Published in final edited form as:
J Phys Chem B. 2013 May 2; 117(17): 4733–4739. doi:10.1021/jp3110369.

Restrained-Ensemble Molecular Dynamics Simulations Based on Distance Histograms from Double Electron-Electron Resonance Spectroscopy

Benoît Roux[*] and **Shahidul M. Islam**
Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, University of Chicago, Chicago, IL 60637, Phone: (773) 493-5303

Abstract

DEER (double electron electron resonance) spectroscopy is a powerful pulsed ESR (electron spin resonance) technique allowing the determination of spin-spin distance histograms between site-directed nitroxide label sites on a protein in their native environment. However, incorporating ESR/DEER data in structural refinement is challenging because the information from the large number of distance histograms is complex and highly coupled. Here, a novel restrained-ensemble molecular dynamics simulation method is developed to incorporate the information from multiple ESR/DEER distance histograms simultaneously. Illustrative tests on 3 coupled spin-labels inserted in T4 lysozyme show that the method efficiently imposes the experimental distance distribution in this system. Different rotameric states of the χ_1 and χ_2 dihedrals in the spin-labels are also explored by restrained ensemble simulations. Using this method, it is hoped that experimental restraints from ESR/DEER experiments can be used to refine structural properties of biological systems.

Keywords

molecular dynamics; potential of mean force; T4 lysozyme; spin-label; side-chain rotamer

I. Introduction

DEER (Double Electron Electron Resonance) is a powerful pulsed double-quantum coherence (DQC) electron spin resonance (ESR) spectroscopy that allows the determination of the nitroxide-nitroxide distance histogram between two methanethiosulfonate spin-labels (MTSSL) attached to a protein via cysteine residues.[1–4] ESR/DEER, in conjunction with site-directed spin labeling (SDSL) techniques,[5,6] now allows one to measure the distance histograms between an unprecedented large number of sites, which requires new computational strategies for structural refinement. For example, up to 51 pairs of experimental ESR/DEER distance histograms have been measured for T4 lysozyme, a small benchmark protein of 165 residues (H. Mchrouab, personal communication). Exploiting the data from ESR/DEER in structural refinement, however, is confronted by several challenges. The measured histograms are information-rich and often have complex structures with multiple peaks. Furthermore, the nitroxide spin-labels (Figure 1) display considerable internal flexibility, [7] which implies that multiple rotameric states must be taken into account and that it is unrealistic to consider a single rotameric state. This uncertainty increases the difficulty in translating the measured distance histograms between spin-labels into a structural refinement restraint. Naively imposing a simple average distance between

roux@uchicago.edu.

C_α backbone atoms is definitely not a valid approximation. Of particular importance, a given spin-label can be involved in multiple distance histograms, which implies that the large body of data is highly coupled. For instance, this precludes an approach where the spin-label pairs could be treated one-by-one. Furthermore, the characterization of the accessible rotamers of the spin-label side chain is difficult.[7–13] In summary, the body of data from ESR/DEER distance histograms is complex, information-rich, highly coupled, and voluminous. While there have been previous efforts to utilize the information from ESR/DEER for protein structure determination and progress has been made using rotamer libraries,[2–4] a general strategy is still lacking.

In familiar energy-based structural refinement algorithms, whether it is based on energy-minimization or molecular dynamics (MD) simulations, one typically introduces a biasing potential that depends on the configuration **x** of the system in order to enforce that a given property $q(\mathbf{x})$ will match the experimental value Q. One of the simplest form is a quadratic potential of the type, $k(q(\mathbf{x})-Q)^2$, with a restraint force constant k. This is the approach that is generally adopted in structural refinement based on X-ray crystallography or in nuclear magnetic resonance (NMR).[14] However, by doing so we unavoidably impose also the magnitude of the fluctuations of the property q, indirectly corrupting the atomic configurations in a way that is arbitrary and undesirable. An alternative strategy is based on the notion of "restrained ensemble" MD simulations introduced by Vendruscolo and coworkers, [15–17] and exploited by Im and co-workers to determine the structure of membrane-bound peptides on the basis of solid state NMR observables.[18–21] The restrained-ensemble MD simulation scheme consists in carrying out parallel MD simulations of N replicas of the basic system in the presence of a biasing potential that enforces the ensemble-average of a given property toward its known experimental value. Working from the perspective of an ensemble is appealing since it attempts to mimic the conditions under which bulk measurements are carried out experimentally. Intuitively, it is hoped that if the ensemble comprises a fairly large number of replicas, the biasing potential enforcing the ensemble-average property constitutes only a mild perturbation acting on any individual replica without causing large unwanted distortions. The idea has been illustrated by showing that solid state NMR properties within the restrained-ensemble MD rapidly converge toward unique statistical distributions as N becomes large.[18–20] Recently, it was demonstrated that the restrained-ensemble MD simulation scheme is formally related to the maximum entropy method for biasing thermodynamic ensembles.[22–24] This helped clarify the general significance of the restrained-ensemble MD simulations, and the underlying conditions for its validity.

Our goal with this article is to exploit these recent developments to formulate a computational approach to utilize the information from multiple ESR/DEER distance histograms in protein structure refinement. Here, an approach based on a mean-field version of the restrained-ensemble MD is elaborated and illustrated to the case of three nitroxide spin-labels attached to the small soluble protein T4 lysozyme. The article is concluded by outlining future applications of this approach in structural refinement.

