DEER-PREdict: Software for Efficient Calculation of Spin-Labeling EPR and NMR Data from Conformational Ensembles

Giulio Tesei[1,◕], João M. Martins[1,◕], Micha B. A. Kunze[1], Yong Wang[1], Ramon Crehuet[1,2], Kresten Lindorff-Larsen[1,*]

1 Structural Biology and NMR Laboratory & the Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen N, Denmark **2** CSIC-Institute for Advanced Chemistry of Catalonia (IQAC), Jordi Girona 18-26, 08034, Barcelona, Spain

◕These authors contributed equally to this work.

* lindorff@bio.ku.dk

Abstract

Owing to their plasticity, intrinsically disordered and multidomain proteins require descriptions based on multiple conformations, thus calling for techniques and analysis tools that are capable of dealing with conformational ensembles rather than a single protein structure. Here, we introduce DEER-PREdict, a software to predict Double Electron-Electron Resonance distance distributions as well as Paramagnetic Relaxation Enhancement rates from ensembles of protein conformations. DEER-PREdict uses an established rotamer library approach to describe the paramagnetic probes which are bound covalently to the protein. DEER-PREdict has been designed to operate efficiently on large conformational ensembles, such as those generated by molecular dynamics simulation, to facilitate the validation or refinement of molecular models as well as the interpretation of experimental data. The performance and accuracy of the software is demonstrated with experimentally characterized protein systems: HIV-1 protease, T4 Lysozyme and Acyl-CoA-binding protein. DEER-PREdict is open source

(GPLv3) and available at github.com/KULL-Centre/DEERpredict and as a Python PyPI package pypi.org/project/DEERPREdict.

Introduction

A detailed understanding of protein function often requires an accurate description of the structure and dynamics of a protein. The characterization of protein complexes as well as multi-domain and disordered proteins is typically achieved by combining experimental techniques of distinct spacial resolution [1]. Among the many different experimental techniques may be used, we focus here on (i) a pulsed electron paramagnetic resonance (EPR) technique called double electron-electron resonance (DEER) and (ii) a nuclear magnetic resonance (NMR) method called paramagnetic relaxation enhancement (PRE). While the two methods differ substantially in their physics and applications, they have in common that they generally involve adding so-called spin-labels to the protein of interest.

DEER, also sometimes known as pulsed electron-electron double resonance (PELDOR), [2–6] relies on probing magnetic dipole-dipole interactions that are sensitive to distributions of residue-residue distances ranging from ∼2 to ∼6 nm [7,8]. For proteins, DEER generally requires site-directed spin labeling (SDSL) to functionalize a pair of selected residues with paramagnetic probes, e.g. 1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl methanethiosulfonate (MTSSL) [4].

PRE NMR also makes use of SDSL to provide information on the average proximity of protein backbone nuclei up to ∼3.5 nm away from the unpaired electron of the paramagnetic probe [9]. The dependence of the rate of relaxation enhancement on the electron-proton distance, r, scales as $\langle r^{-6} \rangle$, making the measurement particularly sensitive to contributions from different probe conformations [9].

Since spin labels are conformationally dynamic, both protein and paramagnetic probes need to be described by conformational ensembles to obtain accurate predictions of DEER and PRE observables from molecular models [10,11].

Molecular dynamics (MD) simulations are one approach to obtain conformational ensembles that model the structure and dynamics of spin-labels for the calculation of EPR and NMR data [12–15]. While such analyses can provide unique insight into the

motions of and interactions between protein and spin-label [16], they may be relatively expensive computationally. Further, many studies integrate results from multiple probe positions, or pairs thereof, which may be difficult to represent in a single MD simulation with explicit representations of the probes.

Another approach is to use conformational analysis of the spin-label combined with modelling of the dynamics [17–20]. Such analyses suggest that the conformational variation of spin-labelled sites is rotameric, i.e. it can be relatively well described by a finite number of defined structures. Thus, in the calculation of DEER data, rapid modeling of dynamic paramagnetic probes was made possible with the introduction of the rotamer library approach (RLA) applied to the MTSSL probe by Polyhach *et al.* [21].

Here, building and expanding on earlier work [3, 21–24], we developed a software tool for fast predictions of DEER and PRE observables from large conformational ensembles using the RLA. We present our implementation, distributed as the DEER-PREdict software, and test it against experimental data on HIV-1 Protease, T4 Lysozyme and the Acyl-CoA-Binding Protein. This software has been previously used for the calculation of both intra- and intermolecular DEER and PRE NMR data [25, 26], and has some overlap with the features in RotamerConvolveMD [22] (github.com/MDAnalysis/RotamerConvolveMD). DEER-PREdict is open-source, documented (deerpredict.readthedocs.io) and open to contributions from the community.

Design and Implementation

DEER-PREdict is written in Python and is available as a Python API, which facilitates its integration within larger data pipelines. Predictions of DEER and PRE data are carried out via the DEERpredict and PREpredict classes. Both classes are initialized with protein structures (provided as MDAnalysis [27] Universe objects) and spin-labeled positions (residue numbers and chain IDs). As shown in the *Results* section, the calculations are triggered by the *run* function, which also sets additional attributes such as the paths of input and output files as well as experiment-specific parameters. Per-frame data is saved in compressed binary files (HDF5 and pickle files) to allow for fast calculations of ensemble averages in reweighting schemes.

For the presented software, we adopt a procedure of rotamer placement and evaluation of labeled sites which is analogous to the RLA of Polyhach *et al.* [21], and we build on this previous work to implement fast calculations of DEER and PRE observables from large structural ensembles, such as MD trajectories.

Rotamer Library Approach

Rotamer libraries have a long history in protein structural analysis [28], with an early being to study side-chain packing [29]. Several other applications of this approach were later employed, e.g. in homology modeling and protein design [30, 31]. In our implementation, the RLA is used to insert the rotamer conformations of a paramagnetic probe at the spin-labeled site and to calculate the Boltzmann weight of each conformer. By default, we use the MTSSL 175 K rotamer library by Polyhach et al. [21], which was filtered to include only the $\chi_1\chi_2$ conformations that are most commonly found in crystal structures of T4 Lysozyme [32]. As shown by Klose *et al.* [23], this selection criterion increases the accuracy of the calculated electron-electron distance distributions. The code is, however, general and it is possible to add other rotamer libraries for MTSSL or other spin-labels.

Following the alignment of the rotamer to the backbone heavy atoms, the calculation of the Boltzmann weights is based on the sum of internal, ϵ_i^{int}, and external, ϵ_i^{ext}, energy contributions. The internal contribution is taken from Polyhach et al. [21] and results from the clustering of representative dihedral combinations from MD simulations. The normalized frequency of each cluster throughout the MD trajectory was used to determine the Boltzmann probability, p_i^{int}, of a given i^{th} state, which was readily converted into the internal energy contribution, ϵ_i^{ext}, via Boltzmann inversion. On the other hand, the external energy contribution is calculated on the fly as the dispersion interaction energy between heavy atoms of rotamer and protein residues within a 1-nm cutoff, using the pairwise 6-12 Lennard-Jones potential of the CHARMM36 force field.

The overall probability of the ith rotamer state is then calculated as

$$p_i = p_i^{int} p_i^{ext} = p_i^{int} \frac{exp\left(-\epsilon_i^{ext}/kT\right)}{Z} \tag{1}$$

where $Z = \sum_i exp\left(-\epsilon_i^{ext}/kT\right)$ is the steric partition function quantifying the fit of the

rotamer in the embedding protein conformation. Low values of Z result from large probe-protein van der Waals interaction energies, suggesting a tight placement of the spin label either due to a displacement of the rotamers or indicative of a wild-type conformation made inaccessible by the presence of the MTSSL probe. Especially in folded proteins, probes located in closely packed regions are likely to induce changes in the ensemble of the spin-labeled protein compared to the native form, and should be avoided in designing SDSL experiments. Therefore, in the calculation of DEER or PRE NMR observables, frames with $Z < 0.05$ are discarded to preclude spurious conformers from contributing to the ensemble average [21]. For the MTSSL 175 K rotamer library, a Z cutoff of 0.05 is compatible with distributions of ϵ_i^{ext} values where at most one of the 46 rotamers has $\epsilon_i^{ext} \approx 3\ k_BT$ while the rest has $\epsilon_i^{ext} \leq 7\ k_BT$. We observed that the results shown in this paper are virtually insensitive to the choice of the Z cutoff between 0.05 and 0.5 (see Figure S1).

Predicting the DEER signal from structural ensembles

Electron-electron distance distributions extracted from DEER experiments have previously routinely been compared with distributions predicted using the RLA implemented in the Matlab-based MMM modeling toolbox (http://www.epr.ethz.ch/software) [10, 33]. Since MMM intrinsically operates on single structures, we and others had to resort to wrapper scripts to compute distance distributions of large ensembles, such as MD trajectories [3, 22, 34]. With the program presented herein, we provide a tool to conveniently predict DEER distance distributions from large conformational ensembles, which can be easily integrated in reweighting schemes such as the Bayesian/Maximum entropy procedure [1, 11, 35, 36].

