ETH zürich

ETH Library

Rotamer libraries of spin labelled cysteines for protein studies

Journal Article

Author(s):
Polyhach, Yevhen; Bordignon, Enrica; Jeschke, Gunnar

Publication date:
2011-02-14

Permanent link:
https://doi.org/10.3929/ethz-a-010784321

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Physical Chemistry Chemical Physics 13(6), https://doi.org/10.1039/C0CP01865A

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Rotamer libraries of spin labelled cysteines for protein studies

Yevhen Polyhach[§], Enrica Bordignon[§], Gunnar Jeschke[*§]

[§]Laboratory of Physical Chemistry, ETH Zürich, Wolfgang-Pauli-Str. 10, 8093 Zürich, Switzerland

[*]To whom correspondence should be addressed. Phone: +41 44 632 5702, Fax: +41 44 633 1448, e-mail: gunnar.jeschke@phys.chem.ethz.ch

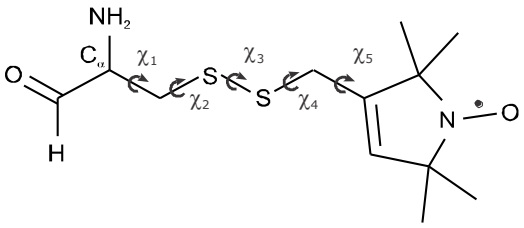
Abstract

Studies of structure and dynamics of proteins using site-directed spin labelling rely on explicit modelling of spin label conformations. The large computational effort associated with such modelling with molecular dynamics (MD) simulations can be avoided by a rotamer library approach based on a coarse-grained representation of the conformational space of the spin label. We show here, that libraries of about 200 rotamers, obtained by iterative projection of a long MD trajectory of the free spin label onto a set of canonical dihedral angles, provide a representation of the underlying trajectory adequate for EPR distance measurements. Rotamer analysis was performed on selected X-ray structures of spin labelled T4 lysozyme mutants to characterize the spin label rotamer ensemble on a single protein site. Furthermore, predictions based on the rotamer library approach are shown to be in nearly quantitative agreement with electron paramagnetic resonance (EPR) distance data on Na^+/H^+ antiporter NhaA and on the light-harvesting complex LHCII whose structures are known from independent cryo electron microscopy and X-ray studies, respectively. Suggestions for the selection of labelling sites in proteins are given, limitations of the approach discussed, and requirements for further development are outlined.

1. Introduction

Structure and dynamics of membrane proteins and protein assemblies can be characterized only by joint utilization of various complementary biophysical techniques. Among those, the combination of site-directed spin labelling (SDSL) with electron paramagnetic resonance (EPR) is becoming increasingly important, as it is not intrinsically limited by size of proteins[1-4]. Small reporter molecules carrying an electron spin (spin labels) introduced into otherwise diamagnetic proteins provide information on their local environment and on the mobility of protein domains to which they are attached. When multiple labels are introduced, distance distributions between them can be measured in the range of 1.5 to 6 nm (in favorable cases up to 8 nm)[5,6]. By measuring a set of distances, a model of an unknown protein state or complex can be obtained when the structure in one or more other states, or the structure of a homologous protein, or the structure of the components of the protein complex are known[7-11].

MTSSL



IA-PROXYL

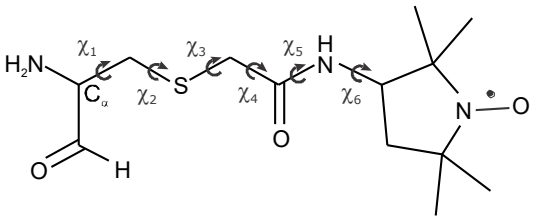


Fig. 1 Molecular constructs for the methanethiosulfonate (MTSSL) and iodoacetamido-PROXYL (IA-PROXYL) spin labelled side chains used in MD simulations. Conformational space is defined by the χ_1-χ_5 and χ_1-χ_6 dihedral angles in case of MTSSL and IA-PROXYL, respectively.

The commonly used spin labels are nitroxide radicals with high conformational flexibility. This flexibility allows for introducing the label with minimal perturbation of the protein structure, as the label side chain can adapt to the free volume in its structural context. However, due to the multiple bond tether, the electron spin residing approximately in the middle of the N-O bond can sample substantial space around the point of attachment (Fig. 1). This has direct effect on the experimental response and is a major problem in the interpretation of distance distributions between spin labels[5,8,12-14].

Although spin labelling is a well established biochemical procedure[15-17], introducing nitroxide radicals into specific sites of a protein is not a trivial task. Trying to place spin labels into spatially restricted positions might lead to narrow distance distributions, but may cause local distortion of the protein structure or result in failure of the labelling. On the other hand, completely unrestricted positions will be easily labelled without affecting the protein structure. However, this may result in broad distributions of label conformations, hence poorly defined distance constraints. Thus, predictions of the conformational distribution of a spin label at a given site are of considerable interest. The computational effort associated with such predictions is a very important issue as well, because generally, the conformational distribution changes when the protein structure changes. Therefore, it may be necessary to repeatedly calculate the conformational distribution of the label when following functional motion of proteins.

Atomistic molecular dynamics (MD) simulations have been often invoked to model spin label conformations together with its protein surrounding when applying SDSL EPR[2,18-21]. Presently, however, such computations have serious limitations. Detailed specification of all interactions governing spin label and protein motion (including solvent) results in a very long computation time

such that medium and large scale conformational motions occurring on a time scale > 200-300 ns become inaccessible. A sampling problem is also encountered in atomistic MD simulations as exhaustive sampling of the conformational space of even relatively small molecules (as nitroxide spin labels) usually cannot be achieved even in trajectories as long as 100 ns[18,22]. Precision of the MD simulations related to the quality of available force fields constitutes another issue. Even with one of the best available force fields for proteins and after overcoming sampling issues, unconstrained MD simulations did not reproduce proton-proton J couplings of the native amino acid side chains obtained by NMR experiments[23]. These J couplings depend on side chain dihedral angle χ_1 via the Karplus relation, hence the distribution of χ_1 in MD trajectories does not agree well with experiment.

Despite all these issues, MD simulations of SDSL EPR experiments were rather successful[18-21]. This indicates that moderate deviations from the true conformational distribution can be tolerated in such simulations and thus justifies development of computationally less expensive approximate approaches. The sophistication of these could then be iteratively increased based on comparison with experimental results. Our own approach is a representation of the conformational space of the spin label side chain by a discrete set of rotameric states – a rotamer library. Experimentally derived very large rotamer libraries for native amino acids are routinely used in modelling the packing of protein side chains[24-27]. Since at present only a few experimental structures of nitroxide-labelled proteins are available, e.g. for several sites in T4 Lysozyme[28-31], derivation of rotamer libraries for spin labels has to rely on modelling. Our first pilot rotamer libraries were created by a systematic conformers search with the MMFF94 force field as implemented in Titan (Wavefunction Inc., Irvine, CA, USA)[5]. For lack of nitroxide parameters in that force field the N-O group was modelled by a keto group C=O. The relative energies of rotamers of the unbound label were neglected and all spin label and protein atoms were kept static during calculation of the interaction energy of a rotamer with the protein. The MTSSL rotamer library created in this way was successfully used to obtain the unknown dimer structure of the Na^+/H^+ antiporter of *Escherichia coli (E. coli)*[8]. The dimer arrangement as determined by EPR-driven molecular modelling was later independently confirmed by a cryo-

electron microscopy (cryo-EM) study[32], indicating that the approximations of the rotamer library approach are tolerable. A detailed comparison of the structures and the discussion of this point are part of the present work.

Despite this initial success, we found that the pilot rotamer libraries did not always provide a sufficient sampling of the conformational space. For instance, in some cases none of the 62 rotamers of the initial MTSSL library would fit into a cavity that was, in principle, sufficiently large to accommodate the spin label side chain. Furthermore, completely ignoring relative energies between the rotameric states appeared to be a too rough approximation, as several studies indicate that certain rotameric states are generally preferred[28-31,33]. In an attempt to address these issues, second-generation rotamer libraries were created using DFT computations[14]. For instance, the MTSSL library with 108 rotamers significantly improved sampling of conformational space. However, the relative rotamer energies obtained with the DFT computations were not realistic, and we obtained much better agreement with experiment by ignoring them. With this library we determined a kinked backbone structure of the transmembrane domain IX of the Na^+/proline symporter PutP of *E. coli*, which closely resembles the structure of the related transmembrane domain in the leucine transporter $LeuT_{Aa}$ of *Aquafex aeolicus*[14]. The same rotamer library was applied to monitor conformational changes of the nucleotide-binding domains during the catalytic cycle of the maltose ATP-binding cassette importer $MalFGK_2$[34].

