Unraveling motion in proteins by combining NMR relaxometry and molecular dynamics simulations: A case study on ubiquitin

Originally published in:
The Journal of Chemical Physics 160(10), https://doi.org/10.1063/5.0188416

Funding acknowledgement:
899683 - FET Open – Novel ideas for radically new technologies (EC)

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Unraveling motion in proteins by combining NMR relaxometry and molecular dynamics simulations: A case study on ubiquitin

Cite as: J. Chem. Phys. 160, 104105 (2024); doi: 10.1063/5.0188416
Submitted: 21 November 2023 • Accepted: 20 February 2024 •
Published Online: 11 March 2024

Candide Champion,[1] Marc Lehner,[1] Albert A. Smith,[2] Fabien Ferrage,[3] Nicolas Bolik-Coulon,[4,5,6] and Sereina Riniker[1,a)]

AFFILIATIONS

[1] Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland
[2] Institute for Medical Physics and Biophysics, Leipzig University, Härtelstrasse 16-18, 04107 Leipzig, Germany
[3] Laboratoire des Biomolécules, LBM, Département de Chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France
[4] Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[5] Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada
[6] Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada

[a)] **Author to whom correspondence should be addressed:** sriniker@ethz.ch

ABSTRACT

Nuclear magnetic resonance (NMR) relaxation experiments shine light onto the dynamics of molecular systems in the picosecond to millisecond timescales. As these methods cannot provide an atomically resolved view of the motion of atoms, functional groups, or domains giving rise to such signals, relaxation techniques have been combined with molecular dynamics (MD) simulations to obtain mechanistic descriptions and gain insights into the functional role of side chain or domain motion. In this work, we present a comparison of five computational methods that permit the joint analysis of MD simulations and NMR relaxation experiments. We discuss their relative strengths and areas of applicability and demonstrate how they may be utilized to interpret the dynamics in MD simulations with the small protein ubiquitin as a test system. We focus on the aliphatic side chains given the rigidity of the backbone of this protein. We find encouraging agreement between experiment, Markov state models built in the χ_1/χ_2 rotamer space of isoleucine residues, explicit rotamer jump models, and a decomposition of the motion using ROMANCE. These methods allow us to ascribe the dynamics to specific rotamer jumps. Simulations with eight different combinations of force field and water model highlight how the different metrics may be employed to pinpoint force field deficiencies. Furthermore, the presented comparison offers a perspective on the utility of NMR relaxation to serve as validation data for the prediction of kinetics by state-of-the-art biomolecular force fields.

I. INTRODUCTION

Biopolymers such as proteins and nucleic acids exist in an ensemble of conformations. They are dynamical entities with biological functions determined by the continuous inter-conversion between states.[1,2] Understanding those dynamics is paramount to elucidate complex biological phenomena, such as enzymatic catalysis, cell signaling, and metabolism, or to engineer proteins for optimized properties or novel functionalities.[3] A wide range of experimental[4] and computational[5,6] methods, often used in conjunction[7] with one another, have been developed over the years to probe various facets of protein dynamics.

Nuclear magnetic resonance (NMR) based methods stand out given their ability to decipher dynamics of physical motions ranging from bond vibrations, side chain rotations, and loop motion to larger conformational changes.[8] The different types of motion

in proteins occur on timescales ranging from picoseconds (ps) to hours, and various experimental methods may be applied to shine light on specific timescales.[9] In particular, extending the domain of applicability of standard relaxation experiments to the nanoseconds–microseconds (ns–μs) regime has remained an open challenge, recently addressed by fast field cycling,[10] high-resolution relaxometry,[11] and other methods.[12–14] While kinetic information is obtained from those experiments, it generally cannot be interpreted directly with mechanistic models at atomic resolution. Given the accessibility of these timescales using all-atom molecular dynamics (MD) simulations,[15] the latter have been combined to complement NMR relaxation experiments.

The simplest combination of relaxation experiments and MD consists in back-calculating relaxation rates from an MD simulation and interpreting the dynamics from the simulation if the agreement is good.[16–18] However, it is customary to compare properties derived from the relaxation rates instead (e.g., order parameter S^2 and correlation time τ), as the latter are more easily interpretable.[19–23] In case the agreement between experiment and simulation is imperfect, Salvi *et al.*[24] have proposed to reweight the MD trajectories to best match the experimental data. We note that this reweighing procedure relies on the correspondence of at least a fraction of the simulation data and the experiments and may be prone to over-fitting. To reduce over-fitting, Kümmerer *et al.*[25] introduced an additional entropy term to the reweighting procedure. Alternatively, it is possible to generate conformational ensembles in accordance with specific NMR observables by biasing the simulation itself, although these procedures bias the resulting dynamics and have, therefore, not been applied in the context of NMR relaxation.[26–28]

The accuracy of the properties predicted from a simulation heavily depends on the quality of the underlying classical force field as well as sufficient sampling. Traditionally, force fields for biomolecular simulations have been parameterized from quantum-mechanical (QM) and/or a wide range of experimental properties (see Refs. 29–31 for recent reviews). In particular, the choice of the backbone and side chain torsion parameters is paramount to producing correct structural ensembles.[32] Torsion parameters are derived from either *ab initio* data,[33,34] experimental data,[35,36] or a mixture of both.[37] In addition to serving as training data, NMR observables such as chemical shifts, J couplings, residual dipolar couplings, and relaxation data have been used to validate force fields *a posteriori*.[38–40] Thereby, both the ability of the force field to reproduce structural properties and the kinetics of inter-conversion can be probed.[41] NMR observables are thus expected to play a growing role in force-field parameterization and validation in coming years.[12] Such validation studies may also expose specific parts of a force field in need of revision. For example, Hoffmann *et al.*[18,42] have recently improved the kinetics of methyl-group rotation in aliphatic side chains by comparison of simulation results with NMR relaxation. Their refitting procedure was based on high-level CCSD(T) torsion profiles and was subsequently refined by Kümmerer *et al.*[43] based on a grid search around the QM derived parameters. The circumstance that refitting was beneficial suggests that *ab initio* data alone are not sufficient to reproduce experimental kinetic rates. However, as the force-field terms are interdependent, the identification of the particular term leading to a discrepancy with experiment is not always straightforward.

In addition to dihedral angles, pairwise non-bonded interactions play a significant role in modeling accurate kinetics and secondary structure propensities of peptides and proteins.[32,44,45] Similarly, interactions at the protein–water interface strongly impact the compactness of the structures produced, in particular for intrinsically disordered proteins.[46] In this light, the choice of water model in a force field is crucial to generate ensembles conforming to experiment. For example, recent adjustments to the Lennard-Jones parameters of water have improved secondary structure propensities (shown independently for both CHARMM36m/TIP3P and AMBER ff99SB-*disp*/TIP4P-D combinations).[37,47] In the context of comparison with NMR relaxation, the water model was shown to strongly influence the tumbling rate of globular proteins in solution.[18,48,49] To investigate the effect that the choice of force field and water model may have on the simulated motion, we perform our analyses in this work for a set of force fields and water models using the motion of ubiquitin on the ps to ns timescale as a test system.

Ubiquitin is a small (76 residues) and relatively rigid protein.[50] Its main biological function is to serve as a biological marker when attached to other substrate proteins to regulate cellular processes, such as protein degradation or activation.[51,52] We chose ubiquitin as a test system in this study because high-quality experimental relaxation data are available in the literature for this protein,[11,13,14,53–56] and different analyses have already been reported.[11,57,58] The aim behind our work is to use this well understood system to compare various state-of-the-art analysis methods and assess their ability to interpret NMR and MD data in tandem. We will focus thereby mostly on the motion of methyl-bearing side chains given the rigidity of the backbone.

II. THEORY

In order to characterize the nature of the motion of a bond-vector $\boldsymbol{\mu}$ (e.g., N–H of the protein backbone), NMR relaxation experiments probe the spectral density function $J(\omega)$ at specific Larmor frequencies ω of the spins associated with the interaction.[59] This spectral density function is the Fourier transform of the time autocorrelation function (or simply "correlation function") and can be extracted from an MD simulation using

$$C(t) = \langle P_2(\boldsymbol{\mu}(\tau) \cdot \boldsymbol{\mu}(t+\tau)) \rangle_\tau, \tag{1}$$

where $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial and $\boldsymbol{\mu}$ is the normalized bond vector of interest. Different methods have been proposed in the literature to compare the calculated and experimental correlation functions (for reviews, see Refs. 12, 60).

A. Model-free approach

The model-free approach pioneered by Lipari and Szabo[61] in 1982 describes the overall motion as a product of correlation functions for overall tumbling and internal motion. Both are assumed to be independent of one another and are fitted with decaying exponential functions. The correlation function for internal motion is written as

$$C_{int}(t) = S^2 + (1 - S^2) \exp(-t/\tau), \tag{2}$$

where S^2 is the generalized order parameter describing the extent of motion present and τ is the effective correlation time. To treat cases in which the relaxation behavior cannot be reproduced with the model-free approach, Clore *et al.*[62] extended the model of internal motion to a bi-exponential function, separating the motion on the fast timescale (subscript f, often ps) and the slow timescale (subscript s, often ns),

$$C_{int}(t) = S_f^2 S_s^2 + \left(1 - S_f^2\right) \exp\left(\frac{-t}{\tau_f}\right) + S_f^2\left(1 - S_s^2\right) \exp\left(\frac{-t}{\tau_s}\right). \quad (3)$$

For MD data, the overall tumbling of the protein can be removed by aligning the protein conformation of each frame to the first frame of the trajectory (typically considering only backbone atoms for the alignment). The correlation time is then obtained using Eq. (1) and fitted to a function such as Eq. (3). The resulting parameters of the correlation functions can be compared to those obtained from the fit of the experimental relaxation data.

B. Isotropic reorientational eigenmode dynamics (iRED)

Assessing internal dynamics may also be performed without alignment of the trajectory by using a method such as isotropic reorientational eigenmode dynamics (iRED).[21,63] In iRED, order parameters are calculated from an eigenvalue decomposition [Eq. (4)] of the covariance matrix $\boldsymbol{M_{ij}}$ (with eigenvalues λ and eigenvectors $\boldsymbol{m}$),

$$S_i^2 = 1 - \sum_{m=6}^{N} \lambda_m |\langle \boldsymbol{m} \rangle_i|^2, \quad (4)$$

obtained from averaging over simulation blocks of chosen length T_{iRED},

$$\boldsymbol{M_{ij}} = \langle P_2(\boldsymbol{\mu_i} \cdot \boldsymbol{\mu_j}) \rangle_{T_{\mathrm{iRED}}}, \quad (5)$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are two different bond vectors originating from the same frame of the trajectory.

