

## Simulation of electron paramagnetic resonance spectra of spin-labeled molecules from replica-exchange molecular dynamics

S. Tyrrell and V. S. Oganessian\*

*School of Chemistry, University of East Anglia, Earlham Road, Norwich NR4 7TJ, United Kingdom*

(Received 24 April 2013; published 3 October 2013)

We report a general approach for the simulation of the electron paramagnetic resonance (EPR) spectra of spin labels attached to peptides and proteins directly from replica-exchange molecular dynamics (REMD) trajectories. Conventional MD trajectories are generally inadequate for the prediction of EPR line shapes since the label can become trapped in one or more of a set of rotameric states, thus preventing both conformational sampling and accurate estimates of the exchange rates between different rotamers. The advantage of using REMD is its ability to provide both efficient conformational sampling and kinetic information for spin-label dynamics. Our approach is illustrated with spin-labeled peptide. This approach to REMD-EPR simulation paves the way for the wider application of MD modeling to the simulation and interpretation of EPR spectra of spin-labeled molecules.

DOI: [10.1103/PhysRevE.88.042701](https://doi.org/10.1103/PhysRevE.88.042701)

PACS number(s): 87.15.M–, 83.10.Mj, 87.15.A–

Recent advances in electron paramagnetic resonance (EPR) instrumentation have made the method of site-directed spin labeling (SDSL) with nitroxide spin probes a valuable and widely used tool in the study of protein global and local motions and conformational changes as well as nanometer-range distance measurements using both continuous-wave (CW) and pulsed EPR (DEER/PELDOR) [1–8]. The most widely used spin label (SL) to date, methanethiosulfonate SL (MTSL) contains a nitroxide radical linked to a protein or peptide cysteine residue by a disulfide bond forming a side chain, usually denoted R1 (Fig. 1) [1–7,9–11]. EPR spectral line shapes contain information about both label and protein dynamics. The internal motional flexibility of the tether of R1, arising from the five dihedral angles connecting R1 to the protein (peptide) backbone, 1–5 in Fig. 1, comprises a total of 108 ( $3 \times 3 \times 2 \times 3 \times 2$ ) rotameric states, which have been shown to be crucial for the line shapes [1–7,9,10]. However, the interpretation of such EPR spectra is not trivial and requires full computer modeling [6,12–19]. It is desirable to be able to predict EPR spectra from the results of molecular-dynamics (MD) simulations of actual structures at the atomistic scale. A common approach is to combine the motions of the SL from MD trajectories with the protein rotation diffusional dynamics modeled by Brownian dynamics (BD) trajectories [15,17]. In the past 20 years, substantial progress has been made in this direction. Simulations of SL contributions to EPR from MD have been attempted by using its probability distribution generated from MD in order to approximate the ordering potential for subsequent calculation using the BD method [6,18]. In a different approach, single MD trajectories have been employed for the direct simulation of EPR [17,19–21]. EPR spectra have also been simulated using multiple MD trajectories for the construction of a discrete stochastic Markov chain model for the rotameric dynamics of R1 [15]. However, while librational motions within each rotameric state of R1 are well captured in conventional MD trajectories, this is not the case for the slow exchange dynamics among rotamers. Regardless of the number of MD trajectories employed,

the SL can become trapped in one of the rotameric states, thereby preventing both efficient conformational sampling and estimation of the exchange rates among different states, thus resulting in an inadequate EPR spectra simulation.

Parallel tempering simulations in the form of replica-exchange MD (REMD) of an extended ensemble of states provide an elegant and powerful way to enhance sampling of states which cannot be sampled efficiently otherwise [22–24]. Equally importantly, it has been shown recently that REMD is not only a powerful method to enhance conformational sampling but also can provide useful kinetic information [25–28]. For example, using REMD-based methods, accurate kinetic information for protein folding and unfolding has been successfully extracted [25,27,28]. In particular, it has been shown that slow transition rates with the time constants of the order of microseconds can be obtained accurately from relatively short discontinuous REMD trajectories with exchanges as short as 5 ps [27].

