Fitting side-chain NMR relaxation data using molecular simulations

Felix Kümmerer[1†], **Simone Orioli**[1,2†], **David Harding-Larsen**[1], **Falk Hoffmann**[3‡], **Yulian Gavrilov**[1], **Kaare Teilum**[1], **Kresten Lindorff-Larsen**[1*]

***For correspondence:** lindorff@bio.ku.dk (KLL)

[†]These authors contributed equally to this work

Present address: [‡]Institute of Biomaterial Science and Berlin-Brandenburg Center of Regenerative Therapies, Helmholtz-Zentrum Geesthacht, Kantstrasse 55, D-14513 Teltow, Germany

[1]Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen. Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark; [2]Structural Biophysics, Niels Bohr Institute, Faculty of Science, University of Copenhagen, Copenhagen, Denmark.; [3]Theoretical Chemistry, Ruhr University Bochum, D-44780 Bochum, Germany

Abstract Proteins display a wealth of dynamical motions that can be probed using both experiments and simulations. We present an approach to integrate side chain NMR relaxation measurements with molecular dynamics simulations to study the structure and dynamics of these motions. The approach, which we term ABSURDer (Average Block Selection Using Relaxation Data with Entropy Restraints) can be used to find a set of trajectories that are in agreement with relaxation measurements. We apply the method to deuterium relaxation measurements in T4 lysozyme, and show how it can be used to integrate the accuracy of the NMR measurements with the molecular models of protein dynamics afforded by the simulations. We show how fitting of dynamic quantities leads to improved agreement with static properties, and highlight areas needed for further improvements of the approach.

Introduction

Proteins are dynamical entities and a detailed understanding of their function and biophysical properties requires an accurate description of both their structure and dynamics. Nuclear magnetic resonance (NMR) experiments, in particular, have the ability to probe and quantify dynamical properties on a wide range of time scales and at atomic resolution. Computationally, molecular dynamics (MD) simulations also make it possible to probe both the structure and dynamics of proteins. Indeed, NMR and MD simulations may fruitfully be combined in a number of ways (***Case, 2002***; ***Lindorff-Larsen et al., 2005***), including using experiments for validating or improving the force fields used in simulations (***Norgaard et al., 2008***; ***Li and Brüschweiler, 2010***, ***2011***), or for using the simulations as a tool to interpret the experiments (***Nodet et al., 2009***; ***Brookes and Head-Gordon, 2016***; ***Chen et al., 2019***; ***Vasile and Tiana, 2019***; ***Bottaro et al., 2020***).

A particularly tight integration between NMR and MD simulations has involved NMR spin relaxation experiments, which probe motions on the ps-to-ns timescales. Indeed, already soon after the first reported MD simulation of a protein (***McCammon et al., 1977***), such simulations were compared to NMR order parameters (***Lipari et al., 1982***). Integration between spin relaxation and MD is aided by the fact that MD can probe the relevant time scales and that the physical processes leading to spin relaxation are relatively well understood, and can therefore be modelled computationally.

The capability to reach given timescales does not, however, automatically translate into a perfect match between experiments and the simulated observables (***van Gunsteren et al., 2018***; ***Bot-***

taro and Lindorff-Larsen, 2018). Indeed, significant discrepancies between the two can emerge because of multiple factors, such as insufficient sampling (*Bernardi et al., 2015*), inaccurate description of the observable, i.e. the so-called forward model (*Cordeiro et al., 2017*), the imperfection of the underlying force fields (*Rauscher et al., 2015*; *Henriques et al., 2015*; *Robustelli et al., 2018*; *Piana et al., 2020*; *Nerenberg and Head-Gordon, 2018*) or a combination of these factors.

NMR spin relaxation experiments are generally interpreted using variants of the so-called model free formalism (*Halle and Wennerström, 1981*; *Lipari and Szabo, 1982*; *Clore et al., 1990*) that interprets the experimental data using generalized order parameters that describe the amplitudes of the motions and time scales associated with those motions. A common approach to compare MD and NMR relaxation experiments involves calculating order parameters from the simulations and comparing to those extracted from the experiments. The results of such comparisons have shown that while MD simulations generally give a relatively good agreement with order parameters for the backbone amide groups, the agreement for methyl bearing side chains is more variable (*Bremi et al., 1997*; *Skrynnikov et al., 2002*; *Best et al., 2005*; *Showalter et al., 2007*; *Liao et al., 2012*; *O'Brien et al., 2016*; *Bowman, 2016*; *Anderson et al., 2020*).

A more detailed and richer combination, however, involves calculating the NMR relaxation rates directly from the simulations and comparing these to experiments. While this approach does not necessarily provide detailed information about the timescales of these motions, it has the advantage of not being dependent on a specific analytical model which might influence the interpretation of the relaxation data. Recently, an approach was developed to calculate deuterium NMR relaxation rates in methyl bearing side chains from MD simulations (*Hoffmann et al., 2018b*, *2020*). Comparison to experiments revealed systematic deviations from simulations with several different force fields, and corrections to the methyl torsion potential were developed based on quantum calculations and shown to increase agreement with experiments (*Hoffmann et al., 2018a*, *2020*). Despite these improvements and a substantial body of work on studies of side chain dynamics (*Lee et al., 1997*; *Palmer III, 1997*; *Ming and Brüschweiler, 2004*; *Cousin et al., 2018*; *Anderson et al., 2020*), the agreement remains imperfect and further improvements would be desirable.

One possible way to reduce the systematic error coming from force field inaccuracies is to introduce a bias in the ensemble generated by MD simulations to improve agreement with the experimental data. Such a bias may either be introduced during the simulation by adding a correction to the force field that depends on the experimental measurements, or after the simulation has been completed through a procedure known as *reweighting*. Using biased simulations it is, for example, possible to construct conformational ensembles that are in agreement with backbone and side chain order parameters from NMR spin relaxation (*Best and Vendruscolo, 2004*; *Lindorff-Larsen et al., 2005*).

Most implementations of such methods for biasing against experiments work by updating some 'prior' information, typically encoded in the MD force field, with information from the experiments by changing the weights of each conformation so as to improve the agreement with experiments when calculated with these new weights. Often these problems are highly underdetermined in that there are many more parameters (weights) to be determined than experimental measurements. Thus, an important ingredient in many methods is a framework to avoid overfitting by balancing the information from the force field with that from the data, and many approaches apply Bayesian or Maximum Entropy formalisms for this purpose (*Bonomi et al., 2017*; *Orioli et al., 2020*).

Although different methods do so in different ways and with different assumptions, most reweighting methods share a key commonality: they can only deal with static (equilibrium) observables, i.e. quantities that can be expressed as ensemble averages over a set of configurations (*Bonomi et al., 2017*; *Orioli et al., 2020*). Evidently, this complicates the generation of conformational ensembles that match the spin relaxation data, which depend both on the amplitudes and time scales of the motions.

A few approaches have been described to fit NMR relaxation data directly. For example, the

isotropic reorientational eigenmode dynamics (iRED) approach (***Prompers and Brüschweiler, 2002***) uses a covariance matrix to calculate spin relaxation, and may be applied to blocked simulations thus indirectly taking into account the time scales of the motions (***Gu et al., 2014***). Recently, a more direct approach has been described to reweight MD simulations against NMR spin relaxation data (***Salvi et al., 2016***, ***2019***). The basic idea is to split one or more simulations into smaller parts (blocks), each of which are long enough to contain information about both the amplitudes and time scales of the motions that lead to spin relaxation. By averaging over such blocks, one can then estimate NMR relaxation parameters that may be compared to experiments. Similar to the reweighting methods described above for individual configurations, one can instead use reweighting of the blocks and thus determine weights that improve agreement with the experimental data without having to decompose it into order parameters. This approach, called ABSURD (Average Block Selection Using Relaxation Data), has been applied to the analysis of backbone NMR relaxation of intrinsically disordered proteins that are difficult to analyse using conventional model free techniques.

We here describe a new approach to interpret NMR relaxation data, focusing our application on side chain dynamics. Our approach builds upon the ABSURD method, and includes two extensions. First, we use the recently described methods to calculate side chain NMR relaxation parameters from molecular dynamics simulations (***Hoffmann et al., 2018b***). Second, we extend ABSURD by including an entropy restraint term in the optimization that helps avoid overfitting. We thus term our method ABSURDer (Average Block Selection Using Relaxation Data with Entropy Restraints). We applied ABSURDer to study the fast timescale dynamics of T4 lysozyme (T4L), a protein which has been the subject of numerous experimental and computational studies. We validate and describe our method using synthetic data and then apply it to data from NMR experiments. We find that reweighting simulations with ABSURDer improves the agreement between methyl relaxation rates in synthetic data sets coming from both the same and a different force field. Furthermore, we find that these reweighted trajectories also show improvements in the agreement of equilibrium observables, such as rotamer distributions. Finally, we use experimental NMR data for reweighting to improve agreement between simulations and experiments. In this case we find smaller improvements are possible, and we discuss possible origins of this observation.

Results and Discussion

Overview of the ABSURDer approach

The workflow in ABSURDer consists of three separate steps (Fig. 1). Steps 1 and 2 correspond to a previously described method for calculating side chain relaxation, and step 3 corresponds to the ABSURD approach with the inclusion of an additional restraint to decrease the risk of overfitting, and to balance the information in the force field with that in the data.

First, we performed three independent $5\,\mu s$-long MD simulations to sample the structure and dynamics of T4L. We here use the recently optimized Amber ff99SB*-ILDN force field with the TIP4P/2005 water model and a modification of the methyl torsion potential, based on CCSD(T) coupled cluster quantum chemical calculations of isolated dipeptides, that improves agreement with NMR relaxation data (***Hoffmann et al., 2018a***). See the Methods section for further details about all simulations.