An advantage of iRED is that it makes no assumption of the existence of a possible molecular reference frame in which external and internal motion may be separable. However, the method only provides a characterization of the extent of motion through an order parameter S_i^2 and does not inform precisely on the associated timescales, i.e., one can only infer that the motion occurs on a timescale smaller than the block size T_{iRED}. Performing the analysis multiple times with different block sizes to obtain the evolution of the order parameters as a function of T_{iRED} may provide an estimate of the timescales involved.[14]

C. Dynamic detector approach

The model-free approach and iRED assume a separation into two or more distinct motions, each of which can be explained with a single exponential decay. To avoid such assumptions, Smith *et al.*[23,64,65] developed the dynamic detector approach, in which the experimental measurements are fitted with multiple detector windows that are optimized to capture most of the motion observed in the NMR experiment. The correlation function of a detector is the integral of the detector's sensitivity θ over all correlation times,

$$C(t) = S^2 + (1 - S^2) \int_0^\infty \theta(\tau_c) \exp(-t/\tau_c) d\tau_c. \quad (6)$$

D. Re-orientational dynamics in MD analyzed for NMR correlation function disentanglement (ROMANCE)

All previously introduced methods suffer from assumptions of the underlying motions being averaged into a certain number of fitting parameters, for example an order parameter and a correlation time, which can be an average over many different motions. An alternative to the previous methods is to use MD simulations to decompose the different motions. In the approach called re-orientational dynamics in MD analyzed for NMR correlation function disentanglement (ROMANCE),[66] the overall, lab-frame recorded motion is separated into system-specific motions in frames aligned to parts of the protein. For example, the methyl rotation of an isoleucine residue is described in a frame aligned to the $C_\gamma - C_\delta$ bond of this residue. To retrieve the total correlation time of the system (C_{Total}, measured in the lab frame), the total rotation of the tensor has to be broken into several steps, defined by the series of reference frames. At each step, a correlation function for that rotation is calculated such that the product of the correlation functions for the individual rotations returns the total correlation function to good approximation,

$$C_{\mathrm{Total}}(t) = C_{\mathrm{Methyl}}(t) \times C_{\chi_2}(t) \times C_{\chi_1}(t) \times C_{\mathrm{Backbone}}(t). \quad (7)$$

In practice, the correlation function in each frame is treated with different motion models to avoid artifacts such as librational motion. The retrieved correlation functions can also be analyzed separately, providing deeper insights into the composition of the average correlation times.

E. Markov state modeling (MSM)

Markov state models (MSMs) have emerged as a powerful technique to describe the dynamics of molecular systems seen in MD simulations.[67–74] In short, MSMs involve the discretization of an MD trajectory (or a set of trajectories) into n microstates, and constructing from it a transition matrix [$\boldsymbol{T_{ij}}$ in Eq. (8)], which describes the evolution of these microstates over time. This is performed by summing all transitions between every pair of states i and j into a counts matrix $\boldsymbol{c_{ij}}$. The frames in the input trajectory are separated by a lag time τ_{lag},

$$\boldsymbol{T_{ij}}(\tau_{lag}) = \frac{\boldsymbol{c_{ij}}(\tau_{lag})}{\sum_{k=1}^{n} \boldsymbol{c_{ik}}(\tau_{lag})}. \quad (8)$$

The microstates are then grouped into metastable sets such that the slowest processes in the system occur between the metastable sets. One significant advantage of MSMs is that the slowest kinetic

information may be recovered as long as a local equilibrium between all states is reached. Kinetic information is typically extracted through implied timescales [ITS, Eq. (9)], which depend on the eigenvalues λ_i of the transition matrix and may not be easily interpretable,

$$t_i = -\frac{\tau_{\text{lag}}}{\ln(\lambda_i)}. \tag{9}$$

Alternatively, it is possible to extract kinetic information related to two specific metastable sets i and j through mean first-passage times [MFPTs, Eq. (10)], which describe the average time to visit state j for the first time if we are currently in state i [by going directly to that state or passing through any other intermediate state(s) k],

$$\mathbb{E}_i[\boldsymbol{T}_j] = \begin{cases} 0, & i = j, \\ 1 + \sum_{k \neq j} \boldsymbol{T}_{i,k} \mathbb{E}_k[\boldsymbol{T}_j], & i \neq j. \end{cases} \tag{10}$$

This recursive formula is solved simultaneously for all MFPTs of the model.

III. METHODS

A. Simulation details

To evaluate the effect of the force field (in particular the torsion parameters) on the comparison with the experimental NMR data, force fields from the AMBER and CHARMM family were selected. In both cases, force-field variants have been published, which were optimized against NMR-derived data: AMBER ff99SB-NMR1-ILDN[36] and CHARMM36m.[37] We also tested other force fields, such as AMBER ff99SB*-ILDN[35] and AMBER ff99SB*-ILDN-q,[45] for which the results are provided in the supplementary material. For comparison, the baseline AMBER force field ff99SB-ILDN[34] was also included. In general, the default water model TIP3P[75] was used, with the exception of C36m-stn-3P where a slightly adjusted TIP3P model was employed.[37] For AMBER ff99SB-NMR1-ILDN, we tested two additional water models (TIP4P[75] and TIP5P[76]) to study solvent effects. Furthermore, AMBER ff99SB-NMR1-ILDN was also simulated with and without the methyl refitting proposed by Hoffmann *et al.*[18,42] The force-field combinations are listed in Table I, in which the short names are used.

The simulations of ubiquitin in water were performed using the GROMACS package,[77,78] version 2021 – *rc*1, compiled with CUDA support and run on NVIDIA GeForce RTX 2080 Ti graphic cards. The system was prepared with GROMACS tools from the crystal structure 1UBQ[79] in the PDB database. The system was parameterized with the given force field and solvated with the chosen water model in a ~6 nm · 6 nm · 6 nm cubic box, followed by an energy minimization. After a short 100 ps NVT equilibration and 100 ps NPT equilibration, the production runs were carried out. The simulations were performed under periodic boundary conditions for 4 μs with a time step of 2 fs, and the output was written every 5000 steps. The bonds involving hydrogen atoms were constrained with the LINCS algorithm[80] with lincs_iter = 1 and lincs_order = 4. The cutoff scheme for the nonbonded interactions was Verlet with a cutoff of 1.0 nm (for all AMBER force fields) and 1.2 nm (for C36m-stn-3P). The Coulomb interactions were calculated using the particle mesh Ewald (PME)[81] algorithm with a Fourier spacing of 0.12 nm. The temperature was controlled using the V-rescale thermostat[82] with a coupling time of 0.1 ps and a reference temperature of 300 K. The pressure was controlled using the Parrinello–Rahman barostat[83] with a coupling time of 2.0 ps, a reference pressure of 1.0 bar, and an isothermal compressibility of 4.5 10^{-5} bar^{-1}. The center of mass motion was removed with the default GROMACS procedure every 100th step.

B. Analysis

For analysis, the protein was centered in each trajectory with the GROMACS command `gmx trjconv -pbc mol -center` and aligned to the first frame with `gmx trjconv -fit rot+trans`.

1. Experimental data

Ubiquitin backbone relaxation data[56] were analyzed using MINOTAUR.[58] For each residue, longitudinal and transverse ^{15}N auto-relaxation rates and ^{15}N–^{1}H dipole–dipole cross-relaxation rates recorded at 14.1, 18.8, and 22.3 T were analyzed together with the high-resolution relaxometry intensities from decays at seven low magnetic fields ranging from 5 to 0.5 T. We used the extended model-free[62] form of the spectral density function as it better reproduced the experimental data compared to the model-free approach.[19] To account for the significant exchange contribution to

TABLE I. Investigated force-field combinations (force field, methyl refitting, and water model) with short names for the simulation of ubiquitin.

Short name	Force field	Methyl refitting[18,42]	Water model
Anmr-stn-3P	AMBER ff99SB-NMR1-ILDN[36]	No	TIP3P
Anmr-met-3P	AMBER ff99SB-NMR1-ILDN[36]	Yes	TIP3P
Anmr-stn-4P	AMBER ff99SB-NMR1-ILDN[36]	No	TIP4P
Anmr-met-4P	AMBER ff99SB-NMR1-ILDN[36]	Yes	TIP4P
Anmr-stn-5P	AMBER ff99SB-NMR1-ILDN[36]	No	TIP5P
Anmr-met-5P	AMBER ff99SB-NMR1-ILDN[36]	Yes	TIP5P
Adef-stn-3P	AMBER ff99SB-ILDN[34]	No	TIP3P
C36m-stn-3P	CHARMM36m[37]	No	TIP3P

the transverse relaxation rates of residues 23 and 25, an additional free parameter α was included in the Markov chain Monte Carlo (MCMC) approach and is defined as $R_2'(B_0) = \alpha B_0^2 + R_2(B_0)$, where B_0 is the static magnetic field, $R_2(B_0)$ is the exchange free expression of the ^{15}N transverse relaxation rate, and R_2' is the expression used to analyze the transverse relaxation rates of residues 23 and 25. The nitrogen chemical shift anisotropy (CSA) was assumed to be axially symmetric with a value of −164 ppm and forming an 18° angle with the N–H bond.[84] The proton CSA was considered fully anisotropic, with σ_{xx} = 14.6 ppm, σ_{yy} = 8.2 ppm, and σ_{zz} = 2.1 ppm, with the x-axis of the CSA tensor being orthogonal to the N–H bond and the y-axis forming a 99° angle with the N–H bond. The N–H bond distance was set to 0.102 nm. The contribution of neighboring protons was considered in the form of two additional protons in the spin system, 0.21 nm away from the amide proton. Only their contribution to the ^{1}H was included. Their CSA was assumed to be axially symmetric, in the alignment with the amide proton and with the value $\frac{1}{2}(\sigma_{xx} + \sigma_{yy} - 2\sigma_{zz})$.[85]

2. iRED

The iRED[21] method was used to extract order parameters describing the backbone motion (N–H) in ubiquitin using the Python package pyDR.[86] We followed recommendations from Gu *et al.*[63] and divided the full trajectory into four blocks of 1 μs each for analysis with iRED. The order parameters reported correspond to the average over the four blocks. The simulations were compared to experimental data acquired by Charlier *et al.*[56] (relaxometry) and fitted with MINOTAUR as described above, as well as Wardenfelt *et al.*[13] [nanoparticle assisted spin relaxation (NASR)].

