Biophysical Journal
Article

Biophysical Society

Determination of Dynamical Heterogeneity from Dynamic Neutron Scattering of Proteins

Derya Vural,[1,2,*] Jeremy C. Smith,[3] and Henry R. Glyde[1]
[1]Department of Physics and Astronomy, University of Delaware, Newark, Delaware; [2]Department of Physics, Giresun University, Giresun, Turkey; and [3]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee

ABSTRACT Motional displacements of hydrogen (H) in proteins can be measured using incoherent neutron-scattering methods. These displacements can also be calculated numerically using data from molecular dynamics simulations. An enormous amount of data on the average mean-square motional displacement (MSD) of H as a function of protein temperature, hydration, and other conditions has been collected. H resides in a wide spectrum of sites in a protein. Some H are tightly bound to molecular chains, and the H motion is dictated by that of the chain. Other H are quite independent. As a result, there is a distribution of motions and MSDs of H within a protein that is denoted dynamical heterogeneity. The goal of this paper is to incorporate a distribution of MSDs into models of the H incoherent intermediate scattering function, $I(Q,t)$, that is calculated and observed. The aim is to contribute information on the distribution as well as on the average MSD from comparison of the models with simulations and experiment. For example, we find that simulations of $I(Q,t)$ in lysozyme are well reproduced if the distribution of MSDs is bimodal with two broad peaks rather than a single broad peak.

INTRODUCTION

Joerg Langowski was a pioneer of combining simulation and experimental neutron-scattering techniques to understand the structure and dynamics of biological systems in solution (1–5). In this spirit, we present here a study aiming to understand and bring insight on dynamical heterogeneity (DH) in proteins through testing models against molecular dynamics simulation and neutron-scattering measurements. This study is dedicated to his memory.

Incoherent neutron-scattering methods and numerical simulation have been widely used to determine motional displacements in proteins (6–27). Enormous progress has been made in measuring, interpreting, and calculating mean-square motional displacements (MSDs) as a function of temperature, the level of hydration, and the level and type of solution and of applied pressure on the protein (10,20,25,28,29). Because the incoherent scattering cross section of hydrogen (H) is large compared with that of other nuclei, the observed MSDs are dominated by those of H. There are typically 1500 H in a protein. These H are in a wide variety of sites, executing a wide spectrum of motional displacements. In this study, our goal is to incorporate a distribution of H MSDs into models of the incoherent intermediate scattering function (ISF), $I(Q,t)$, and the corresponding incoherent dynamical structure factor (DSF), $S(Q,\omega)$, where Q and ω are the wave vector and the energy transfer in the scattering, respectively. The aim is to provide information on these distributions of MSDs by testing the models against numerical simulation of $I(Q,t)$ in lysozyme and $S(Q,\omega)$ observed using neutrons in human acetylcholinesterase (hAChE).

The incoherent ISF, $I(Q,t)$, which can be calculated directly from molecular dynamics (MD) simulations and from which the incoherent $S(Q,\omega)$ observed in neutron-scattering measurements can be obtained, is defined as follows (30,31):

$$I(Q,t) = \frac{1}{N}\sum_{j=1}^{N} I_j(Q,t) = \frac{1}{N}\sum_{j=1}^{N}\left\langle e^{-i\mathbf{Q}\cdot\mathbf{r}_j(t)} e^{i\mathbf{Q}\cdot\mathbf{r}_j(0)}\right\rangle. \quad (1)$$

We restrict the sum over j to a sum over the H in the protein.

Submitted October 26, 2017, and accepted for publication February 12, 2018.

*Correspondence: derya.vural@giresun.edu.tr

Editors: Tamar Schlick and Jill Trewhella.
https://doi.org/10.1016/j.bpj.2018.02.024

To construct a model, we note that $I(Q,t)$ may be formally separated into a time-independent part, $I(Q,t\rightarrow\infty)$, and a time-dependent part (30):

$$I(Q,t) = \frac{1}{N}\sum_{j=1}^{N}\big[I_j(Q,t=\infty) + \big(I_j(Q,t)-I_j(Q,t=\infty)\big)\big]. \quad (2)$$

The time-independent part,

$$E_\infty(Q)\equiv I(Q,t\rightarrow\infty) = \frac{1}{N}\sum_j I_j(Q,t\rightarrow\infty), \quad (3)$$

is denoted the elastic incoherent scattering function (EISF) of the protein. By following (32) and earlier work, we propose a simple model for $I(Q,t)$:

$$I_M(Q,t) = E_\infty(Q) + [1-E_\infty(Q)]C(t) = E_\infty(Q) + [1-E_\infty(Q)]\exp\big(-(\lambda t)^\beta\big). \quad (4)$$

The time-independent part, $E_\infty(Q)$, is exact. In the time-dependent part, the function $C(t)=\exp(-(\lambda t)^\beta)$ is represented by a "stretched" exponential. The model has the correct limits at $t=0$ ($I_j(Q,t=0)=1$ for all Q) and at $t=\infty$. It has a plausible time dependence in between. In Eq. 4, λ and β are parameters to be determined by a best fit of the model $I_M(Q,t)$ to a numerical simulation of $I(Q,t)$.

At least for the proteins staphylococcal nuclease (sNase), bovine pancreatic trypsin inhibitor, hen egg lysozyme, and T4 lysozyme, it has been shown by simulation (33,34) that the EISF of an individual H is well represented by the lowest-order Gaussian term, i.e.,

$$I_j(Q,t\rightarrow\infty) = e^{-Q^2\langle z_j^2\rangle}, \quad (5)$$

where $\langle z_j^2\rangle$ is the MSD of the individual H labeled j along the z axis chosen parallel to $\mathbf{Q}$, i.e., $\langle(\mathbf{Q}\cdot\mathbf{r}_j)^2\rangle = Q^2\langle z_j^2\rangle$. The Gaussian approximation of Eq. 5 is valid beyond low Q out to $Q\gtrsim 4$ Å^{-1} (33,34) for all H except for a few exceptional atoms. These few contribute little to the total EISF of the protein:

$$E_\infty(Q) = I(Q,t\rightarrow\infty) = \frac{1}{N}\sum_j e^{-Q^2\langle z_j^2\rangle}. \quad (6)$$

The total $E_\infty(Q)$ is, however, not a Gaussian because a sum of Gaussians is not a Gaussian unless all the $\langle z_j^2\rangle$ are the same.

In this study, our goal is to explicitly recognize that the MSDs $\langle z_j^2\rangle$ of the individual H are not the same throughout the protein and to propose models of the distribution of $\langle z_j^2\rangle$ in $E_\infty(Q)$. Through fits to simulations and data, the aim is to bring information on the type of distributions that reproduce simulations and experiment well. A model distribution $\rho(\langle z_j^2\rangle)$ can be introduced as the following (19,20,35):

$$E_\infty(Q) = \frac{1}{N}\sum_{j=1}^{N} e^{-Q^2\langle z_j^2\rangle} = \int ds\,\rho(s)\exp\big[-Q^2 s\big], \quad (7)$$

where $s=\langle z_j^2\rangle$. The $\rho(s)$ is a one dimensional (1D) distribution of MSDs along an axis (along Q). We find, for example, that a distribution that has a single broad peak of adjustable position and width does not reproduce numerical simulations of $I(Q,t)$ well. However, a bimodal distribution $\rho(s)$ does reproduce the numerical $I(Q,t)$ well.

Specifically, we substitute the model $E_\infty(Q)$ given by Eq. 7 into $I_M(Q,t)$ in Eq. 4. Parameters in $\rho(s)$ and the parameters λ and β are adjusted and determined by a best fit of $I_M(Q,t)$ to the simulated $I(Q,t)$ as a function of t. This is done for many Q values. A "good" distribution is identified as one that provides a good fit to $I(Q,t)$ and a distribution $\rho(s)$ and average MSD,

$$\langle r^2\rangle = 3\int ds\,\rho(s)s, \quad (8)$$

that is independent of Q. If the motions are isotropic, $\langle z_j^2\rangle = \langle r_j^2\rangle/3$.

