Long-time mean-square displacements in proteins

Derya Vural,[1,*] Liang Hong,[2,3] Jeremy C. Smith,[2,3] and Henry R. Glyde[1]
[1]*Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716-2570, USA*
[2]*Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA*
[3]*Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996, USA*
(Received 28 July 2013; published 8 November 2013)

We propose a method for obtaining the intrinsic, long-time mean square displacement (MSD) of atoms and molecules in proteins from finite-time molecular dynamics (MD) simulations. Typical data from simulations are limited to times of 1 to 10 ns, and over this time period the calculated MSD continues to increase without a clear limiting value. The proposed method consists of fitting a model to MD simulation-derived values of the incoherent intermediate neutron scattering function, $I_{\rm inc}(\mathbf{Q},t)$, for finite times. The infinite-time MSD, $\langle r^2\rangle$, appears as a parameter in the model and is determined by fits of the model to the finite-time $I_{\rm inc}(\mathbf{Q},t)$. Specifically, the $\langle r^2\rangle$ is defined in the usual way in terms of the Debye-Waller factor as $I(\mathbf{Q},t=\infty)=\exp(-Q^2\langle r^2\rangle/3)$. The method is illustrated by obtaining the intrinsic MSD $\langle r^2\rangle$ of hydrated lysozyme powder ($h=0.4$ g water/g protein) over a wide temperature range. The intrinsic $\langle r^2\rangle$ obtained from data out to 1 and to 10 ns is found to be the same. The intrinsic $\langle r^2\rangle$ is approximately twice the value of the MSD that is reached in simulations after times of 1 ns which correspond to those observed using neutron instruments that have an energy resolution width of 1 μeV.

DOI: 10.1103/PhysRevE.88.052706 PACS number(s): 87.14.E−, 87.15.hg, 61.05.F−

I. INTRODUCTION

The mean-square displacement (MSD) of an atom is a fundamental dynamical quantity. In proteins the temperature dependence of the average MSD has been widely used to characterize the internal flexibility of the protein [1–11]. The MSD can be extracted from dynamic neutron scattering experiments. Since the incoherent neutron scattering cross section of the hydrogen (H) nucleus is large, the observed MSD is dominated by the MSD of H in the protein. As a result, the MSD of H in proteins has been extensively investigated by neutron scattering [1–5,12] and the results compared with molecular dynamics (MD) simulation, a technique particularly complementary to dynamic neutron scattering [13]. These studies have been performed as a function of temperature, pressure, and hydration and in a variety of solvents [14].

Most neutron and MD studies to date have investigated the MSD on a picosecond–nanosecond (ps-ns) time scale [1,5,15–28]. At low temperatures ($T<100$ K), a protein is essentially harmonic. As the temperature is increased beyond the harmonic regime, proline puckering transitions and methyl rotations are activated [29]. These onsets are independent of protein hydration. At temperatures $T\simeq 160$–220 K, jumps of nonexchangeable H in the nonproline methylene groups and aromatic phenyl rings dominate the MSD on the time scale of 10^{-9} seconds. For $T\leqslant 220$ K protein flexibility arises chiefly from the hydrophobic and aromatic residues. In contrast, motion of the hydrophilic residues remains suppressed due to stable H-bonding interactions with the neighboring protein residues and hydration water. As T is further increased, at $T_D\sim 180$–220 K, a strongly hydration-dependent increase is found in the localized diffusion of protein nonexchangeable H atoms and in jumps in the hydrophilic side chains. The resulting increase in MSD is denoted the dynamical transition (DT). The jumps in hydrophilic side chains are strongly coupled to the relaxation rates of the H bonds formed with hydration water [29].

MD has also been used to probe the pressure dependence of protein MSDs [24,25,30,31]. These MD studies revealed a qualitative change in the internal protein motions at $p\sim 4$ kb and the existence of two linear regimes in the MSD. The qualitative change is a loss with increasing pressure of high-amplitude, collective protein modes below 2 THz in effective frequency, accompanied by restriction of large-scale solvent translational motion [30]. The DT was found to be pressure independent, indicating that the effective energy barriers separating conformational substates are not significantly influenced by pressure. In contrast, vibrations within substates stiffen with pressure, due to increased curvature of the local harmonic potential in which the atoms vibrate [31].

Given the extensive interest in MSDs, both observed and simulated, it is useful to clarify in detail what is measured and calculated and the possible relation to equilibrium thermodynamics. A global incoherent dynamic structure factor, $S(\mathbf{Q},\omega)$, dominated by H, is observed in neutron scattering measurements. A first consideration is that the H atoms in a protein occupy a spectrum of sites and the fluctuation of H in these sites follows a wide distribution, which can be modeled using a Weibull form [32]. This heterogeneity in the distribution of MSDs leads to dynamic structure factors, $S(\mathbf{Q},\omega)$, that deviate from Gaussian behavior in the scattering wave vector Q. The heterogeneity introduces a correction to fourth order in Q that can be used to extract the variance of the distribution of MSDs [33,34].

Similarly, the global MSD obtained from neutron scattering measurements depends on the energy resolution employed. The MSD is obtained from the elastic ($\omega=0$) component of the resolution-broadened dynamic structure factor, $S_R(\mathbf{Q},\omega=0)$, as

$$\langle r^2\rangle_R=-3\frac{d\ln S_R(\mathbf{Q},\omega=0)}{dQ^2}. \quad (1)$$

*deryavur@gmail.com

The $\langle r^2\rangle_R$ extracted in this way is the MSD after it has had time to develop over a limited time, $0 < t < \tau_R$, only. The time τ_R is set by the width of the energy resolution of the neutron instrument, $\tau_R \simeq \hbar/W$, where W is the full width at half-maximum (FWHM) of the resolution function. Typical instrument energy resolutions lie in the range $100\,\mu\text{eV} > W > 1\,\mu\text{eV}$, which corresponds to evolution times $10\text{ ps} < \tau_R < 1\text{ ns}$. Over this time range the extracted $\langle r^2\rangle_R$ is still increasing with decreasing W (increasing τ_R), indicating that the intrinsic, long-time ($\tau_R \to \infty$) value of $\langle r^2\rangle$ has not been observed [12,17,18,35].

In an earlier paper [36], we proposed a method to extract the intrinsic, long-time MSD $\langle r^2\rangle$ from resolution-dependent data. In the method, the intrinsic $\langle r^2\rangle$ was defined in terms of the $t = \infty$ limit of the incoherent intermediate scattering function (ISF), $I(\mathbf{Q},t)$, as

$$I_\infty = I(\mathbf{Q},t=\infty) = \exp\left(-\tfrac{1}{3}Q^2\langle r^2\rangle\right). \tag{2}$$

The method consists of constructing a model $I(\mathbf{Q},t)$ that includes I_∞ and $\langle r^2\rangle$, calculating the corresponding $S_R(\mathbf{Q},\omega = 0)$ including the resolution width, W, and fitting the model to the observed $S_R(\mathbf{Q},\omega = 0)$. The intrinsic $\langle r^2\rangle$ is obtained from the fit as a fitting parameter. In this way an intrinsic, long time $\langle r^2\rangle$ was obtained from resolution dependent data.

The incoherent ISF, $I_{\text{inc}}(\mathbf{Q},t)$, that is observed in neutron measurements can also be calculated directly from simulations [13]. A comparison between the simulation-derived and the observed MSD can be made by calculating the simulation-derived MSD in exactly the same way that it is obtained from experiments [6,21,35,37]. That is, from the simulated $I_{\text{inc}}(\mathbf{Q},t)$, the incoherent dynamic structure factor $S(\mathbf{Q},\omega)$ is calculated including the instrument resolution width, W. The MSD is obtained from the slope of the calculated resolution-broadened $S_R(\mathbf{Q},\omega)$ at small Q using Eq. (1). This effectively compares the MSD after it has evolved for a specific time τ set by W. In this way excellent agreement between simulated and observed MSDs has been achieved. This also opens the question: Could the simulated intrinsic, long-time $\langle r^2\rangle$ be obtained by fitting a model to finite-time values of a calculated $I_{\text{inc}}(\mathbf{Q},t)$?

To explicitly recognize the time dependence of $\langle r^2\rangle_R$, the concept of neutron time windows was introduced, a concept in which the effects arising from the finite energy resolution are fully integrated [33,38]. The concept has also been incorporated into the formalism that describes the dynamics accompanying the glass transition in molecular systems. When the protein relaxation time decreases with temperature, as it usually does, it has been shown that measurement of $\langle r^2\rangle_R$ over a finite time window can introduce an apparent DT when there is no actual change in the elastic incoherent dynamic structure factor, $S(\mathbf{Q},\omega = 0)$ [33,38]. To avoid this issue, identification and use of an intrinsic, long-time MSD $\langle r^2\rangle$ would be helpful.

Several MD studies have been performed aimed at understanding the origin of elastic neutron scattering from proteins [20–22,29,32–34,38]. These studies have characterized in detail the contributions of time dependence, of non-Gaussian behavior, and of dynamical heterogeneity to elastic scattering on the ps-ns time scale. The strong time dependence of $\langle r^2\rangle_R$ revealed in these studies further opens the question whether an intrinsic long-time, time-independent MSD can be obtained from simulations. Given that folded proteins have well-defined three-dimensional structures, a well-defined long-time, intrinsic MSD would seem to be reasonable and should exist. That is, the internal MSD arising from internal motions in proteins should converge to a plateau as a function of time and a parallel in proteins of the Debye-Waller and x-ray B factors found in the crystalline state should exist. Given this assumption, the question focuses on developing a method to extract the intrinsic $\langle r^2\rangle$ from finite-time simulations.

