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

Markov state models and related Master equation models are powerful tools to study the kinetics of conformational change in molecules. These models segment the conformational space into 'global' configurational states whose mutual exchange estimated using time-resolved data. This approach generally works well when the number of segments is larger than the number of meta-stable states at a given time-resolution, τ, and sufficient data is available. We here describe an alternative approach which leverages a molecular representation based upon several local discretizations. The time-evolution of the local representations is governed by their mutual dependence and their local tendencies to adopt certain configurations both encoded by the model. The advantage of such a localized model is two-fold: firstly, it constitutes a more compact parametric representation of the dynamics and secondly, it enables prediction of transitions to and from molecular configurations not observed in the data used to parameterize the model. We provide an implementation of such a model using a dynamic variant of a Markov random field (MRF) which is a generalization of the Ising model from statistical physics. We present maximum likelihood inference schemes and show through the application to molecular and model systems how the approach enables the discovery of microscopic system configurations not observed in the data used to estimate the model.

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 the transition probabilities between mutually exclusive segments of configuration space[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 intractable. 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.

In this paper, we propose an alternative approach to obtaining statistical models of molecular kinetics. The strategy presented here distinguishes itself compared to the methods discussed above by the way the molecular configurations are represented[8]. We depart from global and mutually exclusive configurational states, instead we consider a configurational state as a concatenation of several locally defined sub-system configurations.

A sub-system can, for instance, be a torsion angle with several rotameric configurations or two residues being in contact or not. 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. A simple example of such a model is a set of independent Markovian models which describe localized conformational changes: rotamer jumps or contact formation/breaking. This example assumes statistical independence of the dynamics of the different sub-systems which is violated in many practical cases, for example the rotameric states of tightly packed side-chains in the hydrophobic core of a protein. To account for statistical dependencies between the sub-systems defined we will need to explicitly model the interaction between the sub-systems.

Markov random fields (MRF) is a class of statistical models which allows us to model the interactions between multiple discrete statistical variables[9]. 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[10]. 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[11,12]. 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). 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.

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 to 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.

Results

Dynamic Markov random fields

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 $\mathbf{s}$ as

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

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. While simulating eq.1 can be performed easily given $\{J_{ij}\}$ and $\{h_i\}$ using one of several sampling schemes[13,14], the inverse problem ie. estimating these parameters from equilibrium samples is not straight forward. This is a result of the complex dependency of the partition function depends on the model parameters. In particular, maximum likelihood inference relies on a 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[11,15].

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 $s_{t,i}$. 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

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

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 a J_{ii} such that sampling according to eq. 2 converges to an equilibrium distribution which is exactly eq. 1.

The dynamics exhibited by dMRFs is closely related to the Glauber dynamics[13], 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

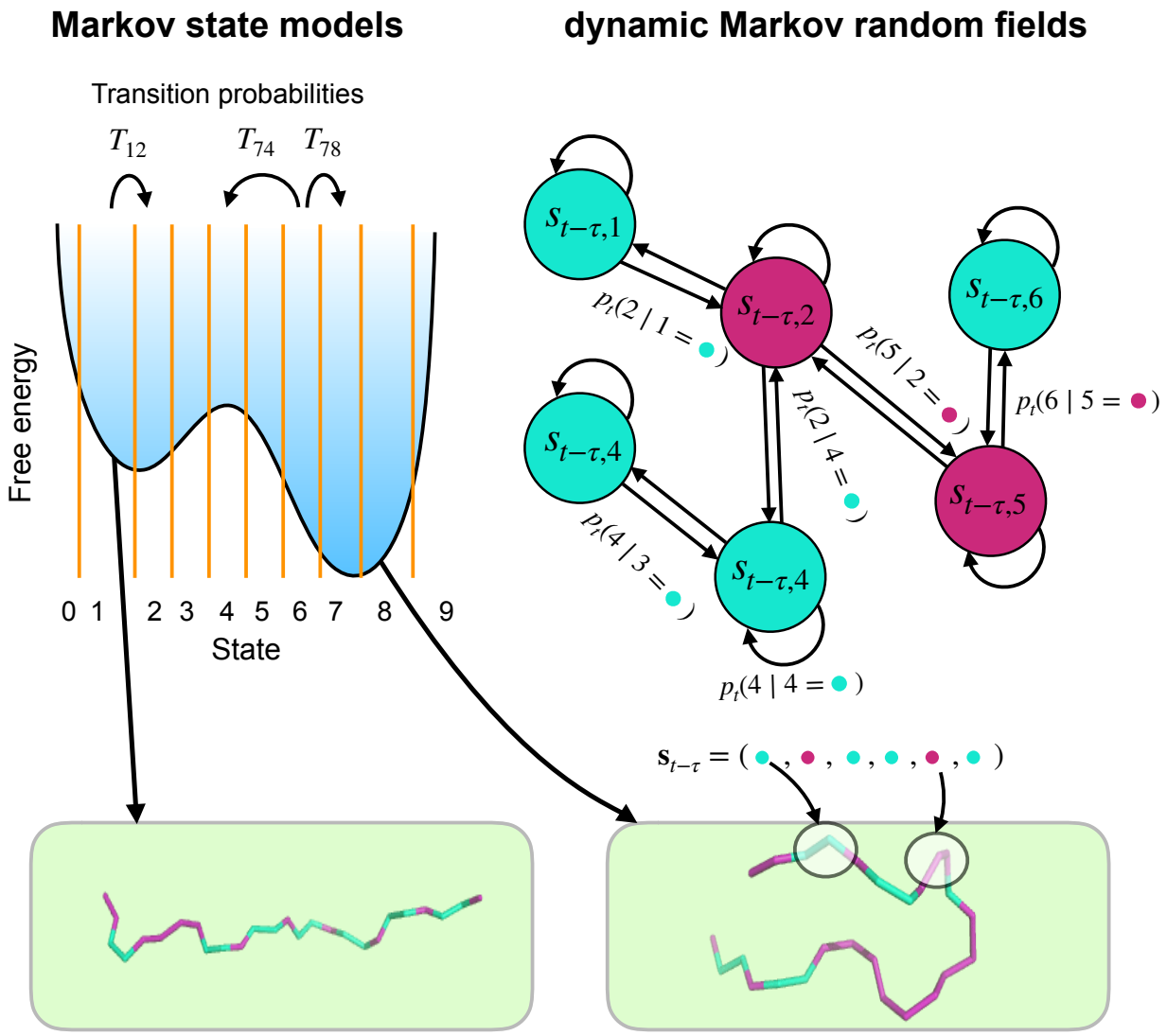

Figure 1: Illustrative comparison of molecular representations in Markov state models and dynamic Markov random fields. Markov state models represent molecular configurations as global discrete states (green rounded boxes) whereas dynamic Markov random fields represent molecular configurations as a concatenation of several local molecular configurations (white circles). The transition probabilities of in between global configurations $i$and j are encoded in a single value T_{ij} in MSMs. In dMRFs the transition probabilities are encoded as coupling between locally defined structural elements.