3. Model-free approach

Motion in the side chains of ubiquitin was probed with the model-free approach. Correlation functions of the internal motion [Eq. (1)] were extracted and fitted to an extended model-free

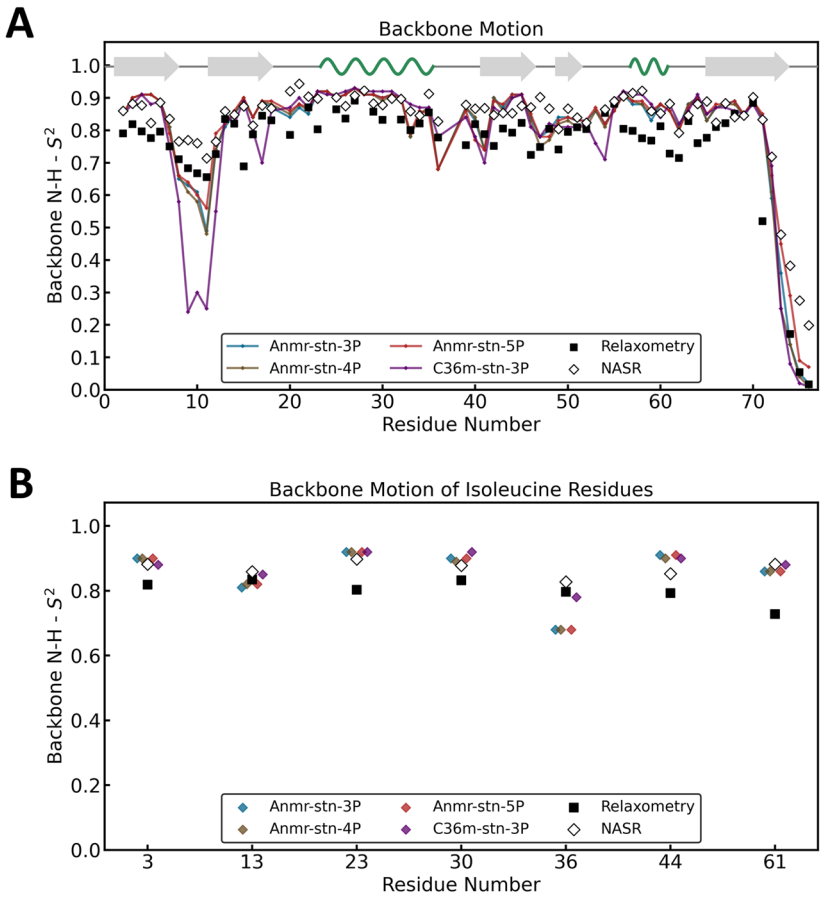

FIG. 1. Comparison of the backbone N–H motion of ubiquitin in MD simulations (colored lines, see Table I) using iRED with two experimental methods: relaxometry[56] (black squares) and NASR[13] (white diamonds). Note that the relaxometry data have been re-analyzed with MINOTAUR.[92] (a) All residues. (b) Isoleucine residues. The results with other force-field combinations are shown in Fig. S2 of the supplementary material.

TABLE II. Average absolute deviations of the order parameters (ΔS^2) between MD simulations and experiment for the backbone N–H motion of ubiquitin.

	ΔS^2 All residues		ΔS^2 All except β_1–β_2 turn		$\Delta S^2 \beta_1$–β_2 Turn	
Force field	Relaxometry	NASR	Relaxometry	NASR	Relaxometry	NASR
Anmr-stn-3P	0.07	0.05	0.07	0.05	0.06	0.14
Anmr-stn-4P	0.07	0.05	0.07	0.05	0.07	0.15
Anmr-stn-5P	0.07	0.05	0.07	0.04	0.06	0.12
C36m-stn-3P	0.09	0.07	0.07	0.05	0.26	0.34

bi-exponential function [Eq. (3)] employing the Python libraries *emcee*[87] and *scipy*.[88] We followed the same procedure as in Ref. 11, with the exception that we did not remove the contributions stemming from fast rotation around the methyl symmetry axis. The fits were performed on each of the four 1 μs blocks, from which we calculated the average value and standard deviation. Order parameters and correlation times for the isoleucine side chains were compared to the experimental relaxometry data from Cousin *et al.*[11] In addition, the same workflow was applied to all other methyl-bearing side chains in ubiquitin (alanine, leucine, valine, and threonine), allowing us to compare them with those acquired experimentally (NASR) and fitted using iRED by Xiang *et al.*[14]

4. Dynamic detector approach

The dynamic detector analysis was performed with six detector windows based on the experimental relaxation data from Charlier *et al.*,[1,56,57] which was pre-processed in Smith *et al.*[57] using the Python package pyDR.[86] The values of both analysis windows were compared to the other methods.

5. ROMANCE approach

The ROMANCE analysis was performed with the Python package pyDR[86] on the full 4 μs trajectory for all isoleucine residues of ubiquitin. The separated motions were each fitted to a single exponential decay, and the correlation times were compared to the other methods.

6. MSM construction

MSMs[70] were constructed in the rotamer space (χ_1 and χ_2 dihedral angles) of each isoleucine residue of ubiquitin using the Python library pyEmma.[89] Dihedral angles were extracted with the MDAnalysis[90] package and served as an input for the structural clustering to discretize the trajectory, which was performed with the energy-based clustering (EBC)[91] algorithm. In this study, the MSMs were constructed from a single long simulation (4 μs) rather than multiple short simulations. A detailed description of the analysis parameters (e.g., lag times) is provided in Table S1 of the supplementary material. Example Jupyter notebooks containing Python code used to generate and plot the data discussed in this study can be found at https://github.com/rinikerlab/MD_and_NMR_Relaxometry.

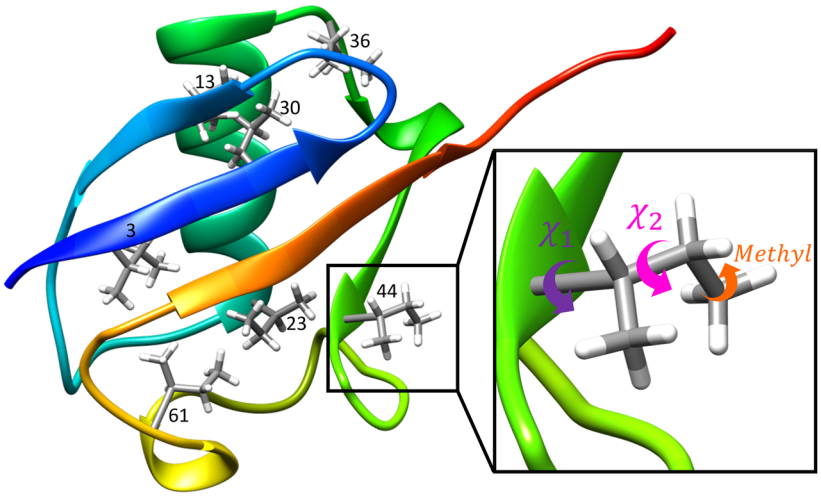

FIG. 2. Visualization of the structure of ubiquitin (PDB code: 1UBQ). All isoleucines are shown as stick representation. The zoom-in on Ile 44 shows the different side chain rotations observed. Note the highly flexible tail of ubiquitin (shown in red) and the flexible loop between residues 7 and 12 (shown in dark blue).

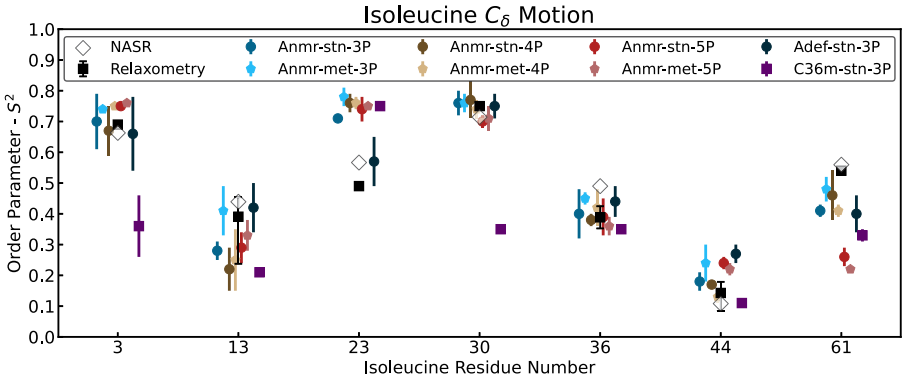

FIG. 3. Comparison of the motion of the seven isoleucine side chains (Ile 3, 13, 23, 30, 36, 44, and 61) of ubiquitin. Data points from the MD simulations (colored points, see Table I) as well as relaxometry data[11] (black squares) of Ile 13, 36, and 44 correspond to the product of S_f^2 and S_s^2 order parameters to describe the full extent of motion. Meanwhile, relaxometry data[11] (black squares) of Ile 3, 23, 30, and 61 as well as the NASR data[14] (white diamonds) correspond to a single overall order parameter S_i obtained via a mono-exponential model-free analysis.

IV. RESULTS AND DISCUSSION

We performed 4 μs simulations of ubiquitin with different force-field combinations (Table I). The protein fold remained stable over the course of the simulation for all setups (Fig. S1 of the supplementary material). In the following, the motions in the simulations on the ps and ns timescales were analyzed with different methods, comparing the results with each other and with experimental NMR relaxation data.

A. Backbone motion

First, we analyzed the backbone motion of ubiquitin by extracting order parameters for the backbone N–H bonds and comparing the computed values with those from two different experimental sources (Fig. 1).