Here we present a general approach for the prediction of EPR motional spectra of SL in biomolecules from REMD trajectories. Using the intervals between exchanges in REMD, we calculate the frequencies of transitions between dihedral rotameric states. We then construct a coarse master equation describing the population of each of the states for SL rotameric dynamics by projecting the dynamical trajectory (DT) onto the 108 coarse-grained states [26]:

$$\frac{dP_i(t)}{dt} = \sum K_{ij} P_j(t) - K_{ji} P_i(t), \quad (1)$$

where  $P_i(t)$  is the probability for being in state  $i$  at time  $t$ , and  $K_{ij}$  is the  $108 \times 108$  rate matrix. The formal solution of Eq. (1) can be presented in a matrix form using a matrix exponential, namely  $\mathbf{P}(t) = \exp(t\mathbf{K})\mathbf{P}(0)$ , where  $\exp(t\mathbf{K})$  is a propagator of the system and for a given lag time  $t$  is defined as the transition probability matrix (TPM)  $\|\exp(t\mathbf{K})\|_{ij}$  [26]. TPM can be estimated from a DT using the branching probability  $S_{ij}$  of changing the state from  $i$  to  $j$ :

$$S_{ij} = \frac{N_{ij}}{\sum_k N_{kj}}, \quad (2)$$

\*v.oganesyan@uea.ac.uk

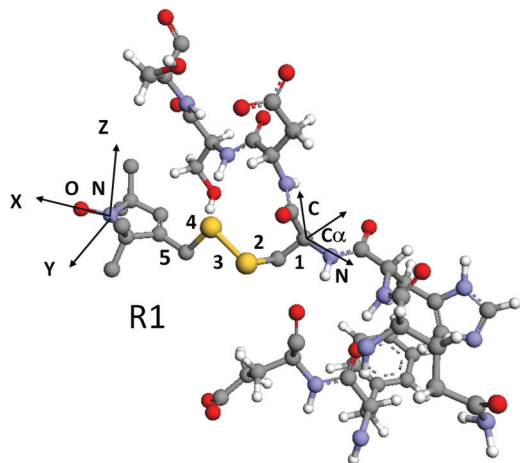


FIG. 1. (Color online) Structure of R1 SL attached to the Cys thiol side chain. Dihedral angles of the tether are numbered. Number 3 corresponds to the -S-S- dihedral angle. Both the magnetic axes of the nitroxide head group of SL and the peptide axes are indicated by arrows, where  $x$  and  $z$  axes of the magnetic frame lie along the N-O direction and perpendicular to the nitroxide plane, respectively.

where  $N_{ij}$  represents the number of transitions in an REMD trajectory that are sampled at intervals  $t$  [15,29] and are calculated from the uninterrupted short segments of the REMD trajectory. The detailed balance is enforced by symmetrizing the transition matrix:  $N_{ij}^{\text{sym}} = \frac{1}{2}(N_{ij} + N_{ji})$ . TPM is then employed in the stochastic kinetic Monte Carlo (KMC) algorithm for generating the time evolution of the population in each of the respective states [30,31].

Our approach is illustrated with both MD and REMD simulations on a model system, a spin-labeled peptide consisting of 16 amino acids LYS-GLU-THR-ALA-LYS-PHE-GLU-ARG-GLN-HIS-CYS-ASP-SER-THR-SER-ALA with the SL attached to thiol side chain of the CYS residue. We have used the GROMACS 4.5.5 package to simulate both MD and REMD trajectories [32] for the spin-labeled peptide in water using the GROMOS96 force field [33] and using the force-field parameters previously developed by us for MTSL SL [19]. To prevent the peptide from unfolding in water, terminal amino acids of both ends were restrained to their initial positions throughout the simulations. The REMD simulation used 35 replicas between 300 and 500 K with temperatures selected in accordance with the previously published approach [34]. The simulation was run to produce a 40 ns trajectory saved at 1 ps intervals with exchanges being attempted every 6 ps. Additionally, several MD simulations 100 ns in length with the temperature set to 300 K with different starting dihedral conformations have been carried out, and a long MD trajectory of 1600 ns was also generated. Figures 2(a) and 2(b) show convergence curves for both long MD and 40 ns REMD simulations, respectively. One can see that MD convergence is achieved at  $\sim 1000$  ns with the maximal deviation of  $\sim 4\%$  of the total population among the states. Because of more efficient sampling, the convergence in REMD is achieved at much shorter times (25 ns) with a maximal deviation of  $\sim 2\%$ .