II. Theoretical Developments

An issue of general interest in computer simulations is to incorporate information from experiments into a structural model. An important caveat in pursuing this goal is to avoid corrupting the resulting model with spurious and arbitrary biases. Let us consider a system that is described by the potential energy $U(\mathbf{X})$, where $\mathbf{X} \equiv \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ represents all atomic coordinates. The equilibrium Boltzmann distribution is,

$$P_0(\mathbf{X}) = \frac{e^{-\beta U(\mathbf{X})}}{\int d\mathbf{X}\, e^{-\beta U(\mathbf{X})}} \quad (1)$$

where $\beta = 1/k_B T$. Without any additional biases, this model based on potential energy $U(\mathbf{X})$ represents all our prior knowledge about the system. Unfortunately, this model is imperfect and, for example, the average over some property $q(\mathbf{X})$,

$$\langle q \rangle = \int d\mathbf{X} q(\mathbf{X})\, P_0(\mathbf{X}) \quad (2)$$

differs from the value Q known from experiment. In this regard, it is natural to attempt to incorporate the experimental information into the model. In practice, this problem can be formulated within the maximum entropy method,[22] whereby one seeks to maximize an excess cross-entropy functional under the constraint that the resulting average must reproduce the experimental data.[23,24] This yields a solution of the general form

$$P(\mathbf{X}) = \frac{e^{-\beta[U(\mathbf{X}) - k_B T \lambda q(\mathbf{X})]}}{\int d\mathbf{X}\, e^{-\beta[U(\mathbf{X}) - k_B T \lambda q(\mathbf{X})]}} \quad (3)$$

where λ is a Lagrange multiplier that must be adjusted to satisfy Eq. (2). It is as if the system is evolving on an effective potential energy surface, $U_{\text{eff}}(\mathbf{X}) = U(\mathbf{X}) - k_B T \lambda_q(\mathbf{X})$.

An interesting special case occurs when the set of properties corresponds to a continuous distribution along some variable ξ that is a function of $\mathbf{X}$. This is the type of data that is provided by ESR/DEER spectroscopy.[1] For example, ξ could be the distance between two atoms or the position of an atom along some axis z. Dividing the values of ξ into a set of L discrete bins $\{\xi^{(i)}\}$ of width $\Delta\xi$, the probability in each bin i can be defined as the average of the population operator,

$$h_i(\xi) = \begin{cases} 1 & \text{if } (i-1)\Delta\xi - \frac{1}{2}\Delta\xi < \xi < (i-1)\Delta\xi + \frac{1}{2}\Delta\xi \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

By construction, the set of function h_i is complete and satisfy the normalization condition, $\Sigma_i h_i = 1$. In the limit that $\Delta\xi$ is very small, the quantity $h_i(\xi)/\Delta\xi$ turns into a delta function. We seek to modify the equilibrium distribution $P_0(\mathbf{X})$ such that the averages $\langle h_i \rangle$ will match the set of experimental values H_i, for $i = 1, \ldots, L$ (also normalized, $\Sigma_i H_i = 1$). This problem can be easily formulated using the maximum cross-entropy method, which yields,

$$P(\mathbf{X}) = \frac{e^{-\beta[U(\mathbf{X}) - k_B T \sum_i \lambda_i h_i(\xi(\mathbf{X}))]}}{\int d\mathbf{X}\, e^{-\beta[U(\mathbf{X}) - k_B T \sum_i \lambda_i h_i(\xi(\mathbf{X}))]}} \quad (5)$$

where λ_i are Lagrange multipliers to be adjusted in order to match the set H_i. The system is evolving on the effective potential energy surface, $U_{\text{eff}} = U - k_B T \Sigma_i \lambda_i h_i$. In the limit where $\Delta\xi$ become very small, the sum of $\lambda_i h_i(\xi)$ support a continuous biasing potential U_{bias} as a function of the coordinate ξ,

$$U_{\text{bias}}(\xi)=-\ln\left[\frac{H(\xi)}{h(\xi)}\right] \quad (6)$$

where $H(\xi)$ is the desired experimental distribution, and $h(\xi)$ is the distribution from the unbiased simulation. $U_{\text{bias}}(\xi)$ is related to the potential of mean force (PMF), $W(\xi)$, of the system along the coordinate ξ, with $W(\xi) - W(\xi^*) = -k_B T \ln[h(\xi)/h(\xi^*)]$, where ξ^* is an arbitrary reference point. This is true because the functions $h_i(\xi)$ form a complete and non-overlapping (orthogonal) basis set along the axis ξ. This shows that the most straightforward route to bias a system toward a single experimental distribution function consists in determining the PMF as a function of ξ, and then biasing the system using the potential U_{bias} defined by Eq. (6). This form is valid for any specific variable that does not involve indirect or hidden correlations, such as any internal degrees of freedom internal in a given residue, or a distance between residue i and j in a protein. This form does not, however, apply to cases where multiple degrees of freedom are superimposed to construct a single average distribution, such as the particle-particle PMF in a liquid for example.

An important extension of the above discussion arises when trying to use the information from multiple experimental ESR/DEER distance distributions to bias the system. Direct application of the maximum entropy principle becomes essentially impractical because it requires the determination of a large number of unknown Lagrange multipliers by solving iteratively until convergence is reached. An alternative strategy is the restrained-ensemble MD simulation scheme.[15–21] It consists in carrying out parallel MD simulations of N replicas of the basic system in the presence of a biasing potential that restrains the ensemble-averaged property toward the experimental value Q. The restrained-ensemble MD simulation is formally equivalent to the maximum entropy principle for biasing an average property toward an experimental value.[22–24] In the restrained-ensemble MD simulations, we consider an ensemble of N parallel systems s, each with the coordinates $\mathbf{X}_s$ and the potential energy $U(\mathbf{X}_s)$, and the ensemble-average of system

$$\bar{q}(\mathbf{X}_1,\ldots,\mathbf{X}_N)=\frac{1}{N}\sum_{s=1}^{N} q(\mathbf{X}_s) \quad (7)$$

is restrained to match the set of experimental distributions by the potential U_{RE}

$$U_{\text{RE}}(\mathbf{X}_1,\ldots,\mathbf{X}_N)=\frac{1}{2}K(\bar{q}(\mathbf{X}_1,\ldots,\mathbf{X}_N)-Q)^2 \quad (8)$$

