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

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

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

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 $\vartheta_{t-\tau,i}^l = h_i - \sum_j 2J_{ij}\mathcal{I}_l(\sigma_{t-\tau,j}) - J_{ij}$, and recover the expression from above,

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

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

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

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

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

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 configurations.

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. This is implemented as part of the graphtime library (http://www.github.com/markovmodel/graphtime/).

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,

$$\ell(\mathbf{J}, \mathbf{h} \mid \mathbf{S}) \quad = \quad \prod_{t=\tau}^{\tau(T-1)} p(\mathbf{s}_{t,i} \mid \mathbf{s}_{t-\tau}), \tag{2}$$

where $p(\mathbf{s}_{t,i} \mid \mathbf{s}_{t-\tau})$ is (1). Similarly, we can write down an a posteriori functional with a Laplacian prior,

$$\mathcal{P}(\mathbf{J},\mathbf{h}\mid\mathbf{S}) \;=\; \left(\prod_{i,j}\frac{\gamma}{2}\exp(-\gamma\,|J_{ij}|)\right)\ell(\mathbf{J},\mathbf{h}\mid\mathbf{S}), \tag{3}$$

which corresponds to a L_1 regularized likelihood with a regularizer γ. Maximization of (2 and 3) can be done using naive gradient ascent, however, this strategy proved computationally inefficient. Throughout this paper, we use the SAGA optimization algorithm[?], as implemented in the scikit-learn python library[?], to optimize (3).

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 intuition for interpretation of numerical results for more interesting cases.

Consider the case with a single binary, uncoupled sub-system σ ($N=1$), 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 (eq. 3 in the main text) to define the transition probability matrix of this system

$$\begin{aligned}\mathbf{T} \;=\; &\begin{pmatrix} \exp[J-h]+\exp[-J+h] & 0 \\ 0 & \exp[-J-h]+\exp[J+h] \end{pmatrix}^{-1} \\ &\times\begin{pmatrix} \exp[J-h] & \exp[-J+h] \\ \exp[-J-h] & \exp[J+h] \end{pmatrix}.\end{aligned} \tag{4}$$

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 σ, Δt defines the time-scale (fx. the trajectory frame saving interval in simulation) and τ is the unit-less integer used when estimating J and h. In other words, $\tau\Delta t$ is the lag-time. When $k\tau\Delta t < 1$ our system is meta-stable on the time-scale of the model (Fig. S1A). Conversely, if $k\tau\Delta t > 1$ we are unable to resolve the dynamics of configurational change in s. The boundary $k\tau\Delta t = 1$ corresponds to a minimal self-coupling of $J_{ii} = \tanh^{-1}(e^{-1})$ (Fig. S1A).

We can compute the same quantity for a single sub-system $\sigma_{i,t}$ 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 $\hat{h}_{i,\mathbf{s}_{t-\tau}}$

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

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. We vary the sub-system couplings $\gamma = \tanh(2J_{ij})$ and α (sub-system configuration change rate) for transition probability matrices which are time-discretization of the Master equation dynamics prescribed by Glauber[?] (Fig S1B,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 S1B,C). In the 9 spin case this limit corresponds to 9 identical copies of the 1 spin case (Fig S1C). 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 S1D). 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

General notes on dMRF estimation

In general the dMRFs presented in this paper are estimated using the L_1 regularized maximum likelihood (or *maximum a posteriori*) scheme defined above. We have implemented a library `graphtime` to achieve this. `graphtime` is a small object oriented library providing basic functionality including estimation and simulation of dMRFs as well as reconstruction of MSMs from dMRFs. The library makes extensive use of the `scikit-learn` and `numpy` for computation. The library is distributed freely under the LGPL and is available at http://www.github.com/markovmodel/graphtime

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.[?] , 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 scenario: 16 trajectories were initialized in the Markov state corresponding to all sub-systems being in state -1, and terminated 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 scenario 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.

In both scenarios each of the 16 data sets were used estimate a dMRF using mildly L_1 regularized maximum likelihood ($\gamma = 1/50$). 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:

$$\sigma(\phi) = \begin{cases} -1 & \text{if } \phi < 0^\circ \\ 1 & \text{if } \phi \geq 0^\circ \end{cases}, \qquad \sigma(\psi) = \begin{cases} -1 & \text{if } \psi < 80^\circ \\ 1 & \text{if } \psi \geq 80^\circ \end{cases} \tag{5}$$

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 L_1 regularized maximum 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 MCMC sampling of MSMs from a posterior distribution defined using the Bayesian scheme[?] , was also carried out using the same lag-time.

