JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Molecular Dynamics of Biomolecules through Direct Analysis of Dipolar Couplings

Simon Olsson,[†,‡] Dariusz Ekonomiuk,[†] Jacopo Sgrignani,[†] and Andrea Cavalli*[,†,§]

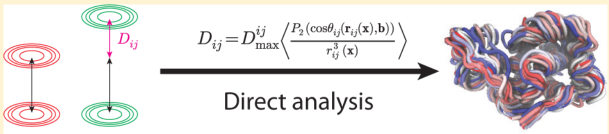
[†]Institute for Research in Biomedicine, Via Vincenzo Vela 6, CH-6500 Bellinzona, Switzerland

[‡]Laboratory of Physical Chemistry, Eidgenössische Technische Hochschule Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

[§]Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW United Kingdom

Supporting Information

ABSTRACT: Residual dipolar couplings (RDCs) are important probes in structural biology, but their analysis is often complicated by the determination of an alignment tensor or its associated assumptions. We here apply the maximum entropy principle to derive a tensor-free formalism which allows for direct, dynamic analysis of RDCs and holds the classic tensor formalism as a special case. Specifically, the framework enables us to robustly analyze data regardless of whether a clear separation of internal and overall dynamics is possible. Such a separation is often difficult in the core subjects of current structural biology, which include multidomain and intrinsically disordered proteins as well as nucleic acids. We demonstrate the method is tractable and self-consistent and generalizes to data sets comprised of observations from multiple different alignment conditions.


D_{ij}
$D_{ij} = D^{ij}_{\max}\left\langle \frac{P_2(\cos\theta_{ij}(\mathbf{r}_{ij}(\mathbf{x}),\mathbf{b}))}{r^3_{ij}(\mathbf{x})} \right\rangle$
Direct analysis

INTRODUCTION

The function of biomolecules is dictated by their ability to change shape over the course of time, that is, their dynamics.[1] This has been observed experimentally for a number of fundamental processes in biology including molecular recognition.[2] In recent years, there have been many significant advances in our understanding of biomolecular dynamics as well as important methodological contributions.[3−8] Still, reports of experimental characterization of biomolecular dynamics at the atomic level remain a rare feat in structural biology.[9,10] Typically, biomolecular dynamics span several spatial and temporal orders of magnitude. Nuclear magnetic resonance spectroscopy (NMR) uniquely provides a wealth of complementary and exquisitely detailed molecular probes which collectively cover most of the time scales relevant to biomolecular dynamics.[11] Residual dipolar couplings (RDCs) constitute one of these, broadly applicable to study structure and dynamics in biological macromolecules.[2,7,12−19] RDCs are measurable given an effective average orientation, or alignment, with respect to an external magnetic field (Figure 1, center).[20] Alignment may be achieved by dissolution in a nematic phase solvent[21] or by strong inherent anisotropic magnetic susceptibility.[22] Often it is possible to acquire multiple sets of RDCs for the same system by changing the experimental conditions, yielding different alignments[21−24] and thus complementary experimental measurements.

Analysis of RDC data in terms of dynamic models is often hampered by the assumption that the studied systems are rigid bodies. This assumption has its origin in the formalism pioneered by Saupe.[25] At its core, the formalism has a tensor, **S**, which describes the degree of order, or alignment, of a molecular frame with respect to an external magnetic field (Figure 1, right). Interestingly, many groups have found the tensor framework to approximate well cases where the aligning body is not strictly rigid;[14,26] this includes multidomain proteins with flexible linkers[27] and disordered proteins.[18,28] However, a number of significant difficulties arise when applied in such cases. First, additional assumptions are often necessary. For globular systems it is assumed that it is possible to separate internal and overall motion.[20] For flexible systems, it is assumed that the system is composed of multiple independently aligning segments,[29] and/or it consists of a mixture of states with different alignment properties.[18,30] Apart from assessing the validity of these assumptions, it is also difficult to gauge exactly when these break down. The second challenge involves the determination or prediction of physically meaningful alignment tensors, which is not always possible,[31] in particular, for highly flexible systems. Consequently, it will be a distinct advantage to mitigate these assumptions and technical difficulties.

With the increasing interest in dynamics in structural biology, the field has witnessed a resurgence of the maximum entropy principle (MEP) the last years.[3−7] The application of MEP to analyze NMR data has a long history in chemical physics, which includes biasing simulations of small organic compounds.[32−38] In structural biology, this principle allows us to derive probability distributions of biomolecular structure from experimental observations of average quantities. This is tremendously useful as it enables us to reconstruct the complex free energy landscape of biomolecules by simply observing their average behavior along a number of experimental coordinates. While MEP like all modeling approaches does not guarantee a *true* solution, it provides the least biased solution given the available information.[39] Consequently, MEP has already seen a

Received: February 5, 2015
Published: April 21, 2015

DOI: 10.1021/jacs.5b01289

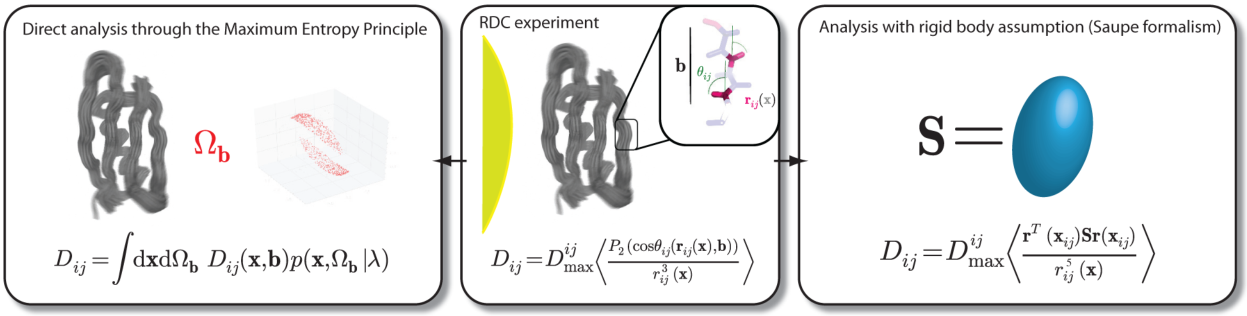

Direct analysis through the Maximum Entropy Principle
$\Omega_\mathbf{b}$
$D_{ij} = \int \mathrm{d}\mathbf{x}\mathrm{d}\Omega_\mathbf{b}\ D_{ij}(\mathbf{x},\mathbf{b})p(\mathbf{x},\Omega_\mathbf{b}|\lambda)$
RDC experiment
b
θ_{ij}
$\mathbf{r}_{ij}(\mathrm{x})$
$D_{ij} = D^{ij}_{\max}\left\langle \frac{P_2(\cos\theta_{ij}(\mathbf{r}_{ij}(\mathbf{x}),\mathbf{b}))}{r^3_{ij}(\mathbf{x})} \right\rangle$
Analysis with rigid body assumption (Saupe formalism)
S =
$D_{ij} = D^{ij}_{\max}\left\langle \frac{\mathbf{r}^T(\mathbf{x}_{ij})\mathbf{S}\mathbf{r}(\mathbf{x}_{ij})}{r^5_{ij}(\mathbf{x})} \right\rangle$