For each trajectory frame or conformation of a given ensemble, the rotamers from the library are placed at the spin-labeled position (Figure 1A) and the distances between all pair combinations of N-O paramagnetic centers are calculated. The resulting matrix of pair-wise distances is then used to compute the distance distribution weighted by the combined probability of each probe conformation, $p_i \times p_j$, with p_i and p_j being the conformation probabilities of rotamers i and j. After averaging over all the frames, a Fast Fourier Transform low pass filter is applied to the distance distribution

for noise reduction [37], resulting in a smooth distribution curve even for the analysis of a single protein conformation (Figure 1B). The average over the trajectory frames can be weighted by a user-specified list of weights e.g. to remove the bias from enhanced sampling simulations.

The dipolar modulation signal can be back-calculated from the distance distribution, $P(r)$, via the following integral [38]

$$S(t) = \int_0^\infty \mathrm{d}r\ P(r)K(r,t). \tag{2}$$

$K(r,t)$ is the DEER kernel

$$K(r,t) = \sqrt{\frac{\pi}{6\omega t}}\left[\cos(\omega)\,\mathrm{Fr\,C}\left(\sqrt{\frac{6\omega t}{\pi}}\right) + \sin(Dt)\,\mathrm{Fr\,S}\left(\sqrt{\frac{6\omega t}{\pi}}\right)\right] \tag{3}$$

where FrC and FrS are Fresnel cosine and sine functions, and ω is the dipolar frequency

$$\omega = \frac{\mu_0}{4\pi\hbar}\frac{\mu_B^2 g^2}{r^3} \tag{4}$$

where μ_0 is the permeability of free space, μ_B is the Bohr magneton and g is the electron g-factor. The experimental DEER time trace can be directly compared with

$$V(t) = 1 + \lambda\left[S(t) - 1\right] \tag{5}$$

where $0 \leq \lambda \leq 1$ quantifies the efficiency of the DEER pump pulse [7].

Prediction of PRE rates and intensity ratios

In analogy to the calculations of electron-electron distances to predict DEER distributions, we extended the use of the RLA to electron-proton separations to improve the accuracy of PRE predictions.

We focus here is on PRE NMR experiments that probe the increase in transverse relaxation rates of any backbone proton due to the dipolar interaction with the unpaired electron of the paramagnetic probe:

$$R_2^{ox} = R_2^{red} + \Gamma_2 \tag{6}$$

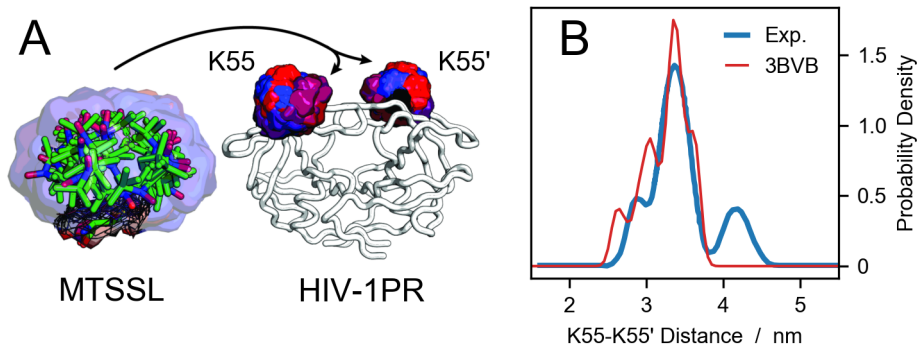


Fig 1. Probe placement scheme and comparison to DEER data. (A) A pool of 46 conformations of the MTSSL probe from the rotamer library are aligned to the backbone of residues K55 and K55' of HIV-1 protease. The color code represent the Boltzmann weights of each rotamer, increasing from blue to red. (B) Electron-electron distance distribution for HIV-1 protease spin labeled at residues K55 and K55'. The blue line is the experimental data from Torbeev *et al.* [39] whereas the red line is the prediction using DEER-PREdict and a crystal structure of HIV-1 protease (PDB code 3BVB).

where R_2^{ox} and R_2^{red} are the transverse relaxation rates in the presence of the spin label in the oxidized or reduced (diamagnetic) state, respectively. We note that it is also possible to measure PREs on other atoms and to probe longitudinal relaxation enhancement, and it would be possible to include such measurements in future versions of DEER-PREdict.

A description of the enhancement of the transverse relaxation due to dipole-dipole interactions in paramagnetic solutions was first proposed by Solomon and Bloembergen [40, 41]

$$\Gamma_2 = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 g^2 \mu_B^2 s_e \left(s_e + 1\right)\left[4J(0) + 3J(\omega_I)\right], \tag{7}$$

where γ_I and ω_I are the gyromagnetic ratio and the Larmor frequency of the proton, respectively, whereas s_e is the paramagnetic spin number, equal to 1/2 for nitroxide probe systems. The spectral density function $J(\omega_I)$ can be described using a model-free formalism [42–45], which takes into account the overall molecular tumbling in the external magnetic field as well as the internal motion of the spin label:

$$J(\omega_I) = \left\langle r^{-6} \right\rangle \left[\frac{S^2\tau_c}{1+\omega_I^2\tau_c^2} + \frac{(1-S^2)\tau_t}{1+\omega_I^2\tau_t^2}\right] \tag{8}$$

where

$$\tau_c = \left(\frac{1}{\tau_r} + \frac{1}{\tau_s}\right)^{-1} \tag{9}$$

and

$$\tau_t = \left(\frac{1}{\tau_r} + \frac{1}{\tau_s} + \frac{1}{\tau_i}\right)^{-1}. \tag{10}$$

τ_r is the rotational correlation time of the protein, τ_s is the effective electron correlation rate and τ_i is the correlation time of the internal motion (effective correlation time of the spin label). For MTSSL probes, $\tau_s \gg \tau_r$ and $\tau_c \approx \tau_r$ [46]. The value of τ_c depends on protein size and structure and is generally of the order of 1–10 ns [24, 47–50]. For τ_i, values between 100 to 500 ps can be assumed, based on e.g. ^{15}N spin relaxation rates and MD simulations [51, 52]. In general τ_c and τ_i can be specified as user input in DEER-PREdict. For the generalized order parameter, S, we use the factorization into contributions from radial and angular internal motions introduced by Brüschweiler *et al.* [44], $S^2 = S^2_{radial} S^2_{angular}$. The expressions for S^2_{radial} and $S^2_{angular}$ were derived from a jump model that treats the N conformers of the rotamer library as N discrete states [24, 45].

$$S^2_{radial} = \frac{\langle r^{-3} \rangle^2}{\langle r^{-6} \rangle} \tag{11}$$

where r is the proton-electron distance and the brackets denote averages over the conformers weighted by the respective Boltzmann weights, p_i, i.e. $\langle r^{-3} \rangle = \sum_i^N r_i^{-3} p_i$ and $\langle r^{-6} \rangle = \sum_i^N r_i^{-6} p_i$.

$$S^2_{angular} = \left\langle \frac{3}{2}\cos^2\Omega - \frac{1}{2} \right\rangle = \sum_{i,j}^{N} \left[\frac{3}{2} \left(\frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_i r_j} \right)^2 - \frac{1}{2} \right] p_i p_j \tag{12}$$

where Ω is the angle between the vectors $\mathbf{r_i}$ and $\mathbf{r_j}$, connecting a backbone proton with the ith and jth rotamer states, respectively. The relaxation enhancement rate for a single protein structure is calculated using Equation 7, and assuming that the motion of the paramagnetic label is much faster than the protein conformational changes, the ensemble average is estimated as

$$\langle \Gamma_2 \rangle = \sum_k^M w_k \Gamma_{2,k}. \tag{13}$$

where M is the number of configurations or frames of the simulation trajectory. In the case of unbiased simulations, the statistical weights, w_l, are simply $1/M$. Optionally, a list of weights can be provided by the user, e.g. to reweight a biased MD simulation [53, 54] or to incorporate the prediction of the PRE rates into a Bayesian/Maximum entropy reweighting scheme [1].

For samples with particularly high PRE rates it can be infeasible to obtain Γ_2 from multiple time-point measurements [55]. In such and other cases, the PRE is sometimes probed indirectly from the ratio of the peak intensities in ^{1}H,^{15}N-HSQC spectra of the spin-labeled protein in the oxidized and reduced state. Assuming that the intensity of the proton magnetization decays exponentially — by transverse relaxation only — during the total INEPT time of the HSQC measurement [56], t_d, the intensity ratio is estimated as

$$\frac{I_{para}}{I_{dia}} = \frac{R_2^{red} \exp\left(-\Gamma_2 t_d\right)}{R_2^{red} + \Gamma_2}. \tag{14}$$

Requirements and Installation

The main requirements are Python 3.6–3.8 and MDAnalysis 1.0 [27, 57]. In an environment with Python 3.6–3.8, DEER-PREdict can readily be installed through the package manager PIP by executing

```
pip install DEERPREdict
```

Package Stability

Tests reproducing DEER and PRE data for the protein systems studied in this article, as well as for a nanodisc [26], are performed automatically using Travis CI (travis-ci.org) every time the code is modified on the GitHub repository. The same tests can also be run locally using the test running tool pytest.