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 $s_{t,i}$ taking the state 1, given $\mathbf{s}_{t-\tau}$ we see that it is sigmoid in $\theta_{t-\tau,i} = \sum_j J_{ij} s_{t-\tau,j} + h_i$,

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

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,

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

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 $s_{i,t}$ 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 $s_{i,t}$ given the regressors $\mathbf{s}_{t-\tau}$[16]. 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, s_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 S1). 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 $i \rightarrow j$ 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.

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 – for molecular systems this is 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

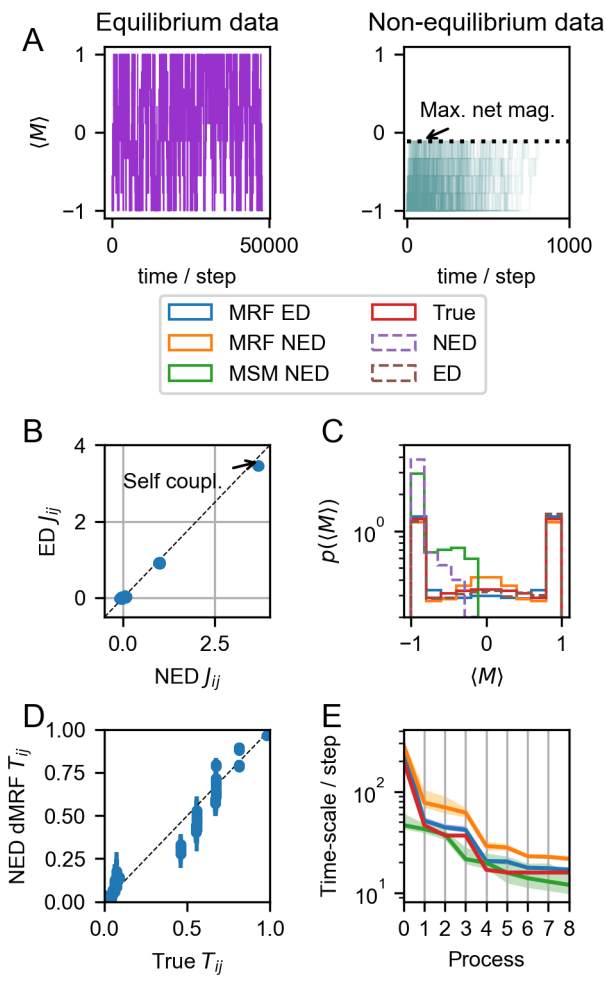

Figure 2: Inference of model parameters for a 1-dimensional 9 spin Ising model. A) Visualization of equilibrium and non-equilibrium data sets used to estimate dMRFs. B) Comparison of coupling parameters estimated using equilibrium and non-equilibrium data sets. C) Analytic stationary distributions of $\langle M \rangle$ and empirical histograms of equilibrium and non-equilibrium data sets. 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.

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 which we designed to be selectively blind towards a particular meta-stable state. We use previously published simulation data of two fast-folding proteins, villin and BBA[17], an all α and α/β folds respectively, to test this.

For villin and BBA we respectively built a five and a four state hidden Markov model (HMM). From the HMM we obtain a meta-stable state assignment for each frame in the

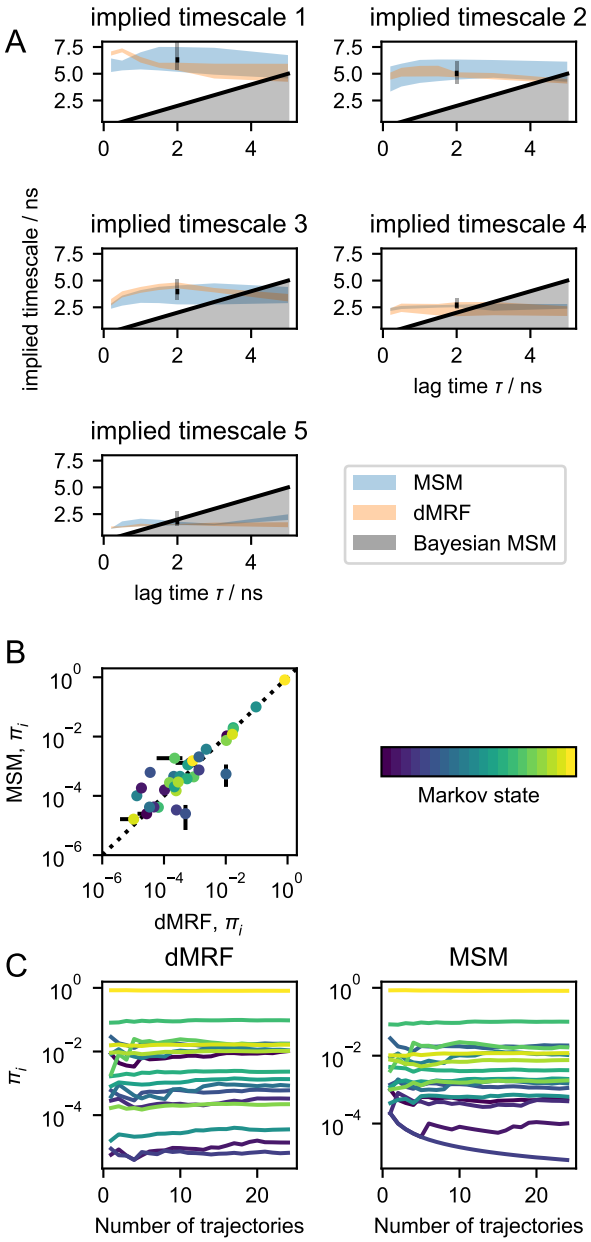

Figure 3: Comparison of dMRF and MSM describing back-bone torsion dynamics of the pentapeptide Trp-Leu-Ala-Leu-Leu (WLALL). A) 95% confidence intervals of MSM (orange) and dMRF (blue) implied time-scales. B) Global state distribution computed according to dMRF and MSMs. C) Convergence of global state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

MD trajectories, which we use to sub-sampled the original MD data, to generate training data for dMRFs construction (see SI for details). From the estimated dMRFs we simulate synthetic trajectories of the time-evolution of the sub-systems. We can directly compare the macroscopic stationary and dynamics properties from these synthetic simulations to the original MD data (Fig. 4).