Second, we divided the total of $15\,\mu s$ simulations into 1500 $10\,ns$ blocks, and calculated NMR relaxation parameters for each of these independently (see Methods, and below for a more detailed discussion on the origin of this block size). Briefly described we: (i) calculate internal correlation functions $C_{int}(t)$ for methyl C-H bonds; (ii) use backbone motions to describe the global overall tumbling motion; (iii) combine internal and global rotational motions to calculate the spectral density function $J(\omega)$ and (iv) extract NMR relaxation parameters from this. We here focus on three NMR relaxation parameters $R(D_z)$, $R(D_y)$ and $R(3D_z^2-2)$, and calculate these rates for each methyl group and each of the 1500 blocks. As done previously (***Hoffmann et al., 2018b***) we exclude residue

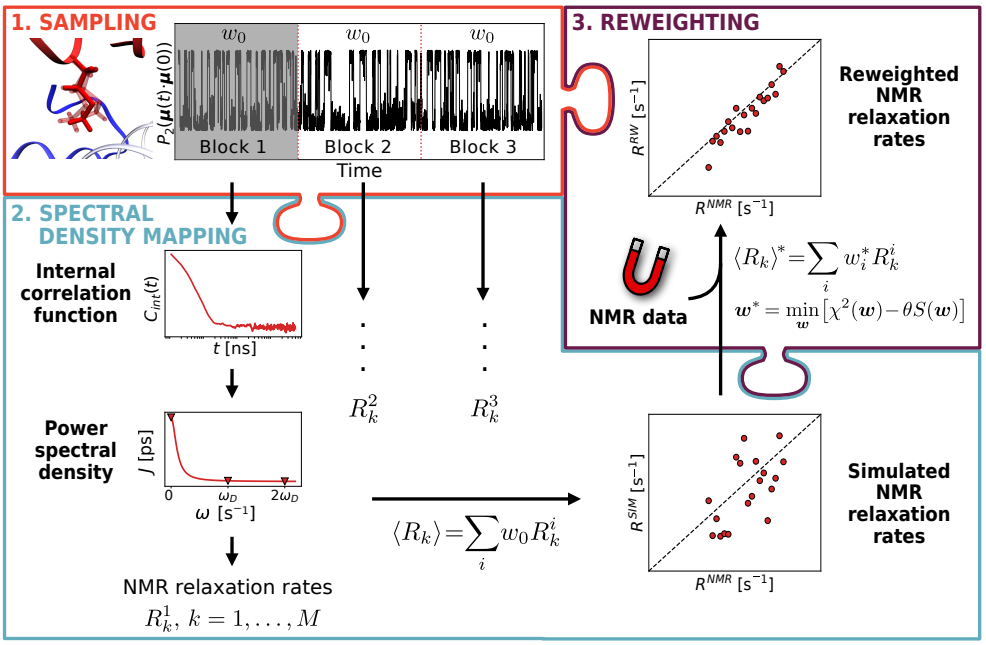


Figure 1. Schematic representation of the ABSURDer workflow. Steps 1 involves sampling protein dynamics using one or more longer MD simulations and calculating bond vector orientations for the resulting trajectories. These are then divided into blocks and in Step 2 we calculate correlation functions, spectral densities and NMR relaxation rates for each block. In Step 3, we optimize the agreement between the average calculated rates and experimentally measured values by changing the weights of the different blocks. For details see main text, methods and code online.

ALA146 from the analyses.

The third step in our approach consists of reweighting the MD simulations to improve agreement with the experimental data. Our approach combines the ABSURD method with concepts from Bayesian/Maximum Entropy ensemble refinement. The goal here is to assign to each of the 1500 $10\,\mathrm{ns}$ simulations a different weight (w_i; $i = 1, \ldots, 1500$) so as to improve agreement with the experimental data, as quantified by the χ^2 between experimental and calculated NMR relaxation rates. Such an approach is, however, prone to overfitting and does not necessarily utilize the information about the conformational landscape encoded in the molecular force field. Thus, we and others have previously described methods that circumvent this problem by adding an entropy regularization to the χ^2 minimization (***Gull and Daniell, 1978***; ***Cesari et al., 2018***; ***Köfinger et al., 2019***; ***Bottaro et al., 2020***; ***Orioli et al., 2020***), by instead minimizing the functional $\mathcal{L}(\mathbf{w}) = \chi^2(\mathbf{w}) - \theta S_{rel}(\mathbf{w})$ (see Methods). In this equation S_{rel} represents the relative entropy that measures how different the weights are from the initial (uniform) weights, and thus how much the final set of simulations differs from the initial set. Typically, we represent this as $\phi_{eff} = \exp(S_{rel})$, which represents the effective fraction of the original 1500 blocks (sub-simulations) that contribute to the final ensemble. In these equations the parameter θ sets the balance between fitting the data (minimizing χ^2) and keeping as much as possible of the original simulation (maximizing ϕ_{eff}). We refer the reader to previous literature about the methods overall and how best to select this parameter (***Andrae et al., 2010***; ***Hummer and Köfinger, 2015***; ***Bottaro et al., 2018***; ***Cesari et al., 2018***; ***Köfinger et al., 2019***; ***Crehuet et al., 2019***; ***Chen et al., 2019***; ***Bottaro et al., 2020***; ***Orioli et al., 2020***).

To describe and understand how well ABSURDer works, we have applied it both to synthetic and experimental data for T4L. We used data of increasing complexity in an attempt to disentangle problems arising from the sampling, the force field, the forward model to calculate relaxation data and our approach to fit the data. In all cases, we employed the 1500 $10\,\mathrm{ns}$ blocks coming from the three $5\,\mu\mathrm{s}$ simulations with the Amber ff99SB*-ILDN force field to fit the data. In our first ap-

plication we created synthetic data using the same force field by running five additional $1\,\mu s$-long simulations. In this case, the only differences between the 'data' and the simulations would arise from insufficient sampling. In our second application, we again created synthetic data, in this case however using the Amber ff15ipq force field (with a modified methyl torsion potential (***Hoffmann et al., 2020***)) and the SPC/E_b water model (***Takemura and Kitao, 2012***; ***Debiec et al., 2016***). We performed three $1\,\mu s$ simulations and used these to generate synthetic NMR relaxation data. This test enables us to examine how the method behaves where there are real differences between the potential used in the simulations we use to fit the data, and that give rise to the 'experimental data'. In contrast to using actual experimental data, however, we here have access to the full underlying ensembles and dynamics, and thus are able to compare e.g. full spectral density function and rotamer distributions. Finally, we applied ABSURDer to experimental NMR relaxation data, providing an example of how the approach might be used in practice. The results of these three levels of complexity are described in the following sections.

Determining block size and assessing convergence

The choice of the block length plays a central role in the calculation of the relaxation rates and in the possibility to optimize the results via reweighting. For this, two opposing effects need to be considered (Fig. 2A). On one hand, a long block size allows for a more correct representation of the correlation functions and thus estimation of the NMR relaxation rates. On the other hand, it is desirable to increase the variation across the different blocks, as this makes it easier to fit the relaxation data by reweighting; however this variance decreases with the length of the blocks (Fig. 2A). Ideally, one would choose a block length that is (i) long enough not to introduce too much bias compared to calculating correlation functions over the full trajectory and (ii) short enough that each block is not just a 'converged' representation of the full trajectory. We chose $10\,\mathrm{ns}$ as a block length that provides a large number of blocks to be used in the fit and still balances these two requirements (Fig. 2A). This choice allows us to calculate the internal time correlation functions of the methyl C-H bonds up to a maximum lag time of $5\,\mathrm{ns}$, as also done previously (***Hoffmann et al., 2020***). We note that the chosen block size is also close to the global tumbling time of the protein ($\tau_R \approx 11\,\mathrm{ns}$). As NMR spin relaxation experiments are mostly sensitive to motions on time scales up to approximately τ_R, this explains why much shorter block sizes do not capture the NMR relaxation parameters accurately. Examining each of the three relaxation rates separately (Fig. S1) we find that in particular $R(D_y)$ is poorly determined using short blocks. Finally, we note that while calculations of correlation functions from just a single $10\,\mathrm{ns}$ of simulation is generally not sufficient to compare to NMR relaxation data, in all our analyses we average over large numbers of blocks. In that case, we find that averaging over 1500 $10\,\mathrm{ns}$ blocks gives very similar results as 3 $5\,\mu s$ blocks (Fig. 3A). This is indeed expected as the division into blocks will only affect the calculations of the correlation that 'cross the boundary' of the blocks.

To assess the level of convergence, we divided our three $5\,\mu s$-long simulations into 15 $1\,\mu s$ segments and employed leave-N-out cross-validation to estimate the impact of the amount of sampling for the estimation of relaxation rates. For each value of N, we calculated the NMR relaxation rates by averaging over the N segments, and then compared to the average over the remaining $15 - N$ segments by calculating the root mean square deviation (RMSD) between the two sets of rates. We repeated these calculations for all possible combinations of leaving out N segments, and the RMSD values were averaged. Finally, we repeated these calculations for all values of N ($N = 1, \ldots, 7$) (Fig. 2B–D). Even at $N = 1$ we find relative low RMSD values when compared to the spread of the rates (Fig. 3A–C), and these errors decrease further as additional sampling is included. Thus, we conclude that using several microseconds of sampling we are able to obtain relatively precise estimates of the relaxation rates, as would be expected given that they report on ps–ns dynamics in the protein.

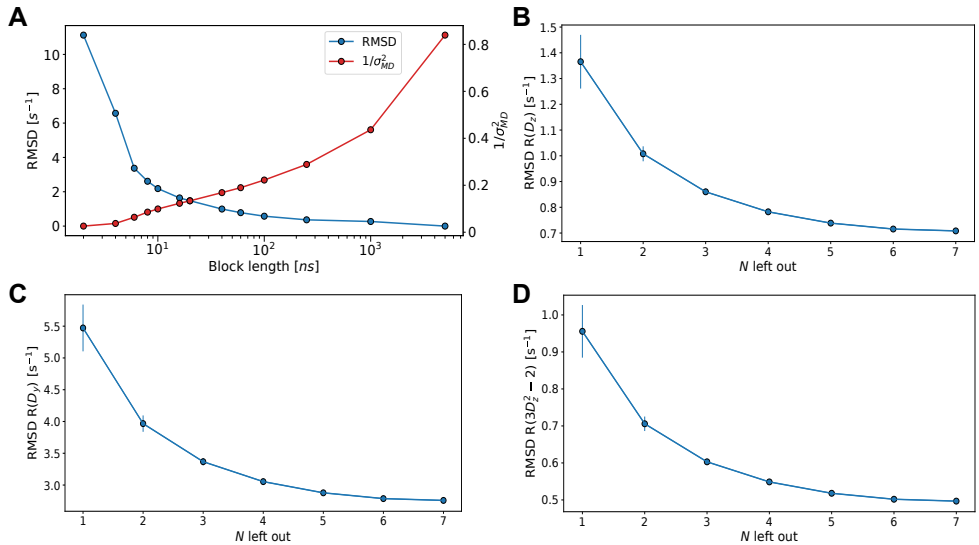


Figure 2. Choice of block size and convergence of NMR relaxation rates. (A) Overall (blue) RMSD of NMR relaxation rates and (red) average inverse variance with respect to the full MD data set as a function of the block length. (B-D) Average RMSD of the three NMR relaxation rates obtained by comparing a set of N $1\,\mu s$ long segments to the rates calculated from the remaining $15 - N$ segments.

Fitting synthetic data generated with the same force field

We ran five $1\,\mu s$-long simulations with the Amber ff99SB*-ILDN force field and calculated NMR relaxation parameters from these. In what follows, this data will be referred to as 'NMR' or 'experiments' to indicate that we treat them as observables from an experiment. We calculated the same observables as the average over 1500 $10\,ns$ blocks coming from the three $5\,\mu s$ simulations with the same force field, and compared them to the synthetic experimental data (Fig. 3A–C). As expected the calculated and experimental values are strongly correlated, in line with the fact that these were generated from the same underlying physical model. The calculated values of a reduced χ^2 were 22, 15 and 17 for $R(D_z)$, $R(D_y)$ and $R(3D_z^2 - 2)$, respectively.

