Dynamic Graphical Models of Molecular Kinetics

Simon Olsson* and Frank Noé*

Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany

E-mail: simon.olsson@fu-berlin.de; frank.noe@fu-berlin.de

Abstract

Detailed atomistic descriptions of molecular dynamics make use of global state variables – that is, a molecular configuration is represented by a single global variable, $\mathbf{x}$. Probabilistic or rate models of molecular kinetcs, such as Markov state models, similarly segment $\mathbf{x}$ into discrete *Markov states*, to facilitate subsequent estimation using time-resolved data. This approach generally works well when states meta-stable at a given time-resolution, τ, are well separated in the discretization and sufficient data is available. Nevertheless, the number of global meta-stable states may grow exponentially with the size of the molecular system, making discretization and estimation difficult. In this work we describe an alternative approach which leverages a molecular representation based upon several local sub-systems. The time-evolution of the subsystems is governed by their mutual dependence and their local tendencies to adopt certain configurations both encoded by the connectivity graph of the model; a *dynamic graphical model*. Such a localized model has a compact parametric form and statistical robustness even with sparse data. Furthermore, dynamic graphical models sidestep the explicit representation of each global state configuration, and thus enable us predict molecular configurations and their dynamics from beyond the observed domain. If each sub-system only depends on a small number of other sub-systems — if the connectivity graph is sparse – we can decompose the graph into dynamically independent clusters of sub-systems. We provide a practical implementation of such a model using a dynamic variant of the Ising model, the dynamic Markov random field (dMRF) and present maximum likelihood inference schemes. With this approach, we find that it is possible to accurately predict realistic global molecular configurations which have not been observed during the model estimation. While the general implementation of dynamic graphical models is ambiguous the one shown here already illustrates the main powers of the approach. We anticipate that the deployment of learning structures with more expressive power can further improve on these results.

Introduction

The function of biological macromolecules including proteins is governed by transitions between meta-stable conformational states. An important goal in molecular biophysics is to obtain accurate and highly resolved models of the kinetics of these conformational transitions. Establishing such models – for example Markov state models (MSM) – for smaller proteins, is becoming increasingly feasible due to improvements in simulation strategies, computer hardware as well as improved statistical estimation methods[1–3]. However, for molecular simulations to have a broader impact within molecular biology, and in particular structural biology, scaling to larger molecular systems is critical.

MSMs and related Master equation-based models describe a Markov jump-process between mutually exclusive configurational states[4,5]. Given all relevant configurational transitions have been sampled, these models can be estimated and thereby provide optimal discrete approximations of the transfer operator, $\mathcal{T}$, governing a systems molecular dynamics[6,7]. However, given a fixed finite time-resolution, τ, the number of molecular configurations which are meta-stable grows exponentially with the size of the molecular system. Generating simulation data which sample all possible configurational transitions thus rapidly becomes in-tractable. The challenge of larger proteins is further exercerbated by the increased over-head due to the large number of particles to be repesented in the simulation box. Consequently, these methods will not scale to large molecular systems.

In this paper, we propose an alternative approach to obtaining statistical models of molecular kinetics. The strategy distinguishes itself compared to the methods discussed above by the way the molecular configurations are represented. We depart from global and mutually exclusive configurational states, instead we consider a configurational state as a concatenation of several locally defined sub-systems and their configurations. With this representation we construct a probabilistic model which gives us a distribution over all the sub-system configurations at a time, given their configuration at previous times – dynamic graphical models (DGM).

Our findings suggests that using localized representations provides an interesting and potentially powerful new way to model molecular kinetics. The localized representation caters to a compact parametric form which can be robustly estimated using relatively little simulation data, yet readily predict beyond the training domain, and can thereby speed up the discovery of meta-stable states. However, ambiguity in the sub-system encoding and similarly in decoding predictions from a sub-system encoded configuration tos a molecular structure remains challenging when deploying dMRFs. Nevertheless, we envision that this work will stimulate further research in to the use of alternative, more statistically efficient ways to encode molecular representation when modeling molecular kinetics.

Dynamic graphical models

Current methods used to simulate and analyse molecular dynamics do not allow statistically sufficient simulations of large molecules as the number of metastable states grow too rapidly with molecule size. The origin of this poor scaling is due to the encoding of molecular configurations using a single state variable, S, which means that the number of configurations which are meta-stable on a time-scale of τ necessarily puts a lower bound on the number of parameters necessary to encode the molecular kinetics.

Here, we propose a new strategy which uses several localized features (*sub-systems*, σ_i) to represent molecular configurations. A global molecular configuration is therefore only indirectly encoded as a concatenation of states of the sub-systems, $S = \{\sigma_i\}_i^N$. The departure from a global state variable used in conventional methodology, necessitates a novel model structure which encodes the time-evolution of the molecular sub-systems. In this work we propose the dynamic graphical models (DGM). DGMs are graphical models which encode the interaction between molecular sub-systems and their time-evolution.

Markov random fields (MRF) is a class of statistical models which allows us to model the interactions between multiple discrete statistical variables[8]. A well-known example of a MRF is the Ising model from statistical physics: a model of N sub-systems on a lattice, each of which adopt one of two configurations, 1 or -1[9]. Each sub-system interacts directly only with its nearest neighbors, and possibly an external field. Although all interactions in the Ising model are local, complicated non-local phenomena arise. Indeed, the Ising model and other closely related models have seen a wide array of applications across the sciences. While these models are powerful, they are notoriously difficult to estimate given a set of empirical observations, without resorting to a number of approximations[10,11]. Nevertheless, even if we succeed in estimating a MRF using any of these methods, the model in itself does not contain a description of the dynamics, limiting us to thermodynamic predictions.

