Mou, T. Kimura, J. Chung, S. Chandrasekar, J. R. Winkler, S. L. Mayo, and S.-o. Shan, Proc. Natl. Acad. Sci. U.S.A. **108**, 6450 (2011).
[7] H. D. Mertens and D. I. Svergun, J. Struct. Biol. **172**, 128 (2010).
[8] S.-R. Tzeng and C. G. Kalodimos, Curr. Opin. Struct. Biol. **21**, 62 (2011).
[9] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, and V. S. Pande, J. Chem. Theory Comput. **9**, 461 (2013).
[10] M. Shirts and V. S. Pande, Science **290** (2000).
[11] D. Shaw, R. Dror, J. Salmon, J. Grossman, K. Mackenzie, J. Bank, C. Young, M. Deneroff, B. Batson, K. Bowers, E. Chow, M. Eastwood, D. Ierardi, J. Klepeis, J. Kuskin, R. Larson, K. Lindorff-Larsen, P. Maragakis, M. Moraes, S. Piana, Y. Shan, and B. Towles, in *High Performance Computing Networking, Storage and Analysis, Proceedings of the Conference on* (2009) pp. 1–11.
[12] B. Hess, J. Chem. Theory Comput. **4**, 116 (2008).
[13] I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson, and G. De Fabritiis, J. Chem. Inf. Model. **50**, 397 (2010), pMID: 20199097.
[14] P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, Nat. Phys. **6** (2010).
[15] T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, Curr. Opin. Struct. Biol. **23**, 58 (2013).
[16] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **107**, 1088 (2010).
[17] V. S. Pande, K. Beauchamp, and G. R. Bowman, Methods **52**, 99 (2010).
[18] K. A. Beauchamp, R. McGibbon, Y.-S. Lin, and V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **109**, 17807 (2012).
[19] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, J. Chem. Phys. **134**, 174105 (2011).
[20] G. R. Bowman, L. Meng, and X. Huang, J. Chem. Phys. **139**, 121905 (2013).
[21] E. Weinan and E. Vanden-Eijnden, J. Stat. Phys. **123**, 503 (2006).
[22] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, Linear Algebra and its Applications **315**, 39 (2000).
[23] R. T. McGibbon and V. S. Pande, J. Chem. Theory Comput. **9**, 2900 (2013).
[24] D. Pelleg, A. W. Moore, *et al.*, in *ICML* (2000) pp. 727–734.
[25] L. R. Rabiner, Proceedings of the IEEE **77**, 257 (1989).
[26] A. R. Liddle, Mon. Not. R. Astron. Soc. Lett. **377**, L74 (2007).
[27] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York Inc., New York, NY, USA, 2001).
[28] A. E. Gelfand and D. K. Dey, J. R. Statistic. Soc. B **56**, pp. 501 (1994).
[29] G. Schwarz, Ann. Stat. **6**, 461 (1978).
[30] H. Akaike, Automatic Control, IEEE Transactions on **19**, 716 (1974).
[31] F. Liu, D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, J. W. Kelly, and M. Gruebele, Proc. Natl. Acad. Sci. U.S.A. **105**, 2369 (2008).
[32] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, Science **330**, 341 (2010).
[33] R. Kannan, L. Lovász, and M. Simonovits, Random Structures & Algorithms **11**, 1 (1997).
[34] M. Simonovits, Math. Program. **97**, 337 (2003).
[35] L. Lovász and S. Vempala, J. Comput. System Sci. **72**, 392 (2006).
[36] C. R. Schwantes and V. S. Pande, J. Chem. Theory Comput. **9**, 2000 (2013).
[37] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, The Journal of Chemical Physics **139**, 015102 (2013).
[38] D. L. Theobald, Acta Crystallographica Section A **61**, 478 (2005).
[39] T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande, J. Am. Chem. Soc. **133**, 18413 (2011).
[40] E. H. Kellogg, O. F. Lange, and D. Baker, J. Phys. Chem. B **116**, 11405 (2012).
[41] K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, and V. S. Pande, Journal of chemical theory and computation **7**, 3412 (2011).
[42] M. Senne, B. Trendelkamp-Schroer, A. S. Mey, C. Schütte, and F. Noé, Journal of Chemical Theory and Computation **8**, 2223 (2012).
[43] G. R. Bowman, D. L. Ensign, and V. S. Pande, Journal of chemical theory and computation **6**, 787 (2010).