Our focus is on the dihedral dynamics of R1, and for this reason the dynamics of the probe is transformed from the laboratory frame to the peptide backbone fixed frame [35] defined by N, C, and  $C\alpha$  atoms shown in Fig. 1.

Dihedral flips are essentially described by discrete Markov chain stochastic dynamics [15]. However, short-time dynamics

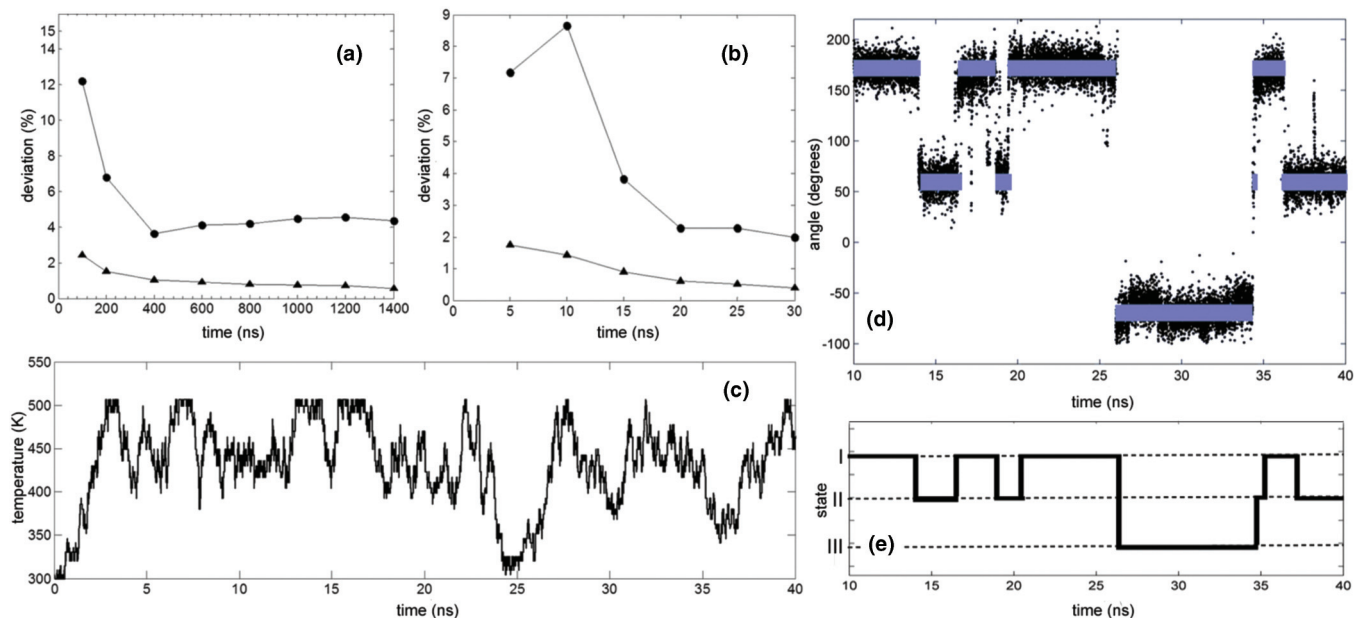


FIG. 2. (Color online) (a) and (b) Convergence of the simulations represented by both the maximal deviation [ $\max_i |P_i^{\text{eq}}(t) - P_i^{\text{eq}}(t_{\text{max}})|$ ] (circles) and mean-square deviation [ $\sqrt{\sum_i [P_i^{\text{eq}}(t) - P_i^{\text{eq}}(t_{\text{max}})]^2 / 108}$ ] (triangles) among the equilibrium populations of states for MD and REMD runs, respectively; (c) temperatures sampled by one of the replicas during REMD; (d) illustration of transition path based assignment of rotameric states using dihedral angle 1 as an example with the assignment of its coarse-grained states I, II, and III corresponding to the mean values of approximately  $180^\circ$ ,  $+60^\circ$ , and  $-60^\circ$ , respectively, shown in (e).