Results

In the following, we present applications of our tool to the prediction of DEER distance distributions and PRE intensity ratios of three folded proteins.

The code snippets reported in this section pertain to DEER-PREdict version 0.1.4.

A Jupyter Notebook to reproduce the results shown below (*article.ipynb*) can be found in the *tests* folder on the GitHub repository. The up-to-date documentation is available at deerpredict.readthedocs.io.

Case study1: DEER data for HIV-1 Protease

HIV-1 protease (HIV-1PR) is a homodimeric aspartic hydrolase involved in the cleavage of the gag-pol polyprotein complex. The inhibition of this process affects the life cycle of the HIV-1 virus, rendering it noninfectious [58]. The HIV-1PR monomer is composed of 99 residues and presents a structurally stable core region (residues 1-43 and 58-99) and a dynamic region characterized by a β-hairpin turn, called the flap (residues 44-57). The active site is located at the interstice between the core regions of the two monomers, in proximity to the catalytic D25 residues. This cavity is closed off by the dynamic flap regions, which are considered to act as a gate controlling the access to the active site. The dynamics of the flap regions are of utmost importance for the development of inhibitors, and have been extensively studied, both experimentally and *in silico* [39, 59–64]. Based on the relative position of the flaps, three main conformational states have been proposed. In X-ray crystallography, the closed state is typically observed for the ligand-bound enzyme (e.g. PDB codes 3BVB [65] and 2BPX [66]), the semi-open state is predominant for the apo form (e.g. PDB code 1HHP [67]) whereas the wide-open state has been observed for variants (e.g. PDB codes 1TW7 [68] and 1RPI [69]) [64]. In DEER measurements, these conformational states can be resolved by spin-labeling sites K55 and K55' (see Figure 1).

In Figure S2, experimental distance distribution and echo intensity signals obtained by Torbeev *et al.* [39] for the apo enzyme are compared with the corresponding predictions of our software based on PDB codes 2BPX, 3HVP and 1TW7. In the experimental distance distribution, the main peak at ~ 3.3 nm corresponds to the closed state whereas the second peak between 4 and 5 nm is characteristic of the wide-open state. The shoulder peak at ~ 2.8 nm has been identified as an open-like state known as the curled/tucked conformation [8, 70, 71]. Although we used single X-ray crystal structures, Figure S2 shows that the RLA results in multimodal distance distributions. For example, the K55-K55' separation between the ammonium groups in

1TW7 is ~ 3.6 nm, which is consistent with the semi-open conformation, however, the distances between the nitroxide groups of the MTSSL conformers span a wide range between 3.3 and 4.4 nm. The echo intensity curves are calculated using Equation 5, where the λ is estimated to 0.0922 by fitting the experimental dipolar evolution function to the corresponding curve derived from the experimental $P(r)$ via Equation 2.

To assess the predictive ability of DEER-PREdict, we generated conformational ensembles of the HIV-1PR homodimer via two different approaches: (a) a single 500-ns unbiased MD simulation, and (b) four independent 125-ns MD simulations restrained with experimental residual dipolar couplings (RDC) data [53, 72] from Roche *et al.* [60, 61] (see SI for methodological details). The initial configuration of our simulations is the X-ray crystal structure of the active-site mutant D25N bound to the inhibitor Darunavir (PDB code 3BVB).

Figure 2 presents a comparison of experimental DEER distance distributions and echo intensity curves with predictions from simulation trajectories of 1,000 frames sampled every 0.5 ns. For a single trajectory, the analysis is performed in 40 s on a 1.7 GHz processor by running the following code:

```
import MDAnalysis
from DEERPREdict.DEER import DEERpredict
u = MDAnalysis.Universe('conf.pdb','traj.xtc')
DEER = DEERpredict(u,residues =[55,55],chains=['A', 'B'],temperature=298)
DEER.run()
```

The third line generates the MDAnalysis Universe object from an XTC trajectory and a PDB topology. The fourth line initializes the DEERpredict object with the spin-labeled residue numbers and the respective chain IDs. The fifth line runs the calculations and saves per-frame and ensemble-averaged data to *res-55-55.hdf5* and *res-55-55.dat*, respectively, as well as the steric partition functions of sites K55 and K55' to the file *res-Z-55-55.dat*.

The results of our unbiased and restrained simulations are in substantial agreement with the findings of Roche *et al.* [60, 61], indicating that the flaps of the inhibitor-free HIV-1PR are predominantly in closed conformation. Compared to the distance distribution calculated from the starting configuration of PDB code 3BVB (see Figure 1), predictions based on MD trajectories more accurately reproduce the shape of

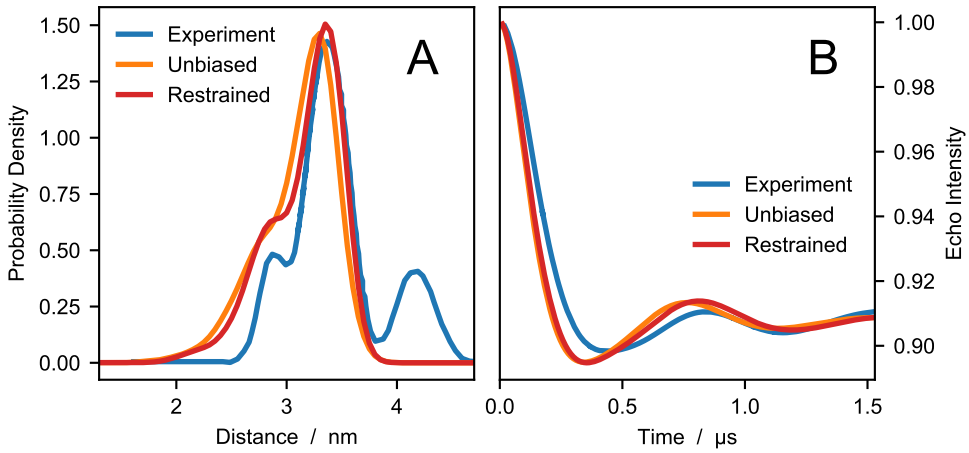


Fig 2. Comparing experiments and simulations for HIV1-PR. DEER distance distributions (A) and echo intensity curves (B) obtained by Torbeev *et al.* [39] from DEER experiments (blue), and calculated using DEER-PREdict from unbiased (orange) and RDC ensemble-biased MD simulations (red).

the shoulder and the main peak of the experimental $P(r)$. Moreover, using the RDC data as restraints leads to a significant improvement in the agreement between simulations and experiments, with the RMSD decreasing from 0.07 for the unbiased to 0.03 for the RDC ensemble-biased simulations. However, in the simulations we do not observe the wide-open state, corresponding to the minor peak at large separations in the experimental $P(r)$. This discrepancy could be due to insufficient sampling or could be attributed to the difference in sequence between the simulated protein (the wild type subtype B from isolate BRU/LAI) and the experimental construct, i.e. M36norleucine, S37N, R41pseudo-homoglutamine, M46norleucine, I63P, I64V, A67α-aminobutyric acid and A95α-aminobutyric acid.

Case study 2: DEER data for T4 Lysozyme

Lysozyme from the T4 bacteriophage (T4L) has long been used as a model system in the study of protein structure and dynamics [73–78]. Here, we focus on the L99A and the triple L99A-G113A-R119P mutants which are structurally similar and mainly differ in the relative populations of their major conformational states. The L99A variant presents a 150 Å^3 hydrophobic pocket capable of binding hydrophobic ligands and has been thoroughly studied to further our understanding of the dynamics and selectivity of the binding pocket [73, 79]. The L99A variant occupies two distinct conformational

states: the ground state (G) and the transient excited state (E), amounting for 97% and 3% of the population, respectively. The large-scale motions converting the G into the E state occur on the millisecond time scale and result in the occlusion of the cavity, which is occupied by the side chain of F114 in the E state [77]. The additional G113A and R119P mutations in the triple-mutant variant interconvert the populations of the conformational states to 4% for the G state and 96% for the E state [77] — note that, here and in the following, we refer to the G and E states based on their structural similarity to the L99A variant rather than on their relative populations. These conformational equilibria have been studied by DEER for various pairs of spin-labeled sites, which effectively resolve the G and E states as separate peaks of the $P(r)$ [78].

