# Combining Experimental and Simulation Data of Molecular Processes via Augmented Markov Models

Simon Olsson, Hao Wu, Fabian Paul, Cecilia Clementi, and Frank Noé

July 4, 2017

SI Appendix

## 1 Theory

### 1.1 Kullback-Leibler optimal coupling between experimental and simulation ensembles

Given data from a simulation ensemble with equilibrium distribution $\boldsymbol{\pi}$ and an unknown experimental ensemble with equilibrium distribution $\hat{\boldsymbol{\pi}}$, we wish to find an expression of $\hat{\boldsymbol{\pi}}$ in terms of $\boldsymbol{\pi}$ which minimizes their Kullback-Leibler functional subject to the appropriate constraints. Performing this optimization is consistent maximum entropy principle. We require $K$ experimentally measured quantities to have expectation values according to $\hat{\boldsymbol{\pi}}$ that match corresponding unknown expectation values $\hat{\boldsymbol{m}}$ and that $\hat{\boldsymbol{\pi}}$ is normalized. We enforce these constraints using $K+1$ Lagrange multipliers - we get:

$$
\begin{aligned}
S &= -\sum_i \hat{\pi}_i \ln\left(\frac{\hat{\pi}_i}{\pi_i}\right) + \sum_k \lambda_k \left(\sum_i \hat{\pi}_i (e_k)_i - \hat{m}_k\right) + \mu\left(\sum_i \hat{\pi}_i - 1\right) \\
\frac{\partial S}{\partial \hat{\pi}_i} &= \sum_k \lambda_k (e_k)_i - \ln\left(\frac{\hat{\pi}_i}{\pi_i}\right) - 1 + \mu = 0 \\
\hat{\pi}_i &= \pi_i \exp\left(\sum_k \lambda_k (e_k)_i\right) \exp(\mu) \Rightarrow \exp(\mu) = \frac{1}{\sum_i \pi_i \exp\left(\sum_v \lambda_v (e_v)_i\right)} \quad \text{by normalization} \\
\hat{\pi}_i &= \frac{\pi_i \exp\left(\sum_v \lambda_v (e_v)_i\right)}{\sum_i \pi_i \exp\left(\sum_v \lambda_v (e_v)_i\right)} \qquad (1)
\end{aligned}
$$

where $\lambda_k$ has to be chosen such that the constraint on the measurable quantities is satisfied, and $\exp(\mu)$ was chosen as to ensure $\hat{\boldsymbol{\pi}}$ is normalized

### 1.2 Augmented likelihood functional

In the main text we present the maximum likelihood minimum Kullback-Leibler divergence functional where the well-known transition probability matrix likelihood has been augmented by a factor introducing the experimental measurements $o_k$ with experimental uncertainties $\sigma_k^2 = 1/2w_k$:

$$
L \propto \left(\prod_{i,j} p_{ij}^{c_{ij}}\right)\left(\prod_k \exp\left(-w_k(\hat{m}_k - o_k)^2\right)\right). \qquad (2)
$$

Without loss of generality we define the symmetric matrix $x_{ij} = \pi_i p_{ij}$. In the following we use the matrix elements $x_{ij}$ as variables, and express the transition probabilities $p_{ij}$ and equilibrium density $\pi_i$ in terms of $x_{ij}$: $p_{ij} = x_{ij}/\pi_i$ , and $\pi_i = \sum_j x_{ij}$.

At a first inspection of Eq. 2, the first factor appears to only depend on the simulation quantities, and the second factor depends on the experimental quantities. However, the factors are coupled through the minimum Kullback-Leibler expression (eq. 1). To make this more explicit we insert expression (1) into the augmented likelihood (3):

$$L \propto \left(\prod_{i,j}\left(\frac{x_{ij}}{\pi_i}\right)^{c_{ij}}\right)\left(\prod_k \exp\left(-w_k(\sum_i \hat{\pi}_i(e_k)_i - o_k)^2\right)\right) =$$
$$\left(\prod_{i,j}\left(\frac{x_{ij}}{\pi_i}\right)^{c_{ij}}\right)\left(\prod_k \exp\left(-w_k\left(\frac{\sum_{s=1}^n (e_k)_s \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)}{\sum_{s=1}^n \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)} - o_k\right)^2\right)\right). \quad (3)$$

If we maximize (3) w.r.t. $\pi$, $p_{ij}$ and $\lambda_k$ we obtain a Markov model integrating both simulation and experimental data: an Augmented Markov Model (AMM). Depending on our particular situation we propose two different algorithms to maximize eq. 3.

### 1.2.1 Estimation: the ideal case

If $o_k$ is within the support of $\boldsymbol{e}_k$ for all $K$ observations, the two factors in the likelihood functional (eq. 2) decouple. Therefore we can estimate $\boldsymbol{\pi}$ first using the reversible estimator of transition probability matrices [1, 2], and then estimate the Lagrange multipliers $\lambda_v$ such that $\hat{\boldsymbol{\pi}} \cdot \boldsymbol{e}_k = o_k$. Then we compute $\hat{\boldsymbol{\pi}}$ using (eq. 1) and compute the AMM by estimating a transition probability matrix with the stationary distribution constrained to be $\hat{\boldsymbol{\pi}}$ [3].

### 1.2.2 General fixed point iteration algorithm

In most practical cases $m_k$ will not be within the support of $\boldsymbol{e}_k$ for all $K$ observations. In this case we need to devise a fixed point iteration algorithm. We start with the log-likelihood:

$$LL = \sum_{i,j} c_{ij} \log x_{ij} - \sum_i c_i \log \pi_i - \sum_k w_k \left(\frac{\sum_{s=1}^n (e_k)_s \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)}{\sum_{s=1}^n \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)} - o_k\right)^2$$

where $c_i = \sum_j c_{ij}$. Both $x_{ij}$ and $\lambda_i$ are unknown, and we want to find these such that the likelihood is maximized.

Taking derivatives of $LL$ with respect to $x_{ij}$, by using the fact that $x_{ij} = x_{ji}$ and $\pi_i = \sum_j x_{ij}$ (that is: $\frac{\partial \pi_i}{\partial x_{kj}} = \delta_{ik} + \delta_{jk}$), we obtain:

$$\frac{\partial LL}{\partial x_{ij}} = \frac{c_{ij} + c_{ji}}{x_{ij}} - \frac{c_i}{\pi_i} - \frac{c_j}{\pi_j} - Q_{ij}(\lambda, \pi) \quad (4)$$

where

$$\begin{aligned} Q_{ij}(\lambda,\pi) &= \frac{\partial}{\partial x_{ij}} \sum_k w_k \left(\frac{\sum_{s=1}^n (e_k)_s \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)}{\sum_{s=1}^n \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)} - o_k\right)^2 = \\ &= 2\sum_k w_k S_k(\lambda,\pi) R_{k,ij}(\lambda,\pi) \end{aligned}$$

with

$$S_k(\lambda,\pi) = \frac{\sum_{s=1}^n (e_k)_s \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)}{\sum_{s=1}^n \pi_s \exp\left(\sum_{v=1}^K \lambda_v (e_v)_s\right)} - o_k = \hat{m}_k - o_k$$

and

$$\begin{aligned}
R_{k,ij}(\lambda,\pi) &= \frac{\partial}{\partial x_{ij}} S_k(\lambda,\pi) \\
&= \frac{\partial}{\partial x_{ij}} \frac{\sum_{s=1}^{n}(e_k)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)}{\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)} \\
&= \frac{(e_k)_i \pi_i \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_i\right) + (e_k)_j \pi_j \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_j\right)}{\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)} \\
&\quad - \frac{\left[\pi_i \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_i\right) + \pi_j \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_j\right)\right] \sum_{s=1}^{n}(e_k)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)}{\left(\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)\right)^2}.
\end{aligned}$$