in MD trajectories are not necessarily Markovian. As a result, the time series can contain many spurious transitions back and forth between states  $i$  and  $j$  before a real transition occurs, leading to the overestimation of the rates of conformational exchanges [15]. To suppress non-Markovian kinetics, we have adapted a transition-based assignment (TBA) method [27,29]. According to this approach, the states are assigned with the help of transition paths which connect well-defined regions, and the transition takes place only when the trajectory crosses from one band to another. This is illustrated in Figs. 2(d) and 2(e) with the assignment of coarse-grained states for dihedral 1 with the bands shown in blue (light gray). Hence fast fluctuations do not contribute to a state change. Previously it has been demonstrated by Buchete and Hummer that both rate matrices and slow relaxation times obtained from REMD runs using TBA are independent of lag times as short as 1 ps [27,29]. In all our simulations, the stringent condition of  $15^\circ$  for the dihedral widths, as reported for peptides [29], was used.

The following procedure was used for generating extended DTs from REMD trajectories. First, for each dihedral angle initial values were assigned based on the ideal reference angles of each rotamer, namely  $60^\circ$ ,  $-60^\circ$ , and  $180^\circ$  for a multiplicity of 3 and  $90^\circ$  and  $-90^\circ$  for a multiplicity 2. Then the coarse graining of the states was refined by performing K-means clustering for the values of all five dihedral angles of the SL calculated from REMD [30,36]. Second, for the entire trajectory, the TBA of states was performed. All the frames with the same rotameric state were stored together in homogeneous pools, one for each state. Finally, after constructing the TPM, the KMC algorithm has been applied to produce long DTs (4000 ns) by generating stochastic dynamical exchanges among the states with frames being selected by iterating over the appropriate pools.

The line shapes of the EPR spectra of nitroxide SLs are determined entirely by the variation in time of the directions of magnetic axes of the nitroxide head group. These are the principal axes of the  $\mathbf{g}$  and  $\mathbf{A}$  tensors which define the Zeeman interaction and the  $^{14}\text{N}$  nuclear hyperfine coupling to the electron spin, respectively. Both tensors are anisotropic, leading to a strong dependence of the EPR resonances on the

direction of the applied magnetic field relative to the principal magnetic axes. EPR spectra were calculated from single DTs using a previously published procedure [19]. In brief, the simulation algorithm is based on the numerical propagation of the full spin density matrix in order to generate the time evolution of the statistically averaged transverse magnetization of the electron spin [6,12]. The Fourier-Laplace transform of the magnetisation trajectory provides the contribution to the EPR spectrum. The total line shape intensity must be averaged over all possible, evenly distributed, orientations of the molecule in the sample.

In the next sections, we compare the results obtained from a 100 ns MD (a typical length employed in previous studies [10, 15,17,19,20]), long 1600 ns MD, and 40 ns REMD trajectories. A converged 1600 ns MD trajectory which sampled all the 108 rotameric states of SL was used as a standard in the analysis of REMD-EPR simulation results. Both long MD and short REMD trajectories indicate that among a total of 108 rotameric states, 24 have negligible populations of  $<0.15\%$ . On the other hand, a 100 ns trajectory provides much poorer sampling of the states with only 42 states sampled. Among them, 31 states have populations of less than 0.15% each.

Figure 3 compares the simulated autocorrelation functions of rotational dynamics of the  $z$ -magnetic axis of the SL and the EPR spectra predicted at two resonance frequencies (bands), respectively, for all three cases. One can see that there is a clear disagreement between the data simulated using 100 and 1600 ns MD trajectories shown as black and red lines, respectively, in Figs. 3(a)–3(c). In particular, EPR line shapes corresponding to 100 ns MD run are substantially broader in both X- and W-bands than in the case of a 1600 ns trajectory, indicating a higher degree of motional restraint of the SL during the former run. It is important to note that 100 ns MD trajectories generated from different starting configurations resulted in significantly different EPR line shapes (not shown here).