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 using the graphtime library.

Hidden Markov model and 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 are encoded into the binary torsion-based sub-systems as for the WLALL peptide (eq. 5) resulting in 66 and 53 active sub-systems in for villin and BBA, respectively. Active sub-system correspond to features which under go at least one transition in their state during the training data. These data-sets were used to generate villin and BBA dMRFs using a $\tau = 300\,\Delta t$, where $\Delta t = 0.2\,\text{ns}$. A regularization constant of $\gamma = \frac{\sqrt{N_{frame}}}{1000}$ was used in both cases, where N_{frames} indicates the number of frames in the data-set.

Enumeration of all the possible Markov states for these systems is intractable, 2^{66} and 2^{53} states respectively. To facilitate HMM estimation we embedded the the discrete features into low-dimensional space using time-lagged independent component analysis (TICA). The resulting 7 and 17 dimensional data for villin and BBA respectively, were clustered into 256 and 384 disjoint clusters using the K-means++ algorithm. The discretized time-series resulting from assigning each molecular dynamics frame to a cluster center was used as a basis to estimate hidden Markov models (HMMs) with 5 and 4 hidden states for villin and BBA respectively. For villin the HMM was estimated with a lag-time of $\tau = 300\,\Delta t$, whereas for BBA $\tau = 1500\,\Delta t$ was used. Each of the hidden states corresponds to a meta-stable molecular configuration.

Using the HMMs we assign each MD frame to one of the meta-stable states using the Viterbi algorithm for both villin and BBA. Using the state assignments we generated data-sets which selectively kept out all observations in one of the meta-stable states while keeping all other observations. This procedure generated 5 and 4 data-sets for villin and BBA respectively. Each of the data-sets are composed of fragments of the original MD data-sets,

Algorithm 1 dMRF to MD reconstruction.The reconstruction error reported is given by the hamming distance evaluated to select the frame.

1. $\mathbf{S} = (\mathbf{s}_0, \mathbf{s}_\tau, ...)$ - contains sub-system encoding of original MD trajectory
2. $\mathbf{X} = (\mathbf{x}_0, \mathrm{x}_\tau, ...)$ - contains all atom representation of original MD trajectory
3. $\hat{\mathbf{S}} \sim p(\mathbf{s}_t \mid \mathbf{s}_{t-\tau})$ - Generate trajectory from dMRF
4. for each $\hat{\mathbf{s}}$ in $\hat{\mathbf{S}}$
   (a) $k \leftarrow \arg\min\left(\mathrm{hamming}(\hat{\mathbf{s}}, \mathbf{S})\right)$ - find closest frame in sub-system encoding
   (b) $\hat{\mathbf{X}}_i \leftarrow \hat{\mathbf{X}}_k$ - copy coordinates from original MD trajectory.

every one of these fragments are required to be at least the length of τ, the lag-time used to estimate the model. Each data-set was used to estimate a dMRF using $\gamma = \frac{\sqrt{N_{frame}}}{1000}$ and $\tau = 300\,\Delta t$ for villin and $\gamma = \frac{\sqrt{N_{frame}}}{2000}$ and $\tau = 400\,\Delta\mathrm{t}$ for BBA. The results presented were robust to minor perturbations in either of these parameters.

Reconstructing molecular trajectories from dMRF trajectories

We reconstruct molecular configurations from the dMRF trajectories by finding the MD frame in a reference simulation with is closest to the synthetic frame. The reported reconstruction errors (ϵ) are the Hamming distance between the synthetic frame and the selected reference MD frame.