where K is a large force constant. In the case of ESR/DEER, the experimental data is a set of distance histogram bins, $H^{ij}(r)$, between ij pairs of spin-labels, where r is the distance between the oxygen atoms of the nitroxide spin-labels, which carries the unpaired electrons. However, a rigorous application of the restrained-ensemble would require extracting a single count for each copy of the system, that is, the distance $|\mathbf{r}_i^s-\mathbf{r}_j^s|$ between spin-labels i and j for the s-th copy. A fair amount of useful information is wasted with this strict application of the restrained-ensemble. Since ESR/DEER experiments are typically carried out for spin-labels that are well separated, by a distance on the order of 10–50 Å, it can be safely assumed that the spin-labels are uncorrelated from one another. Considering the joint probability of the s-th copy of two nitroxides i and j, $\mathscr{P}(\mathbf{r}_i^s,\mathbf{r}_j^s)$, the ensemble average distance histogram of the ij pair is,

$$\bar{h}^{ij}(r)=\frac{1}{N}\sum_{s=1}^{N}\int d\mathbf{r}_i^s d\mathbf{r}_j^s\,\delta\left(|\mathbf{r}_i^s-\mathbf{r}_j^s|-r\right)\,\mathscr{P}(\mathbf{r}_i^s,\mathbf{r}_j^s), \quad (9)$$

which is replaced by the mean-field approximation,

$$\bar{h}^{ij}(r)\approx\frac{1}{N^2}\sum_{s=1}^{N}\sum_{s'=1}^{N}\int d\mathbf{r}_i^s d\mathbf{r}_j^{s'}\,\delta\left(|\mathbf{r}_i^s-\mathbf{r}_j^s|-r\right)\,\mathscr{P}(\mathbf{r}_i^s)\,\mathscr{P}(\mathbf{r}_j^{s'}), \quad (10)$$

The underlying assumption supporting the ansatz $\mathscr{P}(\mathbf{r}_i^s,\mathbf{r}_j^s)\approx\mathscr{P}(\mathbf{r}_i^s)\,\mathscr{P}(\mathbf{r}_j^{s'})$ is that the conformational distribution of the s-th copy of the spin-label i is not affected by the conformational distribution of the s'-th copy of the spin-label j, which is reasonable if the two spin-labels are not in direct contact. Accordingly, the N copies of the two spin-labels yield $N\times N$ separate entries to construct the distance histogram of the ij pair. In other words, it is assumed that the distribution of conformational rotamers of spin-label i is not affected by the conformational rotamers of spin-label j. This is a reasonable approximation as long as the spin-labels i and j are well separated from one another and do not directly collide. Under this condition, the population of accessible rotameric states for spin-labels i and j are statistically independent. Such a treatment of the multiple copies corresponds to a mean-field approximation.[25]

In principle, the discrete counting function based on Eq. (4) could be used to construct the histograms in Eq. (10), although this leads to cumbersome discontinuities. To avoid these problems, a smooth differentiable Gaussian of width σ is used in the construction of the histograms, discretizing the distance r via $r=n\Delta r$, where Δr is the bin width. It follows that the n-th bin of the histogram for the ij pair is calculated as,

$$\bar{h}^{ij}(n)=\frac{1}{N^2}\sum_{s=1}^{N}\sum_{s'=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(n\Delta r-|\mathbf{r}_i^s-\mathbf{r}_j^{s'}|)^2}{2\sigma^2}} \quad (11)$$

where the set of copies for the first spin-label is given by $s\in\{i\}$ and the set of copies for the second spin-label is given by $s'\in\{j\}$. All possible pairs of distances between the copies of two spin-labels are collected into a single histogram, according to a mean-field approximation. The histograms, both from experiment and the restrained-ensemble simulations, are properly normalized, namely, $\Sigma_{\{\text{bin }n\}}\,\bar{h}^{ij}(n)\,\Delta r=\Sigma_{\{\text{bin }n\}}\,H^{ij}(n)\Delta r=1$, for all pairs ij (the Jacobian is implicitly incorporated into the integrand and the histograms have dimension of inverse length). This form has the advantage of allowing continuous first derivatives for the restraining potential. Furthermore, it is also consistent with the notion that the unpaired electron is spatially delocalized over the oxygen of the nitroxide spin-label. The restraining energy is given by,

$$U_{\text{RE}}=\frac{1}{2}K\sum_{\{\text{pair }ij\}}\sum_{\{\text{bin }n\}}\left(\bar{h}^{ij}(n)-H^{ij}(n)\right)^2 \quad (12)$$

where the force constant K has dimension of energy per length square. The first derivative of the restrained-ensemble energy with respect to the position s-th copy of the i-th spin-label, $\mathbf{r}_i^s$, is,

$$\frac{\partial U_{\mathrm{RE}}}{\partial \mathbf{r}_i^s} = K \sum_{j \in \{\text{pair } ij\}} \sum_{s'=1}^{N} \sum_{\{\text{bin } n\}} \left(\overline{h}^{ij}(n) - H^{ij}(n) \right) \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n\Delta r - |\mathbf{r}_i^s - \mathbf{r}_j^{s'}|)^2}{2\sigma^2}} \right) \times \left(\frac{n\Delta r - |\mathbf{r}_i^s - \mathbf{r}_j^{s'}|}{2\sigma^2} \right) \frac{(\mathbf{r}_i^s - \mathbf{r}_j^{s'})}{|\mathbf{r}_i^s - \mathbf{r}_j^{s'}|} \quad (13)$$

All the computational details for testing the restrained-ensemble of ESR/DEER histograms are given in the following section.

