Insight into Structure and Function of *Eco*RI Through Detailed Characterization of Nitroxide Side Chains using Molecular Dynamics Simulations

by

Gayathri Ayasna Rajapakse

B.S., University of Peradeniya, Sri Lanka, 2007

Submitted to the Graduate Faculty of

School of Arts and Science in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

School of Arts and Science

This thesis was presented

by

Gayathri Ayasna Rajapakse

It was defended on

December 7, 2010

and approved by

David Pratt, Ph.D., Professor Department of Chemistry, Faculty of Arts and Sciences

Adrian Michael, Ph.D., Professor Department of Chemistry, Faculty of Arts and Sciences

Thesis Director: Sunil Saxena, Ph.D., Professor, Department of Chemistry, Faculty of Arts and Sciences

Insight into Structure and Function of *Eco*RI Through Detailed Characterization of Nitroxide Side Chains using Molecular Dynamics Simulations

Gayathri Ayasna Rajapakse, M.S.

University of Pittsburgh, 2010

This thesis describes molecular dynamics simulations used in an effort to investigate the orientation and dynamics of the methane thiosulfonate spin label (MTSSL) on solvent exposed sites of the loop (R131C) and β-strand (S180C) of the specific *Eco*RI–DNA complex. Also, the results are compared with those of a well-studied solvent exposed α-helical site (SEHS). By modeling the spin label at these sites using the all-atom simulation, we intend to explain the observed double electron–electron resonance (DEER) distance distribution data of the specific complex of R131C and S180C mutants. *Eco*RI is a restriction endonuclease, which binds and cleaves the specific DNA sequence 5′GAATTC3′. The arms of *Eco*RI are thought to play an important role in its high binding affinity. However, the exact reason(s) why it shows such an extraordinary binding specificity is still being investigated. Therefore, to shed light on the high binding specificity of the enzyme, distance measurements were previously obtained using DEER, providing the structural constraints of the arms of *Eco*RI. Since the DEER experiment provides inter-spin label distances, simulations can be used to de-convolute the spin label contribution to extract the inter-atomic distances (C_α–C_α) of the arms of *Eco*RI. In order to accomplish this, we intend to effectively reproduce the experimental DEER data using molecular dynamics simulations. Utilizing simulations, the structural constraints contained in experimental data can be obtained and compared with the X-ray crystal structure of the specific *Eco*RI–DNA

complex. Although we attempt to explain the DEER data by examining spin-label conformations, the study also involves linking the dynamics of the spin label to the MD-generated distance distributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis is dedicated to the people who have supported me during my two years of graduate school. First I thank my advisor, Professor Sunil Saxena, for his enormous support and dedication in guiding me in the right direction. He supported me throughout my thesis with his patience and knowledge. It was a great pleasure to have worked with him. It is an honor to have Professor David Pratt and Professor Adrian Michael in my thesis committee and I am grateful for their time and advice for my thesis.

I also would like to thank all my past and present group members: Dr. Katherine Stone, Dr. Sangmi Jun, Dr. Zhongyu Yang, Dr. Sharon Ruthstein, Byong-kyu Shin, Jessica Sarver, Ming Ji, Ishara Silva, and Timothy Cunningham. I would especially like to express my sincere gratitude to Jessica Sarver for her commitment, patience, and enthusiasm to train me in spectroscopy and molecular dynamics simulations. She not only helped me improve my presentation skills, but also committed her time to go over my thesis in many instances. She is the one who always encouraged me during hard times. I am also grateful to Byong-kyu Shin for dedicating his time to teach me ESR theory and providing suggestions for my thesis. It is really an honor and a joy to work with them both.

I thank Richard Christy for helping me with the cluster account for my MD simulations. I also thank Jamie Novak at the Writing Center of University of Pittsburgh for correcting my

thesis. Finally, I would like to thank my parents and my sister for their support, patience, and encouragement throughout my graduate studies. Last but not least I thank my friends and relations for their encouragement. I appreciate everything they did to make this thesis complete.

1.0 INTRODUCTION

Protein-DNA interactions are important in understanding the regulation of gene expression, site specific recombination, as well as design of novel therapeutic agents and a new generation of research tools in genetic engineering [1]. The *Eco*RI–DNA complex can be used to study protein-DNA interactions. *Eco*RI is a restriction endonuclease that specifically binds and cleaves a six base pair palindromic DNA sequence 5′GAATTC3′ [2]. Importantly, the enzyme can bind to its correct recognition site up to 50,000–90,000 fold better than noncognate sites, which have one or more incorrect base pairs [3-5]. Therefore, it is of interest to fully understand how such base pair mismatches can drastically change the function of the enzyme.

To gain more clarity on such behavior, a pulsed electron spin resonance (ESR) spectroscopy technique, double electron–electron resonance (DEER), was utilized in conjunction with site-directed spin labeling (SDSL). The obtained distance measurements in the arms of *Eco*RI for specific and noncognate complexes provide the structural constraints of the protein [6]. However, the DEER experiment measures the dipolar interaction between the spin labels and therefore resulting inter-spin label distances. Consequently, the inter-atomic distances associated with the arms of the protein backbone are not readily obtained from the experimental data. In order to paint a more accurate picture of the structure and dynamics of the spin labeled locations of *Eco*RI from ESR measurements, knowledge of the average spin label orientation is required. Therefore, this thesis describes a molecular dynamics simulation method used to model the spin

label at sites 131 (on a loop) and 180 (on a β-strand) to explain the previously obtained DEER distance distribution data of *Eco*RI. Simulations are done to explain only the DEER data of the specific complex for reasons further discussed below.

Our research has three major goals: obtaining backbone constraints of the arms of *Eco*RI by effectively reproducing the experimental DEER distance distribution data; understanding spin-label dynamics and orientations on different protein secondary structures (loop and β-strand), and comparing such results with those of a well studied α-helix. Thus, we show that unlike other simulations done so far, there can be more knowledge obtained than just reproducing the observed DEER data. The study shows the investigation and generation of preferred spin-label orientations on different protein secondary structures, the effect of the initial spin-label orientation on simulation results, and the simulation time length needed to effectively model the spin label by utilizing an all-atom simulation.

1.1 SITE DIRECTED SPIN LABELING (SDSL) - ESR

Site-directed spin labeling (SDSL) is a technique that was pioneered by Dr. W.L. Hubbell and co-workers and has emerged as a valuable method for the investigation of structure and dynamics in a wide range of proteins [7]. The technique is usually used in conjunction with electron spin resonance (ESR) spectroscopy. The SDSL method has evolved over the past decade to a point that, it is possible to routinely determine the elements of secondary structure, including their solvent exposure, and orientations and movements of individual segments of membrane proteins. Several reviews have now appeared that fully discuss the capabilities of SDSL [8-11].

In order to utilize ESR, the system should have an unpaired electron to make the system ESR active. Certain proteins possess endogenous metal ions that are ESR active, however, introduction of an extrinsic spin label is required for proteins, which do not contain such endogenous paramagnetic centers. Thus, introduction of a paramagnetic species is required. 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl-methanethiosulfonate (MTSSL) is the most commonly used spin label in the SDSL technique. In SDSL the nitroxide spin label is designated as R1 [12]. As shown in Figure 1-1, R1 contains a stable unpaired electron which is delocalized between the N-O bond of the nitroxide moiety [13].

Since the spin label specifically reacts with a non-disulfide bonded cysteine of the system, the basic strategy of SDSL involves the substitution of a native residue at the desired site with cysteine through site directed mutagenesis, followed by modification with MTSSL. Thus, solvent accessible free cysteines that are present elsewhere must be mutated to a different amino acid in order to prevent SDSL at those sites. Figure 1-2 illustrates the reaction between MTSSL and the cysteine residue of a protein. As demonstrated in T4 lysozyme and colicin, site-specifically generated cysteine mutants cause little effect on the structure of the protein [14-15].

The intrinsic motion and the orientation of the spin label are defined by its five dihedral angles (Figure1-1). Extensive studies done on the spin label at solvent exposed α-helical sites (SEHS) illustrate that internal motion of this side chain is largely limited to rotations about the last two dihedrals (χ_4 and χ_5) and the remaining dihedrals are fixed on the ESR time scale [13, 16-18]. Although there is a wealth of studies done on spin-label dynamics and orientations on SEHS, enhancing the knowledge on spin label behavior on different secondary structures such as loop and β-strand is also important.

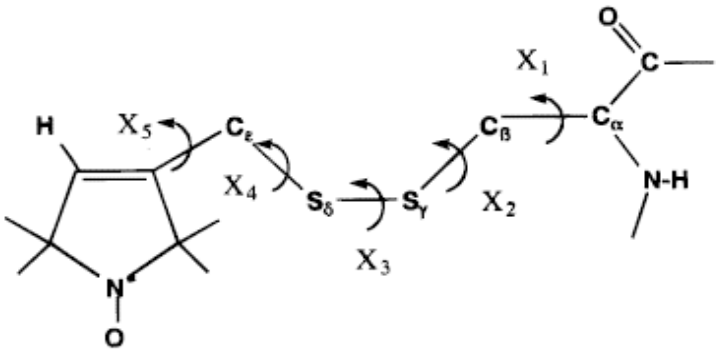
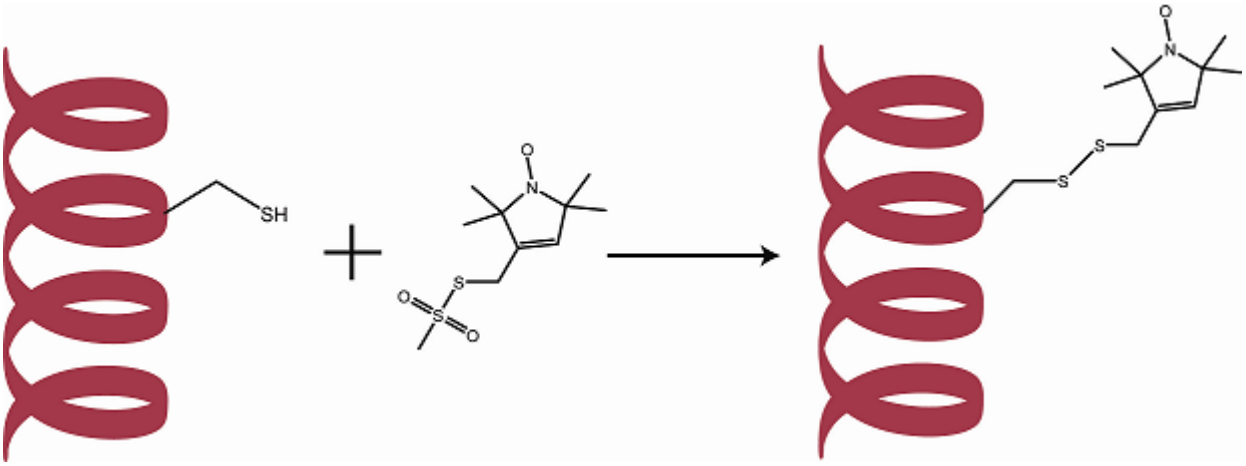


Figure 1-1: Structure of the methane thiosulfonate spin label (MTSSL)-R1. The orientation of the spin label can be defined by the five dihedral angles [13].



MTSSL **Spin-labeled protein**

Figure 1-2: The reaction between the MTSSL and the cysteine of a protein with an α-helix.

1.2 DYNAMICS AND DISTANCE MEASUREMENTS USING ESR

Research on the structure of biological molecules is essential in order to better understand their function. The increasing realization of the significance of structural heterogeneity and disorder-order transitions of biological systems is an important aspect of research that needs to be analyzed. Biophysical methods such as NMR, X-ray crystallography, and fluorescence resonance energy transfer (FRET) are useful in gaining such insight into biological systems like proteins. However, these techniques are limited to systems with small molecular weights, low volumes and high sample concentrations, and biological systems that can be crystallized.

ESR, on the other hand, is sufficiently sensitive to large biomolecular complexes, at low concentrations, overcoming drawbacks experienced by other techniques. The recent advances in SDSL have led to ESR becoming a useful technique in biophysics. The SDSL can be used to understand the conformations and local dynamics of the spin-labeled side chain, including the features of proteins that influence the ESR line shape. Secondly, pulsed techniques can be used to determine long-range distances in given sites of the protein. During the past several years, these technical developments have been used to address several important problems concerning the molecular function of proteins [19].

ESR line shapes of nitroxide spin labels have been extensively analyzed. Continuous wave ESR (CW-ESR) is one of the most common ways to measure dynamics of spin-labeled proteins. The CW-ESR line shape is determined by the motion of the spin label on the nanosecond time scale. The internal dynamic modes of the spin label and the protein backbone contribute to this motion. As a result of the combination of motions, ESR spectrum is generated. The line width of the ESR spectrum has been used as a measurement of spin-label mobility.

Therefore, utilizing various computational analysis programs, the observed CW-ESR spectrum is fitted to obtain quantitative analysis of the dynamics of the spin-labeled protein [12, 20].