Supporting figures

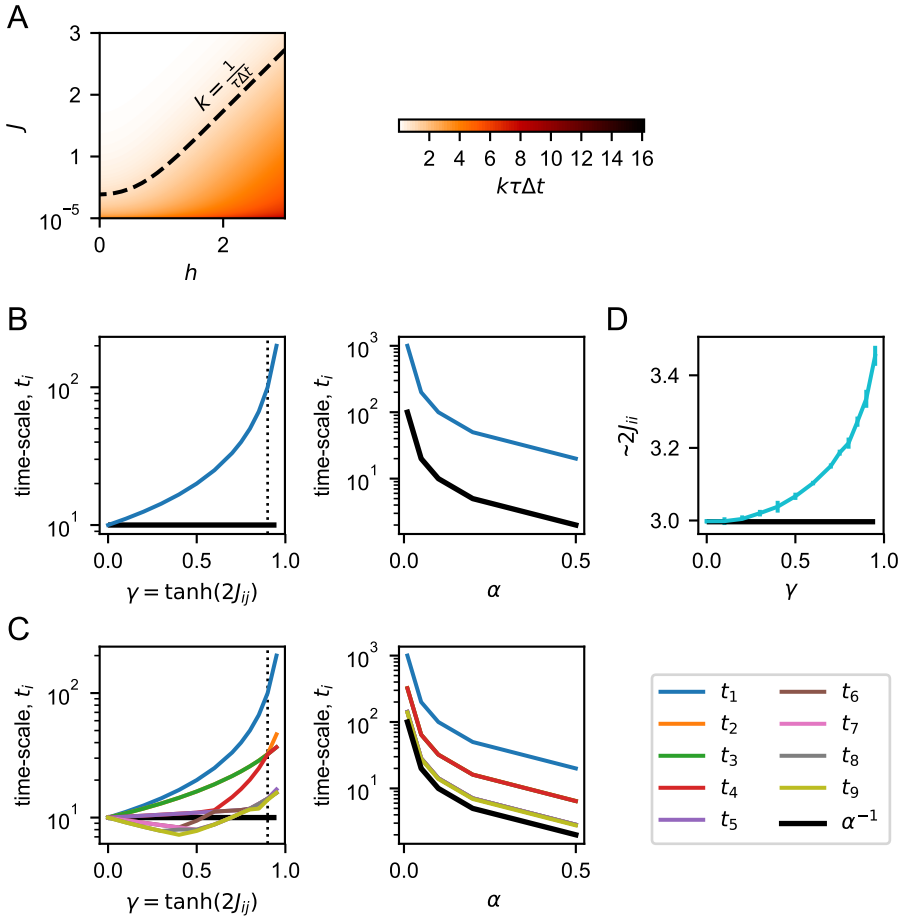

Figure S1: Spectral analysis of dMRF dynamics on 1D Ising models. A) Meta-stability phase-diagram in 1 spin Ising model with dMRF dynamics. B) Implied time-scales in 1 spin Ising model with time-discrete Glauber dynamics as a function of self-coupling (γ) and sub-system configuration change rate (α). C) Implied time-scales in 9 spin 1D Ising model with time-discrete Glauber dynamics as a function of sub-system coupling (γ) and sub-system configuration change rate (α). D) Inferred average self-coupling in dMRFs estimated from 9 spin 1D Ising model data with time-discrete Glauber dynamics.

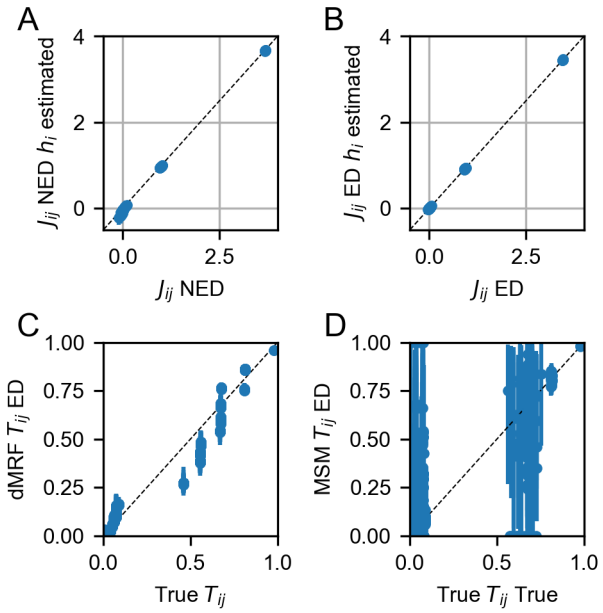

Figure S2: Comparison of parameters estimated for dMRFs and MSMs for a 9 spin 1D Ising model. A and B) comparison of estimated $\{J_{ij}\}$ values for NED and ED data sets when local biases $\{h_i\}$ are estimated or not. C) Scatter-plot of true transition probabilities versus predicted transition probabilities from dMRF estimated using ED. D) Scatter-plot of true transition probabilities versus values from a MSM estimated using ED.