Remarkably, we 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). Reconstructing molecular configurations from the sub-state represention in which the synthetic trajectories are encoded generally has a small error (Fig. 4B), suggesting that the microscopic configurations sampled by dMRFs are realistic. Finally, we inspect the time-correlation function of the Lys 71 ϕ rotameric state in the reference trajectory in villin (Fig. 4C) and .. in BBA () we find that the these are generally predicted well by the dMRFs – however, with a notable exeception. The correlation function computed by the dMRF where state 4 was left out fails to capture the characteristic decay, State 4 for villin corresponds to the folded configuration, and the transition of rotameric states in Lys 71 ϕ are tightly coupled to folding. So although we can predict realistic configurations of the folded without having explicitly observed it is not meta-stable according to the dMRF.

1. Same plots as above for BBA

Discussion and Conclusions

In this paper we introduce dynamic Markov random fields (dMRFs) and in particular their use to model molecular kinetics. The general approach encodes molecular conformations into multiple discrete sub-systems, and the global configurational state is thereby represented by a concatenation of several discrete variables. Using this representation the dMRF approach identifies the probability distribution of all of the sub-systems at a time τ in the future given their current configuration.

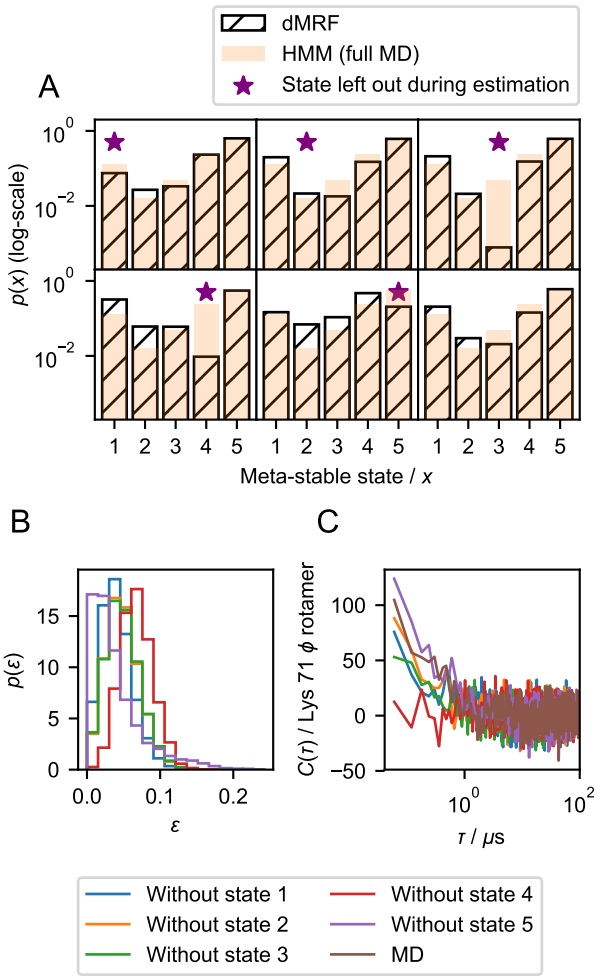

Figure 4: Prediction of macroscopic stationary and dynamic properties of the villin headpiece with dMRFs. A) Meta-stable state distributions sampled by dMRFs estimated leaving data from one of five meta-stable state out during estimation and using the full data set (hatched). Reference distribution estimated using a hidden Markov model (HMM) estimated with the full MD data-set (orange). B) Histograms of reconstruction errors of atomistic models from sub-system encoded trajectories sampled by dMRFs (Hamming distance). C) Auto-correlation function of the rotameric state of the ϕ torsion of Lys 71 in the dMRF models and the simulation data.

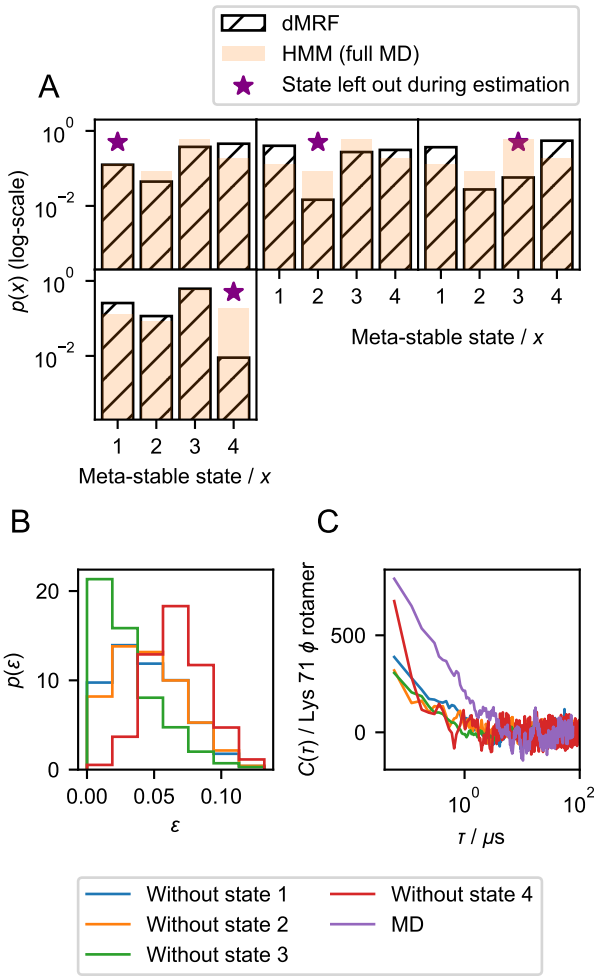

Figure 5: Prediction of macroscopic stationary and dynamic properties of BBA with dMRFs. A) Meta-stable state distributions sampled by dMRFs estimated leaving data from one of four meta-stable state out during estimation and using the full data set (hatched). Reference distribution estimated using a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

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 difficult. Third, the compact parametric form 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[18,19], 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, have seen a surge in interest in the sciences the past few years, including in the context of molecular kinetics[20,21]. A hallmark of deep learning is its ability to learn complicated non-linear functions which fulfill a prespecified set of criteria and sufficient available data[22]. The identification of sub-systems and their discretization is highly non-linear and akin to the featurization, projection and dis-

cretization process whose success critically determines the success of Markov state modelling . The VAMPnets approach recent described illustrated how this the featurization, projection and discretization pipeline of MSM building could be replaced by a deep neutral network [20]. 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. [8] Their method is motivated as a causal model, i.e. the current rotameric configurations causally encodes a probability density of rotameric states at a future time. Effectively the their approach is closely related to the dMRFs: the linear causality is generally implemented as a constrained linear regression problem with a time-lag, whereas the dMRFs can be estimated by an unconstrained logistic/softmax regression problem. The linear causality model sports many of the same advantages as the dMRFs, yet their practical implementation is more ambiguous and generally less practial, compared to dMRFs.