IV. Computational Details

All the simulations were carried out using the program CHARMM.[26] The atomic structure of T4 lysozyme (T4L) determined by X-ray crystallography[27] was taken from the Protein Data Bank (PDB id 2LZM). T4L is a small protein of 165 residues that has often served as a model system to benchmark electron spin resonance (ESR) spectroscopy studies in conjunction with site-directed spin-labeling (SDSL) techniques. [5,6] Three residues, Glu62, Thr109, and Ala134 were mutated to cysteine and nitroxide spin-labels were attached at these positions. A novel energy restraint based on Eq. (12) was implemented in the MMFP module of CHARMM to couple multiple copies and enforce the experimental distributions from ESR/DEER. The analytical first derivative of the new energy restraint was validated by comparison with the finite-difference derivative using the TEST FIRST command. The restrained-ensemble force constant K was set to 10000 (kcal/mol)/Å^2, and the natural spread σ of the Gaussian was set to 1.7 Å. This value is justified because the width of the ESR/DEER histograms is on the order of about 10 Å. The bin width of the histogram was set to 1.0 Å and the histograms run from 1 to 60 Å. Each of the spin-labels was expanded into N copies (N = 1, 10, and 25) using the multiple-copy Locally Enhanced Sampling (LES) method,[25,28,29] while the remainder of the system was simulated with a single copy. The simulations with LES were setup using the REPLICA and BLOCK commands of the program CHARMM.[26] In the LES simulations, the interactions of the replicated spin-labels with the rest of the system were scaled by $1/N$ while the internal interactions of the replicated spin-labels were unscaled. The different replicas of a given spin-label do not interact with one another. The CHARMM force field PARAM27[30] with the CMAP correction for the backbone dihedral[31] was used for the protein. The force field for the spin-labels was taken from the previous work of Sezer et al.[32] A weak harmonic restraint with a force constant of 1 (kcal/mol)/Å^2 was applied to the backbone of the protein to avoid any large displacement away from the X-ray structure. The Adopted Basis Newton-Raphson minimization was performed for 2000 steps to eliminate initial bad contacts between various atoms. The restrained-ensemble system was simulated in vacuum at a temperature of 300K using Langevin dynamics for 20 ns. The friction was set to 1.0 for all non-hydrogen atoms and an integration time step of 1 fs was used. Bonds involving hydrogen atoms were constrained to their equilibrium values using the SHAKE algorithm.[33] For comparison, a conventional (single-copy) MD was also carried out for 20 ns without any restraint. It should be noted that this simulation differs from the restrained-ensemble simulation carried out with a single replica (N = 1) because it does not include any restraint with respect to the ESR/DEER distance histogram data. The nonbonded interactions were smoothly switched off from 10–12 Å using an atom-based cutoff.

IV. Results and Discussion

It was previously demonstrated that the restrained-ensemble MD simulations converges rigorously toward the result from maximum entropy principle in the limit of strong biasing potential and large number of replicas.[24] This is reassuring because it suggests that theoretical constructs built on the restrained-ensemble approach do not corrupt the resulting distribution functions with spurious and unwanted biases, in a maximum entropy sense. As

is clear from the above analysis, applications of the maximum entropy principle requires the determination of several coefficients λ_i, which must be adjusted to match the known experimental data. While the formal task is clear, this optimization process can rapidly become cumbersome and unwieldy, particularly when there is a very large number of data.

The great advantage of an approach based on restrained-ensemble simulations is to obtain a solution that is formally equivalent to that from maximum entropy principle while avoiding the obligation to determine a large number of unknown Lagrange multipliers. While the computational load is certainly larger in the restrained-ensemble simulations because of the multiple copies, the simulation automatically evolves in a way that is consistent with the maximum entropy principle without any additional intervention by the user. This becomes particularly important in the case of ESR/DEER distance histogram data because there can be a huge body of data to take into account. To make the algorithm more efficient, it was assumed that the different spin-labels were uncorrelated from one another. Such a mean-field approximation is reasonable because the spin-labels in ESR/DEER experiments are typically separated by a sufficiently large distance. To illustrate the method, three histograms are used to restrain the motion of three spin-labels introduced at position 62, 109, and 132 in T4 lysozyme (T4L) using site-directed cysteine mutations. The nitroxide spin-label is shown in Figure 1. T4L is a small soluble protein of 165 residues that has long served as a benchmark system for ESR and SDSL techniques. [5,6]

The T4L system is shown in Figure 2. In the left view is shown the protein with single-copy spin-labels. In the right view is shown a typical snapshot taken from the restrained-ensemble simulations. In the case considered here, the 3 spin-labels are well separated, by distances on the order of 20–40 Å. It is worth noting that, while the samples used in ESR/DEER are fast-frozen from room temperature to liquid nitrogen temperature, it makes sense to consider the system at temperature T because the ensemble of spin-label rotamers is rapidly quenched and the distance histograms reflect the distribution at room temperature. Thus, T4L frozen in this manner faithfully maintain the structural properties it possesses in solution at room temperature. Nevertheless, this issue may require further attention in the future, as there are experimental indications that the freezing process may affect the configurational distribution of the spin-labels.[34]

The internal flexibility of the spin-label is of great importance in the interpretation of ESR data in general. Five dihedral angles, $\chi_1 - \chi_5$, connect the nitroxide ring of the spin-label to the protein backbone (Figure 1). The dihedral angle of the disulphide (S-S) bond, χ_3, can adopt one of two configuration, 90° and −90°, which are separated by a large energy barrier. A number of X-ray crystallographic and spectroscopic studies aimed at characterizing the rotameric state of the spin label[7–13] have suggested that the internal dynamics of the spin-label side chain arise predominantly from the dihedral angles χ_4 and χ_5. Variable rotameric states were observed for χ_1 and χ_2 at different sites in T4L, which led to the hypothesis that these dihedrals are hindered by the formation of a hydrogen bond between the sulfur atom of the spin-label and the backbone amide and Cα. [7,13]

Restrained-ensemble MD simulations using 1, 10 and 25 copies for each spin-labels were carried out for 20 ns. The simulation with a single copy corresponds to the standard approach to impose experimental properties. The results from the mean-field restrained-ensemble MD simulations are compared with experimental ESR/DEER data in Figure 3. The restrained simulation converged rapidly toward the experimental histograms. In contrast, the conventional (single copy) unbiased simulations did not accurately reproduce the experimental histograms (this simulation differs from the restrained-ensemble simulation carried out with a single replica with $N = 1$ because it does not include any restraint with respect to the ESR/DEER histograms).