Here, we introduce the third generation of rotamer libraries. These libraries are obtained by projecting conformational dynamics of the free spin label observed in long MD trajectories onto a set of canonical rotamers. First, examples of modelling of the ensembles of spin label rotamers using the newly obtained MTSSL rotamer library are given for both a solvent exposed and a buried site in T4 lysozyme, for which X-ray structural data are available for the two spin-labelled mutants[28,29], and the essential quantities obtained with the rotamer analysis are discussed. Suitability of the rotamer library approach to assist EPR-driven modelling of proteins is demonstrated next. The dimer structure of the NhaA protein of *E. coli,* first obtained using EPR derived distance constraints combined with molecular modelling[8], is compared with the recently obtained NhaA structure using

cryo-EM[32]. In particular, using the new MTSSL rotamer library we simulate double electron electron resonance (DEER) data from the independently obtained cryo-EM structure and compare them to our earlier experimental results. Using the crystal structure of the light-harvesting complex LHCII, DEER time traces were simulated with the new rotamer library for IA-PROXYL and compared to experimental data. Advantages and limitations of the applied modelling methodology are discussed.

2. Methods

2.1 Development of the rotamer libraries

Rotamer libraries for the MTSSL side chain R1 (in the following MTSSL) and for the IA-PROXYL side chain were derived from MD trajectories with a total length of 100 ns. Two temperatures – 298 and 175 K – were chosen for the MD simulations targeting the most commonly used conditions at which EPR experiments on spin labelled proteins are performed. The latter temperature is an estimate of the glass transition temperature of proteoliposome. Although in distance measurements on proteins the DEER experiment is usually done at 50 K, equilibration of the conformational ensemble occurs to a good approximation at the glass transition temperature during shock-freezing of the sample[35]. The methodology of constructing the libraries is general and can be readily repeated at any temperature provided that an MD trajectory with proper sampling can be acquired.

2.1.1 MD simulations

In order to keep the methodology general, the molecular constructs for the MD simulations consisted of the spin label atoms attached to a single isolated residue, Fig. 1. The standard

CHARMM27 force field[36], customized with a recently derived additional set of parameters for R1 side chain atoms[37] was used for MTSSL. For IA-PROXYL, the bending potential for the angle between S-CT2-C atom types, as well as potential for the torsion specified by CT2-S-CT2-C atom types (χ_3 dihedral for PROXYL) had to be added, see Supporting Information. The peptide backbone atoms of the molecular constructs of both MTSSL and IA-PROXYL along with the C_β atom were immobilized during MD runs by assigning to them the *inactive* status available in the Tinker package used for the MD simulations[38]. This provides a more realistic description of spin label side chain dynamics with respect to the backbone atoms, with the latter being also fixed in our computations of the label-protein interaction energy. In order to account for the undersampling (or quasi-ergodicity) problem[18,20-22], MD trajectories were acquired with repeated heating/annealing cycles. For MTSSL, every cycle consisted of a 1000 ps MD run at 600 K (a "hot" run), a subsequent 200 ps thermalization at the target temperature, and a productive 200 ps run at the target temperature during which data were collected at time increments of 1 ps. For IA-PROXYL durations were 500, 200 and 500 ps for the hot, thermalization and the production runs, respectively. The number of cycles was set to 500 for MTSSL and to 250 for IA-PROXYL, so that total trajectory lengths of 100 ns were achieved in both cases. We checked that all transitions between rotameric states of a given single dihedral angle occurred sufficiently often to derive population statistics. Since transitions at the target temperature are very rare only for one of the dihedral angles, χ_3 in MTSSL, sufficient transitions for this individual angle imply that all accessible rotamers were encountered sufficiently often. The heating cycles lead to transitions between a small number of basins that are internally equilibrated at the target temperature. The concept of basins is based on the fact that transitions between different rotamers occur with different rates, which correspond to different heights of energy barriers between them. The lower the barriers, the more frequent the transitions between such states occur during an MD run. A group of rotamers that are all connected via low-barrier pathways occupies the same energy basin. In our context the criterion for two rotamers to belong to the same basin is a transition rate at the target temperature (175 or 298 K) that is fast compared to the total length of the MD trajectory (100 ns). Note that this rate includes transitions between the two

rotamers that proceed via alternative pathways, i.e. via other rotamers. Accordingly, two energy basins are distinct if transitions between any rotamer in the first basin and any rotamer in the second basin at the target temperature are slow compared to the length of the trajectory. Hence, transitions between different basins occur mainly during the hot runs, so that relative populations of the basins correspond to the high temperature. Therefore the set of rotamer populations derived from the MD simulation is only an approximation to the populations at the target temperature. If necessary, this approximation can be improved by replica exchange MD simulations[39], without affecting other steps in our approach. All MD simulations were performed in implicit solvent using the Still solvation model for water[40]. At an earlier stage tests with explicit water solvation were performed for shorter trajectories. Similar dihedral angle distributions were found as with the Still solvation model. As the current rotamer libraries are applied to solvent-accessible, lipid-accessible, and buried labels in the same way, we consider the computational expense of explicit water solvation as unnecessary at this stage.

2.1.2 Projecting an MD trajectory onto a basis set of rotamers

Full length MD trajectories consisting of 100000 frames were reduced to relatively small representative sets of rotamers by projecting the trajectories onto sets of template rotamers (basis sets). At the first stage, distributions of dihedral angles χ_1-χ_5 for MTSSL and χ_1-χ_6 for IA-PROXYL were calculated and displayed as histograms, resulting in dihedral angle profiles of the whole trajectory. Such profiles for MTSSL at 175 K are shown in Fig. 2.

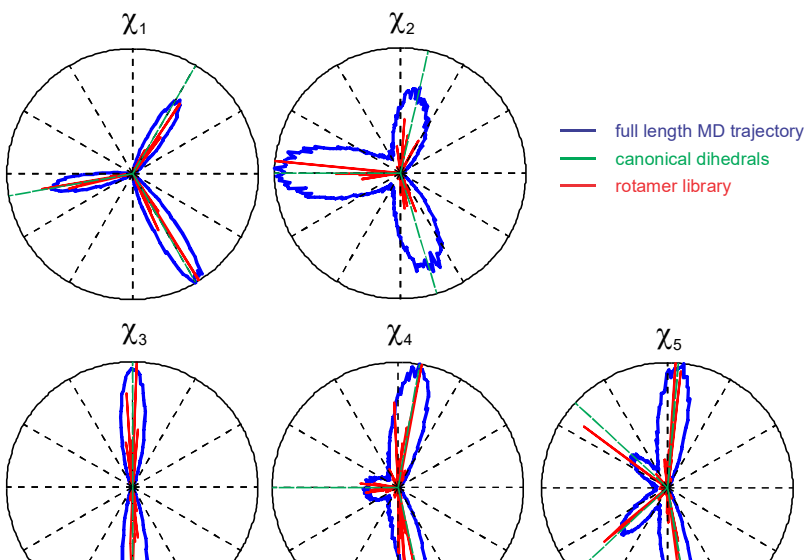


Fig. 2 Distributions of dihedral angles χ_1-χ_5 (*dihedral profiles*) for MTSSL calculated from the full length MD trajectory acquired at 175 K. Canonical dihedral angles values as well as dihedral angles distributions of the rotamers comprising final rotamer library are shown superimposed.

They are similar to profiles obtained in earlier MD simulations for MTSSL attached at a solvent-exposed site of a protein[21]or to an oligo(alanine) α-helix[37]. Each dihedral angle is mainly distributed about a few characteristic values which we term "canonical dihedral angles" (shown in green in Fig. 2). We failed in building the rotamer library from rigid spin label fragments based on all possible combinations of the canonical dihedrals. This failure was due to internal clashes between atoms in a significant number of rotamers constructed in that way (about 50% in case of IA-PROXYL at 298 K), with such clashes being defined as an approach of two atoms to a distance closer than 0.65 times the sum of their van-der-Waals radii. After removal of the internally clashing rotamers, the remaining ones did no longer provide a good representation of the conformational space obtained in the MD simulation. This observation is explained by an intrinsic correlation between dihedrals. Rotamers for which all dihedrals are exactly canonical dihedrals may not necessarily exist.