In addition to the ISF $I_{\text{inc}}(\mathbf{Q},t)$, an MSD can be calculated directly from simulations. This MSD is

$$\Delta^2(t) = \langle[r(t) - r(0)]^2\rangle \equiv \frac{1}{N}\sum_{i=1}^{N}\langle[r_i(t) - r_i(0)]^2\rangle, \tag{3}$$

where $r_i(t)$ is the position of nucleus i in the protein at time t. After long times $t \to \infty$, when the positions $r_i(t)$ and $r_i(0)$ are no longer correlated, $[\langle r(\infty)r(0)\rangle = 0]$, the $\Delta^2(t)$ reduces to

$$\Delta^2(t \to \infty) = \langle r^2(\infty)\rangle + \langle r^2(0)\rangle = 2\langle r^2\rangle_{\text{MD}}. \tag{4}$$

In this way an $\langle r^2\rangle_{\text{MD}} = \Delta^2(t \to \infty)/2$ for H in proteins can be defined and calculated. However, we emphasize that this $\langle r^2\rangle_{\text{MD}}$ is not the same as the intrinsic $\langle r^2\rangle$ defined in terms of $I(\mathbf{Q},t)$ in Eq. (2). First, the average over the nuclei in the protein made when the full $I_{\text{inc}}(\mathbf{Q},t)$ is represented by a global $I(\mathbf{Q},t)$ is not the same as the average over the nuclei made in Eq. (3). Also, the $\Delta^2(t)$ does not converge to a constant, infinite-time value within accessible simulation times of 1 to 100 ns [20,23,24]. Thus the correlations $\langle r(t)r(0)\rangle$ have not vanished and Eq. (4) is not obviously valid within accessible simulation times. Essentially, an $\langle r^2\rangle = \Delta^2(t \to \infty)/2$ cannot be calculated within currently accessible simulation times. Keeping these differences in mind, $\Delta^2(t)/2$ is a physically interesting, time-dependent quantity to calculate. It is especially useful in determining the time scales needed for $\Delta^2(t)/2$ to converge toward a fixed value. However, even the converged value may differ from $\langle r^2\rangle$ in Eq. (2).

In this context, the goal of the present paper is to propose a method for obtaining an intrinsic, long-time, $t \to \infty$, value of the MSD from finite-time simulations. We seek an intrinsic MSD from simulation that is defined exactly as in neutron scattering measurements, i.e., in terms of the global $I(\mathbf{Q},t)$ in Eq. (2). The procedure is to construct a model of the global $I(\mathbf{Q},t)$ which contains $\langle r^2\rangle$ and fit the model to finite-time simulations of $I_{\text{inc}}(\mathbf{Q},t)$. Explicitly, the $I_{\text{inc}}(\mathbf{Q},t)$ observed in neutron scattering experiments and calculated from simulations is [39]

$$I_{\text{inc}}(\mathbf{Q},t) = \frac{1}{N}\sum_{i=1}^{N} b_i^2\langle e^{-i\mathbf{Q}\cdot\mathbf{r}_i(t)}e^{i\mathbf{Q}\cdot\mathbf{r}_i(0)}\rangle. \tag{5}$$

In Eq. (5), b_i is the incoherent scattering length of nucleus i in the protein. As indicated above, the b_i of H is more than 20 times larger than the b_i of other nuclei typically found in proteins. For this reason the scattering from H, which is also almost entirely incoherent, dominates the ISF. In the analysis of neutron scattering experiments it is usual to represent the ISF in Eq. (5) summed over all nuclei by a global $I(\mathbf{Q},t)$ which

represents the whole protein:

$$I(\mathbf{Q},t) = \langle \exp(-i\mathbf{Q}.\mathbf{r}(t)) \exp(i\mathbf{Q}.\mathbf{r}(0)) \rangle. \quad (6)$$

Following the same procedure we construct a model of $I(\mathbf{Q},t)$. The model $I(\mathbf{Q},t)$ is separated into a time-independent part, $I(\mathbf{Q},t=\infty)$, and a time-dependent part, $I'(\mathbf{Q},t)$. The time-independent part, in the Gaussian approximation, is $I_\infty = I(\mathbf{Q},t=\infty)$ given by Eq. (2). I_∞ is the familiar Debye-Waller factor. We define the intrinsic MSD as the $\langle r^2 \rangle$ that appears in I_∞.

To implement the method, we first calculate $I_{\rm inc}(\mathbf{Q},t)$ from MD simulations of a hydrated protein (lysozyme) using Eq. (5). We then fit the model of the global $I(\mathbf{Q},t)$ in Eq. (6), which contains I_∞ and $\langle r^2 \rangle$, to the simulated $I_{\rm inc}(\mathbf{Q},t)$. We treat the $\langle r^2 \rangle$ in the model as a free fitting parameter. In this way we obtain an infinite-time value of the MSD $\langle r^2 \rangle$ from fits to simulation data at finite t.

We test the present method using simulations of lysozyme at several temperatures and of two simulation lengths, $t = 100$ ns and 1 μs. From the simulations, we calculate $I_{\rm inc}(\mathbf{Q},t)$ given by Eq. (5) out to 1 and 10 ns, respectively. From fits of the model $I(\mathbf{Q},t)$ to the calculated $I_{\rm inc}(\mathbf{Q},t)$, we obtain corresponding fitted values of $\langle r^2 \rangle$. We find that $\langle r^2 \rangle$ is the same for the two simulation times, consistent with $\langle r^2 \rangle$ representing a time-independent, long-time MSD. The intrinsic $\langle r^2 \rangle$ is approximately twice $\langle r^2 \rangle_R$, the MSD calculated for motions out to 1.5 ns. A plot of $\langle r^2 \rangle$ versus temperature shows a break in the slope at 140 K and a second DT at $T_D = 220$ K, as has been observed and calculated for proteins experimentally. According to this model, then, the DTs are intrinsic properties of proteins. While the appearance of the DT and transition temperature, T_D, may be modified by experimental time windows (see Sec. V C), the transitions exist independently of finite experimental time windows.

II. MOLECULAR DYNAMICS SIMULATION

Two lysozyme molecules (1AKI [40]) were arbitrarily oriented as shown in Fig. 1 and placed in a simulation box of dimensions 6.5 nm × 3.4 nm × 3.6 nm. The lysozyme molecules inside the simulation box were surrounded by 636 water molecules, corresponding to the hydration level $h = 0.4$ g water/g protein. The box was replicated using periodic boundary conditions to mimic the environment of an experimental powder sample. Similar simulation systems are discussed in the literature [8,26,41–44].

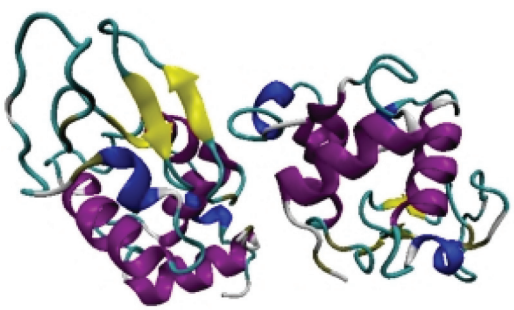

FIG. 1. (Color online) Two lysozyme molecules of random relative orientation selected by GROMACS.

The system was simulated using GROMACS 4.5.1 [45]. The OPLS-AA force field [46] was used for the protein, and the TIP4P force field [47] for the water. The van der Waals interaction was truncated at 1.4 nm, and the electrostatic interaction was represented using the particle mesh Ewald method [48] with a real-space cutoff of 0.9 nm. All bonds including H bonds were constrained with a linear constraints solver algorithm (LINCS) [49]. The energy of the system was first minimized using 50 000 steepest descent steps. The system was then equilibrated in the NVT (mole-volume-temperature) ensemble at each temperature investigated for 10 ns and in the NPT (mole-pressure-temperature) ensemble at 1 bar for 10 ns. The Nose-Hoover algorithm [50] with a coupling time $\tau = 1$ ps and the Parrinello-Rahman algorithm [51] with a coupling time $\tau = 3$ ps were used for the temperature coupling and pressure coupling, respectively.

Simulations of 100-ns length were performed at 18 temperatures between 80 and 300 K. Simulations of 1 μs were made at five temperatures, at 100 K and then in steps of 50 K to 300 K. Data were collected every 10 ps at each temperature for both simulations.

III. THE ISF AND MODEL $I(\mathbf{Q},t)$

To obtain the intrinsic MSD, $\langle r^2 \rangle$, following the procedure outlined in Sec. I, we first calculated the intermediate ISF, $I_{\rm inc}(\mathbf{Q},t)$, defined in Eq. (5). The positions $r_i(t)$ of each nucleus i in the protein were generated in the two MD simulations described above, one of length 100 ns and the other 1 μs. Using the $r_i(t)$, the $I_{\rm inc}(\mathbf{Q},t)$ is calculated directly for times out to 100 ns and 1 μs, respectively. To improve the statistics, each simulation was divided into at least 100 segments and $I_{\rm inc}(\mathbf{Q},t)$ recalculated as an average over these segments. In this way $I_{\rm inc}(\mathbf{Q},t)$ was calculated to 1 ns from the 100-ns MD data and to 10 ns from the 1-μs simulation.

Next we developed a model for the global $I(\mathbf{Q},t)$ defined in Eq. (6) which contains the intrinsic MSD, $\langle r^2 \rangle$, defined in Eq. (2). The model is obtained by first separating $I(\mathbf{Q},t)$ into a time-independent, $t = \infty$, part [$I_\infty = I(\mathbf{Q},t=\infty)$] and a time-dependent part [$I(\mathbf{Q},t) - I_\infty$],

$$I(\mathbf{Q},t) = I_\infty + (I(\mathbf{Q},t) - I_\infty). \quad (7)$$

From Eq. (6),

$$\begin{aligned} I_\infty = I(\mathbf{Q},t=\infty) &= \langle \exp(-i\mathbf{Q}.r(\infty)) \exp(i\mathbf{Q}.r(0)) \rangle, \\ &= \exp\left(-\tfrac{1}{3}Q^2\langle r^2 \rangle + O(Q^4)\right) \end{aligned} \quad (8)$$

is the infinite time limit [36]. To obtain the last expression we assume (i) that $r(\infty)$ and $r(0)$ are completely uncorrelated so that the averages of them are independent, (ii) that the system is translationally invariant in time [no center-of-mass (CM) motion] so that $r(\infty) = r(0)$, and (iii) that in a cumulant expansion of $\langle \exp(-i\mathbf{Q}.r) \rangle$, cumulants beyond the second are negligible. The last assumption is valid if Q is small or if the distribution over r is approximately a Gaussian distribution. The cumulants beyond the second vanish exactly for all Q if the distribution over r is exactly Gaussian.