In previous work (32,34,36) in which DH was ignored, we found that the best-fit average $\langle r^2\rangle$ depended on Q. That is, when $\rho(s)$ in Eq. 7 was approximated by $\rho(s)=\delta(s-s_0)$, where $s_0=\langle r^2\rangle/3$ so that $E_\infty(Q)\rightarrow I_\infty(Q)=e^{-(1/3)Q^2\langle r^2\rangle}$, we found that the best-fit $\langle r^2\rangle$ decreased with increasing Q (see Fig. 1). An MSD $\langle r^2\rangle$ that decreases with increasing Q is often found in fits to data when a model EISF $I_\infty(Q)=e^{-(1/3)Q^2\langle r^2\rangle}$ is assumed (6,16,37–40). A clear goal is also to determine whether a Q-independent distribution, $\rho(s)$, and MSD, $\langle r^2\rangle$, can be obtained when DH is incorporated.

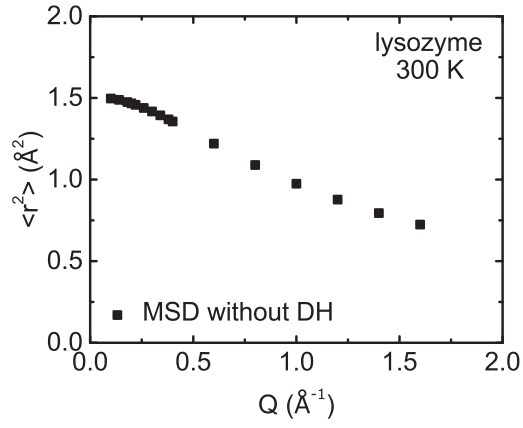

FIGURE 1 MSD in lysozyme obtained in (34) from fitting a model $I(Q,t)$, which neglects dynamical heterogeneity to $I(Q,t)$ calculated from Eq. 1 using molecular dynamics simulation.

The importance of DH in the analysis of neutron-scattering experiments and in the interpretation of MD simulations of proteins has been recognized since the first experiments (6,37,41). As discussed in the comprehensive review by Doster and Settles (29), many experiments were analyzed by assuming a single representative MSD. Inclusion of two MSDs in the EISF, a bimodal distribution, significantly improves the fit to the data (29). Computer simulations (42,43) showed that H in proteins have a broad distribution of MSDs and that analysis of neutron data using a single MSD underestimates the actual MSD. Incorporating a term proportional to Q^4 in the EISF, in addition to the Gaussian Q^2 term, in the analysis provides both an improved average MSD of the nonexchangeable H and a reliable estimate of the variance of the distribution (44). Significant DH was found to arise from methyl group rotations. Meinhold et al. (45) showed that the temperature dependence of the dynamics of mesophilic and thermophilic dihydrofolate reductase derived from elastic neutron-scattering data can be described using a Weibull distribution of MSDs, which is a distribution that resembles those seen in MD simulations. Kneller and Hinsen (19) and Peters and Kneller (20) have analyzed the EISF of hAChE using a Gamma distribution, while Nakagawa et al. (35) considered three model distributions in their analysis of the EISF of sNase. We compare these distributions with the one from our study in the Discussion.

In Materials and Methods, we outline the justification for the Gaussian expression, Eq. 5. We introduce the MSD and a distribution of MSDs calculated by simulation. The DSF that was observed using neutrons, including the instrument resolution that corresponds to $I(Q,t)$, is also presented. In Results, the model $I_M(Q,t)$ is compared with simulations of $I(Q,t)$ in lysozyme, and the model elastic incoherent DSF, $S(Q,\omega=0)$, is compared with $S(Q,\omega=0)$ observed in hAChE. In Discussion, the results are discussed, the present model distribution of MSDs is compared with the distributions found by others, and the future prospects for improving the distribution are presented.

MATERIALS AND METHODS

MD simulation of lysozyme

Two lysozyme molecules (1AKI (46)) were arbitrarily oriented and placed in a simulation box of dimensions 6.5 × 3.4 × 3.6 nm. The lysozyme molecules were surrounded by 636 water molecules corresponding to a 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. The system was simulated using the GROMACS 4.5.1 (47) (University of Groningen, Groningen, the Netherlands). The OPLS-AA force field (48) was used for the protein, and TIP4P (49) was used for the water. The van der Waals interactions were truncated at 1.4 nm, and the electrostatic interactions were represented using the particle mesh Ewald method (50) with a real-space cutoff of 0.9 nm. All bonds, including H bonds, were constrained with a linear constraints solver algorithm (51). The energy of the system was first minimized using 50,000 steepest descent steps. The system was then equilibrated in the mole-volume-temperature ensemble at each temperature investigated for 10 ns and in the mole-pressure-temperature ensemble at 1 bar for 10 ns. The Nosé-Hoover algorithm (52) with a coupling time $\tau = 1$ ps and the Parrinello-Rahman algorithm (53) with a coupling time $\tau = 3$ ps were used for the temperature coupling and pressure coupling, respectively.

Simulations of 100 ns in duration were performed at 300 K. The data were collected every 10 ps.

Model-scattering function including DH

The model incoherent scattering function, $I_M(Q,t)$, including DH in the EISF, $E_\infty(Q)$, was set out in Eq. 4. In this section, we 1) sketch how the expression Eq. 5 for the EISF of an individual H and Eq. 7. for $E_\infty(Q)$ in the model are obtained, 2) introduce MSDs obtained from MD simulations, and 3) present the DSF corresponding to $I_M(Q,t)$, including the neutron-instrument resolution that will be needed in Results.

The expression for $E_\infty(Q) \equiv I(Q, t\to\infty)$ can be obtained by taking the $t = \infty$ limit of $I(Q,t)$ defined in Eq. 1:

$$\begin{aligned} E_\infty(Q) &= \frac{1}{N}\sum_{j=1}^{N}\left\langle e^{-i\mathbf{Q}\cdot\mathbf{r}_j(\infty)}\right\rangle\left\langle e^{i\mathbf{Q}\cdot\mathbf{r}_j(0)}\right\rangle \\ &= \frac{1}{N}\sum_{j=1}^{N}\left\langle e^{-i\mathbf{Q}\cdot\mathbf{r}_j}\right\rangle\left\langle e^{i\mathbf{Q}\cdot\mathbf{r}_j}\right\rangle \\ &= \frac{1}{N}\sum_{j=1}^{N} e^{-\left\langle[\mathbf{Q}\cdot\mathbf{r}_j]^2\right\rangle} = \frac{1}{N}\sum_{j=1}^{N} e^{-Q^2\left\langle z_j^2\right\rangle} \end{aligned} \quad . \tag{9}$$

In these equations, we have used the notion that 1) at $t = \infty$, $r_j(\infty)$ and $r_j(0)$ are statistically independent, 2) $r_j(t)$ is translationally invariant in time (so that $r_j(\infty) = r_j(0) = r_j$) and 3) make a cumulant expansion, $\langle e^{i\mathbf{Q}\cdot\mathbf{r}}\rangle = e^{-(1/2)\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle}$, keeping only the second cumulant. The $\hat{\mathbf{z}}$ axis is chosen parallel to $\mathbf{Q}$. Using properties 1 and 2, we arrive at the second line of Eq. 9. The key approximation is keeping only the second cumulant. As noted above, explicit calculations of the higher cumulants show this is a surprisingly good approximation for an individual H in proteins (33,34) out to $Q \gtrsim 4$ Å^{-1}. The $\langle z_j^2\rangle$ have been shown (33) to be highly isotropic, $e^{-Q^2\langle z_j^2\rangle} = e^{-(1/3)Q^2\langle r_j^2\rangle}$, except for a few H. Using Eq. 9, Eqs. 5 and 6 follow directly.