We here introduce the dynamic MRF (dMRF) as an implementation of DGMs. The dMRF provides a first approximation of a model which uses multiple localized configurations to describe the dynamics of a molecular system. We present statistical estimators of dMRFs and show how dMRFs can recapitulate macroscopic quantities of Markov state models on well sampled data-sets. In cases where less data is available, we find that the dMRFs allow us to predict system configurations which have not not been explicitly observed during the estimation of the model, yet are consistent with other simulation data held aside for validation. Our results demonstrate that it is possible to learn system-specific rules about how different configurational sub-systems interplay.

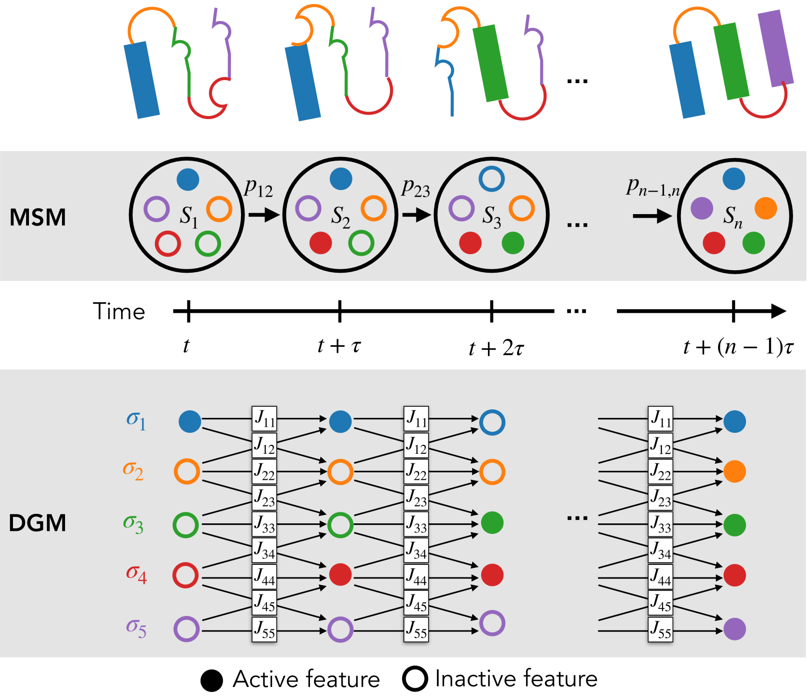

Figure 1: Illustrative comparison of molecular representations in Markov state models (MSM) and dynamic graphical models (DGM). Upper panel shows cartoon representation of different protein conformational states. MIddle panel show MSM where the current Markov state is indicated by, S_i. The lower panel show DGMs where the current state of the system is represented by a concatenation several sub-systems, σ_i. The MSM transition probabilities of in between Markov states S_i and S_j are shown as p_{ij}. The DGM sub-system coupling between σ_i and σ_j is shown as J_{ij}.

Results

Dynamic Markov random fields as an implementation of dynamic graphical models

The Ising model is a simple Hamiltonian model expressing the stationary probability of the configuration of a set of N binary $\{-1, 1\}$-sub-systems $\mathbf{s} = \{\sigma_i\}_{i=1}^{N}$ as

$$p(\mathbf{s}) = \mathcal{Z}^{-1} \exp(\sum_{i=1}^{N} \sum_{j \neq i} J_{ij}\sigma_i\sigma_j + h_i\sigma_i) \tag{1}$$

where the (unit-less) model parameters $J_{ij} = J_{ji}$ and h_i describe the coupling strength between sub-systems i and j and the local field of spin i, respectively, and $\mathcal{Z}$ is the partition function. While simulating eq. 1 can be performed easily given $\{J_{ij}\}$ and $\{h_i\}$ using one of several sampling schemes[12,13], the inverse problem i.e. estimating these parameters from equilibrium samples is not straight forward. This is a result of the complex dependency of the partition function on the model parameters. In particular, maximum likelihood inference relies on a sampling-based Boltzmann learning scheme which is only tractable/accurate in very limiting situations: small N and large numbers of independently and identically drawn samples. Similar restrictions apply to the more general MRFs which allow each sub-system to adopt multiple configurations. Nevertheless, several approximate methods with attractive asymptotic properties have been proposed[10,14].

In the context of molecular simulations and kinetics our data does not consitute independently and identically drawn samples from a configurational equilibrium distribution, rather we have a time-series of a molecular system. In other words, we have samples from a conditional distribution $p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau})$ where $\mathbf{s}_t$ is a representation of the molecular system at time t, comprised of N sub-systems $\sigma_{t,i}$. This conditional probability distribution encodes the dynamics of our molecular system.