The time-dependent part of $I(\mathbf{Q},t)$ has the limits

$$I'(\mathbf{Q},t) = I(\mathbf{Q},t) - I_\infty = \begin{cases} 1 - I_\infty, & t = 0, \\ 0, & t = \infty. \end{cases}$$

We model this by the function

$$I'(\mathbf{Q},t) = (1 - I_\infty)C(t), \tag{9}$$

where $C(t)$ has the limits $C(t = 0) = 1$, $C(t = \infty) = 0$. An example is the stretched exponential function,

$$C(t) = \exp\left(-(\lambda t)^\beta\right), \tag{10}$$

where λ and β are constants. The $C(t)$ represents the decay of correlations in the protein. Collecting, the model is

$$I(\mathbf{Q},t) = I_\infty(\mathbf{Q}) + (1 - I_\infty(\mathbf{Q}))C(t), \tag{11}$$

which is constructed to have the correct limits at $t = 0$ and $t = \infty$ and to have a plausible representation of several motional decay processes at intermediate times described by $C(t)$. We fit the model $I(\mathbf{Q},t)$ in Eq. (11) to the calculated $I_{\rm inc}(\mathbf{Q},t)$ to determine $\langle r^2\rangle$, λ, and β. The model $I(\mathbf{Q},t)$ is the same one we used previously [36] to fit neutron data except that $C(t)$ is a stretched exponential rather than the simple exponential used previously. In fits to neutron data we found [36] that the data were not sufficiently precise to distinguish between a stretched and a simple exponential. In contrast, a simulation-derived $I_{\rm inc}(\mathbf{Q},t)$ is more discriminating.

IV. RESULTS

In this section, we present the fits of the model $I(\mathbf{Q},t)$ given by Eq. (11) to the calculated ISF data for lysozyme. The model of $I(\mathbf{Q},t)$ includes three fitting parameters: the intrinsic MSD, $\langle r^2\rangle$, defined in Eq. (2), the relaxation parameter λ, and the stretched exponential parameter β defined in Eq. (10). The goal is to determine the intrinsic MSD $\langle r^2\rangle$ of H in lysozyme and to obtain values for the relaxation parameters λ and β.

A. Intrinsic MSD

1. 100-ns MD simulation

Figure 2 shows the intermediate scattering function, $I_{\rm inc}(\mathbf{Q},t)$, for times $0 < t < 1$ ns calculated from Eq. (5) using the 100-ns simulation data. Although $I_{\rm inc}(\mathbf{Q},t)$ was calculated at 18 temperatures, only 3 temperatures are shown in Fig. 2. The solid lines are fits of the model $I(\mathbf{Q},t)$ to the calculated $I_{\rm inc}(\mathbf{Q},t)$. The fits are better at higher temperatures than at lower temperatures. Particularly, at temperatures below approximately 170 K, the parameters λ and β that appear in the relaxation function $C(t)$ are not well determined, as we discuss below.

Figures 3 and 4 show the best-fit values of the fitting parameters $\langle r^2\rangle$, λ, and β. From Fig. 3(a), we see that $\langle r^2\rangle$ is Q dependent and is larger and approximately independent of Q at low Q. This Q dependence is similar to that obtained from fits to observed data [to $S(\mathbf{Q},\omega = 0)$]. The $\langle r^2\rangle$ is determined chiefly by the value of $I(\mathbf{Q},t)$ at long t, i.e., by how far $I(Q,\infty)$ lies below $I(Q,t = 0) = 1$.

From Fig. 3(b), we see that λ is also Q dependent, with $\lambda \propto Q^2$ approximately, as found in other simulations [8,9]. We found β to be only weakly dependent on Q and we used an average over several Q values with some adjustments to obtain smooth behavior as a function of temperature. The resulting temperature dependence of β is shown in Fig. 4(b). The λ and β are not well determined at temperatures below approximately 170 K. Essentially, at low temperatures $I_{\rm inc}(\mathbf{Q},t)$ decreases rapidly over a short time t and thereafter it changes slowly. This time dependence is consistent with harmonic motion as shown by Smith *et al.* [19,52] and can be approximately reproduced by a range of λ and β values. The $\langle r^2\rangle$ remains well determined at low temperatures since it is determined chiefly by $I_{\rm inc}(\mathbf{Q},t)$ at long times.

The decrease in the best-fit value of $\langle r^2\rangle$ with increasing Q can have two origins. First, at low Q we are sampling longer-range phenomena. Long-range motions could contribute fully to $\langle r^2\rangle$ at low Q, whereas they could be limited or cut off at high Q, leading to a smaller or limited $\langle r^2\rangle$ at high Q. Since this is a real physical effect, values of $\langle r^2\rangle$ obtained from data [e.g., $S(\mathbf{Q},\omega = 0)$] at low Q ($0 < Q < 0.4$Å^{-1}) are usually selected. Second, and most importantly, Yi *et al.* [34] have shown that $I(\mathbf{Q},t)$ departs from a Gaussian approximation because of dynamical heterogeneity. The heterogeneity apparently introduces a Q^4 term in Eq. (8) which becomes significant at larger Q. This means that an $\langle r^2\rangle$ obtained from data at small Q must be selected.

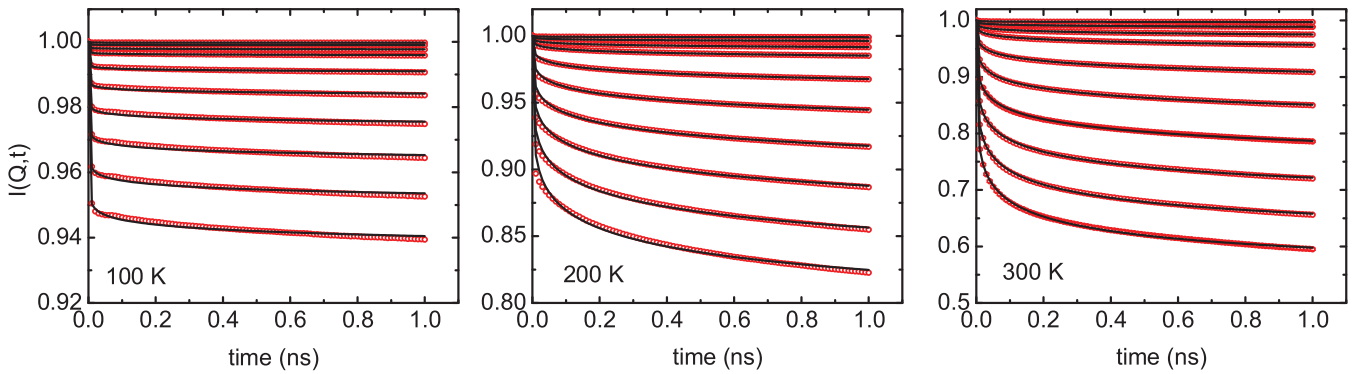

FIG. 2. (Color online) Intermediate scattering function (ISF), $I_{\rm inc}(\mathbf{Q},t)$, for $0 < t < 1$ ns of hydrated lysozyme ($h = 0.4$) obtained from a 100 -ns MD simulation [open (red) circles] and fits of the model ISF $I(\mathbf{Q},t)$ in Eq. (11) [solid (blue) lines] to the $I_{\rm inc}(\mathbf{Q},t)$ at 100, 200, and 300 K. From top to bottom, $Q = 0.1$, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4, and 1.6 Å^{-1}. The error in the ISF is approximately the diameter of the open circles.

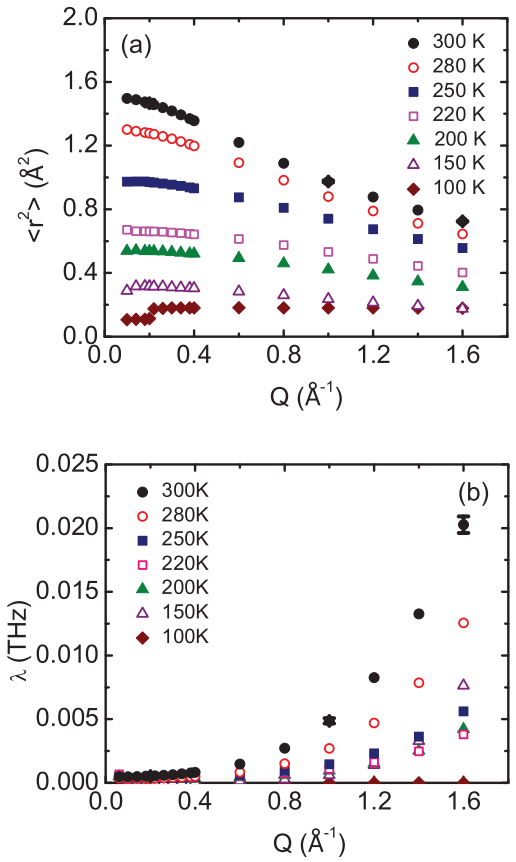

FIG. 3. (Color online) Parameters of the model ISF $I(\mathbf{Q},t)$ of Eq. (11) obtained from fits of the model to the simulations shown in Fig. 2: (a) Intrinsic MSD, $\langle r^2\rangle$, and (b) relaxation parameter, λ, versus Q at temperatures of 100 to 300 K. Error bars on $\langle r^2\rangle$ and λ at 300 K and $Q = 0.2$, 1.0, and 1.6 Å^{-1} are indicated.