Figure 1. Illustration of RDC data from three different perspectives, under aligning conditions (yellow). Left: A joint distribution $p(\mathbf{x}, \Omega_\mathbf{b})$ of structure, $\mathbf{x}$ (gray), and magnetic field orientation, $\Omega_\mathbf{b}$ (red), is constructed via the MEP to model the experimental RDC signal. Center: The experimental RDC data are described well by the secular part of the heteronuclear dipolar interaction Hamiltonian, D_{ij}, between two nuclear spins i and j at strong magnetic fields.[20] Here, $\theta_{ij}(\mathbf{x},\mathbf{b})$ (green) is the angle between the unit interspin unit vector $\mathbf{r}_{ij}(\mathbf{x})$ (magenta), in the molecule $\mathbf{x}$ (gray), and an external magnetic field, $\mathbf{b}$ (black). The interspin distance is denoted $r_{ij}(\mathbf{x})$, while $P_2(x) = (3/2)x^2 - 1/2$ is the second Lagrange polynomial and $D^{ij}_{\max} = -\mu_0\gamma_i\gamma_j\hbar/8\pi^3$ with gyromagnetic ratios γ_i and γ_j of the two nuclei, μ_0 the permeability of vacuum, and Plancks constant, $\hbar$. Right: The Saupe formalism models the experimental RDC data through a rigid-body assumption where a real, traceless and symmetric second rank order tensor $\mathbf{S}$ (navy) relates the alignment of the internal coordinate frame of a molecule to an external magnetic field.

wealth of powerful applications to RDC data of oligosaccharides[15,40,41] as well as folded[7,31] and disordered proteins.[28] Unfortunately these studies either work within the Saupe formalism employing either instantaneous or average alignment tensors or do not accommodate the use of data from multiple alignment conditions.

Here, we demonstrate that it is possible to analyze RDCs directly via an alternative approach which avoids the use of alignment tensors. Through the MEP we derive a joint probability distribution of structure and magnetic field orientation (Figure 1, left). This takes the recently proposed tensor-free 'theta method' from structural biology[31] and previous efforts from chemical physics[34] a step further and allows for integration of data from multiple different alignment conditions. We find our theoretical result is a general framework which facilitates the determination of native state dynamics of globular proteins and interdomain motions in multidomain proteins with flexible linkers without the necessity to deconvolute internal from overall dynamics. The presented approach thus enables the structural biology community to rigorously analyze RDCs in terms of dynamic structural models in a unified and straightforward fashion.

■ THEORY

A Maximum Entropy Approach to the Analysis of RDC Data. In the current context, the aim of modeling using the MEP is to define a normalized probability density function which is the least biased with respect to the Boltzmann distribution of the force field, $E(\mathbf{x})$, and agrees with observed experimental data.[3−7,42] For RDCs, we can construct such a distribution by applying the MEP to the instantaneous expression of the dipolar coupling using vector notation (Figure 1):

$$D_{ij}(\mathbf{r}_{ij}(\mathbf{x}), \mathbf{b}, s) = \frac{sD^{ij}_{\max}}{r^3_{ij}(\mathbf{x})}\left(\frac{3}{2}(\mathbf{b}\cdot\mathbf{r}_{ij}(\mathbf{x}))^2 - \frac{1}{2}\right) \tag{1}$$

A degree of alignment parameter, s, is introduced to account for the attenuation of the observed dipolar signal characterizing the residual dipolar coupling. Physically, s corresponds to the population (or molar fraction) of aligning molecules, $s = a(f + a)^{-1}$, where a and f are the molar concentration of the aligning and isotropically tumbling molecules, respectively. We obtain a distribution of conformations $\mathbf{x}$, and magnetic field orientations $\Omega_\mathbf{b}$, given a vector of N *a priori* unknown Lagrange multipliers $\boldsymbol{\lambda}$ and s, at the inverse temperature $\beta = 1/kT$:[3−7,39,42]

$$p(\mathbf{x}, \Omega_\mathbf{b}|\boldsymbol{\lambda}, s) = Z^{-1}(\boldsymbol{\lambda}, s)\ \exp(-\beta[E(\mathbf{x}) + \sum^{ij\in\mathcal{D}} \lambda_{ij}D_{ij}(\mathbf{r}_{ij}(\mathbf{x}), \mathbf{b}, s)]) \tag{2}$$

Each of the Lagrange multipliers, λ_{ij}, is associated with a corresponding experimental dipolar coupling assignment ij in the set of all experimentally assignable RDC observations $\mathcal{D}$, while $\mathbf{r}_{ij}(\mathbf{x})$ denotes the unit interspin vector of the observation ij and $Z^{-1}(\boldsymbol{\lambda},s)$ is a normalization constant. To satisfy all our requirements for the solution we need to choose $\boldsymbol{\lambda}$ and s in a manner such that $\mathbf{D}^{\lambda,s} \approx \mathbf{D}^{\exp}$, where

$$D^{\lambda,s}_{ij} = \int \mathrm{d}\mathbf{x}\mathrm{d}\Omega_\mathbf{b}\ D_{ij}(\mathbf{r}_{ij}(\mathbf{x}), \mathbf{b}, s)p(\mathbf{x}, \Omega_\mathbf{b}|\boldsymbol{\lambda}, s) \tag{3}$$

are the elements of our vector of back-calculated data $\mathbf{D}^{\lambda,s}$, for a given $\boldsymbol{\lambda}$ and s, while $\mathbf{D}^{\exp}$ is the vector of experimental data, that is, our back-calculated data have to agree with the experimental data. Consequently, we need to estimate $N + 1$ unknown parameters (see the Methods section).

We may extend the formalism to incorporate information from M different alignment conditions, $\mathcal{D}_m$. This is achieved by using a sum of M copies of the Lagrange term in eq 2:

$$p(\mathbf{x}, \Omega_{\mathbf{b}_1},..., \Omega_{\mathbf{b}_M}|\boldsymbol{\lambda}^1, s^1,..., \boldsymbol{\lambda}^M, s^M) = Z(\boldsymbol{\lambda}^1, s^1,..., \boldsymbol{\lambda}^M, s^M)^{-1}$$
$$\exp(-\beta[E(\mathbf{x}) + \sum_{m=1}^{M}\sum^{ij\in\mathcal{D}_m} \lambda^m_{ij}D_{ij}(\mathbf{r}_{ij}(\mathbf{x}), \mathbf{b}_m, s^m)]) \tag{4}$$

Here, each of the M terms corresponds to its own alignment condition and has its own separate magnetic field vector $\mathbf{b}_m$, degree of alignment parameter s^m, and set of Lagrange multipliers $\boldsymbol{\lambda}^m$.