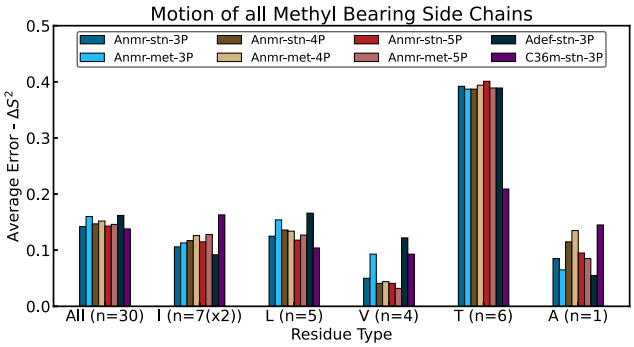

FIG. 4. Comparison of the motion of the methyl-bearing side chains [isoleucine (I), leucine (L), valine (V), threonine (T), and alanine (A)] of ubiquitin with experimental NASR data.[14] The reported values correspond to the average absolute deviations in order parameters (ΔS^2) for each force-field combination (colored bars, see Table I) grouped by amino acid type. The number of data points per amino acid type is given in the label.

The relaxometry and NASR results show similar trends, but interestingly, the order parameters from relaxometry are consistently lower (i.e., more mobile) than those from NASR. The original relaxometry decays were reanalyzed with MINOTAUR, a software application we have recently developed.[92] We cannot exclude experimental artifacts in nitrogen-15 relaxometry experiments possibly due to transient dimerization (the sample concentration was 3 mM). We do not observe significant differences between the AMBER force fields and the different water models. Note that this does not imply that there are generally no differences among the force fields but that the properties of ubiquitin analyzed in this section are not sensitive to them (e.g., other properties may differ, such as the tumbling time, which is known to be sensitive to the water model[48]).

The simulation with the C36m-stn-3P force field (purple line in Fig. 1) displays overall a similar behavior compared to the AMBER force fields, except for a more flexible β_1–β_2 turn between residues 7 and 12. The higher flexibility in the loop region may be because CHARMM36m was purposely fitted to describe intrinsically disordered proteins (as well as folded proteins),[37] whereas AMBER force fields have been found in the past to over-stabilize certain secondary structure elements.[45] The latter issue was addressed in newer versions of this force field.[47] When comparing the computed values to both sources of experimental data, it is not evident which force field is more accurate. For the secondary structures of ubiquitin (α-helices and β-strands), the order parameters from AMBER and CHARMM are similarly high and consistently closer to the NASR measurements (see also Table II). However, for the β_1–β_2 turn (residues 7–12), CHARMM36m is too flexible, while the AMBER simulations reproduce the relaxometry values more closely. These observations highlight how force-field refinement from experimental NMR relaxation data may prove challenging when different experimental sources do not fully agree with one another. To obtain more concrete conclusions about the ability of the available force fields to reproduce such data, a similar comparison on a larger set of diverse (and more flexible) proteins will be needed.

B. Side chain motion

The methyl-bearing side chains of ubiquitin display a wider range of flexibility than the backbone. Our analysis was focused mainly on the isoleucine residues (Fig. 2), which we compared to two independent sets of experimental relaxation data found in the literature (relaxometry[11] and NASR[14]). In this case, the order parameters measured with both experimental methods agree well with one another (Fig. 3).

Overall, we find good agreement between the experimental order parameters and those extracted from the MD simulations, in particular for the residues Ile 13, 36, and 44, for which the ns motion is well defined in the relaxation experiments of Cousin *et al.*[11] We note that error bars around the simulated values can be quite large, indicating that the magnitude of motion varies in the four 1 μs blocks of simulation. This is in line with the findings of Bowman,[93] suggesting that several μs of simulation are required to describe motion

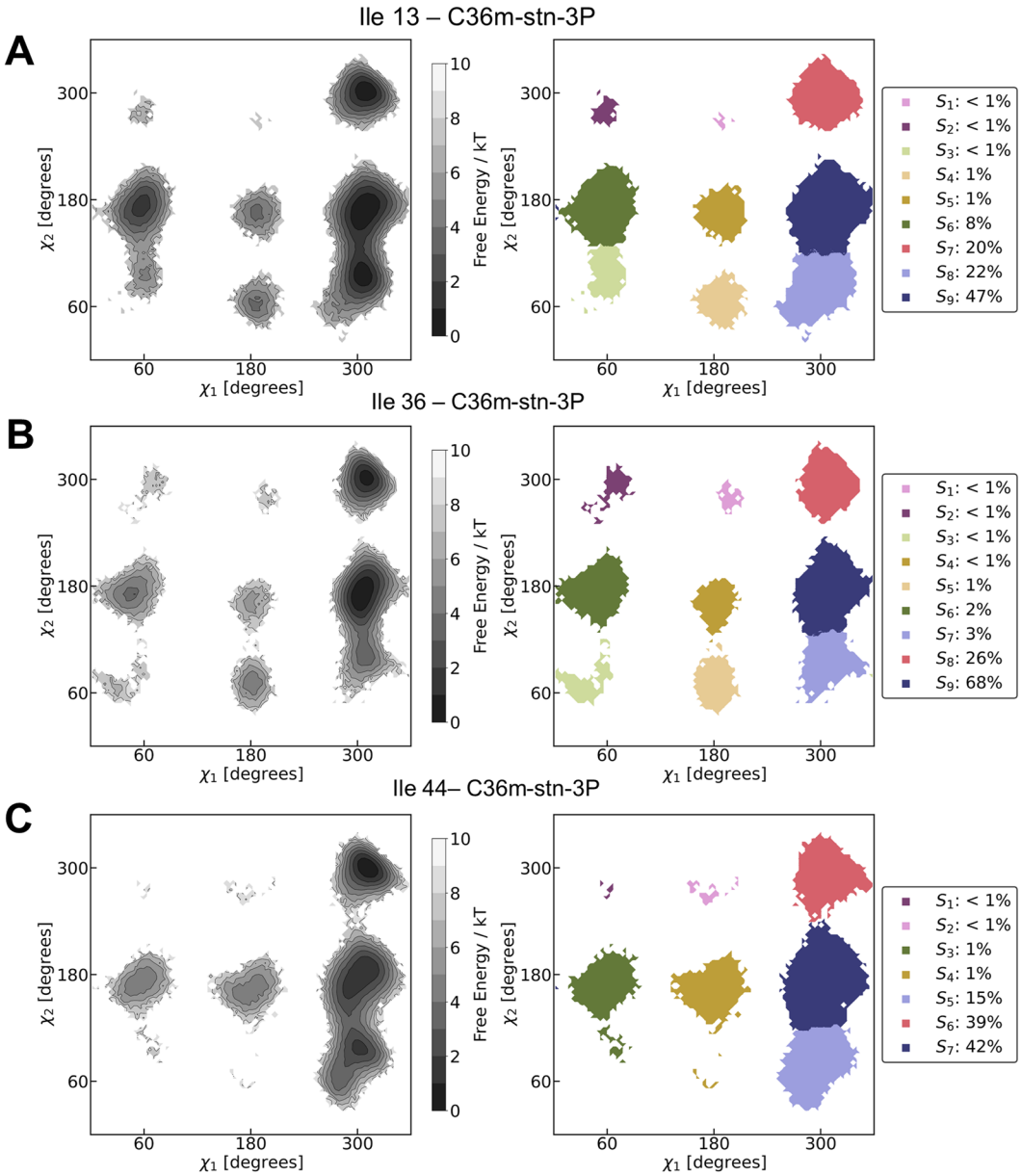

FIG. 5. Rotamer space for Ile 13 (a), 36 (b), and 44 (c). (Left): free energy of the rotamer space shown with a color gradient. (Right): metastable sets (clusters) of the MSM represented in different colors.

on the ns timescale in a converged manner. Rotamer jumps of the isoleucine side chains are rare events on the ns time scale and can be hindered by the protein surroundings. We do not observe any significant effect from the water model. In particular, the solvent exposed residue Ile 44 shows similar dynamics with all three water models studied. For Ile 61, a higher flexibility is seen with TIP5P, but as this residue is buried in the protein interior, the difference is likely incidental due to finite sampling.

The modification of the rotation barriers of the terminal methyl groups (i.e., methyl refitting[18,42]) appears to have little impact on the order parameters (compare lighter and darker versions of each pair of colored points). This suggests that differences in order parameters are driven mainly by differences in the χ_1 and χ_2 dihedral-angle parameters, as seen between C36m-stn-3P and Anmr-stn-(3P/4P/5P) for residues Ile 3 and 30 [and to a lesser degree between Anmr-stn-(3P/4P/5P) and Adef-stn-3P for Ile 23], in agreement with previous work.[58] However, the good agreement observed among all tested force fields for the other residues Ile 13, 36, and 44 indicates that sterics also play a significant role in the observed dynamics (i.e., protein surrounding may hinder rotamer jumps). Altogether, the findings highlight how challenging it is to trace back disagreement with experimental relaxation data to specific force-field terms and parameters.

Finally, we also compared the differences in the overall order parameter ΔS^2 for all methyl-bearing residues between simulation and experimental NASR data (Fig. 4). We find that, on average, the amplitude of reorientational motion of terminal methyl groups is equally well represented by all force fields tested. Only for threonine side chains, there is a substantially larger deviation for the AMBER force fields compared to C36m-stn-3P, suggesting that the χ_1 dihedral-angle parameters of AMBER ff99SB-ILDN may benefit from a refitting, e.g., by following similar procedures as those described by Hoffmann *et al.*[18,42] or Kümmerer *et al.*[43]

The (extended) model-free approach and iRED provide valuable metrics to determine how well the MD simulations match the experimental data. However, these methods do not provide an explanation of the motion at atomic resolution (e.g., rearrangement of a loop and rotation around a specific bond). By constructing MSMs in the χ_1/χ_2 rotamer space of each isoleucine residue, we obtain a more interpretable view of the motion among the different rotameric states. The MSMs of Ile 13, 36, and 44 with the C36m-stn-3P force field are shown in Fig. 5. The corresponding mean first passage times (MFPTs) and implied timescales (ITSs) from the MSMs are compared with the experimental data (using the model-free approach) and ROMANCE in Fig. 6. While implied timescales may be better suited for a comparison with correlation times (given that they describe an exponentially decaying process, see Sec. S2 of the supplementary material), their corresponding eigenvectors may not be easily mapped onto a specific motion (e.g., rotamer jump between two distinct conformers). Therefore, we prefer to present our comparison with MFPTs, which provide a more clearly resolved picture of the dynamics that occur. The MSMs and MFPTs of the other isoleucine residues are provided in Figs. S3–S52 of the supplementary material. Overall, we find remarkable agreement between the MFPTs connecting the most populated macrostates to the experimental model-free correlation times (Fig. 6). Values