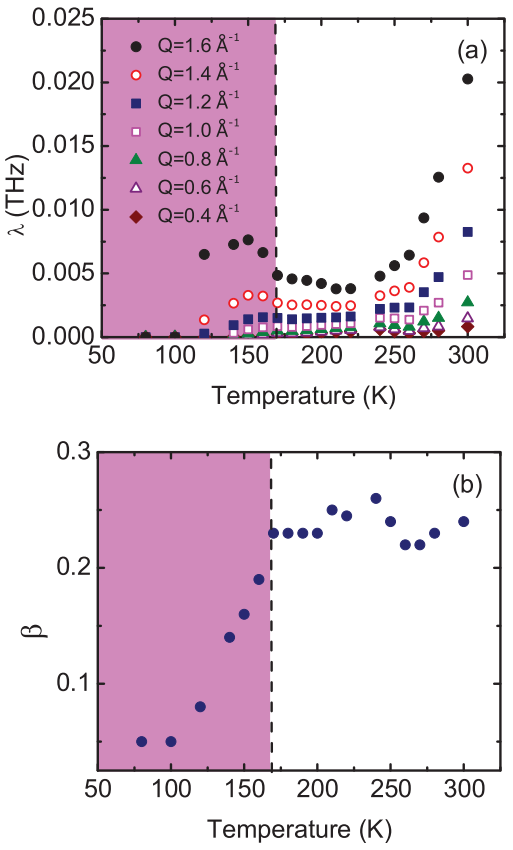

FIG. 4. (Color online) As Fig. 3, for (a) the relaxation parameter λ and (b) the stretched exponential parameter β versus temperature. From bottom to top, $Q = 0.4$, 0.6, 0.8, 1, 1.2, 1.4, and 1.6 Å^{-1}.

Figure 5 shows the intrinsic MSD $\langle r^2\rangle$ obtained from fits to $I_{\rm inc}(\mathbf{Q},t)$ at $Q = 0.2$ Å^{-1}. The plot of $\langle r^2\rangle$ versus temperature shows a break in the slope at around $T = 140$ K. This has been seen previously in simulations and arises from the activation of the dynamics of hydrophobic groups, i.e., the onset of proline puckering and the rotation of methyl groups at around 140 K [6,29]. A break in slope of $\langle r^2\rangle$ versus T near 140 K has also been observed in several proteins [1,6,17,35]. A second break in slope is seen at the well-documented DT, at $T_D \simeq 220$ K, associated with the onset of new higher-amplitude motions of hydrophilic groups in which the hydration water plays a determining role. The intrinsic, long-time $\langle r^2\rangle$ shows the onset of both hydrophobic and hydrophilic (DT) motions.

The error bars of $I(Q,t)$ are approximately the size of the open circles in Fig. 2. Using these error bars, the corresponding error bar in the fitting parameters is shown in Fig. 3. The error bar is calculated using the fitting program. The corresponding error bar in β is approximately ± 0.01. In the fit, we assumed a constant value of $\beta = 0.24$ at all temperatures.

2. 1-μs MD simulation

We turn now to the $I_{\rm inc}(\mathbf{Q},t)$ calculated from the $r_i(t)$ generated in the 1-μs simulations. Five temperatures from 100 to 300 K were simulated. As before, the simulation data were divided into at least 100 segments (time slices), each spanning a time t, $0 < t < 10$ ns, and $I_{\rm inc}(\mathbf{Q},t)$ was calculated as the average over the time slices. The resulting $I(\mathbf{Q},t)$ values are shown in Fig. 6 at three temperatures. The solid line in Fig. 6 is, again, a fit of the model $I(\mathbf{Q},t)$ given by Eq. (11) with $\langle r^2\rangle$, λ, and β treated as free fitting parameters. As in Fig. 2, the fits are better at the higher temperatures, although the fit is better at 200 K than at 300 K for times out to 10 ns.

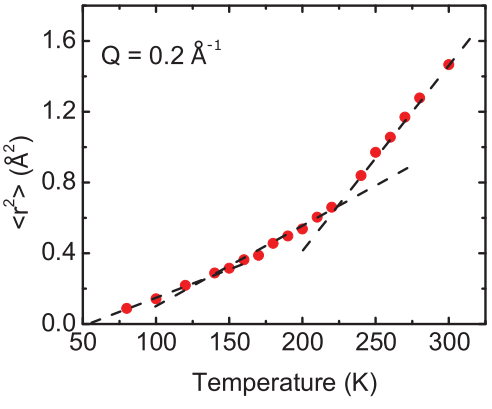

FIG. 5. (Color online) Intrinsic MSD, $\langle r^2\rangle$, defined in Eq. (2), obtained from fits of the model ISF, Eq. (11), to simulations of $I_{\rm inc}(\mathbf{Q},t)$ at $Q = 0.2$ Å^{-1}. The intrinsic $\langle r^2\rangle$ shows a break in slope at $T \simeq 140$ K and $T \simeq 220$ K.

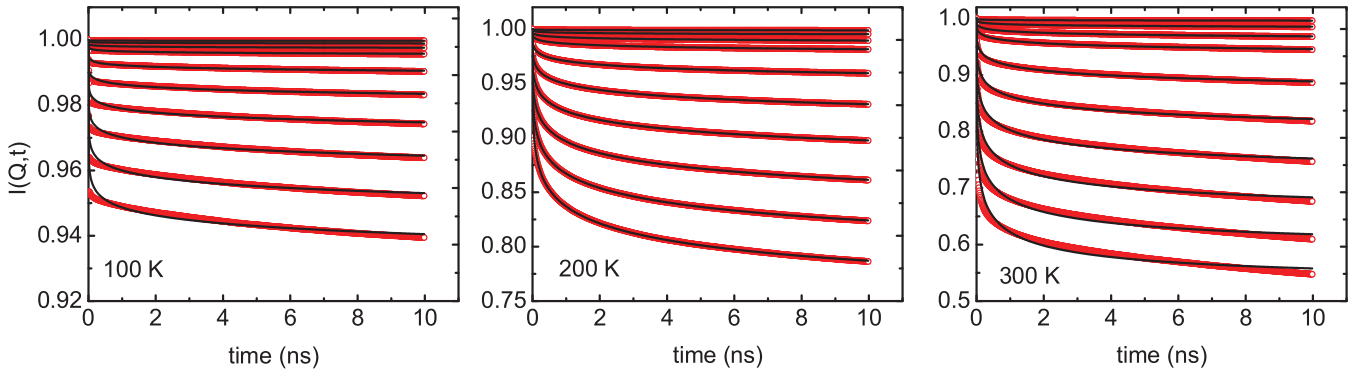

FIG. 6. (Color online) Calculated $I_{\rm inc}(\mathbf{Q},t)$, for $0 < t < 10$ ns, of hydrated lysozyme obtained from a 1-μs MD simulation (open circles) and fits of the model ISF $I(\mathbf{Q},t)$ in Eq. (11) to the data (solid lines) at 100, 200, and 300 K. From top to bottom, $Q = 0.1$, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4, and 1.6 Å^{-1}.

The best-fit values of the intrinsic MSD $\langle r^2\rangle$, λ, and β are shown in Fig. 7. The $\langle r^2\rangle$ decreases with increasing Q, as found for the shorter simulation. The absolute values of $\langle r^2\rangle$ are also consistent with those obtained from the shorter simulation except, possibly, at 100 K. λ^2 is approximately proportional to Q^2, as found in the shorter simulation. The absolute values of λ obtained from fits to $I_{\rm inc}(\mathbf{Q},t)$ over larger times (10 ns) are significantly smaller than those obtained from fits over a shorter time (1 ns) [compare Figs. 3(b) and 7(b)]. The values of β are similar for the two simulation times.

Figure 8 compares the values of $\langle r^2\rangle$ obtained from fits to $I_{\rm inc}(Q,t)$ for times t out to 1 and 10 ns. The agreement of the two is good and excellent at higher temperatures. This suggests that $\langle r^2\rangle$ is, indeed, a long-time ($t \to \infty$) intrinsic value of $\langle r^2\rangle$ that is independent of the time interval of the data from which it is obtained. It also indicates that no new motional process enters the simulations between 100 ns and 1 μs.

B. Simulated MSD $\Delta^2(t)$

In this section we present values of $\Delta^2(t)$ defined by Eq. (3) and calculated directly using the $r_i(t)$ generated in simulations. The sum in Eq. (3) is taken over the nonexchangeable H nuclei only in the lysozyme. That is, all other nuclei in the protein, the H in the hydration water, and the H in the protein that can exchange positions with H in the hydration water (the exchangeable H) are excluded from the sum. The time dependence of $\Delta^2(t)$ and estimated values of $\Delta^2(t = \infty)$ are compared the intrinsic MSD $\langle r^2\rangle$ defined in Eq. (2). The $\langle r^2\rangle$ and $\Delta^2(\infty)/2$ will be the same only if all H are in identical environments in the protein. Specifically, the two will differ when there is dynamical heterogeneity.

Figure 9 shows $\Delta^2(t)/2$ obtained from the 1-μs simulation calculated out to $t = 10$ ns. At $T = 100$ K the $\Delta^2(t)$ appears to have converged after 10 ns and $\Delta^2(t = 10$ ns) and $\langle r^2\rangle$ are quite similar. The $\langle r^2\rangle$ values shown are those from the 100-ns simulations. The $\langle r^2\rangle$ from the 1-μs simulations are slightly larger at 100 and 150 K (see Fig. 8). In contrast, at higher temperatures, $\Delta^2(t)$ has clearly not reached its terminal ($t = \infty$) value after 10 ns. For example, at $T = 300$ K, the intrinsic $\langle r^2\rangle$ is approximately 30%–40% larger than $\Delta^2(t = 10$ ns)/2. This comparison between $\Delta^2(t)/2$ and $\langle r^2\rangle$ is consistent with $\langle r^2\rangle$ representing the intrinsic ($t = \infty$) MSD.

Figure 10 further compares the intrinsic MSD $\langle r^2\rangle$ at $Q = 0.2$ Å^{-1} and the $\Delta^2(t)/2$ at different times t obtained from the 100-ns [Fig. 10(a)] and 1-μs [Fig. 10(b)] simulations. From Fig. 10 we see again that $\Delta^2(t)$ has not converged to a long-time value after $t = 10$ ns except possibly at 100 K. This is especially true at higher temperature. For example, at 300 K the increase in $\Delta^2(t)/2$ between 1 and 10 ns is approximately the same as that between 0.1 and 1 ns, suggesting that convergence is very slow. The $\Delta^2(t)/2$ lie well below the intrinsic $\langle r^2\rangle$, especially at high temperatures.