On the other hand, SDSL with two spin labels at the desired sites of the protein provides a means to measure distances within a biological molecule. Typically two types of distance measurement experiments are performed. The first is a continuous wave (CW) technique in which line broadening due to dipolar interactions is analyzed to extract distances in the range of 8–20 Å. The second approach is based on pulsed ESR techniques which can access longer distances between the two spin labels. During the past years such distance measurement techniques have been applied to structure elucidation of membrane proteins, micelle bound proteins, and nucleic acids [21-23]. Pulsed ESR techniques such as DEER can access distances from 20-80 Å. To measure distances from DEER, the sample is flash frozen from the room temperature. The measured static distance distribution at low temperature can reflect the flexibility of the protein at physiological temperatures. DEER is able to capture a variety of distances between the spin labels resulting, a probability–distance distribution. The average inter-spin label distance can be obtained from this distance distribution. The study in reference [24] demonstrates how distance distributions obtained from DEER can lead to such a clear cut insight into the flexibility of the protein backbone at the location of the spin label.

DEER measures the dipolar interaction between the two unpaired electrons (between the two spin labels) according to the classical equation given below:

$$E = \frac{\mu_0}{4\pi}\left[\frac{(\vec{\mu}_A \cdot \vec{\mu}_B)}{r^3} - \frac{3(\vec{\mu}_A \cdot \vec{r})(\vec{\mu}_B \cdot \vec{r})}{r^5}\right]$$

Where, E is the energy of the interaction, $\vec{\mu_A}$ and $\vec{\mu_B}$ are the magnetic moments of the interacting spins, and r is the distance vector between the two spin labels. The dipolar interaction between the two coupled spins can be shown according to Figure 1-3.

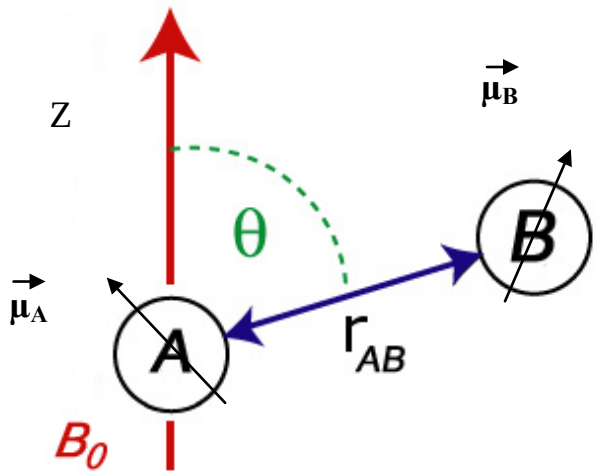


Figure 1-3: The dipolar interaction between the two coupled spins A and B where B_0 is the external magnetic field in the Z direction, $\vec{\mu}_A$ and $\vec{\mu}_B$ are the magnetic moments of the two spins, θ is the angle between the vector connecting the two spins, and r_{AB} is the distance between the two vectors.

The technique measures the distance dependant dipolar interaction. Thus, using the above equation, the relationship between the frequency of the dipolar interaction and the distance between the two spin labels can be derived and is given by the following equation:

$$\omega_{AB} = \frac{\mu_0 \beta_e^2 g_A g_B}{4\pi\hbar} \cdot \frac{\left(3\cos^2\theta_{AB} - 1\right)}{r_{AB}^3}$$

Where, ω_{AB} is the dipolar frequency, μ_0 is the vacuum permeability, g is the g-factor of the electron, β_e is the Bohr magneton, r_{AB} is the inter-spin label distance, and θ is the angle between the inter-spin vector and the external magnetic field. As the ω_{AB} is inversely proportional to r_{AB}^3, from DEER we obtain the inter-spin label distances and not the inter-atomic distances (C_α–C_α) of the sites to which the spin labels are attached.

2.0 SPIN LABEL MODELING

Among all the experimental techniques probing the structure and dynamics of proteins, SDSL-ESR undoubtedly offers information on biological systems with great specificity. The combination of recent advances in experimental methods for generating doubly spin-labeled proteins and the analysis of spectral data have opened the door to a whole new class of structural problems. Such advancements have offered the opportunity to bridge the gap between the structure and function through the elucidation of the dynamics of structural changes.

However, the complexity of the spin label, the finite length introduced by the spin label (5–12 Å) and the five degrees of freedom allowing it to generate a number of conformations, is a major concern [25]. Thus, obtaining detailed dynamics of the protein backbone from the ESR spectrum is difficult. The majority of existing spin-label dynamics and orientation studies done so far are in relation to explaining CW-ESR line shape [26-34]. The various computational methods developed in explaining CW-ESR provide an insight to the reorientational dynamics of the attached spin label.

DEER in contrary provides a variety of inter-nitroxide distances relating to different conformations of the spin label due to the aforementioned complexity of the probe. Provided that the spin label has so much flexibility, a variety of inter-spin label distances is possible which is reflected as a probability distance distribution rather than just a single distance between them. Importantly these distance distributions are a convolution of both backbone constraints and the

intrinsic spin-label motion. Thus, it is necessary to disentangle this backbone motion from that of the spin label. In fact it is equally interesting and significant to study both dynamics and conformations of the spin label to relate that information to obtain protein dynamics and constraints. Therefore, to maximally exploit the information present in ESR spectra and DEER data, quantitative knowledge about the intrinsic flexibility of the spin label is required.

2.1 WHAT IS KNOWN ABOUT THE SPIN LABEL

Interpretation of the DEER distance distributions and CW-ESR line shape derived from SDSL is quite challenging due to the length and the flexibility of the spin label mentioned above [13]. However, computational analysis such as *ab-initio* calculations shows that the spin label prefers certain orientations and dynamics at a given site of the protein [16, 18]. The most commonly used MTSSL has five dihedral angles from χ_1 through χ_5, which can be rotated independent of the protein backbone (Figure 1-1). Depending on these angles, the unpaired electron that is localized on the nitroxide bond can vary between 5–12 Å from the C_α of the protein backbone to which the spin label is attached [29]. Extensive studies have been done on the spin label by various computational techniques with the purpose of rationalizing side chain mobility on ESR line shapes of SEHS.

According to Tombolato *et. al.*, spin-label dynamics on a SEHS can be described by transitions between conformers [15]. The transitions are two types-torsional oscillations and conformational jumps-and depend on the time scale of the motion and the energy between each transition. As a result of low energy barriers, bond rotations enable the torsional oscillations to

occur over time scales shorter than a nanosecond. The oscillations are characterized by smaller amplitudes on the order of 10^{o} or less. The conformational jumps, on the other hand, are characterized by larger amplitude changes in a slower time scale than torsional oscillations. Conformational jumps are determined by two factors: the height of the energy barrier and frictional effects introduced by rotation of the spin label in the solution environment. All dihedral angles of the spin label show torsional oscillations. However, conformational jumps about the most proximal dihedral angles, χ_1 to χ_3, are unlikely due to the presence of high torsional barriers and/or frictional effects. On the other hand, the distal dihedral angles, χ_4 and χ_5, show much mobility as they possess low energy barriers and experience minimal frictional effects. Therefore, the spin label mobility is mostly determined by χ_4 and χ_5 angles. Such mobility results in motional averaging of the spectra which is explained by the χ_4/χ_5 model. Nevertheless, the nitroxide ring can have hydrophobic interactions with the adjacent amino acid side chains of the protein possibly restricting the motion of χ_4 and χ_5 to a certain extent.

The work from Tombolato *et. al*. further shows that at an SEHS the S_δ of the disulfide bond interacts with C_α of the protein backbone, thus constraining the rotations about χ_1 and χ_2 (Figure 1-1) [16]. According to the χ_4/χ_5 model, the inter-residue immobilization of the disulfide by the backbone effectively locks the conformational transitions about the first two dihedral angles, χ_1 and χ_2. The *ab-inito* calculations confirm that χ_1 usually adapts -60° (-g) and 180° (t) where as χ_2 can be -75° (-g), +75° (+g) and 180° (t). The letters “g and t” refer to Gauche and Trans conformations, respectively. The χ_3 dihedral angle primarily exists in either $\pm 90^{o}$. The inter-conversion between the two major conformers of the disulfide is slow due to the high torsional energy barrier, 6–7 kcal/mol, and significant variations besides $\pm 90^{o}$ are rarely seen

[16]. Due to this high energy barrier of the isomerization of χ_3, the chain motion is largely determined by χ_4 and χ_5, thus the given name.

Hubbell and co-workers obtained the crystal structures of the spin-labeled T4 lysozyme, which support the computational analysis of the MTSSL [13]. For each X-ray structure investigated, the disulfide bond of the spin label was well-resolved indicating that the disulfide group is localized in space. This in turn implies the values of χ_1 and χ_2 dihedral angles are constrained. According to the crystal structures of the T4 lysozyme, both χ_1 and χ_2 mostly prefer {–g, –g} and {t, +g} configuration to a lesser extent, and the preferred state is justified by a weak interaction between S_δ and C_α–H. The mildly acidic C_α–H is capable of forming hydrogen bonds with S_δ which may act as a hydrogen bond acceptor [13]. Together the crystallographic data also suggest that the presence of the disulfide linkage and its interaction with the main chain atoms effectively locks the C_α–C_β–S_γ–S_δ atom group in position and relinquish motional averaging to the terminal bonds χ_4 and χ_5.

Significant progress has been made so far at SEHS, using mutagenesis, X-ray crystallography, and quantitative analysis of ESR spectra of spin-labeled proteins by modeling techniques. Nevertheless, there is a dearth of information about spin-label dynamics on a β-strand and a flexible loop region of proteins. An attempt has been made by Lietzow and Hubbell to extract dynamics of the spin label on a β-strand of cellular retinol-binding protein [35]. Essentially, the structural constraint on spin-label motions on a β-strand is different from that of an α-helix. The nearest neighboring side chains of amino acids in α-helices show little interaction with the spin label in most solvent exposed sites. This minimal interaction is due to large spacing of the side chains because they project radially outward from each other. In contrast, the nearest neighbor side chains in a flat β-strand are more closely spaced than in an α-helix and lie

essentially parallel to one another. However, in the case of a twisted β-strand, the side chains point away from each other minimizing interactions between them. Using molecular models to compare to the observed ESR spectra, the study shows that for a solvent exposed β-strand the missing hydrogen bond (HB) neighbors permit {+g, -g} and {–g, t} conformations. The study further demonstrates that the nitroxide motion can be ordered due to the interaction of the disulfide with the main chain atoms, particularly N–H$^{\cdots}$S$_{\delta}$ interaction. Strand twisting can reduce non hydrogen bonding (NHB) neighbor interactions allowing {–g, -g} configuration as well, which is also preferred in SEHS [35].

It is evident that the majority of these studies done on spin-label dynamics involve α-helices. Given that there is little known about spin-label dynamics on a β-strand and hardly any for a loop, *Eco*RI offers the benefit of investigating dynamics and orientations of the spin label on different protein secondary structures. Thus, this study will be an initiative to model the spin label on different secondary structures of the protein and compare them to SEHS.

2.2 MD SIMULATIONS AND METHODS

Over the past decade, considerable work has been devoted to developing a theoretical framework for the conformational analysis of the spin label with the purpose of characterizing its geometrical and motional parameters and analyzing ESR line shape [27, 32, 36]. Some of the existing computational approaches, based on Stochastic Liouville Equation, employ various models for the molecular motions of the spin label and the protein backbone such as rigid-body diffusion for the protein combined with Brownian ‘‘diffusion in a cone’’ for the spin probe, ‘‘two-site jumping’’ motion, Brownian dynamics, and finally, atomically detailed MD [31, 37-38].

Recent work done in order to reproduce inter-spin label distances obtained by DEER have used varying approaches [39-41]. Generally, the contribution of the backbone can be known if the contribution from the spin label is eliminated. A previous study done by Jeschke and co-workers demonstrated, for a series of shape persistent biradicals, the end-to-end distance distributions of the backbone can be faithfully recovered by fitting the label-to-label distance distributions of the spin labels extracted by MD simulations [41]. The simulations were performed in vacuum describing the system as a coarse-grained model and treating the spin labels as internally rigid segments. Thus, they indicate that simulations in vacuum are a good model for investigating the conformational distribution of the spin label. However, developing a reliable method in order to model the spin label can take different approaches such as all-atom, united atom simulations, and simulations using different force fields [23, 42-43]. A study done on the force field impact and spin label modeling show that subtle site-dependent differences in spin label rotations are more readily captured in the AMBER99 than in the OPLS/AMBER. However, the study further shows that sampling and equilibration are found to be better with the

OPLS/AMBER force field at equal trajectory lengths [30]. In addition to using different force fields, techniques such as Monte Carlo simulations have also been introduced prior to MD.

A recent MD study done in vacuum explains the distance measurements obtained by DEER between the primary quinine electron acceptor of *Rhodobacter sphaeroides* and the nitroxide spin label. The simulations were run for 6 ns at 600 K and demonstrate a good agreement between the experimental data and MD [43]. Fajer and co-workers also performed a similar study on AntR monomer and Myosine proteins in order to explain DEER data by introducing a rotamer search prior to MD. Favorable rotamers were chosen from a rotamer search and short 1 ns MD simulations were performed at 300 K using them [39-40]. The study reveals that satisfactory agreement between the experimental data and MD generated distance distributions cannot be assured from short MD runs due to insufficient sampling of the spin label.

All these approaches are intended to offer dynamics and orientations of the spin label, however, correlating orientations and dynamics of the spin label with that of the inter-spin label distances generated in MD has not been explicitly studied yet. Thus far, simulations performed to model the spin label use various conditions such as high temperatures, implicit/explicit solvent, vacuum conditions, as well as varied time lengths. Although there is no common approach used to model the spin label, insufficient sampling is usually a major problem associated with these techniques. While performing simulations in vacuum and/or high temperatures increases sampling, these conditions may not reflect the true dynamics of the spin label or the protein of interest. Further, straightforward conclusions cannot be made claiming that the results agree or disagree with the experiment solely based on running short simulations for just a single starting orientation of the spin label. In fact, to accept the obtained results from the simulations discussed above, it is necessary to assess the reliability of different approaches, which is very complicated.