There is a tight connection between the ergodic dynamics of a system and its (non)equilibrium steady-state distribution. Therefore, if we are able to successfully model the conditional distribution, $p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau})$, we can make predictions about the steady-state properties: at equilibrium, $p(\mathbf{s})$. Here, we model the conditional transition probability $p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau})$ using the simplest dynamic Markov random field (dMRF), a "dynamic Ising model". The vector $\mathbf{s}_t$ encodes the configuration of N (binary: $\{-1, 1\}$) subsystems at time t. We have

$$
\begin{aligned}
p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau}) &= \mathcal{Z}^{-1} \exp \left\{ \sum_{i=1}^{N} \sigma_{t,i} (\sum_{j=1}^{N} J_{ij} \sigma_{t-\tau,j} + h_i) \right\} \\
&= \mathcal{Z}^{-1} \exp \left\{ \mathbf{s}_t^{\intercal} \mathbf{J} \mathbf{s}_{t-\tau} + \mathbf{s}_t^{\intercal} \mathbf{h} \right\} \qquad (2)
\end{aligned}
$$

where $\mathbf{J} \in \mathbb{R}^{N \times N}$ is a matrix of coupling parameters and $\mathbf{h}$ is a vector of local fields and $\mathcal{Z}$ is the partition function. Unlike in eq. 1, the couplings J_{ij} are not necessarily equal to J_{ji} and the 'self-couplings' (J_{ii}) take on finite values. In the case where $\mathbf{J}$ is symmetric ($J_{ij} = J_{ji}$) we can choose J_{ii} such that sampling according to eq. 2 converges to an equilibrium distribution which is exactly eq. 1. In the Supporting information we discuss and analyse the dMRF model parameters further, and show their dependence on intrinsic system properties (Fig. S1).

The dynamics exhibited by dMRFs is closely related to the Glauber dynamics[12], commonly employed to simulate the Ising model. Glauber dynamics is time-continuous and involves at most one sub-system changing its state in δt and all the sub-systems have a fixed “spin-flip” rate of k. In contrast the dMRF dynamics is time-discrete and allows for several sub-systems to change their configuration in a time τ, and the “spin-flip” rates are not necessarily uniform and encoded into the 'self-couplings', J_{ii} (Supporting Information).

If we consider the conditional probability of a single sub-system $\sigma_{t,i}$ taking the state 1, given $\mathbf{s}_{t-\tau}$ we see that it is sigmoid in $\theta_{t-\tau,i} = \sum_j J_{ij} \sigma_{t-\tau,j} + h_i$,

$$
\begin{aligned}
p(s_{t,i} = 1 \mid \mathbf{s}_{t-\tau}) &= \frac{\exp[\theta_{t-\tau,i}]}{\exp[\theta_{t-\tau,i}] + \exp[-\theta_{t-\tau,i}]} \\
&= \frac{1}{1 + \exp[-2\theta_{t-\tau,i}]}.
\end{aligned}
$$

Note that the partition function is analytically tractable!

Since the marginal probability distributions of the sub-systems of $\mathbf{s}_t$ are conditionally independent given $\mathbf{s}_{t-\tau}$, the joint probability density is simply their product,

$$p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau}) = \prod_i^N p(\sigma_{t,i} \mid \mathbf{s}_{t-\tau}) = \prod_i^N \frac{1}{1 + \exp[-\sigma_{t,i} 2\theta_{t-\tau,i}]}. \tag{3}$$

Our statistical model of the global time-evolution of $\mathbf{s}_t$ is in other words composed of N independent models each of which encode the configuration probabilities of a sub-system $\sigma_{i,t}$ given the previous configuration of all sub-systems, $\mathbf{s}_{t-\tau}$. These properties makes estimation of $\mathbf{J}$ and $\mathbf{h}$ in eq. 3 by maximum likelihood straight forward: we solve N independent logistic regression problems each of which model the outcome of a sub-system $\sigma_{i,t}$ given the regressors $\mathbf{s}_{t-\tau}$[15]. Throughout this paper, we use a LASSO/L_1 regularized estimator, which corresponds to performing maximum *a posteriori* estimation with a Laplacian prior on J_{ij} centered at 0 (Supporting information).

This dynamic Ising model can readily be generalized to cases where each sub-system can adopt more than two states, σ_i. In the supporting information we discuss this generalization and show how estimation of such a model corresponds to performing N independent softmax regression problems on a simulation trajectory against itself with a time-lag of τ (Supporting information).

Inference of unbiased couplings from biased data

So far we have described a new dynamical model of which allows us to model a time-discrete evolution of a molecular system using a localized molecular representation. A model of this kind allows us to estimate unbiased coupling between sub-systems J_{ij} given a biased data-set, if the sub-system couplings are not a function of the global configuration of the sub-systems. To test this assertion we first consider the one dimensional Ising model. The Ising model has two meta-stable global configurations when simulated at sub-critical temperatures using Glauber dynamics. The two meta-stable configurations correspond to all sub-systems adopting the same configuration.

We generate non-equilibrium (NED) and equilibrium (ED) data sets of the Ising model, by performing several realizations of the Glauber dynamics with all sub-systems initialized in state -1. In the NED, we terminate the simulation if the net magnetization becomes larger than 0 – i.e. when more than half of all the sub-systems are in the 1 configuration. Consequently, several global configurational states have not been observed in the data set, among these one of the meta-stable states. In the ED data simulations are run with unbiased Glauber dynamics with a length as to match the sampling statistics of the NED data (Fig 2A). Using a regularized maximum likelihood scheme we estimate two models: NED dMRF and ED MRF respectively, along with a MSM of the NED. We find that dMRFs estimated using the NED and ED data-sets converge to the same sub-system couplings (Fig 2B) regardless of whether we choose to estimate the external field parameters, h_i, or not (Fig S2). This means that the unbiased sub-system couplings can be recovered from a biased data-set, in turn suggesting that global state distributions can be estimated from biased data.