Figure 11 shows the $\Delta^2(t)$ at 200 and 300 K over the time range $0 < t < 1$ ns as calculated from the 100-ns simulation and the 1-μs simulation. The $\Delta^2(t)$ obtained from data taken out to 1 μs is somewhat smaller than that from the 100-ns simulation. This suggests that there may be a structural change in the time scale between 100 ns and 1 μs. However, these small differences do not appear to affect $I_{\rm inc}(\mathbf{Q},t)$ or the fitted intrinsic $\langle r^2\rangle$ significantly. Values of $\Delta^2(t)$ obtained from the 1-μs simulation with and without the CM motion subtracted are also shown. The contribution of the CM motion to $\Delta^2(t)$ is small. Values of $r_i(t)$ corrected for CM motion were used to calculate $I_{\rm inc}(\mathbf{Q},t)$.

V. DISCUSSION

The aim of the present paper is to propose a method to obtain the intrinsic, long-time MSD in proteins from finite-time simulations. The intrinsic MSD $\langle r^2\rangle$ is defined as the $\langle r^2\rangle$ that appears in the infinite time limit of the incoherent ISF given by Eq. (2), often referred to as the Debye-Waller factor. The method consists of calculating the ISF $I_{\rm inc}(\mathbf{Q},t)$ from a simulation and fitting a model $I(\mathbf{Q},t)$ which contains $\langle r^2\rangle$ to the calculated $I_{\rm inc}(\mathbf{Q},t)$. The resulting intrinsic $\langle r^2\rangle$ exhibits two interesting features: (i) the $\langle r^2\rangle$ is independent of the simulation time used to calculate $I_{\rm inc}(\mathbf{Q},t)$, at least up to 1 μs, and (ii) the $\langle r^2\rangle$ shows a clear break in the slope of $\langle r^2\rangle$ vs T at the DT and a second break at a lower temperature, $T \simeq 140$ K. The intrinsic $\langle r^2\rangle$ shows the same breaks in slope that are found in time-limited MSD and observed in experiments. This suggests that a DT is an intrinsic property of proteins, not simply an artifact of finite instrument resolution and limited time windows.

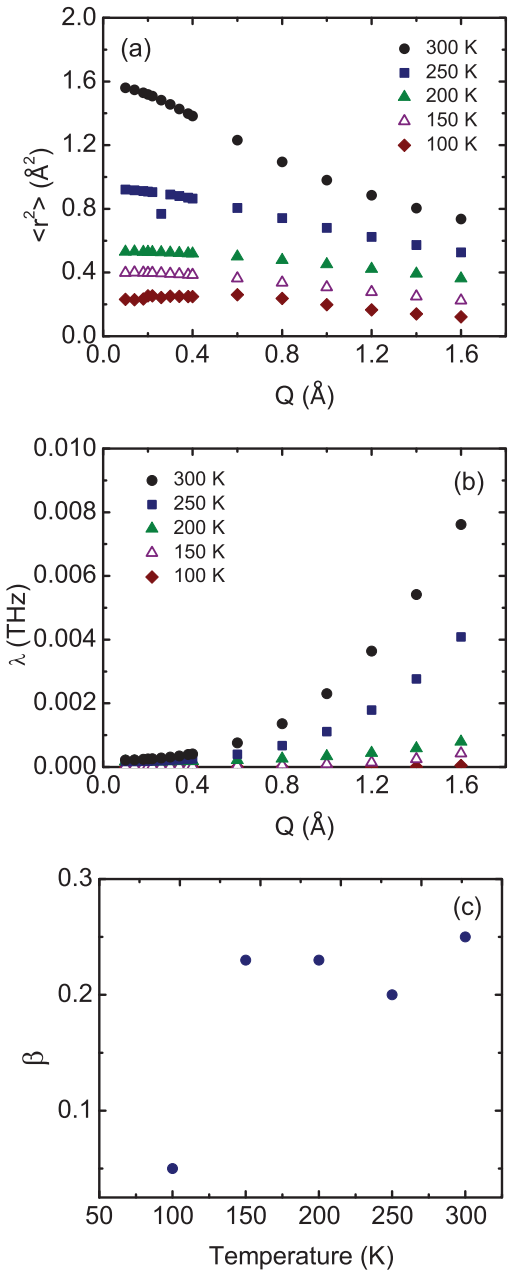

FIG. 7. (Color online) Parameters of the model ISF, Eq. (11), obtained from the fits of $I(\mathbf{Q},t)$ to the calculated $I_{\rm inc}(\mathbf{Q},t)$, for $0 < t < 10$ ns, shown in Fig. (6): (a) intrinsic MSD, $\langle r^2\rangle$, (b) relaxation parameter, λ, and (c) stretched exponential parameter, β, at five temperatures: 100, 150, 200, 250, and 300 K.

A. Comparison with existing MSD for lysozyme

To place the present intrinsic $\langle r^2\rangle$ in context with the existing MSD, first, we compare $\langle r^2\rangle$ with the resolution-broadened MSD $\langle r^2\rangle_R$, both calculated from the present simulations of lysozyme. A resolution-broadened $\langle r^2\rangle_R$ is an MSD that has developed over a finite time only. This time, τ_R, is determined by the resolution width as $\tau_R = (8\ln 2)^{1/2}\hbar/W$ for a Gaussian resolution function. The smallest FWHM readily available today is $W = 1$ μeV ($\tau_R = 1.5$ ns). By using the same simulation for both $\langle r^2\rangle$ and $\langle r^2\rangle_R$, we can isolate the impact of a finite resolution width.

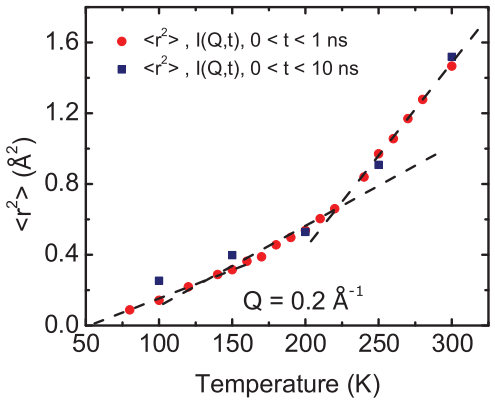

FIG. 8. (Color online) Intrinsic MSD $\langle r^2\rangle$ versus temperature obtained from fits to the $I_{\rm inc}(\mathbf{Q},t)$ at $Q = 0.2$ Å^{-1} obtained from (i) 100 ns (filled circles) and (ii) 1-μs MD simulations (filled squares). The $\langle r^2\rangle$ is largely independent of the simulation time fitted.

Simulated and observed MSDs are usually compared by comparing resolution-broadened, $\langle r^2\rangle_R$, MSDs that have developed over the same time period [6,35,37]. Excellent agreement between simulated and observed MSD has been obtained in this way. Specifically, the resolution-broadened MSD, $\langle r^2\rangle_R$, is obtained from Eq. (1) in which $S_R(\mathbf{Q},\omega)$ is the observed, resolution-broadened dynamic structure factor,

$$S_R(Q,\omega) = \frac{1}{2\pi}\int dt\,\exp(i\omega t)I(Q,t)R(t), \qquad (12)$$

and $R(t)$ is the Fourier transform of the instrumental resolution function in time. $R(t)$ is typically a Gaussian, $R(t) = \exp(-\frac{t^2}{2\tau_R^2})$. The resolution function cuts off $I(\mathbf{Q},t)$ after a time $\tau_R = (8\ln 2)^{1/2}\hbar/W$. In experiments, the observed $\langle r^2\rangle_R$ is obtained by inserting the observed $S_R(\mathbf{Q},\omega = 0)$ in Eq. (1). In simulations the calculated ISF is inserted in Eq. (12) to obtain $S_R(\mathbf{Q},\omega = 0)$ and then $\langle r^2\rangle_R$ is again obtained using Eq. (1). In this way an $\langle r^2\rangle_R$ that has evolved over a time τ_R is compared. We follow exactly this procedure to calculate $\langle r^2\rangle_R$ from the present simulations by substituting our model $I(\mathbf{Q},t)$ into Eq. (12).

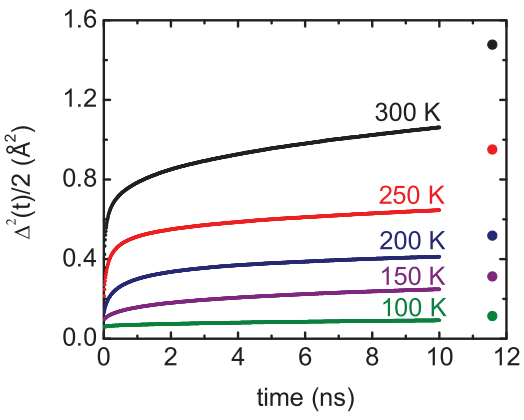

FIG. 9. (Color online) MSD $\Delta^2(t)/2$ defined in Eq. (3) of nonexchangeable hydrogen versus time at five temperatures, from 100 to 300 K. Filled circles are the corresponding intrinsic MSDs $\langle r^2\rangle$ at each temperature.

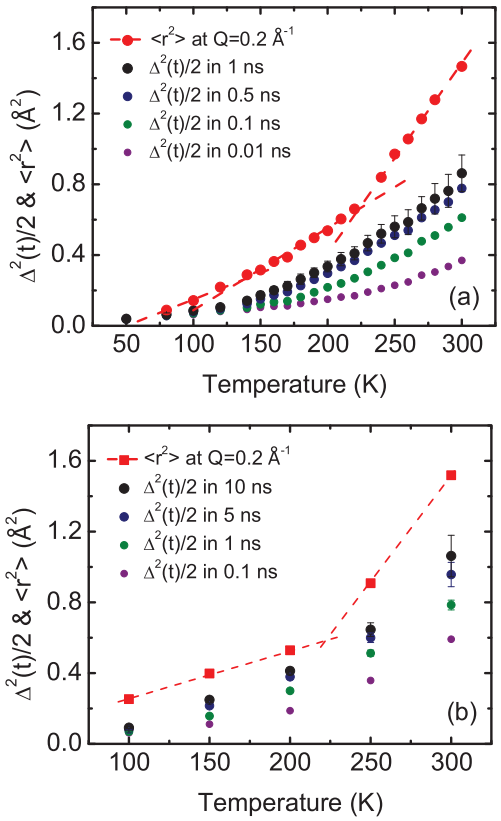

FIG. 10. (Color online) Comparison of the intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2$ Å^{-1} and the MSD $\Delta^2(t)/2$; (a) out to times 0.01, 0.1, 0.5, and 1 ns obtained from the 100-ns MD simulation and (b) out to times 0.1, 1, 5, and 10 ns, obtained from the 1-μs MD simulation.