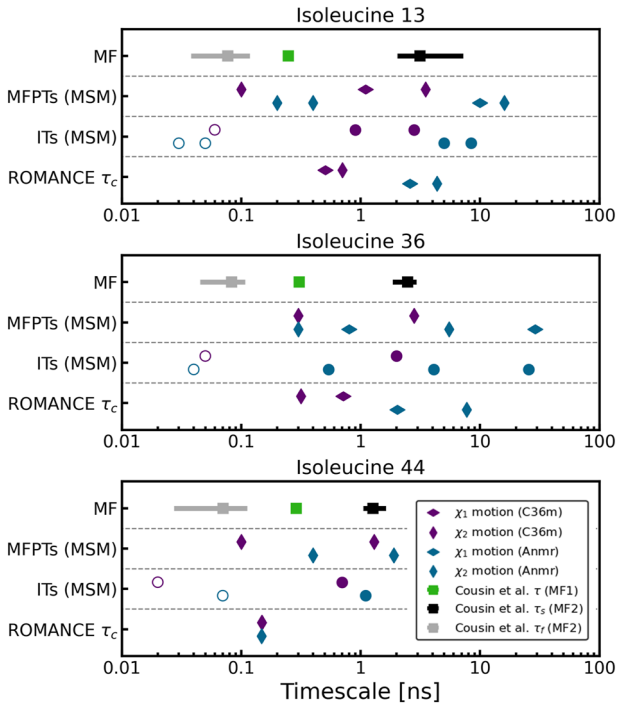

FIG. 6. Comparison of the experimental and computed side chain timescales of motion for Ile 13, 36, and 44 in ubiquitin. For these three residues, the ns motion was well characterized by relaxometry experiments.[11] The mean first passage times (MFPTs) for the transitions connecting the most populated states and correlation times obtained with ROMANCE are shown for the MD simulations with Anmr-stn-3P (blue diamonds) and C36m-stn-3P (purple diamonds). The implied timescales whose amplitude was greater than 0.05 are shown in circles (empty circles if the motion is below the lag time, indicating that the value should be interpreted with caution). The lag times used to construct the MSMs were 0.1, 0.25, and 0.25 ns with AMBER and 0.1, 0.25, and 0.05 ns with CHARMM for Ile 13, 36, and 44, respectively. Experimental correlation times were obtained from mono-exponential (MF1, green squares) or bi-exponential (MF2, gray/black squares) model-free fits.[11]

extracted from the simulations with the C36m-stn-3P force field lie closer to the experimental values and are consistently faster than those with the Anmr-stn-3P force field. Interpretation is, however, complicated by the fact that MFPTs do not describe an exponentially decaying process, and the corresponding implied timescales are typically lower than the MFPTs (by a factor of roughly half). Overall, we find that MFPTs are useful to obtain an idea of the general timescale at which specific rotamer jumps occur, without necessarily providing a precise quantitative value. We also note that the MFPTs for the fastest motion can be overestimated as they cannot be lower than the lag time used to construct the MSM [e.g., Ile 44 where the lowest MFPTs are close to the lag time (0.05 ns for CHARMM and 0.25 ns for AMBER) in Fig. 6]. MSMs are typically used to inform on the slowest motion of the system, which can be determined more accurately with a larger lag time, at the cost of describing the faster motion of the system.

For Ile 44, the faster correlation time can be attributed to transitions between the *gauche*$_+$ rotamer (metastable set S_5) and the *trans* rotamer (S_7), whereas the slower correlation time corresponds to the transitions between the *gauche*$_-$ (S_6) and the *trans* (S_7) rotamers [Fig. 5(c)]. Since both of these transitions correspond to a χ_2 rotation, the ROMANCE approach does not separate them and, instead, returns a correlation time closer to the mono-exponential fit (Fig. 6).

In addition, we note that experimental correlation times (e.g., τ_s for Ile 13 and 36) do not necessarily correspond to a single motion and could arise from averaging of multiple transitions occurring on the same timescale. The larger experimental error for τ_s of Ile 13 could thus be attributed to the fact that nearly all the nine rotamer states are populated, leading to many types of motion on the ns timescale. The flexibility of the nearby loop (residues 7–12) may also potentially impact the dynamics, although both the iRED and ROMANCE analyses [Figs. 1(b) and 7] do not suggest more backbone motion for Ile 13 compared to the other isoleucines.

For Ile 23, the presence of a single dominant rotamer (*gauche*$_-$) with rare transitions (slower than the tumbling time) to other rotameric states (Figs. S11 and S36 of the supplementary material) may explain why motion in the low ns timescale was not observed experimentally. However, one has to keep in mind that the simulated order parameters of Ile 23 (Fig. 3) were overestimated (i.e., more rigid) compared to experiment, which suggests that the transitions may be underestimated in the simulations.

While the MSMs provide a distinct separation of the motion in the χ_1/χ_2 rotamer space, a clear advantage of ROMANCE lies

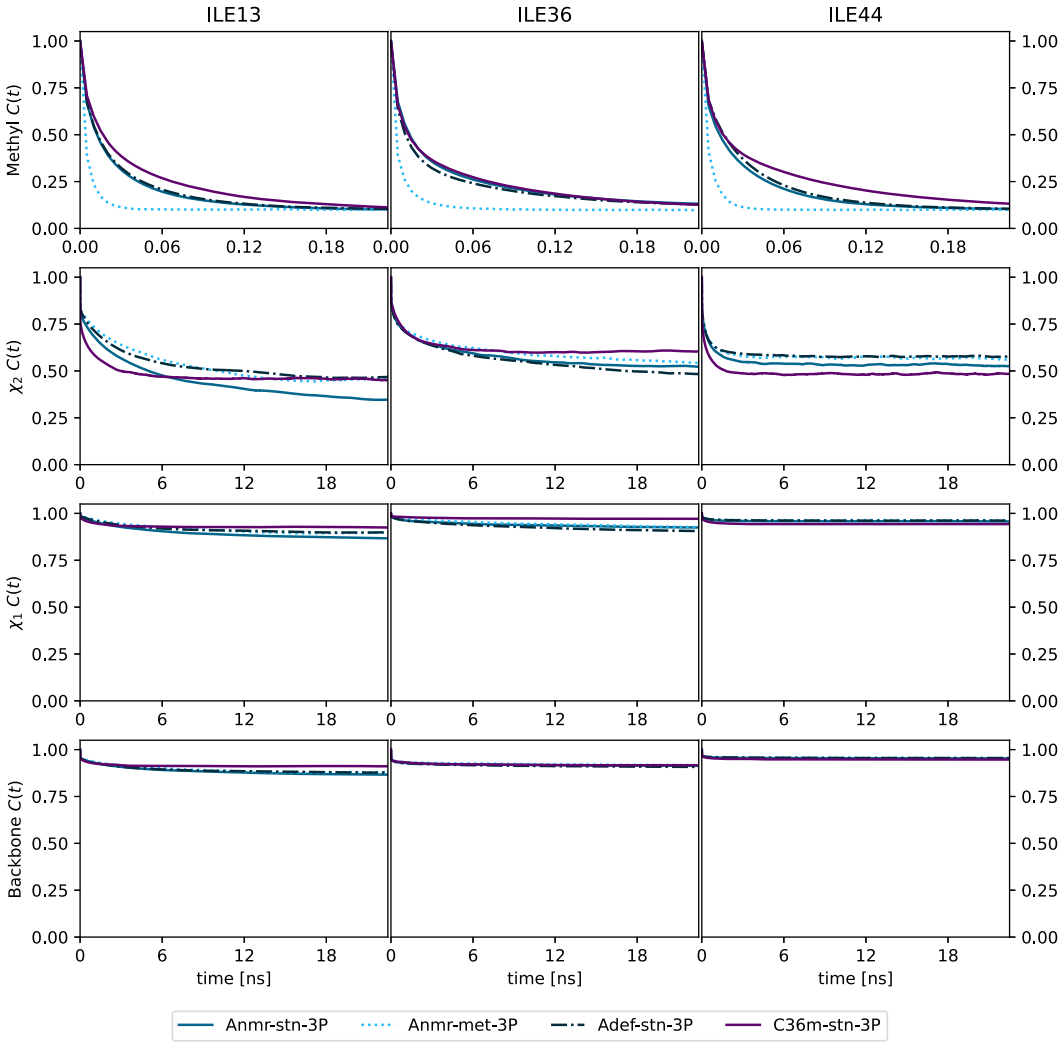

FIG. 7. Time evolution of the correlation functions extracted with ROMANCE for Ile 13, 36, and 44, separated into four different motions (methyl rotation, χ_2-rotation, χ_1-rotation, and backbone motion) for the force fields Anmr-stn-3P, Anmr-met-3P, Adef-stn-3P, and C36m-stn-3P.

in its ability to separate the internal side chain motion from more global secondary-structure rearrangements. For isoleucines, the side chain motion is separated into four different rotation frames: backbone motion, χ_1-rotation, χ_2-rotation, and methyl rotation. The results for the three residues Ile 13, 36, and 44 are shown in Fig. 7 for four selected set-ups. The results for the other force fields are provided in Fig. S53 of the supplementary material. For the backbone motion, similar trends can be observed as in other methods, such as iRED [Fig. 1(b)], where Ile 13 shows slightly more motion since it is close to the flexible loop (residues 7–12), although overall backbone motions are similar to each other. Again, the force-field differences are quite small, with the exception of C36m-stn-3P, which shows less backbone motion for Ile 13 and a faster correlation time. Also for the χ_1 and χ_2 motions on the ns timescale, faster correlation times are observed with C36m-stn-3P, which is in agreement with the MSM analysis shown in Fig. 6. The fitted correlation times and S^2 order parameters from the ROMANCE approach can be found in Tables 26–28 of the supplementary material. However, since there is only a mono-exponential fit for each χ motion and no separation of the different rotamer states, the extracted correlation times give an averaged representation of the χ motion compared to the MSM method and do not exactly match the MFPTs. ROMANCE also includes small amplitude motions around the individual angles χ_1 and χ_2, which are removed in the MSM through the assignment to individual states. These faster motions also bias the mono-exponential fit toward shorter correlation times. Similarly, the differences in the S^2 order parameter of the backbone motion might arise from the different decomposition of the motion between ROMANCE and iRED.