In MD simulations, the MSD of a single H nucleus j after a time t can be calculated as

$$\Delta_j^2(t) = \left\langle\left[r_j(t) - r_j(0)\right]^2\right\rangle. \tag{10}$$

In Eq. 10, $r_j(t)$ is the position of the H nucleus j in the protein at time t. At $t\to\infty$, the correlation between the positions $r_j(t)$ and $r_j(0)$ is lost, as noted above, and the $\Delta_j^2(t)$ reduces to

$$\Delta_j^2(t=\infty) = \left\langle r_j^2(\infty) - r_j^2(0)\right\rangle = 2\left\langle r_j^2\right\rangle. \tag{11}$$

The average MSD of H in the protein, $\langle r^2\rangle_{MD}$, at long time can also be calculated as

$$\overline{\Delta}^2(t=\infty) = \frac{1}{N}\sum_{j=1}^{N}\Delta_j^2(t=\infty) = 2\left\langle r^2\right\rangle_{MD}. \tag{12}$$

In actual simulations, $\Delta_j^2(t)$ is calculated at long times (e.g., 1 ns), and it is again assumed that the correlations between $r_j(t)$ and $r_j(0)$ are lost.

A histogram of MSDs can be calculated as

$$\langle r^2\rangle_{MD} = \frac{1}{N}\sum_j \left\langle r_j^2(t)\right\rangle = \frac{1}{N}\sum_j \frac{1}{2}\Delta_j^2(t) = \int d\left(\frac{1}{2}\Delta^2(t)\right)\rho_{MD}\left(\frac{1}{2}\Delta^2(t)\right)\frac{1}{2}\Delta^2(t). \quad (13)$$

In Eq. 13, $\rho_{MD}((1/2)\Delta^2(t))$ is the histogram of MD MSDs. Fig. 2 shows the histogram for lysozyme at 300 K. This is a histogram of 1D MSDs along a radial variable (assumed to be isotropic). The time t (1 ns) is assumed to be long enough that the $\Delta_j^2(t)$ are approaching their long time, nearly stationary values (see Fig. 11 of (32)). The distribution $\rho_{MD}((1/2)\Delta^2(t))$ is on the same length scale as that of $\rho(s)$ but is not defined by the same relation as $\rho(s)$.

The model DSF corresponding to the model ISF, $I_M(Q,t)$, is defined as follows:

$$S(Q,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\,\exp(i\omega t) I_M(Q,t). \quad (14)$$

In a neutron-scattering measurement, we observe $S_R(Q,\omega)$, which is the $S(Q,\omega)$ broadened by the neutron instrument resolution function, $R(\omega)$:

$$S_R(Q,\omega) = \int_{-\infty}^{\infty} d\omega' S(Q,\omega')R(\omega-\omega'). \quad (15)$$

To calculate an $S_R(Q,\omega)$ for comparison with measurements on IN13 at the Institut Laue Langevin (Grenoble, France), we assume the following Gaussian instrument resolution function:

$$R(\omega) = \left[2\pi\sigma^2\right]^{-1/2}\exp\left(-\frac{1}{2}\frac{\omega^2}{\sigma^2}\right). \quad (16)$$

We set the width of the Gaussian at $\sigma = 1/\tau = (W/\hbar)/(8\ln 2)^{1/2}$, where W is the fullwidth at half-maximum of the energy resolution of IN13 and where $W = 8$ μeV, giving $\sigma \sim 0.005$ ps^{-1}.

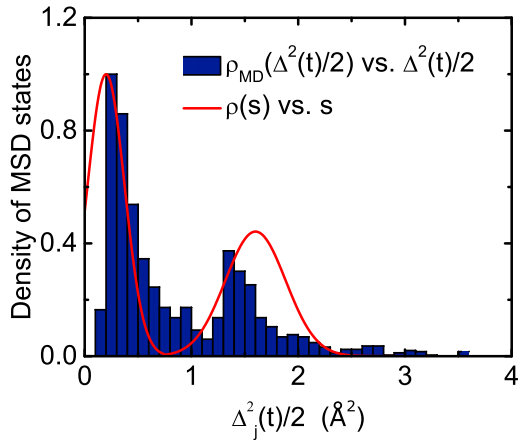

FIGURE 2 The histogram of hydrogen (H) mean-square displacements (MSDs) in lysozyme calculated from MD simulation, $\langle r^2\rangle_{MD} = \langle\Delta^2(t)\rangle/2$ at $t = 1$ ns, compared with the model $\rho(s)$ given by Eq. 23 (*red solid line*). The parameters in $\rho(s)$ are chosen to get the best fit to the histogram: $s_1 = 0.25$ Å^2, $s_L = 0.76$ Å^2, $\Delta s = 1.4$ Å^2, $s_2 = s_1 + \Delta s$, $\sigma_1^2 = 0.03$ Å^4, and $\sigma_2^2 = 0.08$ Å^4. To see this figure in color, go online.

Substituting $S(Q,\omega)$ from Eq. 14 into Eq. 15 and expressing $R(\omega)$ in terms of its Fourier transform $R(t)$ leads to

$$S_R(Q,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\,\exp(i\omega t) I_M(Q,t)R(t), \quad (17)$$

where $R(t) = \exp(-(1/2)\sigma^2t^2)$. The elastic limit of Eq. 17 is

$$S_R(Q,\omega=0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt I_M(Q,t)R(t). \quad (18)$$

Given $I_M(Q,t)$, this is the expression we use to evaluate $S_R(Q,\omega=0)$ for comparison with experiment.

$S_R(Q,\omega=0)$ in the limit of high instrument resolution ($\sigma^2\to 0$), which is often used in presenting data, can be obtained readily from Eq. 18. For example, substituting $I_M(Q,t)$ given by Eq. 4 into Eq. 18 gives the following:

$$S_R(Q,\omega=0) = E_\infty(Q)\left(2\pi\sigma^2\right)^{-1/2} + [1-E_\infty(Q)] \times \frac{1}{2\pi}\int_{-\infty}^{\infty} dtC(t)\exp\left(-\frac{1}{2}\sigma^2t^2\right). \quad (19)$$

Since $E_\infty(Q) = 1$ at $Q = 0$, the $Q\to 0$ limit of $S_R(Q,\omega=0)$ is $S_R(Q=0,\omega=0) = (2\pi\sigma^2)^{-1/2}$. $S_R(Q,\omega=0)$ normalized by $S_R(Q=0,\omega=0)$, the form in which data is often presented, is therefore defined as the following:

$$\frac{S_R(Q,\omega=0)}{S_R(Q=0,\omega=0)} = E_\infty(Q) + [1-E_\infty(Q)]\left(2\pi\sigma^2\right)^{1/2} \times \frac{1}{2\pi}\int_{-\infty}^{\infty} dtC(t)\exp\left(-\frac{1}{2}\sigma^2t^2\right). \quad (20)$$

In the limit of high energy resolution ($\sigma\to 0$), where $S_R(Q,\omega=0)\to S(Q,\omega)$, the normalized $S(Q,\omega)$ is defined as the following:

$$\frac{S(Q,\omega=0)}{S(Q=0,\omega=0)} = E_\infty(Q) = I(Q,t=\infty). \quad (21)$$

That is, in the limit of high resolution, the normalized elastic DSF, $S(Q,\omega=0)/S(Q=0,\omega=0)$, is the EISF, $I(Q,t=\infty)$ (6,19,28,33,36). Eq. 21 can be derived generally without the use of a model $I(Q,t)$ or $R(\omega)$. In Eq. 21, the EISF (6) is often approximated as $E_\infty(Q)\simeq e^{-(1/3)Q^2\langle r^2\rangle}$ to provide a simple, transparent presentation of data in terms of $\langle r^2\rangle$:

$$\frac{S(Q,\omega=0)}{S(Q=0,\omega=0)} \simeq e^{-(1/3)Q^2\langle r^2\rangle}. \quad (22)$$

Use of Eq. 22 neglects DH and assumes infinitely high instrument resolution so that $\langle r^2\rangle$ in principle is the long-time limit of the MSD. However, in practice, because the instruments have a finite energy resolution, a fit of Eq. 22 to data yields an $\langle r^2\rangle$ that has developed over a finite time only.