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 github: ....

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

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

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

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

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

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

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

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

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

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

(18) 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.

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

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

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

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

(23) Defazio, A.; Bach, F.; Lacoste-Julien, S. In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., Weinberger, K. Q., Eds.; Curran Associates, Inc., 2014; pp 1646–1654.

(24) Pedregosa, F. et al. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(25) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. *J. Chem. Phys.* **2015**, *143*, 174101.

Supporting information

Theory and numerical results

Dynamic Markov random fields with Q_i sub-system configurations

The dynamic Ising model can readily be generalized to cases where each sub-system can take on more than two configuration. In the Ising Hamiltonian the product of the two spin states results in an effective *exclusive-or* operation ($\oplus$), whose output is mapped to $\{-1, 1\}$. We here ensure consistency with this by mapping the output of an exclusive-or operation on the two spin through the linear map, $f(x) : 2x - 1$. The same map is used to ensure the local field contribution is similarly mapped to $\{-1, 1\}$. We obtain the transition density for the q spin states case

$$
\begin{aligned}
p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau}) &= \mathcal{Z}^{-1} \exp[\sum_{i}^{N}\sum_{j}^{N}\sum_{l}^{q-1}\sum_{k}^{q-1} J_{i,j}^{l,k}\left(f(s_{t,i} \oplus s_{t-\tau,j}) + h_i^l f(\mathcal{I}_l(s_{t,i})\right)] \\
&= \mathcal{Z}_a^{-1} \exp\left(\sum_{i}^{N}\sum_{l}^{q-1}\mathcal{I}_l(s_{t,i})\sum_{j}^{N}\sum_{k}^{q-1} 2J_{i,j}^{l,k}(1 - \mathcal{I}_k(s_{t-\tau,j})) - J_{i,j}^{l,k} + 2h_i^l\mathcal{I}_l(s_{t,i}) - h_i^l\right) \\
&\quad \mathcal{Z}_b^{-1} \exp\left(\sum_{i}^{N}\sum_{l}^{q-1}(1 - \mathcal{I}_l(s_{t,i}))\sum_{j}^{N}\sum_{k}^{q-1} 2J_{i,j}^{l,k}\mathcal{I}_k(s_{t-\tau,j}) - J_{i,j}^{l,k} + 2h_i^l\mathcal{I}_l(s_{t,i}) - h_i^l\right)
\end{aligned}
$$

where we have introduced the indicator function $\mathcal{I}_k(x) = \begin{cases} 1 & \text{if } x = k \\ 0 & \text{else} \end{cases}$. We see that the transition density factors into two mutually exclusive parts. Let us consider first the binary case ($q = 2$): we define the variable $\vartheta_{t-\tau,i}^l = h_i - \sum_j 2J_{ij}\mathcal{I}_l(s_{t-\tau,j}) - J_{ij}$, and recover the expression from above,

$$p(s_{t,i}=1 \mid \mathbf{s}_{t-\tau}) \;=\; \frac{1}{1+\exp[-2\vartheta^{1}_{t-\tau,i}]}.$$

More generally the transition probability density of a single sub-system takes on the form,

$$p(s_{t,i}=l \mid \mathbf{s}_{t-\tau}) \;=\; \frac{\exp[h_i^l+\sum_j^N\sum_k^{q-1}J_{i,j}^{l,k}-2J_{i,j}^{l,k}\mathcal{I}_k(s_{t-\tau,j})]}{1+\sum_m^{q-1}\exp[h_i^m+\sum_j^N\sum_k^{q-1}J_{i,j}^{m,k}-2J_{i,j}^{m,k}\mathcal{I}_k(s_{t-\tau,j})]}.$$

As for the $q=2$ case, the transition probabilities are conditionally independent, however, there are now q different outcomes:

$$p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau}) = \prod_{i=1}^{N}\frac{\exp[h_i^{s_{t,i}}+\sum_j^N\sum_k^{q-1}J_{i,j}^{s_{t,i},k}-2J_{i,j}^{s_{t,i},k}\mathcal{I}_k(s_{t-\tau,j})]}{1+\sum_m^{q-1}\exp[h_i^m+\sum_j^N\sum_k^{q-1}J_{i,j}^{m,k}-2J_{i,j}^{m,k}\mathcal{I}_k(s_{t-\tau,j})]}, \tag{4}$$

where $s_{t,i}$ indicates the configuration of sub-system i at time t. Note, that it not necessary for different sub-systems to have the same number of configuration.

Maximum likelihood estimation of $\{J_{ij}^{kl}\}$ and $\{h_i^l\}$ given a time-series of sub-system configurations, $\mathbf{S}=\{\mathbf{s}_0,\mathbf{s}_\tau,\ldots,\mathbf{s}_{T\tau}\}$, corresponds to solving N soft-max auto-regression problems, which is the multi-category generalization of logistic regression. There are, similarly, efficient schemes available to solve problems of this type.

Maximum Likelihood estimation of dMRFs

Given a time-series of state-configurations, $\mathbf{S}=\{\mathbf{s}_0,\mathbf{s}_\tau,\ldots,\mathbf{s}_{T\tau}\}$, the likelihood function of the couplings and fields is given by,

$$\ell(\mathbf{J},\mathbf{h} \mid \mathbf{S}) \;=\; \prod_{t=\tau}^{\tau(T-1)} p(\mathbf{s}_{t,i} \mid \mathbf{s}_{t-\tau}), \tag{5}$$

where $p(\mathbf{s}_{t,i} \mid \mathbf{s}_{t-\tau})$ is (4). Similarly, we can write down an a posteriori functional with a Laplacian prior,

$$\mathcal{P}(\mathbf{J}, \mathbf{h} \mid \mathbf{S}) \;=\; \left(\prod_{i,j} \frac{\gamma}{2} \exp(-\gamma \, |J_{ij}|) \right) \ell(\mathbf{J}, \mathbf{h} \mid \mathbf{S}), \tag{6}$$

which corresponds to a L_1 regularized likelihood with a regularizer γ. Maximization of (5 and 6) can be done using naive gradient ascent, however, this strategy proved computationally inefficient. Throughout this paper, we use the SAGA optimization algorithm [23], as implemented in the scikit-learn python library [24], to optimize (6).