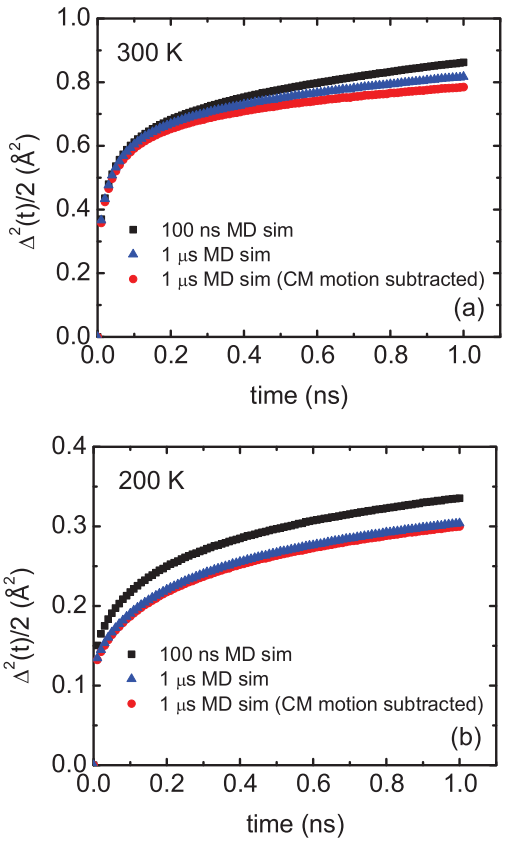

FIG. 11. (Color online) MSD $\Delta^2(t)/2$ of nonexchangeable hydrogen versus time up to 1 ns as calculated from MD simulations of 100 ns [(black) squares] and 1 μs with [(red) circles] and without [(blue) triangles] CM motion subtracted at (a) 300 K and (b) 200 K.

The intrinsic $\langle r^2 \rangle$ and the resolution-broadened $\langle r^2 \rangle_R$ of lysozyme are compared in Fig. 12. The $\langle r^2 \rangle$ is found to be approximately twice the $\langle r^2 \rangle_R$ for a resolution width $W = 1$ μeV ($\tau_R = 1.5$ ns). The ratio $\langle r^2 \rangle / \langle r^2 \rangle_R$ is approximately independent of temperature for $T > 150$ K. The $\langle r^2 \rangle$ remains well above $\langle r^2 \rangle_R$ for $W = 0.1$ μeV ($\tau_R = 15$ ns), a resolution approximately 10 times higher than that available today. Physically, we expect resolution broadening to reduce $\langle r^2 \rangle_R$ below $\langle r^2 \rangle$ if τ_R is less than or comparable to the longest relaxation time $\tau = \lambda^{-1}$ of the protein. From Figs. 3 and 7 we see that, at low Q and temperatures above 170 K, $\tau = \lambda^{-1} \geqslant 1$ ns. The τ is somewhat longer at low temperatures. On this basis we expect resolution broadening to be important at $W = 1$ μeV. The degree of broadening depends sensitively on the functional form of $C(t)$ in the model $I(\mathbf{Q},t)$, as discussed below.

Second, we compare the present simulated $\langle r^2 \rangle_R$ with previous simulated values of $\langle r^2 \rangle_R$ for lysozyme. Roh *et al.* [6] have calculated $\langle r^2 \rangle_R$ at low Q and $W = 1$ μeV from their simulations of lysozyme hydrated to $h = 0.43$. The Roh *et al.* $\langle r^2 \rangle_R$ and the present $\langle r^2 \rangle_R$ at low Q ($Q = 0.2$ Å^{-1}) and $W = 1$ μeV for $h = 0.40$ are compared in Fig. 13. The agreement is excellent since the present $\langle r^2 \rangle_R$ is somewhat lower as expected since the present h is lower and the MSD is very sensitive to h.

Third, we compare the present intrinsic $\langle r^2 \rangle$ with the $\langle r^2 \rangle_R$ observed experimentally in lysozyme using an energy resolution width $W = 1$ μeV. The observed $\langle r^2 \rangle_R$ of lysozyme at four hydration levels and the present $\langle r^2 \rangle$ are shown in Fig. 14. The observed $\langle r^2 \rangle_R$ values are very sensitive to the hydration level for $T > T_D$. In Fig. 14 the present intrinsic $\langle r^2 \rangle$ for $h = 0.4$ lies above but close to the observed $\langle r^2 \rangle_R$ for $h = 0.45$ and significantly higher than the observed $\langle r^2 \rangle_R$ for lower hydrations, as expected. Comparing the calculated $\langle r^2 \rangle_R$ in Fig. 13 with the observed $\langle r^2 \rangle_R$ in Fig. 14, we see that the calculated $\langle r^2 \rangle_R$ lie somewhat below but close to the

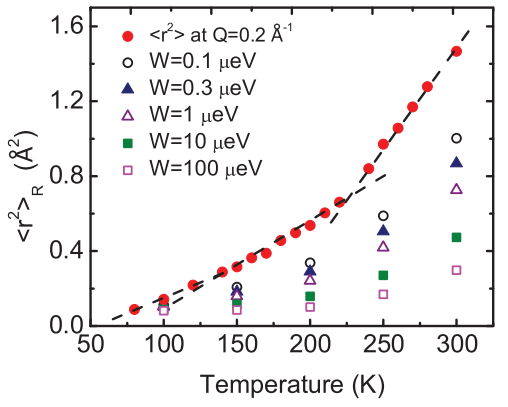

FIG. 12. (Color online) Present intrinsic MSD $\langle r^2 \rangle$ at $Q = 0.2$ Å^{-1} and resolution-broadened MSD $\langle r^2 \rangle_R$ calculated from the same model for the energy resolution widths, $W = 0.1$, 0.3, 1, 10, and 100 μeV.

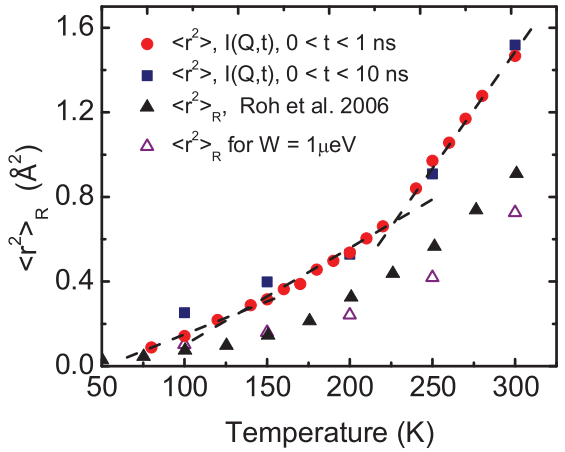

FIG. 13. (Color online) Present intrinsic MSD $\langle r^2\rangle$ at $Q = 0.2$ Å^{-1} and present resolution-broadened MSD $\langle r^2\rangle_R$ ($W = 1$ μeV) (open triangles) for lysozyme at $h = 0.40$ compared with the simulated $\langle r^2\rangle_R$ at $W = 1$ μeV of Roh *et al.* [6] for lysozyme at $h = 0.43$ (filled triangles).

experimental values for similar levels of hydration. Broadly the agreement between the simulated and the observed $\langle r^2\rangle_R$ values is very good, in terms of both the absolute value and the temperature dependence.

B. Sensitivity of $\langle r^2\rangle$ to the model $C(t)$

The difference between the intrinsic $\langle r^2\rangle$ and the resolution-broadened $\langle r^2\rangle_R$, such as shown in Fig. 13, is sensitive to the functional form of $C(t)$ used to describe the dynamic correlations in the protein in the model $I(\mathbf{Q},t)$. Four forms of $C(t)$ are compared in Fig. 15. The stretched exponential function for $C(t)$ given by Eq. (10) that we have used here provides a reasonable fit of the model $I(\mathbf{Q},t)$ to the calculated $I_{\rm inc}(\mathbf{Q},t)$ provided the parameter β is small, i.e., $\beta = 0.23$. In Fig. 15 we see that the stretched exponential for a small β and the Mittag-Leffler function [24] have long-range tails reaching out to times a factor of 10 beyond $t = \lambda^{-1}$. This means that correlations persist for times well beyond $\tau = \lambda^{-1}$. For this reason when a stretched exponential with a small β is used we

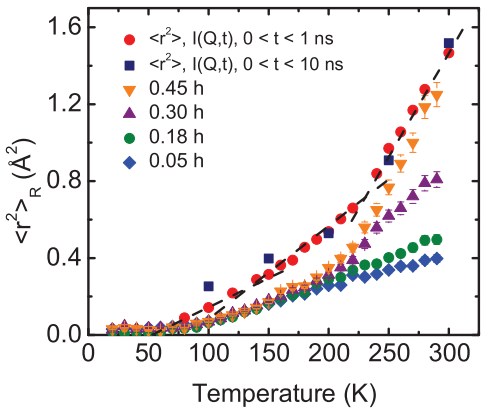

FIG. 14. (Color online) Present intrinsic MSD $\langle r^2\rangle$ at $Q = 0.2$ Å^{-1} for lysozyme ($h = 0.4$), as in Fig. 13, compared with the experimental resolution-broadened MSD $\langle r^2\rangle_R$ for $W = 1$ μeV for lysozyme at different hydration levels (h) observed by Roh *et al.* [6].