To characterize the effect of the number of copies of the spin-labels used in the restrained-ensemble MD simulation, we calculated the rotameric states of the spin-label at the 3 sites in T4L. The distribution of the χ_1 and χ_2 dihedrals is displayed as scatter plots in Figure 4 for spin-labels at position 62, 109 and 134 in T4L. For all 3 sites in T4L, the number rotameric states visited by the simulations is the smallest with a single replicas and becomes roughly equivalent with 10 or 25 replicas, indicating that the multiple-copy representation has converged for the present system.

Restraint-ensemble simulations performed with 10 and 25 copies of each spin-label are found to explore more rotameric states along χ_1 and χ_2 dihedrals when compared to those obtained from conventional single-copy restrained MD simulation. This suggests that the dihedrals are still flexible, even though distance distributions are restrained to experimental distributions in restraint-ensemble simulation. Usually the rotameric states in the spin-label side chains are denoted by *p*, *t*, and *m*, depending on the range of dihedral angles from 0° to 120°, 120° to −120° and −120° to 0°, respectively. spin-label 62 prefer mostly the *mm* (78%) and *mp* (18%) rotamers. In contrast, spin-labels 109 and 134 visit multiple rotameric states, with *mt* (31%), *pp* (21%), *pm* (13%), *mp* (13%), *mm* (12%) and *tm* (6%) rotamers for spin-label 109, and *mt* (42%), *mm* (25%), *pt* (14%) and *tt* (10%) for spin-label 134. The restrained-ensemble MD simulation method enforces the experimental distance histogram while allowing reasonable fluctuations of the dihedral angles of the spin-labels.

The present treatment implicitly considered the real-space pair-distance ESR/DEER histograms as the “raw input” information from experiments. In reality, such distance histograms are obtained by post-processing the measured time-dependent ESR/DEER signal.[35] A legitimate question is whether one should construct the restrained-ensemble simulation methodology on the basis of the actual time-dependent EPR/DEER experimental signal. The observed EPR/DEER signal, which depends on multiple factors, can be expressed as $V^{ij}(t) = B(t)[1 - \lambda F^{ij}(t)]$.[35] Of primary interest is the function $F^{ij}(t)$, which corresponds to the magnetic interaction between spins *i* and *j*. In addition, the constant λ is related to the depth of modulation of the signal and depends on the fraction of doubly labeled protein in the sample. Lastly, there is the background decaying function $B(t)$, which depends on intermolecular spin-spin coupling with distant neighbors. $B(t)$ is sensitive to the dimensionality of the spatial distribution of labeled proteins and is generally unknown. There is a linear relationship between the time-dependent signal, $F^{ij}(t)$, and the distance distribution $H^{ij}(r)$, which can be expressed in discrete form as,

$$F^{ij}(t_n) = \sum_{r_m} M_{\{t_n, r_m\}} H^{ij}(r_m) \quad (14)$$

where the operator $M_{\{t_n, r_m\}}$ includes, among other things, the effect of averaging the spin-spin vector $(\mathbf{r}^i - \mathbf{r}_j)$ over all possible orientations. Determining the real-space histograms $H^{ij}(r)$ from the time-dependent signal $F^{ij}(t)$ amounts to a difficult inverse problem, which is typically is solved by using Tikhonov regularization.[35] Clearly, converting $V^{ij}(t)$ into $H^{ij}(r)$ requires careful considerations of multiple factors pertaining to the sample preparation and experimental conditions. The validity of a structural model can be ascertained by examining if the time-dependent signal reproduced experimental ESR/DEER data. For example, Boura et al[36] generated a large number of possible conformations from unbiased simulations, and then constructed a weighted ensemble by scoring those conformations in post-analysis according to their ability to reproduce the time-dependent ESR/DEER signal. In contrast, the goal of the present multiple-copy restrained-ensemble simulation method is to actively drive the protein conformation toward a 3D structure that is consistent with the data. While a post-analysis scoring method can work if the correct structure is found among the trial set of

conformation, to have a practical structural refinement algorithm, it is necessary to have the ability to bias the dynamical evolution of the system toward conformations that are consistent with the input data. To go from ESR/DEER measurements to the 3D protein requires two very different and difficult operations: (1) processing of the time signal, and (2) refining the 3D protein structure from correlated data. While it is conceivable that one could use the signal $V^{ij}(t)$, or the post-processed function $F^{ij}(t)$, directly into some kind of restrained-ensemble MD simulation for the purpose of structural refinement, we see no clear advantage in combining two difficult operations into a single step. In effect, the distance histograms offer a nice intermediate stage allowing one to get a sense of the spatial information provided by the data, and see if the post-processing treatment of the signal has been carried out properly.

V. Summary

A multiple-copy mean-field restrained-ensemble method was implemented to incorporate the information from multiple ESR/DEER distance histograms simultaneously. Illustrative tests on 3 coupled spin-labels in T4L show that the method efficiently imposes the experimental distance distribution rapidly in this system. Future directions of the present methodology will be to fully explore the consequences of incorporating all the ESR/DEER data on the rotamer distribution of the spin-labels in T4L. While the present illustrative simulations were carried out in vacuum, it will certainly be critical to consider the effect of explicit solvent in future studies. It might also be of interest to try to reduce the large sulphur-sulphur dihedral barrier to increase the rotameric interconversion rates of the nitroxide spin-label using enhanced sampling techniques such as accelerated MD (aMD).[37] The restrained-ensemble MD simulation methodology described here could be extended to other experimental properties and observables, such as X-ray and neutron scattering profiles, NMR, resonance energy transfer to give a few examples. Future applications will be aimed at determining the population of nitroxide spin-label rotamers using ESR/DEER experiments.