RESULTS

Model distribution of MSDs

In Model Scattering Function Including DH, we proposed a model of the incoherent ISF given by Eqs. 4 and 7 that

contains DH. The DH was represented by a distribution of MSDs, $\rho(s)$, where $s = \langle z_j^2 \rangle$ is a continuous variable representing the 1D MSD along an axis (z) chosen parallel to the wave-vector transfer Q. Our goal is to determine $\rho(s)$ from fits of the model to MD simulations and to neutron-scattering (NS) data with the criteria that there should be a good fit and that a distribution $\rho(s)$ and an average MSD that is independent of Q should be obtained.

If $\rho(s)$ is simply a delta function $\rho(s) = \delta(s - s_0)$ with $s_0 = \langle r^2 \rangle/3$, then the DH model, Eqs. 4 and 7, reduces to a model without DH: Eq. 11 of (32). The model without DH used previously led to the Q-dependent $\langle r^2 \rangle$ shown in Fig. 1. A simple model for $\rho(s)$ beyond a delta function is a Gaussian, as set out in Appendix A. With a Gaussian $\rho(s)$, we obtain a reasonable fit of the model $I_M(Q,t)$ to the simulated $I(Q,t)$, but the average MSD and parameters in the Gaussian are Q-dependent (see Appendix A). An extension is a Gaussian $\rho(s)$ with a tail reaching to higher s values, as discussed in Appendix B. This model also provided a good fit but again with Q-dependent parameters.

Guided by the histogram shown in Fig. 2, we introduce a bimodal $\rho(s)$ consisting of two Gaussians, given by

$$\rho(s) = \begin{cases} \exp\left(-(s-s_1)^2/2\sigma_1^2\right) & 0 \le s \le s_L \\ A\exp\left(-(s-s_2)^2/2\sigma_2^2\right) & s_L < s \le \infty \end{cases}, \quad (23)$$

where $A = \exp(-(s_L - s_1)^2/2\sigma_1^2)\exp((s_L - s_2)^2/2\sigma_2^2)$ and $s_2 = s_1 + \Delta s$.

The parameters in the model are s_1, Δs, s_L, σ_1, and σ_2. A normalized distribution $\rho_N(s)$ was calculated as $\rho_N(s) = \rho(s)/\int ds\, \rho(s)$. This model with parameters chosen to represent lysozyme is shown in Fig. 2. In the following sections, we show fits of the model $I_M(Q,t)$ obtained using the bimodal $\rho(s)$ to MD simulations and to NS data.

Comparison with MD simulation

The model ISF $I_M(Q,t)$ given by Eqs. 4 and 7 using the distribution of MSDs $\rho(s)$ defined in Eq. 23 is fitted to the $I(Q,t)$ calculated from Eq. 1 using MD simulation. Equation 1 is calculated out to 1 ns using an $r_j(t)$ generated in an MD simulation of lysozyme out to 100 ns. Fig. 3 (*upper frame*) shows the fit of Eq. 4 to the numerical $I(Q,t)$. The model parameters s_1, s_L, and λ were treated as free parameters adjusted to obtain the best fit, giving $s_L = 0.68$ Å^2 and s_1 and λ shown in Figs. 3 and 4. The other parameters were fixed at $\beta = 0.24$, $\Delta s = 1.4$ Å^2, $s_2 = s_1 + \Delta s$, $\sigma_1^2 = 0.03$ Å^4, and $\sigma_2^2 = 0.08$ Å^4. The fit is good except at the highest Q values. Fig. 3 (*middle*) shows $\rho(s)$ as a function of s for several Q values. The weight in the second peak of $\rho(s)$ increases if s_1 increases. Fig. 3 (*bottom*) shows the parameter s_1 as a function of Q. Clearly, $\rho(s)$ and s_1 change little with Q.

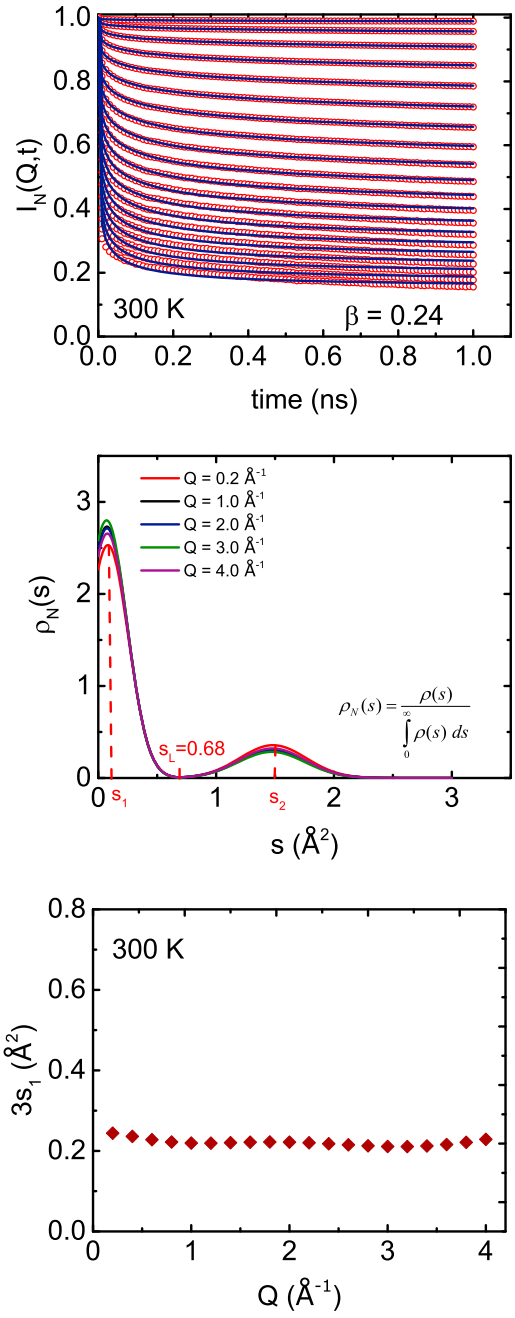

FIGURE 3 (*Top*) The ISF $I(Q,t)$ calculated using Eq. 1 from an MD simulation of lysozyme (*open circles*) and a fit of the model $I_M(Q,t)$ given by Eqs. 4, 7, and 23 (*solid line*). Shown is $I(Q,t)$ at 20 Q values from 0.2 to 4 Å^{-1} in steps of 0.2 Å^{-1}, top to bottom. (*Middle*) The $\rho_N(s)$ vs. s for the Q values is shown: 0.2, 1.0, 2.0, 3.0, and 4.0 Å^{-1}. (*Bottom*) The best-fit parameter $3s_1$ vs. Q is shown. To see this figure in color, go online.