Here, we compare DEER distance distributions calculated with DEER-PREdict for two pairs of probe positions (D89C–T109C and T109C–N140C) with the corresponding experimental data by Lerch*et al.* [78]. First, we calculate the $P(r)$ of the single states using PDB code 3DMV for the G states and PDB codes 2LCB and 2LC9 for the E states of single and triple mutants, respectively. Second, the $P(r)$'s are linearly combined based on the experimentally derived ratios of G and E populations (97:3 for L99A and 4:96 for L99A-G113A-R119P) [77]. Additionally, we predict DEER distance distributions from previously reported metadynamics MD simulations of L99A and L99A-G113A-R119P [75]. In these calculations, the average over the trajectory is weighted by $\exp\left(F_{bias}/k_BT\right)$, where F_{bias} is the final static bias for each frame and k_BT is the thermal energy. The analysis of a trajectory of 6,670 frames is performed in 11 min on a 1.7 GHz processor executing the following lines of code:

```
1 import MDAnalysis
2 from DEERPREdict.DEER import DEERpredict
3 import numpy as np
4 u = MDAnalysis.Universe('conf.pdb','traj.xtc')
5 for residues in [[89, 109],[109, 140]]:
6     DEER = DEERpredict(u,residues=residues,temperature=298)
7     DEER.run(weights=np.exp(Fbias/(0.298*8.3145)))
```

In line six, we specify the positions of the spin-labels and the temperature at which the metadynamics simulations were performed. In line seven, we provide the weights of each trajectory frame, generated from the array of F_{bias} values.

Figure 3 shows a comparison between the experimental distance distributions

obtained by Lerch *et al.* [78] and our predictions. In general, the calculated distributions fall within the experimental ranges of inter-probe distances and are particularly accurate for the D89C–T109C spin-labeled pair in metadynamics simulations. The sharper shape of the experimental $P(r)$'s, relative to the calculated distributions, could be due to the cryogenic temperatures at which DEER experiments are conducted, whereas simulations were performed at room temperature.

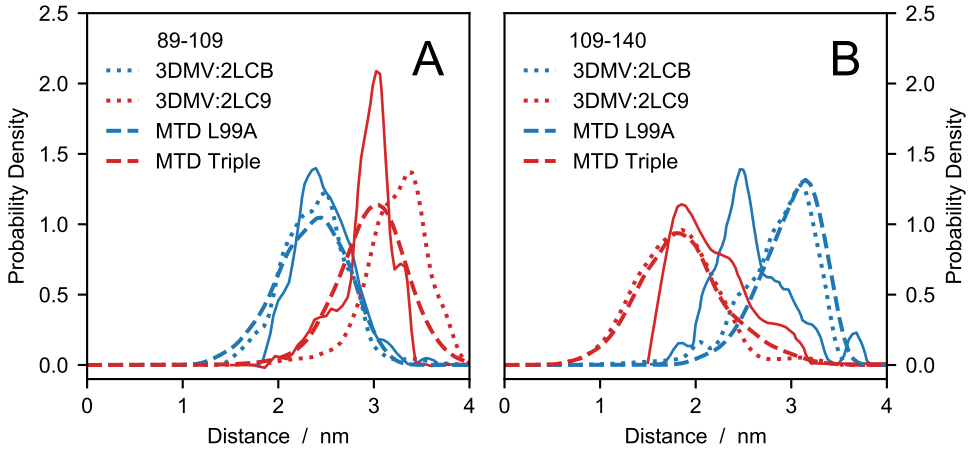


Fig 3. Comparing experiments with simulations and structures of T4 Lysozyme variants. DEER distance distributions for probe positions (A) D89C–T109C and (B) T109C–N140C of the single (blue) and the triple variant (red). Solid lines are the experimental data by Lerch et al. [78], dotted lines are calculated from PDB codes and dashed lines are predictions from metadynamics (MTD) simulations by Wang and coworkers [75].

Case study 3: PRE data for Acyl-CoA-Binding Protein

The RLA is well known in the EPR community and generally favored over e.g. a Cα-based approach as discussed elsewhere [3, 10, 23]. In the presented software, we apply the same improved modeling of the probe flexibility also to the prediction of PRE rates and intensity ratios.

Our test data is the PRE data for the bovine Acyl-coenzyme A Binding Protein (ACBP) reported by Teilum *et al.* [48]. In this study the structural behavior of ACBP under native and mildly-denaturing conditions was investigated via the SDSL of five positions in the amino acid the sequence: T17C, V36C, M46C, S65C and I86C. Here, we focus on the native state of ACBP for which an NMR structure comprising 20 conformers has been refined from residual dipolar couplings (RDC) and deposited in the

Protein Data Bank (PDB code 1NTI). Figure 4 shows a comparison between the experimental data and the intensity ratios calculated from the Γ_2 values averaged over the 20 conformations of the PDB entry. A good overall agreement is achieved across the different probe positions. Notably, using the RDC-refined structure, we reproduce most of the structural features observed in the PRE experiments, including the proximity of residues 24, 27, 31 and 34 to the spin-labeled residue 86, which is consistent with a helix-turn-helix motif. The predicted intensity ratios are generated in 2 s on a 1.7 GHz processor executing the following code:

```
1 import MDAnalysis
2 from DEERPREdict.PRE import PREpredict
3 u = MDAnalysis.Universe('1nti.pdb')
4 for res in [17,36,46,65,86]:
5     PRE = PREpredict(u,res,temperature=298,atom_selection='H')
6     PRE.run(tau_c=2e-09,tau_t=2*1e-10,delay=1e-2,r_2=12.6,wh=750)
```

At line three, we load PDB code 1NTI as an MDAnalysis Universe object. We then use a for loop to calculate the PRE data from the distances between amide protons and the spin-label N-O group at five different positions along the amino acid sequence. In the last line we specify $\tau_c = 2\,\mathrm{ns}$, $\tau_t = 0.2\,\mathrm{ns}$, $t_d = 10\,\mathrm{ms}$, $R_2 = 12.6\,\mathrm{s}^{-1}$ and $\omega_I = 2\pi \times 750\,\mathrm{MHz}$. Per-frame and ensemble-averaged PRE data are automatically saved to files named *res-*.pkl* and *res-*.dat*, respectively, whereas per-frame steric partition functions are saved to *res-Z-*.dat*.

The RLA analysis also provides an indication on how well the structure can accommodate the paramagnetic probe at a given position. For example, Figure S3 shows that, while for most spin-label sites the steric partition function, Z, varies between 1 and 1.5, for the placement of the probe at residue 86, Z drops below the cutoff of 0.05 proposed by Polyhach *et al.* [21] in five conformers out of 20. The significantly lower Z values indicate that residue 86 is in a tightly packed region in the protein structures and that spin-labeling position 86 can lead to structural deformations or changes in the populations of the underlying ensemble. This observation is consistent with stability experiments performed by Teilum *et al.*on wild type and spin-labeled mutants [48], showing that I86C is the least stable of the studied spin-labeled mutants.

Although the RLA assumes that the overall protein conformation is unaffected by

the presence of the spin-label, the case of the I86C mutant of ACBP highlights how DEER-PREdict can help to identify spin-labeled sites in folded proteins that are likely to violate this assumption.

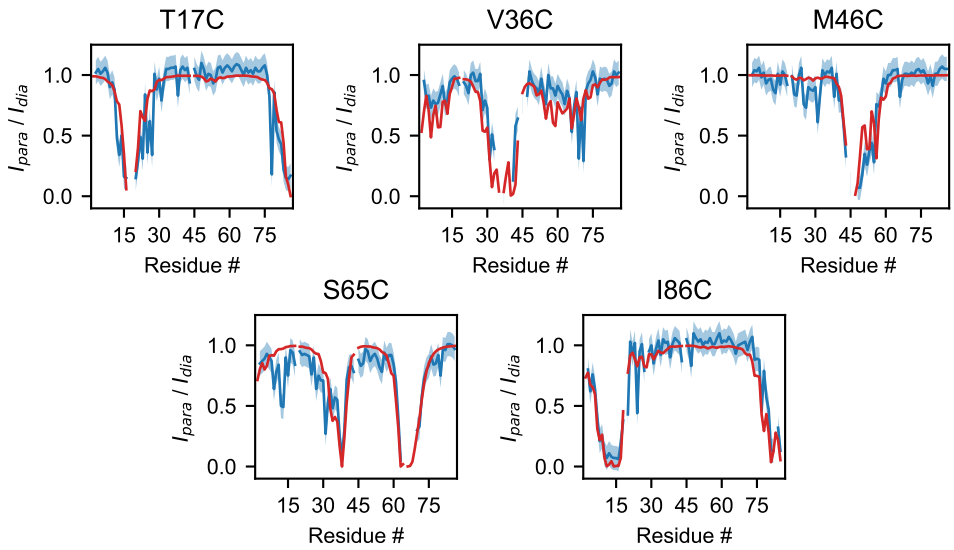


Fig 4. Calculated and experimental PRE HSQC intensity ratios for T17C, V36C, M46C, S65C and I86C. Blue lines represent the experimental data [48], with the associated ±0.1 error shown by the blue shaded areas. Red lines represent intensity ratios calculated from PDB code 1NTI with $\tau_c = 2\,\mathrm{ns}$, $\tau_t = 0.2\,\mathrm{ns}$, $t_d = 10\,\mathrm{ms}$, $R_2 = 12.6\,\mathrm{s}^{-1}$.