Instead, a small or moderate deviation from the first canonical values of, for example, χ_1 may be correlated to a similar deviation of the first canonical value of χ_2 and yet another systematic deviation from the second canonical value of χ_2. Hence, the canonical dihedrals are averages over the total conformational ensemble and are not necessarily representative for the sub-ensemble that corresponds to a single rotamer. Therefore, canonical dihedrals were used only as a set of template

rotamers around which MD frames were subsequently clustered. For each MD frame, the closest template rotamer was identified based on the least square deviation between the dihedrals of the current frame and the template rotamer. The frame was then assigned to a group of structures belonging to that template rotamer.

After partitioning the whole trajectory into sub-ensembles in that way, the dihedrals of a given template rotamer were redefined by averaging over all frames assigned to that rotamer. In general, these averaged dihedrals do not correspond to the dihedrals in any individual frame in the trajectory. However, they are a better representation of a group of similar structures than the canonical values. Hence, clustering of the trajectory frames was repeated with respect to the redefined template rotamers and average dihedrals were recomputed for the new partitioning. From each cluster of MD frames, the one with the smallest mean square deviation from the average dihedrals was taken as a representative structure and thus as one rotamer in the library. The number of MD frames assigned to each cluster reflects the statistical weight of the corresponding rotamer in the underlying MD trajectory. After normalization by the total number of frames it is equal to an internal population p_i^{in} of the rotamer at the target temperature.

An initial estimate of the completeness of the representation of the full-length MD trajectory by the corresponding set of rotamers is illustrated in Fig. 2 where dihedrals of the rotamers representing the MTSSL library at 175 K, scaled according to their internal populations, are superimposed onto the dihedral profiles of the whole MD trajectory. Results for rotamer libraries of MTSSL at 298 K as well as for PROXYL at 175 K and 298 K are of similar quality, see Supporting Information.

2.1.3 Completeness control of the MD projection

The rotamer libraries are primarily applied to EPR distance distribution measurements. From this point of view, completeness of the rotamer representation was tested more rigorously by comparing DEER data for pairs of unrestrained spin labels simulated either via the rotamer library or

via the underlying MD trajectory. For this, the coordinates of a residue can be expressed in a local frame constructed from the coordinates of the backbone residue atoms (Fig. 3a). The *x* axis of this frame coincides with the C_α-N bond, and the *xy* plane is the peptide plane defined by the C_α, N and carbonyl C atoms. Without losing generality, the first residue of the pair can be fixed at the coordinate origin (C_α atom), whereas coordinates of the second residue are computed by translation and subsequent rotation of its local frame (Fig. 3b).

Several randomly selected relative positions and orientations of the local frames, characterized by a set of three Euler angles and a translation vector, were generated and subsequently populated with the MD trajectory frames as well as with the rotamers of the corresponding rotamer library. Pairs of coordinate frames from the trajectories were randomly selected using a Monte-Carlo strategy, as complete coverage would correspond to $5 \cdot 10^9$ frame pairs. Both the distance distributions and the corresponding DEER time traces were simulated. Contributions from every randomly selected pair of MD trajectory frames were added with equal weights, whereas those of rotamer pairs from the rotamer library were weighted by the product of the internal populations of the rotamers. DEER form factors were computed from distance distributions using the get_DEER.m subroutine from the freely available DEERAnalysis2009 package [41] and assuming a modulation depth of 0.4, as it is commonly encountered in DEER measurements at X-band frequencies. Such simulations for two representative relative positions of residues, calculated from the MD trajectory for MTSSL at 175 K, are shown in Fig. 3c together with those calculated using the corresponding rotamer library.

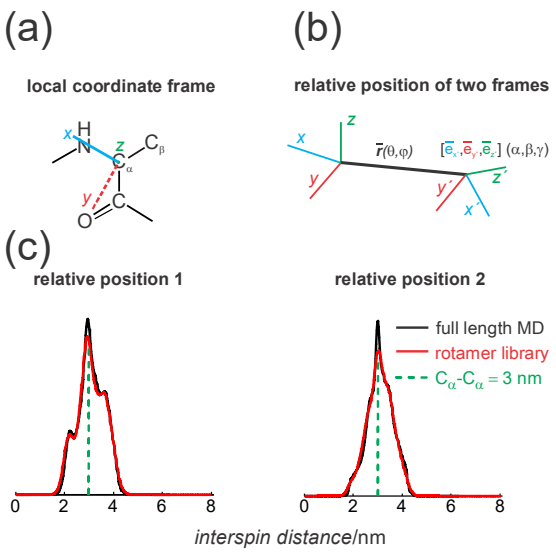


Fig. 3 Completeness check for the MTSSL rotamer library derived from the 175 K MD trajectory. For two relative positions of residues, distance distributions and associated DEER time traces calculated from the full length MD trajectory are compared to those calculated using the corresponding rotamer library. Relative position 1: θ=60°, φ=47°, α=17°, β=38°, γ=51°; relative position 2: θ=0°, φ=90°, α=180°, β=0°, γ=0°. At both relative positions, the magnitude of the translation vector $\vec{r}(\theta,\varphi)$ is kept the same (to 3 nm).

Small differences are seen in the distance distributions, but these differences are practically invisible in the DEER form factors. As the form factors are closer to the primary data than the distance distributions, the rotamer library represents the full MD trace within the actual precision of the DEER experiment. In other words, since the transformation of DEER form factors to distance distributions is a moderately ill-posed problem[42], the technique is unable to detect the small differences between the rotamer representation and the MD trajectory. Similar results were obtained for MTSSL at 298 K as well as for PROXYL at both 175 and 298 K (not shown). Hence, our procedure of reducing a full-length MD trajectory to a small, representative set of rotameric states – the rotamer library – can be judged to provide a complete description of the conformational distribution for a spin label that does not interact with neighboring residues in a protein. This procedure is fast and can be repeated to create improved rotamer libraries whenever MD trajectories of better quality become available. The description thus obtained is complete within the precision of pulsed dipolar EPR distance measurements, not necessarily in the context of other experiments. Note

finally, that the procedure is also quite robust: virtually identical rotamer sets were obtained by varying the initial choice of the canonical dihedral angle values within a reasonable range.

2.2 Rotamer analysis at a labelling site

Internal populations p_i^{in} reflect corresponding internal relative energies ε_i^{in} of the rotameric states of the library. If the internal energy of the most populated rotamer with population $p_{\max}$ is set to zero, relative energies of all other rotamers can be computed by Boltzmann inversion: $p_i^{in} / p_{\max} = \exp(-\varepsilon_i^{in} / kT)$.

Another important energy contribution arises from interaction of the label with its environment in the protein. Within our approach, this external energy term is computed assuming fixed positions for the interacting atoms. More precisely, the external rotamer energy of the *i*-th rotamer ε_i^{ext} is calculated solely as a sum of pairwise Lennard-Jones interactions (using the CHARMM force field parameterization) between all spin label and protein atoms. To speed up computations, a distance cutoff is assumed for the pairwise interaction. Being a library specific quantity, the cutoff radius was empirically defined as the distance from the C^{α} atom of the residue to be mutated to the most distant spin label atom among all the rotamers in the library, plus 4 Å. Protein atoms that are further away from the spin label C^{α} atom than this cutoff radius are neglected in the computation of the external energy. Furthermore, all pairwise interactions involving hydrogen atoms are neglected. As atoms always undergo some librational motion, and as conformational dynamics is represented discretely in terms of rigid rotamers, the pairwise potentials need to be somewhat softer than what would be expected from theory. Such a use of empirically softened Lennard-Jones potentials is common practice when predicting protein side chain packing[24-27]. In our approach, we soften interatomic potentials by scaling down the equilibrium interatomic distance in the Lennard-Jones term by a "forgive" factor. We found that experimental DEER data are best reproduced with our libraries when using forgive factors in the range of 0.5 to 0.6. In the following we use a value of 0.5.

By adding the external and internal rotamer energy terms, the probability of a particular rotamer to occur is given by: $p_i = p_i^{in} \cdot p_i^{ext} = p_i^{in} \cdot \exp(-\varepsilon_i^{ext} / kT) / Z$, where Z is the statistical partition function $Z = \sum_i p_i^{in} \cdot \exp(-\varepsilon_i^{ext} / kT)$, with i indexing all rotamers. Both the internal and external rotamer populations have to be calculated at the same temperature T in order to obtain properly weighted statistical distributions of the rotamers. Rotamer libraries for MTSSL and IA-PROXYL derived from the MD trajectories acquired at 298 K and 175 K are then directly applicable for modelling conformations of the spin labels at room temperature and for membrane proteins in liposomes at 50 K, respectively. For soluble proteins in glassy frozen mixtures of water and glycerol, a temperature of 210-240 K may be more appropriate, depending on the glycerol content. However, given the other approximations in the approach, we believe that the 175 K library is still a reasonable choice for soluble proteins. Note also that the glass transition of the solvation shell of a protein in aqueous solution is close to 175 K[43].