Acknowledgments

The work was carried out in the context of the *Membrane Protein Structural Dynamics Consortium* which is funded by grant U54-GM087519 from the National Institute of Health (NIH). The experimental ESR/DEER data on the 3 sites in T4L were generously provided by Richard Stein and Hassane McHaourab.

References

1. Borbat PP, Mchaourab HS, Freed JH. Protein Structure Determination Using Long-Distance Constraints from Double-Quantum Coherence ESR: Study of T4 Lysozyme. J Am Chem Soc. 2002; 124:5304. [PubMed: 11996571]
2. Bhatnagar J, Freed JH, Crane BR. Rigid Body Refinement of Protein Complexes with Long-Range Distance Restraints from Pulsed Dipolar ESR. Two-Component Signaling Systems, Pt B. 2007; 423:117.
3. Kazmier K, Alexander NS, Meiler J, Mchaourab HS. Algorithm for Selection of Optimized EPR Distance Restraints for De Novo Protein Structure Determination. J Struct Biol. 2011; 173:549. [PubMed: 21074624]
4. Hirst SJ, Alexander N, Mchaourab HS, Meiler J. Rosettaepr: An Integrated Tool for Protein Structure Determination from Sparse EPR Data. J Struc Biol. 2011; 173:506.
5. Hubbell WL, McHaourab HS, Altenbach C, Lietzow MA. Watching Proteins Move Using Site-Directed Spin Labeling. Struct. 1996; 4:779.
6. Hubbell WL, Gross A, Langen R, Lietzow MA. Recent Advances in Site- Directed Spin Labeling of Proteins. Curr Opin Struct Biol. 1998; 8:649. [PubMed: 9818271]

7. Sezer D, Freed JH, Roux B. Multifrequency Electron Spin Resonance Spectra of a Spin-Labeled Protein Calculated from Molecular Dynamics Simulations. J Am Chem Soc. 2009; 131:2597. [PubMed: 19191603]
8. Mchaourab HS, Steed PR, Kazmier K. Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR. Spectroscopy Struct. 2011; 19:1549.
9. Mchaourab HS, Kalai T, Hideg K, Hubbell WL. Motion of Spin-Labeled Side Chains in T4 Lysozyme: Effect of Side Chain Structure. Biochem. 1999; 38:2947. [PubMed: 10074347]
10. Mchaourab HS, Lietzow MA, Hideg K, Hubbell WL. Motion of Spin-Labeled Side Chains in T4 Lysozyme, Correlation with Protein Structure and Dynamics. Biochem. 1996; 35:7692. [PubMed: 8672470]
11. Guo ZF, Cascio D, Hideg K, Hubbell WL. Structural Determinants of Nitroxide Motion in Spin-Labeled Proteins: Solvent-Exposed Sites in Helix B of T4 Lysozyme. Prot Sci. 2008; 17:228.
12. Guo ZF, Cascio D, Hideg K, Kalai T, Hubbell WL. Structural Determinants of Nitroxide Motion in Spin-Labeled Proteins: Tertiary Contact and Solvent-Inaccessible Sites in Helix G of T4 Lysozyme. Prot Sci. 2007; 16:1069.
13. Fleissner MR, Bridges MD, Brooks EK, Cascio D, Kalai T, Hideg K, Hubbell WL. Structure and Dynamics of a Conformationally Constrained Nitroxide Side Chain and Applications in EPR Spectroscopy. Proc Natl Acad Sci USA. 2011; 108:16241. [PubMed: 21911399]
14. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & Nmr System: A New Software Suite for Macromolecular Structure Determination. Acta Crystallogr D Biol Crystallogr. 1998; 54:905. [PubMed: 9757107]
15. Lindorff-Larsen K, Best RB, Vendruscolo M. Interpreting Dynamically-Averaged Scalar Couplings in Proteins. J Biomol NMR. 2005; 32:273. [PubMed: 16211481]
16. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous Determination of Protein Structure and Dynamics. Nature. 2005; 433:128. [PubMed: 15650731]
17. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between Native Ensembles and Experimental Structures of Proteins. Proc Natl Acad Sci USA. 2006; 103:10901. [PubMed: 16829580]
18. Lee J, Chen JH, Brooks CL, Im WP. Application of Solid-State Nmr Restraint Potentials in Membrane Protein Modeling. J Mag Res. 2008; 193:68.
19. Jo S, Im W. Transmembrane Helix Orientation and Dynamics: Insights from Ensemble Dynamics with Solid-State Nmr Observables. Biophys J. 2011; 100:2913. [PubMed: 21689524]
20. Kim T, Jo S, Im W. Solid-State Nmr Ensemble Dynamics as aMediator between Experiment and Simulation. Biophys J. 2011; 100:2922. [PubMed: 21689525]
21. Im W, Jo S, Kim T. An Ensemble Dynamics Approach to Decipher Solid-State Nmr Observables of Membrane Proteins. Biochim Biophys Act. 2012; 1818:252.
22. Jaynes ET. Information Theory and Statistical Mechanics. Phys Rev. 1957; 106:620.
23. Pitera JW, Chodera JD. On the Use of Experimental Observations to Bias Simulated Ensembles. J Chem Theory Comput. 2012; 8:3445.
24. Roux B, Weare J. On the Statistical Equivalence of Restrained-Ensemble Simulations with the Maximum Entropy Method. J Chem Phys. 2013; 138:084107. [PubMed: 23464140]
25. Elber R, Karplus M. Enhanced Sampling in Molecular-Dynamics - Use of the Time- Dependent Hartree Approximation for a Simulation of Carbon-Monoxide Diffusion through Myoglobin. J Am Chem Soc. 1990; 112:9161.
26. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. Charmm: The Biomolecular Simulation Program. J Comput Chem. 2009; 30:1545. [PubMed: 19444816]
27. Weaver LH, Matthews BW. Structure of Bacteriophage-T4 Lysozyme Refined at 1.7 a Resolution. J Mol Biol. 1987; 193:189. [PubMed: 3586019]

28. Roitberg A, Elber R. Modeling Side-Chains in Peptides and Proteins - Application of the Locally Enhanced Sampling and the Simulated Annealing Methods to Find Minimum Energy Conformations. J Chem Phys. 1991; 95:9277.