For certain given approach, its consistency and validity should be thoroughly investigated on the system of interest.

Although there are discrepancies associated with different techniques used, these early studies provide the basis for further development of methods to simulate ESR spectra and possibly fit data to determine both the average inter-spin distance and the effective relative orientations of the spin labels using simple models of spin-label dynamics. Alternatively, the results of accurate simulations of spin-label dynamics and orientations may be used in the analysis of experimental data.

2.2.1 Metropolis Monte Carlo Minimization Conformational Search (Rotamer Search)

As mentioned before, a common problem encountered in the majority of the simulation techniques is insufficient sampling. Of course, an important requirement of trajectories is a representative sampling of the conformational space explored by the spin label. However, in a given trajectory, the spin label could remain trapped in a minimum, depending on the starting orientation of the spin label. Therefore, this problem should be carefully addressed to obtain reliable information from MD. The accurate selection of spin-label conformations with respect to the protein frame is an ever-present challenge in computational modeling. In the past, several methods have been used to determine the spin-label orientation in the molecular frame of reference [44]. In an effort to develop a tool to effectively screen and produce a set of spin-label conformations with respect to the protein backbone, Fajer and co-workers proposed a Metropolis Monte Carlo Minimization (MMCM) technique [33]. MMCM is a stochastic approach that searches the conformational space accessible to the spin label at a given site of the protein data

bank (PDB) structure [33, 40, 45]. Rotamer search is the method by which the MMCM algorithm results in different spin-label conformations. Rotamer search generates all possible rotamers of the spin label for a particular starting orientation at a given site of the protein, enables us to choose spin-label conformers that might not sample in MD. Thus, introducing this method prior to MD is expected to improve sampling.

Precisely, rotamer search generates a library of spin-label conformations with various energies by randomly rotating the dihedral angles of the spin label. For each iteration, dihedral angles from χ_1 through χ_5 of the spin label are rotated by random amounts. The generated rotamers will be accepted or rejected based on the Metropolis criterion. During the rotamer search, the protein backbone is kept fixed and only the spin label is allowed to move (rigid cage assumption). The MMCM approach calculates the potential energy of the spin label from the bond stretching, angle bending, dihedral, van der Waals, electrostatics, and hydrogen bonding potentials. Fajer and co-workers demonstrated that when MMCM is introduced prior to MD simulation, there is a good correlation between the measured ESR distances and the simulated inter-spin label distances [42]. In all of their rotamer search-based studies they chose the low energy conformers of the spin label generated by rotamer search to use in MD simulations for a more effective sampling.

3.0 *ECO*RI–DNA SPECIFIC COMPLEX

Restriction endonuclease is a type of enzyme that cuts a double-stranded DNA at a specific recognition nucleotide sequence known as a restriction site. The restriction site is usually a 4–8 base-pair-long palindromic sequence [2]. There are several types of restriction endonucleases, and this thesis involves a study of a type II restriction endonuclease which is found in bacterial cells and used as a defense mechanism against invading viral DNA. By recognizing and binding to the specific viral DNA, these enzymes can cleave that viral DNA and prevent the replication of the restriction site within the bacterial cell. However, since the specific sites of the bacterial genome are chemically modified, the restriction enzyme can discriminate between the viral genome and that of the bacterium [46-47]. Usually the enzyme requires a co-factor such as Mg^{2+} in order to cleave the DNA strands. Type II restriction endonucleases have been widely studied due to their importance in gene analysis and cloning work. In addition, these enzymes have been the subject of extensive genetic, biochemical, and biophysical research for more than a decade attributing to their extraordinary sequence selectivity and high binding specificity [48-49]. Therefore, these enzymes are used as important models to understand protein-DNA interactions.

*Eco*RI is a type II restriction endonuclease that binds to a specific hexanucleotide sequence and cleaves at the indicated positions as shown below [2, 50].

The enzyme is found in *E. coli* bacterial cells and it has to locate the above DNA sequence out of 49,000 base pairs in the viral genome, making its binding specificity crucial [51]. Although the specific sites in the *E. coli* DNA are chemically modified to protect against cleavage, there are 18,000–21,000 unprotected miscognate sites (different from the specific sequence by one base pair) that could be cleaved by *Eco*RI with a second-order rate constant [3]. However, there are ~4.6 million base pairs in the *E. coli* genome, ~4.5 million of which are nonspecific sequences (different from the specific sequence by two or more base pairs) [52]. The active enzyme is a homodimer composed of two identical monomers, each being 31 kDa in molecular weight [53]. *Eco*RI is one of the first restriction endonucleases to be crystallized, and thus there is a vast amount of structural and thermodynamical data collected on the specific complex [3, 5, 53-55]. Upon binding to the specific DNA, the so-called “arms” of *Eco*RI wrap around the sequence facilitating restraint of the DNA. The arm region consists of two distinct components: inner and outer arms (Figure 3-1). During complex formation, *Eco*RI undergoes disorder-order transition as observed in the crystal structure [56]. The enzyme is known to demonstrate a high binding affinity towards the specific DNA sequence as mentioned earlier. Previous studies demonstrate that both direct (protein-base) and indirect (protein-phosphate and DNA conformation) contacts are responsible for the high binding affinity [3]. However, both crystal structures of the free and

the bound enzyme and ethylation interference footprints (DNA phosphate) of the specific and miscognate complexes indicate that these arms play a role in DNA binding [3].

In particular the arms show crucial phosphate contacts at:

pXpGAApTTC

X is any nucleotide base in the flanking sequence and p represents the phosphate group at those positions. The phosphate contacts at the indicated positions facilitate base recognition and help to orient the protein recognition elements within the major groove of the DNA. These contacts act as clamps to stabilize the kinked DNA conformation [3]. The DNA backbone footprints suggest that these phosphate contacts are missing in *Eco*RI miscognate complexes. Although there is a wealth of information on the specific *Eco*RI–DNA complex, the stringent binding affinity demonstrated by *Eco*RI is still of major interest. In order to investigate what makes the enzyme show such a remarkable binding affinity, the structure of *Eco*RI–DNA specific complex has to be compared with that of the noncognate (miscognate and nonspecific) complexes. Since there are no crystal structures available for the noncognate complexes of *Eco*RI, there is little or no prior structural information on these complexes. Thus, understanding high binding affinity of *Eco*RI is not possible. In the realm of restriction endonucleases there are ~3600 known restriction endonucleases with 73 crystal structures, 38 of which are in complex with specific DNA sequence. However, there are only 4 crystal structures available for miscognate or nonspecific complexes in the protein data bank, which do not include *Eco*RI [57]. Therefore, knowing that SDSL-ESR is capable of offering structural constraints of biological molecules, DEER was utilized to make distance measurements in the arm region of the *Eco*RI–DNA complexes of specific and noncognate [6]. The measurements are expected to provide more

knowledge on these noncognate complexes to be able to compare them to the specific complex in order to shed light on the roles of the arms and relate to the differences in their binding affinities.

Figure 3-1: The specific *Eco*RI–DNA complex (1CKQ): Inner arm (orange) and outer arm (green) wrap around the DNA (grey) holding it against the main domain of the enzyme [56].

3.1 DEER DISTANCE DISTRIBUTIONS

Our group has previously obtained distance measurements using DEER on the specific and noncognate complexes of two *Eco*RI mutants: R131C-in the inner arm and S180C-in the outer arm (Figure 3-2) [6]. Due to the homodimeric nature of *Eco*RI, spin labeling at one site produces two spin-labeled sites making it feasible to use DEER [53]. The distance distributions corresponding to the DEER data of these mutants were generated using Jeschke's DeerAnalysis program [25]. The Figure 3-2 shows the probability–distance distributions of the specific complexes only, since the MD study focuses on explaining only those of the specific R131C and S180C complexes [6]. The DEER distribution data overall suggest that the *Eco*RI arms wrap the DNA and are similarly oriented in both specific and noncognate complexes. This in turn implies that the DNA in specific and noncognate complexes occupies roughly the same binding cleft of the enzyme [6]. In general DEER data demonstrate the flexibility of the arms of *Eco*RI of the specific and noncognate complexes.

As mentioned earlier, obtaining structural constraints of the protein requires the understanding of spin-label conformations in relation to their dynamics. Therefore, the rest of the thesis will talk about a molecular dynamic simulation technique used in an attempt to explain previously obtained experimental DEER distance distribution data of *Eco*RI. All MD simulations performed were in full collaboration with Jessica Sarver of the Saxena group. My primary role has been generating different initial orientations of the spin label, performing rotamer search on them, and generating PDB structures for subsequent MD simulations. I have been closely working with her during the development of the procedure and analysis of the results.

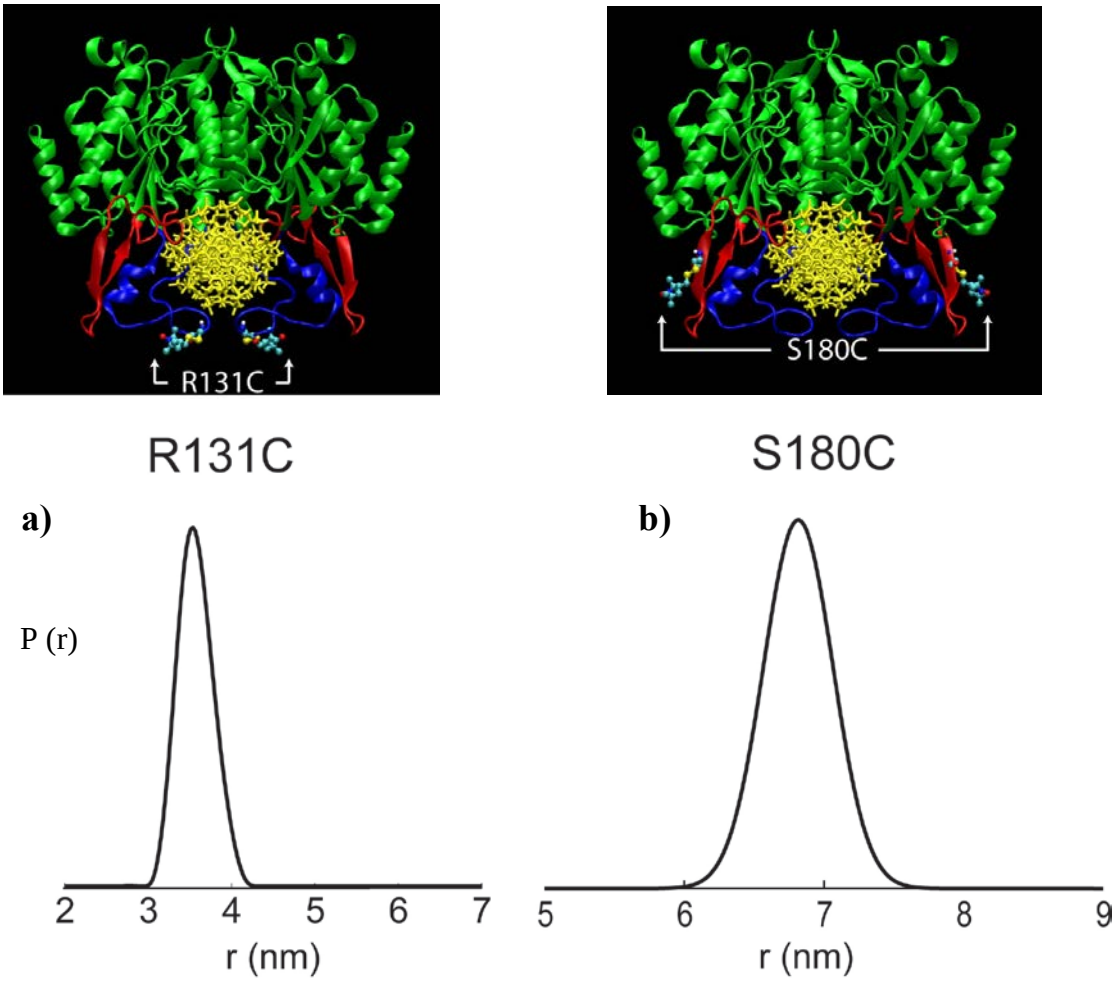


Figure 3-2: The probability–distance distribution data (obtained by DEER) of a) R131C (inner arm) and b) S180C (outer arm) of the specific *Eco*RI-DNA complex. R131C shows an average inter-spin label distance of 35 Å and S180C shows an average inter-spin label distance of 68 Å [6, 56].

3.2 MOLECULAR DYNAMICS SIMULATIONS PERFORMED PREVIOUSLY ON DEER DISTRIBUTION DATA

To extract protein backbone constraints of the arms of *Eco*RI from the observed DEER distance distributions, the spin labels at R131C and S180C sites were modeled previously through all-atom simulations [6]. The simulations were performed on the specific R131C and S180C complexes since the crystal structure is available only for the specific complex. Previous MD simulations executed on R131C possesses an average inter-spin label distance that agrees with the experimental DEER results. However, the simulated distance distribution is significantly broader than the experiment. On the other hand, MD failed to explain the experimental data of S180C (Figure 3-3) and appears bimodal. The MD-generated average inter-spin label distance of S180C has been found to differ from that of the experimental by 1–3 Å. In addition, the simulated distance distribution is narrower than the experimental distance distribution. The simulations were performed starting with an arbitrary orientation of the spin label on both sites, in explicit solvent, for 10 ns.