The Rigid-Body Limit Enables Computation of Saupe Tensors. The framework defined above yields a distribution of magnetic field orientation and the molecular structure. Thus, we may compute RDC data directly independently of the definition of an explicit molecular frame. We now consider the case addressed in the Saupe tensor formalism where, $\mathbf{x}$, is a rigid body, $\hat{\mathbf{x}}$. Under this condition the maximum entropy distribution, eq 2, becomes a probability density function of magnetic field orientations given $\hat{\mathbf{x}}$, $p(\Omega_\mathbf{b}|\hat{\mathbf{x}}, \boldsymbol{\lambda}, s)$. Following

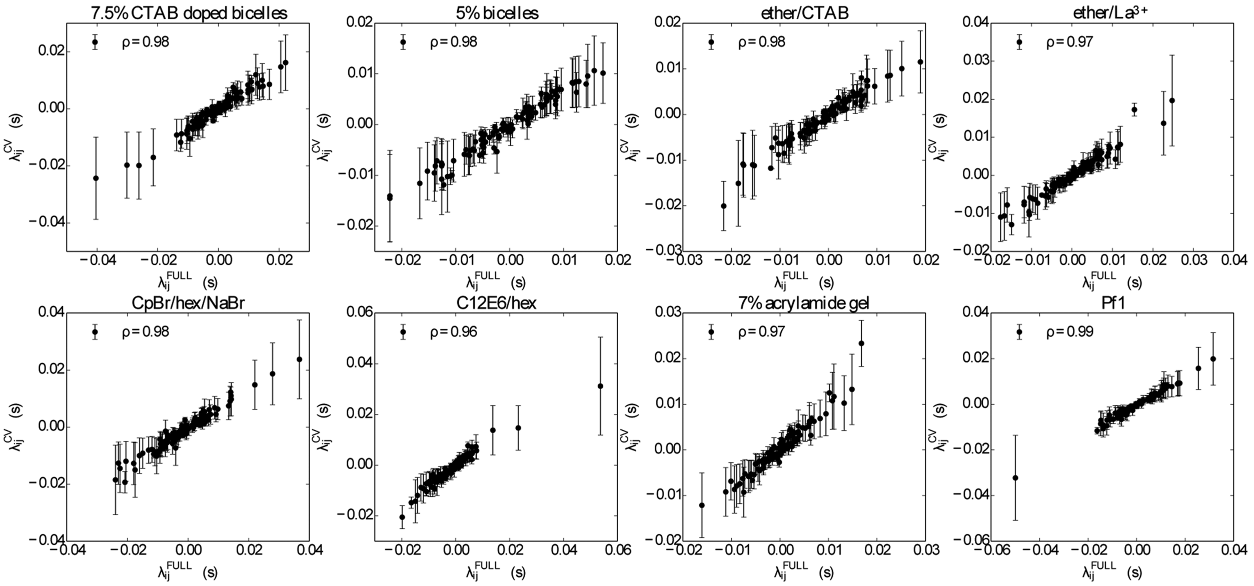


Figure 2. Comparison of Lagrange multipliers estimated using a 5-fold cross validation ($\lambda_{ij}^{\mathrm{CV}}$) and the corresponding values generated using the full data set ($\lambda_{ij}^{\mathrm{FULL}}$) for each of the eight alignment media with data on lysozyme considered herein. Correlation coefficients (ρ) of the series are inset in the upper left corner of each subplot. The error bars denote the standard deviation of the mean of each Lagrange multiplier as computed using the four independent estimates obtained in the 5-fold cross validation.

estimation of $\boldsymbol{\lambda}$ and s, we may compute a Saupe tensor directly from this distribution by using the expectation:

$$\mathbf{S} = \frac{1}{2}\int \mathrm{d}\Omega_{\mathbf{b}}(3\mathbf{b}\otimes\mathbf{b} - \mathbf{I})p(\Omega_{\mathbf{b}}|\hat{\mathbf{x}}, \boldsymbol{\lambda}, s) \tag{5}$$

where ⊗ denotes the tensor product and $\mathbf{I}$ is the 3×3 identity matrix, and $p(\Omega_{\mathbf{b}}|\hat{\mathbf{x}}, \boldsymbol{\lambda}, s)$ is the distribution of orientations of the magnetic field relative to the orientation of $\hat{\mathbf{x}}$.

■ METHODS

Estimation of the Parameters $\boldsymbol{\lambda}$ and s. The unknown parameters $\boldsymbol{\lambda}$ and s can be computed by maximizing the agreement of the back-calculated RDCs with experimental data. In other words, we are looking for the set of parameters $\boldsymbol{\lambda}$ and s that minimizes the expression:

$$\chi^2 = \sum^{ij\in\mathcal{D}} (D_{ij}^{\mathrm{exp}} - D_{ij}^{\lambda,s})^2 \tag{6}$$

or

$$\chi^2 = \sum^{ij\in\mathcal{D}} \frac{(D_{ij}^{\mathrm{exp}} - D_{ij}^{\lambda,s})^2}{\sigma^2} \tag{7}$$

if quantitative information about experimental uncertainty, σ, is available.

It is possible to estimate the degree of alignment parameter, s, *a priori* via the histogram method or extensions thereof.[43,44] Here, we compute s at each estimation step as the value that minimizes eq 6 or 7, i.e.

$$s^{(n)} = \underset{s}{\mathrm{argmin}} \sum^{ij\in\mathcal{D}} (D_{ij}^{\mathrm{exp}} - D_{ij}^{\lambda,s})^2 \tag{8}$$

Given eq 6 or 7 we can find the optimal $\boldsymbol{\lambda}$ by a simple steepest descent in the space of parameters:

$$\lambda_{ij}^{(n+1)} = \lambda_{ij}^{(n)} + 2\Delta t\left(D_{ij}^{\mathrm{exp}} - D_{ij}^{\lambda^{(n)},s^{(n)}}\right)\frac{\partial}{\partial\lambda_{ij}^{(n)}}D_{ij}^{\lambda^{(n)},s^{(n)}} \tag{9}$$

where $\lambda^{(n)}$ and $s^{(n)}$ are the parameters after the n-th iteration, Δt is the step size parameter scaled by σ^{-2}, and the partial derivatives are

$$\begin{aligned}\frac{\partial}{\partial\lambda_{ij}}D_{ij}^{\lambda,s} &= -\beta\left(\int \mathrm{d}\mathbf{x}\mathrm{d}\Omega_{\mathbf{b}}\ p(\Omega_{\mathbf{b}}, \mathbf{x}|\lambda, s)D_{ij}^{\lambda,s}(\mathbf{r}_{ij}(\mathbf{x}), \mathbf{b}, s)^2 - (D_{ij}^{\lambda,s})^2\right)\\ &= -\beta\,\mathrm{Var}\left(D_{ij}^{\lambda,s}\right)\end{aligned} \tag{10}$$

Given our experimental RDC data are conditionally independent we are guaranteed to obtain at a unique set of Lagrange multipliers, regardless of starting conditions, as the optimization problem is convex.[45] If the data are not strictly so the maximum entropy distribution will still be uniquely defined, but we will converge to a space of equivalent Lagrange multipliers.[46,47]

Molecular Dynamics Simulations. All simulations were carried out in the `almost`[48] molecular dynamics (MD) framework using the Amber03 force field[49] and a generalized Born implicit solvation model.[50] We employed SHAKE constraints on all bonds involving protons[51] and 2 fs timesteps. Simulations were kept at 300 K using a Berendsen thermostat.[52] The simulations were started from structures deposited in the Protein Data Bank with accession codes 1E8L[53] (for lysozyme simulations) or 1Q6H[54] (for sFkpA simulations) following an energy minimization in the force field by steepest decent. All threads were initialized with from the same coordinates but different initial velocities (sampled according to the Maxwell−Boltzmann distribution at 300 K) and different random seeds. The force field was restrained by the RDC data using eqs 2 and 4 yielding a full pseudo-energy:

$$E(\mathbf{x}, \Omega_{\mathbf{b}_1},..., \Omega_{\mathbf{b}_M}) = E(\mathbf{x}) + \sum_{m=1}^{M}\sum^{ij\in\mathcal{D}_m} \lambda_{ij}^{m}D_{ij}(\mathbf{r}_{ij}(\mathbf{x}), \mathbf{b}_m, s^m) \tag{11}$$