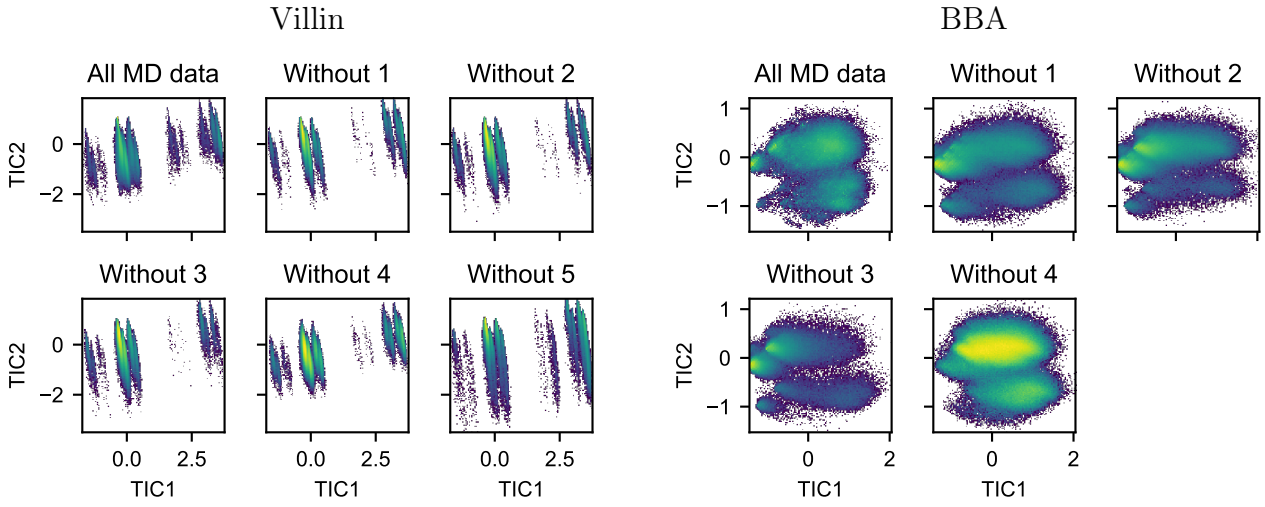

Figure S3: Configurational space sampling in the two principal time-lagged independent components. Comparison of spatial sampling by molecular dynamics trajectories and trajectories sampled according dMRF models estimated on subsets of the molecular dynamics data. The subsets each have the meta-stable state indicated by the plot sub-title left out.