The above expression is related to the fluctuations of the observables in the Markov states. It can be recast in more compact form in terms of the experimental state probabilities, $\hat{\pi}_i$, by using expression (1) and $\hat{m}_k$:

$$R_{k,ij}(\lambda,\pi) = (e_k)_i \hat{\pi}_i + (e_k)_j \hat{\pi}_j - (\hat{\pi}_i + \hat{\pi}_j)\,\hat{m}_k = \hat{\pi}_i\,((e_k)_i - \hat{m}_k) + \hat{\pi}_j\,((e_k)_j - \hat{m}_k)$$

and

$$Q_{ij}(\lambda,\pi) = 2\sum_k w_k S_k(\lambda,\pi) R_{k,ij}(\lambda,\pi) = 2\sum_{k=1}^{K} w_k\,(\hat{m}_k - o_k)\,[\hat{\pi}_i\,((e_k)_i - \hat{m}_k) + \hat{\pi}_j\,((e_k)_j - \hat{m}_k)]\,.$$

This expression is consistent with the intuition that if an observable is not expected to vary significantly across the different microstates of the system, its experimental measurement is not very useful to correct the statistics.

By taking the derivatives of $LL$ with respect to $\lambda_i$, we obtain:

$$\begin{aligned}
\frac{\partial LL}{\partial \lambda_i} &= -2\sum_k w_k \frac{\partial}{\partial \lambda_i}\left(\frac{\sum_{s=1}^{n}(e_k)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)}{\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)} - o_k\right)^2 \\
&= -2\sum_k w_k S_k(\lambda,\pi) G_{k,i}(\lambda,\pi)
\end{aligned} \tag{5}$$

with $S_k(\lambda,\pi)$ defined above and

$$\begin{aligned}
G_{k,i}(\lambda,\pi) &= \frac{\partial}{\partial \lambda_i} S_k(\lambda,\pi) = \\
&= \frac{\partial}{\partial \lambda_i} \frac{\sum_{s=1}^{n}(e_k)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)}{\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)} = \\
&= \frac{\sum_{s=1}^{n}(e_k)_s (e_i)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)}{\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)} + \\
&- \frac{\left(\sum_{s=1}^{n}(e_k)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)\right)\left(\sum_{s=1}^{n}(e_i)_s \pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)\right)}{\left(\sum_{s=1}^{n}\pi_s \exp\left(\sum_{v=1}^{K}\lambda_v (e_v)_s\right)\right)^2}.
\end{aligned}$$

The above expression also takes a more compact form in terms of the corrected densities, and it's related to the covariance of the observables over the Markov states:

$$G_{k,i}(\lambda,\pi) = \sum_{s=1}^{n}(e_k)_s (e_i)_s \hat{\pi}_s - \left(\sum_{s=1}^{n}(e_k)_s \hat{\pi}_s\right)\left(\sum_{s=1}^{n}(e_i)_s \hat{\pi}_s\right) = \hat{m}_{ki} - \hat{m}_k \hat{m}_i$$

where $\hat{m}_{ki} = \sum_{s=1}^{n} (e_k)_s (e_i)_s \hat{\pi}_s$.

Setting the derivatives $\frac{\partial LL}{\partial x_{ij}}$ equal to 0 and solving for $x_{ij}$ yields the equations:

$$\frac{c_{ij} + c_{ji}}{x_{ij}} = \frac{c_i}{\pi_i} + \frac{c_j}{\pi_j} + Q_{ij}(\lambda, \pi)$$

or

$$x_{ij} = \frac{c_{ij} + c_{ji}}{\frac{c_i}{\pi_i} + \frac{c_j}{\pi_j} + Q_{ij}(\lambda, \pi)}. \tag{6}$$

These equations are not easily solved directly, as the expression on the right hand side of (6) is also a non-trivial function of $x_{ij}$. However, they can be solved iteratively, as discussed below. From (6) we can construct the fixed point iteration for $\boldsymbol{\pi}$:

$$\pi_i = \sum_j x_{ij} = \sum_j \frac{c_{ij} + c_{ji}}{\frac{c_i}{\pi_i} + \frac{c_j}{\pi_j} + Q_{ij}(\lambda, \pi)} \tag{7}$$

Setting to $\frac{\partial LL}{\partial \lambda_i}$ to 0 and solving for $\lambda_i$ yields:

$$\sum_k w_k S_k(\lambda, \pi) G_{k,i}(\lambda, \pi) = 0 \tag{8}$$

which can be updated with a Newton step.

### 1.2.3 Estimation algorithm

The equations above lead to the following algorithm:

1. Initialize $\boldsymbol{\pi}$ using the reversible transition probability matrix estimator with $c_{ij}$ and set all $\lambda_i = 0$.
2. Iterate until convergence:
   (a) Update $\boldsymbol{\pi}$ for fixed $\{\lambda_i\}$ (thus, fixed $\hat{\boldsymbol{\pi}}$ and $\{\hat{m}_k\}$) using fixed-point step (Eq. 7).
   (b) Update $\{\lambda_i\}$ for fixed $\boldsymbol{\pi}$ using a Newton step (Eq. 8).
   (c) Compute updated $\hat{\boldsymbol{\pi}}$ and $\{\hat{m}_k\}$ for fixed $\boldsymbol{\pi}$ with updated $\{\lambda_i\}$ (Eq. 1 and $\hat{m}_k$).
3. Compute $\hat{\boldsymbol{P}} = \arg\max_{\boldsymbol{P}}(\boldsymbol{C} \mid \hat{\boldsymbol{\pi}}, \boldsymbol{P})$, i.e. maximum likelihood MSM for given stationary distribution, $\hat{\boldsymbol{\pi}}$.

### 1.2.4 Implementation details

Covergence of the Lagrange multiplier $\lambda_l$ is when the change $\left|\frac{\hat{m}_l^{(i+1)} - \hat{m}_l^{(i)}}{\sigma_l}\right| < 5\%$, where $\hat{m}_l^{(i)}$ is $\hat{m}_l$ in the $i$-th iteration and $\sigma_l$ is the experimental uncertainty. When the Lagrange multipliers have converged, $\boldsymbol{\pi}$ and $\hat{\boldsymbol{\pi}}$ are updated until the change in $LL$ is less than $10^{-3}$.