29. Simmerling C, Fox T, Kollman PA. Use of Locally Enhanced Sampling in Free Energy Calculations: Testing and Application to the Alpha -¿Beta Anomerization of Glucose. J Am Chem Soc. 1998; 120:5771.

30. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All- Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102:3586.

31. Mackerell AD, Feig M, Brooks CL. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. J Comput Chem. 2004; 25:1400. [PubMed: 15185334]

32. 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 Alpha-Helix. J Phys Chem B. 2008; 112:5755. [PubMed: 18412413]

33. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comput Phys. 1977; 23:327.

34. Georgieva ER, Roy AS, Grigoryants VM, Borbat PP, Earle KA, Scholes CP, Freed JH. Effect of Freezing Conditions on Distances and Their Distributions Derived from Double Electron Electron Resonance (Deer): A Study of Doubly-Spin-Labeled T4 Lysozyme. J Magn Res. 2012; 216:69.

35. Jeschke G. Deer Distance Measurements on Proteins. Ann Rev Phys Chem. 2012; 63:419. [PubMed: 22404592]

36. Boura E, Rozycki B, Herrick DZ, Chung HS, Vecer J, Eaton WA, Cafiso DS, Hummer G, Hurley JH. Solution Structure of the Escrt-I Complex by Small-Angle X-Ray Scattering, EPR, and Fret Spectroscopy. Proc Natl Acad Sci USA. 2011; 108:9437. [PubMed: 21596998]

37. Hamelberg D, Mongan J, McCammon JA. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. J Chem Phys. 2004; 120:11919. [PubMed: 15268227]

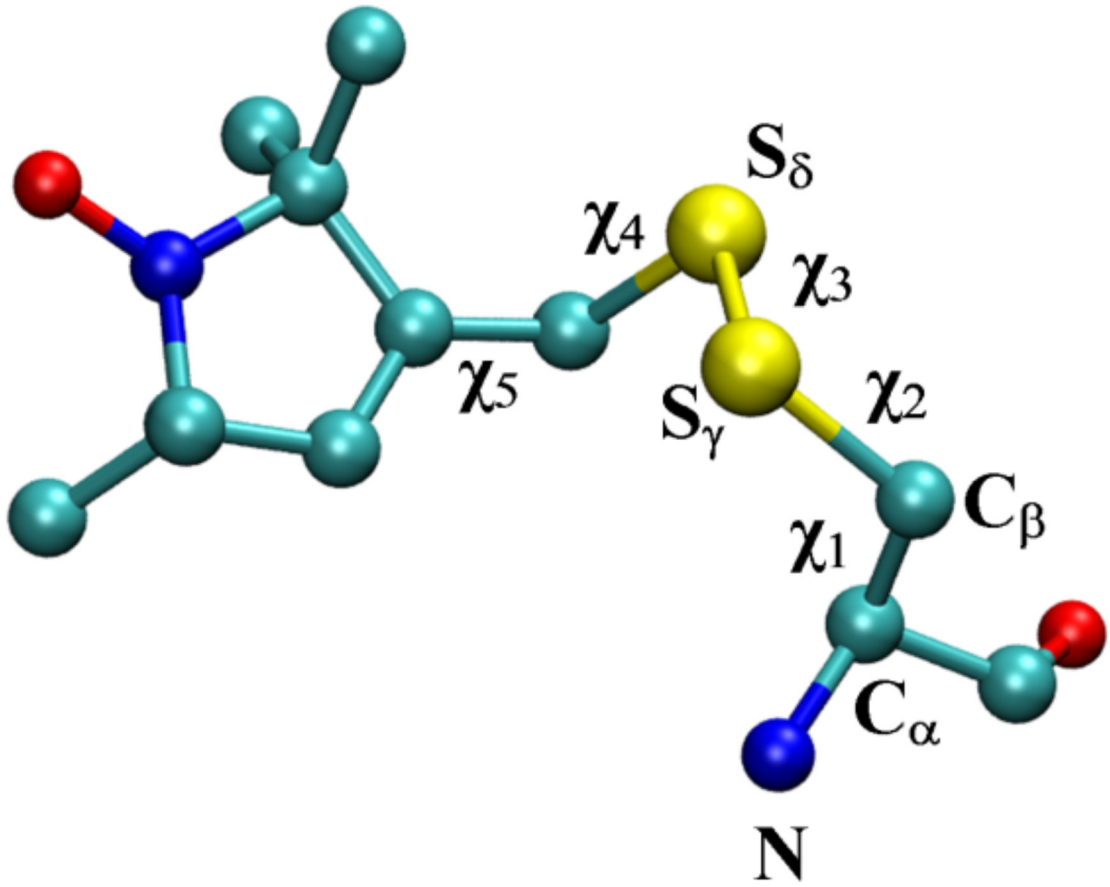

Figure 1.
Representation of the spin-labeled side chain MTSSL (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl-methanethiosulfonate) to a cysteine through a disulfide bond. The five dihedral angles are denoted by χ_1 to χ_5.