At any other temperature, the whole procedure would have to be repeated– computation of an MD trajectory and iterative projection to a rotamer library. Rescaling of rotamer populations derived at some other reference temperature via the Boltzmann law generally does not work. For example, high temperature internal rotamer populations cannot be obtained from a finite-length low temperature MD trajectory simply because the rotameric states needed may not be sampled at low temperature. More generally, spin label rotamers derived here are essentially a coarse-grained representation of the underlying MD trajectory – they correspond to averages over sub-ensembles. The individual states in these sub-ensembles have different energies. Therefore, internal rotamer populations will not in general scale according to the Boltzmann law, as a sum of exponential functions with different exponents cannot be written as an exponential function. A more rigorous discussion of temperature and state dependence of coarse graining can be found in[44].

The partition function Z depends on the construction of the library. For a given library, Z is a measure of the tightness of the labelling site. A small number Z indicates high net interaction energy between the label and the protein environment, i.e. an unfavorable site. For the MTSSL 175 K

rotamer library, we consider sites with $Z < 0.05$ as tight sites where labelling may fail or lead to significant distortion of the protein structure. Such tight sites should be experimentally avoided.

On the other hand, very loose sites are not favorable either, as they correspond to broad conformational distributions, thus broad spatial distributions of the electron spin. This limits the precision of structural models derived from pulsed dipolar EPR distance measurements. The spatial distribution of the electron spin can be quantified by the root mean square deviation (r.m.s.d) of the N-O bond midpoint from its average value (*NO rmsd*):

$$\sigma_{NO} = \left(0.005 nm^2 + N/(N-1)\sum p_i[(x_i - \langle x \rangle)^2 + (y_i - \langle y \rangle)^2 + (z_i - \langle z \rangle)^2]\right)^{1/2}$$

in which the offset of 0.005 nm^2 accounts for the libration motion of the nitroxide such that σ_{NO} is not equal to zero even when only one rotamer is populated. Small values of the *NO rmsd* associated with at least moderate values of Z indicate favorable labelling positions.

Finally, the rotamer analysis returns a number of significant rotamers as those accounting for 99.5% of the total population at each site. This figure is particularly useful to identify loose positions, characterized by a large number of rotamers with similar low populations. The number of significant rotamers is also useful to indicate tight positions, where the prediction of the average N-O bond midpoint position may not be very reliable. If only one rotamer or a small number of rotamers are significantly populated, the predicted position will be very sensitive to the approximations of a rigid protein and to the quality of the underlying structure. Furthermore, errors in relative internal rotamer energies, resulting from limited precision of the underlying force field can then lead to substantial prediction errors. This leads to the conclusion, which may at first seem counterintuitive, to avoid very tight sites in structural modelling. Pairs of such sites are expected to provide very precisely known label-to-label distances. However, the large uncertainties in translating theses distances to backbone-to-backbone distances overcompensate this advantage, at least with the current quality of our predictions. For sites with moderate spatial restrictions, and thus a moderate number of significantly populated rotamers, the latter uncertainties partially average. A similar argument has been put forward in an earlier experimental study[11]. A more detailed discussion on the

information about sites in membrane proteins that can be obtained by rotamer analysis can be found in[45].

2.3 Implementation of the rotamer library approach into the MMM package

The rotamer library approach for modelling spin label conformations is implemented into the open-source package Multiscale Modeling of Macromolecules (MMM)[46]. This is a Matlab (Mathworks™) based collection of routines accessible through a graphical user interface dedicated to model structure and structural changes of proteins and protein assemblies based on experimental restraints from multiple techniques. Rotamer libraries for MTSSL and IA-PROXYL derived from the MD trajectories at 298 and 175 K are available in this package. In order to perform rotamer analysis, a template protein structure in PDB format has to be loaded into the program. Single or multiple sites (up to the entire sequence) can be selected for rotamer analysis with the desired type of spin label. Rotamer analysis on a single site takes about 15 seconds for a solvent exposed position and up to about 50 seconds for a tightly packed position in the protein core using a common desktop computer. The computed rotamer distributions can be used to calculate distance distributions between pairs of sites which are automatically transformed into dipolar evolution time traces (form factors) of the DEER experiment. Direct comparison of such simulations with experimental data is possible within MMM. Protein and label visualizations in this work were prepared by the graphical means of MMM.

3. Results and discussion

In order to demonstrate performance of the rotamer analysis at a mutation position in a protein as well as to show how essential quantities such as the partition function Z, number of significant rotamers and the *NO rmsd* can be used for a preliminary assessment of the rotamer ensemble, T4

lysozyme was chosen for which several crystal structures containing attached MTSSL are available. Usage of the rotamer libraries in EPR-assisted modelling of proteins – the primary application of the developed approach – is shown in the Section 3.2 and 3.3 on the examples of sodium-proton exchanger NhaA and the light-harvesting complex LHCII proteins, respectively.

3.1 Spin label at restricted and loose positions in protein

The solvent exposed loop site 82 in T4 lysozyme (PDB# 1ZYT)[28] and the solvent-inaccessible buried position 118 (PDB# 2NTH)[29] serve as examples of spatially free and spatially restricted labelling positions, respectively. The PDB files were used as input for *in silico* labelling based on the rotamer library. Cysteines that already carried the spin label were mutated back to the original residues by editing the PDB file. Crystal water was ignored in the calculation of the spin label rotamer populations. Both positions were checked for the possible influence of crystals contacts on the determination of the crystal structure rotamers. For that, the unit cell was first reconstructed using information available in the PDB files. Subsequently copies of the unit cell were generated to ensure that all neighboring positions of the parent T4L molecule were occupied. The results of the rotamer analysis on the parent molecule isolated and in the presence of all of its neighbours in the crystal were identical for both positions 118 and 82. The analysis for position 118 with the MTSSL library at 175 K revealed 37 significant rotamers covering 99.5% of the total rotamer population with partition function Z = 0.86. These rotamers are shown in Fig. 4a with transparency encoding populations p_i along with the MTSSL rotamer from the crystal structure.

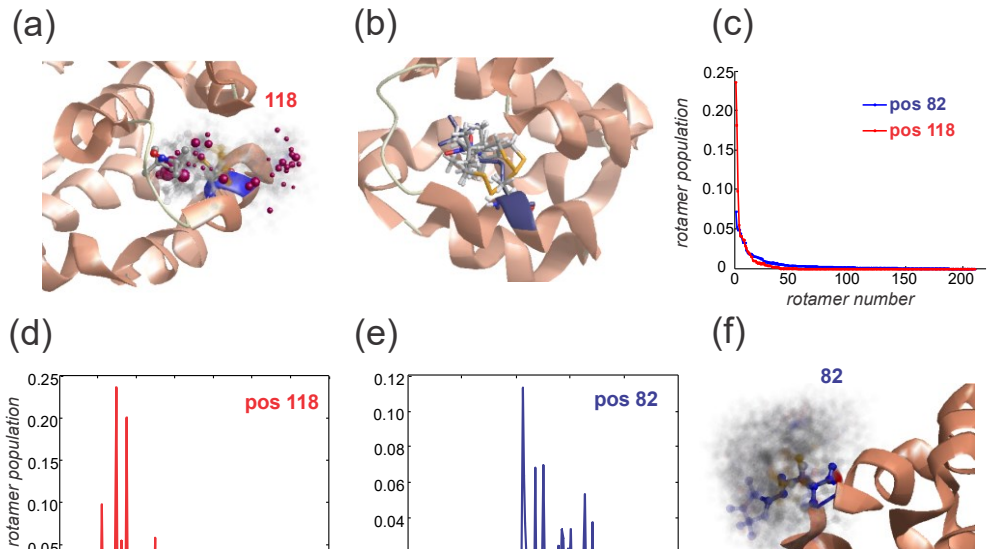


Fig. 4 Spin label at sites with different flexibility in T4 lysozyme. (**a**) MTSSL rotamers from the MTSSL 175 K rotamer library (transparent) and crystal structure rotamer (PDB # 2NTH, ball and stick) at position 118. Purple balls show positions of the electron spins scaled according to populations of the respective rotamers. (**b**) Three most populated rotamers and the crystal structure rotamer (blue) at position 118 in a stick representation. (**c**) Population density distribution for MTSSL rotamers at positions 118 and 82. (**d-e**) Plots of rotamer population vs $\Delta r_{NO}=r_{NO}-\langle r_{NO}\rangle$ for positions 118 and 82 respectively. (**f**) MTSSL rotamers (transparent) and crystal structure rotamer (PDB #1ZYT, ball and stick, blue) at position 82.