With the estimated models we reconstruct global MSMs by enumerating all possible global transitions $i \to j$ of the system and compute their transition probabilities, T_{ij}, using eq. 3. We use these MSMs to directly compare the stationary (thermodynamic) and dynamic properties of our estimated model against a 'true' analytical MSM reference used to generate the data. Remarkably, we find the distribution of the net system magnetizations, $\langle M \rangle$, of the estimated models closely resemble the reference distribution (Fig 2C). Conversely, a MSM estimated using the NED, fails to capture global configurations which have not been directly observed in the training data (Fig 2C). Similarly, direct comparison of global transition probabilities, T_{ij}, predicted from the dMRF correlate well with their true reference values, albeit with some biases (Fig 2D). Finally, the true global relaxation time-scales are recapitulated well by the dMRF models (Fig 2E) – however, for the NED dMRF the time-scales are systematically slowed, as the self-couplings are over-estimated (Fig 2B). Nevertheless, the NED MSM completely fails to capture the slowest process as it has not been observed in the training data.

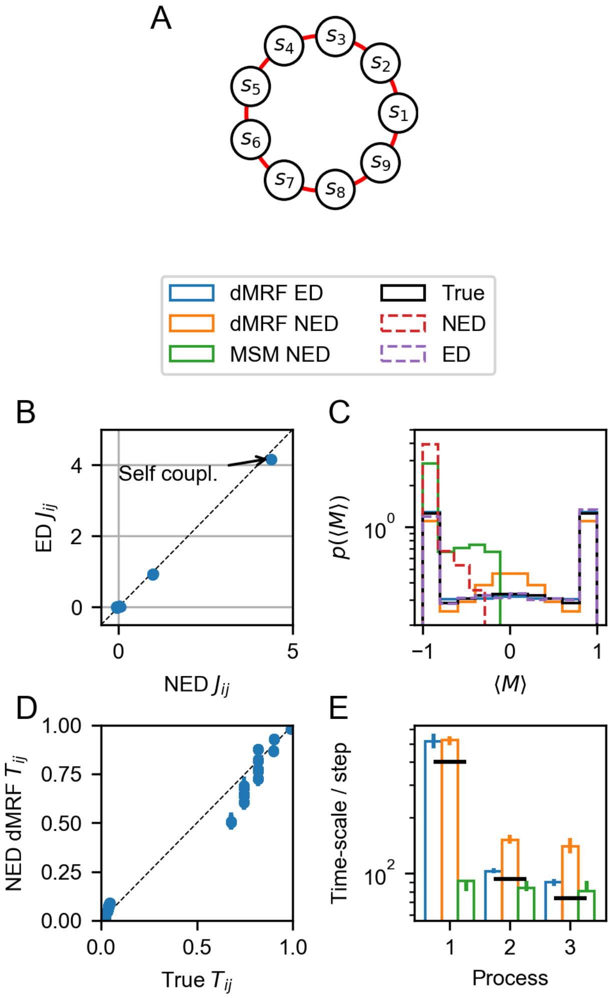

Figure 2: Inference of model parameters for a 1-dimensional 9 spin Ising model. A) Illustration of the 9-spin 1D Ising model with periodic boundary conditions. B) Comparison of coupling parameters estimated using equilibrium and non-equilibrium data sets. C) Analytic stationary distributions of $\langle M \rangle$ and empirical histograms of equilibrium and non-equilibrium data sets. Shown dMRF predictions are for models where the local fields $\{h_i\}$ are not estimated. D) Comparison of global configurational state transition probabilities predicted in dMRF trained on non-equilibrium data set against the true reference values. E) Implied time-scales of estimated models and the true reference.

dMRFs as models of molecular dynamics: comparison to MSMs

Above we show that estimated dMRFs can predict important meta-stable configurations which have not been observed during training. However, for the example above, we have an optimal encoding of the sub-systems (the spins) with truly Markovian dynamics and we know that the sub-system couplings are independent on the global state – for molecular systems this is generally not the case. We used molecular dynamics simulation data of a small penta-peptide system (WLALL) to test whether dMRFs could be a viable scheme for modelling molecular kinetics. This system is sufficently small to enable detailed microscopic analysis with MSMs while exerting meta-stability.

We encode the back-bone torsions (ϕ,ψ) of WLALL into 8 sub-systems, each of which has 2 states, corresponding to different rotameric basins in the Ramachandran plane (Methods). We estimate MSMs and dMRFs using this representation, however, for the MSMs we enumerate all the possible (2^8) global configurations of the sub-systems as independent Markov states. By enumerating all possible sub-system transitions the dMRFs can be used to reconstruct a MSM whose properties can be compared directly to the directly estimated MSM. We find that the implied time-scales computed from dMRFs match those of the MSMs well at multiple lag-times (Fig. 3A). Similarly, the stationary probabilities of the Markov states match closely, in particular for high probability states (Fig. 3B).