## 1.3 Special case of the augmented likelihood functional: state counts only

In some simulation cases, including generalized ensemble methods (meta-dynamics/local-elevation, replica exchange etc) and Markov chain Monte Carlo methods, we only have access to unbiased state counts, $c_i$. In these cases the first factor of the likelihood functional reduces to a categorial distribution:

$$L \propto \left(\prod_i \pi_i^{c_i}\right) \left(\prod_k \exp\left(-w_k(\hat{m}_k - o_k)^2\right)\right). \tag{9}$$

We can maximize this for $\boldsymbol{\pi}$ and $\lambda_v$ through the log-likelihood:

$$LL = \sum_i c_i \log \pi_i - \sum_k w_k \left(\frac{\sum_{s=1}^{n} (e_k)_s \pi_s \exp\left(\sum_{v=1}^{K} \lambda_v (e_v)_s\right)}{\sum_{s=1}^{n} \pi_s \exp\left(\sum_{v=1}^{K} \lambda_v (e_v)_s\right)} - o_k\right)^2, \tag{10}$$

we get

$$
\begin{aligned}
\frac{\partial LL}{\partial \pi_i} &= \frac{c_i}{\pi_i} - \sum_k w_k \left( \frac{(e_k)_i \pi_i \exp\left(\sum_{v=1}^{K} \lambda_v (e_v)_s\right)}{\sum_{s=1}^{n} \pi_s \exp\left(\sum_{v=1}^{K} \lambda_v (e_v)_s\right)} - o_k \right)^2 \\
&= \frac{c_i}{\pi_i} - 2 \sum_k w_k \hat{\pi}_i (\hat{m}_k - o_k)[(e_k)_i - \hat{m}_k] \\
\pi_i &= \frac{c_i}{2 \sum_k w_k (\hat{m}_k - o_k) \hat{\pi}_i [(e_k)_i - \hat{m}_k]}.
\end{aligned}
$$

Similarly, for the Lagrange multipliers we get,

$$
\begin{aligned}
\frac{\partial LL}{\partial \lambda_v} &= \sum_k w_k \left( \frac{(e_k)_i \pi_i \exp\left(\sum_{v=1}^{K} \lambda_v (e_v)_s\right)}{\sum_{s=1}^{n} \pi_s \exp\left(\sum_{v=1}^{K} \lambda_v (e_v)_s\right)} - o_k \right)^2 \\
0 &= 2 \sum_k w_k (\hat{m}_k - o_k).
\end{aligned}
$$

Iterating these two equations will until convergence will allow us to get an estimate of $\hat{\boldsymbol{\pi}}$. Alternatively, if unbiased samples of $\boldsymbol{\pi}$ can be recovered from the sampling approach, we may use an importance sampling scheme to generate samples from $\hat{\boldsymbol{\pi}}$ [4].

### 1.4 Bayesian error estimation in AMMs

Following convergence of the Lagrange multipliers, errors of AMM are estimated through Markov Chain Monte Carlo sampling. We assume a sparsity inducing prior on the transition probabilities and sample stationary distributions conditioned on the transition counts, $\mathbf{C}$ [5], the estimated Lagrange multipliers and experimental restraints. Effectively, we are generating samples from the following posterior distribution

$$
p(\hat{\boldsymbol{\pi}} \mid \mathbf{C}, \mathbf{o}, \boldsymbol{\sigma}) \propto \prod_k \mathcal{N}(o_k \mid \hat{m}_k, \sigma_k) p(\boldsymbol{\pi} \mid \mathbf{C}). \tag{11}
$$

Here, $\mathbf{o}$ is the vector of experimental data with associated uncertainties $\boldsymbol{\sigma}$, and $\mathcal{N}(\cdot)$ denotes the normal distribution. The posterior samples are then used to compute an ensemble of transition probability matrices as above.

### 1.5 Correlated experimental errors

If knowledge of the correlation structure of experimental observables is available, these can be taken into account by replacing the assumption of independent normal data in the experimental factor of the likelihood, with one allowing for expression of the full covariance structure $\boldsymbol{\Sigma}$: a multi-variate Normal. In this case the likelihood becomes

$$
L \propto \left( \prod_{i,j} p_{ij}^{c_{ij}} \right) \exp(-\frac{1}{2} (\hat{\mathbf{m}} - \mathbf{o})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\hat{\mathbf{m}} - \mathbf{o})). \tag{12}
$$

subject to the same maximum entropy coupling functional (eq. 1). Further, $\hat{\mathbf{m}}$ and $\mathbf{o}$ are vectors of predicted and experimental experimental observables and $\boldsymbol{\Sigma}$ is the full experimental covariance matrix.

## 2 SI Materials and Methods

### 2.1 Protein folding model system simulation and analysis details

200 trajectories of 10000 steps were generated for each model and used for the analysis shown in the main text. Mean first passage times of folding and unfolding were computed as entering the fully folded state starting from the fully unfolded state, and vice versa, as previously described [6]. The AMMs were estimated using the average helicity in the 'true' model, 59.788 with an uncertainty of 0.299, as a restraint. The model is based upon one from Ref. [7].

## 2.2 Augmented Markov model and Markov state model estimation details of Ubiquitin

All pairwise C$\alpha$-C$\alpha$ distances and $\phi$,$\psi$ backbone dihedral angles - the molecular features - were extracted from two 1 ms trajectories of Ubiquitin strided into 5 nanosecond steps. The simulations were carried out as reported previously in the CHARMM22* [8, 9] and CHARMM-*h [10, 11]* force fields. These molecular features embedded into the slowest 6 time-lagged independent components (lag-time: 0.75 $\mu s$) [12] and this space was then segmented into 256 cluster using k-means clustering. Count matrices were generated for each of the 1 ms trajectories independently (lag-time: 100 ns) which were used to estimate the two Markov state models, implied time-scale and self-consistency checks were carried out (Fig. S1). The augmented Markov models were estimated by maximizing the likelihood given in equation 3, using the same count matrices used for estimation the corresponding MSMs [13]. In addition to the count matrix, experimental residual dipolar couplings (RDC) and $\mathrm{N^H - H^\alpha}$ $^3$J couplings couplings were used as input. The inclusion of an experimental observable, $o_k$, involves computation of the experimental observable in each Markov state, $\mathbf{e_k}$. For $^3$J couplings couplings the Karplus relation was used with previously published Karplus parameters [14]. For the RDCs the frames of the MD trajectories were superimposed to the frame which minimizes the sum of C$\alpha$ RMSD to all other frames, and one alignment tensor was estimated for each of the 36 alignment conditions as previously described [15]. Each of the alignment tensors were then used to compute Markov state averaged RDCs. Finally, we set uncertainties, $\sigma_k$, uniformly to 0.5 Hz to accommodate for any potential noise in the experiment as well as any random errors associated with prediction of Markov state observables. Estimated AMMs were robust with respect to variation of error parameters between 0.2 and 2.0 Hz. Bayesian error estimates, are reported as confidence intervals of 50 posterior samples. For AMMs the samples were generated using Eq. (11) and for MSMs they were generated as previously described [5].