The magnetic field orientations $\Omega_{\mathbf{b}}$ were updated following each MD step by 10 Monte Carlo (MC) steps. A MC step consisted of the proposal new $\Omega_{\mathbf{b}}'$ from the uniform distribution on the unit sphere followed by an evaluation according to Metropolis' criterion[55] using the potential expression (eq 11):

$$\Delta E_i = E(\mathbf{x}, \Omega_{\mathbf{b}_1},..., \Omega_{\mathbf{b}_i}',..., \Omega_{\mathbf{b}_M}) - E(\mathbf{x}, \Omega_{\mathbf{b}_1},..., \Omega_{\mathbf{b}_i},..., \Omega_{\mathbf{b}_M}) \tag{12}$$

Here, we refer to a MD/MC step as to one MD step followed by 10 MC steps; the MC steps were done independently for each of the alignment conditions, i. The parameters $\boldsymbol{\lambda}$ and s were estimated in an iterative fashion using 384 independent short simulations run in parrallel (4 ps for sFkpA and 4 ps for lysozyme) for each estimation step. For lysozyme, estimation was also performed with 64 independent simulations to test stability. The moments in eq 9 used for updating the $\boldsymbol{\lambda}$ and s parameters were computed using every 20th (for lysozyme) or 10th (for sFkpA) MD/MC step of all 384 simulations. The parameters $\boldsymbol{\lambda}$ and s were updated every 1000 MD/MC steps for both sFkpA and lysozyme. Following a burn-in period of 50 and 60 $\boldsymbol{\lambda}$, s-estimation steps for lysozyme and sFkpA, respectively, we collect a production ensemble (65 steps for sFkpA and 50 steps for lysozyme). Due to finite sampling errors the parameters will fluctuate stochastically around their average true values. For this reason we continue updating the parameters during the production phase of the simulation as in the burn-in phase. Atomic coordinates were saved every 2000th MD/MC step.

In general, we find that increasing the length of short simulation has only a minor influence on the convergence rate whereas increasing the number of independent simulations increases the convergence rate and overall stability. Recent methodological developments may provide the means for future improvements in the parameter estimation.[56]

Cross-Validation Tests of Lysozyme. The 5-fold cross-validation was performed using an estimation procedure identical to that described for lysozyme above. A total of 100 estimation steps were carried out per cross-validation. A total of 76,800 frames were generated for each cross-validation and the last 38,400 were used to compute Q-factors as well as ^3J couplings reported in the results section. Average $\boldsymbol{\lambda}^m$ and s^m values shown in Figure 2 and Supporting Table 1 were computed using the final estimate.

Unrestrained Simulation. Unrestrained reference ensemble of lysozyme was generated as in the protocol above albeit without employing the RDC restraints and using 16 independent molecular simulations running for a total of 124 ns.

Estimation of the Probability Density Function of Magnetic Field Orientations in a Rigid-Body Frame. Estimation of $p(\Omega_b|\hat{\mathbf{x}}, \boldsymbol{\lambda}, s)$ was performed by eliminating the MD steps from the procedure above. Consequently, we simply performed MC sampling of the magnetic field orientations. The Lagrange multipliers $\boldsymbol{\lambda}$ and the parameter s were estimated using 50000 MC steps per estimation step, for a total of 200 steps. Convergence was monitored every second estimation step. After convergence, s, was kept fixed, another 200 steps were performed to ensure stability. This procedure was carried out for GB3 and lysozyme, respectively. For GB3 the result from the last estimation step was used to calculate the values in Table 1, using eq 5. Similarly, the final estimates ($\lambda_{ij}^{\text{STATIC}}$) were used for comparison with the corresponding dynamic ones in Supporting Figure 3.

Table 1. Analysis of Saupe Alignment Tensors in Seven Different GB3 Constructs with Different Alignment Properties[a] (see refs 23 and 58)

GB3 mutant	ρ	$d_F(\mathbf{S}_{\text{SVD}}, \mathbf{S}_{\text{ME}})$
K19AD47K	0.999953	0.403
K19AT11K	0.999923	0.332
K19ED40N	0.999962	0.307
K19EK4A-C-His6	0.999939	0.425
K19EK4A-N-His6	0.999991	0.289
K19EK4A	0.999956	0.402
WT	0.984370	1.184

[a] Correlation coefficient (ρ) and Frobenius distance ($d_F(\mathbf{S}_{\text{SVD}}, \mathbf{S}_{\text{ME}})$) of Saupe tensors computed using the reorientation expectation $\mathbf{S}_{\text{ME}}$ (eq 5) and using singular value decomposition $\mathbf{S}_{\text{SVD}}$, respectively. The (Frobenius) distance between matrices A and B is given by $d_F(A, B) = \|A - B\|_F$, where $\|U\|_F = (\text{tr}(UU^T))^{1/2}$. tr(·) is the matrix trace, and U^T the matrix transpose of U.

RESULTS

The Rigid-Body Limit of the Maximum Entropy Formalism of RDCs. To gain some intuition about the presented formalism we consider the limit of where the structure, $\mathbf{x}$, is a rigid body, $\hat{\mathbf{x}}$. This limit corresponds to the one treated in the classic Saupe tensor formalism. Furthermore, as described above, in this case we may compute Saupe tensors following the estimation of the parameters $\boldsymbol{\lambda}$ and s. Doing so we compare Saupe tensors computed in this fashion to those computed with the conventional fitting procedure,[57] using experimental N–H RDCs of the third immunoglobulin binding domain of protein G (GB3) recorded in seven different alignment conditions.[23,58] In all cases, we find the tensors to be identical, if we consider the statistical uncertainty (Table 1).

Robust Generation of a Native-State Ensemble of Lysozyme Using Data from Multiple Different Alignment Conditions. The native state dynamics of hen lysozyme has been thoroughly characterized by experiment[24] and simulation.[14,59] Consequently, it constitutes an ideal test-case for our newly established framework. Using eq 4 we analyze the native-state dynamics of lysozyme using H–N RDCs acquired in eight different alignment media.[24] Following 50 estimation steps the $\boldsymbol{\lambda}^m$ (see Supporting Figure 1) and s^m parameters stabilize. This enables us to generate an ensemble of lysozyme in excellent agreement with the RDC data, as well as, complementary ^3J coupling[60] and NOE data,[53] see Tables 2 and 3 and Supporting Figure 2a.