As opposed to MSMs, dMRFs yield a transition probability for all possible sub-system configurations, whereas MSMs only provide information about directly observed transitions – similarly, dMRFs (here, $N_{\mathrm{dMRF}} = 8^2 + 8$) are parametrically more compact compared to MSMs (here, $N_{\mathrm{MSM}} = 2^{16}$). Intuitively, we would anticipate the smaller number of parameters enable more precise estimation of parameters with less data, effectively meaning that dMRFs are more data-efficient that MSMs. The smaller error-bars on the dMRF implied time-scales suggest that this may be the case (Fig. 3B). However, to more quantitatively compare the data-efficiency of MSMs and dMRFs we estimated multiple dMRFs and MSMs with increasing volumes of molecular dynamics data. Using the estimated models we compute the stationary probabilities of Markov states visited in all trajectories (Fig. 3C). We find the predicted stationary probabilities converge at lower data-volumes for dMRFs compared to MSMs in particular for low-probability states. However, since there is a discrepancy between the predicted stationary probabilities for the low probability states and we do not have an exhaustive data-set at hand, we cannot gauge which of the predictions are more accurate.

Prediction of unobserved meta-stable molecular configurations

Above we saw how the dMRF framework can recapitulate the dynamics of MSMs, and provide predictions of stationary probabilities of unobserved global configurations. To more systematically test the predictive power of dMRFs in predicting meta-stable configurations not observed during simulation we estimated several dMRFs designed to be selectively blind towards a particular meta-stable state. We specifically data of two fast-folding proteins previously published, villin and BBA[16], an all α and an α/β folds respectively.

For villin and BBA we respectively built a five and a four state hidden Markov model (HMM)[17]. Using the HMM we assign the MD trajectories to meta-stable states. We use the meta-stable assignments to sub-sample the original data to generate data-sets which are specifically blind towards one of the 5 or 4 meta-stable states. These sub-sampled data-sets are used for estimating dMRFs each of which in turn are 'blind' to one meta-stable configuration (see SI for details). From the estimated dMRFs we simulate synthetic trajectories of the time-evolution of the sub-systems. We find that the dMRFs generally sample the same configurational space as the full MD simulations, yet with skewed populations (Fig. S3). This suggests that although the predicted thermodynamics in the blinded dMRFs is not quantitatively accurate, the sampled space is realistic. This finding is mirrored by the fact that statistical properties of the configurations sampled by the dMRF closely agree with their MD counterparts (Figs. S4 and S5)

Next we quantitatively compare the macroscopic stationary and dynamics properties from these synthetic simulations to HMMs estimated the full MD data (Fig. 4). Remarkably, we

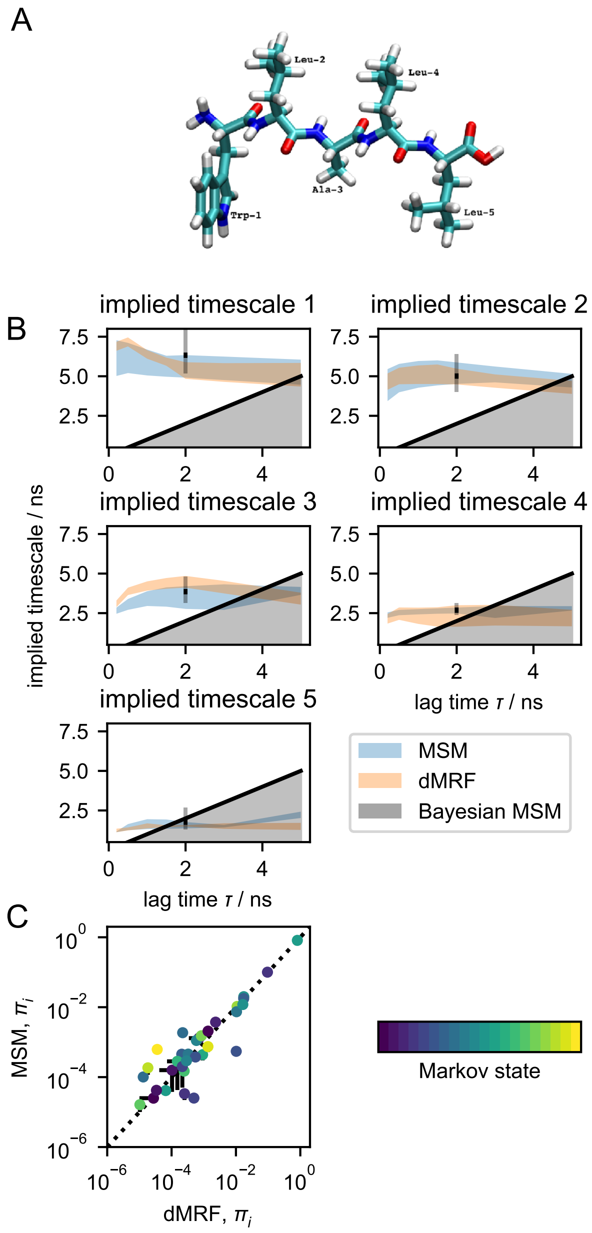

Figure 3: Comparison of dMRF and MSM describing back-bone torsion dynamics of the pentapeptide Trp-Leu-Ala-Leu-Leu (WLALL). A) Molecular structure render of the WLALL peptide in stick representation. B) 95% confidence intervals of MSM (orange) and dMRF (blue) implied time-scales. C) Global state distribution computed according to dMRF and MSMs. Confidence intervals in A and B are computed using bootstrap with replacement. Markov states shown in C are those observed in all 24 trajectories; dMRF state probabilities are renormalized to this set.

find that the synthetic trajectories sample configurations which have not been observed during training, and in the majority of cases with a population comparable to the MD simulation (Fig. 4A/D). Reconstructing molecular configurations from the sub-system represention in which the synthetic trajectories are encoded generally has a small error (Fig. 4B/E), suggesting that the microscopic configurations sampled by dMRFs indeed are realistic. Inspecting the reconstructed molecular trajectories we see that the meta-stable configurations generally are meta-stable, that is, they remain in the same meta-stable state for several frames (Fig. S6).