In Fig. 4 (*top*), we compare the average MSD of lysozyme at 300 K obtained from Eq. 8 including DH and with the MSD obtained excluding heterogeneity shown in Fig. 1. Although there is some variation of $\langle r^2 \rangle$ with Q when DH is included, the value of $\langle r^2 \rangle$ at $Q \simeq 0$ and at $Q = 4$ Å^{-1} are approximately the same. Hence, an MSD that is independent of Q within 10% and that has no systematic

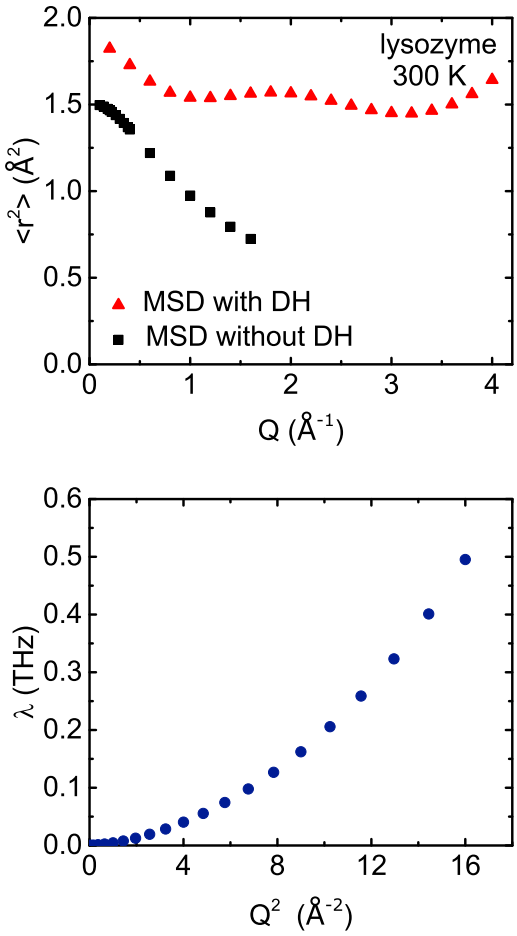

FIGURE 4 (*Top*) The average MSD obtained from fitting a model that includes DH (Eqs. 4, 7, and 23) to simulations of $I(Q,t)$ for lysozyme (*triangles*) and the MSD obtained from fitting a similar model (34) without DH (*squares*). (*Bottom*) The parameter λ obtained for the model including DH for the fits is shown in Fig. 3 (*top*). To see this figure in color, go online.

Q dependence can be obtained by including DH. In contrast, the MSD without DH clearly decreases uniformly with increasing Q. The bottom frame in Fig. 4 shows the λ values obtained from the fits in Fig. 3.

Comparison with NS measurements

In this section, we test the model $I_M(Q,t)$ including DH, Eq. 4, by calculating the corresponding elastic DSF, $S_R(Q,\omega=0)$, and comparing it with NS measurements made on hydrated powders of the protein hAChE. The measurements were made by Peters et al. (20) using IN13 at the Institut Laue Langevin. The model $S_R(Q,\omega=0)$, given by Eq. 17, was fitted to the observed $S_R(Q,\omega=0)$ as a function of Q. The model distribution of MSDs, $\rho(s)$, is again given by Eq. 23. In the fit of the model $I_M(Q,t)$, the parameters, λ, β, s_1, and s_L in $\rho(s)$ were treated as free parameters to be determined by the best fit to the NS data. The best fit s_L was 0.70 Å^2 rather than 0.68 Å^2 for lysozyme. The remaining parameters in $\rho(s)$ were held at the same values as those used for lysozyme: $\Delta s = 1.4$ Å^2, $s_2 = s_1 + \Delta s$, $\sigma_1^2 = 0.03$ Å^4, and $\sigma_2^2 = 0.08$ Å^4.

Fig. 5 shows the data and the fitted $S_R(Q,\omega=0)$ and the fitting parameters $3s_1$ and λ at five temperatures. The fit to the data is clearly good. The best-fit stretched exponential parameters β were 0.22, 0.24, 0.22, 0.28, and 0.24 at 220,

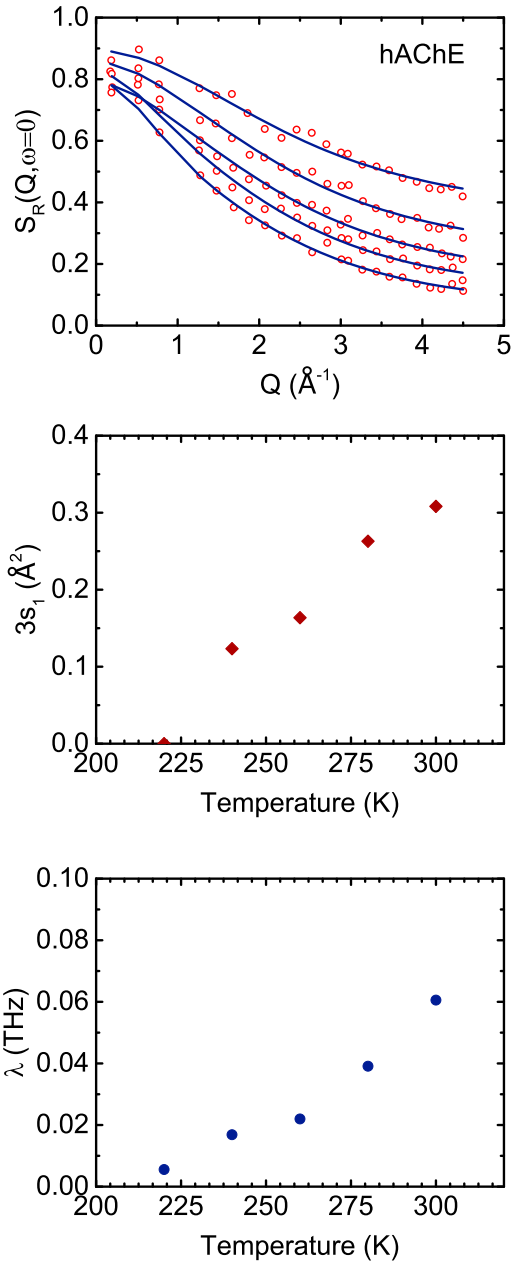

FIGURE 5 (*Top*) The $S_R(Q,\omega=0)$ observed by Peters et al. (20) in hAChE (*open circles*) and the fit of our model (Eqs. 4, 7, 18, and 23) (*solid line*), which includes DH to the observed $S_R(Q,\omega=0)$. From top to bottom, the temperatures are 220, 240, 260, 280, and 300 K. (*Middle* and *bottom*) The parameter $3s_1$ and the relaxation parameter λ obtained from fits of the model $S_R(Q,\omega=0)$ to the experimental data are shown. To see this figure in color, go online.

240, 260, 280, and 300 K, respectively, which are all close to the best-fit value of $\beta = 0.24$ found from fits to simulations of lysozyme at 300 K. The remaining parameters in $\rho(s)$ were the same as those used for lysozyme, as shown in Fig. 2. The best-fit values of $3s_1$ and λ clearly depend on temperature, as expected.

Fig. 6 shows the distribution of MSDs, $\rho_N(s)$, and the average MSD $\langle r^2 \rangle$ defined in Eq. 8 of hAChE at five temperatures. The temperature dependence of $\rho_N(s)$ follows from that of s_1. As s_1 increases with temperature, the population of the second peak at large MSDs increases.

The averaged MSD obtained from Eq. 8 and the MSD obtained from Eq. 22 are compared in Fig. 6. At temperatures higher than 250 K, the averaged MSD obtained from Eq. 8 including DH is larger than that obtained when neglecting DH. Incorporating DH leads to a more pronounced dynamical transition in the MSD at $T \simeq 250$ K.

DISCUSSION

In this study, we have proposed a model of the ISF of H in lysozyme, $I_M(Q, t)$, that includes DH. The DH is represented by a distribution of H MSDs in the protein. The distribution is used to calculate the EISF, $E_\infty(Q)$, given by Eq. 7, that appears as part of the model ISF. The model is fitted to numerical values of the ISF calculated from MD simulations of lysozyme to assess how well the model can describe the ISF.

If the distribution of MSDs is bimodal (two peaks), as shown in Fig. 2, then we find that we can reproduce the numerical $I(Q,t)$ well and obtain a model distribution $\rho(s)$ and an average MSD, $\langle r^2 \rangle$, that is independent of Q for $0 < Q < 4$ Å^{-1} within 10%. There is no systematic dependence of the average MSD on Q. In particular, $\langle r^2 \rangle$ at $Q \simeq 0$ and $Q \simeq 4$ Å^{-1} are approximately the same. The roughly Q-independent MSD are obtained, including DH. $\langle r^2 \rangle \simeq 1.5$ Å^2 between $0.5 < Q < 0.35$ Å^{-1} is approximately equal to the MSD obtained at $Q \to 0$ using models that ignore DH (see Fig. 4 (*top*)). With other distributions, e.g., a single Gaussian or a Gaussian with a tail, the best-fit distributions were a function of Q.