We then asked whether we could improve the agreement between experiments and simulations by changing the weights of each of the 1500 blocks to reweight the calculated observables. We thus determined a set of weights that minimize the functional $\mathcal{L}(\mathbf{w}) = \chi^2(\mathbf{w}) - \theta S_{rel}(\mathbf{w})$ at different values of θ and plot the resulting χ^2 vs. ϕ_{eff} (Fig. 3A–C; insets). It is clear that as θ is decreased to put greater weight on fitting the data, the agreement between experiment and simulation can be improved substantially, though at the cost of utilizing only a fraction of the input simulations. For simplicity, we here opted to use a value θ that gives rise to $\phi_{eff} \approx 0.2$, though our conclusions are relatively robust to this choice. At this level of fitting the agreement between experiment and simulations improves substantially (reduced χ^2 approximately 3, 1 and 2 for $R(D_z)$, $R(D_y)$ and $R(3D_z^2-2)$, respectively). We note also that $\phi_{eff} \approx 0.2$ represents a total of $3\,\mu s$ of MD simulation, which in itself is sufficient to obtain relatively converged values (Fig. 2B–D).

In addition to reweighting the simulations using all three types of relaxation rates ($R(D_z)$, $R(D_y)$ and $R(3D_z^2-2)$), we also performed reweighting using each rate individually or in pairs, and used the remaining rate(s) for cross validation (Fig. S2). Overall, we find that fitting one or two rates leads to improvements also in rates that are not used in reweighting. For certain rates and very aggressive reweighting (low values of ϕ_{eff}) we observe, however, an increase in the cross validated rates. This behaviour is expected as the three rates in part depend and report on the same properties of the spectral density function. Indeed, it is clear that $R(D_y)$ is less correlated with the two other rates, which is likely due to the fact that only this relaxation rate depends on the spectral density at zero frequency ($J(0)$). Nevertheless, the general improvement in cross validation down to $\phi_{\text{eff}} \approx 0.2$

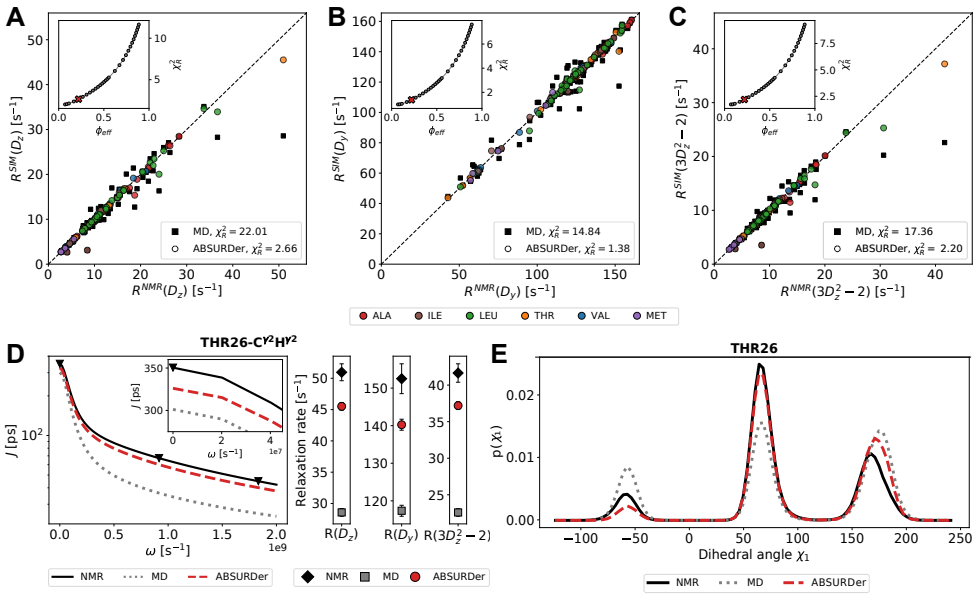


Figure 3. ABSURDer applied to synthetic data generated with the same force field. (A – C) Comparison between 'experiment' and simulation for $R(D_z)$, $R(D_y)$ and $R(3D_z^2 - 2)$ both before and after reweighting with ABSURDer. The insets show the behaviour of the respective reduced χ^2 vs. ϕ_{eff} during the reweighting using different values of θ. The chosen $\theta = 300$ (resulting in $\phi_{eff} \approx 0.2$) is shown as a red cross. (D) The spectral density function and $R(D_z)$, $R(D_y)$ and $R(3D_z^2 - 2)$ of THR26-$C^{\gamma 2}H^{\gamma 2}$ before and after reweighting with ABSURDer. $J(0)$, $J(\omega)$ and $J(2\omega)$ are shown as black triangles. The errorbars represent the standard error of the mean from averaging over the 1500 $10\,\mathrm{ns}$ sub-trajectories. (E) The χ_1 angle probability density distribution of THR26 before and after reweighting with ABSURDer.

provides another reason why we chose this value for our analyses.

While the NMR relaxation parameters are the experimental observables, they are calculated from the simulations via the spectral density functions, $J(\omega)$. Spectral density functions are often estimated from experiments using simplified and analytically tractable functions that balance the complexity of the motions and the number of free parameters that can be estimated from limited experiments. Using synthetic data gives us the possibility of comparing the spectral density functions from the simulations, both before and after reweighting, with those that give rise to the experiments. We here exemplify this using a single methyl group from THR26 (Fig. 3D); the code and data to generate plots for all residues are available online (github.com/KULL-Centre/papers/tree/master/2020/AB kummerer-orioli-et-al-2020). The observed NMR relaxation parameters depend on the spectral density at three frequencies ($\omega = 0$, ω_D and ω_{2D}, respectively) and it is thus not surprising that fitting to this data improves agreement at these frequencies. Nevertheless, it is also clear that our reweighting of the trajectories leads to a general improvement between the calculated spectral density function and that which was used to generate the synthetic data. As it is clear from this example (Fig. 3D), and from the overall results on all rates (Fig. 3A–C), it is also evident that discrepancies between experiments and simulations remain. While it is possible to fit the data more closely by decreasing θ this comes with the risk of potential overfitting (Fig. S2) and diminished trust of the input simulations.

The observed NMR relaxation parameters depend on a complex combination of both overall tumbling and various types of internal motions. In addition to the fast rotation of the methyl groups, this includes fluctuations of the backbone and motions both within and between rotameric states. It has previously been shown that rotamer jumps, in particular when this occurs on a timescale faster than rotational tumbling, can contribute substantially to side chain relaxation (***Best et al., 2005***; ***Hu et al., 2005***). We thus calculated the side chain dihedral angles and compared the values before and after reweighting to those from the simulation used to generate the synthetic

data. Highlighting again THR26 we see that, due to imperfect sampling of rotamer distributions, even in multi-microsecond simulations, there are clear differences between the distribution of the χ_1 dihedral angle in the simulations used to generate the data compared to that used to fit it (Fig. 3E). Particularly important, however, we also see a clear improvement after the reweighting. Thus, we find that by reweigting a 'dynamic' property (NMR relaxation rates) we also find improvement in a 'static' property such as the distributions of dihedral angles. Similar results were found for a wide range of dihedral angles including also in the backbone (Fig. S4) (see also github.com/KULL-Centre/papers/tree/master/2020/ABSURDer-kummerer-orioli-et-al-2020).

Fitting synthetic data generated with a different force field

The calculations above demonstrate how ABSURDer can be used to fit NMR data using MD simulations, but neglects the challenges arising from imperfect force fields. Thus, we increased the complexity of our comparisons by generating synthetic data with a different force field than that used to fit the data. We thus ran three $1\,\mu s$-long simulations with the Amber ff15ipq force field and calculated synthetic NMR relaxation parameters from these. We again used ABSURDer to fit the 1500 $10\,ns$ blocks to this data. In this case, imperfect agreement arises both due to insufficient sampling and differences in the underlying potential, leading to differences both in the dynamics and thermodynamic averages.

This added complexity is evident when we compare 'experimental' and calculated NMR relaxation rates, which show substantial differences and reduced χ^2 values that are about two orders of magnitude greater than in the situation described above (177, 156 and 135 for $R(D_z)$, $R(D_y)$ and $R(3D_z^2 - 2)$, respectively) (Fig. 4A–C). We can improve the agreement by applying the reweighting procedure, but in contrast to the situation described above, it is more difficult to obtain a good agreement. Thus, when selecting θ to give $\phi_{eff} = 0.2$ we reduce the χ^2 values by about two-fold, compared to the about 10-fold in the case described above. We observe a similar behaviour also when only subsets of rate types are employed for fitting (Fig. S3). Nevertheless, it is clear that by changing the weights of the 1500 blocks we can improve the agreement between experiment and simulations (Fig. 4A–C). Looking at the different methyl groups, we do not find that any specific residue gives rise to substantially worse agreement than others.

We again examined the consequences of the reweighting on the dihedral angles distributions, in this case focusing on ILE9 (Fig. 4D). The NMR relaxation experiments probe both the dynamics of the C^γ and C^δ in isoleucine residues, and the χ_1 and χ_2 dihedrals may display substantial couplings. We thus calculated the two-dimensional probability densities before and after reweighting and compared to the distribution from the Amber ff15ipq simulations that we used to generate the synthetic data. Although systematic differences remain between the reweighted and 'experimental' probability densities, it is clear that ABSURDer substantially corrects the probability densities. For example, the dominant rotamer in the Amber ff15ipq simulation used to calculate the synthetic NMR relaxation has χ_1 as gauche$^+$ and χ_2 as gauche$^-$ and has a population of 0.46. The Amber ff99SB*-ILDN simulation has a population of 0.25 for this rotamer, a value that increases to 0.41 after reweighting. At the same time, the population of the gauche$^+$/gauche$^+$ rotamer, which is almost not populated in the Amber ff15ipq simulation, decreases from 0.42 to 0.33. We note that although it is not unexpected, it is not trivial that a reweighting procedure employing kinetic data is able to correct static quantities such as equilibrium probability density. As above, we also find improvements in the backbone dihedral angles (Fig. S5). This example shows how the application of ABSURDer may help correct for inaccurate rotamer distributions in the simulations, though how much depends both on the information in the experimental data, and the sampling of the prior (force field). Indeed, no population shift can be obtained through reweighting if one of the rotameric states of interest is never observed during a simulation.