Table 2. Quantitative Evaluation of Lysozyme Models

	current study	1E8L[53]	DeSimone[i]	Amber03
^3JHα–HN[60] (RMSD, Hz)	0.95[k]/0.95[l]	1.32[k]/1.36[l]	0.67/–	1.28[k]/1.40[l]
H–N RDC[a] (Q)	0.036	0.079[j]	0.171	0.534[j]
H–N RDC[b] (Q)	0.048	0.082[j]	0.142	0.554[j]
H–N RDC[c] (Q)	0.042	0.355[j]	0.138	0.562[j]
H–N RDC[d] (Q)	0.072	0.279[j]	0.184	0.518[j]
H–N RDC[e] (Q)	0.045	0.340[j]	0.221	0.586[j]
H–N RDC[f] (Q)	0.046	0.324[j]	0.159	0.58[j]
H–N RDC[g] (Q)	0.064	0.443[j]	0.219	0.552[j]
H–N RDC[h] (Q)	0.076	0.355[j]	0.196	0.518[j]

[a] Alignment media: 7.5% CTAB doped bicelles. [b] Alignment media: 5% bicelles. [c] Alignment media: ether/CTAB. [d] Alignment media: ether/La^{3+}. [e] Alignment media: CpBr/hex/NaBr. [f] Alignment media: C12E6/hex. [g] Alignment media: 7% acrylamide gel. [h] Alignment media: Pf1. [i] All values are as reported in ref 14. [j] Computed using ensemble averaged SVD.[16] [k] Computed using Karplus parameters as reported and described in ref 61. [l] Computed using Karplus parameters reported described in ref 61.

We further validated our ensemble model and the parameter estimation procedure by a 5-fold cross-validation test. Here, we generated five randomized, mutually exclusive training and test sets comprised of 80% and 20% of the original data, respectively. We subsequently repeated the estimation of $\boldsymbol{\lambda}^m$ and s^m using each of the five training sets independently resulting in five different ensembles (CVn, n = 1, ..., 5). We held the corresponding test sets aside for validation. During estimation we find that both the training set (Q_{work}) and test set (Q_{free}) stabilize rapidly to values that suggest good

Table 3. Average Violations of 523 Backbone−Backbone NOE Derived Distance Restraints in Seven Ensembles of Lysozyme[a]

model	average violation (Å)	*p* value
CV1	0.1711	–
CV2	0.1681	–
CV3	0.1645	–
CV4	0.1682	–
CV5	0.1671	–
no data	0.1792	5.993×10^{-7}
all RDC data	0.1592	10^{-4}

[a] The *p*-values of the unrestrained (no data) and restrained with all RDC data (all RDC data) are computed relative to a null-hypothesis defined by the cross-validation ensembles (CV1−5).

agreement. Not surprisingly, we find that Q_{free} values are systematically larger than their corresponding Q_{work} values, see Table 4. However, there are no signs of overfitting as we do not see any systematic increases in the Q_{free} statistics as a function of the estimation step (see Supporting Figure 2b−f). In addition, the ^3J coupling and NOE data back-calculated from the cross-validation ensembles in general shows an inferior correlation with experimental data when compared to the ensemble generated using the full data set (See Tables 2, 3, and 4). Lagrange multipliers and degree of alignment parameters estimated using the full data set and those computed from the cross-validation estimation correlate surprisingly well, see Figure 2 and Supporting Table 1.

Finally, we characterize our ensemble in terms of a pincer motion (see Figure 3) as described previously.[14] The distribution of this angle, θ, is in qualitative agreement with a previous report.[14] However, our ensemble suggests that a slightly wider spectrum of pincer angles is sampled while agreeing with ^3J coupling data either on par[14] or better agreement when compared to other models (Table 2). Additionally, we compared free and antibody bound states of lysozyme deposited in the Protein Data Bank in terms of θ and their radius of gyration, R_g. The antibody bound forms of lysozyme cluster in two discrete classes, one which is similar to the *apo* forms, and one which appears to be a closed state (Figure 3B). Conformations similar to both states are consistently sampled in our MD simulations using the full data set, either of the five cross-validation data sets or only the Amber03 force field (see Figure 3B and Supporting Figure 4). However, the unrestrained simulation samples an overly wide distribution of the pincer angle, which could also explain the inferior agreement with experimental data.

Refining Dynamic Models of Intra- And Interdomain Flexibility. The tensor-free formalism presented here is independent of separation of internal and overall dynamics which is known to often hamper analysis of data on flexible system such as multidomain proteins. For example, the analysis of RDC data in these cases involves manual expert assessments, often guided by complementary experiment, which in turn are used as the basis to infer a motional model. Using the presented framework we may supersede such manual hypothesis driven inference of dynamics in flexible biomolecules. As an example we here turn to the ATP-independent chaperone and peptidyl-prolyl *cis*/*trans*-isomerase (PPIase), FkpA-ΔCT (sFkpA). The dimeric native state of sFkpA is stabilized by intertwined helices in dimerization domain of two independent chains. Through a flexible α-helix (helix III) the dimerization domains are connected to their corresponding catalytic domains. In full, sFkpA appears as in characteristic V-shape (see Figure 4).

We apply our framework to analyze the interdomain flexibility in sFkpA following the protocol discussed above, using data from two alignment conditions ($C_{12}E_5$/hexanol/H_2O liquid crystals and Pf1 filamentous bacteriophages).[27] Similarly for lysozyme, the parameters $\boldsymbol{\lambda}$ and *s* stabilizes after approximately 50 estimation steps. Due to the relatively anisotropic shape of sFkpA the degree of alignment parameters *s* (0.55% and 0.39%) are approximately 30−100% larger than those observed for lysozyme.[62] From step 60 and on we generate a production ensemble with average *Q*-factor suggesting excellent agreement (Q_{average} = 0.06) between the generated conformational ensembles and the experimental data (Supporting Figure 5). We find that the two C-terminal binding domains undergo significant uncorrelated motions with respect to the N-terminal dimerization domains, (Figures 4 and 5).

Since sFkpA is a symmetric homodimer the RDC data used here is an average of the two chains as they have degenerate chemical shifts. Therefore, we use the same data sets to restrain each of the chains in the dimer. Thus, we have an internal litmus test of convergence: agreement of the Lagrange multipliers obtained in each of the independent chains. The correlation coefficient of the Lagrange multipliers obtained in the two chains is 0.99 for the data obtained in both of the alignment media considered here. Similarly, the degree of alignment parameters are both within a 8% relative error of each other in the two chains.

Table 4. Quantitative Evaluation of the 5-Fold Cross-Validation Analysis of Lysozyme

	CV1	CV2	CV3	CV4	CV5
^3JHα−HN[60] (RMSD, Hz)	0.991[i]/0.996[j]	0.962[i]/0.97[j]	0.981[i]/0.987[j]	1.009[i]/1.013[j]	0.965[i]/0.971[j]
H−N RDC[a] ($Q_{\text{work}}/Q_{\text{free}}$)	0.036/0.29	0.033/0.196	0.033/0.298	0.036/0.23	0.036/0.183
H−N RDC[b] ($Q_{\text{work}}/Q_{\text{free}}$)	0.042/0.292	0.041/0.201	0.046/0.221	0.043/0.209	0.048/0.25
H−N RDC[c] ($Q_{\text{work}}/Q_{\text{free}}$)	0.038/0.247	0.041/0.236	0.044/0.239	0.039/0.131	0.042/0.22
H−N RDC[d] ($Q_{\text{work}}/Q_{\text{free}}$)	0.065/0.384	0.068/0.448	0.057/0.174	0.057/0.328	0.068/0.207
H−N RDC[e] ($Q_{\text{work}}/Q_{\text{free}}$)	0.043/0.27	0.039/0.182	0.041/0.334	0.042/0.253	0.042/0.158
H−N RDC[f] ($Q_{\text{work}}/Q_{\text{free}}$)	0.039/0.398	0.046/0.308	0.04/0.133	0.036/0.433	0.045/0.136
H−N RDC[g] ($Q_{\text{work}}/Q_{\text{free}}$)	0.076/0.506	0.075/0.18	0.075/0.455	0.073/0.301	0.073/0.234
H−N RDC[h] ($Q_{\text{work}}/Q_{\text{free}}$)	0.064/0.374	0.064/0.53	0.069/0.495	0.056/0.409	0.067/0.306