According to a recent work done by Fajer and co-workers, these differences can be attributed to insufficient sampling, in which several local minima of rotamers of the spin label are not being accessed in a single MD trajectory [58-59]. In other words, the system does not exhaustively search the conformational space available for the spin-labeled protein in 10 ns. The lack of sampling can be reflected in not only the average inter-spin label distance, but also the distance distribution as well. However, R131C resides on a loop whereas S180C resides on a β-strand. Thus, R131C could be sampling more orientations due to the high flexibility of the loop region compared to the β-strand. Certainly, agreement between the experiment and MD simulations has proven to be a bottleneck for such ESR measurements in a variety of systems. In

order to alleviate the problem of sampling and extract the entire distribution of the spin label movements, all possible conformers must be sampled.

Although simulations are useful to characterize the spin label with respect to the protein backbone, running MD simulations is still technically demanding as they are limited by the quasi ergodicity problem [59]. The basic idea of ergodicity is that if one allows the system to evolve over time indefinitely, the system will eventually pass through all possible states. Unfortunately, assessing the ergodicity of a single trajectory can be difficult depending on the starting orientation of the spin label [60]. Thus, efficient sampling can be obtained by performing multiple simulations for different starting orientations of the spin label and there by sampling conformations that are not sampled in a single MD trajectory. In addition, to effectively model the average spin-label orientation and dynamics, adequate time (~100 ns) is needed for simulations. Running long simulations will increase the probability of resolving the most favorable average nitroxide orientation(s).

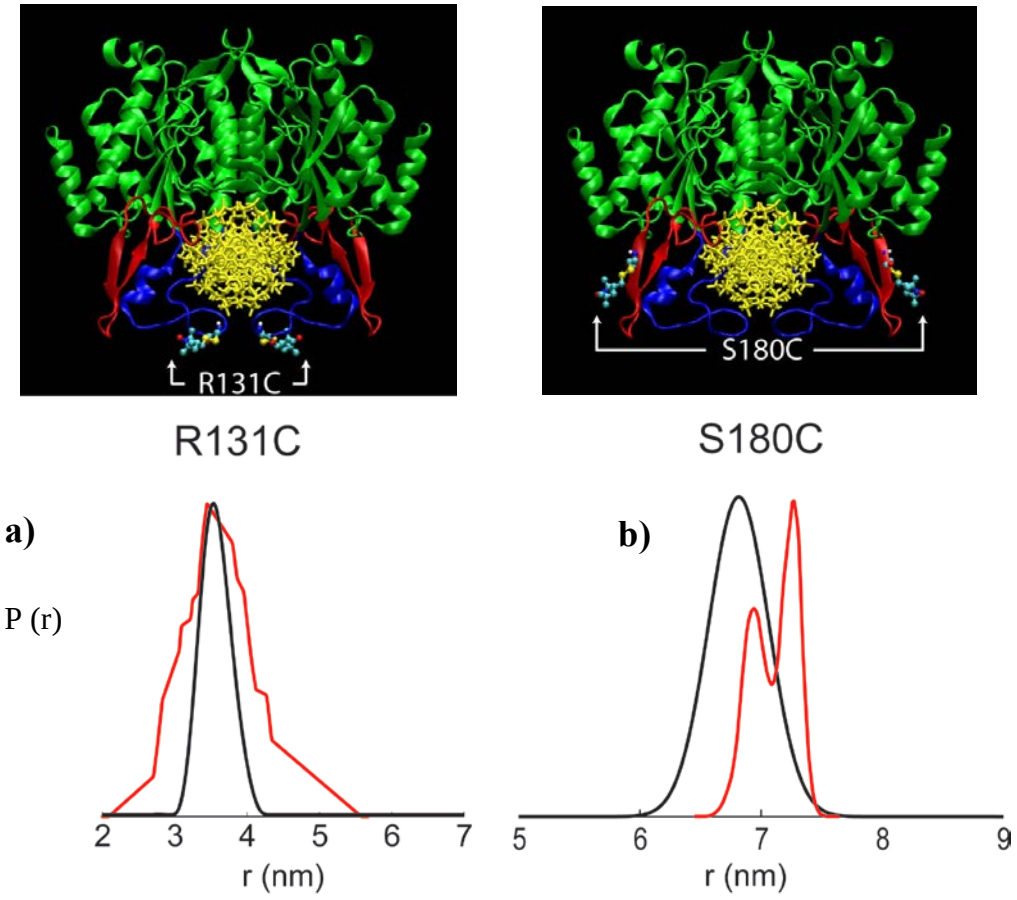


Figure 3-3: The MD-generated distance distributions overlaid on the experimental distance distributions of a) R131C and b) S180C (**MD** and **Experimental**) [6, 56].

4.0 COMPUTATIONAL ANALYSIS ON DEER DATA

Since part of the problem in previous MD results was insufficient sampling, following the work of Fajer and co-workers, our study introduces a rotamer search prior to MD. As described below, spin-label libraries are generated for different starting orientations, and favorable rotamers are then chosen to run independent simulations. The conformational search depends on the initial orientation of the spin label used. By choosing rotamers with different orientations, we expect them to sample spin-label conformations that are not sampled during a single trajectory (if the spin label is trapped in a minimum), thus ensuring exhaustive sampling. In addition, in order to ensure better and quality results from our simulations compared to previous MD, appropriate changes are made to the MD procedure as described below. The new MD protocol includes additional steps. The simulation protocol is kept consistent for all the constructed PDB structures.

4.1 MMCM (ROTAMER SEARCH) ON R131C AND S180C

Residues R131 and S180 of the *Eco*RI PDB structure 1CKQ [56] were mutated using a mutagenesis program in visual molecular dynamics (VMD) followed by an energy minimization using the CHARMM program. Then, rotamer search was performed for the given spin-label orientation. Since *Eco*RI is a homodimer, rotamer search was performed separately for each monomer of R131C and S180C mutants. For each site, ~2000 rotamers were generated to ensure the exhaustive survey of the rotamer space. To further ensure the thorough sampling of the spin label, nine more independent rotamer searches were performed on different starting orientations of the spin label for both R131C and S180C. The search was executed at 300 K.

To run subsequent MD simulations, PDB structures were constructed for both R131C and S180C choosing low energy rotamers. Rotamers belonging to different rotamer families (different orientations) were chosen to construct different PDB structures for the mutants. Figure 4-1 shows two different rotamer families generated in rotamer search. Each family differed by no more than ~1–3 kcal/mol, therefore still making the rotamers energetically favorable. To save computer costs in MD, the PDB structures were generated by combining coordinates of the spin labels on both 131 and 180 sites of the protein (Figure 4-2). In this way ten PDB structures were constructed. At spin-label positions 131 and 180, the distance between the spin labels was verified as longer than 15 Å as simulations involved calculation of long range interactions like electrostatics within a range of 10 Å. Since the samples used in the DEER technique are flash frozen from the room temperature, it is assumed that the sample represents the most stable room temperature conformations of the spin labels. Therefore, by choosing the low energy stable rotamers of the spin label at room temperature (300 K) and performing subsequent MD simulations on them it is expected that the MD results will be close to the experimental data.

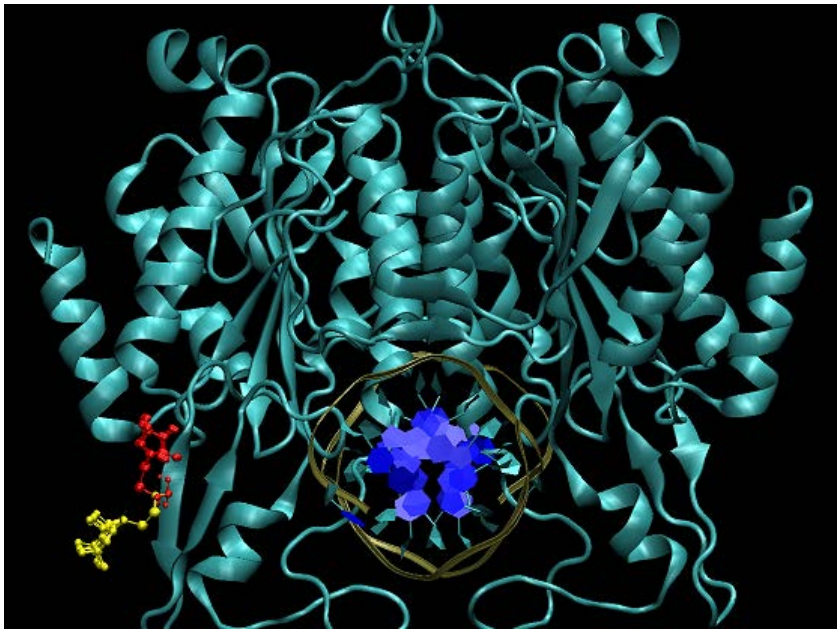


Figure 4-1: Rotamers belonging to different rotamer families (red and yellow) generated by rotamer search [56].

4.2 MD SIMULATIONS

To relieve the rigid-cage assumption and sample the local environment at the found minima, we ran MD simulations on all ten PDB structures using NAMD. Simulations employed the CHARMM22 force field. Each system was solvated in a cubic water box containing ~9226 TIP3P water molecules and Na^+ and Cl^- ions equivalent to a concentration of 0.22 M and neutral, with the box extending 8 Å away from the furthest atom of the system. Constructed PDB structures were energy-minimized again by keeping the protein, DNA, and the spin-label orientation fixed in order to remove strong van der Waals forces that would otherwise distort the system. For all simulations, the temperature was raised to 300 K in increments of 10 K over 10 ps time interval followed by an equilibration phase in an NPT ensemble. The simulations were run using a 2 fs time step in NVT conditions. The temperature and pressure were maintained at 300 K and 1 atm, respectively, using Nose-Hover method in which Langevin dynamics are used to control pressure and temperature. Particle Mesh Ewald (PME) was used for the calculation of long-range electrostatic interactions. The production phase of MD simulation was 30 ns for each system. All the simulations were performed on the Pople at the Pittsburgh Super Computer (PSC) and used 24,000 CPU hours.

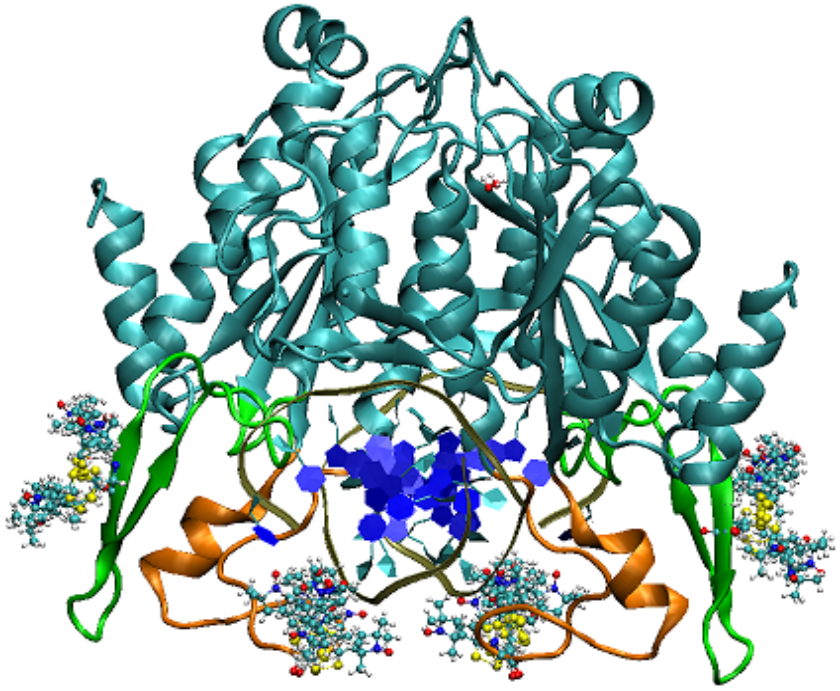


Figure 4-2: The different spin-label orientations chosen for R131C (orange) and S180C (green) sites for subsequent MD simulations [56].

5.0 RESULTS AND DISCUSSION

The spin-labeled mutants generated through exhaustive analysis of rotamer searches were used in subsequent 30 ns MD simulations (Figure 4-2). Analysis of the root-mean-square deviation (RMSD) of the protein backbone from the initial crystal structure provides a quality check of the simulations. As shown in Figure 5-1, the RMSD of the protein backbone is found to stabilize after 10 ns. A similar equilibration is observed for the other independent MD runs (data not shown). Therefore, only the last 20 ns of the simulation were used for the analysis. From these MD simulations, the following plots were generated for both R131C and S180C mutants: trajectories of the nitroxide distances and dihedral angles, inter-spin label distance distributions, correlation between the dihedral angles, and probability–chi angle distributions. A collection of data allowed for examining the efficiency of sampling, variation of spin label dihedral angles over time, and the evaluation of possible preferred conformers of the spin label to be able to compare them to SEHS.