Comparison with Cβ-based PRE Predictions

As previously discussed, the explicit treatment of the paramagnetic probe may be crucial for the accurate back-calculation of DEER data, and even more so for PRE predictions, due to the $\langle r^{-6} \rangle$-dependence of the PRE.

A common way to restrain MD simulations or to back-calculate PRE experimental data without explicitly simulating the paramagnetic probe is to approximate the electron location to the position of the Cβ atom of the spin-labeled residue [80]. The advantage of this approach is that (a) multiple labeling sites can be analyzed in a single simulation and (b) the explicit atom is present in the simulation making the calculation of PREs straightforward. Cβ-based calculations may, however, be prone to over- or underestimating electron-proton distances by several angstroms, thereby introducing a systematic error. Figure 5 shows that the Cβ approximation may overestimate the effect of transient tertiary contacts on the PRE rates of ACBP, yielding experimentally

consistent predictions only when the time scales for the reorientational dynamics is artificially made fast. With respect to the RLA, the Cβ approach also leads to increased RMSD between the predicted and the experimental intensity ratio of ACBP (see Figure S4). Moreover, when using the RLA, the τ_c values that minimise the RMSD are found in the range of typical reorientational correlation time constants for proteins of $\sim$ 100 residues (2–5 ns) whereas the Cβ-based approach underestimates the optimal τ_c.

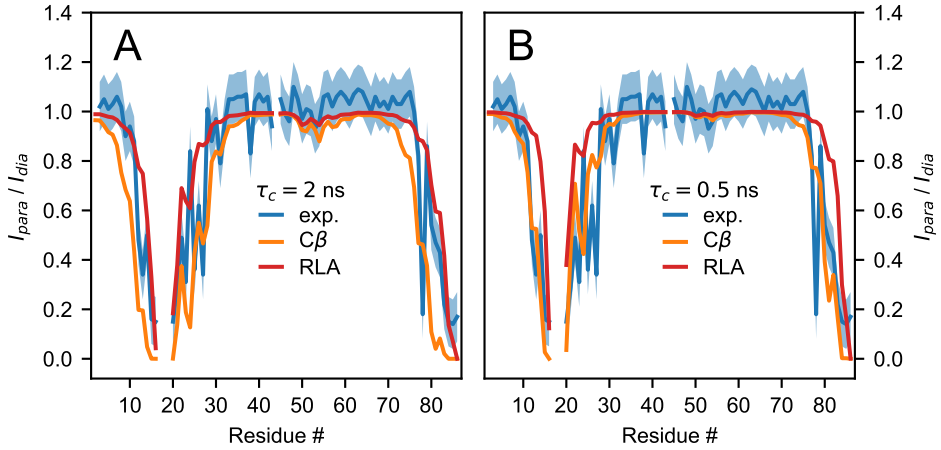


Fig 5. PRE intensity ratios for ACBP spin labeled at position 65 calculated for (A) $\tau_c = 2$ ns and (B) $\tau_c = 0.5$ ns. Blue lines represent the experimental data [48], with the associated ± 0.1 error shown by the blue shaded areas. Orange and red lines represent Cβ-based and RLA-based predictions, respectively.

Conclusion

We have introduced an open-source software with a fast implementation of the RLA in tandem with protein ensemble averaging, for the calculation of DEER and PRE data. Using three examples, we have highlighted the capabilities of our implementation: (a) the extension of the RLA for DEER data from a protein ensemble and (b) the calculation of PRE rates and intensity ratios with the same approach.

The structural interpretation of DEER and PRE measurements requires an accurate treatment of the structure and conformational heterogeneity of the spin labels. In the presented software, this is achieved using the RLA and, in the case of the PRE, a model-free approach to describe the dynamics. Relative to simulations of the explicitly spin-labeled mutants, the RLA presents the particular advantage of enabling the

prediction for multiple SDSL experiments from a single simulation of the wild type sequence.

DEER-PREdict bridges *in vitro* and *in silico* structural biology in two ways:

1. It improves the structural interpretation of experimental DEER and PRE results by providing rapid and accurate predictions of electron-electron distances and PRE effects from conformational ensembles. Therefore, the program can serve as a tool for validating molecular models and force fields as well as for interpreting experimental results.

2. It can be used to rapidly scan a number of candidate sites and guide the selection of adequate positions for the investigation of specific conformational states. By inspecting the steric partition function provided by our tool, it is also possible to predict if a position in the sequence is likely to accommodate the paramagnetic probe within the wild-type structure. If models for distinct conformational states are available, DEER-PREdict may also be used to determine sites where experimental measurements may be used to distinguish between the conformational states. Both of these features hold the potential of facilitating the design and enhance the success rate of time- and labor-intensive SDSL experiments.

Availability and Future Directions

The software is implemented using the popular trajectory analysis package MDAnalysis, version 1.0 [27] and is available on GitHub at github.com/KULL-Centre/DEERpredict. DEER-PREdict is also distributed as a PyPI package (pypi.org/project/DEERPREdict) and archived on Zenodo (DOI: 10.5281/zenodo.3968394). DEER-PREdict and MDAnalysis are published and distributed under GPL licenses, version 3 and 2, respectively.

DEER-PREdict has a general framework and can be readily extended to encompass non-protein biomolecules as well as additional rotamer libraries of paramagnetic groups. Moreover, the software can be augmented with a module to predict Förster resonance energy transfer data, combining the insertion routines already implemented for MTSSL

probes with existing rotamer libraries for fluorescent dyes.

Acknowledgments

M.B.A.K. acknowledges funding from the Lundbeck Foundation. R.C. acknowledges funding from MINECO (CTQ2016-78636-P). K.L.-L. acknowledges funding via a Sapere Aude Starting Grant from the Danish Council for Independent Research and the Lundbeck Foundation BRAINSTRUC initiative in structural biology.

Author Contributions

Conceptualization: Giulio Tesei, João M. Martins, Micha B. A. Kunze, Kresten Lindorff-Larsen

Data curation: Giulio Tesei, João M. Martins, Micha B. A. Kunze, Yong Wang

Funding acquisition: Ramon Crehuet, Kresten Lindorff-Larsen

Software: Giulio Tesei, João M. Martins, Micha B. A. Kunze, Ramon Crehuet

Supervision: Micha B. A. Kunze, Kresten Lindorff-Larsen

Writing — original draft: Giulio Tesei, João M. Martins, Micha B. A. Kunze, Kresten Lindorff-Larsen

Writing — review & editing: Giulio Tesei, João M. Martins, Micha B. A. Kunze, Yong Wang, Ramon Crehuet, Kresten Lindorff-Larsen

Supporting Information

Molecular Dynamics Simulations

HIV-1 protease

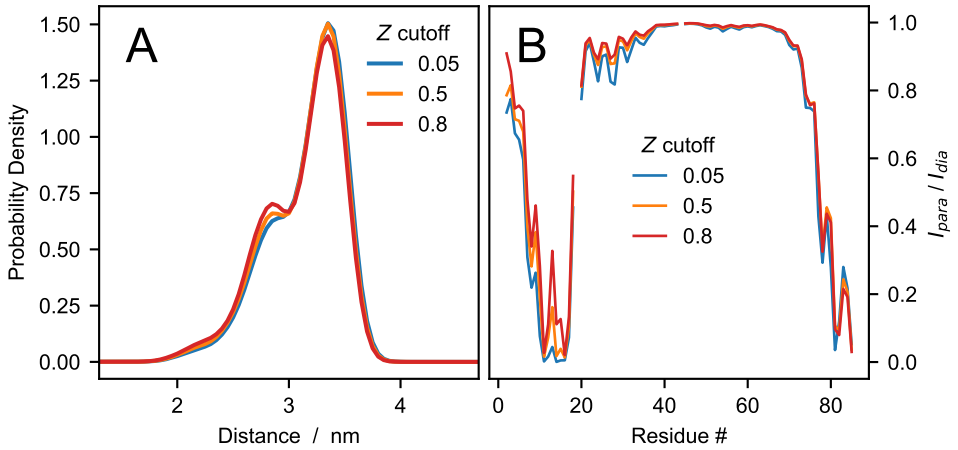
All simulations were performed using GROMACS 5.1 [81] with the PLUMED 2 [80] plugin. Unbiased and RDC-biased metadynamics [72, 82] simulations where performed starting from a closed conformer after removal of the inhibitor from the X-ray crystal structure (PDB code 3BVB) [83]. The protein was simulated in a cubic box with a side length of 8.677 nm containing 22,228 water molecules, 59 sodium cations, and 67 chloride anions. Although the K55C mutations have been shown to have a negligible impact on enzymatic activity [59], we maintained the lysine residues at the spin-labeled positions with the aim of capturing the conformational ensemble of the wild type.