To test our REMD-EPR simulation approach, we first generated a long 4000 ns DT from the TPM obtained from the original 1600 ns MD trajectory which can be viewed as a special case of a continuous REMD trajectory with

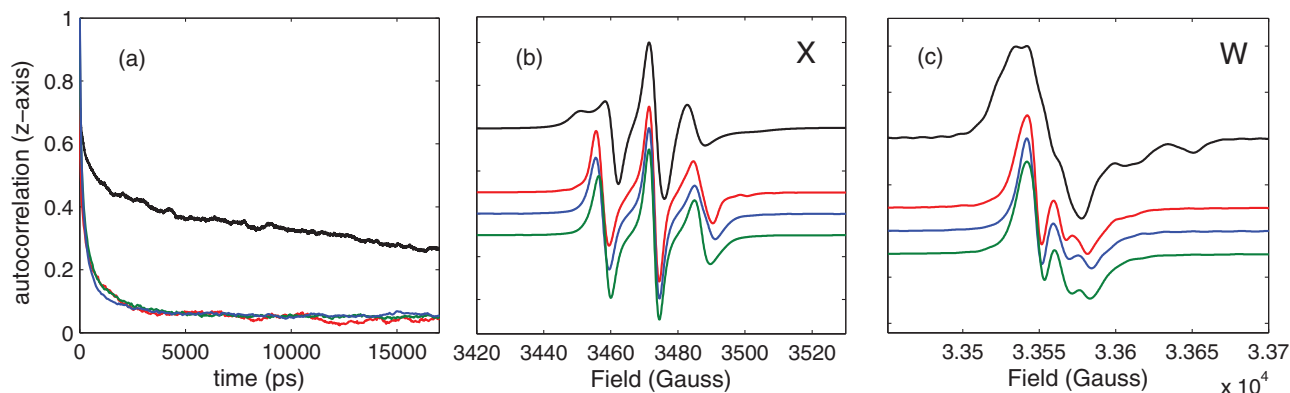


FIG. 3. (Color) (a) Rotational autocorrelation functions for the  $z$  axis of the nitroxide frame of the SL; (b) and (c) show EPR spectra simulated for X- (9.5 GHz) and W-band (94 GHz) frequencies, respectively. Black, red, blue, and green lines correspond to the correlation functions and spectra calculated using 100 ns MD, 1600 ns MD, MD-KMC, and REMD-KMC trajectories, respectively. EPR spectra are normalized to their maximum intensities.

no exchange boundaries. Both the correlation function and EPR spectra calculated from this MD-KMC DT are shown as blue lines in Fig. 3 and are in excellent agreement with the ones calculated directly from the 1600 ns MD trajectory.

Finally, the TPM has been calculated from the 40 ns REMD trajectory concatenated for 300 K. In the REMD run, there were no real exchanges recorded which connect any two states with different -S-S- angles. As a result, the TPM has a block diagonal form with two subblocks mixing only the states corresponding to different values of a -S-S- dihedral angle. It has been observed previously that a very low exchange rate between the two conformations of the disulfide torsion angle is an intrinsic property of R1 and is due to a high potential barrier to rotation [15,19]. This is confirmed by the results from the long 1600 ns MD run indicating only four changes for this angle. Thus two independent subtrajectories of appropriate lengths proportional to the relevant total populations of the two -S-S- states were generated from the corresponding subblocks and concatenated into a single 4000 ns REMD-KMC DT. The resulting single discontinuity in REMD-KMC DT at  $\sim 2000$  ns is too slow a dynamical change on the EPR time scale to have any effect on the line shape. The resulting correlation function and predicted EPR spectra of both bands, shown as green lines in Fig. 3, are in agreement with the ones calculated directly from the 1600 ns MD trajectory. EPR spectra simulated from both 1600 ns MD and REMD correspond to characteristic line shapes reported in the literature for the solvent exposed sites of the spin-label attachment [1].

It is instructive to look at how adequately the equilibrium populations of conformational states are represented in all three cases considered, namely 100 ns MD, 1600 ns MD, and 40 ns REMD trajectories. For solvent exposed or partially exposed sites of SL attachments, rotational flips of dihedrals 4 and 5 occur on the subnanosecond to few nanosecond time scale and usually are well captured in a single 100 ns MD trajectory [15,19]. In contrast, the dynamics of dihedrals 1–3 are generally too slow to be adequately represented by 100 ns trajectories, with dihedral 3 rarely seen on such a time scale [15,19]. Figure 4 shows equilibrium populations  $P_{eq}$  of 18 rotameric states ( $3 \times 3 \times 2$ ) corresponding to dihedral angles 1–3 [Figs. 4(a), 4(b), and 4(c) show results for 100 ns MD, 1600 ns MD, and 40 ns REMD runs, respectively]. It also presents schematic representations of sampled SL conformations corresponding to these states averaged over librational motions and the motions of the remaining dihedrals 4 and 5. The results clearly show that the 100 ns trajectory does not sample well accessible states, with only eight of them seen compared to all the states visited in the long MD run. In the 100 ns run among eight sampled states, two make the dominant contribution for 80% of the sampling time. In contrast, a relatively short 40 ns REMD run is able to adequately reproduce the states including the ability to sample both rotameric subspaces associated with the -S-S- dihedral angle. Small differences observed between some corresponding populations of rotameric states are within the accuracy of convergence of the simulations. Importantly, the label in our model is solvent exposed with broad distribution among rotameric states resulting in the low value of the order parameter ( $\sim 0.05$ ). As a result, small differences in population