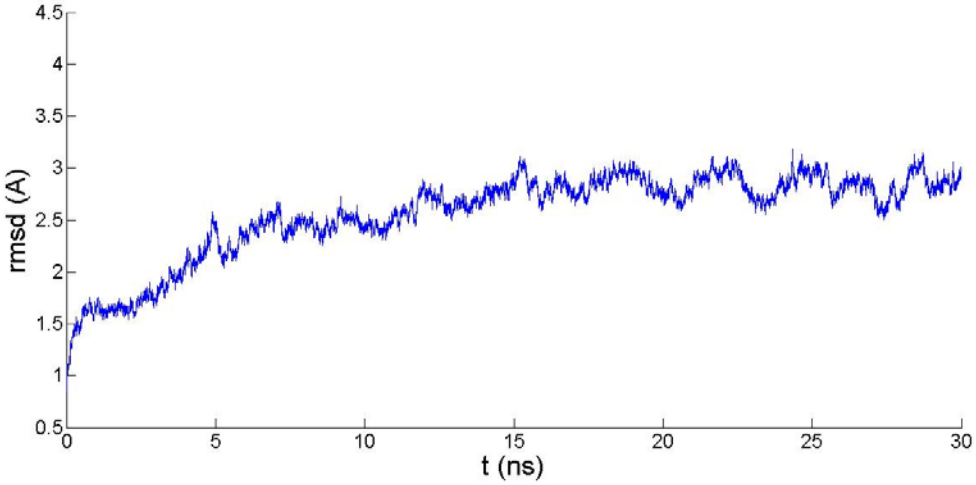
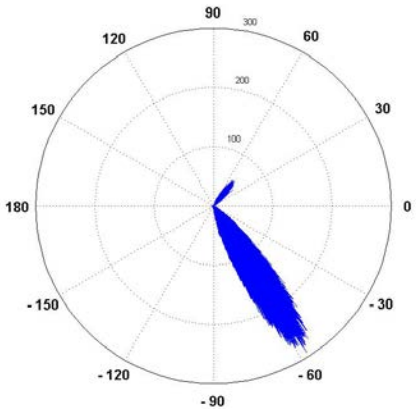
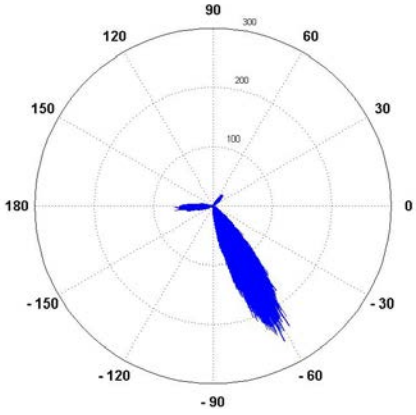
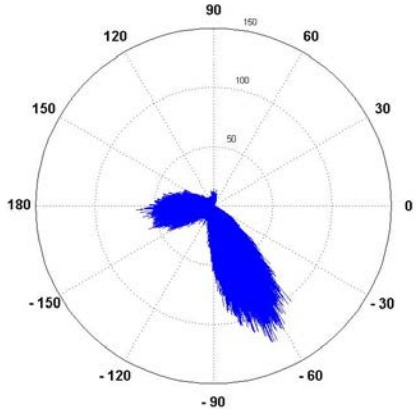
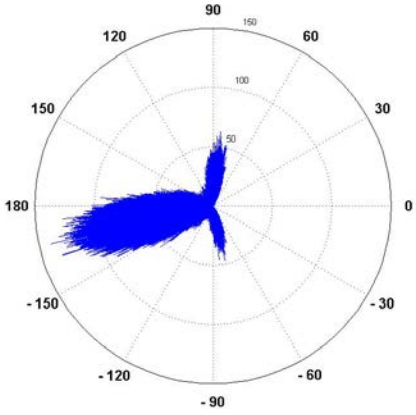
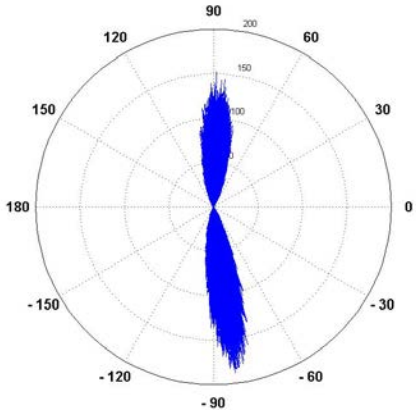
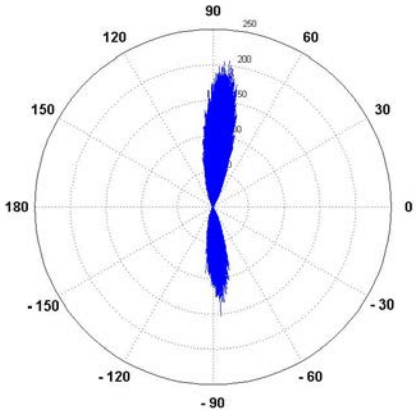


Figure 5-1: The root-mean-square deviation (RMSD) of the protein backbone in one of the MD trajectories. The first 10 ns of the simulation are not equilibrated.

5.1 SPIN LABEL DYNAMICS ON LOOP AND BETA STRAND

Molecular dynamics simulations performed in explicit solvent seem to show preferred conformations of the spin label when it is residing on a loop or a β-strand. According to the χ_4/χ_5 model of an SEHS, the first three dihedral angles of MTSSL (χ_1, χ_2 and χ_3) rarely undergo conformational jumps [13, 16]. Figure 5-2 shows the polar plots of the five dihedral angles χ_1, χ_2, χ_3, χ_4, and χ_5 of the spin label residing on a loop and a β-strand. The five dihedral angles shown in the Figure 5-2 are extracted from all the ten independent MD simulations for both sites of the homodimeric *Eco*RI–DNA complex for the last 20 ns of the simulation. Although over the 30 ns time period, χ_1 of R131C is quite steady at -60°, χ_1 of S180C frequently prefers -75° with jumps to 180° as well. As the simulations progress, the majority of S180C simulations preferentially shows -75° for χ_1. In general, we expect the spin label on a loop region to show more flexibility than that of a β-strand. The restricted mobility of χ_1 at R131C could be due to the close proximity of the spin label to the DNA. The preference for χ_1 at either -60° or -75°, as opposed to 180°, is due to the close proximity of S_δ and C_α–H, possibly resulting the $S_\delta \cdots H–C_\alpha$ interaction. This observation is consistent with the work done by Hubbell and co-workers.

On the other hand, time trajectories of χ_2 (data not shown) show much mobility (short-lived transitions) for both R131C and S180C, being incompatible with the χ_4/χ_5 model. Apparently, there is an unexpected mobility associated with the most proximal dihedral angles of the spin label, χ_1 and χ_2 in our simulations. Nevertheless, Figure 5-2 clearly shows that χ_2 of R131C prefers -75° with significant jumps to 180°, whereas χ_2 of S180C prefers 180° with jumps to +75° and a minute number of jumps to -75°.

	R131C	S180C
χ_1		
χ_2		
χ_3		

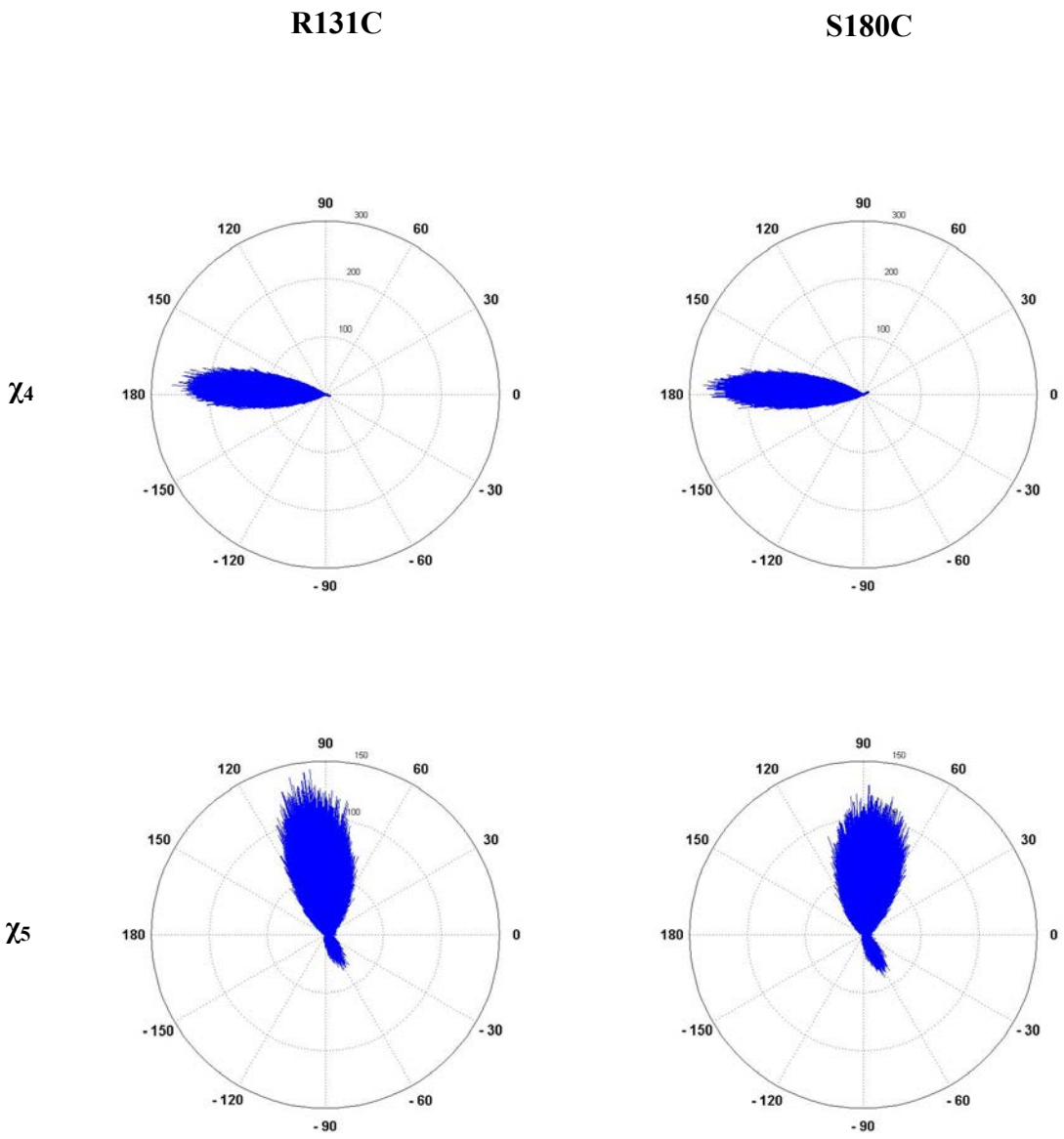


Figure 5-2: Polar plots of the five dihedral angles of the spin label at R131C (loop) and S180C (β-strand) extracted from all ten independent MD simulations for both sites of the homodimeric *Eco*RI–DNA complex for the last 20 ns of the simulation.

According to Hubbell and co-workers, a twisted β-strand should reduce NHB interactions with β-branched amino acid residues permitting {-60°, -60} for χ_1 and χ_2, respectively, given the predicted interaction between S_δ and N–H. Careful analysis of the twisted β-strand in our simulations demonstrates that regardless of the presence of a valine residue, adjacent to the spin label at S180C, χ_1 and χ_2 favor {-75°, 180°}, respectively. Hence, longer simulations are needed to observe whether χ_2 demonstrates any compatibility to what was suggested by Hubbell and co-workers.

The χ_3 angle behaves according to the suggested χ_4/χ_5 model and prefers either +90° or -90° (Figure 5-2). However, the time trajectories of χ_3 (data not shown) show that the dihedral angle rarely goes from +90° to -90° or vice versa during the 30 ns simulation. As the energy barrier for the inter-conversion between +90° and -90° conformations is high according to previous studies, the probability that one could observe this inter-conversion in MD simulations is low. Compared to the suggested χ_4/χ_5 model on an SEHS, χ_4 shows a lower mobility during the whole 30 ns time period, preferring 180°. This behavior of χ_4 is incompatible with the proposed χ_4/χ_5 model. The difference in the flexibility of χ_2 and χ_4 angles can be clearly seen in Figure 5-2. On the other hand, χ_5 variation is consistent with the χ_4/χ_5 model, showing much flexibility.

The investigation of the contribution of the spin-label conformations to the observed average inter-spin label distances obtained by MD is important. The most probable conformers of the spin label for each site were extracted based on the results of our correlation plots. These probable conformers were then generated in VMD using the mutagenesis program (Table 1). The Figure 5-3 shows the generated frozen spin-label conformations on both sites based on the values reported in Table 1. The color coding in Table 1 represents different conformations of the spin label in Figure 5-3. Careful analysis of the spin label conformations in Figure 5-3 shows that

some of these conformations could possibly have interactions with neighboring amino acid side chains of the protein. The probability of sampling a certain spin label conformation depends on the type of the interaction between the spin label and the neighboring amino acid side chain(s).

The generated frozen spin-label conformations enabled us to measure the inter-spin label distances as reported in Table 1. The reported values are some examples of possible inter-spin label distances that can be resulted from the preferred conformations of the MTSSL at R131C and S180C sites. These inter-spin label distances are compared to the average inter-spin label distances generated in MD. The comparison is important to understand the preference for a certain average inter-spin label distance observed in MD at R131C and S180C sites. In addition, it is also interesting to examine how much these observed inter-spin label distances in MD deviate from the measured inter-spin label distances of the frozen conformations. The inter-spin label distances shown in Table 1 are based on the assumption that both the spin labels have the same conformation at the sites of interest. It is possible that MD could be sampling average inter-spin label distances on or around the values given in Table 1. Since MD involves the dynamic motion of the whole protein backbone, the inter-spin label distances can be different from those reported in the table.