[a] Alignment media: 7.5% CTAB doped bicelles. [b] Alignment media: 5% bicelles. [c] Alignment media: ether/CTAB. [d] Alignment media: ether/La^{3+}. [e] Alignment media: CpBr/hex/NaBr. [f] Alignment media: $C_{12}E_6$/hex. [g] Alignment media: 7% acrylamide gel. [h] Alignment media: Pf1. [i] Computed using least-squares fitting as described in ref 61. [j] Computed using Karplus parameters reported described in ref 61.

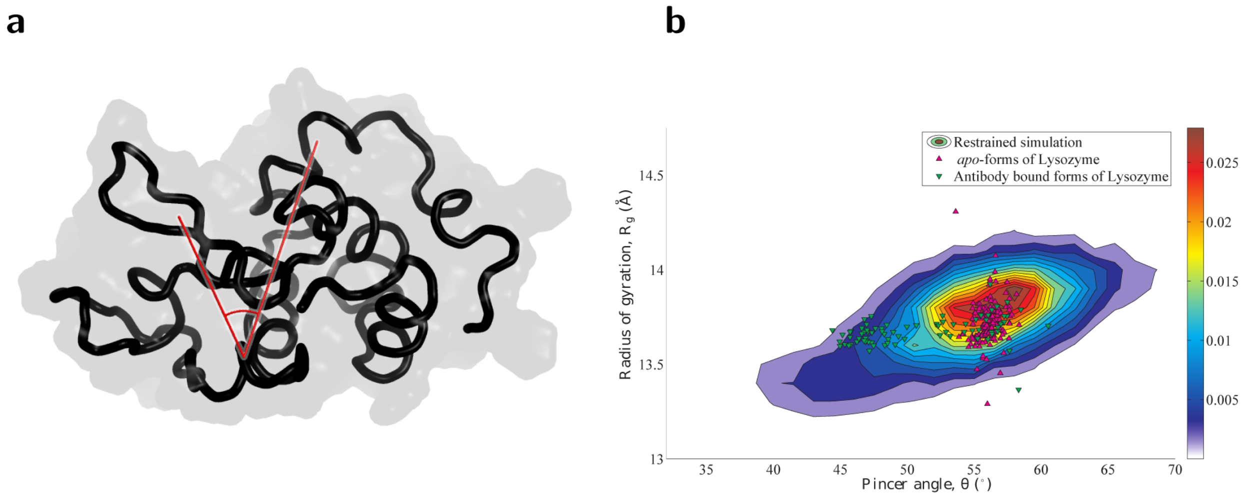


Figure 3. (a) Schematic illustration of pincer angle θ (red) between the center of mass of domains A and B interconnected by a hinge domain. The domains are defined by Cα atoms of residues (111–114), (80–84, 90–93), and (44–45, 51–52) for domain A, hinge, and B, respectively. (b)Two-dimensional contour plot of the probability density distribution of θ and Cα radius of gyration, R_g, in the maximum entropy restrained simulation. Scatter points illustrate *apo* (pink upward-pointing triangles) and *holo* (green downward-pointing triangles) forms of lysozyme deposited in the protein data bank.

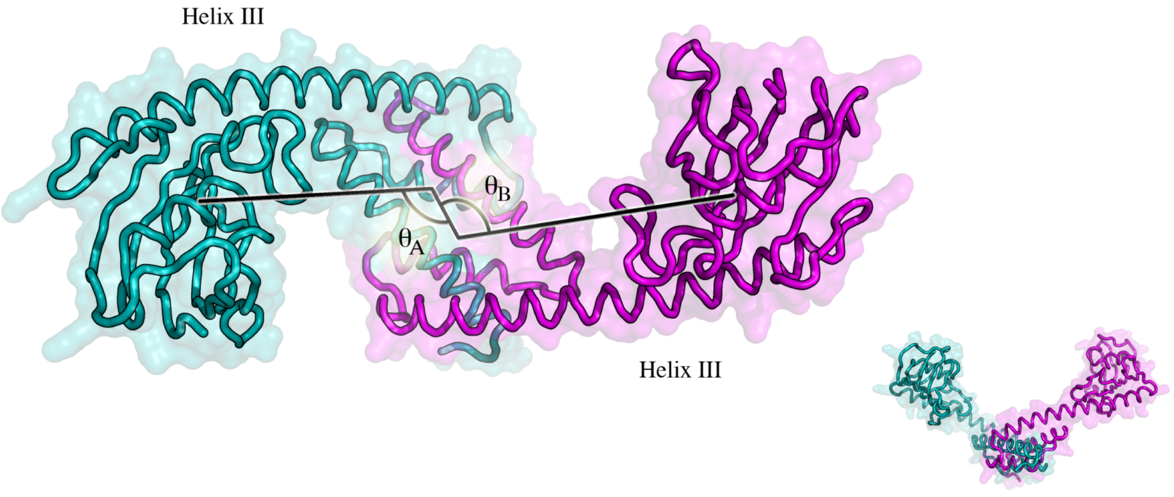


Figure 4. Top view of sFkpA with a schematic illustration of the interdomain flexibility measures used to assess qualitative agreement of the sFkpA ensemble with NMR relaxation data in helix III. Each of the two chains (A, B) in are highlighted with cyan and purple colors, respectively. θ_A and θ_B are defined as the angles between vectors spanned by the center of mass between Cα atoms of residues 14–84 of chain A and B and residues 85–224 in chain A and B, respectively. Small side view render of sFkpA is shown in the lower right corner.

■ DISCUSSION

We present a framework to analyze RDC data independently of the definition of alignment tensors and a separation of internal and overall dynamics. Consequently, we may rigorously treat data recorded on dynamically disparate systems in a unified manner. Briefly, the framework is quantitatively summarized in the eqs 2 and 4 and effectively constitutes a probability density function based on a molecular mechanics force field and a weighted sum of angular terms coupling interatomic spin vectors $\mathbf{r}_{ij}(\mathbf{x})$ to an external magnetic field, **b**. In this manner, the approach integrates the geometrical information held within the RDC data without making prior assumptions beyond the common: the observed data represents an average quantity, eq 1 realistically describes the experimental data, and finally the alignment is partial and thus only a fraction of the molecules are aligning.

We use the thoroughly characterized protein hen lysozyme to test and validate the presented framework. We are able to generate ensembles which are in excellent agreement with the data from eight different alignment data simultaneously. Additionally, we show through a 5-fold cross-validation test that the estimation of the unknown Lagrange multipliers and degree of alignment parameter is highly robust. The robustness depends on the number of independent simulations used in each estimation step, but the overall results are the same (see Supporting Figure 1). Our cross-validation test further shows no signs of overfitting; the sheer consequence was lower overall quality of the cross-validation ensembles as compared to the ensemble generated using the full data set as seen through validation using complementary NOE and ^3J coupling data (Tables 2, 3, and 4). Consequently, the approach will also have a distinct advantage in the case where only sparse data is available, as it will facilitate a balanced analysis of all the experimental data with minimal risk of overfitting.