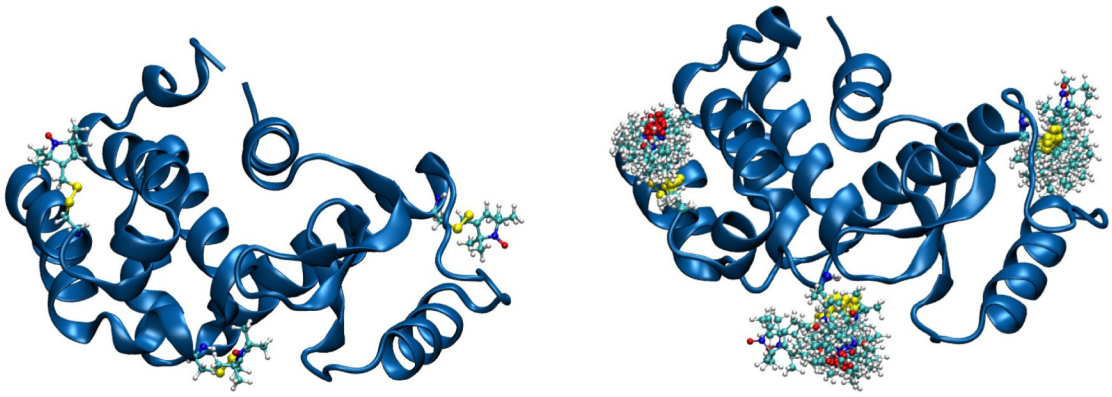

Figure 2.
T4 lysozyme with nitroxide spin-labels at sites 62, 109 and 132. Left is shown the system with single copies, and right is shown the system with 25 replicas of the spin-label at each site.

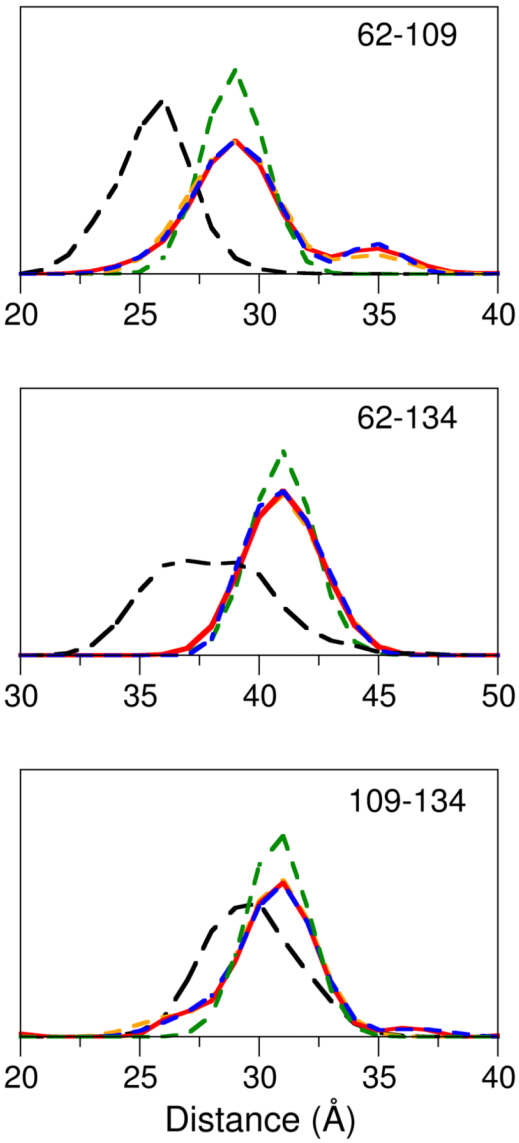

Figure 3.
Distance histograms for 3 correlated sites (62–109, 62–134, and 109–134) from ESR/DEER experiments (red line), unrestrained MD (black dashed line) and the restrained-ensemble simulations performed with 1 (green dashed line), 10 (yellow dashed line) and 25 (blue dashed line) copies of the spin-label side chain.

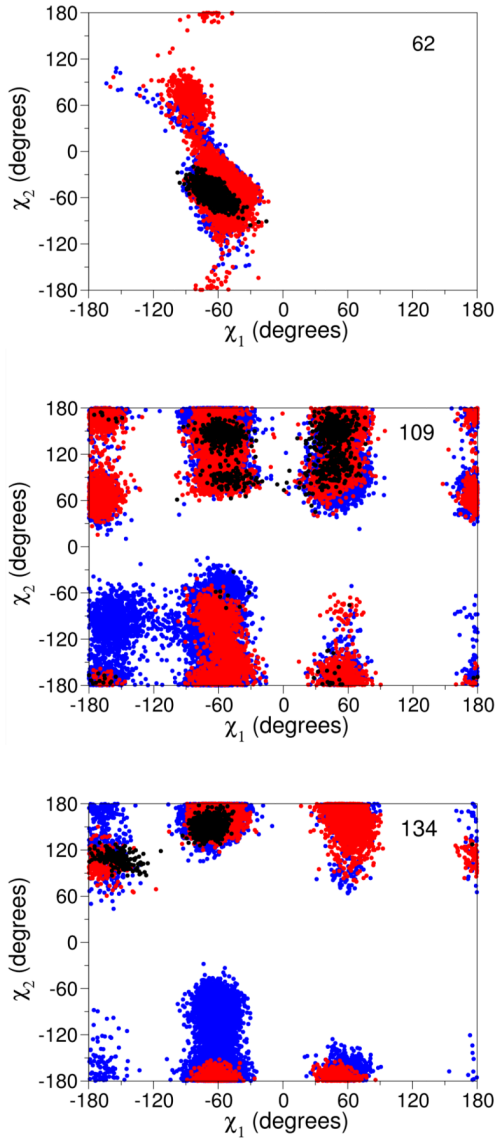

Figure 4.

Rotamer distribution along χ_1 and χ_2 at site 62 (upper panel), 109 (middle panel) and 134 (lower panel) in T4L obtained from the retrained-ensemble MD simulations performed with 1 (black dots), 10 (red dots) and 25 (blue dots) copies of each spin-labels side chains. The units of the dihedral angles are in degrees. For each snapshot during the simulation, a total of 1, 10 and 25 dihedral values were extracted, providing a total of 2000, 20000 and 500000 data points from the entire simulation.