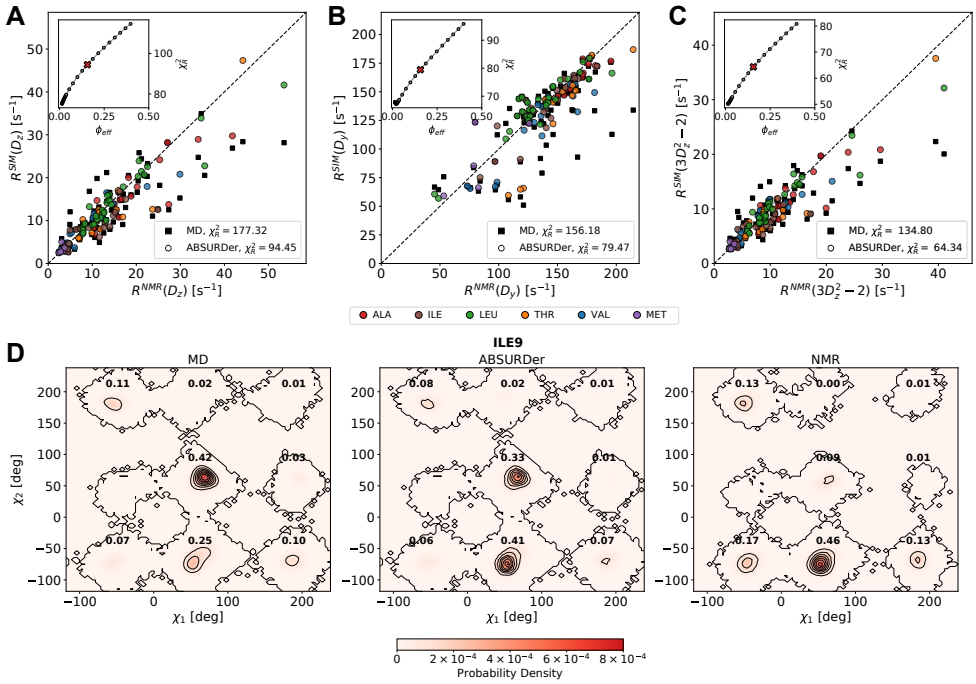


Figure 4. ABSURDer applied to synthetic data generated with different force fields. (A - C) $R(D_z)$, $R(D_y)$ and $R(3D_z^2 - 2)$ from both data sets are compared before and after reweighting with ABSURDer. The insets show the behaviour of the respective reduced χ^2 vs. ϕ_{eff} during the reweighting using different values of θ. The chosen $\theta = 4500$ corresponding to $\phi_{eff} \approx 0.2$ is shown as a red cross. (D) Probability density distribution of the χ_1 and χ_2 angles of ILE9 before and after applying ABSURDer, and the corresponding 'experimental' data. Numbers in bold indicate probabilities of the corresponding rotamer states.

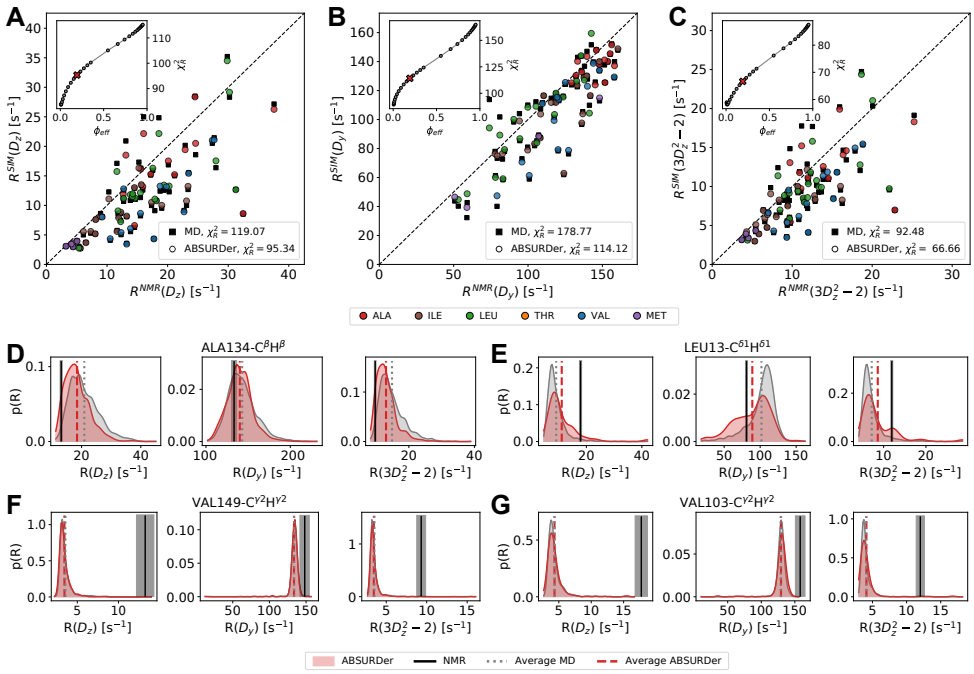


Figure 5. ABSURDer applied to experimental data. (A - C) $R(D_z)$, $R(D_y)$ and $R(3D_z^2-2)$ from both data sets are compared before and after reweighting with ABSURDer. The insets show the behaviour of the respective reduced χ^2 vs. ϕ_{eff} during the reweighting using different values of θ. The chosen $\theta = 1400$ corresponding to $\phi_{eff} \approx 0.2$ is shown as a red cross. (D-G) Distributions of $R(D_z)$, $R(D_y)$ and $R(3D_z^2-2)$ over the blocks for ALA134-C$^\beta$H$^\beta$ (D), LEU13-C$^{\delta1}$H$^{\delta1}$ (E), MET102-C$^\epsilon$H$^\epsilon$ (F) and VAL103-C$^{\gamma2}$H$^{\gamma2}$ (G) before and after reweighting with ABSURDer. Vertical lines represent the average relaxation rate for the respective data set.

Fitting experimental data

Having used synthetic data to show how ABSURDer makes it possible to fit MD simulations using NMR relaxation experiments, we now proceed to fit to experimental data on T4L. Before discussing the results, however, some comments are in order. In both cases above where we employed synthetic data to mimic experimental NMR relaxation rates, we used relaxation rates for all 100 methyl groups in T4L (apart from ALA146). In practice, however, not all of these can be measured accurately e.g. due to overlapping peaks or artifacts from strong ^{13}C–^{13}C coupling in certain residues (***Hoffmann et al., 2018b***). Thus, below we use only the 73 methyl groups whose relaxation rates were recently measured (***Hoffmann et al., 2018b***). To examine the effect of optimizing against a subset of residues we first repeated the reweighting procedure using the synthetic data, but restricting to the set of 73 methyl groups, and used the remaining 27 as a test set (Fig. S6). For both sources of synthetic data we find that optimizing on the 73 methyl groups lead to improvements in the remaining 27, unless aggressive reweighting to low ϕ_{eff} is employed.

We thus proceeded to use the previously measured data (***Hoffmann et al., 2018b***) recorded at $950\,\text{MHz}$, and compared these to the NMR relaxation parameters from the three $5\,\mu\text{s}$ simulations, calculated as averages over the 1500 $10\,\text{ns}$ blocks (Fig. 5A–C). We find a reasonable agreement (reduced χ^2 values of about 119, 179 and 92 for $R(D_z)$, $R(D_y)$ and $R(3D_z^2-2)$, respectively), in line with similar calculations previously reported from ten $0.3\,\mu\text{s}$ simulations. In contrast to the results described above for the synthetic data we observe, however, some systematic differences with the calculated rates on average being 16% lower than experiments.

We applied ABSURDer to improve agreement with the NMR experiments. When fitting to $\phi_{eff} = 0.2$ we are only able to obtain modest improvements of the calculated rates, decreasing the χ^2 values by on average 30%. The difficulty in fitting this dataset also occurs when fitting only a single or two of the three types of relaxation rates, and as above $R(D_y)$ appears to be less correlated with

the other two rates (Fig. S7). We also examined the agreement for each type of amino acid and find that in particular relaxation rates in valine methionine methyl groups are poorly estimated and difficult to reweight (Fig. S8).

We then asked the question why some but not other relaxation rates and methyls can be improved by ABSURDer. Reweighting approaches such as ABSURDer rely on finding a subensemble (or subset of trajectories) that fits the experimental data better than the full ensemble when averaged uniformly. Thus, the successful application requires that the calculated rates for the different blocks can be combined (linearly) to fit the data. We thus calculated the distribution of the relaxation rates over the 1500 blocks for selected methyl groups and compared the averages before and after reweighting to the experiments (Fig. 5D–G). For some residues and relaxation rates, exemplified by ALA134-$C^{\beta}H^{\beta}$ and LEU13-$C^{\delta1}H^{\delta1}$ in Fig. 5D–E, we find a relatively broad distribution where the relaxation rates calculated for the different blocks overlap with the experimental rates. In these cases it is possible to improve agreement by increasing the weights of some blocks and decreasing the weights of others. In several other cases, exemplified by MET102-$C^{\epsilon}H^{\epsilon}$ and VAL103-$C^{\gamma2}H^{\gamma2}$ in Fig. 5F–G, the experimental relaxation rates lie outside the range sampled in our (blocks of) MD simulations. In such cases, reweighting cannot improve agreement substantially, because no linear combination of the blocks can get close to the experiments.

Conclusions

We have here presented ABSURDer, an extension of the ABSURD approach (***Salvi et al., 2016***), and applied it to NMR relaxation measurements of side chain dynamics. We have described and validated ABSURDer using synthetic data and applied it to experimental NMR data. As expected, for the three levels of 'complexity' we obtain different levels of agreement after reweighting between simulated and 'experimental' data.

When we used the same force field to generate synthetic data as that used to fit the data (ff99SB*-ILDN) we find good agreement even before reweighting, in line with the fact that the calculations of the relaxation parameters are relatively precise even using only a few microseconds of simulations (Fig. 2). Nevertheless, we can improve agreement further by reweighting and find, for example, improvements in dihedral angle distributions when we fit against the relaxation rates. As expected from the fact that we fit against the relaxation rates, we also find overall improvement of the spectral density functions after reweighting.

The overall observations were similar when we generated synthetic data using Amber ff15ipq and fitted using simulations generated with ff99SB*-ILDN. Here, we are still able to achieve a substantial improvement in the quality of the fit (around a 50% decrease in the overall χ^2). Notably, also in this case the overall agreement between 'experimental' and simulated spectral densities and rotamer distributions increases upon reweighting.

When we fitted the experimental NMR data to the simulations with ff99SB*-ILDN we obtain a more modest improvement (an overall 30% decrease in the χ^2). Examining the probability distribution of rates over the different blocks and residues, we find that while some of these are broad and highly overlapping with the experimental average, many others are sharp and incompatible with the experimentally determined rates. Clearly, it is difficult to reweight when there is only little overlap between the experimental value and the rates calculated from the different blocks. This problem is exacerbated by the fact that we fit all rates simultaneously leading to a problem akin to the 'curse of dimensionality'.

The natural question arises whether it is possible to increase the agreement between simulated and experimental NMR relaxation rates further. As described in the Introduction, the main sources of error between simulated and experimental data are: (i) insufficient sampling, (ii) force field inaccuracies, and (iii) remaining errors in the forward model used to estimate the relaxation rates from the trajectories. In the absence of a measure of ground truth it may, however, be difficult to disentangle these effects.