Third, the system was equilibrated for 10 ns with a 2-fs time step in the NPT ensemble, using the Berendsen barostat [84] with 0.5-ps coupling constant and isothermal compressibility of 4.5e-5 bar^{-1}. Starting from the equilibrated structure, a production run of 500 ns was performed in the NVT ensemble for the unbiased MD simulation. For the restrained simulations, we obtained the backbone N-H RDCs for the inhibitor-free HIV-1PR from Roche *et al.* [61] and applied a linear potential (force constant 25,000 kJ/mol) between the experimental data and the RDCs calculated as averages over 4 independent simulations. Each replica was simulated for 125 ns in the NVT ensemble. We used the AMBER ff99SB*-ILDN force field [85, 86] for all simulations. First, the structure was minimized with the steepest descent algorithm to a tolerance of $10\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ with restrained water molecules. Second, we simulated the system the for 5 ns in the NVT ensemble using the leap-frog integrator with a time step of 1 fs. A 1.2 nm cutoff was used for van der Waals interactions, with a force-switch modification at 1.0 nm. Coulomb interactions were treated with the particle-mesh Ewald method [87] with direct space cutoff of 1.2 nm. The temperature was set to 298 K using the v-rescale thermostat [88] with a 5 ps coupling constant. All the bonds involving hydrogen atoms were constrained using the LINCS algorithm [89].

T4 Lysozyme

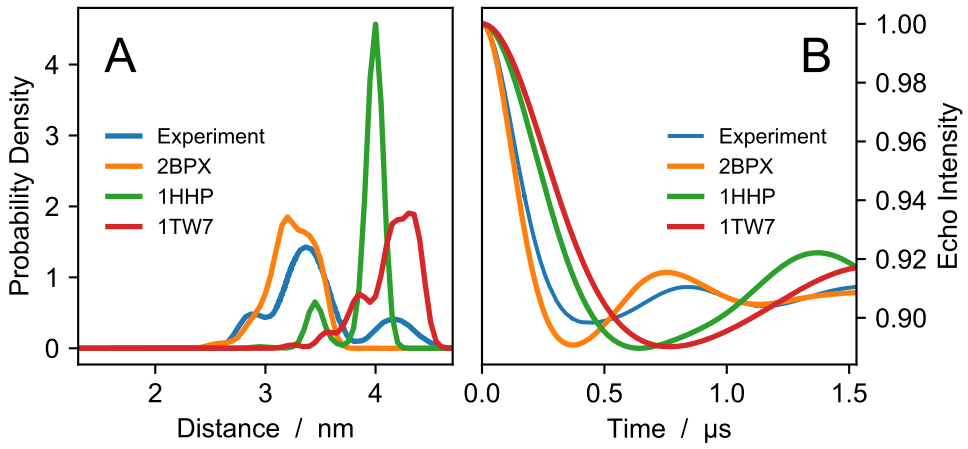
The simulations of the L99A single mutant and the L99A-G113A-R119P triple mutant of T4 Lysozyme analysed in this study have been reported previously by Wang *et al.* [75]. The X-ray crystal structure of PDB code 3DMV was used as the initial configuration for the G states of both single and triple mutants. Simulations of the E states were started from chemical-shift-derived Rosetta-calculated structures, i.e. PDB code 2LCB and PDB code 2LC9 for single and triple mutant, respectively. Further simulations details can be found in the original article [75].

Fig. S1 Influence of the Z cutoff on predicted DEER and PRE NMR data.



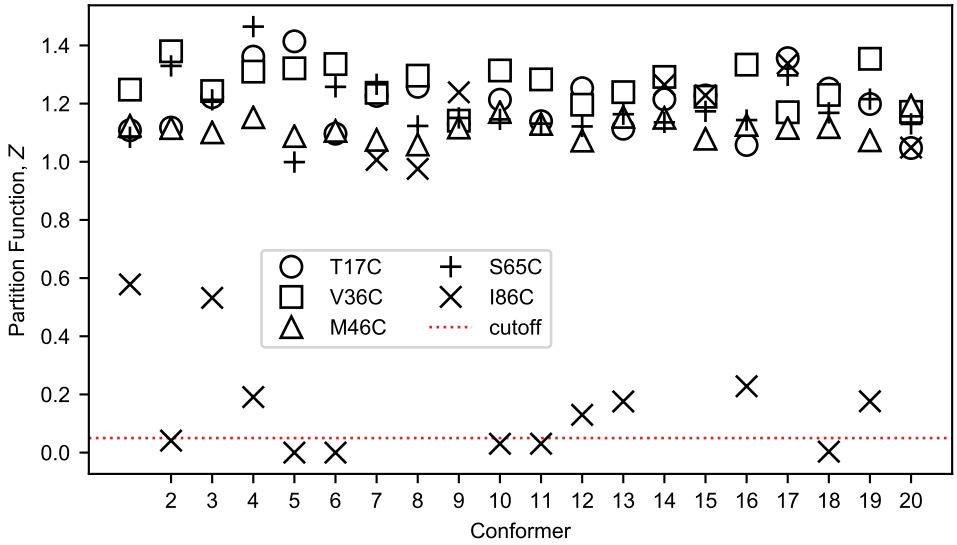
(A) DEER distance distributions calculated from RDC ensemble-biased MD simulations of HIV-1PR. (B) Predicted intensity ratios for ACBP spin-labeled at position 86 obtained from PDB code 1NTI with $\tau_c = 2\,\mathrm{ns}$, $\tau_t = 0.2\,\mathrm{ns}$, $t_d = 10\,\mathrm{ms}$, $R_2 = 12.6\,\mathrm{s}^{-1}$. DEER and PRE predictions are performed using three different cutoff values of the steric partition function, Z, namely 0.05 (blue lines), 0.5 (orange lines) and 0.8 (red lines).

Fig. S2 Comparison of DEER data from Torbeev *et al.* [39] with X-ray crystal structures deposited in the Protein Data Bank.



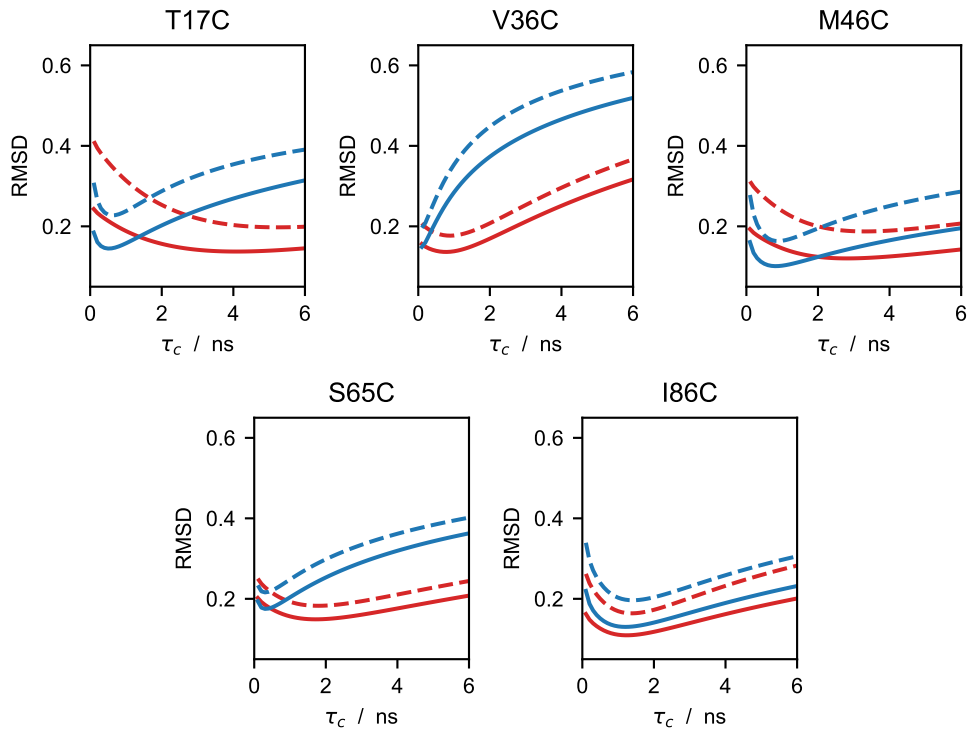
DEER distance distributions (A) and echo intensity curves (B) obtained by Torbeev *et al.* [39] from DEER experiments (blue), and calculated using X-ray crystal structures representative of closed (PDB code 2BPX, orange), semi-open (PDB code 1HHP, green) and wide-open (PDB code 1TW7, red) HIV-1PR conformations.

Fig. S3 Steric partition function quantifying the fitness of the rotamers at the spin-labeled site.