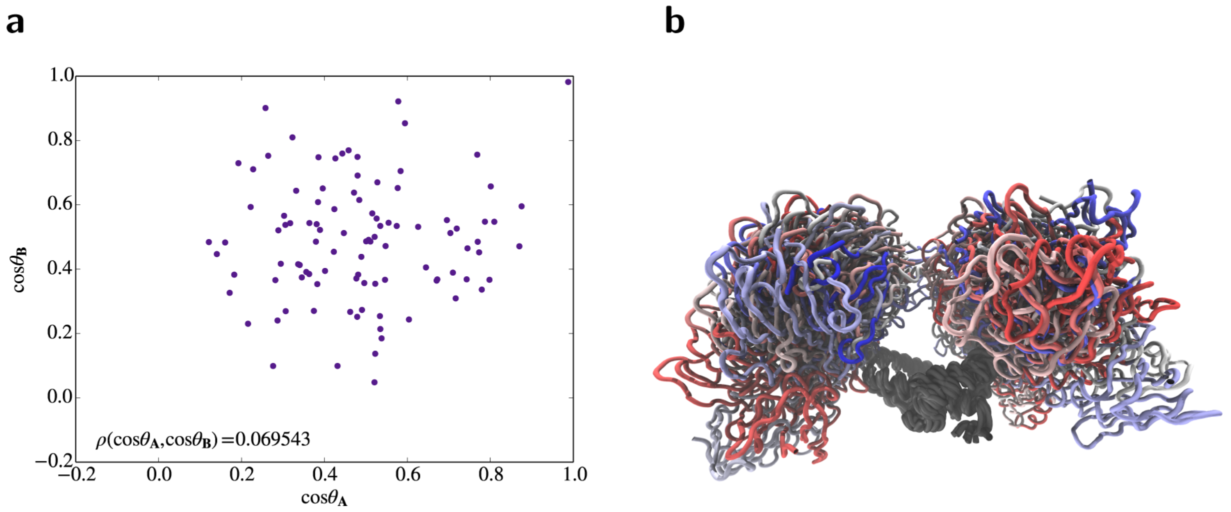


Figure 5. Illustration of uncorrelated motion observed in the sFkpA ensemble generated in the present study. (a) Scatterplot of the cosines of angles θ_A and θ_B as defined in Figure 4. Inset value (ρ) is the computed Pearson's correlation coefficient using 100 randomly selected structures from the full ensemble. (b) Ray-trace rendering of 50 randomly selected structures from the sFkpA ensemble, with the dimerization domains are shown in gray and the catalytic domains are shown with a diverging color gradient.[66]

The Lagrange multipliers here measure the amount information one needs to add to the prior structural model (e.g., a force field or a static structure) in order to obtain agreement with the given data. To explore this we compared parameters estimated when we assume a static structural model to those obtained with a dynamic one (see Supporting Figure 3). We observe two differences between the Lagrange multipliers. First, the overall scale of the Lagrange multipliers in the static case are approximately an order of magnitude larger, with the exception of the two data sets used to refine the static structure. Second, we see a very poor correlation between the two different sets of Lagrange multipliers. Consequently, this directly shows that one needs to introduce much more bias (information) to enable the inherently dynamically averaged RDC data to agree with a static model. Furthermore, it suggests that it is not generally possible to obtain reasonable, or even relative, estimates of the Lagrange multipliers by using a static structure.

Having thoroughly scrutinized the presented framework using lysozyme, we move on to study the case of sFkpA where the separation of internal and overall dynamics is not as clear. Indeed, we are able to generate an ensemble in excellent quantitative agreement with the RDC data. Previously, Hu et al. reported that it was necessary to manually decompose the catalytic and dimerization domains from each other to enable fitting of Saupe tensors which were in reasonable, albeit not ideal agreement with the data. The decomposition was in turn guided by complementary NMR relaxation data which suggested that helix 3 (see Figure 4) has a number of highly flexible residues (residues 84−91). The ensemble we present is in qualitative agreement with this data also. In fact, we observe uncorrelated motions between the two C-terminal domains with respect to the dimerization domains (see Figure 5). Hu et al. proposed the C-terminal domains move independently of each other to accommodate wide substrate specificity and subsequent chaperone action through a polymorphic chaperone−substrate interaction surface.[27] However, as the authors pointed out, direct observation of such motion was complicated by necessity to deconvolute internal from overall motion. The ensemble presented here is an atomic detail model of such motion which we obtain independently of such deconvolution. Finally, we observe considerable intradomain flexibility, in particular in the, N-terminal dimerization domain with a mean backbone RMSD of 2.6 Å. This is consistent with a previous observation that the N-terminal dimerization domain, unlike the C-terminal binding domain, might be insufficiently represented by the published crystal structure. Flexibility-driven mechanisms have been reported for a number of other ATP-independent chaperones.[63−65] This indicates that the ensemble we present here may be a biologically relevant one.

■ CONCLUSION

In conclusion, we present a new framework for direct, tensor-free, analysis of RDC data in terms of dynamic models of biological macromolecules which holds the Saupe tensor formalism as a special case. Further, we demonstrate the methods broad applicability and tractability through two examples were we robustly determine native state ensembles of two structurally and dynamically different proteins. Specifically, we show that flexible multidomain systems may be analyzed in a fashion completely analogous to that of relatively rigid proteins. We anticipate the approach to extend beyond the type of systems addressed here and, in particular, excel for very flexible systems where deconvolution of internal and overall motion is not possible.

The method has been implemented in the freely available open source frameworks for molecular simulations `almost`[48] and `phaistos`[67] as well as the `PLUMED2` framework.[68] We have made a number of example scripts publicly available which should make it easy for the scientific community to adopt the presented methodology at https://github.com/cavallilab/meprdc.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional figures and tables showing convergence of Lagrange multipliers and Q factors in parameter estimation, joint probability densities of pincer-angle and R_g and auxiliary results. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b01289.

■ AUTHOR INFORMATION

Corresponding Author

*andrea.cavalli@irb.usi.ch

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Beat Vögeli for critically reading the manuscript and many stimulating discussions. S.O. is funded by an Independent Postdoc grant from The Danish Council for Independent Research Natural Sciences (ID: DFF-4002-00151).

■ REFERENCES

(1) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679−85.

(2) Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133*, 10336−9.

(3) Pitera, J. W.; Chodera, J. D. J. *J. Chem. Theory Comput.* **2012**, *8*, 3445.

(4) Roux, B.; Weare, J. *J. Chem. Phys.* **2013**, *138*, 084107.

(5) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 094112.

(6) Olsson, S.; Frellsen, J.; Boomsma, W.; Mardia, K. V.; Hamelryck, T. *PLoS One* **2013**, *8*, e79439.

(7) Olsson, S.; Vögeli, B. R.; Cavalli, A.; Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K.; Hamelryck, T. *J. Chem. Theory Comput.* **2014**, *10*, 3484−3491.