The same bimodal model distribution of MSDs with minor variations was used to calculate the EISF of hAChE. Good agreement with neutron measurements (20) was obtained. This suggests that a bimodal distribution may be common to the two proteins, but see the discussion below.

Comparison of distributions

In their extensive review, Doster and Settles (29) noted that a bimodal distribution of MSDs of the form $\rho(s) = A_1\delta(s - z_1^2) + A_2\delta(s - z_2^2)$, where A_1 to z_2 are parameters, significantly improved fits to observed EISFs over a single MSD model. They also extracted a general distribution that had two broad peaks from the observed incoherent ISF. Nakagawa et al. (35) defined a distribution of MSDs, $\rho(s)$, in terms of the EISF in exactly the form used here in Eq. 7. They considered a bimodal $\rho_B(s)$ (sum of two delta functions as above), a Gaussian $\rho_G(s)$, and an exponential $\rho_E(s)$. The EISF, $E_\infty(Q)$, of sNase was calculated and compared with the experiment. The $\rho_B(s)$ and $\rho_G(s)$ reproduced the observed EISF equally well and the $\rho_E(s)$ somewhat less well.

Kneller and Hinsen (19) and Peters and Kneller (20) define a distribution of MSDs also in exactly the form of Eq. 7. They use a Gamma distribution, $\rho(\lambda) = (\beta\lambda)^\beta exp[-\beta\lambda]/\lambda\Gamma(\beta)$, where $\Gamma(\beta)$ is the Gamma function. In $\rho(\lambda)$, $\lambda = s/\sigma^2$ is a dimensionless variable, σ is a convenient length parameter, and β and σ are adjustable parameters. For $\beta = \infty$, $\rho(\lambda) = \delta(\lambda - 1)$ (a delta function) for $\beta \simeq 10$, $\rho(\lambda)$ is approximately Gaussian with a tail, and as β goes to zero, $\rho(\lambda)$ approaches an exponential. They also employ a shifted version in which λ starts at a small finite value rather than zero. They are able to

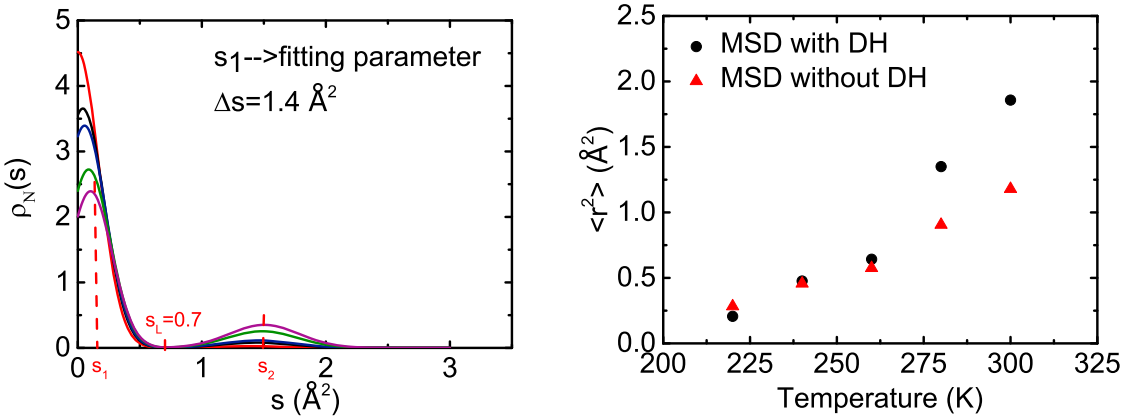

FIGURE 6 (*LHS*) The normalized distribution of MSDs, $\rho_N(s)$, obtained by fitting Eq. 18 to the experimental $S_R(Q, \omega = 0)$ for hAChE shown in Fig. 5 for five different temperatures: 220, 240, 260, 280, and 300 K. The height of the second peak in $\rho_N(s)$ is largest at 300 K. (*RHS*) The MSD obtained using Eq. 8 including DH (*solid circles*) and the MSD without DH obtained by fitting Eq. 22 to the observed $S_R(Q, \omega = 0)$ in the Q-range are shown: 0.2–0.8 Å^{-1} (*solid triangles*). To see this figure in color, go online.

get a good fit to the EISF of hydrated hAChE out to $Q = 4.5$ Å^{-1}. They find that the increase in the average MSD at the dynamical transition (DT) is larger when a distribution of MSDs is used. For example, the distribution broadens (β decreases) as temperature is increased through the DT approaching distributions found in simulations at higher temperatures (42).

In this study, in addition to fitting a model EISF to experiment or a simulation, we have calculated a model ISF as a function of time for a range of Q values and fitted the model to numerical simulations of ISF. We have required that the distribution $\rho(s)$ and the average MSD, as physical quantities, be independent of Q. Only the bimodal distribution with each component represented by a Gaussian met this requirement. The $\rho(s)$ was guided by and is quite similar to the distribution generated by the MD simulation in Fig. 2.

Improved distributions

This model $\rho(s)$ can be significantly improved. The present parameterization is somewhat rigid. For example, much of the increase of the average MSD with increasing temperature arises from an increase in occupation of the second peak at larger MSD rather than adjustment of the peak positions or widths. Also, we have introduced DH only in the MSD. The relaxation parameter λ in the model ISF is still assumed to be the same for all the H, whereas a distribution of λ is anticipated. However, these results show that a physically realistic, Q-independent average MSD can be obtained if DH is incorporated in the model EISF.

Including DH also offers new opportunities. For example, in addition to a obtaining a single MSD from a measurement, information on the distribution of MSDs within a protein and its hydration water can be obtained. Similarly, the distribution of MSDs may be modified and changed by changing the hydration level or by adding solutes to the hydration water. In Fig. 6 (*right-hand side* (*RHS*)), we see that the DT is more pronounced if the data is analyzed with a more flexible model that incorporates DH. That is, when DH is incorporated, the average MSD is found to be larger at higher temperatures (20,42). Thus, the motional displacements that are especially sensitive to temperature and may lead to function above T_D may be specifically identified and more precisely represented when DH is incorporated. Including a distribution of MSDs in the analysis of neutron data of other large molecules may also be rewarding.

Heterogeneity of H in lysozyme

Returning to lysozyme, Fig. 2 shows a histogram of the MSDs of the individual H in lysozyme obtained from MD simulation, $\langle r^2\rangle_{MD} = \langle \Delta^2(t)\rangle/2$ at $t = 1$ ns, as discussed in Materials and Methods. The histogram may be divided into three parts: a large peak at low MSDs ($0 < \langle r^2\rangle_{MD} <$ 1 Å^2) that we denote "Peak 1," a second smaller peak at larger MSDs ($1 < \langle r^2\rangle_{MD} < 2$ Å^2) ("Peak 2"), and a tail reaching to larger MSDs ($\langle r^2\rangle_{MD} > 2$ Å^2) that we denote the "Tail." An interesting question is which H in lysozyme have MSDs that contribute to "Peak 1," "Peak 2," and the "Tail." Table 1 lists the number of H with MSDs in each part of the histogram from each molecular group. For example, Table 1 shows that almost all of the H in *NH* (270 of 288 H) contributes to "Peak 1." Indeed, the MSDs in "Peak 1" arise chiefly from H in *NH*, *CH*, and CH_2. In contrast, few H (only 5 out of 408) in NH_3 and CH_3 have small MSDs that contribute to "Peak 1." Particularly interesting is that all of the H in the methyl group CH_3 have large MSDs that contribute to either "Peak 2" or the "Tail." None of the H in CH_3 contributes to "Peak 1." Thus, there is substantial heterogeneity in the dynamics of H, with H in different units having quite different MSD values (e.g., 94% of H in *CH* have small MSDs, and all the H in CH_3 have large MSDs).