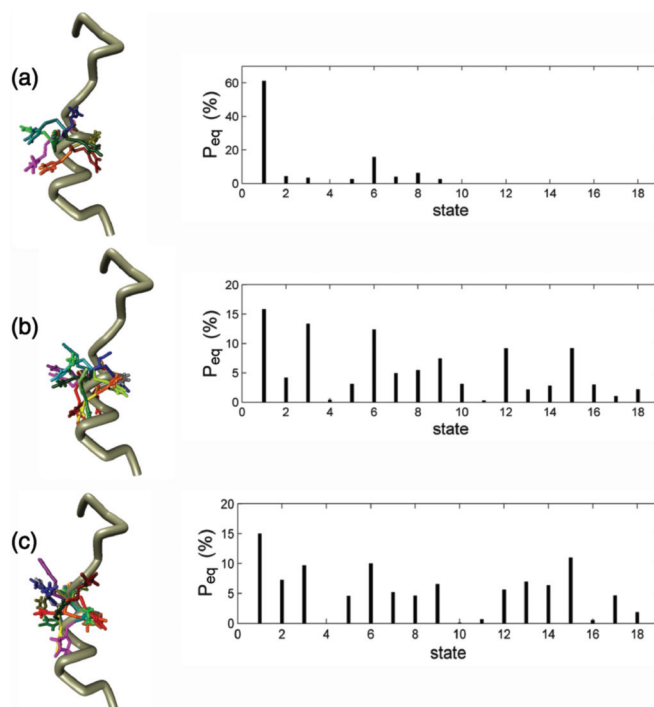


FIG. 4. (Color online) Parts (a), (b), and (c) show conformations of 18 SL states arising from the dihedral angles 1–3 generated from 100 ns MD, 1600 ns MD, and 40 ns REMD runs, respectively. Insets on the right show equilibrium populations of these states calculated from the corresponding runs.

among multiple rotameric states have a negligible effect on the simulated EPR spectra. After coarse graining of the states by TBA in the REMD trajectory, one can estimate the averaged lifetime of a particular state  $i$  from the mean duration of all visits to this state. The calculated lifetimes show good agreement between the 1600 ns MD and REMD runs with the lifetimes averaged among all 18 states to be 1025 and 990 ps, respectively. This is different from the 100 ns trajectory, which has a mean lifetime of 640 ps among the sampled states.

In conclusion, we have reported an approach for predicting the EPR spectral line shapes arising from the motions of the nitroxide SL using REMD. Our approach is general and is applicable to any SL. The advantage of using relatively short REMD runs [the 40 ns REMD run with parallel replicas was  $>50$  times faster compared to the sequential 1600 ns MD one] for simulating EPR spectra of SL motions opens the prospect of applying the REMD-EPR method to a wide range of molecular systems. Since multiple replicas with different temperatures are used in REMD, prediction of variable-temperature EPR spectra from a single REMD run is also possible. The REMD-EPR approach can also be adopted for an accurate calculation of the rotameric distribution of SLs, which is important for the analysis of distance distribution measurements by pulsed EPR.

V.S.O. thanks the Leverhulme Trust (F/00204/AT) and the EPSRC (GR/H020411/1) for financial support.