Steric partition function calculated from rotamer-protein van der Waals interactions for five spin-labeled mutants of ACBP. The horizontal dashed line indicates the cutoff used in the criterion for discarding protein conformations where the placement of the rotamer is characterized by steric clashes with the surrounding residues.

Fig. S4 Comparison of optimal τ_c values for RLA vs. Cβ-based approach. This figure illustrates the systematic time-scale difference when using explicit MTSSL probes instead of approximating the location of the unpaired electron with the position of the Cβ atom of the spin-labeled residue. Compared with the C_β-based approach, the RLA improves the accuracy in reproducing the experimental PRE data. The optimal τ_c values are also closer to the experimentally derived value of 4 ns [48] when using the RLA prediction whereas the minima of the Cβ-based RMSD are generally located at lower τ_c values.



Dependence on τ_c of the RMSD between experimental and predicted PRE ratios of ACBP. Red and blue lines are obtained using the RLA and approximating the electron location with the position of the Cβ atom, respectively. Solid and dashed lines represent the RMSD values calculated from all the data points and from intensity ratios in the dynamic range $0.1 < I_{para} \ / \ I_{dia} < 0.9$.

References

1. Orioli S, Larsen AH, Bottaro S, Lindorff-Larsen K. How to learn from inconsistencies: Integrating molecular simulations with experimental data. In: Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly. Elsevier; 2020. p. 123–176. Available from: `https://doi.org/10.1016/bs.pmbts.2019.12.006`.

2. Pannier M, Veit S, Godt A, Jeschke G, Spiess HW. Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins. Journal of Magnetic Resonance. 2000;142(2):331–340.

3. Bagnéris C, Rogala KB, Baratchian M, Zamfir V, Kunze MBA, Dagless S, et al. Probing the Solution Structure of IκB Kinase (IKK) Subunit γ and Its Interaction with Kaposi Sarcoma-associated Herpes Virus Flice-interacting Protein and IKK Subunit β by EPR Spectroscopy. Journal of Biological Chemistry. 2015;290(27):16539–16549.

4. Klare JP, Steinhoff HJ. Spin labeling EPR. Photosynthesis Research. 2009;102(2-3):377–390.

5. Phan G, Remaut H, Wang T, Allen WJ, Pirker KF, Lebedev A, et al. Crystal structure of the FimD usher bound to its cognate FimC–FimH substrate. Nature. 2011;474(7349):49–53.

6. Klose D, Voskoboynikova N, Orban-Glass I, Rickert C, Engelhard M, Klare JP, et al. Light-induced switching of HAMP domain conformation and dynamics revealed by time-resolved EPR spectroscopy. FEBS Letters. 2014;588(21):3970–3976.

7. Edwards TH, Stoll S. A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy. J Magn Reson. 2016;270:87–97.

8. Casey TM, Fanucci GE. Spin labeling and Double Electron-Electron Resonance (DEER) to Deconstruct Conformational Ensembles of HIV Protease. In: Methods in Enzymology. Elsevier; 2015. p. 153–187. Available from: `https://doi.org/10.1016/bs.mie.2015.07.019`.

9. Clore GM, Iwahara J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev. 2009;109(9):4108–4139.

10. Jeschke G. MMM: A toolbox for integrative structure modeling. Protein Science. 2017;27(1):76–85.

11. Reichel K, Stelzl LS, Köfinger J, Hummer G. Precision DEER distances from spin-label ensemble refinement. The journal of physical chemistry letters. 2018;9(19):5748–5752.

12. Budil DE, Sale KL, Khairy KA, Fajer PG. Calculating slow-motional electron paramagnetic resonance spectra from molecular dynamics using a diffusion operator approach. The Journal of Physical Chemistry A. 2006;110(10):3703–3713.

13. Ding F, Layten M, Simmerling C. Solution structure of HIV-1 protease flaps probed by comparison of molecular dynamics simulation ensembles and EPR experiments. Journal of the American Chemical Society. 2008;130(23):7184–7185.

14. Sezer D, Freed JH, Roux B. Parametrization, molecular dynamics simulation, and calculation of electron spin resonance spectra of a nitroxide spin label on a polyalanine α-helix. The journal of physical chemistry B. 2008;112(18):5755–5767.

15. Xue Y, Skrynnikov NR. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, 15N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? Journal of the American Chemical Society. 2011;133(37):14614–14628.

16. Sasmal S, Lincoff J, Head-Gordon T. Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid-β. Biophysical journal. 2017;113(5):1002–1011.

17. Robinson B, Slutsky L, Auteri F. Direct simulation of continuous wave electron paramagnetic resonance spectra from Brownian dynamics trajectories. The Journal of chemical physics. 1992;96(4):2609–2616.

18. Steinhoff HJ, Hubbell WL. Calculation of electron paramagnetic resonance spectra from Brownian dynamics trajectories: application to nitroxide side chains in proteins. Biophysical journal. 1996;71(4):2201–2212.

19. Tombolato F, Ferrarini A, Freed JH. Dynamics of the nitroxide side chain in spin-labeled proteins. The Journal of Physical Chemistry B. 2006;110(51):26248–26259.

20. Tombolato F, Ferrarini A, Freed JH. Modeling the effects of structure and dynamics of the nitroxide side chain on the ESR spectra of spin-labeled proteins. The Journal of Physical Chemistry B. 2006;110(51):26260–26271.

21. Polyhach Y, Bordignon E, Jeschke G. Rotamer libraries of spin labelled cysteines for protein studies. Phys Chem Chem Phys. 2011;13(6):2356–2366. doi:10.1039/c0cp01865a.

22. Stelzl LS, Fowler PW, Sansom MSP, Beckstein O. Flexible Gates Generate Occluded Intermediates in the Transport Cycle of LacY. J Mol Biol. 2014;426(3):735–751.

23. Klose D, Klare JP, Grohmann D, Kay CWM, Werner F, Steinhoff HJ. Simulation vs. Reality: A Comparison of In Silico Distance Predictions with DEER and FRET Measurements. PloS One. 2012;7(6):e39492.

24. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, et al. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. J Am Chem Soc. 2010;132(24):8407–8418.

25. Milkovic NM, Thomasen FE, Cuneo MJ, Grace CR, Martin EW, Nourse A, et al. Interplay of folded domains and the disordered low-complexity domain in mediating hnRNPA1 phase separation. bioRxiv. 2020;doi:10.1101/2020.05.15.096966.

26. Bengtsen T, Holm VL, Kjølbye LR, Midtgaard SR, Johansen NT, Tesei G, et al. Structure and dynamics of a nanodisc by integrating NMR, SAXS and SANS experiments with molecular dynamics simulations. bioRxiv. 2019;doi:10.1101/734822.

27. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. Journal of Computational Chemistry. 2011;32(10):2319–2327. doi:10.1002/jcc.21787.

28. Dunbrack Jr RL. Rotamer libraries in the 21st century. Current opinion in structural biology. 2002;12(4):431–440.

29. Ponder JW, Richards FM. Tertiary templates for proteins. J Mol Biol. 1987;193(4):775–791.

30. Bower MJ, Cohen FE, Dunbrack RL. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. J Mol Biol. 1997;267(5):1268–1282.

31. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. Protein Sci. 1995;4(10):2006–2018.

32. Fleissner MR, Cascio D, Hubbell WL. Structural origin of weakly ordered nitroxide motion in spin-labeled proteins. Protein Sci. 2009;18(5):893–908.

33. Ibáñez LF, Jeschke G, Stoll S. DeerLab: A comprehensive toolbox for analyzing dipolar EPR spectroscopy data. Magnetic Resonance Discussions. 2020;doi:10.5194/mr-2020-13.

34. Göddeke H, Timachi MH, Hutter CAJ, Galazzo L, Seeger MA, Karttunen M, et al. Atomistic Mechanism of Large-Scale Conformational Transition in a Heterodimeric ABC Exporter. Journal of the American Chemical Society. 2018;140(13):4543–4551.

35. 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.

36. Bottaro S, Bengtsen T, Lindorff-Larsen K. Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. In: Methods in Molecular Biology. Springer US; 2020. p. 219–240. Available from: `https://doi.org/10.1007/978-1-0716-0270-6_15`.

37. Hayes J. Some observations on digital smoothing of electroanalytical data based on the Fourier Transformation. Analytical Chemistry. 1973;45(2):277–284.

38. Worswick SG, Spencer JA, Jeschke G, Kuprov I. Deep neural network processing of DEER data. Science Advances. 2018;4(8):eaat5218. doi:10.1126/sciadv.aat5218.

39. Torbeev VY, Raghuraman H, Mandal K, Senapati S, Perozo E, Kent SBH. Dynamics of "Flap" Structures in Three HIV-1 Protease/Inhibitor Complexes Probed by Total Chemical Synthesis and Pulse-EPR Spectroscopy. J Am Chem Soc. 2009;131(3):884–885.