Spectral analysis of dMRFs

Performing a general symbolic analysis of the spectral properties of dMRFs is not trivial as the transition matrices describing all by the simplest systems (with 1 and 2 binary subsystems) are intractable. In this section we show the simplest case analytically, to build some intution to interpret numerical results for more interesting cases.

Consider the case with a single binary, uncoupled sub-system s ($N = 1$), here the possible transitions full system are $(-1 \to -1), (-1 \to 1), (1 \to -1)$ and $(1 \to 1)$. We may use the expression of transition probabilities from the main text (eq. 3) to define the transition probability matrix of this system

$$\begin{aligned} \mathbf{T} \;=\; & \begin{pmatrix} \exp[J-h] + \exp[-J+h] & 0 \\ 0 & \exp[-J-h] + \exp[J+h] \end{pmatrix}^{-1} \\ & \times \begin{pmatrix} \exp[J-h] & \exp[-J+h] \\ \exp[-J-h] & \exp[J+h] \end{pmatrix}. \end{aligned} \tag{7}$$

The eigenvalues of $\mathbf{T}$ are

$$\lambda_0 = 1, \quad \lambda_1 = \frac{\sinh(2J)}{\cosh(2h) + \cosh(2J)} \overset{h=0}{\Longrightarrow} \tanh(J),$$

where the zeroth eigenvalue corresponds to the eigenvector encoding the stationary state ($\mathbf{v}_0 = (v, v)$), and the first eigenvalue corresponds to the eigenvector encoding the in the sub-system changing configuration operation ($\mathbf{v}_1 = (-w, w)$), respectively. From the latter of these eigenvalues we can compute $k\tau\Delta t$ as,

$$\begin{aligned} k\tau\Delta t &= -\log(\lambda_1) \\ &= -\log\left(\frac{\sinh(2J)}{\cosh(2h) + \cosh(2J)}\right), \end{aligned}$$

and for $h = 0$

$$k\tau\Delta t = -\log\left(\tanh(J)\right),$$

where k is the average rate of configurational change in s, Δt defines the time-scale (fx. the trajectory frame saving interval in simulation) and τ is the unitless integer used when estimating J and h. When $k\tau\Delta t < 1$ our system is meta-stable on the time-scale of the model (Fig. 6A). Conversely, if $k\tau\Delta t > 1$ we are unable to resolve the dynamics of configurational change s. The boundary $k\tau\Delta t = 1$ corresponds to a minimal self-coupling of $J_{ii} = \tanh^{-1}(e^{-1})$ (Fig. 6A).

We can compute the same quantity for a single sub-system $s_{i,t}$ in the context $N - 1$ other sub-systems in a particular configuration a certain sub-system $\mathbf{s}_{t-\tau}$ by identifying an effective h, as $\hat{h}_{i,\mathbf{s}_{t-\tau}}$

$$
\begin{aligned}
p(s_{t,i} \mid \mathbf{s}_{t-\tau}) &\propto \exp\left(s_{t,i}(\sum_j J_{ij} s_{t-\tau,j} + h_i)\right) \\
&= \exp\left(s_{t,i} J_{ii} s_{t-\tau,i} + s_{t,i} \underbrace{(\sum_{j \neq i} J_{ij} s_{t-\tau,j} + h_i)}_{\hat{h}_{i,\mathbf{s}_{t-\tau}}}\right) \\
&= \exp\left(s_{t,i}(J_{ii} s_{t-\tau,j} + \hat{h}_{i,\mathbf{s}_{t-\tau}})\right).
\end{aligned}
$$

In other words, all sub-systems with non-zero couplings to sub-system i will affect the meta-stability of sub-system i in a manner which is dependent on the global configuration $\mathbf{s}_{t-\tau}$ and the self-coupling J_{ii}.

Numerical analysis of 1 and 9 spin 1D Ising models with Glauber dynamics

We numerically compute the characteristic time-scales ($t = 1/k$) for 1 and 9 spin 1D Ising models with periodic boundaries as a function of $\gamma = \tanh(2J_{ij})$ and α (sub-system configuration change rate) from transition probability matrices which are time-discretizations of the Master equation dynamics prescribed by Glauber[13] (Fig 6B,C). The parameter α is the the same for all sub-systems, $h_i = 0$ and J_{ij} is the same value for all (i, j) where $i = (j + 1) \mod N$, that is, all adjacent pairs of sub-systems. For the 1 spin case and only this case, this is J_{ii}.

In both cases we see the time-scales approach to basal rate α as $J_{ij} \to 0$ (Fig 6B,C). In the 9 spin case this limit corresponds to 9 identical copies of the 1 spin case (Fig 6C). The characteristic time-scales change non-linearly as a function of γ; in the 9-spin case the different time-scales cross suggesting that different global configurational transitions become relatively slower or faster depending on the sub-system coupling strength J_{ij}. As a function of α the time-scales decrease as $t_i \propto \frac{1}{\alpha}$.

If we inspect the estimated J_{ii} as $\gamma \to 0$ we see how the self-couplings converge to the value anticipated for α (Fig 6D). The self-couplings J_{ii} increase non-linearly with γ corresponding to a reduction of the rate of change in the sub-systems. This observation is consistent with our intuition from above and the characteristic time-scales. The estimated J_{ii} values shown are means and standard deviations from dMRFs estimated using trajectory realizations ($2 \cdot 10^5$ steps each) of the 9 spin Ising model.

Methods

dMRF analysis of 1D Ising model with Glauber dynamics

A reference MSM describing Glauber dynamics of a 9 spin 1D Ising model was constructed by discretization of the master equation in ref.[13], with $\gamma = \tanh(2J_{ij}) = 0.95$ and flip-rate $\alpha = 0.1\,\mathrm{step}^{-1}$. The MSM encoded the transition probabilities between the 512 global Markov states of the 9 spin Ising model in a 512×512 transition matrix. We generated two data-sets using this MSM. The first data set emulates a non-equilibrium scenerio: 16 trajectories were initialized in the Markov state corresponding to all sub-systems being in state -1, and termininated if the average of the sub-system states in the next step was larger than 0, however, with a maximum length of 2000 steps (approximately ten times the slowest implied time-scale). The second data-set emulates an equilibrium scenerio were the 16 simulations were uninhibited, however, limited in length to match those of the non-equilibrium case to ensure consistent statistics. The initialization in the second case was identical to the non-equilibrium case.