(8) Beauchamp, K. A.; Pande, V. S.; Das, R. *Biophys. J.* **2014**, *106*, 1381−90.

(9) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128−32.

(10) Vögeli, B.; Kazemi, S.; Güntert, P.; Riek, R. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1053−7.

(11) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964−72.

(12) Bouvignies, G.; Bernadó, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885−90.

(13) Clore, G. M.; Schwieters, C. D. *J. Mol. Biol.* **2006**, *355*, 879−86.

(14) de Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M. *Biochemistry* **2013**, *52*, 6480−6.

(15) Thaning, J.; Stevensson, B.; Ostervall, J.; Naidoo, K. J.; Widmalm, G.; Maliniak, A. *J. Phys. Chem. B* **2008**, *112*, 8434−6.

(16) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158−4159.

(17) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471−5.

(18) Esteban-Martín, S.; Fenwick, R. B.; Salvatella, X. *J. Am. Chem. Soc.* **2010**, *132*, 4626−32.

(19) Fisher, C. K.; Huang, A.; Stultz, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 14919−27.

(20) Prestegard, J. H.; al Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424.

(21) Bax, A. *Protein Sci.* **2003**, *12*, 1−16.

(22) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279−83.

(23) Yao, L.; Bax, A. *J. Am. Chem. Soc.* **2007**, *129*, 11326−7.

(24) Higman, V. A.; Boyd, J.; Smith, L. J.; Redfield, C. *J. Biomol. NMR* **2011**, *49*, 53−60.

(25) Saupe, A. *Naturforscher* **1964**, *19a*, 161−171.

(26) Salvatella, X.; Richter, B.; Vendruscolo, M. *J. Biomol. NMR* **2008**, *40*, 71−81.

(27) Hu, K.; Galius, V.; Pervushin, K. *Biochemistry* **2006**, *45*, 11983−91.

(28) Sanchez-Martinez, M.; Crehuet, R. *Phys. Chem. Chem. Phys.* **2014**, *16*, 26030−9.

(29) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7804−5.

(30) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407−18.

(31) Camilloni, C.; Vendruscolo, M. *J. Phys. Chem. B* **2015**, *119*, 653−661.

(32) Catalano, D.; Forte, C.; Veracini, C. A.; Zannoni, C. *Isr. J. Chem.* **1983**, *23*, 283−289.

(33) Catalano, D.; di Bari, L.; Veracini, C. A.; Shilstone, G. N.; Zannoni, C. *J. Chem. Phys.* **1991**, *94*, 3928−3935.

(34) Catalano, D.; Emsley, J. W.; la Penna, G.; Veracini, C. A. *J. Chem. Phys.* **1996**, *105*, 10595−10605.

(35) Berardi, R.; Spinozzi, F.; Zannoni, C. *J. Chem. Phys.* **1998**, *109*, 3742−3759.

(36) Stevensson, B.; Sandström, D.; Maliniak, A. *J. Chem. Phys.* **2003**, *119*, 2738−2746.

(37) Celebre, G.; Cinacchi, G. *J. Chem. Phys.* **2006**, *124*, 176101.

(38) Thiele, C. M.; Schmidts, V.; Böttcher, B.; Louzao, I.; Berger, R.; Maliniak, A.; Stevensson, B. *Angew. Chem., Int. Ed. Engl.* **2009**, *48*, 6708−12.

(39) Jaynes, E. T. In *Probability Theory: The Logic of Science*; Bretthorst, G. L., Ed.; Cambridge University Press: Cambridge, U.K., 2003.

(40) Stevensson, B.; Landersjö, C.; Widmalm, G.; Maliniak, A. *J. Am. Chem. Soc.* **2002**, *124*, 5946−5947.

(41) Landersjö, C.; Stevensson, B.; Eklund, R.; Ostervall, J.; Söderman, P.; Widmalm, G.; Maliniak, A. *J. Biomol. NMR* **2006**, *35*, 89−101.

(42) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. *PLoS Comput. Biol.* **2014**, *10*, e1003406.

(43) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159−62.

(44) Habeck, M.; Nilges, M.; Rieping, W. *J. Biomol. NMR* **2008**, *40*, 135−44.

(45) Agmon, N.; Alhassid, Y.; Levine, R. *J. Comput. Phys.* **1979**, *30*, 250 − 258.

(46) Procaccia, I.; Shimoni, Y.; Levine, R. D. *J. Chem. Phys.* **1976**, *65*, 3284−3301.

(47) Alhassid, Y.; Agmon, N.; Levine, R. *Chem. Phys. Lett.* **1978**, *53*, 22−26.

(48) Fu, B.; Sahakyan, A. B.; Camilloni, C.; Tartaglia, G. G.; Paci, E.; Caflisch, A.; Vendruscolo, M.; Cavalli, A. *J. Comput. Chem.* **2014**, *35*, 1101−1105.

(49) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999−2012.

(50) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297−304.

(51) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.

(52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(53) Schwalbe, H.; Grimshaw, S. B.; Spencer, A.; Buck, M.; Boyd, J.; Dobson, C. M.; Redfield, C.; Smith, L. J. *Protein Sci.* **2001**, *10*, 677−88.

(54) Saul, F. A.; Arié, J.-P.; Vulliez-le Normand, B.; Kahn, R.; Betton, J.-M.; Bentley, G. A. *J. Mol. Biol.* **2004**, *335*, 595−608.

(55) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.

(56) Habeck, M. *Phys. Rev. E* **2014**, *89*, 052113.

(57) Losonczi, J. A.; Andrec, M.; Fischer, M. W.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334 − 342.

(58) Yao, L.; Vögeli, B.; Ying, J.; Bax, A. *J. Am. Chem. Soc.* **2008**, *130*, 16518−20.

(59) McCammon, J. A.; Gelin, B. R.; Karplus, M.; Wolynes, P. G. *Nature* **1976**, *262*, 325−6.
(60) Smith, L. J.; Sutcliffe, M. J.; Redfield, C.; Dobson, C. M. *Biochemistry* **1991**, *30*, 986−996.
(61) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *J. Biomol. NMR* **2005**, *32*, 273−280.
(62) Louhivuori, M.; Otten, R.; Lindorff-Larsen, K.; Annila, A. *J. Am. Chem. Soc.* **2006**, *128*, 4371−6.
(63) Tapley, T. L.; Körner, J. L.; Barge, M. T.; Hupfeld, J.; Schauerte, J. A.; Gafni, A.; Jakob, U.; Bardwell, J. C. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5557−62.
(64) Quan, S.; Wang, L.; Petrotchenko, E. V.; Makepeace, K. A.; Horowitz, S.; Yang, J.; Zhang, Y.; Borchers, C. H.; Bardwell, J. C. *Elife* **2014**, *3*, e01584.
(65) Quan, S.; Koldewey, P.; Tapley, T.; Kirsch, N.; Ruane, K. M.; Pfizenmaier, J.; Shi, R.; Hofmann, S.; Foit, L.; Ren, G.; Jakob, U.; Xu, Z.; Cygler, M.; Bardwell, J. C. A. *Nat. Struct. Mol. Biol.* **2011**, *18*, 262−9.
(66) Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33−38.
(67) Boomsma, W.; et al. *J. Comput. Chem.* **2013**, *34*, 1697−1705.
(68) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604−613.