40. Solomon I. Relaxation Processes in a System of Two Spins. Phys Rev. 1955;99(2):559.

41. Bloembergen N. Proton Relaxation Times in Paramagnetic Solutions. J Chem Phys. 1957;27(2):572–573.

42. 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. doi:10.1021/ja00381a009.

43. Olejniczak ET, Dobson CM, Karplus M, Levy RM. Motional averaging of proton nuclear Overhauser effects in proteins. Predictions from a molecular dynamics simulation of lysozyme. J Am Chem Soc. 1983;106:1923–1930.

44. Brueschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR. Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution. J Am Chem Soc. 1991;114:2289–2302.

45. Iwahara J, Schwieters CD, Clore GM. Ensemble Approach for NMR Structure Refinement against 1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule. J Am Chem Soc. 2004;126(18):5879–5896.

46. Clore GM, Iwahara J. Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. Chemical Reviews. 2009;109(9):4108–4139. doi:10.1021/cr900033p.

47. Bibow S, Ozenne V, Biernat J, Blackledge M, Mandelkow E, Zweckstetter M. Structural Impact of Proline-Directed Pseudophosphorylation at AT8, AT100, and PHF1 Epitopes on 441-Residue Tau. Journal of the American Chemical Society. 2011;133(40):15842–15845. doi:10.1021/ja205836j.

48. Teilum K, Kragelund BB, Poulsen FM. Transient Structure Formation in Unfolded Acyl-coenzyme A-binding Protein Observed by Site-directed Spin Labelling. J Mol Biol. 2002;324(2):349–357.

49. Liu W, Liu X, Zhu G, Lu L, Yang D. A Method for Determining Structure Ensemble of Large Disordered Protein: Application to a Mechanosensing Protein. Journal of the American Chemical Society. 2018;140(36):11276–11285. doi:10.1021/jacs.8b04792.

50. Gomes GNW, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, et al. Integrating multiple experimental data to determine conformational ensembles of an intrinsically disordered protein. bioRxiv. 2020;doi:10.1101/2020.02.05.935890.

51. Khan SN, Charlier C, Augustyniak R, Salvi N, Déjean V, Bodenhausen G, et al. Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. Biophysical Journal. 2015;109(5):988–999. doi:10.1016/j.bpj.2015.06.069.

52. Rezaei-Ghaleh N, Parigi G, Zweckstetter M. Reorientational Dynamics of Amyloid-β from NMR Spin Relaxation and Molecular Simulation. The Journal of Physical Chemistry Letters. 2019;10(12):3369–3375. doi:10.1021/acs.jpclett.9b01050.

53. Camilloni C, Cavalli A, Vendruscolo M. Replica-Averaged Metadynamics. J Chem Theory Comput. 2013;9(12):5610–5617.

54. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;.

55. Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, et al. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. Molecular Cell. 2018;69(3):465–479.e7. doi:10.1016/j.molcel.2017.12.022.

56. Battiste JL, Wagner G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry. 2000;39(18):5355–5365.

57. Richard J Gowers, Max Linke, Jonathan Barnoud, Tyler J E Reddy, Manuel N Melo, Sean L Seyler, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In: Sebastian Benthall, Scott Rostrup, editors. Proceedings of the 15th Python in Science Conference; 2016. p. 98 – 105.

58. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, et al. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci U S A. 1988;85(13):4686–4690.

59. Blackburn ME, Veloro AM, Fanucci GE. Monitoring inhibitor-induced conformational population shifts in HIV-1 protease by pulsed EPR spectroscopy. Biochemistry. 2009;48(37):8765–8767.

60. Roche J, Louis JM, Bax A. Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. ChemBioChem. 2015;16(2):214–218.

61. Roche J, Louis JM, Bax A, Best RB. Pressure-induced structural transition of mature HIV-1 Protease from a combined NMR/MD simulation approach. Proteins. 2015;83(12):2117–2123.

62. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. J Am Chem Soc. 2006;128(9):2812–2813.

63. Huang X, Britto MD, Kear-Scott JL, Boone CD, Rocca JR, Simmerling C, et al. The Role of Select Subtype Polymorphisms on HIV-1 Protease Conformational Sampling and Dynamics. J Biol Chem. 2014;289(24):17203–17214.

64. Liu Z, Casey TM, Blackburn ME, Huang X, Pham L, de Vera IMS, et al. Pulsed EPR characterization of HIV-1 protease conformational sampling and inhibitor-induced population shifts. Physical Chemistry Chemical Physics. 2016;18(8):5819–5831. doi:10.1039/c5cp04556h.

65. Sayer JM, Liu F, Ishima R, Weber IT, Louis JM. Effect of the Active Site D25N Mutation on the Structure, Stability, and Ligand Binding of the Mature HIV-1 Protease. Journal of Biological Chemistry. 2008;283(19):13459–13470. doi:10.1074/jbc.m708506200.

66. Munshi S, Chen Z, Li Y, Olsen DB, Fraley ME, Hungate RW, et al. Rapid X-ray diffraction analysis of HIV-1 protease–inhibitor complexes: inhibitor exchange in single crystals of the bound enzyme. Acta Crystallographica Section D Biological Crystallography. 1998;54(5):1053–1060. doi:10.1107/s0907444998003588.

67. Spinelli S, Liu QZ, Alzari PM, Hirel PH, Poljak RJ. The three-dimensional structure of the aspartyl protease from the HIV-1 isolate BRU. Biochimie. 1991;73(11):1391 – 1396. doi:https://doi.org/10.1016/0300-9084(91)90169-2.

68. Martin P, Vickrey JF, Proteasa G, Jimenez YL, Wawrzak Z, Winters MA, et al. "Wide-Open" 1.3 ÅStructure of a Multidrug-Resistant HIV-1 Protease as a Drug Target. Structure. 2005;13(12):1887–1895. doi:10.1016/j.str.2005.11.005.

69. Logsdon BC, Vickrey JF, Martin P, Proteasa G, Koepke JI, Terlecky SR, et al. Crystal Structures of a Multidrug-Resistant Human Immunodeficiency Virus Type 1 Protease Reveal an Expanded Active-Site Cavity. Journal of Virology. 2004;78(6):3123–3132. doi:10.1128/jvi.78.6.3123-3132.2004.

70. Heaslet H, Rosenfeld R, Giffin M, Lin YC, Tam K, Torbett BE, et al. Conformational flexibility in the flap domains of ligand-free HIV protease. Acta Crystallographica Section D Biological Crystallography. 2007;63(8):866–875. doi:10.1107/s0907444907029125.

71. Huang X, Britto MD, Kear-Scott JL, Boone CD, Rocca JR, Simmerling C, et al. The Role of Select Subtype Polymorphisms on HIV-1 Protease Conformational Sampling and Dynamics. Journal of Biological Chemistry. 2014;289(24):17203–17214. doi:10.1074/jbc.m114.571836.

72. Camilloni C, Vendruscolo M. A tensor-free method for the structural and dynamical refinement of proteins using residual dipolar couplings. J Phys Chem B. 2015;119(3):653–661.

73. Eriksson AE, Baase WA, Wozniak JA, Matthews BW. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature. 1992;355(6358):371–373.

74. Kato H, Vu ND, Feng H, Zhou Z, Bai Y. The folding pathway of T4 lysozyme: an on-pathway hidden folding intermediate. J Mol Biol. 2007;365(3):881–891.

75. Wang Y, Papaleo E, Lindorff-Larsen K. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. eLife. 2016;5. doi:10.7554/elife.17505.

76. Mulder FA, Hon B, Muhandiram DR, Dahlquist FW, Kay LE. Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. Biochemistry. 2000;39(41):12614–12622.

77. Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, et al. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature. 2011;477(7362):111–114.

78. Lerch MT, López CJ, Yang Z, Kreitman MJ, Horwitz J, Hubbell WL. Structure-relaxation mechanism for the response of T4 lysozyme cavity mutants to hydrostatic pressure. Proc Natl Acad Sci U S A. 2015;112(19):E2437–46.

79. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, et al. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science. 1992;255(5041):178–183.

80. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Computer Physics Communications. 2014;185(2):604–613.

81. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1(2):19–25.

82. Cavalli A, Camilloni C, Vendruscolo M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. J Chem Phys. 2013;138(9).

83. Sayer JM, Liu F, Ishima R, Weber IT, Louis JM. Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. J Biol Chem. 2008;283(19):13459–13470.

84. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics. 1984;81(8):3684–3690. doi:10.1063/1.448118.

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

86. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix—Coil Transition of Polypeptides. J Phys Chem B. 2009;113(26):9004–9015.

87. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 1993;98(12):10089–10092. doi:10.1063/1.464397.

88. Bussi G, Donadio D, Parrinello M. Canonical Sampling Through Velocity Rescaling. J Chem Phys. 2007;126(1):014101. doi:10.1063/1.2408420.

89. Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal of Chemical Theory and Computation. 2008;4(1):116–122. doi:10.1021/ct700200b.