Finally, we inspect the time-correlation function (TCF) of the rotameric state of a backbone torsion angle in the reference simulations of villin and BBA and compare them to corresponding TCFs from the dMRF trajectories (Fig. 4C/F). We find these generally agree well, although the dMRFs tend to over-estimate the relaxation rates slightly. This is particularly clear for the TCFs calculated from dMRFs blinded to the folded configurations (state red/4 in both cases).

Discussion and Conclusions

MSMs rely on a global discretization of the molecules' state space. In other words, the state of the molecule is described by a single state variable which evolves as a Markov jump process in time[2]. To obtain an accurate model of the timescales larger than τ, the number of states that are metastable at this timescale defines the minimum number of states, and thus also parameters, that must be used in an MSM[4,18]. Thus, being able to parametrize an MSM from a realistic amount of simulation data requires that the number of metastable states is not too large.

Indeed, MSMs were particularly successful for protein systems that are either small or sufficiently co-operative to possess only few metastable states[19–22]. This approach is not scalable to large molecular systems. Multi-domain or multi-protein systems likely possess

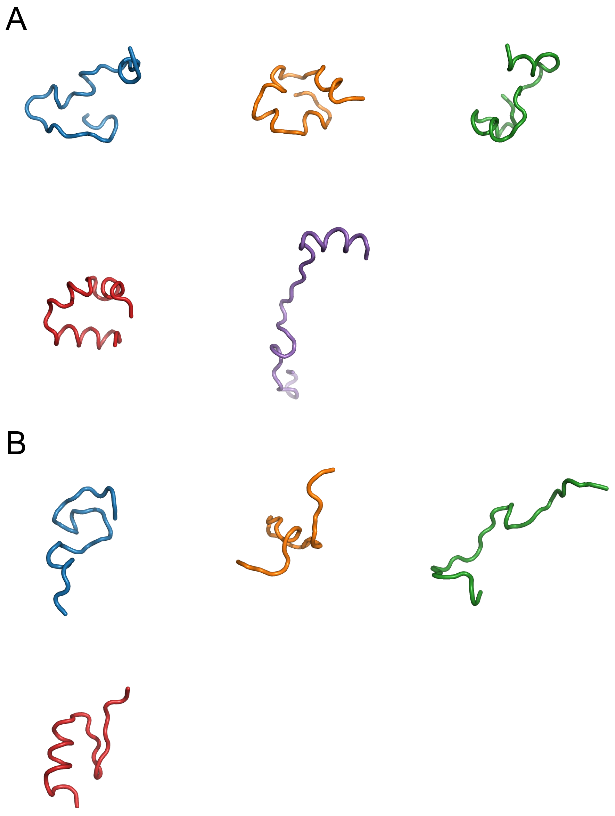

Figure 4: Molecular renders of meta-stable configurations identified by HMMs of villin (A) and BBA (B), using a ribbon representation. Color/meta-stable state relationship: Blue/1, Orange/2, Green/3, Red/4 Purple/5.