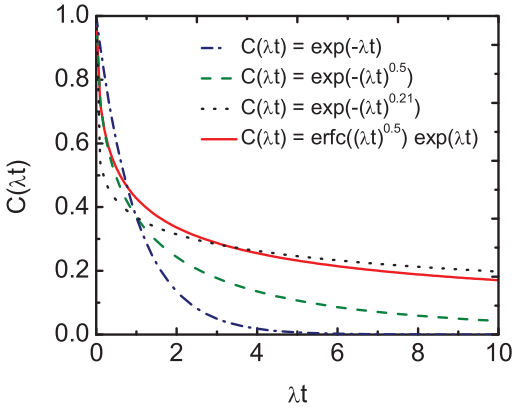

FIG. 15. (Color online) Comparison of relaxation functions: a simple exponential [dashed-dotted (blue) line], a stretched exponential with $\beta = 0.5$ [dashed (green) line], a stretched exponential with $\beta = 0.21$ [dotted (black) line], and the Mittag-Leffler function [solid (red) line].

expect $\langle r^2\rangle$ to lie above $\langle r^2\rangle_R$ even when $\lambda^{-1} \simeq \tau_R$, as found here in Fig. 13. In contrast, if $C(t)$ is represented by a simple exponential or a stretched exponential with a large value of β, the correlations die out rapidly on a time scale $\tau = \lambda^{-1}$. It was not possible to obtain a good fit to $I_{\rm inc}(\mathbf{Q},t)$ using a simple exponential or a large β. Thus correlations that persist to long times, $t = 10\lambda^{-1}$, appear to be a feature of lysozyme.

In an earlier study [36], we proposed a method to obtain the intrinsic MSD in proteins from fits to experiment, to observed resolution-broadened $S_R(\mathbf{Q},\omega = 0)$. The model $I(\mathbf{Q},t)$ employed was the same as that used here in Eq. (11). The model $I(\mathbf{Q},t)$ was Fourier transformed [see Eq. (12)] to obtain $S_R(\mathbf{Q},\omega = 0)$. However, $C(t)$ was represented by a simple exponential, chosen because the experimental data were not very discriminating and fits using a simple exponential and a stretched exponential could not be distinguished. The ratio $\langle r^2\rangle/\langle r^2\rangle_R$ obtained from fits to data at $W = 1$ μeV was approximately 1.0–1.2 rather than a factor of two as found here. MD simulation-derived $I_{\rm inc}(\mathbf{Q},t)$ are more discriminating. We believe that the present $C(t)$ and ratio $\langle r^2\rangle/\langle r^2\rangle_R$ are more accurate because the $C(t)$ obtained from simulations is more accurate. Hence, the method proposed in Ref. [36] to obtain $\langle r^2\rangle$ from experiment needs to be upgraded by replacing the exponential $C(t)$ with a stretched exponential, with β set at approximately 0.23.

The model could conceivably be further refined by using more sophisticated expressions for $C(t)$ that combine vibrational motion at short times and diffusion at longer times. Also, the expression for the time-dependent part of $I(\mathbf{Q},t)$ in Eq. (9) may be too simple.

C. Dynamical transition and impact of instrument resolution

Despite decades of experimental and theoretical studies, the physical origin of the DT remains debated. It has been ascribed [53] to a sudden change of effective elasticity in proteins [2], to the onset of motions of specific side groups, e.g., methyl group rotations [54], and to a glass transition or a phase transition in the hydration water [4,55] and interpreted as an apparent effect arising because the MSD is observed with a finite instrument

resolution width [38,56–58]. In this section we discuss the impact of observing the DT using an instrument having a finite energy resolution width, W, within the present model. First, the intrinsic MSD $\langle r^2\rangle$ that we have found in this paper for lysozyme, shown in Figs. 5 and 8, displays a clear DT at a transition temperature, T_D, of 220 K. This result suggests that, within the rigor of simulations, the DT is an intrinsic property of a protein. The DT is not simply an artifact of observing the MSD with an instrument having a finite W and a limited time window. However, the change of slope of the MSD at T_D can be modified and T_D shifted to a higher temperature when the DT is observed with a finite W [12,17,18,35] as emphasized recently [59].

When W is finite, the MSD $\langle r^2\rangle_R$ defined in Eqs. (1) and (12), rather than $\langle r^2\rangle$, is observed. In Eq. (12), motions in the protein can contribute to the $\langle r^2\rangle_R$ for a limited time, $\tau_R \simeq \hbar/W$, only. Since motions over a limited time window τ_R are included, $\langle r^2\rangle_R$ is always smaller than $\langle r^2\rangle$ at a given temperature. As a result, the T_D in $\langle r^2\rangle_R$ is shifted to a higher temperature. Using the present model, this shift is illustrated in Fig. 16, where the T_D is explicitly identified. For example, when observed with an instrument for which $W = 100$ μeV ($\tau_r = 15$ps for a Gaussian resolution function), the apparent T_D is shifted to 240 K.

The degree of impact of W on $\langle r^2\rangle_R$ depends on the rate at which correlations decay in the protein. In the present model, the decay rate depends on the magnitude of the parameter λ and on the functional form of $C(t)$. The present stretched exponential $C(t)$ has a long-time tail (see Fig. 15). This means that the reduction of $\langle r^2\rangle_R$ below $\langle r^2\rangle$ begins at small values of W (long τ_R). But the rate of change of $\langle r^2\rangle_R$ with W is gradual. In an earlier model [36], $C(t)$ was described by a simple exponential which falls rapidly with t (see Fig. 15). For this $C(t)$, the reduction of $\langle r^2\rangle_R$ below $\langle r^2\rangle$ begins at a larger value of W and thereafter the reduction increases rapidly with increasing W. We reproduce the $\langle r^2\rangle_R$ obtained for an exponential $C(t)$ in Fig. 17. In Fig. 17, T_D was identified simply as the temperature at which the calculated $\langle r^2\rangle$ begins to increase rapidly with temperature, i.e., the temperature at which the slope of $\langle r^2\rangle$ vs T changes markedly. The T_D

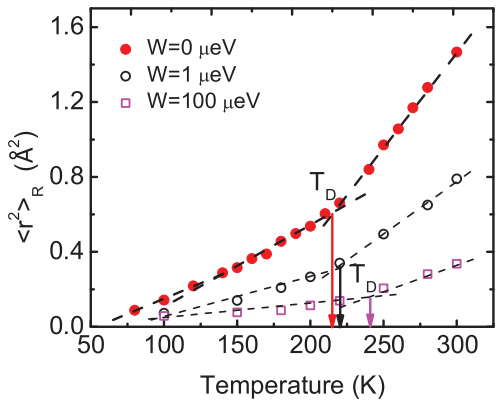

FIG. 16. (Color online) Intrinsic MSD ($W = 0$) and resolution-broadened MSD $\langle r^2\rangle_R$ for $W = 1$ μeV and $W = 100$ μeV obtained from Eqs. (1) and (12) using the present model $I(\mathbf{Q},t)$ (reproduced from Fig. 12), with the dynamical transition temperature T_D identified.

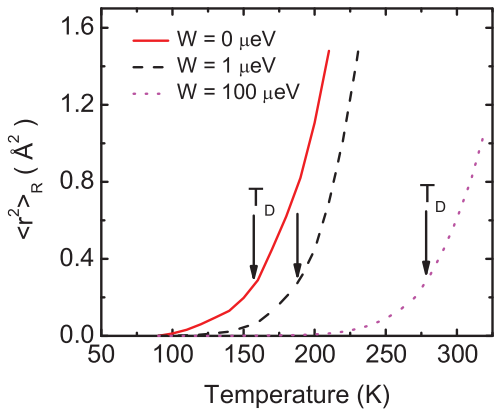

FIG. 17. (Color online) Intrinsic MSD ($W = 0$) and resolution-broadened MSD $\langle r^2\rangle_R$ for $W = 1$ μeV and $W = 100$ μeV with the T_D identified. MSDs are obtained from fits to experiment in Ref. [36] using a model, $I(\mathbf{Q},t)$, that has a simple exponential decay function $C(t) = \exp[-\lambda t]$.

increases rapidly with increasing W, which vividly illustrates the dependence of T_D on W. Using this simple model, the increase in the apparent T_D with increasing W can also be readily understood. In the model, $\langle r^2\rangle_R/\langle r^2\rangle = [1 + \frac{W}{I_\infty \lambda}]^{-1} \simeq [1 - \frac{W}{I_\infty \lambda} + \cdots]$. The parameter λ increases with increasing T. Thus for a given W, the ratio $W/(I_\infty\lambda)$ decreases with increasing temperature and $\langle r^2\rangle_R/\langle r^2\rangle$ is larger at higher temperatures. Thus $\langle r^2\rangle_R$ is decreased least by finite resolution at the highest temperatures.

As illustrated by these models, a DT is readily observed with an instrument having a finite resolution width. The chief impact of a finite W is to shift the apparent T_D to a higher temperature.

D. MSD calculated from simulations

We have also evaluated the MSD $\Delta^2(t)$ for H in lysozyme defined in Eq. (3) from the present simulations. The $\Delta^2(t)/2$ will be the same as the intrinsic $\langle r^2\rangle$ only (i) if the $\Delta^2(t)$ has reached its long-time, converged value so that the correlations are 0 as discussed in Eq. (4), and (ii) if all the H's in the protein are in identical environments so that $I_{\rm inc}(\mathbf{Q},t)$ reduces to the model $I(\mathbf{Q},t)$. Also, in the present work, only the nonexchangeable H nuclei are included in $\Delta^2(t)$, while the intrinsic $\langle r^2\rangle$ is obtained from a fit to an $I_{\rm inc}(\mathbf{Q},t)$ which includes all nuclei. We expect the latter difference to be minimal and it would be straightforward to include only the H nuclei in $I_{\rm inc}(\mathbf{Q},t)$ if desired. With these caveats, we have compared the $\Delta^2(t)/2$ with $\langle r^2\rangle$ in Figs. 9 and 10. At low temperatures (e.g., 100 K) where diffusion is expected to be less important, we find that $\Delta^2(t)/2$ appears to have converged after 10 ns and approaches $\langle r^2\rangle$ reasonably well. However, at higher temperatures (e.g., 250 K), $\Delta^2(t)/2$ has not converged to a constant after 10 ns and lies well below $\langle r^2\rangle$. From these comparisons it would be interesting to evaluate $\Delta^2(t)$ out to longer times to determine whether it converges and to reveal the dynamics contributing. For example, at 300 K, nearly translational diffusion may be possible for some H's in the protein that are near the surface or near hydration water. It would be interesting to exclude these

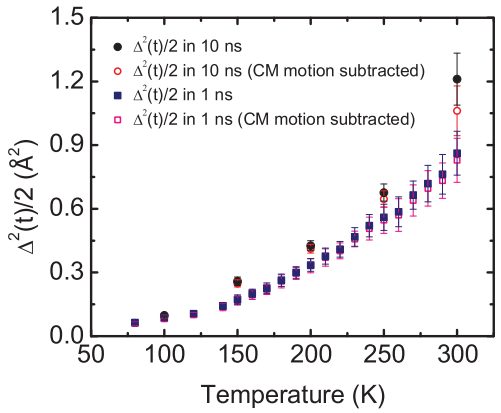

FIG. 18. (Color online) MSD $\Delta^2(t)/2$, with and without the CM motion subtracted after $t = 1$ ns (100-ns MD simulation) and $t = 10$ ns (1-μs MD simulation).