## 2.3 Back-calculation of NMR observables

As the lag-time of the Markov models exceeds the rotational correlation time of Ubiquitin at the temperatures considered here, we may accurately approximate the exact nuclear Overhauser enhancements between a spin-pair $i$ by $e\mathrm{NOE_i} \propto -\hat{\boldsymbol{\pi}} \cdot \mathbf{r_i^{-6}}$ [16]. Here, $\mathbf{r_i^{-6}}$ is the vector of the inter-atomic distance to the minus sixth power between the atoms involved in the given eNOE. Expressions for cases where this condition is not the fulfilled are given elsewhere [17]. Cross-correlated relaxation rates between the inter-atomic vectors A-B and C-D were computed as CCR $\propto P_2(\cos\theta)$ where $P_2(\cdot)$ is the second Lagrange polynomial and $\theta$ is the angle between the A-B and C-D vectors. This expression assumes isotropic rotational diffusion. Expressions for anisotropic rotational diffusion are given elsewhere [18]. $R_{1\rho}$ relaxation dispersion data were computed as recently described [19] using the $^1\mathrm{H^N}$ chemical shifts computed by CAMSHIFT [20, 21].

## 2.4 Comparisons of Markov state models and augmented Markov models

To compare two models 1 and 2, either MSMs and AMMs, a set of microstates visited in both models, $\mathcal{K}$, is identified, the stationary distributions of each of the models were renormalized on this set yielding the distributions $\rho_1$ and $\rho_2$ defined on the common set. Using the Kullback-Leibler divergence

$$\mathcal{L}(\rho_1, \rho_2) = \sum_{i \in \mathcal{K}} \rho_1^i \ln \frac{\rho_1^i}{\rho_2^i}$$

we can compute the symmetric measure, the Jensen-Shannon divergence (JSD) as

$$\mathcal{H}(\rho_1, \rho_2) = \frac{1}{2}\left(\mathcal{L}(\rho_1, \varrho) + \mathcal{L}(\rho_2, \varrho)\right), \quad \varrho^i = \frac{1}{2}(\rho_1^i + \rho_2^i),$$

which measures how 'similar' two distributions are to each other. The JSD is dominated by differences high-density areas an will not be sensitive to subtle changes sparsely populated states. Considering the uncertainties in the inferred models this is a desirable property as it leads to a more robust measure. Confidence intervals were computed by using the JSD for pairs of independently sampled models; for AMMs the scheme described above was used to generate samples, for MSMs a previously described Bayesian scheme was used [5].

| | *Unfolded* | $a$ | $b$ | $c$ | $a/b$ | $a/c$ | $b/c$ | $a/b/c$ |
|---|---|---|---|---|---|---|---|---|
| $U(\mathcal{S})$, True | 0.0 | -1084.29425612 | -1332.24398379 | -1648.54618261 | -1766.78439442 | -2094.3271071 | -2365.00620538 | -2873.88298642 |
| $U(\mathcal{S})$, $\Delta\Delta G_{\text{fold}} = -0.095\,kT$ | 0.0 | -1084.32136348 | -1332.27728989 | -1648.58739626 | -1766.82856403 | -2094.37946528 | -2365.06533054 | -2873.95483349 |
| $U(\mathcal{S})$, $\Delta\Delta G_{\text{fold}} = -0.475\,kT$ | 0.0 | -1084.4297929 | -1332.41051429 | -1648.75225088 | -1767.00524247 | -2094.58889799 | -2365.30183116 | -2874.24222179 |
| $U(\mathcal{S})$, $\Delta\Delta G_{\text{fold}} = -1.03\,kT$ | 0.0 | -1084.56532968 | -1332.57704479 | -1648.95831916 | -1767.22609052 | -2094.85068888 | -2365.59745693 | -2874.60145717 |
| $U(\mathcal{S})$, $\Delta\Delta G_{\text{fold}} = -2.04\,kT$ | 0.0 | -1084.83640325 | -1332.91010578 | -1649.3704557 | -1767.66778662 | -2095.37427065 | -2366.18870848 | -2875.31992791 |
| $U(\mathcal{S})$, $\Delta\Delta G_{\text{fold}} = -3.04\,kT$ | 0.0 | -1085.10747681 | -1333.24316678 | -1649.78259225 | -1768.10948272 | -2095.89785243 | -2366.77996003 | -2876.03839866 |
| $U(\mathcal{S})$, $\Delta\Delta G_{\text{fold}} = -4.07\,kT$ | 0.0 | -1085.37855038 | -1333.57622777 | -1650.19472879 | -1768.55117881 | -2096.42143421 | -2367.37121159 | -2876.75686941 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $S(\mathcal{S})$ | 10.3804 | 6.76878 | 5.9408 | 4.8846 | 4.5025 | 3.409 | 2.50553 | 0.8109 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| helicity | 0. / 2.5 | 11. / 5 | 22. / 5 | 33. / 2.5 | 55. / 5 | 66 / 5 | 77. / 5 | 88. / 2.5 |
| quench | 5. / 0.8 | 5. / 1.6 | 5. / 1.6 | 15. / 0.8 | 5. / 0.64 | 10. / 0.64 | 10. / 0.64 | 8.3 / 1.6 |

Table S1: Table of parameters of protein folding models. The first seven rows are enthalpic contributions, the eighth row contains the entropic contribution in each of the states (same for all models). The lower two rows are the observables of each of the states. All entalphies are in $\text{kcal}\,\text{mol}^{-1}$ and entropies are in $\text{kcal}\,\text{mol}^{-1}\,\text{K}^{-1}$. For the observables, helicity and quenching, mean / standard deviation of the observables in each state is given.

## 2.5 Protein folding model system parameters

## 2.6 Bayes factor model selection using $R_{1,\rho}$ relaxation dispersion data

We use the Bayes factor [22] to evaluate whether $R_{1,\rho}$ data predicted by the AMMs, $\mathbf{a}$, is better at describing the experimental data $\mathbf{r}$ considering the experimental uncertainty $\boldsymbol{\sigma}$ when compared to the MSM, $\mathbf{m}$. We compute these using a numerical approximation of the integral of the likelihood of observing the predicted $R_{1,\rho}$ data under a normal error model, given the data. We obtain the expression

$$\text{Bayes factor} \approx \log \sum_i \mathcal{N}(\mathbf{a}^i \mid \mathbf{r}, \boldsymbol{\sigma}) - \log \sum_i \mathcal{N}(\mathbf{m}^i \mid \mathbf{r}, \boldsymbol{\sigma}) \tag{13}$$

where the $i$ is the index of the sampled models. AMMs are sampled according to (eq. 11) and MSMs are sampled as previously described [5]. We here use 50 samples for each model type. The predicted data for the AMMs and MSMs is scaled and an intercept is added according to a previously described scheme [19], prior to the calculation of the Bayes factors.