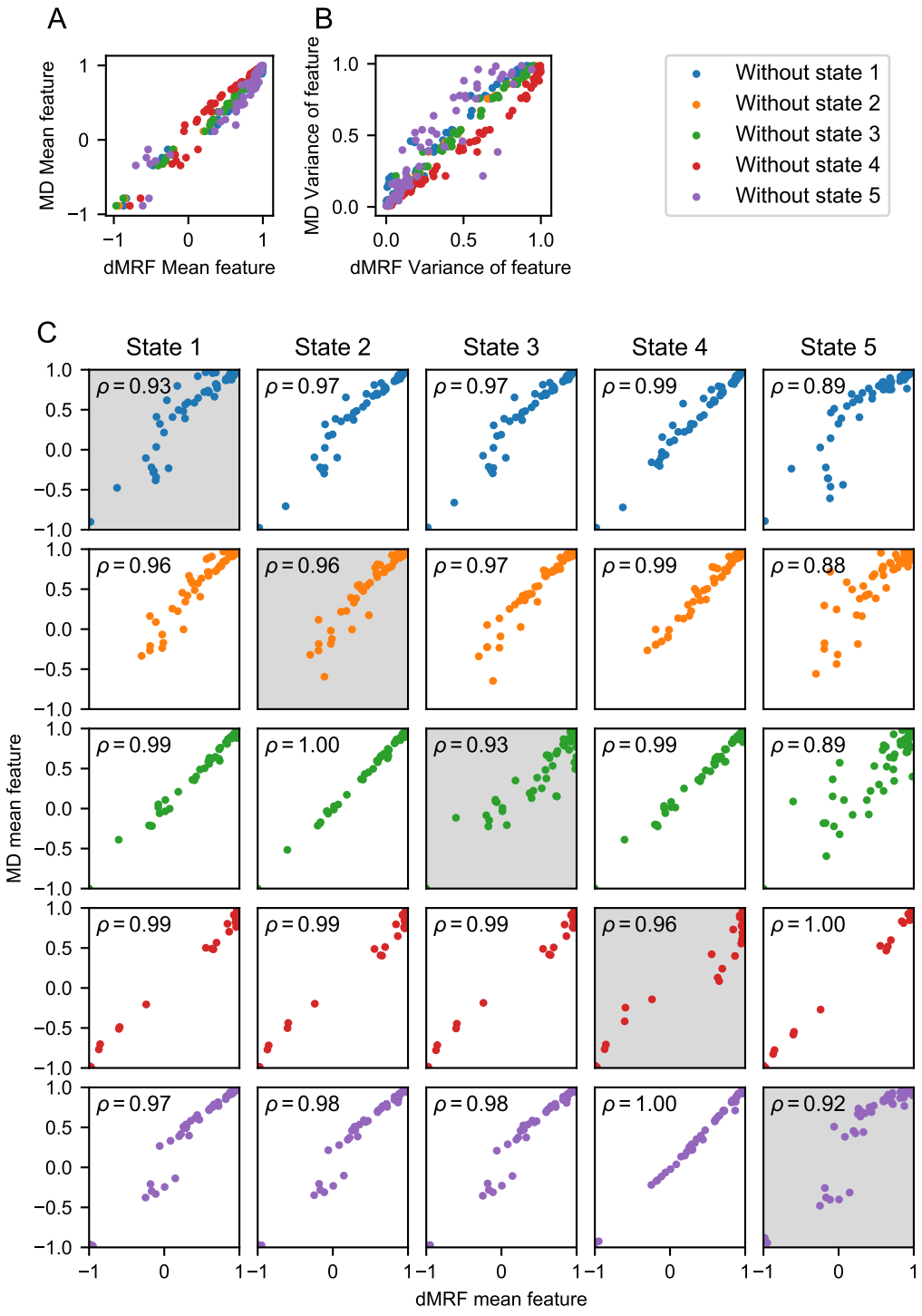

Figure S4: Comparison of sub-system statistics in dMRFs and MD data for the villin headpiece. A) Correlation of mean sub-system configuration predicted by dMRFs and observed in MD. B) Correlation of the variance of sub-system configurations predicted by dMRFs and observed in MD. C) Correlation of average sub-system predictions from dMRF with corresponding values observed in MD.