Our analysis of errors due to sampling (Fig. 2B–D) shows that the rates calculated from the $15\,\mu s$ of simulation are rather precise compared e.g. to the deviation between experimental and calculated rates (Fig. 5). We have also calculated the total weight of the blocks coming from each of the three $5\,\mu s$ simulations after reweighting, and while these sums differ from the expected average of 0.33, none of them are greater than 0.5 (Tab. S1). While these observations suggest that sampling is not the main source of deviation, the situation during reweighting is more complicated. Indeed, the selection of the block size demonstrates one of the complicating factors (Fig. 2A); using a too short block size leads to systematic errors while using too long blocks leads to 'local convergence' and thus a narrow distribution of rates across the blocks. While a $10\,ns$ block size strikes a reasonable balance between these opposing factors it is clear that, in many cases, there is not sufficient overlap between the distribution of calculated rates and the experimental values (Fig. 5). Clearly, additional sampling to obtain more blocks could help alleviate this situation by increasing the number of samples in the tails of these distributions, though in many cases the deviations are so large that substantially more sampling would be needed. However, regardless of the amount of sampling, the choice to cut the trajectory into blocks naturally introduces a pre-averaging of the dynamics at the sub-nanosecond scale; therefore, if the dynamics at short timescales is inaccurate because of force field imprecision, there is nothing ABSURDer can do to fix it. One future solution to this problem could be to use blocks of different sizes and combine the derived data. In this way one could fit motions on fast timescales by using small blocks and slower motions with longer blocks. Such an approach could be used together with relaxation data from a wider range of magnetic field strengths to probe motions on a range of time scales (***Cousin et al., 2018***).

As reweighting methods such as ABSURDer rely on an overlap between experiment and calculated parameters prior to reweighting, they are aided by good agreement before reweighting (***Hummer and Köfinger, 2015***; ***Orioli et al., 2020***; ***Larsen et al., 2020***). Indeed the extent of reweighting needed to obtain good agreement is related to a measure of the error in the force field (***Qian, 2001***; ***Orioli et al., 2020***). Thus, as starting point for our calculations we used force fields that have recently been shown to give improved agreement to side chain relaxation data (***Hoffmann et al., 2018a***,b, ***2020***). Those studies used quantum-level calculations to modify the barriers for the methyl spinning, and demonstrated improved agreement with experimental NMR relaxation rates. More specifically, it was shown that methyl spinning barriers where generally too high, and that a small decrease (obtained by fitting to coupled cluster quantum chemical calculations of isolated dipeptides) lead to increased spinning rates and better agreement with experimental data in both T4L and ubiquitin (***Hoffmann et al., 2018a***,b, ***2020***). Other effects might, however, contribute to the imperfect agreement between experiments and simulations. For example, the quantum calculations did not include information about how methyl spinning rates are influenced e.g. by side-chain packing or tunneling effects (***Chatfield and Wong, 2000***; ***Chatfield et al., 2004***).

A final source of error between experiment and simulation that may also prevent reweighting is the accuracy of the forward model used to calculate the relaxation rates from simulations. Here we have used a recently described approach to calculate deuterium relaxation rates from MD simulations (***Hoffmann et al., 2018a***,b). While the approach is highly accurate, there are some sources of errors including for example the treatment of tumbling motions, non-symmetry of partially deuterated methyl groups, finite sampling of the fastest dynamics (we save frames every $1\,ps$) and fitting of the correlation functions to a fixed number of exponentials.

In summary, we have presented ABSURDer and applied it to study motions in methyl containing residues in a folded protein. The approach is general and may be extended to other types of residues and nuclei, to other NMR parameters that depend on relaxation rates such as nuclear Overhauser enhancements, and even to data beyond those generated by NMR such as e.g. fluorescence correlation and neutron spin-echo spectroscopy.

Methods

Sampling

The starting structure for our MD simulations was the X-ray structure of the cystein-free T4L SER44GLY mutant (PDB 107L) (***Blaber et al., 1993***) where we changed GLY44 back to a serine. We performed three different sets of simulations: Three $5\,\mu s$ long ("MD Data" in all three analyses) and five $1\,\mu s$ long ("NMR Data" in the first synthetic data set) simulations, both using the AMBER ff99SB*-ILDN / TIP4P-2005 protein force field (***Lindorff-Larsen et al., 2010***; ***Hornak et al., 2006***; ***Best and Hummer, 2009***), and three $1\,\mu s$ long simulations using the AMBER ff15ipq / SPC/E_b (***Debiec et al., 2016***) force field ("NMR Data" in the second synthetic data set). We applied a modification to the methyl rotation barriers to both force fields as recently described (***Hoffmann et al., 2018a***, ***2020***). In the case of the AMBER ff15ipq simulation, we used the Amber input preparation module LEAP from AmberTools17 (***Case et al., 2017***) to set up the system and we afterwards converted the topology to GROMACS file format using ParmED (***Swails et al., 2010***).

All MD simulations were carried out with GROMACS v2018.1. We used a periodic truncated dodecahedron of $400\,nm^3$ volume as a simulation box, keeping 1.2 nm distance from the protein which was centered in the cell. We solvated the system with 12230 TIP4P-2005 (***Abascal and Vega, 2005***) water molecules and 12247 SPC/$E_b$water molecules, respectively, retaining crystal waters. To neutralize the system and simulate it in a physiological salt concentration, we also added $150\,mmol\,L^{-1}$ Na^+/Cl^- ions. We minimized the systems with 50000 steps of steepest descent and equilibrated them for $200\,ps$ in the *NPT* ensemble, using harmonic position restraints (force constants of $1000\,kJ\,mol^{-1}\,nm^{-2}$) on the heavy atoms of the protein. Equations of motion were integrated using the leap-frog algorithm. We set a cut-off of $1\,nm$ for the Van der Waals and Coulomb interactions and employed Particle-Mesh Ewald summation for long-range electrostatics (***Essmann et al., 1995***) with 4th order cubic interpolation and $0.16\,nm$ Fourier grid spacing. We ran the production simulations in the *NPT* ensemble, using the velocity re-scaling thermostat (***Bussi et al., 2007***) with a reference temperature of $300\,K$ and thermostat timescale of $\tau_T = 1\,ps$, and the Parrinello-Rahman barostat with a reference pressure of $1\,bar$, a barostat timescale $\tau_P = 2\,ps$ and an isothermal compressibility of $4.5 \times 10^{-5}\,bar^{-1}$. Finally, we saved the protein coordinates every $1\,ps$ and, after running the simulations, we removed the overall tumbling of the protein by fitting to a reference structure.

Spectral Density Mapping

We used a previously described approach (***Hoffmann et al., 2018b***) to calculate NMR relaxation rates from computed spectral densities. The original work used multiple $300\,ns$-long simulations, however, we analysed our simulations in a block-wise fashion, using $10\,ns$ long, non-overlapping blocks. This yielded 1500 blocks for the set of 3x $5\,\mu s$-long simulations, 500 blocks for the set of 5x $1\,\mu s$-long simulations and 300 blocks for the 3x $1\,\mu s$ long simulations. We started by calculating the internal time correlation function (TCF), *i.e.* without global tumbling, of the three C_{methyl}-H^i_{methyl} ($i = 1, 2, 3$) bond vectors up to a maximum lag-time of $5\,ns$ for all side-chain methyl groups. Next, we fitted the internal TCFs with six exponential functions and an offset,

$$C_{\mathrm{int,exp}}(t) = \sum_{i=1}^{6} A_i e^{-t/\tau_i} + S^2_{\mathrm{long}}, \tag{1}$$

where $\sum_{i=1}^{6} A_i + S^2_{\mathrm{long}} = 1, 0 \leq A_i \leq 1, 0 \leq S^2_{\mathrm{long}} \leq 1, \tau_i \geq 0$ $(i = 1, \ldots, 6)$ and S^2_{long} is the long-time limit order parameter. We note that one should not attribute physical meaning to these amplitudes and timescales (***Bremi et al., 1997***). Since the protein was found to be best represented by an axially symmetric tumbling model (***Hoffmann et al., 2018b***), we introduced methyl-specific global tumbling times $\tau_{R,i}$ by multiplying the internal TCF with a single-exponential thus, yielding the total TCF:

$$C(t) = C_{\mathrm{int,exp}}(t) e^{-t/\tau_{R,i}}. \tag{2}$$

In this case, we applied the experimental methyl group-specific tumbling times, which, for the synthetic data sets, we extracted based on analysing backbone N–H dynamics (***Maragakis et al.,***

2008; *Hoffmann et al., 2018b*). Briefly, we calculated TCFs for the backbone N-H bond vectors and fitted them to the three-parameter Liparo-Szabo model:

$$C(t) = S^2 \mathrm{e}^{-t/\tau_\mathrm{c}^\mathrm{eff}} + \left(1 - S^2\right) \mathrm{e}^{-t/\tau_\mathrm{red}} \tag{3}$$

with $\tau_\mathrm{red} = \left(\tau_\mathrm{c}^\mathrm{eff} \tau_\mathrm{f}\right) / \left(\tau_\mathrm{c}^\mathrm{eff} + \tau_\mathrm{f}\right)$ and where S^2, τ_f and $\tau_\mathrm{c}^\mathrm{eff}$ are used as fitting parameters. Next, we used the fitted rotational tumbling times and the initial structure of the protein, translated so that its center of mass is located at the origin (PDBinertia, http://comdnmr.nysbc.org), to calculate the principal axis frame (Quadric (*Lee et al., 1997*)). Finally, we extracted the principal values D_{xx}, D_{yy} and D_{zz} from the diffusion tensor and calculated the methyl-specific rotational diffusion constants $D_i = \mathrm{Tr}\mathscr{D}/3$ from which we calculated the methyl-specific tumbling times $\tau_{R,i} = 1/(6D_i)$.

After introducing $\tau_{R,i}$, we transformed the total TCF into a spectral density function:

$$J(\omega) = \sum_{i=1}^{6} \frac{A_i \tau_i^\mathrm{eff}}{1 + (\omega \tau_i^\mathrm{eff})^2} + \frac{S_\mathrm{long}^2 \tau_{\mathrm{R},i}}{1 + (\omega \tau_{\mathrm{R},i})^2}, \tag{4}$$

where $\tau_i^\mathrm{eff} = (\tau_i \tau_{\mathrm{R},i})/(\tau_i + \tau_{\mathrm{R},i})$. Finally, we calculated the relaxation rates directly from $J(0)$, $J(\omega_\mathrm{D})$, $J(2\omega_\mathrm{D})$ via

$$R\left(D_z\right) = \frac{1}{32} \left(\frac{qQe^2}{\hbar}\right)^2 \left[J\left(\omega_\mathrm{D}\right) + 4J\left(2\omega_\mathrm{D}\right)\right], \tag{5}$$

$$R\left(D_y\right) = \frac{1}{32} \left(\frac{qQe^2}{\hbar}\right)^2 \left[9J(0) + 15J\left(\omega_\mathrm{D}\right) + 6J\left(2\omega_\mathrm{D}\right)\right], \tag{6}$$

$$R\left(3D_z^2 - 2\right) = \frac{1}{32} \left(\frac{qQe^2}{\hbar}\right)^2 \left[3J\left(\omega_\mathrm{D}\right)\right], \tag{7}$$

where $\left(qQe^2/\hbar\right)^2$ is the quadrupolar coupling constant of deuterium. We used $145.851\,\mathrm{MHz}$ as the Larmor frequency ω_D of deuterium for all data sets which corresponds to a Bruker magnetic field strength of $22.3160\,\mathrm{T}$ that was also used to measure the experimental NMR relaxation rates.