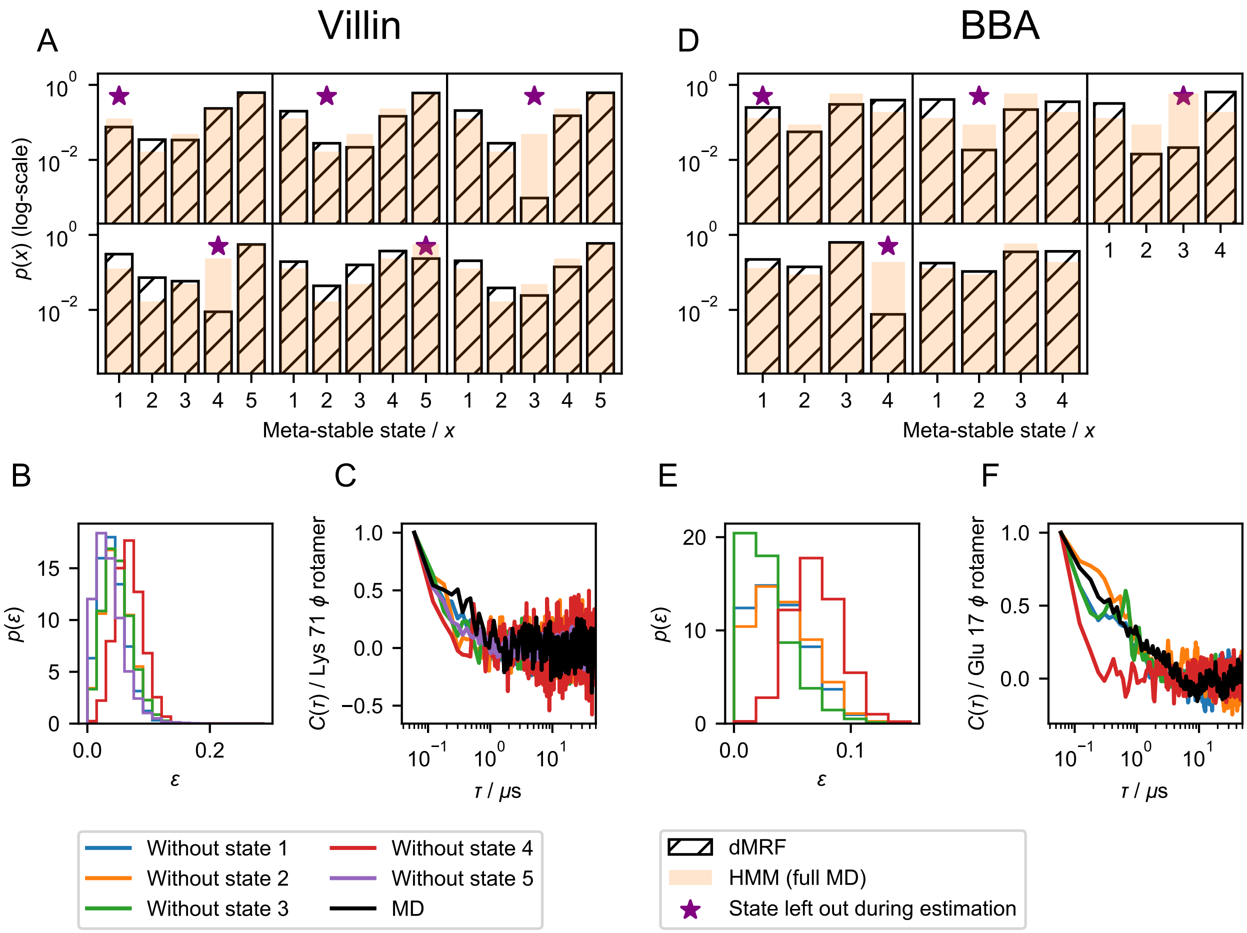

Figure 5: Prediction of macroscopic stationary and dynamic properties of fast folding proteins Villin and BBA with dMRFs. A/D) Meta-stable state distributions sampled by dMRFs estimated leaving data from one of four meta-stable state out during estimation and using the full data set (hatched). Reference distribution estimated using a hidden Markov model (HMM) estimated with the full MD data-set (orange). B/E) Histograms of reconstruction errors of atomistic models from sub-system encoded trajectories sampled by dMRFs (Hamming distance). C) Normalized auto-correlation function of the rotameric state of the ϕ torsion of Lys 71 in Villin as sampled by the dMRF models and in the simulation data. F) Normalized auto-correlation function of the rotameric state of the ϕ torsion of Glu 17 in BBA as sampled by the dMRF models and in the simulation data.

at least partially uncoupled subunits with internal metastable states[23]. A molecular system characterized by N independent subunits with M metastable states each will have an exponential number, M^N, of metastable states.

In this paper, we introduce the concept of dynamic graphical models (DGM) and their use to model molecular kinetics. The general approach encodes molecular conformations into multiple sub-systems; global configurational states are therefore encoded indirectly as a concatenation of several local variables. As DGMs only model the pair-wise interactions between sub-systems an exponential number of global configurations can be represented with a quadratic number of parameters. We illustrate dynamic graphical models using a generalization of the dynamic Ising model — the dynamic Markov random fields (dMRFs).

In general, we find the dMRFs, in spite of their compact parametric form can approximate well the dynamic behavior of several molecular systems. When gauged against MSMs, dMRFs predict similar characteristic time-scales and stationary distributions. Beyond the capabilities of MSMs, dMRF also allow for the prediction of transitions to and from unobserved molecular conformations and their stationary probabilities. Specifically, we found that dMRFs can predict the predict the prescence of meta-stable configurations which have not been observed during model estimation, although their absolute probabilities are not always accurately captured.

With the change in representation necessary to use dMRF come new technical challenges and limitations. First, identifying important sub-systems and how these should be encoded is not trivial. Specfically, we currently do not have a governing principle to gauge different sub-system encodings against each other *a priori*. Similarly, going from an encoded representation back to a molecular configuration (decoding) is generally a difficult problem. Second, although efficient regularized maximum likelihood estimators are readily available, optimally chosing regularization factors remains heuristic. Third, the compact parametric form of the dMRFs comes with limited expressive power. It is specifically assumed that the sub-system couplings are *not* a function of the global configuration – if this is violated we cannot expect to obtain quantitatively predictive models using dMRFs. Finally, unlike for MSMs[3] the integration of experimental data into the estimation of dMRFs is currently not possible. Similarly prediction of experimental observables will, unlike for MSMs[24,25], involve sampling dMRFs which could be both time-consuming and difficult, in particular when no sub-system decoding is possible.

Machine learning, and in particular deep learning, has seen a surge in interest in the sciences the past few years, including in the context of molecular kinetics[26,27]. A hallmark of deep learning is its ability to learn complicated non-linear functions which fulfill a pre-specified set of criteria given sufficient available data[28]. The identification of sub-systems and their discretization is highly non-linear and akin to the featurization, projection and discretization process whose success critically determines the success of Markov state modelling . The VAMPnets approach recently illustrated how the featurization, projection and discretization pipeline of MSM building could be replaced by a deep neutral network[26]. We envision that a related approach can be taken in the context of identifying sub-systems in molecular systems for dMRFs and thereby possibly side-step this currently cumbersome and manual process.