- [1] L. J. Berliner, *Spin Labelling: The Next Millennium* (Plenum, New York, 1998).
- [2] W. L. Hubbell, D. S. Cafiso, and C. Altenbach, *Nat. Struct. Biol.* **7**, 735 (2000).
- [3] Z. F. Guo, D. Cascio, K. Hideg, and W. L. Hubbell, *Protein Sci.* **17**, 228 (2008).
- [4] P. P. Borbat, A. J. Costa-Filho, K. A. Earle, J. K. Moscicki, and J. H. Freed, *Science* **291**, 266 (2001).
- [5] G. Jeschke, *ChemPhysChem* **3**, 927 (2002).
- [6] H.-J. Steinhoff and W. Hubbell, *Biophys. J.* **71**, 2201 (1996).
- [7] G. E. Fanucci and D. S. Cafiso, *Curr. Opin. Struct. Biol.* **16**, 644 (2006).
- [8] Y. E. Nsmelov, C. B. Karim, L. Song, P. G. Fajer, and D. D. Thomas, *Biophys. J.* **93**, 2805 (2007).
- [9] D. Sezer, J. H. Freed, and B. Roux, *J. Am. Chem. Soc.* **131**, 2597 (2009).
- [10] C. Beier and H.-J. Steinhoff, *Biophys. J.* **91**, 2647 (2006).
- [11] E. Stendardo, A. Pedone, P. Cimino, M. C. Menziani, O. Crescenzi, and V. Barone, *Phys. Chem. Chem. Phys.* **12**, 11697 (2010).
- [12] B. H. Robinson, L. J. Slutsky, and F. P. Auteri, *J. Chem. Phys.* **96**, 2609 (1992).
- [13] D. E. Budil, S. Lee, S. Saxena, and J. H. Freed, *J. Magn. Reson., Ser. A* **120**, 155 (1996).
- [14] P. Hakansson, P.-O. Westlund, E. Lindahl, and O. Edholm, *Phys. Chem. Chem. Phys.* **3**, 5311 (2001).
- [15] D. Sezer, J. H. Freed, and B. Roux, *J. Phys. Chem. B* **112**, 11014 (2008).
- [16] V. S. Oganessian, *J. Magn. Reson.* **188**, 196 (2007).
- [17] S. C. DeSensi, D. P. Rangel, A. H. Beth, T. P. Lybrand, and E. J. Hustedt, *Biophys. J.* **94**, 3798 (2008).
- [18] D. E. Budil, K. L. Sale, K. A. Khairy, and P. G. Fajer, *J. Phys. Chem. A* **110**, 3703 (2006).
- [19] V. S. Oganessian, *Phys. Chem. Chem. Phys.* **13**, 4724 (2011).
- [20] E. Kuprusevicius, G. F. White, and V. S. Oganessian, *Faraday Discuss.* **148**, 283 (2011).
- [21] V. S. Oganessian, E. Kuprusevicius, H. Gopee, A. N. Cammidge, and M. R. Wilson, *Phys. Rev. Lett.* **102**, 013005 (2009).
- [22] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- [23] D. Paschek and A. E. Garcia, *Phys. Rev. Lett.* **93**, 238105 (2004).
- [24] M. Andrec, A. K. Felts, E. Gallicchio, and R. M. Levy, *Proc. Natl. Acad. Sci. USA* **102**, 6801 (2005).
- [25] D. van der Spoel and M. M. Seibert, *Phys. Rev. Lett.* **96**, 238102 (2006).
- [26] S. Sriraman, I. G. Kevrekidis, and G. Hummer, *J. Phys. Chem. B* **109**, 6479 (2005).
- [27] N.-V. Buchete and G. Hummer, *Phys. Rev. E* **77**, 030902 (2008).
- [28] S. C. Yang, J. N. Onuchic, A. E. Garcia, and H. Levine, *J. Mol. Biol.* **372**, 756 (2007).
- [29] N.-V. Buchete and G. Hummer, *J. Phys. Chem. B* **112**, 6057 (2008).
- [30] A. B. Bortz, M. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- [31] S. A. Serebrinsky, *Phys. Rev. E* **83**, 037701 (2011).
- [32] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- [33] E. Lindahl, B. Hess, and D. van der Spoel, *Phys. Chem. Chem. Phys.* **7**, 306 (2001).
- [34] A. Patriksson and D. van der Spoel, *Phys. Chem. Chem. Phys.* **10**, 2073 (2008).
- [35] B. K. P. Horn, *J. Opt. Soc. Am.* **4**, 629 (1987).
- [36] J. A. Hartigan, *Clustering Algorithms* (Wiley, New York, 1975).