For each of the 16 trajectories for each scenerio we estimated a dMRF using mildly L_1 regularized maximum likelihood with ($\gamma = 1/50$). The dMRFs were used to compute means and confidence intervals shown.

dMRF and MSM construction for WLALL peptide

A total of 24, 500 nanosecond molecular dynamics simulations of the WLALL penta-peptide were encoded into 8 binary sub-systems by discretizing the back-bone torsion angles into rotameric states. Specifically, the following discretization maps were used:

$$\sigma(\phi) = \begin{cases} -1 & \text{if } \phi < 0^\circ \\ 1 & \text{if } \phi \geq 0^\circ \end{cases}, \qquad \sigma(\psi) = \begin{cases} -1 & \text{if } \psi < 80^\circ \\ 1 & \text{if } \psi \geq 80^\circ \end{cases} \tag{8}$$

for all back-bone torsions apart from the N-terminal ψ.

For MSM construction all possible global configurations of the local sub-systems (a total of $2^8 = 256$) were enumerated and each of the simulation frames were assigned to a global Markov state. From the 24 trajectories we generated 5 groups, each of which had its own 20% of the trajectories left-out. These data-sets were used for estimating 5 maximum likelihood MSMs and 5 maximum likelihood dMRFs which were used to compute means and confidence intervals shown. MSMs and dMRFs with a lag-time of 2 ns were used for comparison of the predicted stationary distributions and implied time-scales. For reference Bayesian posterior samples using the full MD data-set was also carried out[25] using the same lag-time.

Finally, we estimated 24 further dMRFs and MSMs with lag time 2 ns each of which include one trajectory more than the previous starting with 1 randomly selected. The inclusion of data is done randomly without replacement, but is consistent between the dMRFs and MSMs shown. From these dMRFs and MSMs global stationary distributions of Markov states were computed and compared.

All dMRFs were estimated with a regularizer $\gamma = 1$; varying the regularizer up to one order of magnitude did not change the dMRFs microscopic properties in any appreciable way. Estimation was performed by solving 8 logistic regression problems (one per sub-system) for each dMRF using the SAGA solver as implemented in the scikit-learn python library.

dMRF construction for Villin and BBA fast folders

We use the simulation data of the fast-folders villin and BBA generously made available to us by DE Shaw Research (New York, NY). These trajectories were encoded into torsion-based sub-systems as for the WLALL peptide (eq. 8).

Supporting figures

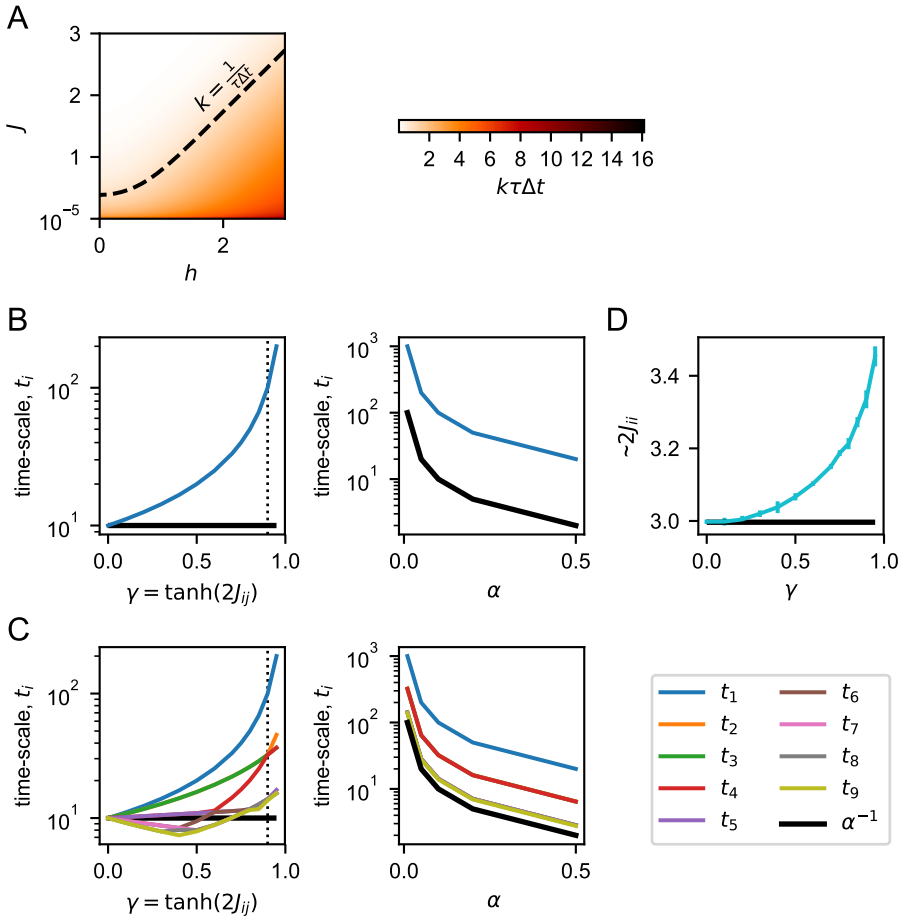

Figure 6: Spectral analysis of dMRF dynamics on 1D Ising models. A) Meta-stability phase-diagram in 1 spin Ising model with dMRF dynamics. B) Implied time-scales in 1 spin Ising model with time-discrete Glauber dynamics with as a function of self-coupling (γ) and subsystem configuration change rate (α). C) Implied time-scales in 9 spin 1D Ising model with time-discrete Glauber dynamics with as a function of sub-system coupling (γ) and subsystem configuration change rate (α). D) Inferred average self-coupling in dMRFs estimated from 9 spin 1D Ising model data with time-discrete Glauber dynamics.