Gerber and Horenko recently described a linear probabilistic model which aims to predict the time-evolution of several binary random variables.[29] Their method is motivated as a Granger causality model, i.e. the current rotameric configurations causally encodes a probability density of rotameric states at a future time. Their model can be seen as a special case of the DGMs which is implemented by a constrained linear regression problem with a time-lag.

We anticipate the use of dMRFs or related methods can be broadly applied in biophysics. The compact parametric form allows us to model molecular systems which are significantly larger than what would be tractable using for example MSMs. Furthermore, a sparse subsystem coupling matrix $\mathbf{J}$ could enable a spatial decomposition or large molecular systems into conditionally independent fragments, more amenable for simulation. The ability of dMRFs to predict unobserved global configuration could futher be used to in an adaptive simulation setting, where previously unobserved states are reconstructed and used to seed new molecular simulations.

A library for estimation and analysis of dMRFs along with example notebooks to reproduce results from select figures of this manuscript is available on as part of the graphtime library: http://www.github.com/markovmodel/graphtime/

Synopsis Discovery of meta-stable configurational states of large biomolecules is an important problem in biophysics. We describe a procedure to speed up this discovery process by learning how local structural elements interact and how these interaction influence the temporal behavior of the global configurational states.

Supporting Information Available

A detailed report of the practical estimation of models presented in the work including hyper-parameter choices can be found in the supporting information. Further, a theoretical generalization of dMRF to cases where sub-system may adopt more than two discrete states along with their constrained and unconstrained maximum likelihood estimators is available. Finally auxillary results including figures and movies illustrating properties of the estimated models may are shown in the supporting material. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Acknowledgement

We thank Fabian Paul, Hao Wu, Christoph Wehmeyer, Illia Horenko, Moritz Hoffmann and Tim Hempel for stimulating discussions. We thank Nuria Plattner for providing simulation data for the penta-peptide system and DE Shaw Research for providing the simulation data for the villin headpiece and BBA. We gratefully acknowledge funding from the Alexander von

Humboldt foundation (Postdoctoral fellowship to SO) and the European Research Council (ERC Consolidator Grant ScaleCell to FN).

References

(1) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *J. Chem. Phys.* **2009**, *131*, 124101.

(2) Bowman, G. R., Pande, V. S., Noé, F., Eds. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation.*; Advances in Experimental Medicine and Biology; Springer Heidelberg, 2014; Vol. 797.

(3) Olsson, S.; Wu, H.; Paul, F.; Clementi, C.; Noé, F. *Proceedings of the National Academy of Sciences* **2017**, *114*, 8265–8270.

(4) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.

(5) Husic, B. E.; Pande, V. S. *Journal of the American Chemical Society* **2018**, *140*, 2386–2396.

(6) Noé, F.; Nüske, F. *Multiscale Modeling & Simulation* **2013**, *11*, 635–655.

(7) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. *Journal of Chemical Theory and Computation* **2014**, *10*, 1739–1752.

(8) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer Science Business Media, 2006.

(9) Lenz, W. *Phys. Z.* **1920**, *21*, 613–615.

(10) Ravikumar, P.; Wainwright, M. J.; Lafferty, J. D. *Ann. Statist.* **2010**, *38*, 1287–1319.

(11) Parise, S.; Welling, M. In *Advances in Neural Information Processing Systems 19*; Schölkopf, B., Platt, J., Hoffman., T., Eds.; 2006.

(12) Glauber, R. J. *Journal of Mathematical Physics* **1963**, *4*, 294–307.

(13) Kawasaki, K. *Physical Review* **1966**, *145*, 224–230.

(14) Roudi, Y.; Tyrcha, J.; Hertz, J. *Physical Review E* **2009**, *79*.

(15) Roudi, Y.; Hertz, J. *Physical Review Letters* **2011**, *106*.

(16) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517–520.

(17) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. *J. Chem. Phys.* **2013**, *139*, 184114.

(18) Sarich, M.; Noé, F.; Schütte, C. *SIAM Multiscale Model. Simul.* **2010**, *8*, 1154–1177.

(19) Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011–19016.

(20) Plattner, N.; Doerr, S.; Fabritiis, G. D.; Noé, F. *Nature Chemistry* **2017**, *9*, 1005–1011.

(21) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. *Nature Chemistry* **2013**, *6*, 15–21.

(22) Bowman, G. R.; Geissler, P. L. *Proceedings of the National Academy of Sciences* **2012**, *109*, 11681–11686.

(23) Faelber, K.; Posor, Y.; Gao, S.; Held, M.; Roske, Y.; Schulze, D.; Haucke, V.; Noé, F.; Daumke, O. *Nature* **2011**, *477*, 556–560.

(24) Noe, F.; Doose, S.; Daidone, I.; Lollmann, M.; Sauer, M.; Chodera, J. D.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 4822–4827.

(25) Olsson, S.; Noé, F. *J. Am. Chem. Soc.* **2017**, *139*, 200–210.

(26) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. *Nature Communications* **2018**, *9*.

(27) Wu, H.; Mardt, A.; Pasquali, L.; Noe, F. Deep Generative Markov State Models. arXiv:1805.07601, 2018.

(28) Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press, 2016; `http://www.deeplearningbook.org`.

(29) Gerber, S.; Horenko, I. *Proceedings of the National Academy of Sciences* **2014**, *111*, 14651–14656.