The most populated computed rotamers occupy the same hydrophobic cage as the single rotamer modelled into the electron density cloud obtained by X-ray diffraction. Furthermore, the N-O bond midpoint positions of these rotamers are close to the one observed in the crystal structure. In particular, the three most populated rotamers, which together contribute 52% of the total population, closely match the location of the X-ray rotamer (Fig. 4b), although the sets of dihedral angles differ. Location of the N-O bond midpoint averaged over the three rotamers deviates from the one of the crystal structure rotamer by 2.6 Å. For this position, computed rotamer populations are distributed very unevenly (red line in Fig. 4c), with the leading rotamer contributing almost 25%. It is interesting to note that despite the apparent tightness of the cage, the position 118 is not a particularly restricted one, as is already apparent from the partition function. We find an *N-O rmsd* value of 0.49 nm showing that the remaining low population rotamers, whose sum population is 48%, feature a broad

spatial distribution. Plotting rotamer populations against $\Delta r_{NO} = \vec{r}_{NO} - \langle \vec{r}_{NO} \rangle$ indeed reveals a narrowly peaked distribution from the leading rotamers on top of a rather broad background, Fig. 4d. The broad background largely stems from rotamers that do not occupy the cage that is occupied by the X-ray rotamer. Protein and side chains conformational sampling in crystal structures may be modulated relative to the native environment (as shown for example by Cafiso and co-workers[47]), and the spin labels "trapped" in aqueous frozen solutions might not always correspond to the crystal side chains at those sites. To test whether the predicted conformations are populated in frozen aqueous solutions, as opposed to the crystalline protein, distance distribution measurements on double mutants involving site 118 would need to be performed.

A different situation is encountered for position 82. This site is located on a turn connecting two helices (long helix Thr59-Asn81 and short helix Lys82-Ser90). The spin label is strongly solvent exposed and there are no other parts of the protein structure in the vicinity of the labelling site which would restrict spin label flexibility. Rotamer analysis revealed that 163 out of the 210 rotamers in the library are significantly populated with a partition function Z = 1.27. Uncertainty of the position of the electron spin as qualified by the *N-O rmsd* is 0.59 nm. Spatial occupancy of the rotamers is not at all compact in that case. Compared to position 118, population density at position 82 is distributed much more evenly over the rotameric states; there are no clearly leading rotamers at that site, Fig. 4c. The distribution of Δr_{NO} over the rotamers is broad and more homogeneous, it does not exhibit a clear sharp-peak character, and on average it has much stronger off-zero shift compared to the distribution at position 118, Fig. 4e. The resulting rotamer distribution for position 82 is shown in Fig. 4f. The X-ray study also revealed larger spatial freedom of the spin label at position 82 compared to that at position 118: the electron density was only resolved up to C_ε atom of the spin label, whereas the R1 side chain at position 118 was modelled in its entirety. Both the 82R1 and 118R1 structures correspond to a temperature of 100 K[28,29].

Nevertheless, the situations in the crystal appear to correspond to narrower conformational distributions than are predicted by the rotamer library approach, in particular for 82R1. Again,

distance distribution measurements on double mutants involving position 82 in a frozen solution could help to resolve this issue.

3.2 Structure of Na+/H+ antiporter from EPR restrains and from the cryo-EM map

For the main sodium-proton exchanger NhaA in the inner membrane of *E. coli* the arrangement of the native dimeric state was first obtained by combining site-directed spin labelling, pulsed EPR distance measurements using the DEER experiment, and simulation of the spin label conformations with the first-generation rotamer library approach[8]. The structure of the dimer was solved by rigid-body docking of NhaA monomer molecules (PDB # 1ZCD, resolution 3.45 Å)[48], assuming C_2 symmetry in the homodimer. Altogether nine primary experimental DEER time traces for pairs of MTSSL-labelled sites were used as restraints. The NhaA dimer geometry found in the EPR study (Fig. 5, orange) features a very small dimer interface.

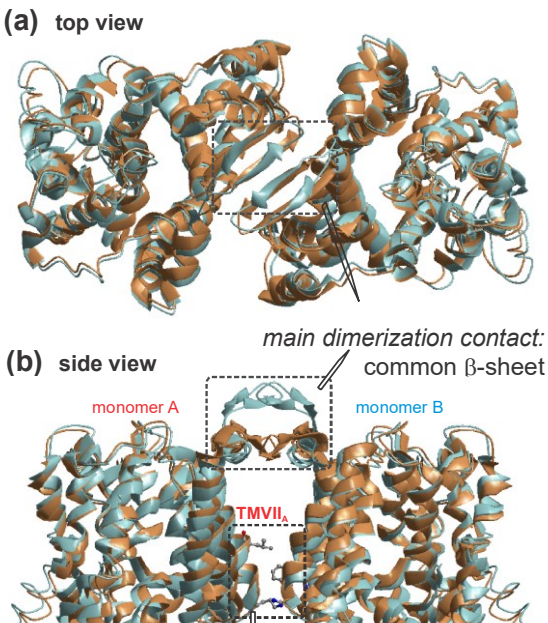


Fig. 5 Structures of sodium-proton exchanger NhaA from E. coli obtained by EPR (orange) and by cryo-EM (steelblue) are shown superimposed. (**a**) View from the periplasmic side along the membrane normal, and (**b**) – perpendicular to the membrane normal. Minor dimerization contact at the cytoplasmic part of the protein is shown in insert in **b**.

Essentially, two contacts were revealed: a main one – between two β-hairpins at the periplasmic side of each monomer forming a shared solvent exposed β-sheet, as well as a minor hydrophobic contact at the cytoplasmic part between Leu210 of one monomer with Trp258 of the other molecule and a hydrogen bond between Arg204 and a backbone oxygen of the other molecule, Fig. 5. Physiological relevance of the β-sheet contact was later confirmed by biochemical experiments showing that deletions or mutations in that region can destabilize the dimer and impair viability of the bacteria under high-salt conditions[49]. The basic arrangement of the NhaA dimer was later independently confirmed by using the 3.45-Å X-ray structure of the NhaA monomer as a basis for molecular replacement in the 7-Å cryo-EM map acquired on 2D crystals[32]. The all-atom superposition, done with MMM, of the newer cryo-EM structure (PDB # 3FI1) with our EPR-

derived structure gives an r.m.s.d. of 2.00 Å. Details of the minor dimerization contact on the cytoplasmic side of the membrane were confirmed as well, Fig. 5.

The most essential difference between the two dimer structures is in the main dimerization contact region, Fig. 5b. In both structures the contact between the β-hairpins of the two monomers was found to be essential for dimer formation by a shared β-sheet. However, due to the absence of the structure restraints in that region in the EPR study, the original conformation of the β-hairpins in the monomers of the template X-ray structure was kept unchanged (rigid-body approximation). In the cryo-EM study, electron density corresponding to the β-sheet was found in a different location. Adjusting the mismatch between the original conformations of the β-hairpins in the crystal structure of the NhaA monomer and the experimental electron density results in a more solvent exposed conformation of the shared β-sheet.

Although we cannot exclude differences between the structures in liposomes (EPR study) and 2D crystals at different pH values, it appears likely that the original conformations of the β-hairpins in the monomer X-ray structure are due to a packing artifact that transferred into the EPR structure via the rigid-body approximation. This indicates that the conformation of potentially mobile domains needs to be characterized by strategically placed spin labels in future studies and that flexible docking or structure fitting from EPR data needs to be developed.

The EM structure, which was obtained independently from our EPR structure and on the basis of a completely different set of constraints, opens up the possibility to test the prediction power of the rotamer library approach. For this we assume that the dimer structure in shock-frozen liposomes, corresponding to the conditions of the DEER experiments, is the same as the one in the 2D crystals on which EM was performed. If that is the case, it should be possible to predict experimental DEER data from the EM structure without any further structure fitting. Such a simulation provides a serious test for the reliability that can be expected for rotamer library simulations of the label-to-label distance distributions in membrane proteins.