# 3 Supporting data

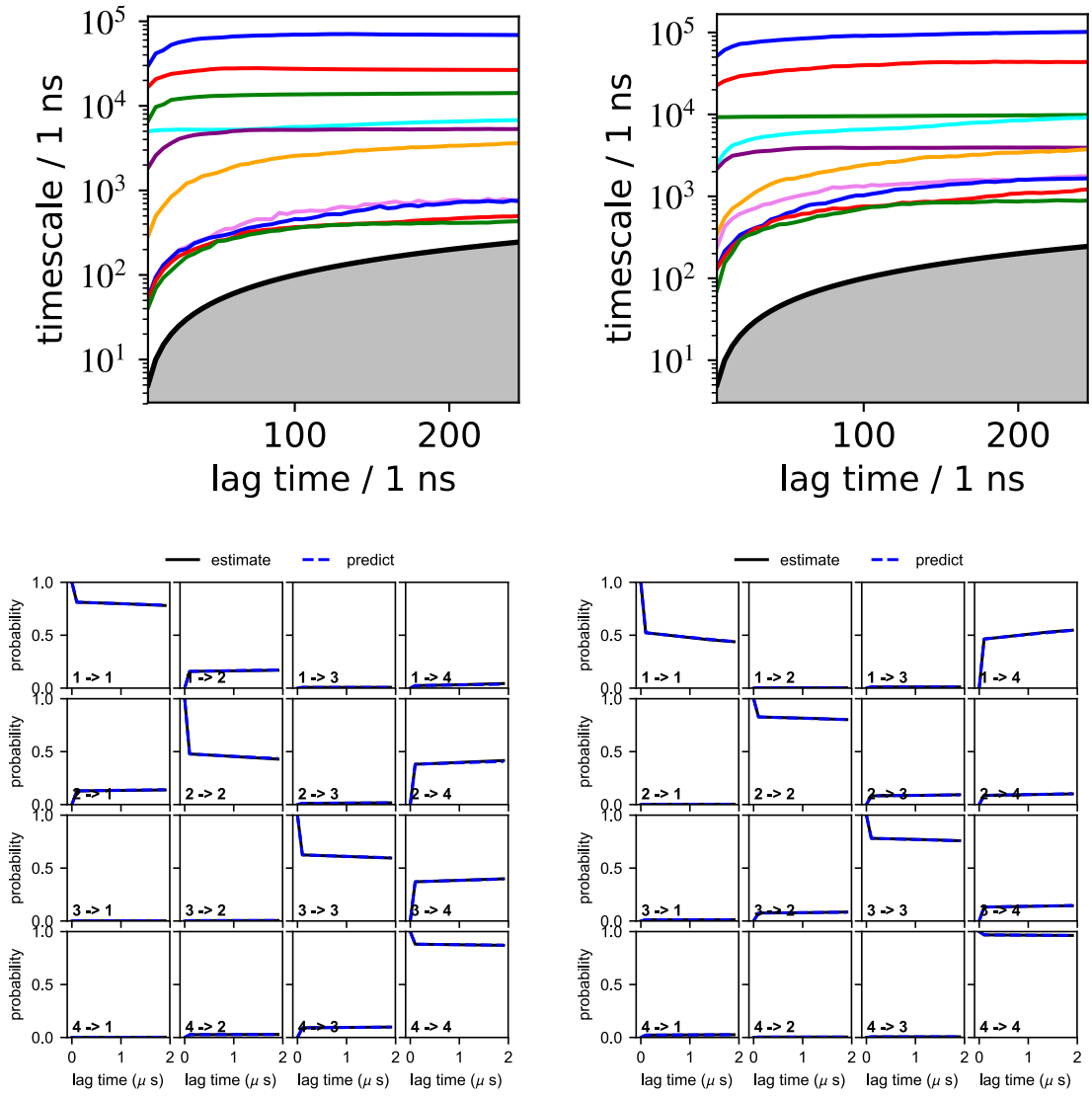


Figure S1: Upper: Implied time-scale as a function of MSM lag-time for CHARMM22* (left) and CHARMM-$h$ (right). Each colored line corresponds to the correlation timescale, computed as $\tau_i = -\frac{\tau_{\mathrm{MSM}}}{\log|\lambda_i|}$ where $\lambda_i$ is the $i$'th largest Eigenvalue of the transition probability matrix determined using a lag-time of $\tau_{\mathrm{MSM}}$. Lower: Chapman-Kolmogorov test for CHARMM22* (left) and CHARMM-$h$ (right). Each of the panels show the (estimated (from MD data) and predicted (using a MSM with lagtime: 100 ns) transition probability between 4 meta-stable sets automatically determined using PCCA.

| | C22* MSM | C22* AMM | C-$h$ MSM | C-$h$ AMM |
|---|---|---|---|---|
| RDCs ($Q$) * | $0.318^{0.319}_{0.316}$ | $0.294^{0.294}_{0.294}$ | $0.282^{0.283}_{0.281}$ | $0.279^{0.279}_{0.279}$ |
| $^3J$ (Hz) * | $0.999^{0.999}_{0.999}$ | $0.987^{0.987}_{0.987}$ | $0.942^{0.942}_{0.941}$ | $0.937^{0.938}_{0.936}$ |
| eNOEs ($\rho$) | $0.907^{0.908}_{0.907}$ | $0.908^{0.909}_{0.908}$ | $0.911^{0.912}_{0.911}$ | $0.913^{0.913}_{0.912}$ |
| CCRs ($\rho$) | $0.42^{0.42}_{0.41}$ | $0.41^{0.41}_{0.41}$ | $0.46^{0.47}_{0.45}$ | $0.48^{0.49}_{0.48}$ |
| $\tau_1$ ($\mu s$) | $70.04^{78.04}_{59.59}$ | $12.83^{14.}_{11.88}$ | $106.27^{135.98}_{80.22}$ | $22.32^{24.77}_{16.17}$ |
| $\tau_2$ ($\mu s$) | $27.42^{30.9}_{24.61}$ | $11.38^{12.29}_{10.16}$ | $43.31^{60.75}_{35.20}$ | $17.85^{18.94}_{14.61}$ |
| $\tau_3$ ($\mu s$) | $13^{15.77}_{10.24}$ | $9.12^{9.86}_{7.89}$ | $10.72^{13.57}_{8.33}$ | $13.92^{14.98}_{10.13}$ |
| $\tau_4$ ($\mu s$) | $8.81^{6.82}_{5.11}$ | $3.44^{3.81}_{2.95}$ | $7.01^{8.75}_{5.86}$ | $8.38^{9.28}_{7.22}$ |

Table S2: Comparison of AMMs and MSMs of ubiquitin built using simulation data from the CHARMM22* (C22*) and CHARMM-$h$ (C-$h$) force fields. Agreement with residual dipolar coupling data is computed as the average Q-factor ($Q = \frac{rms(D_{\mathrm{exp}} - D_{\mathrm{calc}})}{rms(D_{\mathrm{exp}})}$) [23] of experimental data from 36 alignment conditions. Individual Q-factors for all the data sets are shown in Figs. S2-S3. Agreement with eNOE data (backbone-backbone amide proton eNOE at 307K [24]), CCR data [25] ($R_{\mathrm{HN,C\alpha H\alpha}}$, 308K) was computed using Pearsons correlation coefficient. The four slowest correlation timescales, are computed as $\tau_i = -\frac{\tau_{\mathrm{MSM}}}{\log|\lambda_i|}$ where $\lambda_i$ is the $i$-th largest Eigenvalue of the transition probability matrix determined using a lag-time of $\tau_{\mathrm{MSM}} = 100\,\mathrm{ns}$. Upper and lower bounds of a 95% confidence interval are shown in super- and sub-scripts, respectively.