The ROMANCE approach is the only method tested in this study that provides insights into the methyl rotation of the isoleucine residues. For the methyl rotation, we clearly see a difference from the methyl refitting proposed by Hoffmann *et al.*[18,42] As expected, the refitted methyl rotation leads to a faster decay of the correlation function due to the lower energy barrier.

Next, all methods were compared to the dynamic detector approach. The results of an analysis with six detector windows are shown in Fig. 8. The sensitivity of the detectors can be seen in the top panel, while the individual responses to each detector for each isoleucine residue are shown in the bar plots at the bottom. The bar plots show some difference in the correlation-time distributions and intensity of the motions of the isoleucines, but most detectors have a rather small overall response. The main observation is that the force fields have similar responses but, especially for slower motions, diverge from the experimental responses. However, the correlation times analyzed with the other methods (ROMANCE and MFPTs of the MSMs) lie in the same dominant detector window and only minor shifts can be observed between the three isoleucines of interest (Ile 13, 36, and 44).

Finally, we have compared the populations of the different isoleucine rotamers to those predicted based on experimental chemical shifts in Ref. 94 and from explicit models of motion in Ref. 58 (Fig. 9). In general, we find good agreement for Ile 3, 13, 36, and 61 (two most populated rotamers are the same, and differences stem mainly from the minor rotamers). Our results are also in line (identical primary rotamer in χ_1 dimension) with previous reports

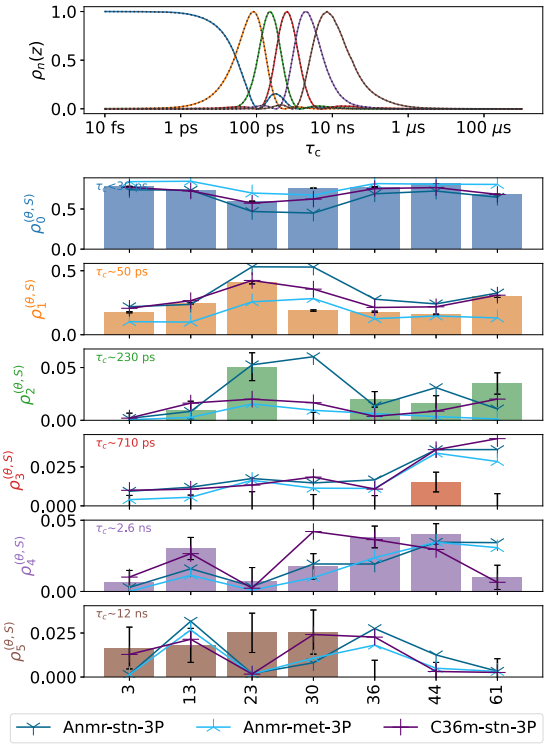

FIG. 8. Dynamic detector approach with six detector windows for all isoleucine residues of ubiquitin for three different force fields in comparison with experimental data. (Top): distribution of the four detector windows. (Bottom): response of the different isoleucine residues to the corresponding detector extracted from experiment (colored bars) and response obtained from the MD simulations (colored lines).

based on J-couplings and residual dipolar couplings, although these experiments did not provide any resolution in the χ_2 dimension.[95,96] For Ile 23, 30, and 44, we observed that the chemical-shifts method predicts rotamer $\{g_p, t\}$ to be significantly populated, in contrast to our simulations and the explicit model[58] (which was constructed based on rotamers sampled in a 1 μs MD simulation with an AMBER force field). It is not readily determinable which experimental method is correct, as only four rotameric states were allowed to be populated in the chemical-shifts model ($\{g_p, t\}$, $\{g_m, t\}$, $\{g_m, g_m\}$, and $\{t, t\}$), which seems inappropriate for the solvent exposed, more freely rotatable Ile 44 for which all three rotamers along the χ_2 axis ($\{g_m, t\}$, $\{g_m, g_m\}$, and $\{g_m, g_p\}$) were populated in our simulations. Insufficient sampling of the Ile 23 side chain rotations also prevents us from conclusively determining which of the two rotamers ($\{g_m, g_p\}$ or $\{g_p, t\}$) is the second most populated after $\{g_m, g_m\}$. Once again, this comparison highlights the challenges when validating MD simulations with NMR experiments. In addition to populations, explicit models of motion also report on the kinetics of transitions. However, the rotamer jumps included in the explicit model sometimes differ from the MSMs described above. For example, the third

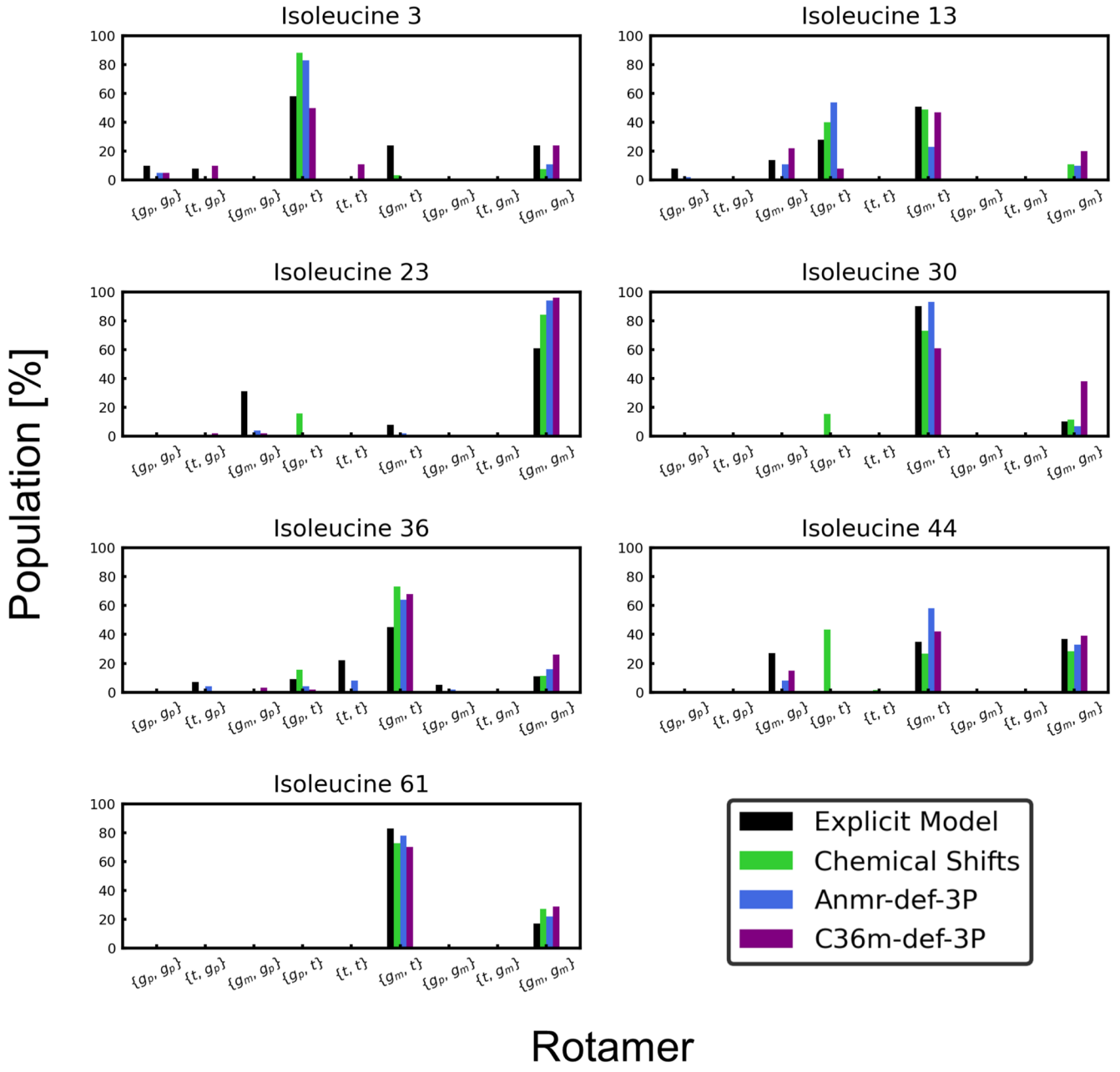

FIG. 9. Populations of the nine isoleucine rotameric states extracted from the MD simulations with two different force fields (blue and purple), experimental chemical shifts (green, Ref. 94), and explicit models of motion of NMR relaxation (black, Ref. 58). Each rotamer corresponds to a set of $\{\chi_1, \chi_2\}$ dihedral angles, labeled *gauche*$_+$ (g_p), *trans* (t), and *gauche*$_-$ (g_m), corresponding to dihedral angles of 60, 180, and 300°, respectively.

most populated rotamer of Ile 13 ($\{g_m, g_m\}$) is absent from the explicit model, although it corresponds to the slower ns motion in the MSM. For this reason, the comparison of rates of transitions between MSMs and explicit models is not straightforward (see Fig. S54 of the supplementary material).

V. CONCLUSIONS

In this study, we have compared eight sets of 4 μs MD simulations of ubiquitin (different force fields and water models) with the results from backbone and side chain NMR relaxation experiments. The analysis was performed with multiple methods, i.e.,

model-free approach, iRED, MSMs, ROMANCE, and the dynamic detector approach, to determine their relative strengths, differences, and the areas of applicability when relating data extracted from MD simulations to NMR relaxation experiments. Despite few recent examples,[58,97] MSMs have not yet been extensively applied to interpret NMR relaxation data. Based on our work and recent trends in the literature, we envision that MSMs derived from MD simulations will become a relevant tool to provide mechanistic insights into NMR relaxation.