We used this approach for each side-chain methyl group and each of the 1500 blocks to calculate sets of the three relaxation rates, which we employed as a input to the ABSURDer reweighting. When generating synthetic data, we calculated average rates over the blocks, and we used the standard error of the mean over the blocks as the errors when calculating χ^2 and for reweighting with ABSURDer.

Reweighting

Our reweighting approach builds upon the previously described ABSURD method (*Salvi et al., 2016*), where a longer set of trajectories are divided into *blocks*, each of which is long enough to estimate the relaxation rates of interest. Finally, a weighted average of the rates is performed over the blocks, where the weights, **w**, are obtained through the optimisation of a functional which determines the agreement between simulated and experimental rates. In particular, the ABSURD functional is given by

$$\mathcal{R}(\mathbf{w}) = \sum_{r=1}^{N_r} \sum_{n=1}^{M} \left[\frac{R_{r,n}^\mathrm{exp} - \sum_{i=1}^{N} w_i R_{r,n}^i}{\sqrt{(\sigma_{r,n}^\mathrm{exp})^2 + (\sigma_{r,n}^\mathrm{calc})^2}}\right]^2, \quad \sum_{i=1}^{N} w_i = 1 \tag{8}$$

where r ranges over the N_r types of experimental NMR relaxation rates, M is the overall number of measured rates and N represents the number of blocks which the trajectory has been cut into, $\sigma_{r,n}^\mathrm{exp}$ is the experimental error on the n-th rate of the r-th type, $\sigma_{r,n}^\mathrm{calc}$ is the standard error on the simulated rates, estimated by averaging all the n-th rates of the r-th type over the different blocks. The optimal set of weights is obtained as the corresponding minimum in weight space, $\mathbf{w}^* = \min_\mathbf{w} \mathcal{R}(\mathbf{w})$. Each weight is associated to a block from the trajectory and it encodes the relevance of each block to the dynamical ensemble.

The functional in Eq. 8 is unregularised, which sometimes may lead to overfitting (***Orioli et al., 2020***). In the original implementation of ABSURD, this was avoided by extensive cross-validation using multiple relaxation rates, measured at multiple magnetic fields. A complementary approach is to use the MD simulations as a statistical prior and to balance the trust in the data and the experiments (***Orioli et al., 2020***).

We take this latter approach by introducing an extension to the ABSURD functional (which we denominate ABSURDer) to make it less prone to overfitting. In particular, we add a Shannon relative-entropy regularization $S(\mathbf{w})$ (***Kullback and Leibler, 1951***; ***Gull and Daniell, 1978***) and the new functional takes the following form:

$$\mathcal{T}(\mathbf{w}) = \sum_{r=1}^{N_r}\sum_{n=1}^{M}\left[\frac{R_{r,n}^{\text{exp}} - \sum_{i=1}^{N} w_i R_{r,n}^i}{\sqrt{(\sigma_{r,n}^{\text{exp}})^2 + (\sigma_{r,n}^{\text{calc}})^2}}\right]^2 + \theta\sum_{i=1}^{N} w_i \log\left(\frac{w_i}{w_i^0}\right) = \chi^2(\mathbf{w}) - \theta S(\mathbf{w}) \tag{9}$$

where $w_i^0 = 1/N$ represent the initial, uniform weights provided by the prior. The parameter θ, is used to set the balance between our trust in the prior and in the experimental data. In particular, for $\theta \to \infty$, an infinite trust is put on the prior and the weights w_i stay at their initial values. In the opposite limit, $\theta \to 0$, all the trust is put on the experimental data and the weights are allowed to change freely to accommodate this trust. In this case, the functional $\mathcal{T}(\mathbf{w})$ reduces to that used in ABSURD. The two limits of θ can be seen also from parameter $\phi_{\text{eff}}(\mathbf{w}) = \exp(S(\mathbf{w}))$, which provides a measure of the effective fraction of blocks retained for a given set of weights. In particular, for $\theta \to \infty$ we have $\phi_{\text{eff}}(\mathbf{w}) = 1$, as $w_i \simeq w_i^0$. On the other hand, for $\theta \to 0$ and in the case where the prior is very different from the experimental data, $\phi_{\text{eff}}(\mathbf{w})$ may take on very small values.

We minimised functional $\mathcal{T}(\mathbf{w})$ using the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS-B) in its SciPy implementation (***Virtanen et al., 2020***). In each reweighting run we employed 30 values of θ selected in the range $[0, 8000]$. To accelerate convergence of the optimization, we started by optimising the $\theta = 8000$ functional using $\mathbf{w}_0$ as a guess set of weights, then we employed the obtained optimal weights as guess for the successive minimisation run and so on, iteratively. We stress that the value of $\mathbf{w}_0$ in Eq. 9 remained unchanged throughout the minimisation, regardless of the choice of the guess input weights. The ABSURDer code is available under a GNU GPL v3.0 license at github.com/KULL-Centre/ABSURDer and scripts and data to generate the results in this paper are available at github.com/KULL-Centre/papers/tree/master/2020/ABSURDer-kummerer-orioli-et-al-2020.

Acknowledgments

We are grateful to Profs. Lars V. Schäfer and Frans A.A. Mulder for discussions, help and comments on the manuscript. We acknowledge support by a grant from the Lundbeck Foundation to the BRAINSTRUC structural biology initiative (155-2015-2666, to K.L.-L.), the NordForsk Nordic Neutron Science Programme (to K.L.-L.), the Carlsberg Foundation (CF17-0491, to Y.G.), and the Novo Nordisk Foundation (NNF15OC0016360, to K.T. and K.L.-L.).

References

Abascal JL, Vega C. A general purpose model for the condensed phases of water: TIP4P/2005. The Journal of chemical physics. 2005; 123(23):234505.

Anderson JS, Hernández G, LeMaster DM. 13C NMR Relaxation Analysis of Protein GB3 for the Assessment of Side Chain Dynamics Predictions by Current AMBER and CHARMM Force Fields. Journal of Chemical Theory and Computation. 2020; 16(5):2896–2913.

Andrae R, Schulze-Hartung T, Melchior P. Dos and don'ts of reduced chi-squared. arXiv preprint arXiv:10123754. 2010; .

Bernardi RC, Melo MC, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochimica et Biophysica Acta (BBA)-General Subjects. 2015; 1850(5):872–877.

Best RB, Clarke J, Karplus M. What contributions to protein side-chain dynamics are probed by NMR experiments? A molecular dynamics simulation analysis. Journal of molecular biology. 2005; 349(1):185–203.

Best RB, Hummer G. Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides. The journal of physical chemistry B. 2009; 113(26):9004–9015.

Best RB, Vendruscolo M. Determination of protein structures consistent with NMR order parameters. Journal of the American Chemical Society. 2004; 126(26):8090–8091.

Blaber M, Zhang XJ, Matthews BW. Structural basis of amino acid alpha helix propensity. Science. 1993; 260(5114):1637–1640.

Bonomi M, Heller GT, Camilloni C, Vendruscolo M. Principles of protein structural ensemble determination. Current opinion in structural biology. 2017; 42:106–116.

Bottaro S, Bengtsen T, Lindorff-Larsen K. Integrating molecular simulation and experimental data: a Bayesian/maximum entropy reweighting approach. In: *Structural Bioinformatics* Springer; 2020.p. 219–240.

Bottaro S, Bussi G, Kennedy SD, Turner DH, Lindorff-Larsen K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. Science advances. 2018; 4(5):eaar8521.

Bottaro S, Lindorff-Larsen K. Biophysical experiments and biomolecular simulations: A perfect match? Science. 2018; 361(6400):355–360.

Bowman GR. Accurately modeling nanosecond protein dynamics requires at least microseconds of simulation. Journal of computational chemistry. 2016; 37(6):558–566.

Bremi T, Brüschweiler R, Ernst RR. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: application to phenylalanines in antamanide. Journal of the American Chemical Society. 1997; 119(18):4272–4284.

Brookes DH, Head-Gordon T. Experimental inferential structure determination of ensembles for intrinsically disordered proteins. Journal of the American Chemical Society. 2016; 138(13):4530–4538.

Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. The Journal of chemical physics. 2007; 126(1):014101.

Case D, Cerutti D, Cheatham III T, Darden T, Duke R, Giese T, Gohlke H, Goetz A, Greene D, Homeyer N, et al. AMBER 2017. San Francisco: University of California. 2017; .

Case DA. Molecular dynamics and NMR spin relaxation in proteins. Accounts of chemical research. 2002; 35(6):325–331.

Cesari A, Reißer S, Bussi G. Using the maximum entropy principle to combine simulations and solution experiments. Computation. 2018; 6(1):15.

Chatfield DC, Augsten A, D'cunha C. Correlation times and adiabatic barriers for methyl rotation in SNase. Journal of Biomolecular NMR. 2004; 29(3):377–385.

Chatfield DC, Wong SE. Methyl motional parameters in crystalline l-alanine: molecular dynamics simulation and NMR. The Journal of Physical Chemistry B. 2000; 104(47):11342–11348.

Chen Pc, Shevchuk R, Strnad FM, Lorenz C, Karge L, Gilles R, Stadler AM, Hennig J, Hub JS. Combined small-angle X-ray and neutron scattering restraints in molecular dynamics simulations. Journal of chemical theory and computation. 2019; 15(8):4687–4698.

Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. Journal of the American Chemical Society. 1990; 112(12):4989–4991.

Cordeiro TN, Chen Pc, De Biasio A, Sibille N, Blanco FJ, Hub JS, Crehuet R, Bernadó P. Disentangling polydispersity in the PCNA- p15PAF complex, a disordered, transient and multivalent macromolecular assembly. Nucleic acids research. 2017; 45(3):1501–1515.

Cousin SF, Kadeřávek P, Bolik-Coulon N, Gu Y, Charlier C, Carlier L, Bruschweiler-Li L, Marquardsen T, Tyburn JM, Brüschweiler R, Ferrage F. Time-resolved protein side-chain motions unraveled by high-resolution relaxometry and molecular dynamics simulations. Journal of the American Chemical Society. 2018; 140(41):13456–13465.

Crehuet R, Buigues PJ, Salvatella X, Lindorff-Larsen K. Bayesian-Maximum-Entropy reweighting of IDP ensembles based on NMR chemical shifts. Entropy. 2019; 21(9):898.

Debiec KT, Cerutti DS, Baker LR, Gronenborn AM, Case DA, Chong LT. Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. Journal of chemical theory and computation. 2016; 12(8):3926–3947.