Because the H in CH_3 have large MSDs at 300 K, these H contribute significantly to the average MSD of the whole protein at 300 K. The onset of large amplitude average MSDs at higher temperatures in proteins (e.g., at the DT, $T_D \simeq$ 180–220 K, and at somewhat lower temperatures) has been identified at least in part with the onset of methyl group rotations (54). Certainly, Table 1 shows that H in CH_3 contribute large MSDs to the average at 300 K. In future work, it would be interesting to determine the histogram of MSDs at lower temperatures and to further identify what fraction of the increase in the average MSD with increasing temperature can be attributed to H in CH_3 and other units.

CONCLUSION

To close, we return to an issue of long-standing and continuing interest: is the EISF, Eq. 21, a Gaussian in Q (i.e., $I(Q, t = \infty) = e^{-Q^2\langle z^2\rangle}$, where $\langle z^2\rangle$ is a constant) or not, and if not, why not? Nakagawa et al. (35) and Smith and collaborators (41–45) explicitly introduced the DH of H and a distribution of MSDs to explain and reproduce non-Gaussian EISFs. The EISF (see Eq. 9) is the sum of

TABLE 1 The Histogram of the MSDs of H in Lysozyme Calculated Using MD Has Two Peaks and a Tail at Higher MSD

Unit	Total	Peak 1	Peak 2	Tail
NH	288	270	16	2
NH_2	156	62	64	30
NH_3	42	5	24	13
CH	406	380	22	4
CH_2	620	525	84	11
CH_3	366	0	293 (80%)	73 (20%)
OH	40	26	12	2
Total	1918	1268	515	135

This table lists the number of H with MSDs in Peak 1 (largest peak), in Peak 2 (smaller peak), and in the Tail. Of the 620 H in CH_2, 508 have MSDs that lie in Peak 1, 82 have MSDs in Peak 2, and 30 have MSDs in the Tail.

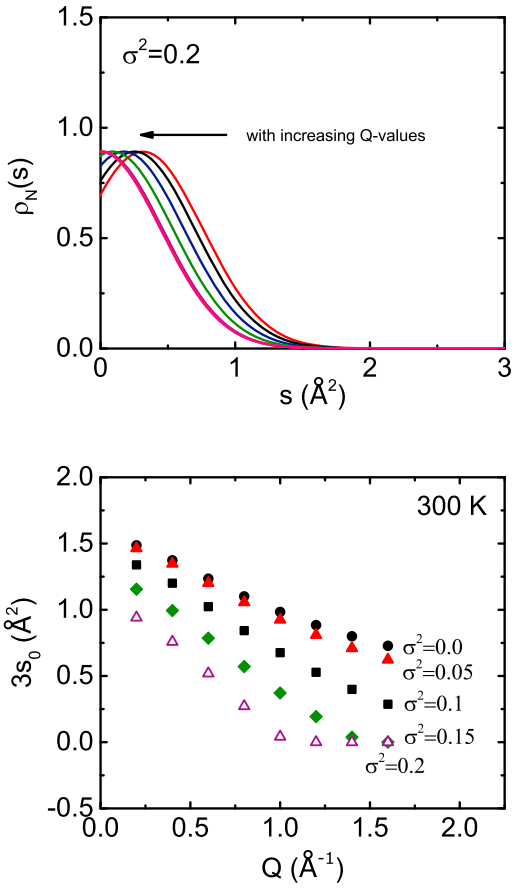

FIGURE 7 (*Top*) The best-fit Gaussian $\rho_G^N(s)$ at eight Q-values, from 0.2 to 1.6 Å^{-1} in steps of 0.2 Å^{-1}, with σ set at $\sigma^2 = 0.2$ Å^4. (*Bottom*) The best-fit s_0 in Eq. 24 is shown at eight Q values with σ^2 set at the values shown (in Å^4). To see this figure in color, go online.

the Debye-Waller factors (squared) of the individual constituents (dominated by H) of the protein, $I(Q, t = \infty) = (1/N)\sum_{j=1}^{N}\langle e^{-i\mathbf{Q}\cdot\mathbf{r}_j}\rangle\langle e^{i\mathbf{Q}\cdot\mathbf{r}_j}\rangle$. Nakagawa et al. (35) wrote the sum in the form of Eq. 7 and calculated the EISF of sNase from it. They found that the non-Gaussian EISF of sNase could be well reproduced using both a two-MSD (two–Debye-Waller factor) model or a model Gaussian distribution of MSDs. Similarly, Peters and Kneller (20) found that the observed, non-Gaussian EISF of hAChE shown in Fig. 5 could be reproduced using Eq. 7 and a Gamma distribution for $\rho(s)$. Clearly, the observed, non-Gaussian EISF can be reproduced using the two-state model or by introducing DH and using quite different distributions of MSDs, $\rho(s)$. Thus, whereas the EISF is critically important, its shape is not sufficiently unique to distinguish well between different models of DH.

For this reason, we turned to the full $I(Q,t)$ as a function of time at several Q values to investigate models of DH. In summary, we found that a two-MSD or two–Gaussian distribution $\rho(s)$ that reproduces the EISF well could not reproduce $I(Q,t)$ well at several Q values. Only the $\rho(s)$ shown in Fig. 2 was able to reproduce the MD-simulated $I(Q,t)$ well. In the future, we hope to find other functions that can be used to distinguish clearly between different models of DH and to improve the model of $\rho(s)$.

APPENDIX A: GAUSSIAN DISTRIBUTION FUNCTION

In this appendix, we consider a simple Gaussian representation of the distribution of the MSDs, $\rho(s)$, appearing in Eq. 7:

$$\rho_G(s) = \sqrt{\frac{1}{2\pi\sigma^2}}\exp\left[-\frac{(s-s_0)^2}{2\sigma^2}\right], \tag{24}$$

where s_0 and σ are parameters. Because the MSDs must be positive, we limit $\rho_G(s)$ to lie between 0 and ∞ and normalize it as

$$\rho_G^N(s) = \frac{1}{A}\rho_G(s), \tag{25}$$

where $A = \int_0^\infty ds\,\rho(s)$.

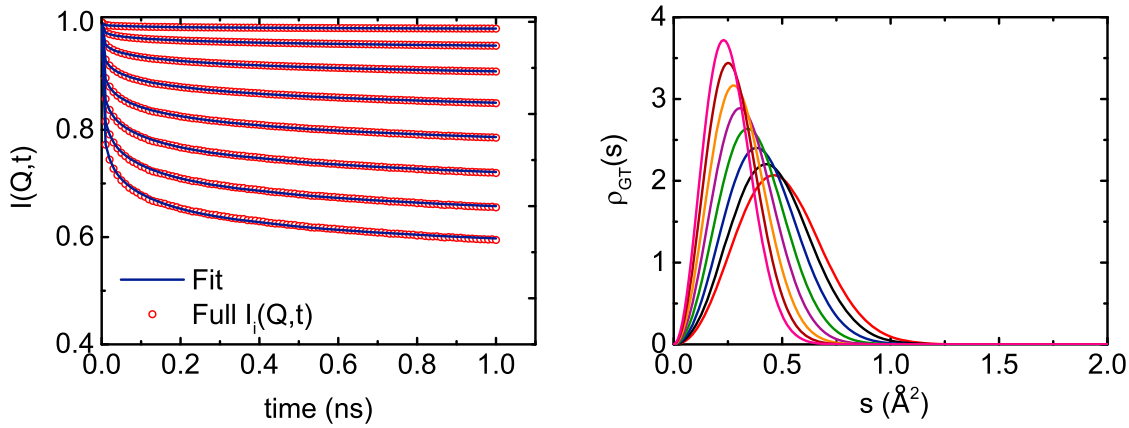

FIGURE 8 (*LHS*) $I(Q,t)$ of lysozyme calculated using MD simulation (*open circles*) and a fit of the model $I_M(Q,t)$, Eq. 4, obtained using Eq. 7 and $\rho_{GT}(s)$ in Eq. 26 to the simulated $I(Q,t)$ for eight Q values, 0.2 to 1.6 Å^{-1} from top to bottom. (*RHS*) The corresponding best-fit $\rho_{GT}(s)$ is shown. The broadest $\rho_{GT}(s)$ fits $Q = 0.2$ Å^{-1}. To see this figure in color, go online.