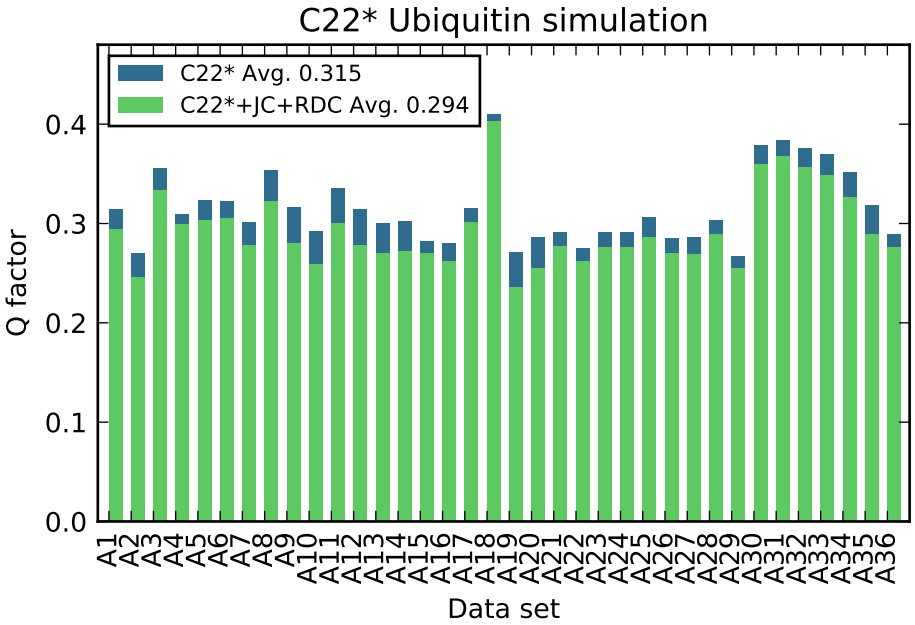


Figure S2: Agreement of MSMs(C22*) and AMMs (C22*+RDC+JC) built with the CHARMM-22* force field with individual residual dipolar coupling (RDC) datasets. The AMMs include all RDC data plus a $N^H - H^\alpha$ $^3$J-coupling dataset from Maltsev *et al.*[26]. Dataset names are as in *Lange et al.* [27].

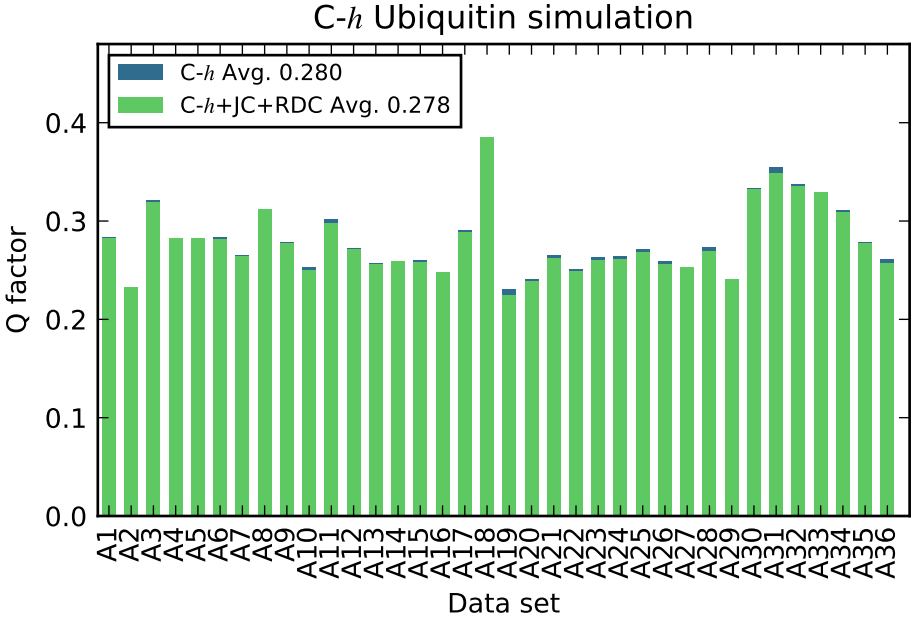


Figure S3: Agreement of MSMs(C-$h$) and AMMs (C-$h$+RDC+JC) built with the CHARMM-$h$ force field and individual residual dipolar coupling (RDC) datasets. The AMMs include all RDC data plus a $N^H - H^\alpha$ $^3$J-coupling dataset from Maltsev *et al.*[26]. Dataset names are as in Lange *et al.* [27].