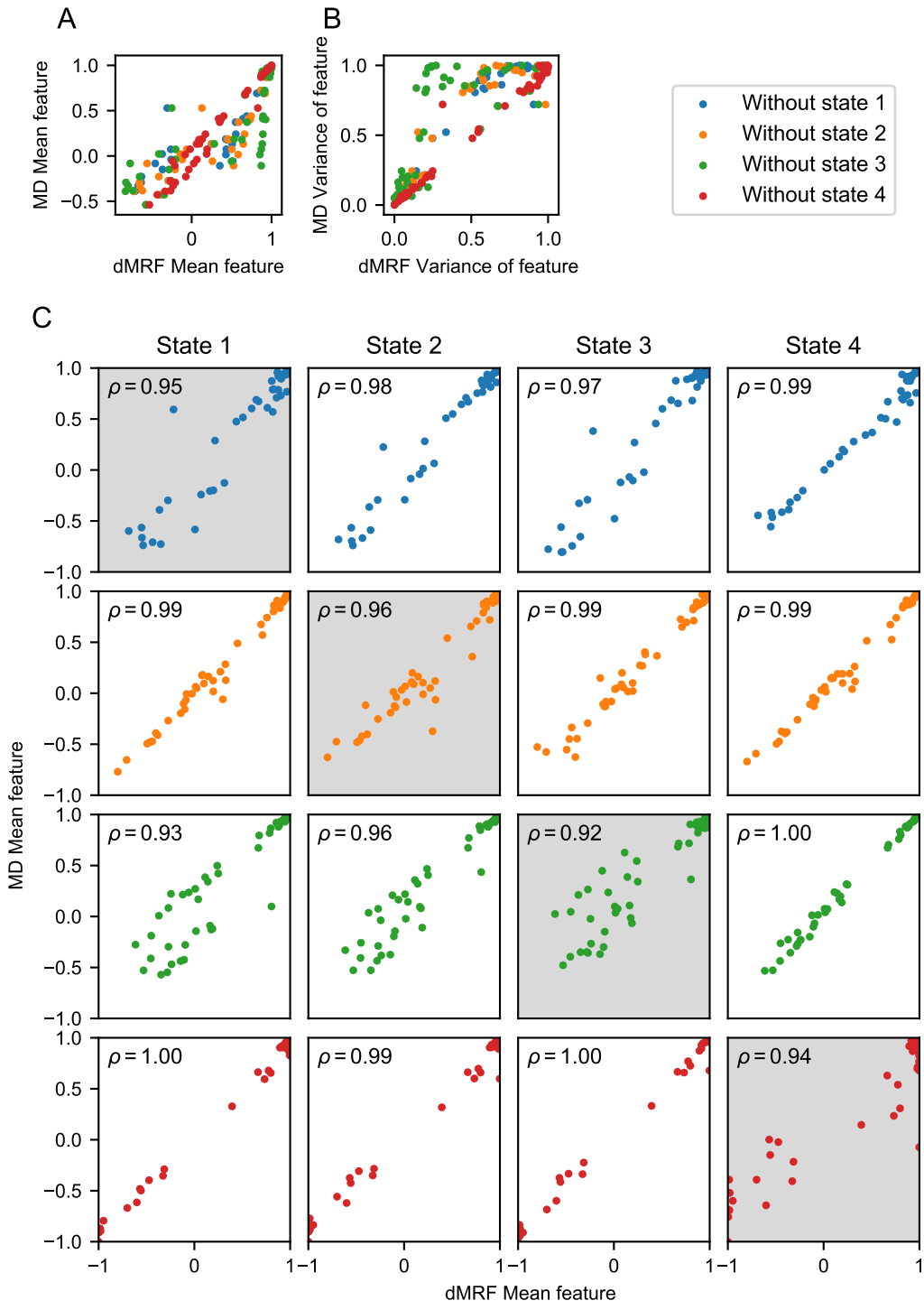

Figure S5: Comparison of sub-system statistics in dMRFs and MD data for BBA. A) Correlation of mean sub-system configuration predicted by dMRFs and observed in MD. B) Correlation of the variance of sub-system configurations predicted by dMRFs and observed in MD. C) Correlation of average sub-system predictions from dMRF with corresponding values observed in MD.

Villin BBA

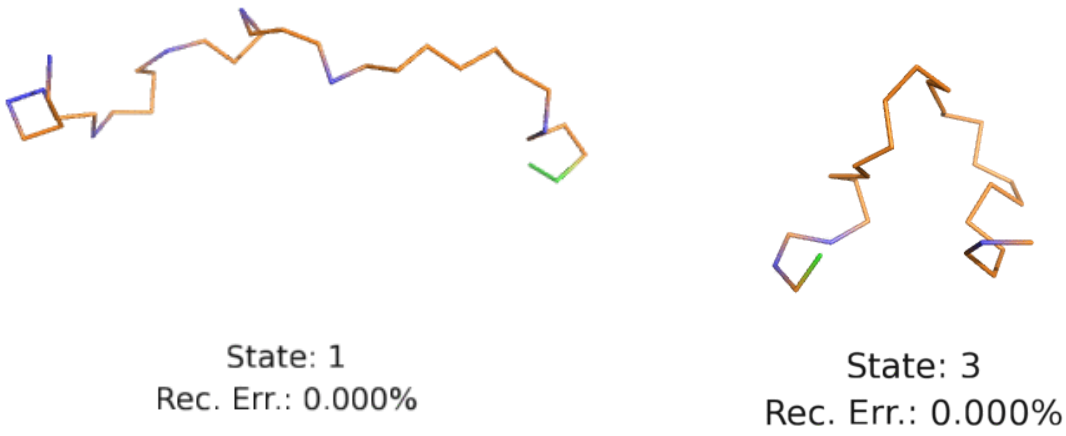

Figure S6: 250 frames reconstructed from dMRF simulations of models estimated without folded state (state 4). GIF files showing animation can be found in other supporting material.

References

() 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.

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

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

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