Given the rigidity of the backbone of ubiquitin, we focused our analysis mainly on the methyl-bearing side chains, in particular the isoleucine residues. While we found an overall good agreement between the simulations and experiments, the use of MSMs and ROMANCE enhanced the interpretation of relaxation data by allowing us to precisely assign specific rotamer transitions, leading to the NMR relaxation data. The ROMANCE correlation times extracted were, in general, shorter (resulting from motions not captured by the MSMs) than the MFPTs extracted from the MSMs and could not distinguish when two motions occurred within the same degree of freedom (e.g. for Ile 44). Interestingly, for Ile 44, we found that the ROMANCE correlation times matched the mono-exponential fit to experimental data, whereas MSMs could decouple the two individual transitions and matched the bi-exponential fit of experimental relaxation rates well. We found remarkable agreement between the MFPTs corresponding to the most frequent rotamer transitions of Ile 13, 36, and 44, for which ps and ns motion had been unambiguously determined experimentally,[11] although the comparison remains qualitative because MFPTs do not describe an exponentially decaying process. To further distinguish the advantages and disadvantages of the different analysis methods, a similar comparison is needed in the future with more flexible proteins than ubiquitin. For example, the dynamic detector approach may be better suited to analyze a protein system in which motion occurs on more separated timescales.

Overall, we found the AMBER force fields tested to perform very similarly. Most surprisingly, the choice of the water model did not seem to significantly influence the motion of the backbone or side chains of ubiquitin. Meanwhile, faster or more extensive dynamics were observed in the simulations with CHARMM36m, in terms of both backbone motion (loop involving residues 7–12) and side chain motion. As the two experimental techniques (relaxometry and NASR) showed significant differences in the backbone motion, it was not possible to ascertain whether one force field reproduced experiment better than the other. Only one clear difference could be seen for threonine residues, where CHARMM36m gave a much lower error ($\Delta S2$) when compared to experiment, indicating potential issues with the threonine χ_1 dihedral-angle parameters in the AMBER force fields. We note that it is generally non-trivial to determine which force-field parameters lead to particular differences in the dynamics. Using NMR relaxation data as a fitting target to parameterize or refine force fields for biomolecular simulations may thus be challenging. Nevertheless, the comparison with NMR data is excellent for validation. Finally, we note that the convergence of the dynamics in the low ns timescale requires multiple μs of simulation (as previously discussed in Ref. 93), pointing to the extent of simulation data that will be needed when validating the ability of force fields to describe the dynamics of even slower timescales (e.g., μs).

SUPPLEMENTARY MATERIAL

The supplementary material contains additional details, tables, and figures for the analysis of backbone motion, Markov state models, ROMANCE analysis, and comparison with explicit models.

ACKNOWLEDGMENTS

The authors thank Rafael Brüschweiler and Cyril Charlier for sharing experimental data, R. B. for helpful discussions, and Franz Waibl for his review of the manuscript. M. L., C. C., S. R., and F. F. acknowledge the financial support by the European Union through its FET-Open call, Grant agreement 899683 (HIRES-MULTIDYN). A. A. S. acknowledges the financial support from the Deutsche Forschungsgesellschaft (DFG Grant No. 450148812). N. B. C. acknowledges the CIHR for a postdoctoral fellowship.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

C.C. and M.L. are contributed equally to this work.

Candide Champion: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Marc Lehner**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Albert A. Smith**: Data curation (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Validation (supporting); Visualization (supporting); Writing – review & editing (equal). **Fabien Ferrage**: Funding acquisition (supporting); Investigation (supporting); Methodology (supporting); Validation (supporting); Writing – review & editing (equal). **Nicolas Bolik-Coulon**: Data curation (equal); Methodology (equal). **Sereina Riniker**: Conceptualization (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The simulation setup and all analysis methods used for this work are open source and available on GitHub (https://github.com/rinikerlab/MD_and_NMR_Relaxometry).

REFERENCES

[1]C. Charlier, S. F. Cousin, and F. Ferrage, "Protein dynamics from nuclear magnetic relaxation," Chem. Soc. Rev. **45**, 2410–2422 (2016).

[2]M. D. Miller and G. N. Phillips, "Moving beyond static snapshots: Protein dynamics and the protein data bank," J. Biol. Chem. **296**, 100749 (2021).

[3]B. E. Clifton, D. Kozome, and P. Laurino, "Efficient exploration of sequence space by sequence-guided protein engineering and design," Biochemistry **62**, 210–220 (2023).

[4]M. Grimaldo, F. Roosen-Runge, F. Zhang, F. Schreiber, and T. Seydel, "Dynamics of proteins in solution," Q. Rev. Biophys. **52**, e7 (2019).

[5]T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu, "Principles and overview of sampling methods for modeling macromolecular structure and dynamics," PLoS Comput. Biol. **12**, e1004619 (2016).

[6]C. Narayanan, K. Bafna, L. D. Roux, P. K. Agarwal, and N. Doucet, "Applications of NMR and computational methodologies to study protein dynamics," Arch. Biochem. Biophys. **628**, 71–80 (2017).

[7]K. Grohe, S. Patel, C. Hebrank, S. Medina, A. Klein, P. Rovó, S. K. Vasa, H. Singh, B. Vögeli, L. V. Schäfer, and R. Linser, "Protein motional details revealed by complementary structural biology techniques," Structure **28**, 1024–1034.E3 (2020).

[8]A. A. Kawale and B. M. Burmann, "Characterization of backbone dynamics using solution NMR spectroscopy to discern the functional plasticity of structurally analogous proteins," STAR Protoc. **2**, 100919 (2021).

[9]G. Ortega, M. Pons, and O. Millet, "Protein functional dynamics in multiple timescales as studied by NMR spectroscopy," Adv. Protein. Chem. Struct. Biol. **92**, 219–251 (2013).

[10]G. Parigi, E. Ravera, M. Fragai, and C. Luchinat, "Unveiling protein dynamics in solution with field-cycling NMR relaxometry," Prog. Nucl. Magn. Reson. Spectrosc. **124–125**, 85–98 (2021).

[11]S. F. Cousin, P. Kadeřávek, N. Bolik-Coulon, Y. Gu, C. Charlier, L. Carlier, L. Bruschweiler-Li, T. Marquardsen, J.-M. Tyburn, R. Brüschweiler, and F. Ferrage, "Time-resolved protein side-chain motions unraveled by high-resolution relaxometry and molecular dynamics simulations," J. Am. Chem. Soc. **140**, 13456–13465 (2018).

[12]O. Stenström, C. Champion, M. Lehner, G. Bouvignies, S. Riniker, and F. Ferrage, "How does it really move? Recent progress in the investigation of protein nanosecond dynamics by NMR and simulation," Curr. Opin. Struct. Biol. **77**, 102459 (2022).

[13]S. Wardenfelt, X. Xiang, M. Xie, L. Yu, L. Bruschweiler-Li, and R. Brüschweiler, "Broadband dynamics of ubiquitin by anionic and cationic nanoparticle assisted NMR spin relaxation," Angew. Chem. **133**, 150–154 (2021).

[14]X. Xiang, A. L. Hansen, L. Yu, G. Jameson, L. Bruschweiler-Li, C. Yuan, and R. Brüschweiler, "Observation of sub-microsecond protein methyl-side chain dynamics by nanoparticle-assisted NMR spin relaxation," J. Am. Chem. Soc. **143**, 13593–13604 (2021).

[15]R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, "Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald," J. Chem. Theory Comput. **9**, 3878–3888 (2013).

[16]T. Bremi, R. Brüschweiler, and R. R. Ernst, "A protocol for the interpretation of side-chain dynamics based on NMR relaxation: Application to phenylalanines in antamanide," J. Am. Chem. Soc. **119**, 4272–4284 (1997).

[17]N. R. Skrynnikov, O. Millet, and L. E. Kay, "Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions," J. Am. Chem. Soc. **124**, 6449–6460 (2002).

[18]F. Hoffmann, M. Xue, L. V. Schäfer, and F. A. Mulder, "Narrowing the gap between experimental and computational determination of methyl group dynamics in proteins," Phys. Chem. Chem. Phys. **20**, 24577–24590 (2018).

[19]G. Lipari, A. Szabo, and R. M. Levy, "Protein dynamics and NMR relaxation: Comparison of simulations with experiment," Nature **300**, 197–198 (1982).

[20]R. B. Best, J. Clarke, and M. Karplus, "What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis," J. Mol. Biol. **349**, 185–203 (2005).

[21]J. J. Prompers and R. Brüschweiler, "General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation," J. Am. Chem. Soc. **124**, 4522–4534 (2002).

[22]J. R. Allison, S. Hertig, J. H. Missimer, L. J. Smith, M. O. Steinmetz, and J. Dolenc, "Probing the structure and dynamics of proteins by combining molecular dynamics simulations and experimental NMR data," J. Chem. Theory Comput. **8**, 3430–3444 (2012).

[23]A. A. Smith, M. Ernst, S. Riniker, and B. H. Meier, "Localized and collective motions in HET-s(218-289) fibrils from combined NMR relaxation and MD simulation," Angew. Chem. **131**, 9483–9488 (2019).

[24]N. Salvi, A. Abyzov, and M. Blackledge, "Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation," J. Phys. Chem. Lett. **7**, 2483–2489 (2016).

[25]F. Kümmerer, S. Orioli, D. Harding-Larsen, F. Hoffmann, Y. Gavrilov, K. Teilum, and K. Lindorff-Larsen, "Fitting side-chain NMR relaxation data using molecular simulations," J. Chem. Theory Comput. **17**, 5262–5275 (2021).

[26]A. E. Torda, R. M. Scheek, and W. F. van Gunsteren, "Time-dependent distance restraints in molecular dynamics simulations," Chem. Phys. Lett. **157**, 289–294 (1989).

[27]A. M. Bonvin, R. Boelens, and R. Kaptein, "Time- and ensemble-averaged direct NOE restraints," J. Biomol. NMR **4**, 143–149 (1994).

[28]R. B. Best and M. Vendruscolo, "Determination of protein structures consistent with NMR order parameters," J. Am. Chem. Soc. **126**, 8090–8091 (2004).

[29]P. S. Nerenberg and T. Head-Gordon, "New developments in force fields for biomolecular simulations," Curr. Opin. Struct. Biol. **49**, 129–138 (2018).

[30]S. Riniker, "Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview," J. Chem. Inf. Model. **58**, 565–578 (2018).

[31]D. van der Spoel, "Systematic design of biomolecular force fields," Curr. Opin. Struct. Biol. **67**, 18–24 (2021).

[32]W. Kang, F. Jiang, and Y.-D. Wu, "How to strike a conformational balance in protein force fields for molecular dynamics simulations?," WIREs Comput. Mol. Sci. **12**, e1578 (2022).