The majority of the conformers prefer χ_5 at +75°, and thus Table 1 provides distances corresponding to χ_5 = +75°. From Figure 5-3, two types of rotamer families can be identified for R131C and three types of rotamer families can be identified for S180C. These rotamer families can be sampled or observed in either experiment or MD if the differences in the inter-spin label distances between them are significantly high. In the table, “p” is when χ_3 = +90°, and “m” is when χ_3 = -90° (the designation is based on reference [18]). It is interesting to observe substantially different inter-spin label distances between rotamer families due to changes in χ_2

(Table 1). However, running longer simulations on all the PDB structures will resolve the most preferred conformation of the spin label on both sites, making it feasible to predict the average inter-spin label distance.

Table 1: Different rotamer families generated with their corresponding inter-spin label distances for R131C and S180C at $\chi_3 = \pm 90°$ and keeping χ_5 at +75°. (p: $\chi_3 = +90°$ and m: $\chi_3 = -90°$).

Site	χ_1	χ_2	χ_3	χ_4	χ_5	Inter-spin label distance (Å)
R131C (loop)	-60°	-75°	±90°	180°	+75°	p 31, m 33
	-60°	180°	+90°	180°	+75°	p 45
S180C (β-strand)	-75°	180°	±90°	180°	+75°	p 73, m 72
	+45°	180°	±90°	180°	+75°	p 73, m 72
	180°	180°	±90°	180°	+75°	p 73, m 72
	180°	+75°	+90°	180°	+75°	p 59
	-75°	+75°	+90°	180°	+75°	p 62
	+45	-75	-90	180	+75	p60

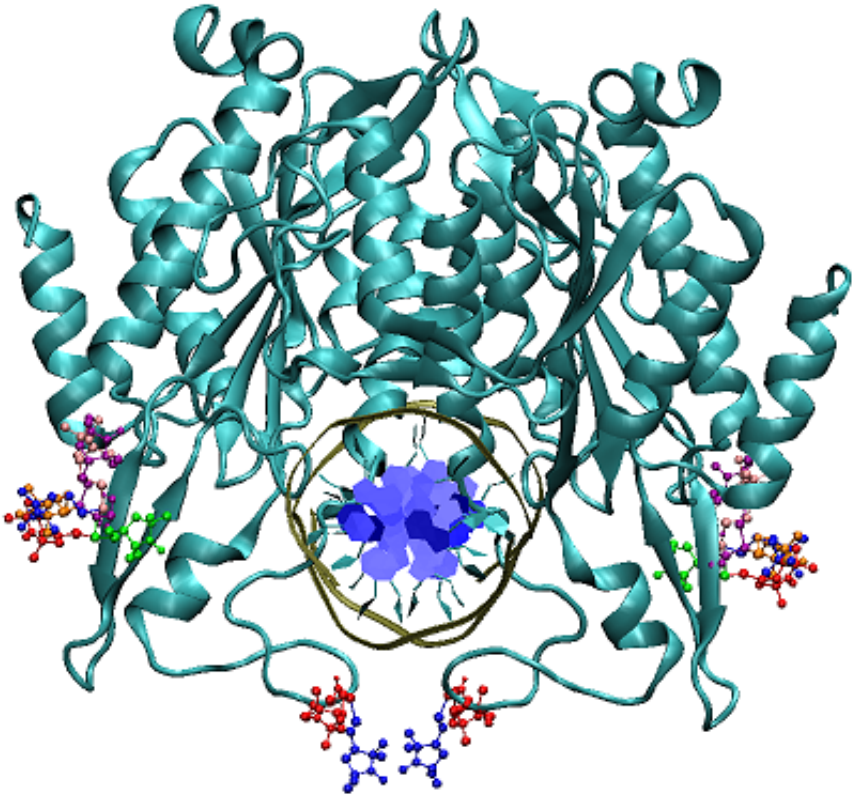


Figure 5-3: The generated frozen spin-label conformations of R131C and S180C based on the values reported in Table 1 [56].

5.2 SIMULATED DISTANCE DISTRIBUTIONS

We compared the modeled and experimentally derived distance distributions of the R131C and S180C sites of *Eco*RI. Regardless of the initial orientation of the spin label, the modeled average inter-spin label distances are longer than those of the experimental data by 1–4 Å for R131C. However, satisfactory agreement is seen for S180C (Figure 5-4). A similar trend, according to Figure 5-2, is seen for the rest of the nine independent simulations performed on R131C and S180C. The analysis of Freed and co-workers on spin-label dynamics studies suggests that a single spin label can have a conformational distribution between 0.2 nm and 0.6 nm [61]. According to our simulated distance distributions, the inter-spin label distances are broader than what is suggested in reference [61], implying that distance distributions have a higher contribution from the backbone than from the spin label.

For both R131C and S180C, the modeled distance distributions are broader than the previous MD results. The broader distance distributions could be a result of running longer simulations, allowing the spin label to sample more. However, overall the S180C distance distributions are relatively narrower than the R131C distance distributions. Moreover, the examination of the nitroxide trajectories of all ten independent simulations illustrates that the starting orientation of the spin label is not trivial in these MD simulations, as different orientations result in different average inter-spin label distances (Figure 5-5). This is specifically true for R131C. A more reasonable interpretation for such difference, as mentioned earlier, could be the differences in secondary structures. The loop region is more flexible, making the spin label move all around, whereas the β-strand is more rigid, limiting the spin-label motion.

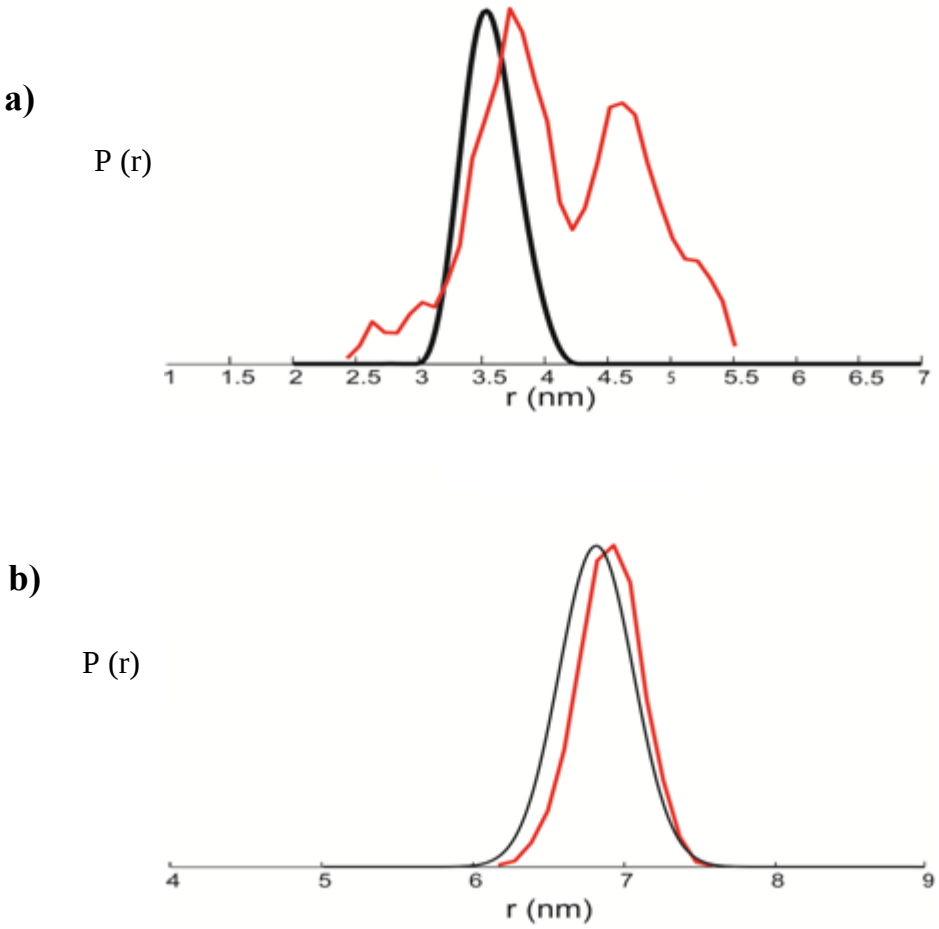


Figure 5-4: The simulated probability–distance distributions overlaid on experiment of a) R131C and b) S180C (**MD** and **Experiment**) [6].

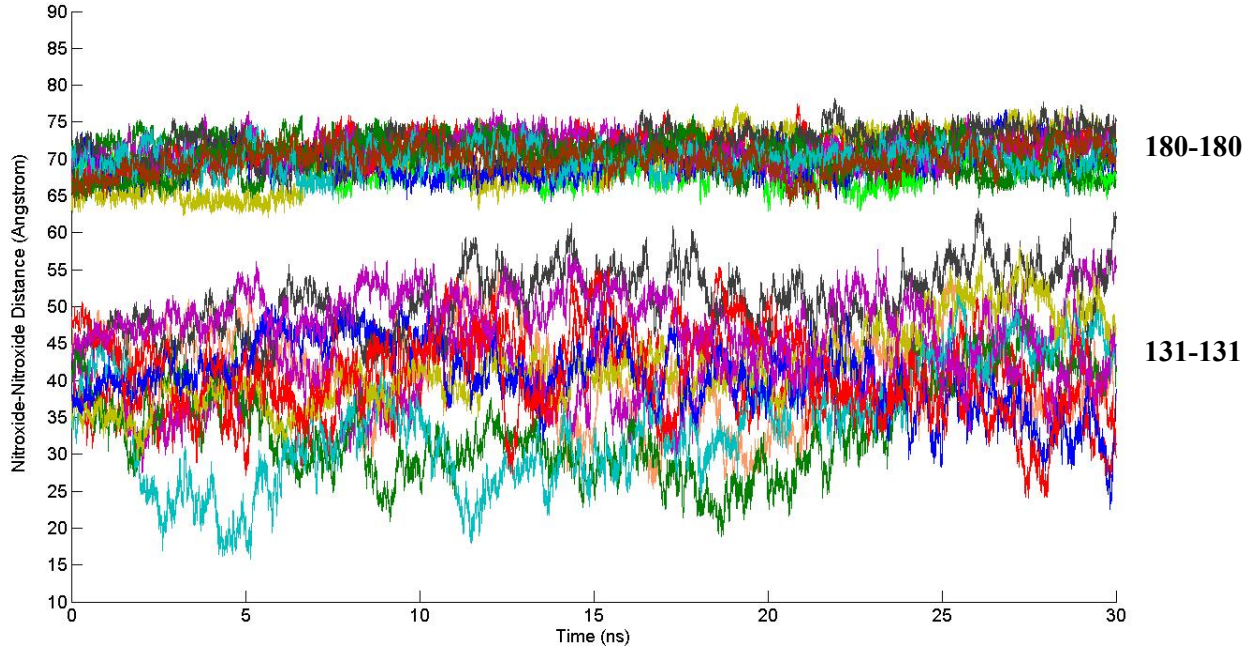


Figure 5-5: Nitroxide–nitroxide distance trajectories for 30 ns of MD simulations performed on specific *Eco*RI–DNA complex crystal structure with spin-labeled sites at 131 and 180. Ten independent simulations were run using different starting orientations of the spin label (Different colors represent the independent simulations performed).

In the case of R131C, the majority of the trajectories sample the experimental distance at a certain time (Figure 5-5). According to Table 1, the spin label at 131 can result in two distinct rotamer families with considerably different inter-spin label distances (31 Å and 45 Å). Hence, different and broader distance distributions of R131C could be a result of the flexibility of the loop region and the sampling of a combination of these different rotamer families in our simulations. On the other hand, the spin label at 180 samples several more probable conformers compared to R131C, but the majority of the sampled conformers have similar inter-spin label distances (Table 1). In addition, the average inter-spin label distance observed in all the independent MD simulations is close to the inter-spin label distance resulted by the majority of the spin-label conformers at 180 (~72 or 73 Å). Thus, the observed narrow and similar distance distributions of S180C could be a result of the rigidity of the β-strand as mentioned above and the high probability sampling of the conformers with 72 or 73 Å average inter-spin label distance.

Even though S180C results are quite satisfactory, the question remains whether the system has reached ergodicity. True ergodicity occurs when the dynamic average of a simulation approaches the thermodynamic average. Such behavior only occurs when the simulation sufficiently samples the conformational space [59]. One way of tracking ergodicity is by observing convergence of the inter-spin label distance distributions. However, in order to reliably track ergodicity, we should observe the convergence of the conformers sampled in MD. Thus, by running longer simulations, we can ensure whether the highest probable conformers sampled for R131C and S180C over the 30 ns time period are retained or reduced to a single conformer. It should be noted that multiple conformers can result in similar average inter-spin label distances that agree with the experimental data. However, multiple conformations may also create

difficulty in finding the most probable corresponding rotamer of the spin label from the experiment. In addition, the highest probable conformation(s) sampled in MD may not be the highest probable conformation from the experiment.