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of chemical physics. 1995; 103(19):8577–8593.

Gu Y, Li DW, Brüschweiler R. NMR order parameter determination from long molecular dynamics trajectories for objective comparison with experiment. Journal of chemical theory and computation. 2014; 10(6):2599–2607.

Gull SF, Daniell GJ. Image reconstruction from incomplete and noisy data. Nature. 1978; 272(5655):686–690.

van Gunsteren WF, Daura X, Hansen N, Mark AE, Oostenbrink C, Riniker S, Smith LJ. Validation of molecular simulation: an overview of issues. Angewandte Chemie International Edition. 2018; 57(4):884–902.

Halle B, Wennerström H. Interpretation of magnetic resonance data from water nuclei in heterogeneous systems. The Journal of Chemical Physics. 1981; 75(4):1928–1943.

Henriques J, Cragnell C, Skepö M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. Journal of chemical theory and computation. 2015; 11(7):3420–3431.

Hoffmann F, Mulder FA, Schäfer LV. Predicting NMR relaxation of proteins from molecular dynamics simulations with accurate methyl rotation barriers. The Journal of Chemical Physics. 2020; 152(8):084102.

Hoffmann F, Mulder FA, Schäfer LV. Accurate methyl group dynamics in protein simulations with AMBER force fields. The Journal of Physical Chemistry B. 2018; 122(19):5038–5048.

Hoffmann F, Xue M, Schäfer LV, Mulder FA. Narrowing the gap between experimental and computational determination of methyl group dynamics in proteins. Physical Chemistry Chemical Physics. 2018; 20(38):24577–24590.

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics. 2006; 65(3):712–725.

Hu H, Hermans J, Lee AL. Relating side-chain mobility in proteins to rotameric transitions: insights from molecular dynamics simulations and NMR. Journal of biomolecular NMR. 2005; 32(2):151–162.

Hummer G, Köfinger J. Bayesian ensemble refinement by replica simulations and reweighting. The Journal of chemical physics. 2015; 143(24):12B634_1.

Köfinger J, Stelzl LS, Reuter K, Allande C, Reichel K, Hummer G. Efficient ensemble refinement by reweighting. Journal of chemical theory and computation. 2019; 15(5):3390–3401.

Kullback S, Leibler RA. On information and sufficiency. The annals of mathematical statistics. 1951; 22(1):79–86.

Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, Lindorff-Larsen K. Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. PLoS computational biology. 2020; 16(4):e1007870.

Lee LK, Rance M, Chazin WJ, Palmer AG. Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13Cα nuclear spin relaxation. Journal of biomolecular NMR. 1997; 9(3):287–298.

Li DW, Brüschweiler R. NMR-based protein potentials. Angewandte Chemie International Edition. 2010; 49(38):6778–6780.

Li DW, Brüschweiler R. Iterative optimization of molecular mechanics force fields from NMR data of full-length proteins. Journal of chemical theory and computation. 2011; 7(6):1773–1782.

Liao X, Long D, Li DW, Brüschweiler R, Tugarinov V. Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group. The Journal of Physical Chemistry B. 2012; 116(1):606–620.

Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature. 2005; 433(7022):128–132.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function, and Bioinformatics. 2010; 78(8):1950–1958.

Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. Journal of the American Chemical Society. 1982; 104(17):4546–4559.

Lipari G, Szabo A, Levy RM. Protein dynamics and NMR relaxation: comparison of simulations with experiment. Nature. 1982; 300(5888):197–198.

Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, Arkin IT, Jensen MØ, Xu H, Trbovic N, Friesner RA, et al. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. The Journal of Physical Chemistry B. 2008; 112(19):6155–6158.

McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977; 267(5612):585–590.

Ming D, Brüschweiler R. Prediction of methyl-side chain dynamics in proteins. Journal of biomolecular NMR. 2004; 29(3):363–368.

Nerenberg PS, Head-Gordon T. New developments in force fields for biomolecular simulations. Current opinion in structural biology. 2018; 49:129–138.

Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. Journal of the American Chemical Society. 2009; 131(49):17908–17918.

Norgaard AB, Ferkinghoff-Borg J, Lindorff-Larsen K. Experimental parameterization of an energy function for the simulation of unfolded proteins. Biophysical journal. 2008; 94(1):182–192.

O'Brien ES, Wand AJ, Sharp KA. On the ability of molecular dynamics force fields to recapitulate NMR derived protein side chain order parameters. Protein Science. 2016; 25(6):1156–1160.

Orioli S, Larsen AH, Bottaro S, Lindorff-Larsen K. Chapter Three - How to learn from inconsistencies: Integrating molecular simulations with experimental data. In: Strodel B, Barz B, editors. *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*, vol. 170 of Progress in Molecular Biology and Translational Science Academic Press; 2020.p. 123 – 176. http://www.sciencedirect.com/science/article/pii/S1877117319302121, doi: https://doi.org/10.1016/bs.pmbts.2019.12.006.

Palmer III AG. Probing molecular motion by NMR. Current opinion in structural biology. 1997; 7(5):732–737.

Piana S, Robustelli P, Tan D, Chen S, Shaw DE. Development of a force field for the simulation of single-chain proteins and protein-protein complexes. Journal of Chemical Theory and Computation. 2020; .

Prompers JJ, Brüschweiler R. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. Journal of the American Chemical Society. 2002; 124(16):4522–4534.

Qian H. Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. Physical Review E. 2001; 63(4):042103.

Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmüller H. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. Journal of chemical theory and computation. 2015; 11(11):5513–5524.

Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. Proceedings of the National Academy of Sciences. 2018; 115(21):E4758–E4766.

Salvi N, Abyzov A, Blackledge M. Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. The journal of physical chemistry letters. 2016; 7(13):2483–2489.

Salvi N, Abyzov A, Blackledge M. Solvent-dependent segmental dynamics in intrinsically disordered proteins. Science advances. 2019; 5(6):eaax2348.

Showalter SA, Johnson E, Rance M, Brüschweiler R. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. Journal of the American Chemical Society. 2007; 129(46):14146–14147.

Skrynnikov NR, Millet O, Kay LE. Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions. Journal of the American Chemical Society. 2002; 124(22):6449–6460.

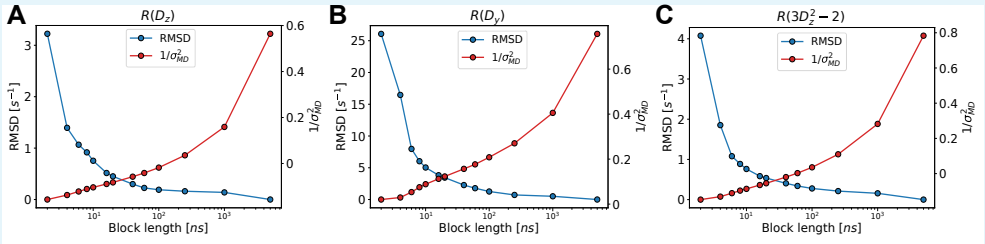
Swails J, Hernandez C, Mobley D, Nguyen H, Wang L, Janowski P, ParmEd; 2010.

Takemura K, Kitao A. Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. The Journal of Physical Chemistry B. 2012; 116(22):6279–6287.

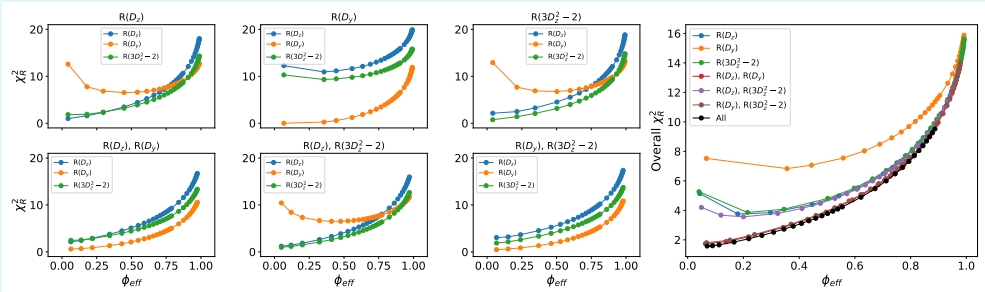
Vasile F, Tiana G. Determination of structural ensembles of flexible molecules in solution from NMR data undergoing spin diffusion. Journal of chemical information and modeling. 2019; 59(6):2973–2979.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Jarrod Millman K, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey C, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020; 17:261–272. doi: https://doi.org/10.1038/s41592-019-0686-2.

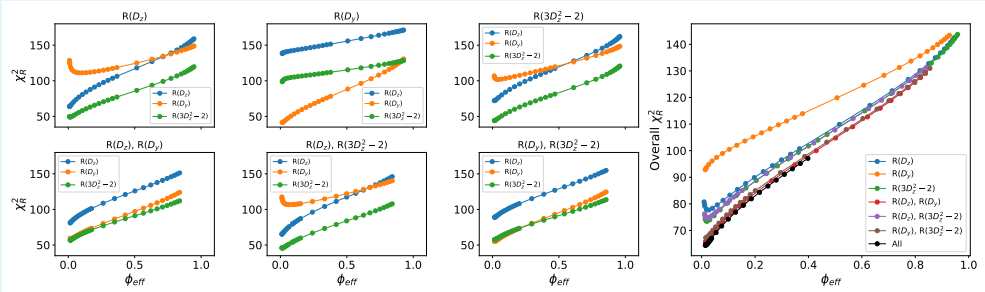
Supporting Material



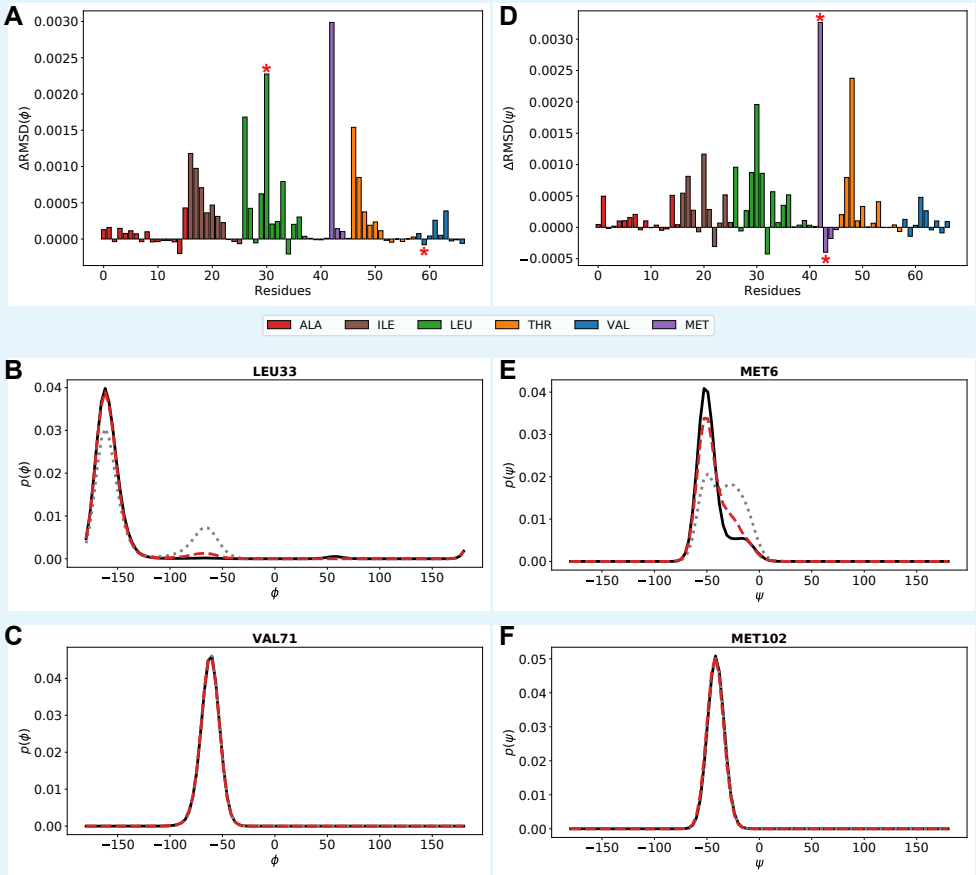
Supporting Figure 1. Assessment of convergence and effect of block size. RMSD of NMR relaxation rates (blue) and average inverse variance with respect to the full MD dataset (red) is shown as a function of the block length.