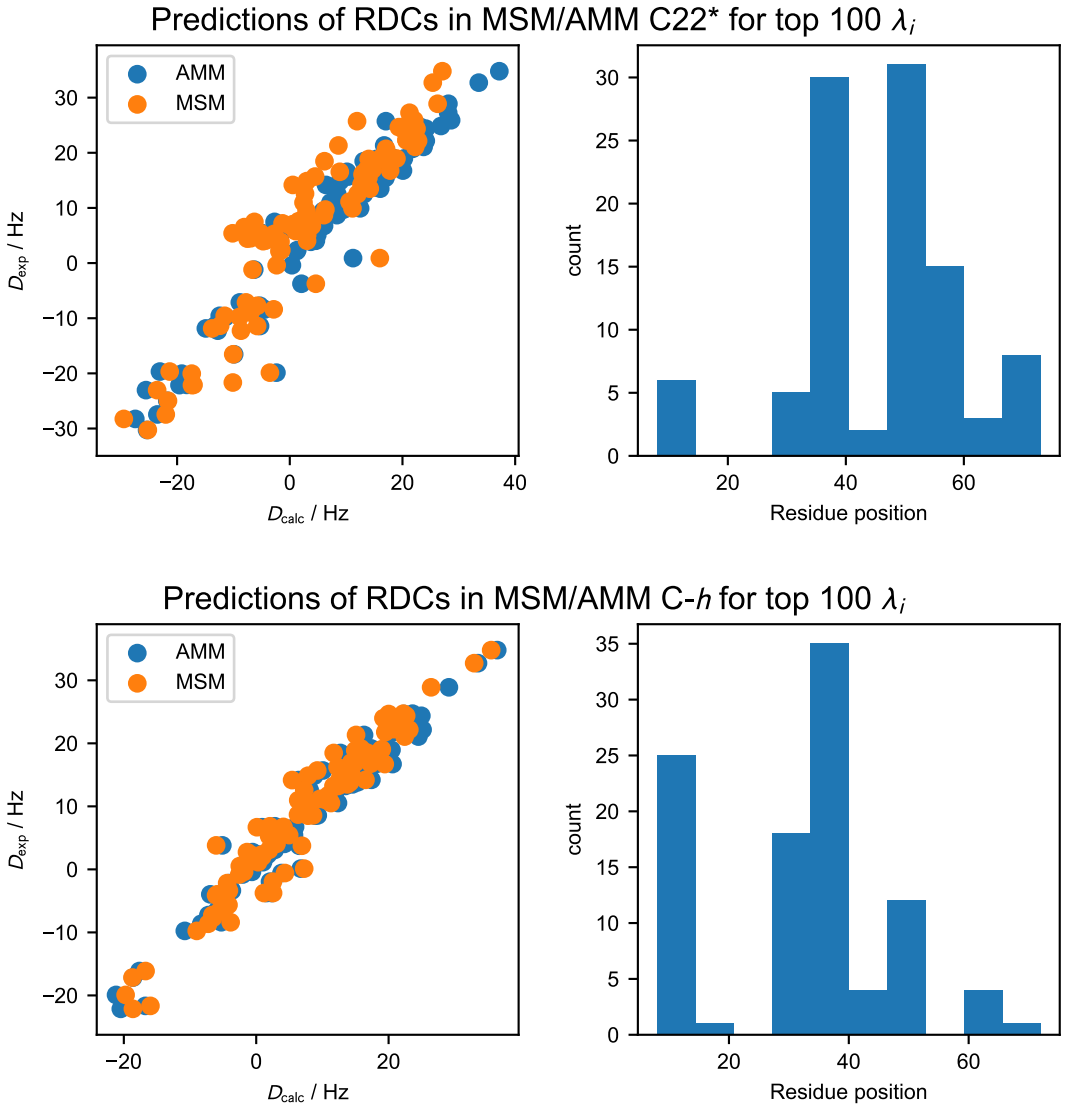


Figure S4: Right: Predictions of residual dipolar couplings associated with the 100 largest Lagrange multipliers. Left: the sequential distribution of the residual dipolar couplings with the 100 largest Lagrange multipliers. Top for C22*, bottom for C-$h$.

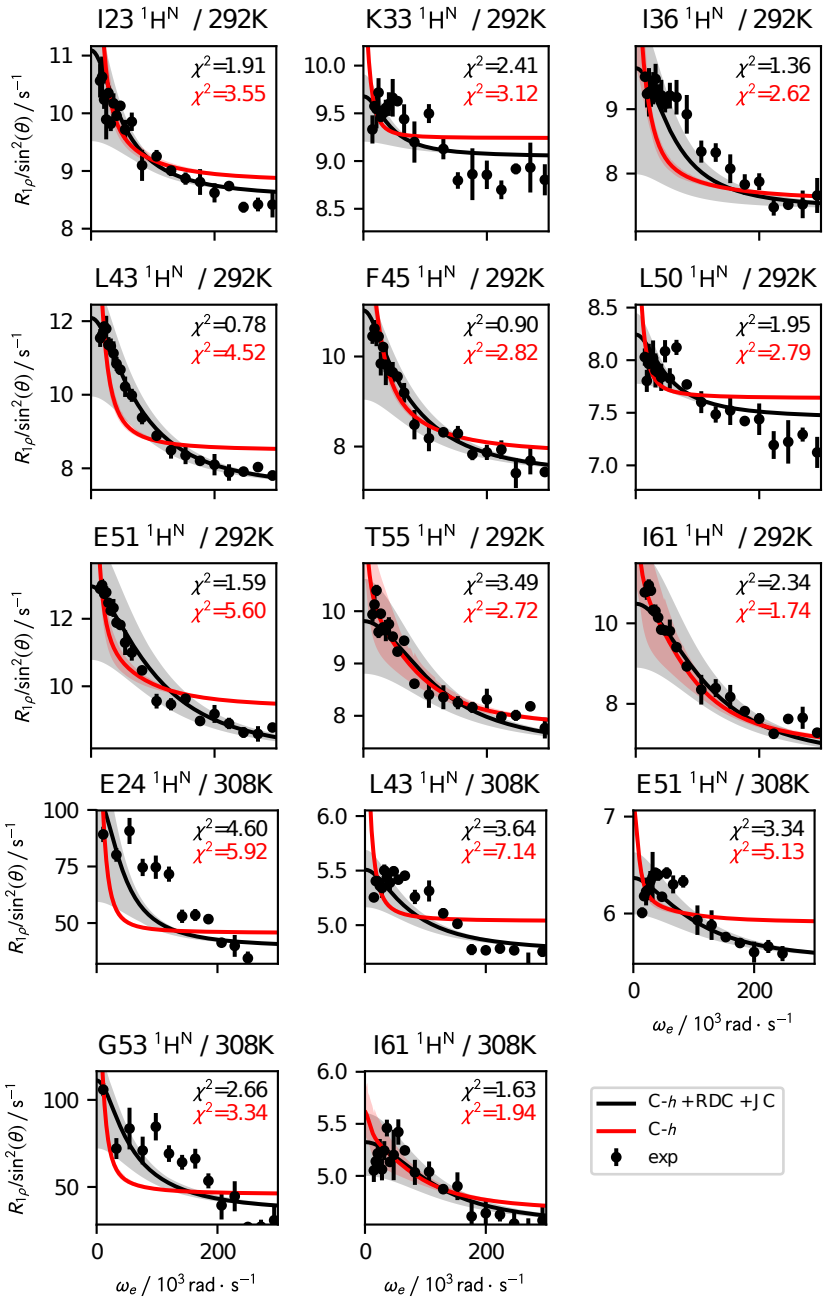


Figure S5: Comparison of experimental $R_{1\rho}$ data at 292K and 308K with MSMs(C-$h$) and AMMs (C-$h$+RDC+JC) based on simulation data from CHARMM-$h$ at 300K. The AMMs also use RDC and $^3$J-coupling data as described in the main text. Shaded area around the solid lines represents a 95% confidence interval. $\chi^2$ values are computed to the sample mean (solid lines).