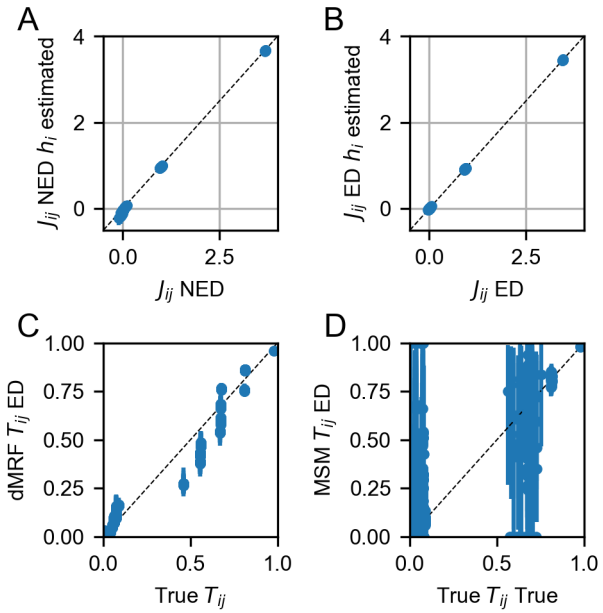

Figure 7: Comparison of parameters estimated for dMRFs and MSMs for the 9 spin 1D Ising model. A and B) comparison estimated $\{J_{ij}\}$ values for NED and ED data sets when local biases were estimated or not. C) Scatter-plot of true transition probabilities versus predicted transition probabilities from dMRF estimated using ED. D) Scatter-plot of true transition probabilities versus values from a MSM estimated using ED.

Supporting GIFs

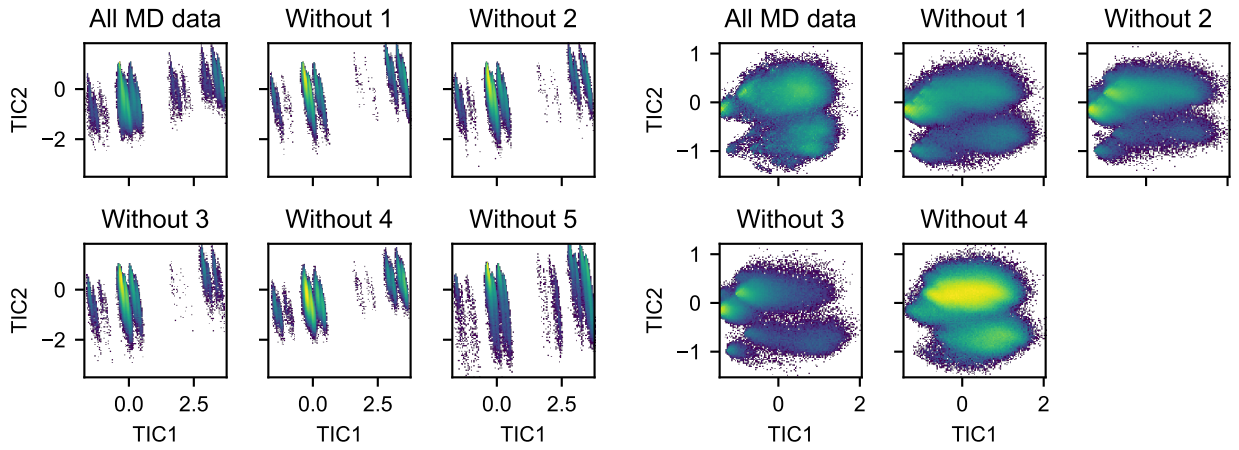

Figure 8: Configurational space sampling in two time-lagged independent components. Comparison of molecular dynamic trajectories and trajectories sampled according dMRF models estimated on subsets of the molecular dynamics data. Left: Villin headpiece, Right: BBA.

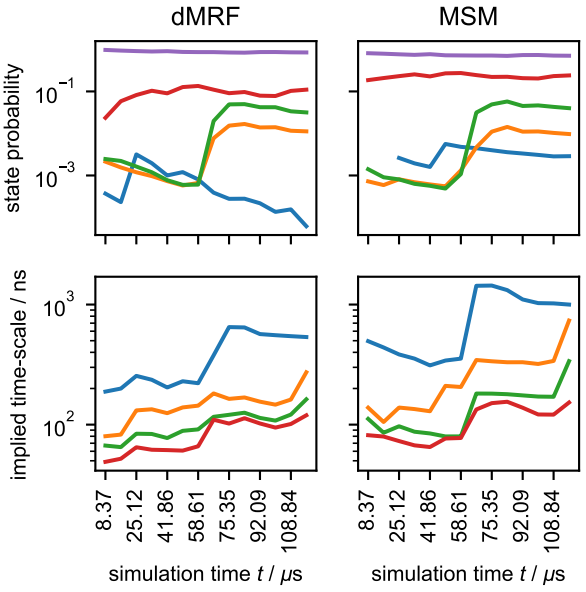

Figure 9: Convergence of stationary and dynamic properties in dMRFs and MSMs estimated with increasing volumes of simulation data. (Same for BBA missing, will prep when mature model is available)