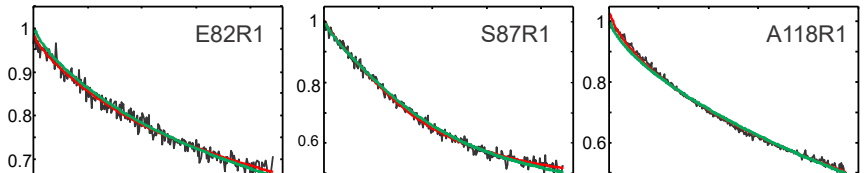


Fig. 6 DEER signals simulated with MTSSL rotamer library derived from the 175 K MD trajectory using cryo-EM structure of NhaA dimer (PDB #3FI1) as a template (green); best fit DEER signals (red) and experimental DEER traces (black) of the original EPR-based structure determination of the NhaA dimer molecule[8].

With the 175 K rotamer library of MTSSL and using the EM structure of NhaA dimer as a template, sites used in the EPR study were labelled *in silico*, interspin distance distributions were constructed and converted into DEER form factors. These were used to simulate DEER time traces by fitting modulation depth and the decay rate constant of the background function to the experimental DEER traces of the EPR study. In Fig. 6, DEER signals simulated in that way (green lines) are superimposed with the experimental DEER traces (black lines) measured between corresponding positions in the protein. Additionally, the best fit DEER traces obtained in the original EPR-based structure determination of the NhaA dimer are shown for reference (red lines).

The overall agreement between the sets of simulated and experimental DEER signals is good. Both the mean distances between sites and the associated distribution widths are generally well reproduced. Some overestimate of spin label flexibility at site 221 results in less pronounced oscillations of the simulated DEER signal (green line), as was already seen in the original best fit trace (red line). Nevertheless, the mean interspin distance is predicted correctly both now and in the original study. Position 177 is the only one for which a more significant deviation is encountered in the simulation based on the EM structure, although even this deviation is not severe. This deviation could be either due to a difference between the EM structure and the structure in liposomes, as measured by EPR, or due to different sampling of the conformational space at site 177 with the newly derived MTSSL rotamer library compared to the older MMFF94 based library. Detailed analysis reveals that the new library with a larger number of rotamers leads to a more complete sampling of the free space, so that additional rotamers become populated at position 177. This causes the slight short-distance shift of the mean interspin distances and the corresponding increase of the dipolar frequency in the simulated DEER signal. Note that the apparently better fit with the smaller first-generation library at this site is accidental. The error compensation resulting in this better agreement is not likely to be general and thus not a reason for preferring the old library.

This change is not expected to significantly influence the structural model of the NhaA dimer, which was originally largely overdetermined by using nine distance restraints for fitting of four parameters to minimize the effects due to uncertainty in rotamer prediction. Moreover, a systematic variation of the number of the distance constraints was used in order to estimate the final precision of the structure ensemble of the molecule, i.e. the constraint 177R1-177R1 is missing in several structures of the ensemble[8].

3.3 Rotamer library predictions for a “rigid” protein structure

In general, the agreement between simulation and experiment for the NhaA dimer is better than would be expected from comparison of the ensemble of rotamers predicted for T4 lysozyme with the single crystal rotamer revealed in the X-ray structures of the spin labelled mutants. While this may indicate that membrane proteins in frozen liposome suspensions and probably glassy frozen solutions of soluble proteins feature broader conformational distributions of side chains than crystalline proteins, we may not exclude that part of the success of the NhaA simulations is due to another kind of error compensation: the rotamer library predicts a broader conformational distribution than is realistic and this compensates for the protein backbone flexibility that we neglect in our computations. If this is the case, rotamer predictions for proteins that are more rigid than the membrane transporter NhaA should exhibit larger deviations. Aiming to check this hypothesis, we simulated distance distributions and associated DEER time traces between sites of the very rigid light-harvesting complex LHCII, using the IA-PROXYL rotamer library derived from the corresponding 175 K MD trajectory, and compared them with experimental DEER data, Fig. 7.

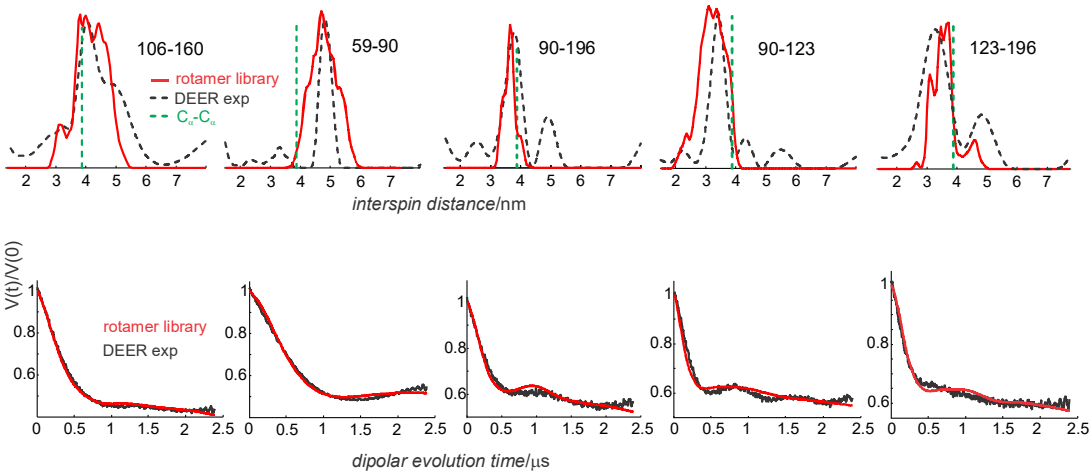


Fig. 7 Simulated interspin distance distributions (upper panel) and associated DEER signals (lower panel) for light-harvesting complex LHCII calculated with IA-PROXYL rotamer library derived from the 175 K MD trajectory versus experimental data. C_α-C_α distances determined from the template crystal structure (PDB #2BHW) are shown for reference.

The light harvesting complex can be categorized as "rigid" due to the presence of chlorophyll and carotenoid cofactors fixed in the protein in a specific geometry to allow proper energy transfer. Arguments for such rigidity were also advanced on the basis of the far-reaching agreement of crystal structures of LHCII from two different organisms and in different types of crystals[50]. The crystal structure of LHCIIb (PDB #2BHW) was used as a template for atom positions during the simulations and DEER time traces were simulated as described above for the case of NhaA. While in some cases the width of the distance distributions is somewhat overestimated indeed and correspondingly the simulated DEER traces appear oversmoothed (pairs 90-123 and particularly 59-90), the deviations are neither strong nor systematic, such that the predictions made for LHCII are of similar quality as for NhaA. Rotamer library predictions for site V196 used for comparison with experimental DEER data in the present work, were successfully assisting a study of folding kinetics of LHCII performed with electron spin echo envelope modulation (ESEEM).[51] For further checks and improvement of the rotamer library approach, more experimental data and corresponding simulations are needed. For that purpose, our methodology is made publicly available via the open-source program package MMM[46]. In fact, with a sufficiently large body of distance distribution data for independently known structures, it would become possible to treat the set of rotamers as a basis set and adjust relative populations of basins or even individual rotamers by fitting of experimental data. Such an approach would be analogous to deriving relative populations for native side chain rotamers from crystal structures.[24-26]

Conclusion and outlook

Having a fast and reliable approach for prediction of spin label conformations facilitates application of SDSL EPR methods in protein studies. Such an approach can be based on representation of conformational space of the spin label by a discrete set of rotamers and prediction of the conformational distribution by calculating a statistical ensemble of the rotamers at the site of

interest. We have shown that a small library, obtained by iterative projection of a long MD trajectory onto a set of canonical rotamers, is a good representation of the entire trajectory. Libraries for MTSSL and IA-PROXYL were constructed that are suitable for modelling results of SDSL EPR at ambient (298 K) and cryogenic (50-80 K) temperatures. The whole set of rotamers along with the associated populations allows for fast prediction of distance distributions and corresponding DEER form factors for pairs of sites. Additionally, the partition function Z of the rotamer ensemble, the spatial distribution of the electron spin position, characterized by the *NO rmsd*, and the number of significant rotamers, obtained upon the rotamer analysis of an individual site, provide information on tightness of the site and can thus be used in selecting sets of mutation positions in a protein. Both very tight and very loose sites should be avoided in spin labelling.

Performance of the rotamer library was tested on T4 lysozyme for which crystal structures containing MTSSL at different positions are available. In accordance with the X-ray data, solvent exposed 82R1 site was correctly differentiated from the much tighter solvent-inaccessible 118R1 site as a result of the rotamer analysis using MTSSL library derived from 175 K MD trace. However, for both positions, spin label rotamer ensembles with different degree of compactness were predicted instead of one single rotamer of the X-ray study at each site.