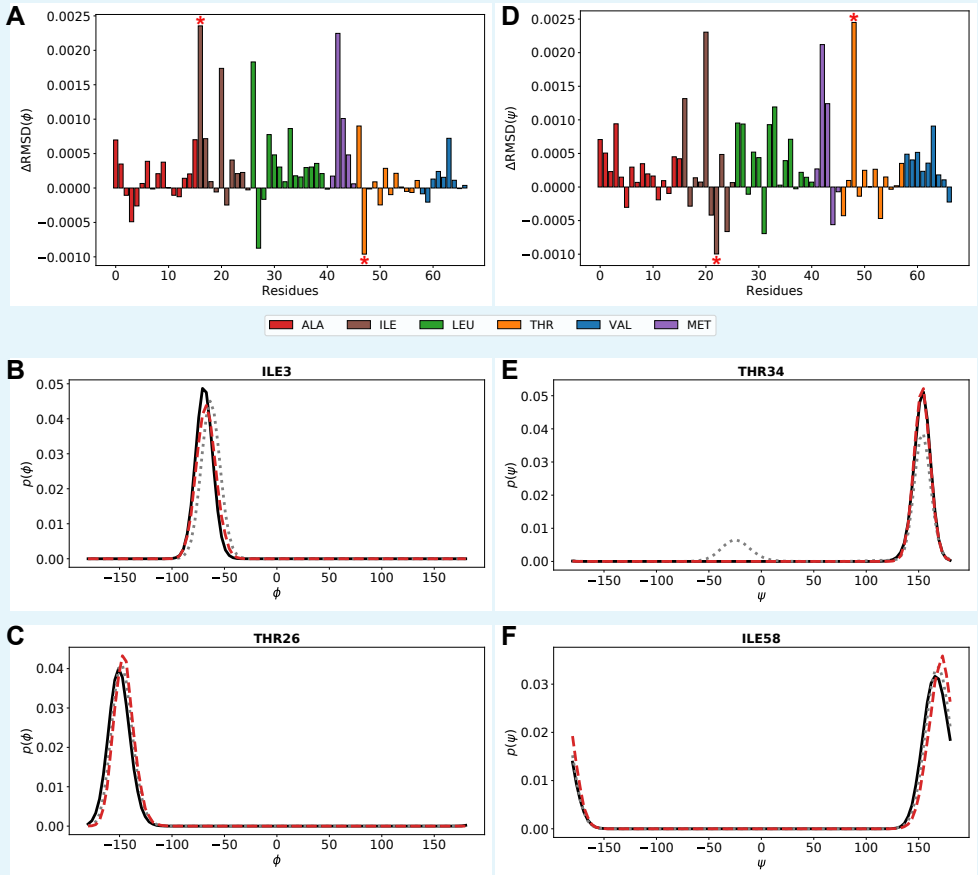
Supporting Figure 2. Cross validation of ABSURDer reweighting when synthetic data were generated using the Amber ff99SB*-ILDN force field. The panels show $\chi^2_R(\phi_{\text{eff}})$ curves for each of the three types of NMR relaxation rates. Each panel differs by which data was used in reweighting (label above panel), and we show the results using all six possible combinations of the three rates with the large panel corresponding to all three rates.



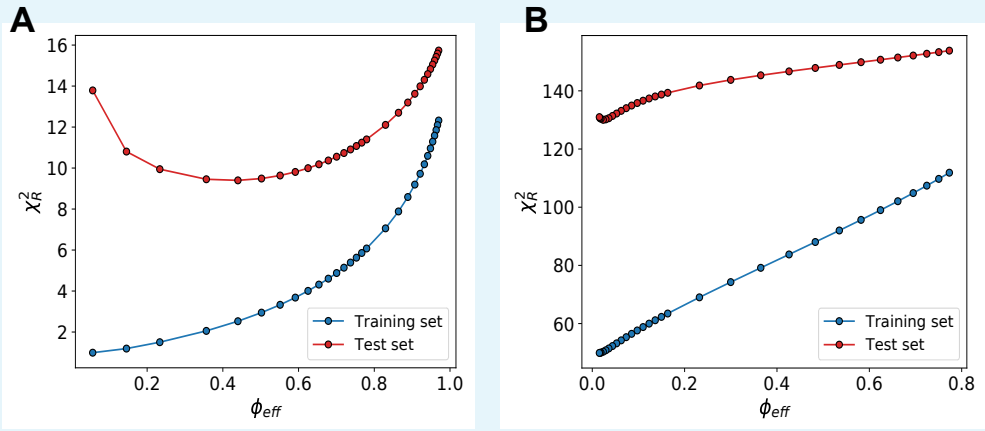
Supporting Figure 3. Cross validation of ABSURDer reweighting when synthetic data were generated using the Amber ff15ipq force field. The panels show $\chi^2_R(\phi_{\text{eff}})$ curves for each of the three types of NMR relaxation rates. Each panel differs by which data was used in reweighting (label above panel), and we show the results using all six possible combinations of the three rates with the large panel corresponding to all three rates.



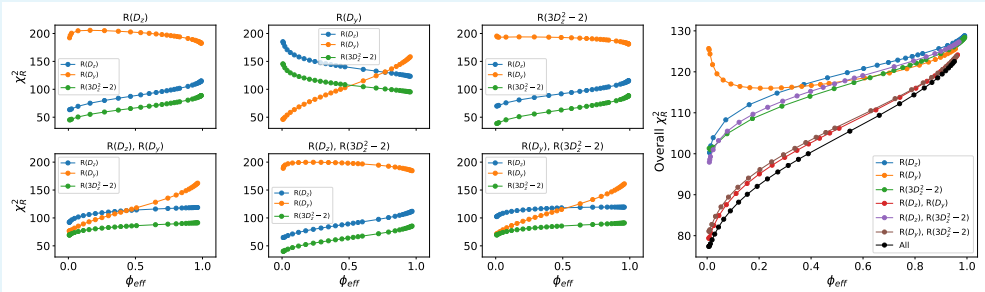
Supporting Figure 4. Effect of ABSURDer reweighting on backbone dihedral angle distributions. Difference RMSD (ΔRMSD) between distributions for the backbone (A) ϕ and (B) ψ dihedral angles. We calculated the RMSD of the dihedral angle distributions before and after reweighting, in both cases using the distribution used to generate the synthetic data generated by Amber ff99SB*-ILDN as the reference. Thus, large positive values of ΔRMSD indicate highly improved distributions after reweighting whereas negative values indicate distributions worsened by the application of ABSURDer. The bars are colored by residue type and asterisks mark residues highlighted in panels C–F. Examples of residues for which ABSURDer (C,D) substantially improve the dihedral angle distributions and (E,F) worsens them.



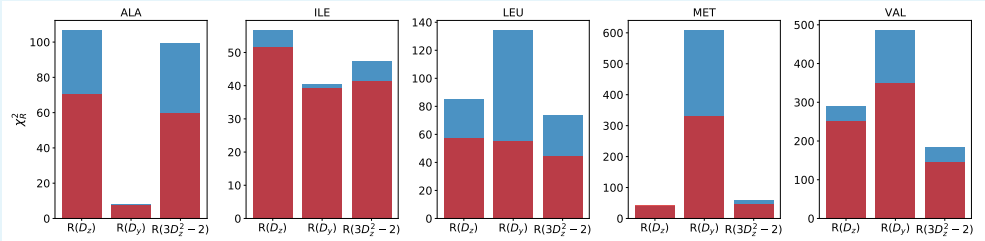
Supporting Figure 5. Effect of ABSURDer reweighting on backbone dihedral angle distributions. Difference RMSD (ΔRMSD) between distributions for the backbone (A) ϕ and (B) ψ dihedral angles. We calculated the RMSD of the dihedral angle distributions before and after reweighting, in both cases using the distribution used to generate the synthetic data generated by Amber ff15ipq as the reference. Thus, large positive values of ΔRMSD indicate highly improved distributions after reweighting whereas negative values indicate distributions worsened by the application of ABSURDer. The bars are colored by residue type and asterisks mark residues highlighted in panels C–F. Examples of residues for which ABSURDer (C,D) substantially improve the dihedral angle distributions and (E,F) worsens them.



Supporting Figure 6. Effect of ABSURDer reweighting on NMR relaxation rates not included in the optimisation. We used ABSURDer with synthetic data to fit the relaxation rates for the (blue) 73 methyl groups that have been probed experimentally, and (red) cross-validate across the remaining 27 methyl groups. (A) Synthetic data generated with Amber ff99SB*-ILDN. (B) Synthetic data generated with Amber ff15ipq.



Supporting Figure 7. Effect of ABSURDer reweighting on experimental data when only a subset of NMR relaxation rate types are used. $\chi^2_R(\phi_{\text{eff}})$ curves for the three kinds of rate separately obtained by fitting with respect to all the possible combinations of rate types (6 panels on the left) and the corresponding average (overall χ^2_R, right panel).



Supporting Figure 8. Effect of ABSURDer on the different residue types. Residue specific χ^2_R between calculated and experimental NMR relaxation rates before (blue) and after (red) reweighting. The results are obtained by averaging the residuals over the 15 ALA, 20 ILE, 5 MET, 14 VAL and 20 LEU residues. Note that each panel uses different scales for χ^2_R.

Supporting Table 1. Sum of the weights for blocks in each of the three independent $5\,\mu s$ long trajectories.

	Trajectory 1	Trajectory 2	Trajectory 3
Same force fields	0.21	0.5	0.29
Different force fields	0.28	0.45	0.27
Experimental	0.38	0.28	0.34