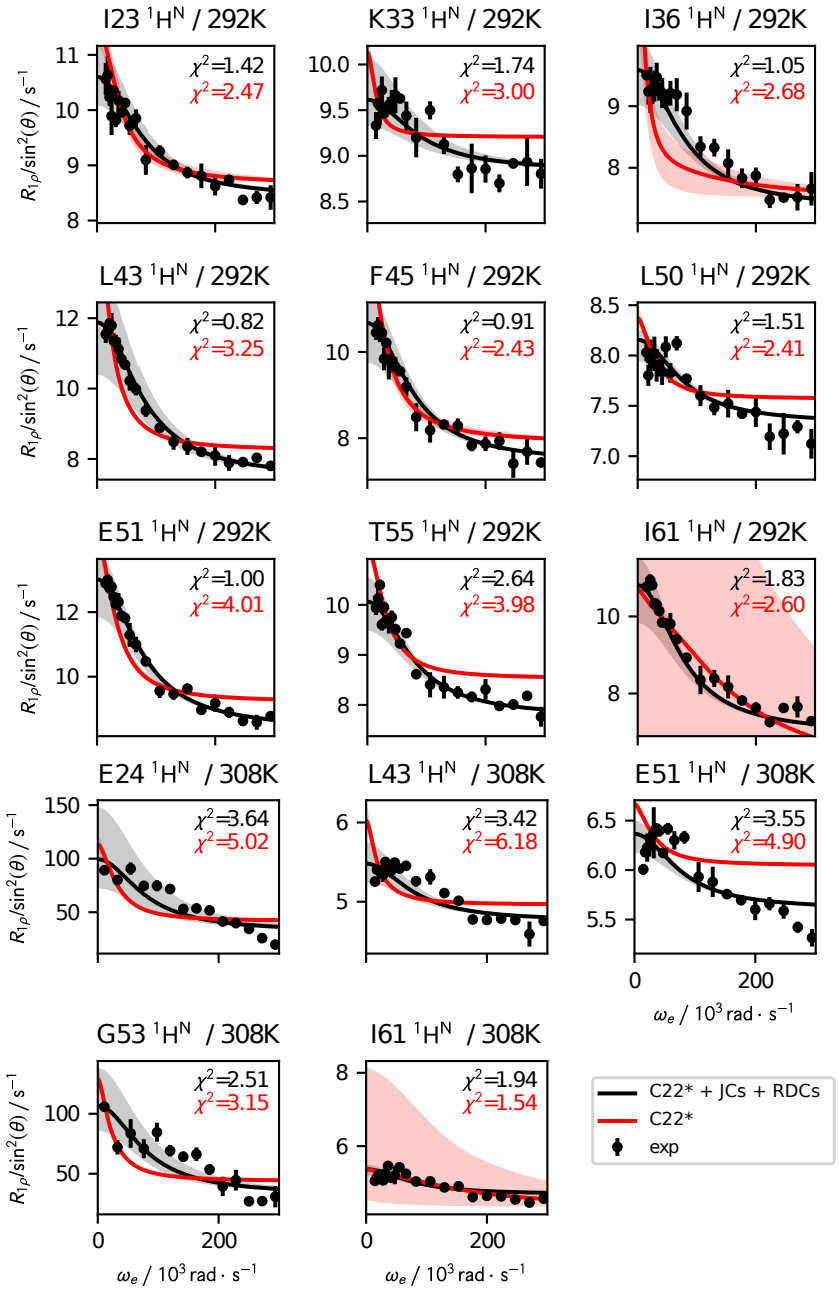


Figure S6: Comparison of experimental $R_{1\rho}$ data at 292K and 308K with MSMs(C22*) and AMMs (C22*+RDC+JC) based on simulation data from CHARMM22* at 300K. The AMMs also use RDC and $^3$J-coupling data as described in the main text. Shaded area around the solid lines represents a 95% confidence interval. $\chi^2$ values are computed to the sample mean (solid lines).

# References


[1] Prinz JH et al. (2011) Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* 134(17):174105.

[2] Bowman GR, Beauchamp KA, Boxer G, Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* 131:124101.

[3] Trendelkamp-Schroer B, Noé F (2013) Efficient bayesian estimation of markov model transition matrices with given stationary distribution. *J. Chem. Phys.* 138(16):164113.

[4] Olsson S et al. (2014) Probabilistic determination of native state ensembles of proteins. *J. Chem. Theory Comput.* 10(8):3484–3491.

[5] Trendelkamp-Schroer B, Wu H, Paul F, Noé F (2015) Estimation and uncertainty of reversible markov models. *J. Chem. Phys.* 143(17):174101.

[6] Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* 106(45):19011–19016.

[7] Prinz JH, Keller B, Noé F (2011) Probing molecular kinetics with markov models: metastable states, transition pathways and spectroscopic observables. *Phys. Chem. Chem. Phys.* 13(38):16912.

[8] Piana S, Lindorff-Larsen K, Shaw DE (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal* 100(9):L47–L49.

[9] Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* 110(15):5915–5920.

[10] Sborgi L et al. (2015) Interaction networks in protein folding via atomic-resolution experiments and long-time-scale molecular dynamics simulations. *J. Am. Chem. Soc.* 137(20):6506–6516.

[11] Lindorff-Larsen K, Maragakis P, Piana S, Shaw DE (2016) Picosecond to millisecond structural dynamics in human ubiquitin. *J. Phys. Chem. B* 120(33):8313–8320.

[12] Pérez-Hernández G, Paul F, Giorgino T, Fabritiis GD, Noé F (2013) Identification of slow molecular order parameters for markov model construction. *J. Chem. Phys.* 139(1):015102.

[13] Scherer MK et al. (2015) PyEMMA 2: A software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* 11(11):5525–5542.

[14] Vögeli B, Ying J, Grishaev A, Bax A (2007) Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. *J. Am. Chem. Soc.* 129(30):9377–9385.

[15] Showalter SA, Brüschweiler R (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J. Am. Chem. Soc.* 129(14):4158–4159.

[16] Tropp J (1980) Dipolar relaxation and nuclear overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. *J. Chem. Phys.* 72(11):6035–6043.

[17] Brüschweiler R, Case DA (1994) Characterization of biomolecular structure and dynamics by NMR cross relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.* 26:27–58.

[18] Vögeli B (2010) Comprehensive description of NMR cross-correlated relaxation under anisotropic molecular tumbling and correlated local dynamics on all time scales. *J. Chem. Phys.* 133(1):014501.

[19] Olsson S, Noé F (2017) Mechanistic models of chemical exchange induced relaxation in protein NMR. *J. Am. Chem. Soc.* 139(1):200–210.

[20] Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate predictions of protein nmr chemical shifts from interatomic distances. *J. Am. Chem. Soc.* 131(39):13894–5.

[21] Fu B et al. (2014) Almost: An all atom molecular simulation toolkit for protein structure determination. *J. Comput. Chem.* 35(14):1101–1105.

[22] Kass RE, Raftery AE (1995) Bayes factors. *JASA* 90(430):773–795.

[23] Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* 120(27):6836–6837.

[24] Leitz D, Vögeli B, Greenwald J, Riek R (2011) Temperature dependence of $^1H^N$ – $^1H^N$ distances in ubiquitin as studied by exact measurements of NOEs. *J. Phys. Chem. B* 115(23):7648–7660.

[25] Fenwick RB et al. (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J. Am. Chem. Soc.* 133(27):10336–10339.

[26] Maltsev AS, Grishaev A, Roche J, Zasloff M, Bax A (2014) Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based liquid crystalline phase. *J. Am. Chem. Soc.* 136(10):3752–3755.

[27] Lange OF et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320(5882):1471–1475.