Robustness of the approach and its use in structural studies of proteins were discussed on the example of the Na^+/H^+ antiporter NhaA protein for which a set of nine DEER distance measurements and an independently obtained cryo-EM structure exist. Despite generally good agreement, the comparison suggests that because of the approximations involved in the rotamer library approach, structural modelling relies on overdetermination of the structure by distance constraints rather than on the extreme accuracy of the single constraint predictions.

Due to the low computational costs of the rotamer library approach the prediction of spin label conformations is no longer a bottleneck studies of protein structure by EPR, not even if multistep fitting with repeated spin label modelling is required. The rotamer library approach as described here is implemented in the open source software package MMM.

Acknowledgment

We kindly acknowledge Peter Fajer and Denis Sezer for fruitful discussions. Supported by grants from Swiss National Foundation (200021_121579) and from Deutsche Forschungsgemeinschaft (JE 246/3-2).

Electronic supplementary information (ESI) available: Details on additional parameterization of the CHARMM27 force field parameters for PROXYL; dihedral angle profiles for the full length MD trajectories for MTSSL and PROXYL at 175 and 298 K along with corresponding dihedral angle plots for the final rotamer libraries of the labels.

References

1. L. Columbus, W. Hubbell, *J. Am. Chem. Soc.*, 2002**, 27**, 288-295.

2. D. Fangyu, M. Layten, C. Simmerling, *J. Am. Chem. Soc.*, 2008, **130**, 7184-7185.

3. M. Kim,G. E. Fanucci, D. S. Cafiso, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 11975-11980.

4. M. Swanson, V. Kathirvelu, T. Maitan, F. Frerman, G. Eaton, S. Eaton, *J. Am. Chem. Soc.* 2009, **131**, 15978-15979.

5. G. Jeschke, Y. Polyhach, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1895-1910.

6. O. Schiemann, T. Prisner, *Q. Rev. Biophys.*, 2007, **40**, 1-53.

7. E. Perozo, D. Cortes, L. Cuello, *Nat. Struct. Biol.*, 1998, **5**, 459-469.

8. D. Hilger, Y. Polyhach, E. Padan, H. Jung, G. Jeschke, *Biophys. J.*, 2007, **93**, 3675-3683.

9. J. Bhatnagar, P. Borbat, A. Pollard, A. Bilwes, J. Freed, B. Crane, *Biochemistry*, 2010, **49**, 3824-3841.

10. E. Georgieva, T. Ramlall, P. Borbat, J. Freed, D. Eliezer, *J. Am. Chem. Soc.*, 2008, **130**, 12856-12857.

11. C. Altenbach, A. Kusnetzow, O. Ernst, K. Hofmann, W. Hubbell, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 7439-7444.

12. K Sale, L. Song, Yi-S. Liu, E. Perozo, P. Fajer, *J. Am. Chem. Soc.*, 2005, **127**, 9334-9335.

13. P. Borbat, H. Mchaourab, J. Freed, *J. Am. Chem. Soc.*, 2002, **124**, 5304-5314.

14. D. Hilger, Y. Polyhach, H. Jung, G. Jeschke, *Biophys. J.*, 2009, **96**, 217-225.

15. J. Klare, H.-J. Steinhoff, *Photosynth. Res.*, 2009, **102**, 377-390.

16. C. Altenbach, S. Flitsch, H. Khorana, W. Hubbell, *Biochemistry*, 1989, **28**, 7806-7812.

17. M. Hemminga, in *ESR Spectroscopy in Membrane Biophysics*, ed. M. H. Hemminga and L. J. Berliner, Springer, 2007, ch. 1, pp. 1-5.

18. D. Sezer, J. Freed, B. Roux, *J. Am. Chem. Soc.*, 2009, **131**, 2597-2605.

19. D. Budil, K. Sale, K. Khairy, P. Fajer, *J. Phys. Chem. A*, 2006, **110**, 3703-3713.

20. S. DeSensi, D. Rangel, A. Beth, T. Lybrand, E. Hustedt, *Biophys. J.*, 2008, **94**, 3798-3809.

21. C. Beier, H.-J. Steinhoff, *Biophys. J.*, 2006, **91**, 2647-2664.

22. M. Fajer, H. Li, W. Yang, P. Fajer, *J. Am. Chem. Soc.*, 2007, **129**, 13840-13846.

23. J. Allison, W. van Gunsteren, *Chem. Phys. Chem.*, 2009, **10**, 3213-3228.

24. M. Bowler, F. Cohena, R. Dunbrack, *J. Mol. Biol.*, 1997, **267**, 1268-1282.

25. R. Peterson, P. Dutton, J. Wand, *Protein Sci.*, 2004, **13**, 735-751.

26. J. Ponder, F. Richards, *J. Mol. Biol.*, 1987, **193**, 775-791.

27. D. Gordon, S. Marshall, S. Mayo, *Curr. Opin. Struct. Biol.*, 1999, **9**, 509-513.

28. M. Fleissner, D. Cascio, W. Hubbell, *Protein Sci.*, 2009, **18**, 893-908.

29. Z. Guo, D. Cascio, K. Hideg, T. Kalai, W. Hubbell, *Protein Sci.*, 2007, **16**, 1069-1086.

30. Z. Guo, D. Cascio, K. Hideg, T. Kalai, W. Hubbell, *Protein Sci.*, 2007, **17**, 228-239.

31. R. Langen, O. Kyoung Joon, D. Cascio, W. Hubbell, *Biochemistry*, 2000, **39**, 8396-8405.

32. M. Appel, D. Hizlan, K. Vinothkumar, C. Ziegler, W. Kühlbrandt, *J. Mol. Biol.*, 2009, **386**, 351-365.

33. A. Polimeno, J. Freed, *J. Phys. Chem.*, 1995, **99**, 10995-11006.

34. M. Grote, E. Bordignon, Y. Polyhach, G. Jeschke, H.-J. Steinhoff, E. Schneider, *Biophys. J.*, 2008, **95**, 2924-2938.

35. G. Jeschke, M. Sajid, M. Schulte, N. Ramezanian, A. Volkov, H. Zimmermann, A. Godt, *J. Am. Chem. Soc.*, in press.

36. M. Schlenkrich, J. Brichmann, A. MacKerell, M. Karplus, in *Biological Membranes: A Molecular Prospective from Computation and Experiment*, Birkhauser, Boston, 1996; pp 31-81.

37. D. Sezer, J. Freed, B. Roux, *J. Phys. Chem. B.*, 2008, **112**, 5755-5767.

38. J. Ponder and F. Richards, *J. Comput. Chem.*, 1987, **8**, 1016-1024.

39. D. J. Earl, M. W. Deem, *Phys. Chem. Chem. Phys*., 2005, **7**, 3910-3916.

40. W. Still, A. Tempczyk, R. Hawley, T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127-6129.

41. G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. Timmel, D. Hilger, H. Jung, *Appl. Magn. Reson.*, 2006, **30**, 473-498.

42. G. Jeschke, A. Koch, U. Jonas, A. Godt, *J. Magn. Reson.*, 2002, **155**, 72-82.

43. W. Doster, *Biochim. Biophys. Act*., 2010, **1804**, 3-14.

44. Q. Sun, J. Ghosh, R. Faller, in *Coarse-Graining of Condensed Phase and Bimolecular Systems*, ed. G. Voth, CRC Press, 2009, pp 69-82.

45. Y. Polyhach, G. Jeschke, *Spectroscopy*, 2010, **24**, in press.

46. www.epr.ethz.ch/software/index

47. D. M. Freed, P. S. Horanyi, M. C. Wiener, D. S. Cafiso, *Biophys. J.*, 2010, **99**, 1604 - 1610.

48. C. Hunte, E. Screpanti, M. Venturi, A. Rimon, E. Padan, H. Michel, *Nature*, 2005, **435**, 1197-1202.

49. K. Herz, A. Rimon, G. Jeschke, E. Padan, *J. Biol. Chem.*, 2009, **284**, 6337-6347.

50. T. Barros, W. Kuhlbrandt, *Biochim. Biophys. Act.*, 2009, **1787**, 753-772.

51. A. Volkov, C. Dockter, Y. Polyhach, H. Paulsen, G. Jeschke, *J. Phys. Chem. Lett.*, 2010, **1**, 663-667.

Electronic supplementary information

To accompany

Rotamer libraries of spin labelled cysteines for protein studies

Yevhen Polyhach[§], Enrica Bordignon[§], Gunnar Jeschke[*§]

[§]Laboratory of Physical Chemistry, ETH Zürich, Wolfgang-Pauli-Str. 10, 8093 Zürich, Switzerland

Customization of the CHARMM27 force-field for IA-PROXYL

The bending potential for the angle between S-CT2-C atom types, as well as potential for the torsion specified by CT2-S-CT2-C atom types (χ_3 dihedral for IA-PROXYL) are missing in the standard CHARMM27 force field. In the case of IA-PROXYL, the former is needed to account for bending of the S_γ-C_δ-C_ε bond, whereas the latter specifies torsion about the χ_3 dihedral angle, see Fig. S1.