H's from $\Delta^2(t)$. In this regard it is also important to exclude the CM motion, which becomes important at higher temperatures and longer times as shown in Fig. 18.

VI. CONCLUSION

We have proposed a procedure to obtain the intrinsic, long-time MSD in proteins from finite-time simulations. The intrinsic MSD represents the equilibrium MSD, as would be predicted by statistical mechanics and the energy landscape, assuming that the protein does not go through major structural changes. The specific MSD investigated is the one determined in neutron scattering measurements. The intrinsic MSD is calculated from simulations of 100 ns and 1 μs and found to be independent of simulation time. The intrinsic, long-time MSD in lysozyme is found to be approximately twice the MSD that develops after a time of 1.5 ns, as would be observed using neutron instruments with an energy resolution width of $W = 1$ μeV. The intrinsic MSD shows the same breaks in slope with temperature as does the finite-time MSD. The ratio of the intrinsic to the finite-time MSD is sensitive to the model functions (e.g., stretched exponentials) used to describe the motions in the protein as well as to the decay times of the motions themselves.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge valuable discussions with Mark Johnson, Giuseppi Zaccai, and Dominique Bicout. This work was supported by the DOE, Office of Basic Energy Sciences, under Contract No. ER46680 (D.V. and H.R.G.) and by NSF Grant No. MCB-0842871 (L.H. and J.C.S.).

[1] W. Doster, S. Cusack, and W. Petry, Nature (London) **337**, 754 (1989).
[2] G. Zaccai, Science **288**, 1604 (2000).
[3] J. Fitter, R. E. Lechner, G. Buldt, and N. A. Dencher, Proc. Natl. Acad. Sci. USA **93**, 7600 (1996).
[4] S.-H. Chen, L. Liu, E. Fratini, P. Baglioni, A. Faraone, and E. Mamontov, Proc. Natl. Acad. Sci. USA **103**, 9012 (2006).
[5] J. H. Roh, V. N. Novikov, R. B. Gregory, J. E. Curtis, Z. Chowdhuri, and A. P. Sokolov, Phys. Rev. Lett. **95**, 038101 (2005).
[6] J. H. Roh, J. E. Curtis, S. Azzam, V. N. Novikov, I. Peral, Z. Chowdhuri, R. B. Gregory, and A. P. Sokolov, Biophys. J. **91**, 2573 (2006).
[7] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik, and G. Zaccai, Q. Rev. Biophys. **35**, 327 (2002).
[8] M. Tarek and D. J. Tobias, Biophys. J. **79**, 3244 (2000).
[9] S.-H. Chen, M. Lagi, X.-Q. Chu, Y. Zhang, C. Kim, A. Faraone, E. Fratini, and P. Baglioni, Spectroscopy **24**, 1 (2010).
[10] R. M. Daniel, J. C. Smith, M. Ferrand, S. Héry, R. Dunn, and J. L. Finney, Biophys. J. **75**, 2504 (1998).
[11] R. V. Dunn, V. Reat, J. Finney, M. Ferrand, J. C. Smith, and R. M. Daniel, Biochem. J. **346**, 355 (2000).
[12] R. M. Daniel, R. V. Dunn, J. L. Finney, and J. C. Smith, Annu. Rev. Biophys. Biomol. Struct. **32**, 69 (2003).
[13] J. C. Smith, Q. Rev. Biophys. **24**, 227 (1991).
[14] J. A. Rupley and G. Careri, Adv. Protein Chem. **41**, 37 (1991).
[15] W. Doster, S. Cusack, and W. Petry, Phys. Rev. Lett. **65**, 1080 (1990).
[16] M. Ferrand, A. J. Dianoux, W. Petry, and G. Zaccai, Proc. Natl. Acad. Sci. USA **90**, 9668 (1993).
[17] M. Jasnin, L. van Eijck, M. M. Koza, J. Peters, C. Laguri, H. Lortat-Jacob, and G. Zaccai, Phys. Chem. Chem. Phys. **12**, 3360 (2010).
[18] H. Nakagawa, H. Kamikubo, and M. Kataoka, Biochim. Biophys. Acta **1804**, 27 (2010).
[19] J. Smith, K. Kuczera, and M. Karplus, Proc. Natl. Acad. Sci. USA **87**, 1601 (1990).
[20] J. A. Hayward and J. C. Smith, Biophys. J. **82**, 1216 (2002).
[21] J. A. Hayward, J. L. Finney, R. M. Daniel, and J. C. Smith, Biophys. J. **85**, 679 (2003).
[22] J. A. Hayward, R. M. Daniel, J. L. Finney, and J. C. Smith, Chem. Phys. **292**, 389 (2003).
[23] V. Hamon, P. Calligari, K. Hinsen, and G. R. Kneller, J. Non-Cryst. Solids **352**, 4417 (2006).
[24] V. Calandrini, V. Hamon, K. Hinsen, P. Calligari, M.-C. Bellissent-Funel, and G. R. Kneller, Chem. Phys. **345**, 289 (2008).
[25] V. Calandrini and G. R. Kneller, J. Chem. Phys. **128**, 065102 (2008).
[26] T. E. Dirama, G. A. Carri, and A. P. Sokolov, J. Chem. Phys. **122**, 244910 (2005).
[27] T. E. Dirama, J. E. Curtis, G. A. Carri, and A. P. Sokolov, J. Chem. Phys. **124**, 034901 (2006).
[28] A. Lerbret, F. Affouard, P. Bordat, A. Hédoux, Y. Guinet, and M. Descamps, Chem. Phys. **345**, 267 (2008).
[29] Y. Miao, Z. Yi, D. C. Glass, L. Hong, M. Tyagi, J. Baudry, N. Jain, and J. C. Smith, J. Am. Chem. Soc. **134**, 19576 (2012).
[30] L. Meinhold and J. C. Smith, Phys. Rev. E **72**, 061908 (2005).
[31] L. Meinhold, J. C. Smith, A. Kitao, and A. H. Zewail, Proc. Natl. Acad. Sci. USA **104**, 17261 (2007).

[32] L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney, and J. C. Smith, Biophys. J. **94**, 4812 (2008).
[33] T. Becker and J. C. Smith, Phys. Rev. E **67**, 021904 (2003).
[34] Z. Yi, Y. Miao, J. Baudry, N. Jain, and J. C. Smith, J. Phys. Chem. B **116**, 5028 (2012).
[35] K. Wood, S. Grudinin, B. Kessler, M. Weik, M. Johnson, G. R. Kneller, D. Oesterhelt, and G. Zaccai, J. Mol. Biol. **380**, 581 (2008).
[36] D. Vural and H. R. Glyde, Phys. Rev. E **86**, 011926 (2012).
[37] M. Tarek, G. J. Martyna, and D. J. Tobias, J. Am. Chem. Soc. **122**, 10450 (2000).
[38] T. Becker, J. A. Hayward, J. L. Finney, R. M. Daniel, and J. C. Smith, Biophys. J. **87**, 1436 (2004).
[39] T. Róg, K. Murzyn, and G. R. Kneller, J. Comp. Chem. **24**, 657 (2003).
[40] P. J. Artymiuk, C. C. F. Blake, D. W. Rice, and K. S. Wilson, Acta. Cryst. B **38**, 778 (1982).
[41] L. Hong, N. Smolin, B. Lindner, A. P. Sokolov, and J. C. Smith, Phys. Rev. Lett. **107**, 148102 (2011).
[42] L. Hong, X. Cheng, D. C. Glass, and J. C. Smith, Phys. Rev. Lett. **108**, 238102 (2012).
[43] M. Lagi, X. Chu, C. Kim, F. Mallamace, P. Baglioni, and S. H. Chen, J. Phys. Chem. B **112**, 1571 (2008).
[44] A. Oleinikova, N. Smolin, I. Brovchenko, A. Geiger, and R. Winter, J. Phys. Chem. B **109**, 1988 (2005).
[45] B. Hess, C. Kutzner, D. V. Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
[46] W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).
[47] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, J. Chem. Phys. **120**, 9665 (2004).
[48] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[49] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comp. Chem. **18**, 1463 (1997).
[50] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[51] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[52] J. Smith, S. Cusack, U. Pezzeca, B. R. Brooks, and M. Karplus, J. Chem. Phys. **85**, 3636 (1986).
[53] S. Khodadadi, A. Malkovskiy, A. Kisliuk, and A. P. Sokolov, Biochim. Biophys. Acta **1804**, 15 (2010).
[54] A. L. Lee and A. J. Wand, Nature (London) **411**, 501 (2001).
[55] W. Doster, J. Non-Cryst. Solids **357**, 622 (2011).
[56] R. M. Daniel, J. L. Finney, V. Reat, R. Dunn, M. Ferrand, and J. C. Smith, Biophys. J. **77**, 2184 (1999).
[57] S. Khodadadi, S. Pawlus, J. H. Roh, V. G. Sakai, E. Mamontov, and A. P. Sokolov, J. Chem. Phys. **128**, 195106 (2008).
[58] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and R. Young, Proc. Natl. Acad. Sci. USA **101**, 14408 (2004).
[59] G. Schiro, F. Natali, and A. Cupane, Phys. Rev. Lett. **109**, 128102 (2012).