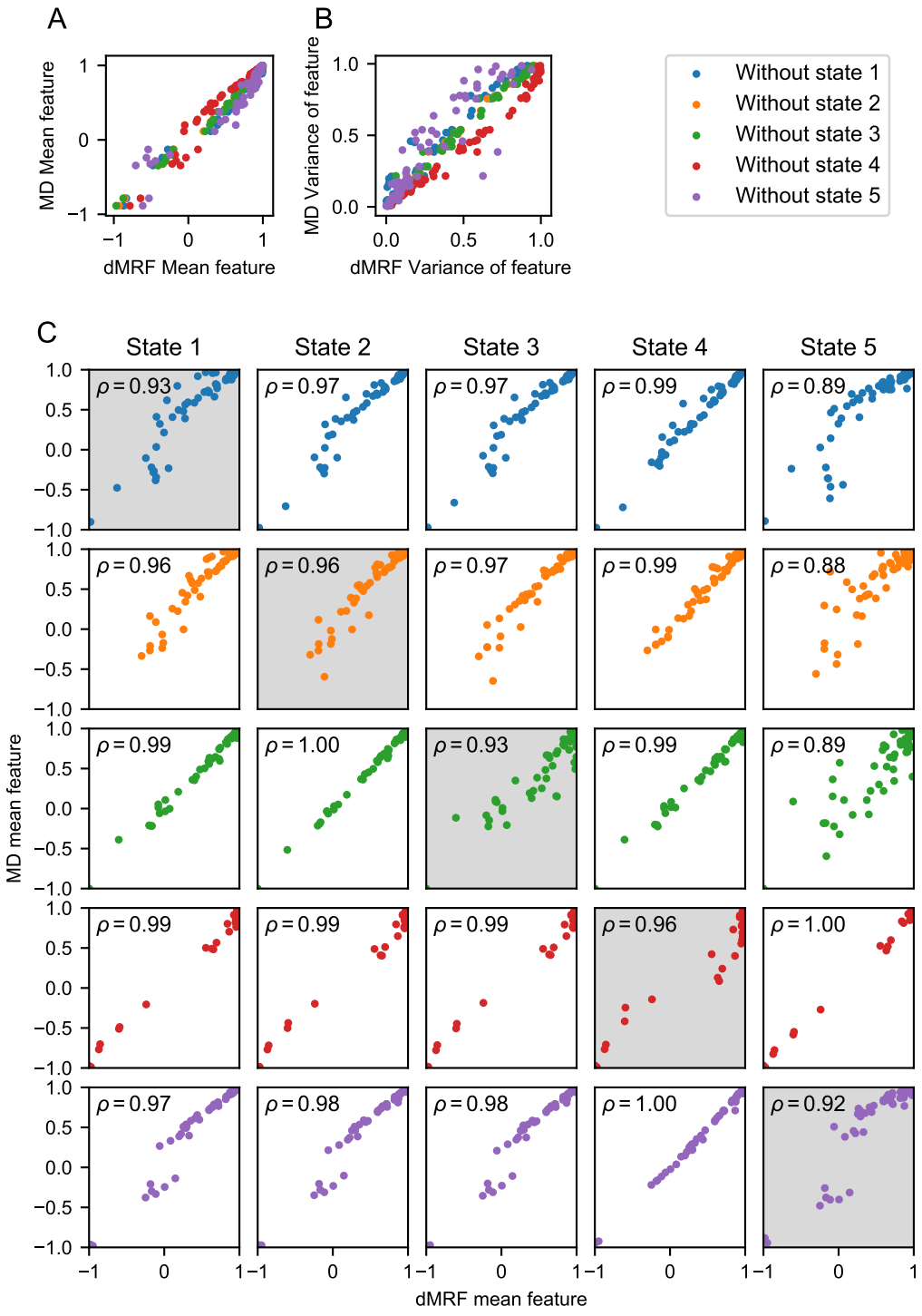

Figure 10: Comparison of sub-system statistics in dMRFs and MD data for the villin headpiece. A) Correlation of mean sub-system configuration predicted by dMRFs and observed in MD. B) Correlation of the variance of sub-system configurations predicted by dMRFs and observed in MD. C) Correlation of average sub-system predictions from dMRF with corresponding values observed in MD.

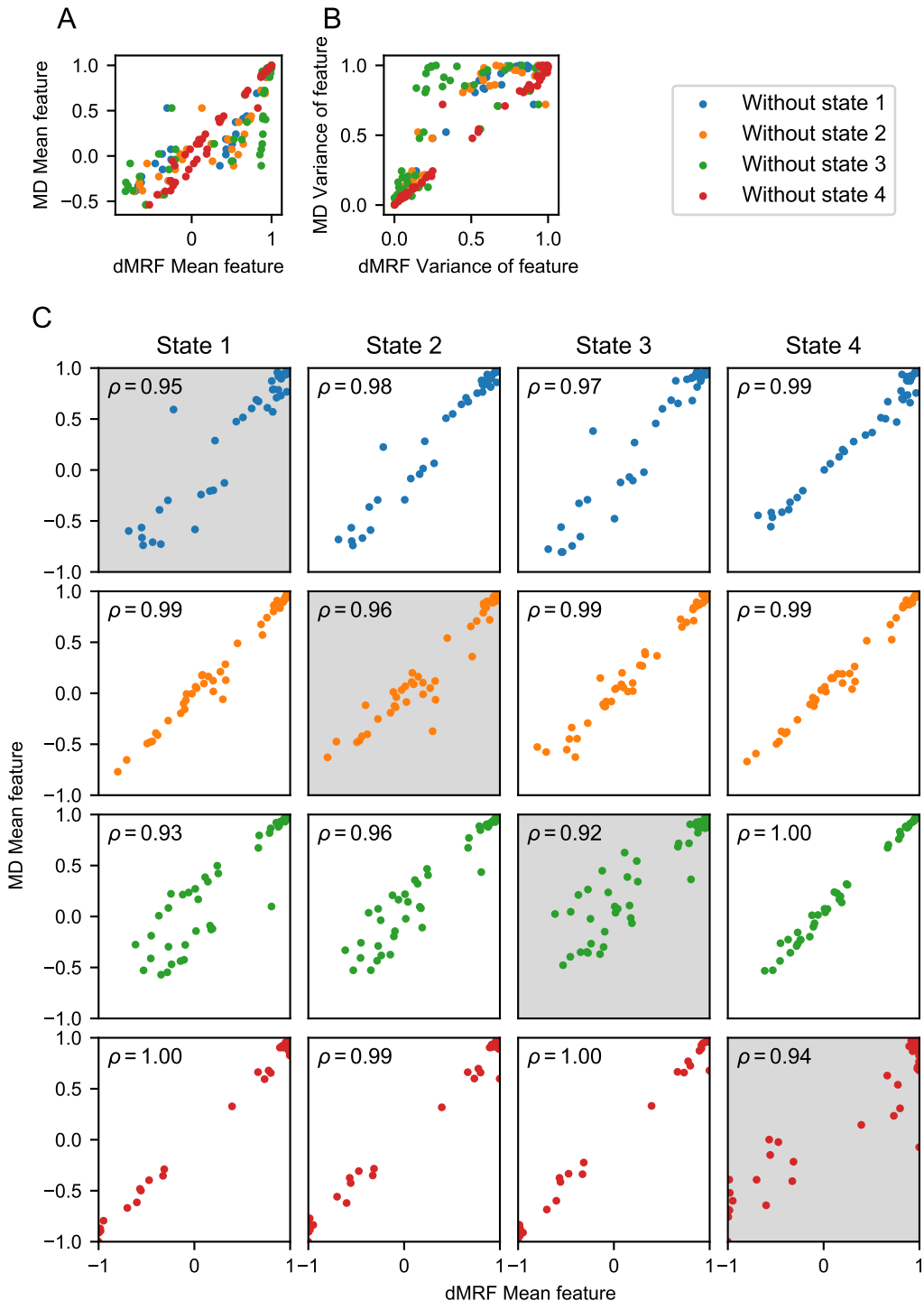

Figure 11: Comparison of sub-system statistics in dMRFs and MD data for BBA. A) Correlation of mean sub-system configuration predicted by dMRFs and observed in MD. B) Correlation of the variance of sub-system configurations predicted by dMRFs and observed in MD. C) Correlation of average sub-system predictions from dMRF with corresponding values observed in MD.