Since one of our goals in the project is to obtain backbone constraints of the arms of *Eco*RI, we generated the inter-atomic distance distributions (C_α–C_α) of the residues to which the spin labels are attached. Figure 5-6 shows the corresponding inter-atomic distance distributions extracted for the two sites. However, our ten independent simulations result in different inter-atomic distance distributions for R131C, whereas they result in similar inter-atomic distance distributions for S180C (data not shown). According to the X-ray crystal structure of the specific *Eco*RI–DNA complex, the inter-atomic distances (C_α–C_α) of R131C and S180C are 29 and 56 Å, respectively. However, our simulations result in average inter-atomic distances that deviate significantly from those obtained from the X-ray crystal structure (Figure 5-6). Also, the time trajectories of C_α–C_α distances of R131C and the RMSD of the inner arm (from the initial crystal structure) seem to fluctuate substantially compared to those of S180C in our simulations (data not shown). The loop flexibility could be a reason for the average inter-atomic distances of R131C (obtained from independent MD runs) to be significantly different from those of the available X-ray crystal structure. It is obvious from Figure 5-6 that the majority of the contribution in the experimental data is coming from the backbone distribution. Therefore, we could safely assume that the DEER distribution data have a higher contribution from the backbone rather than from the spin label itself. In order to confirm the obtained results, the simulated inter-spin label distance distributions should satisfactorily reproduce the experimental data with reasonable convergence.

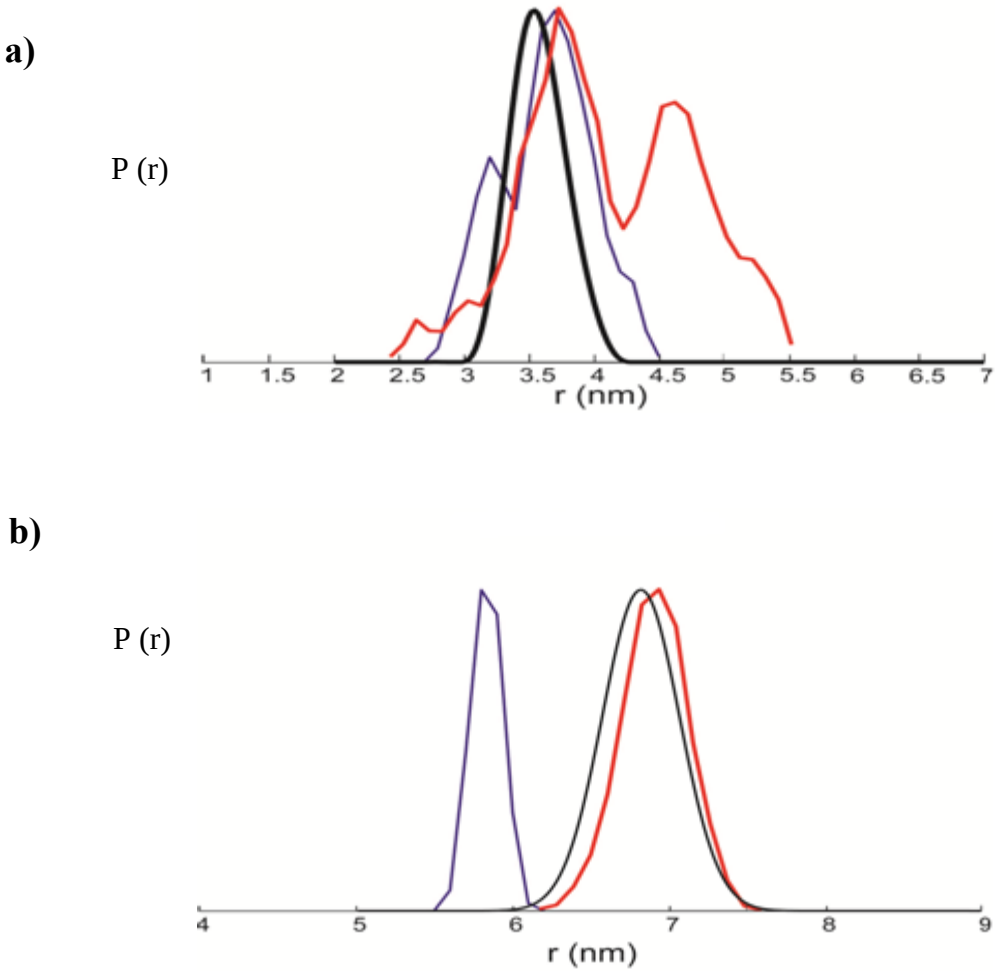


Figure 5-6: The corresponding inter-atomic distance distributions (blue), overlaid on the simulated inter-spin label distance distributions (red) and experimental (black) of a) R131C and b) S180C [6].

5.3 THE COMPARISON OF SPIN LABEL DYNAMICS BETWEEN LOOP OR BETA STRAND AND SOLVENT EXPOSED HELICAL SITE (SEHS)

Based on all ten simulations, the preferred rotamers of the spin label on a loop and a β-strand were obtained and compared to those of an SEHS. According to previous work of Freed and co-workers, the most probable conformers of the spin label for an SEHS are given in Table 2 [16]. Only the first three rotamer families of the spin label residing on a β-strand (from Table 1) are taken into consideration due to their high probability. It is worth reminding at this point that the transition rates between a pair of conformers are determined by the energetic barrier and frictional effects of the dihedral angles in the solvent environment. Based on earlier studies on an SEHS, the rotations around χ_1 and χ_2 angles are slower compared to χ_4 and χ_5 dihedral angles. However, time trajectories of our simulations show that the conformational jumps involved with χ_1 of a spin label on a β-strand and especially χ_2 of a spin label on both loop and β-strand occur at a considerably higher rate compared to an SEHS (data not shown). According to work done by Freed and co-workers, χ_1 of an SEHS exhibits an energy barrier of ~6 kcal/mol for the transition from -60º to 180º, which is significantly high to cross [16]. In our simulations S180C show a minute number of jumps to 180º implying that there can be a similar energy barrier associated with the transition from -60º to 180º for a spin label residing on a β-strand (Figure 5-2). At an SEHS, a high energy barrier is also associated with the transition from -60º to +60º (~7 kcal/mol). According to Figure 5-2, χ_1 of R131C shows a higher number of jumps to -60º (-75º in the case of S180C) than to +60º, further verifying the computational analysis of Freed and co-workers.

Freed and co-workers further show that at an SEHS, χ_2 exhibits an energy barrier of ~6 kcal/mol for the transition from -75º to +75º. Based on Figure 5-2, the number of jumps

to -75º is higher than that of +75º particularly for R131C and such behavior of the spin label can be possibly due to a similar energy barrier demonstrated between -75º to +75º configurations for an SEHS. The variation of χ_3 is similar to what is seen for an SEHS, preferring either +90º or -90º, and the inter-conversion between these two conformations is rarely seen as suggested by the χ_4/χ_5 model. The mobility of χ_4 is smaller for both sites compared to that of an SEHS and seen mostly at 180º. Yet, χ_5 transitions demonstrate a behavior that is similar to that of an SEHS.

Table 2: Comparison of highest probable conformers of the spin label on α-helix [16], β-strand and loop regions.

Site	χ_1	χ_2	χ_3	χ_4	χ_5
α-helix	-60°	180°	+90°	180°	±77°
	-60°	-75°	-90°	180°	±77°
	-60°	180°	+90°	+75°	+8°
β-strand	-75°	180°	±90°	180°	+75°
	180°	180°	±90°	180°	+75°
	+45°	180°	±90°	180°	+75°
Loop	-60°	-75°	±90°	180°	+75°
	-60°	180°	+90°	180°	+75°

6.0 SUMMARY

All-atom MD simulations performed so far offer interesting information on spin-label dynamics and orientations on a loop and a β-strand. The study has the advantage of examining the effects of atomic details in explicit solvent. In our simulations, the restrictions of the rotational motions of the spin-label arise from forces between the spin-label atoms and the neighboring protein residues, as well as solvent molecules. Specifically, our simulations account for the hydrophobic interactions between the bulky methyl groups of the spin label and the neighboring amino acid side chains and steric effects responsible for spin label torsions. Thus, our simulations offer a more realistic picture of the spin label behavior. In the analysis of the dihedral angles, we have found a considerable difference in the highest probable conformations of the spin label on different secondary structures of the protein. Our results on spin-label dynamics do not necessarily comply with the χ_4/χ_5 model proposed for SEHS. As previously stated, there are many conditions that vary between MD simulations performed in modeling the spin label. However, one of the major concerns associated with simulations is the fact that there is no standard force field used by any group. The force field CHARMM22 was developed several years ago from which the bonds, angles, and the dihedral angles of the spin label were modeled according to previous research [62]. This is the force field that our group has used so far in MD simulations. In addition, Roux and co-workers recently developed a force field from *ab-initio*

calculations. This force field was generated to give special attention to the calculation of the energetics of bond rotation in the spin label.

In our simulations χ_1 of S180C and χ_2 of both R131C and S180C show multiple transitions between rotameric states. Roux and co-workers state that the absence of sulfur lone pairs in force fields like CHARMM22 makes it impossible to reproduce the S_δ-H-C_α interaction in MD simulations unlike in *ab-initio* calculations [18]. For this reason, the force field we used may not be able to model this interaction to sufficiently immobilize the first two dihedral angles, χ_1 and χ_2. The force field parameters are presumably overestimating the rotational mobility about χ_1 and χ_2 in our simulations. Roux and co-workers further show that careful parameterization of the force field reproduced the spin-label dynamics on an SEHS calculated from the *ab-initio* method [18]. As a result, a better correlation between χ_1 and χ_2 dihedral angles was obtained. Possibly the force field developed by Roux and co-workers might be better in replicating the experimental results than the CHARMM22 force field used in our simulations. However, the variations could be due to the differences in the constraints of the secondary structures. Even though the χ_4/χ_5 model was developed based on SEHS, the compatibility of the model for a spin label on a loop and a β-strand is still a question.

One of the major concerns in this project is to improve sampling by running longer and independent simulations. Performing simulations with different spin-label orientations has resulted in multiple conformations of the spin label on a loop and β-strand (Table 1). Therefore, by running independent simulations, our results demonstrate an improvement in sampling. Even though the loop region shows more flexibility compared to the β-strand, the different number of rotamers generated by the spin label on a loop is less than that of a β-strand. This could be due to the close proximity of the spin label on the loop to the DNA, thus restricting its mobility.

Simulations after 30 ns on S180C adequately reproduce the experimental data, but no satisfactory agreement is seen for R131C. For different starting orientations of the spin label, R131C generates distance distributions different from each other although S180C distance distributions are similar to one another (data not shown). As mentioned earlier, such differences could result from the dissimilarities between the constraints of the secondary structures. Although we have gathered a significant amount of information to paint a picture on the behavior of the spin label on different secondary structures, it is worth reminding that MD-generated distance distributions only supplement the DEER data.

Furthermore, these simulations can be helpful to study spin-label dynamics and orientations on a loop and a β-strand and can also be used to predict the expected ESR line shape. However, in order to agree with all the results, MD should be run for a longer time period, allowing the simulations to resolve the most favorable conformation(s). To address the question of whether the current MD simulation protocol overestimates the rotational mobility about χ_1 and χ_2, MD simulations are currently being carried out using the force field developed by Roux and co-workers. The new simulations will also examine the behavior of χ_4 to obtain more information on the flexibility of this dihedral angle on different secondary structures. By running simulations using the two different force fields, we will be able to investigate the overall effects variation in parameterization causes on the spin-label dynamics and average structure. Last but not least, our exhaustive simulations will help towards developing an MD procedure in the future that can be applied to different biological systems in order to reliably model the nitroxide spin-label orientations and dynamics so that atomistic details of the system can be well understood.

BIBLIOGRAPHY

1. Dai, X., *Purifacation of Wild Type EcoRI Endonuclease and EcoRI Endonuclease RS187 Crystal Growth of WT EcoRI Endonuclease-DNA Complex and Endonuclease RS187-DNA Complex*. 2007, University of Pittsburgh: Pittsburgh.
2. Hedgpeth, J.G., H. M.; Boyer, H. W., *DNA Nucleotide Sequence Restricted by the RI Endonuclease.* Proc. Natl. Acad. Sci., 1972. **69**: p. 3448-3452.
3. DAVID R. LESSER, M.R.K., LINDA JEN-JACOBSON, *The Energetic Basis of Specificity in the Eco RI Endonuclease-DNA Interaction.* SCIENCE, 1990. **250**.
4. Paul J. Sapienza, 2 John M. Rosenberg,1 and Linda Jen-Jacobson1, *Structural and Thermodynamic Basis for Enhanced DNA Binding by a Promiscuous Mutant EcoRI Endonuclease.* Structure, 2007. **15**: p. 1368–1382.
5. Paul J. Sapienza, C.A.d.T., William H. McCoy IV Samyukta V. Jana and Linda Jen-Jacobson*, *Thermodynamic and Kinetic Basis for the Relaxed DNA Sequence Specificity of "Promiscuous" Mutant EcoRI Endonucleases.* J. Mol. Biol., 2005: p. 1–18.
6. Katherine M. Stone, J.E.T., Jessica Sarver, Paul J. Sapienza, Sunil Saxena, and Linda Jen-Jacobson, *Electron Spin Resonance Shows Common Structural Features for Different Classes of EcoRI–DNA Complexes.* Angew. Chem. Int. Ed., 2008. **47**: p. 10192 –10194.
7. Altenbach, C.F., S. L.; Khorana, H. G.; Hubbell, W. L., , *Structural Studies on Transmembrane Proteins, Spin Labeling of Bacteriorhodopsin Mutants at Unique Cysteines.* Biochemistry, 1989. **28**: p. 7806-7812.
8. Feix JB, K.C., *Site-directed spin labeling of membrane proteins and peptide-membrane interactions. In Spin Labeling: The Next Millenium*. 1998. p. 251-81.
9. Christian Altenbach, S.L.F., H. Gobind Khorana, and Wayne L. Hubbell, *Structural Studies on Transmembrane Proteins. 2. Spin Labeling of Bacteriorhodopsin Mutants at Unique Cysteines.* 1989. **28**: p. 7806-7812.
10. Altenbach, W.L.H.a.C., *Investigation of structure and dynamics in membrane proteins using site-directed spin labeling.* Structural Biology, 1994. **4**: p. 566-573.
11. Hubbell WL, G.A., Langen R, Lietzow M. A *Recent advances in sitedirected spin labeling of proteins.* Curr. Opin. Struct. Biol., 1998. **8**: p. 649–56.
12. Hubbell, L.C.a.W.L., *A new spin on protein dynamics.* Elsevier Science, 2002. **27**: p. 288-295.
13. Ralf Langen, Kyoung Joon Oh, Duilio Cascio, and Wayne L. Hubbell, *Crystal Structures of Spin Labeled T4 Lysozyme Mutants: Implications for the Interpretation of EPR Spectra in Terms of Structure†*. Biochemistry, 2000. **39**: p. 8396-8405.
14. Mchaourab, H.S., Lietzow, M., Hideg, K., and Hubbell, W.L.,, *Motion of spin labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics.* Biochemistry, 1996. **35**: p. 7692-7704.