[33]V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters," Proteins **65**, 712–725 (2006).

[34]K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, "Improved side-chain torsion potentials for the amber ff99SB protein force field," Proteins **78**, 1950–1958 (2010).

[35]R. B. Best and G. Hummer, "Optimized molecular dynamics force fields applied to the helix–coil transition of polypeptides," J. Phys. Chem. B **113**, 9004–9015 (2009).

[36]D.-W. Li and R. Brüschweiler, "NMR-based protein potentials," Angew. Chem. **122**, 6930–6932 (2010).

[37]J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, Jr., "CHARMM36m: An improved force field for folded and intrinsically disordered proteins," Nat. Methods **14**, 71–73 (2017).

[38]K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data," PLoS One **7**, e32131 (2012).

[39]K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, "Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements," J. Chem. Theory Comput. **8**, 1409–1414 (2012).

[40]O. F. Lange, D. Van Der Spoel, and B. L. De Groot, "Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data," Biophys. J. **99**, 647–655 (2010).

[41]A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard, and R. M. Lanigan, "Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics," Proteins: Struct., Funct., Bioinf. **82**, 195–215 (2014).

[42]F. Hoffmann, F. A. A. Mulder, and L. V. Schäfer, "Accurate methyl group dynamics in protein simulations with AMBER force fields," J. Phys. Chem. B **122**, 5038–5048 (2018).

[43]F. Kümmerer, S. Orioli, and K. Lindorff-Larsen, "Fitting force field parameters to NMR relaxation data," J. Chem. Theory Comput. **19**, 3741–3751 (2023).

[44]J. Yoo and A. Aksimentiev, "Refined parameterization of nonbonded interactions improves conformational sampling and kinetics of protein folding simulations," J. Phys. Chem. Lett. **7**, 3812–3818 (2016).
[45]R. B. Best, D. de Sancho, and J. Mittal, "Residue-specific α-helix propensities from molecular simulation," Biophys. J. **102**, 1462–1467 (2012).
[46]J. Mu, H. Liu, J. Zhang, R. Luo, and H.-F. Chen, "Recent force field strategies for intrinsically disordered proteins," J. Chem. Inf. Model. **61**, 1037–1047 (2021).
[47]P. Robustelli, S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," Proc. Natl. Acad. Sci. U. S. A. **115**, E4758–E4766 (2018).
[48]K. Takemura and A. Kitao, "Water model tuning for improved reproduction of rotational diffusion and NMR spectral density," J. Phys. Chem. B **116**, 6279–6287 (2012).
[49]O. H. S. Ollila, H. A. Heikkinen, and H. Iwai, "Rotational dynamics of proteins from spin relaxation times and molecular dynamics simulations," J. Phys. Chem. B **122**, 6559–6569 (2018).
[50]A. Hershko and A. Ciechanover, "The ubiquitin system," Annu. Rev. Biochem. **67**, 425–479 (1998).
[51]Z. J. Chen and L. J. Sun, "Nonproteolytic functions of ubiquitin in cell signaling," Mol. Cell **33**, 275–286 (2009).
[52]K. N. Swatek and D. Komander, "Ubiquitin modifications," Cell Res. **26**, 399–422 (2016).
[53]B. Brutscher, R. Brüschweiler, and R. R. Ernst, "Backbone dynamics and structural characterization of the partially folded a state of ubiquitin by ^{1}H, ^{13}C, and ^{15}N nuclear magnetic resonance spectroscopy," Biochemistry **36**, 13043–13053 (1997).
[54]X. Zhang, X. Sui, and D. Yang, "Probing methyl dynamics from ^{13}C autocorrelated and cross-correlated relaxation," J. Am. Chem. Soc. **128**, 5073–5081 (2006).
[55]X. Liao, D. Long, D.-W. Li, R. Brüschweiler, and V. Tugarinov, "Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group," J. Phys. Chem. B **116**, 606–620 (2012).
[56]C. Charlier, S. N. Khan, T. Marquardsen, P. Pelupessy, V. Reiss, D. Sakellariou, G. Bodenhausen, F. Engelke, and F. Ferrage, "Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry," J. Am. Chem. Soc. **135**, 18665–18672 (2013).
[57]A. A. Smith, M. Ernst, B. H. Meier, and F. Ferrage, "Reducing bias in the analysis of solution-state NMR data with dynamics detectors," J. Chem. Phys. **151**, 034102 (2019).
[58]N. Bolik-Coulon, O. Languin-Cattoën, D. Carnevale, M. Zachrdla, D. Laage, F. Sterpone, G. Stirnemann, and F. Ferrage, "Explicit models of motion to understand protein side-chain dynamics," Phys. Rev. Lett. **129**, 203001 (2022).
[59]V. A. Jarymowycz and M. J. Stone, "Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences," Chem. Rev. **106**, 1624–1671 (2006).
[60]K. Zumpfe and A. A. Smith, "Model-free or not?," Front. Mol. Biosci. **8**, 727553 (2021).
[61]G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity," J. Am. Chem. Soc. **104**, 4546–4559 (1982).
[62]G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn, "Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins," J. Am. Chem. Soc. **112**, 4989–4991 (1990).
[63]Y. Gu, D.-W. Li, and R. Brüschweiler, "NMR order parameter determination from long molecular dynamics trajectories for objective comparison with experiment," J. Chem. Theory Comput. **10**, 2599–2607 (2014).
[64]A. A. Smith, M. Ernst, and B. H. Meier, "Because the light is better here: Correlation-time analysis by NMR spectroscopy," Angew. Chem., Int. Ed. **56**, 13590–13595 (2017).
[65]A. A. Smith, M. Ernst, and B. H. Meier, "Optimized 'detectors' for dynamics analysis in solid-state NMR," J. Chem. Phys. **148**, 045104 (2018).
[66]A. A. Smith, "Interpreting NMR dynamic parameters via the separation of reorientational motion in MD simulation," J. Magn. Reson. Open **10–11**, 100045 (2022).
[67]C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, "A direct approach to conformational dynamics based on hybrid Monte Carlo," J. Comput. Phys. **151**, 146–168 (1999).
[68]W. C. Swope, J. W. Pitera, and F. Suits, "Describing protein folding kinetics by molecular dynamics simulations. 1. Theory," J. Phys. Chem. B **108**, 6571–6581 (2004).
[69]G. R. Bowman, X. Huang, and V. S. Pande, "Using generalized ensemble simulations and Markov state models to identify conformational states," Methods **49**, 197–201 (2009).
[70]G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, "Progress and challenges in the automated construction of Markov state models for full protein systems," J. Chem. Phys. **131**, 124101 (2009).
[71]V. S. Pande, K. Beauchamp, and G. R. Bowman, "Everything you wanted to know about Markov state models but were afraid to ask," Methods **52**, 99–105 (2010).
[72]J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: Generation and validation," J. Chem. Phys. **134**, 174105 (2011).
[73]J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," Curr. Opin. Struct. Biol. **25**, 135–144 (2014).
[74]B. E. Husic and V. S. Pande, "Markov state models: From an art to a science," J. Am. Chem. Soc. **140**, 2386–2396 (2018).
[75]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).
[76]M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," J. Chem. Phys. **112**, 8910–8922 (2000).
[77]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1-2**, 19–25 (2015).
[78]P. Bauer, B. Hess, and E. Lindahl (2022). "GROMACS 2022.1 manual," Zenodo.
[79]S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, "Structure of ubiquitin refined at 1.8 Å resolution," J. Mol. Biol. **194**, 531–544 (1987).
[80]B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "LINCS: A linear constraint solver for molecular simulations," J. Comput. Chem. **18**, 1463–1472 (1997).
[81]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," J. Chem. Phys. **103**, 8577–8593 (1995).
[82]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**, 000001 (2007).
[83]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**, 7182–7190 (1981).
[84]K. Loth, P. Pelupessy, and G. Bodenhausen, "Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy," J. Am. Chem. Soc. **127**, 6062–6068 (2005).
[85]S. F. Cousin, P. Kadeřávek, N. Bolik-Coulon, and F. Ferrage, "Determination of protein ps-ns motions by high-resolution relaxometry," in *Protein NMR: Methods and Protocols* (Springer, 2018), pp. 169–203.
[86]K. Zumpfe and A. A. Smith (2023). "pyDR," https://github.com/alsinmr/pyDR.
[87]D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC hammer," Publ. Astron. Soc. Pac. **125**, 306–312 (2013).
[88]P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in python," Nat. Methods **17**, 261–272 (2020).
[89]M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noé, "PyEMMA 2: A software package for estimation, validation, and analysis of Markov models," J. Chem. Theory Comput. **11**, 5525–5542 (2015).

[90]N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," J. Comput. Chem. **32**, 2319–2327 (2011).
[91]M. Thürlemann and S. Riniker, "Energy-based clustering: Fast and robust clustering of data with known likelihood functions," J. Chem. Phys. **159**, 024105 (2023).
[92]N. Bolik-Coulon, M. Zachrdla, G. Bouvignies, P. Pelupessy, and F. Ferrage, "Comprehensive analysis of relaxation decays from high-resolution relaxometry," J. Magn. Reson. **355**, 107555 (2023).
[93]G. R. Bowman, "Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation," J. Comput. Chem. **37**, 558–566 (2016).
[94]L. Siemons, B. Uluca-Yazgi, R. B. Pritchard, S. McCarthy, H. Heise, and D. F. Hansen, "Determining isoleucine side-chain rotamer-sampling in proteins from ^{13}C chemical shift," Chem. Commun. **55**, 14107–14110 (2019).
[95]J. J. Chou, D. A. Case, and A. Bax, "Insights into the mobility of methyl-bearing side chains in proteins from $^{3}J_{CC}$ and $^{3}J_{CN}$ couplings," J. Am. Chem. Soc. **125**, 8959–8966 (2003).
[96]A. A. A. I. Ali, F. Hoffmann, L. V. Schäfer, and F. A. A. Mulder, "Probing methyl group dynamics in proteins by NMR cross-correlated dipolar relaxation and molecular dynamics simulations," J. Chem. Theory Comput. **18**, 7722–7732 (2022).
[97]S. Olsson and F. Noé, "Mechanistic models of chemical exchange induced relaxation in protein NMR," J. Am. Chem. Soc. **139**, 200–210 (2017).