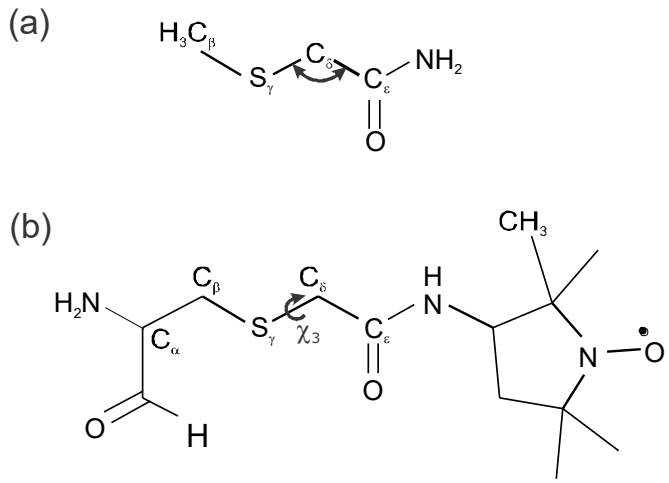


Fig. S1 (**a**) Molecular fragment used for determination of the bending potential for the angle between S-CT2-C atom types. (**b**) IA-PROXYL spin label.

The S-CT2-C bending potential was obtained by fitting vibrational frequencies of the molecular fragment (Fig. S1a) calculated by Tinker to the corresponding frequencies calculated by density functional theory (as implemented in the ORCA package[1]). The potential for the χ_3 torsion corresponding to the CT2-S-CT2-C atom types was not explicitly parameterized but rather replaced with that for CT3-S-CT2-CT2 atoms types. Inspection of other CHARMM27 torsion parameters suggests that similar replacements are commonplace.

Summary for MD simulations and projection of the full length MD trajectory on the set of rotamers for MTSSL and IA-PROXYL

Distributions of dihedral angles χ_1-χ_5 for MTSSL and χ_1-χ_6 for IA-PROXYL were calculated from the corresponding full-length MD trajectories acquired at 175 and 298 K. In the form of histograms or dihedral angle profiles, such distributions are shown as blue lines in Fig. S2 and Fig. S3 for MTTSL and IA-PROXYL respectively. Additionally, canonical dihedral angle values (shown green) as well as dihedral angles of the rotamers forming final rotamer libraries (shown red) are displayed superimposed on the corresponding dihedral angle profiles.

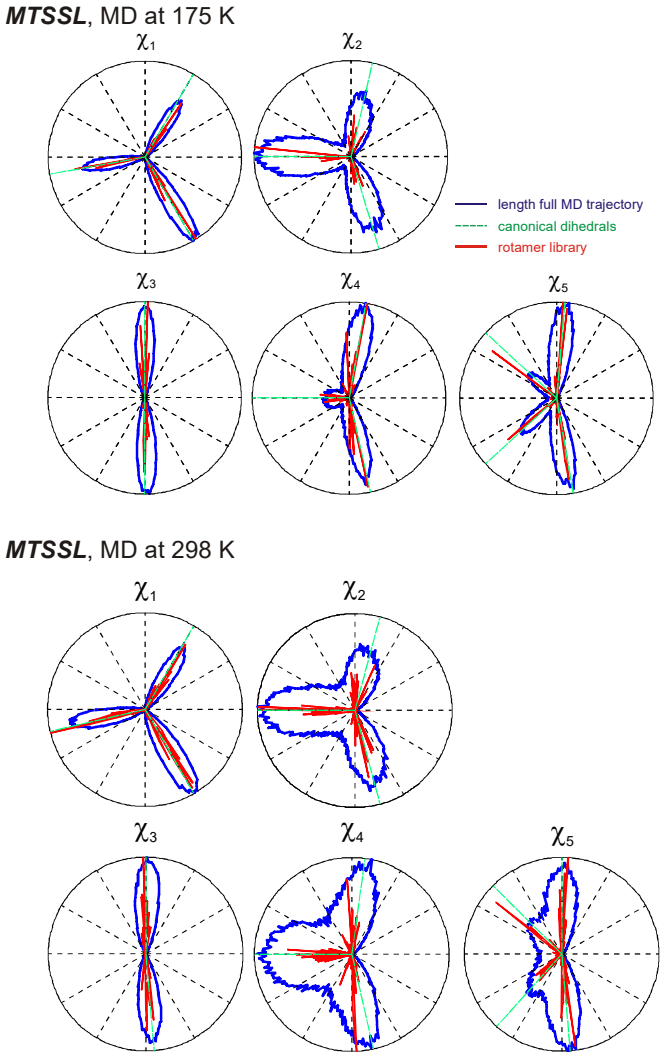


Fig. S2 Distributions of dihedral angles χ_1-χ_5 (*dihedral profiles*) for MTSSL calculated from the full length MD trajectories acquired at 175 and 298 K. Canonical dihedral angles values as well as dihedral angles distributions of the rotamers comprising corresponding final rotamer libraries are shown superimposed.

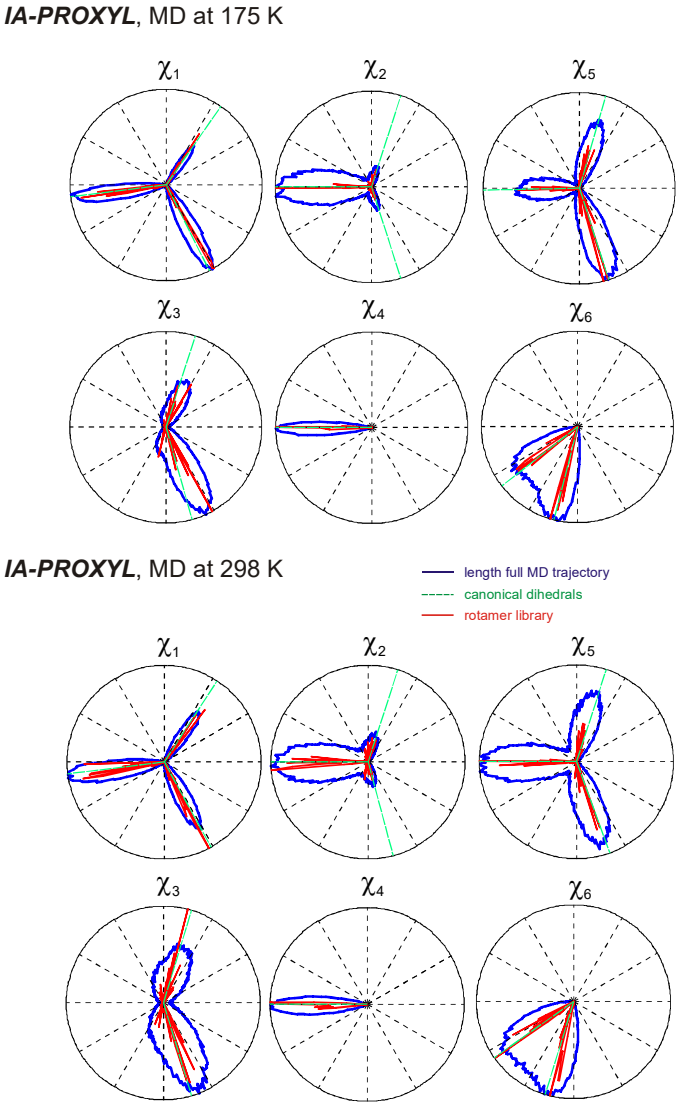


Fig. S3 Distributions of dihedral angles χ_1-χ_6 (dihedral profiles) for IA-PROXYL calculated from the full length MD trajectories acquired at 175 and 298 K. Canonical dihedral angles values as well as dihedral angles distributions of the rotamers comprising corresponding final rotamer libraries are shown superimposed.

References.

1. F. Neese, *Coord. Chem. Rev.,* 2009**, 253**, 526-563.