15. Todd, A.P., Chong, J., Levinthal, F., Levinthal, C., and Hubbell, W.L., *Site-Directed Mutagenesis of Colicin E1 provides specific attachment sites for spin labels whose spectra are sensitive to local conformation.* Proteins, 1989. **6**: p. 294-305.
16. Fabio Tombolato and Alberta Ferrarini, J.H.F., *Dynamics of the Nitroxide Side Chain in Spin-Labeled Proteins.* J. Phys. Chem. B, 2006. **110**: p. 26248-26259.
17. Fabio Tombolato and Alberta Ferrarini, J.H.F., *Modeling the Effects of Structure and Dynamics of the Nitroxide Side Chain on the ESR Spectra of Spin-Labeled Proteins.* J. Phys. Chem. B, 2006. **110**: p. 26260-26271.
18. Deniz Sezer, J.H.F., and Benoît Roux, *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**: p. 5755-5767.
19. Cafiso, G.E.F.a.D.S., *Recent advances and applications of site-directed spin labeling.* Structural Biology, 2006. **16**: p. 644–653.
20. Lazaridisy, M.S.a.T., *Modeling a Spin-Labeled Fusion Peptide in a Membrane: Implications for the Interpretation of EPR Experiments.* Biophysical Journal, 2007. **92**: p. 10–22.
21. Marco Bortolus, F.T., Isabella Tessari, Marco Bisaglia, Stefano Mammi, and A.F. Luigi Bubacco, and Anna Lisa Maniero, *Broken Helix in Vesicle and Micelle-Bound r-Synuclein: Insights from Site-Directed Spin Labeling-EPR Experiments and MD Simulation* J. AM. CHEM. SOC., 2008. **130**: p. 6690–6691.
22. Qi Cai, A.K.K., Wayne L. Hubbell, Ian S. Haworth, Gian Paola C. Gacho, and K.l.n.H. Ned Van Eps, Eric J. Chambers and Peter Z. Qin, *Site-directed spin labeling measurements of nanometer distances in nucleic acids using a sequence-independent nitroxide probe.* Nucleic Acids Research, 2006: p. 1–9.
23. Hang Yu, Y.M., Lars Nordenskiold and Gerhard Stock, *Influence of Nitroxide Spin Labels on RNA Structure: A Molecular Dynamics Simulation Study.* J. Chem. Theory Comput., 2008. **4**: p. 1781–1787.
24. Soraya Pornsuwan, G.B., Christian E. Schafmeister, and Sunil Saxena, *Flexibility and Lengths of Bis-peptide Nanostructures by Electron Spin Resonance.* J. AM. CHEM. SOC., 2006. **128**: p. 3876-3877.
25. Polyhach, G.J.a.Y., *Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance.* Phys. Chem. Chem. Phys., 2007. **9**: p. 1895–1910.
26. Deniz Sezer, J.H.F., and Benoît Roux, *Multifrequency Electron Spin Resonance Spectra of a Spin-Labeled Protein Calculated from Molecular Dynamics Simulations.* J. AM. CHEM. SOC., 2009. **131**: p. 2597–2605.
27. Zhichun Liang, Y.L., Jack H. Freed, Linda Columbus and Wayne L. Hubbell, *A Multifrequency Electron Spin Resonance Study of T4 Lysozyme Dynamics Using the Slowly Relaxing Local Structure Model.* J. Phys. Chem. B, 2004. **108**: p. 17649-17659.
28. David E. Budil, K.L.S., Khaled A. Khairy, and Peter G. Fajer, *Calculating Slow-Motional Electron Paramagnetic Resonance Spectra from Molecular Dynamics Using a Diffusion Operator Approach.* J. Phys. Chem. A, 2006. **110**: p. 3703-3713.
29. Susan C. DeSensi, D.P.R., y Albert H. Beth,z Terry P. Lybrand,* and Eric J. Hustedtz, *Simulation of Nitroxide Electron Paramagnetic Resonance Spectra from Brownian Trajectories and Molecular Dynamics Simulations.* Biophysical Journal, 2008. **94**: p. 3798–3809.

30. Stoica, I., *Force field impact and spin-probe modeling in molecular dynamics simulations of spin-labeled T4 lysozyme.* J Mol Model, 2005. **11**: p. 210–225.
31. Stoica, I., *Using Molecular Dynamics To Simulate Electronic Spin Resonance Spectra of T4 Lysozyme.* J. Phys. Chem. B, 2004. **108**: p. 1771-1782.
32. Ziwei Zhang, M.R.F., Dmitriy S. Tipikin, Zhichun Liang, Jozef K. Moscicki, Keith A. Earle, Wayne L. Hubbell, and Jack H. Freed, *Multifrequency Electron Spin Resonance Study of the Dynamics of Spin Labeled T4 Lysozyme.* J. Phys. Chem. B, 2010.
33. Ken Sale, C.ı.S.a., Kim A. Sharp, K´alm´an Hideg, and Peter G. Fajer,, *Structural Determination of Spin Label Immobilization and Orientation: A Monte Carlo Minimization Approach.* Journal of Magnetic Resonance, 2002. **156**: p. 104–112.
34. Deniz Sezer, J.H.F., and Benoit Roux, *Using Markov Models to Simulate Electron Spin Resonance Spectra from Molecular Dynamics Trajectories.* J. Phys. Chem. B, 2008. **112**: p. 11014–11027.
35. Hubbell, M.A.L.a.W.L., *Motion of Spin Label Side Chains in Cellular Retinol-Binding Protein: Correlation with Structure and Nearest-Neighbor Interactions in an Antiparallel β-Sheet.* Biochemistry, 2004. **43**: p. 3137-3151.
36. Leslie J. Schwartz, A.E.S., and Jack H. Freed, *Analysis of electron spin echoes by spectral representation of the stochastic Liouville equation.* J. Chern. Phys., 1982. **77**: p. 5410–5425.
37. Hubbell, H.-J.S.a.W.L., *Calculation of Electron Paramagnetic Resonance Spectra from Brownian Dynamics Trajectories: Application to Nitroxide Side Chains in Proteins.* Biophysical Journal, 1996. **71**: p. 2201-2212.
38. Deniz Sezer, J.H.F., and Benoît Roux, *Simulating electron spin resonance spectra of nitroxide spin labels from molecular dynamics and stochastic trajectories*
J. Chem. Phys, 2008. **128**: p. 165106.
39. K. Ilker Sen, § Timothy M. Logan, and Piotr G. Fajer, *Protein Dynamics and Monomer-Monomer Interactions in AntR Activation by Electron Paramagnetic Resonance and Double Electron-Electron Resonance†.* Biochemistry, 2007. **46**: p. 11639-11649.
40. Piotr G Fajer, M´at´e Gyimesi, Andras M´aln´asi-Csizmadia, Clive R Bagshaw, K Ilker Sen1 and Likai Song, *Myosin cleft closure by double electron–electron resonance and dipolar EPR.* J. Phys.: Condens. Matter, 2007. **19**: p. 285208(10pp).
41. Adelheid Godt, M.S., Herbert Zimmermann, and Gunnar Jeschke, *How Flexible Are Poly(paraphenyleneethynylene)s?* Angew. Chem. Int. Ed., 2006. **45**: p. 7560 –7564.
42. Ken Sale, § Likai Song, Yi-Shiuan Liu, Eduardo Perozo, and Piotr Fajer, *Explicit Treatment of Spin Labels in Modeling of Distance Constraints from Dipolar EPR and DEER.* J. AM. CHEM. SOC., 2005. **127**: p. 9334-9335.
43. Igor V. Borovykh, S.C., Prasad Gajula, Peter Gast Heinz-Jurgen Steinhoff, Martina Huber, *Distance between a native cofactor and a spin label in the reaction centre of Rhodobacter sphaeroides by a two-frequency pulsed electron paramagnetic resonance method and molecular dynamics simulations.* Journal of Magnetic Resonance, 2006. **180**: p. 178–185.
44. Fajer, P.G., *Determination of spin-label orientation within the myosin head.* Proc. Natl. Acad. Sci., 1994. **91**: p. 937–941.
45. FAJER, P.G., *Determination of spin-label orientation within the myosin head.* Biophysics, 1994. **91**: p. 937-941.
46. Campbell, N.A.R., J. B., ed. sixth edition ed. Biology. 2002.

47. Mathews, V.H.a.A., *Biochemistry*. 3rd ed. 2004.
48. A Grigorescu., M.H., P.A. Wilkosz, K. Chandrasekhar., and J.M. Rosenberg, *In Restriction Endonucleases.* 2004: p. 137-177.
49. Jacobson, L.J., *Protein-DNA Recognition Complexes:Conservation of Structure and Protein-DNA Complexes.* SCIENCE, 1997. **305**: p. 386-389.
50. Heitman, J., *How EcoRI Endonuclease Recognizes and Cleaves DNA.* BioEssays 1992. **14**: p. 445.
51. F. Sanger, A.R.C., G.F. Hong, D.F.Hill, G.B. Petersen, *Nucleotide Sequence of Bacteriophage lambda DNA.* J. Mol. Biol. , 1982. **162**: p. 729-773.
52. F.R., B., *The complete Genome Sequence of Escherichia coli K-12.* Science, 1997. **277**: p. 1453-1462.
53. Rosenberg, J.M., *Structure and function of restriction endonucleases.* Structural Biology, 1991. **1**: p. 104-113.
54. Linda Jen-Jacobson, Lisa E. Engler, J.T.A., Michael R. Kurpiewski, and Arabela Grigorescu, *Thermodynamic Parameters of Specific and Nonspecific Protein-DNA Binding.* SUPRAMOLECULAR CHEMISTRY, 2000. **12**: p. 143-160.
55. McClarin, J.A.F., C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., Rosenberg, J. M., *Structure of the DNA-EcoRI Endonuclease Recognition Complex at 3 Å Resolution.* Science, 1986. **234**: p. 1526 - 1542.
56. Grigorescu, A., *Structural and Energetic Determinants of the Binding Specificity of EcoRI Endonuclease.* , in *Biological Sciences*. 2003, University of Pittsburgh: Pittsburgh.
57. H.Berman, J.W.Z.F., G. Gilliland, T. Bhat, I.S. Weissig and P.Bourne, Nucleic Acids Research, 2000. **28**: p. 235.
58. Ken Sale, L.S., Yi-Shiuan Liu, Eduardo Perozo, and Piotr Fajer, *Explicit Treatment of Spin Labels in Modeling of Distance Constraints from Dipolar EPR and DEER.* J. AM. CHEM. SOC., 2005. **127**: p. 9334-9335.
59. Mikolai I. Fajer, Hongzhi Li, Wei Yang, and Piotr G. Fajer, *Mapping Electron Paramagnetic Resonance Spin Label Conformations by the Simulated Scaling Method.* J. AM. CHEM. SOC., 2007. **129**: p. 13840-13846.
60. Alan Grossfield, S.E.F., and Michael C. Pitman, *Convergence of Molecular Dynamics Simulations of Membrane Proteins.* PROTEINS: Structure, Function, and Bioinformatics, 2007. **67**: p. 31–40.
61. P. P. Borbat, H.S.M.a.J.H.F., J. Am. Chem. Soc., 2002. **124**: p. 5304–5314.
62. Sangmi Jun, J.S.B., Mike Yonkunas, Rob Coalson, and Sunil Saxena, *Unfolding of Alanine-Based Peptides Using Electron Spin Resonance Distance Measurements.* Biochemistry, 2006. **45**: p. 11666-11673.
63. Ken Sale, C.ı.S.a., Kim A. Sharp, K´alm´an Hideg, and Peter G. Fajer, *Structural Determination of Spin Label Immobilization and Orientation: A Monte Carlo Minimization Approach*[1]. Journal of Magnetic Resonance, 2002. **156**: p. 104–112.