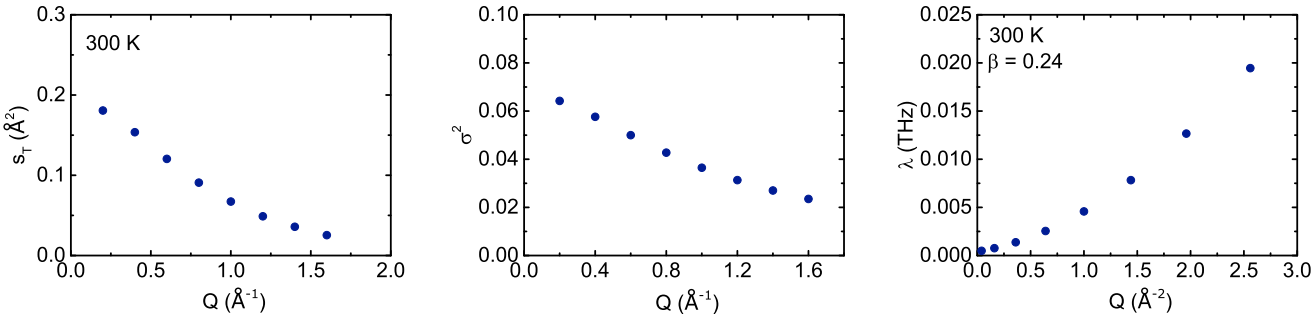

FIGURE 9 The best-fit parameters s_T, σ^2, and λ obtained from fits shown in Fig. 8. To see this figure in color, go online.

We fit the model $I_M(Q,t)$ of Eq. 4, obtained using Eq. 7 and the Gaussian $\rho_G^N(s)$ in Eq. 25, to the MD simulation of $I(Q,t)$ for lysozyme at 300 K. To display the dependence on s_0, s_0 is treated as a free-fitting parameter, and σ^2 is held fixed at 0.2 Å^4. Fig. 7 (*top*) shows the resulting best-fit distribution of MSDs, $\rho_G^N(s)$, at eight Q values. The $\rho_G^N(s)$ is clearly Q dependent. Fig. 7 (*bottom*) shows the best-fit values of s_0 as a function of Q for several σ^2 values. The best-fit s_0 clearly decreases with increasing Q. Allowing σ to vary in the fit does not change this basic character. A simple Gaussian is not a good representation of $\rho(s)$.

APPENDIX B: GAUSSIAN DISTRIBUTION WITH TAIL

We consider the Gaussian distribution $\rho_G(s)$ proposed in Appendix A multiplied by a factor of s^2; specifically,

$$\rho_{GT}(s) = A\sqrt{\frac{1}{2\pi\sigma^2}}s^2\exp\left[-\frac{(s-s_T)^2}{2\sigma^2}\right], \qquad (26)$$

where A is again a normalization constant. The aim is to extend the Gaussian $\rho_G(s)$ to higher s (a tail) and to have a function that goes to zero at $s = 0$. For this distribution, $E_\infty(Q)$ in Eq. 7 is defined as the following:

$$E_\infty(Q) = \int ds\, \rho_{GT}(s)\exp\left[-Q^2 s\right]$$

$$= A\sqrt{\frac{1}{2\pi\sigma^2}}\int_0^\infty dss^2\exp\left[-\frac{(s-s_T)^2}{2\sigma^2}\right]\exp\left[-Q^2 s\right]$$

$$= A\sqrt{\frac{1}{2\pi\sigma^2}}\exp\left(-s_T Q^2 + \frac{2Q^4\sigma^2}{4\alpha}\right) \qquad (27)$$

$$\int_0^\infty dss^2\exp\left[-\frac{\left(\left(s-\left(s_T-\frac{Q^2}{2\alpha}\right)\right)^2\right.}{2\sigma^2}\right]$$

The model $I_M(Q,t)$, Eq. 4, calculated using $\rho_{GT}(s)$, is fitted to the simulated $I(Q,t)$ of lysozyme at 300 K with s_T, σ, and λ treated as free-fitting parameters. The β was fixed at 0.24. As shown in Fig. 8 (*left-hand side* (*LHS*)), the fit is good at all Q. However, as seen in Fig. 8 (*RHS*), the model $\rho_{GT}(s)$ depends on Q. It is peaked at lower s and is much narrower at higher Q values. The best-fit values of s_T, σ^2, and λ are shown in Fig. 9. The best-fit values s_T and σ^2 clearly depend on Q. Thus, although a good fit to the MD simulation of $I(Q,t)$ can be obtained, the model and average MSDs extracted remain Q dependent, as found when DH is neglected entirely.

AUTHOR CONTRIBUTIONS

D.V., J.C.S., and H.R.G. designed the research and wrote the manuscript. D.V. performed the calculations.

ACKNOWLEDGMENTS

H.R.G. gratefully acknowledges the hospitality of the Theory Group at the Insitut Laue Langevin, where part of this manuscript was written.

This work was supported by the Department of Energy, Office of Basic Energy Sciences under contract number ER46680 (D.V. and H.R.G.) and by the Genomic Science Program, Office of Biological and Environmental Research, United States Department of Energy under contract FWP ERKP752 (J.C.S.).

REFERENCES

1. Timmins, P. A., J. Langowski, and R. S. Brown. 1988. An elongated model of the Xenopus laevis transcription factor IIIA-5S ribosomal RNA complex derived from neutron scattering and hydrodynamic measurements. *Nucleic Acids Res.* 16:8633–8644.
2. Krueger, S., G. Zaccai, …, M. Gellert. 1990. Neutron and light-scattering studies of DNA gyrase and its complex with DNA. *J. Mol. Biol.* 211:211–220.
3. Lindner, P., H. R. Bewersdorff, …, R. Oberthr. 1990. Drag-reducing surfactant solutions in laminar and turbulent flow investigated by small-angle neutron scattering and light scattering. *Prog. Colloid Polym. Sci.* 81:107–112.
4. Hammermann, M., N. Brun, …, J. Langowski. 1998. Salt-dependent DNA superhelix diameter studied by small angle neutron scattering measurements and Monte Carlo simulations. *Biophys. J.* 75: 3057–3063.
5. Hammermann, M., K. Tóth, …, J. Langowski. 2000. Salt-dependent compaction of di- and trinucleosomes studied by small-angle neutron scattering. *Biophys. J.* 79:584–594.
6. Doster, W., S. Cusack, and W. Petry. 1989. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature.* 337:754–756.
7. Doster, W., S. Cusack, and W. Petry. 1990. Dynamic instability of liquidlike motions in a globular protein observed by inelastic neutron scattering. *Phys. Rev. Lett.* 65:1080–1083.
8. Daniel, R. M., J. C. Smith, …, J. L. Finney. 1998. Enzyme activity below the dynamical transition at 220 K. *Biophys. J.* 75:2504–2507.

9. Tarek, M., and D. J. Tobias. 2000. The dynamics of protein hydration water: a quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophys. J.* 79:3244–3257.
10. Zaccai, G. 2000. How soft is a protein? A protein dynamics force constant measured by neutron scattering. *Science.* 288:1604–1607.
11. Roh, J. H., V. N. Novikov, …, A. P. Sokolov. 2005. Onsets of anharmonicity in protein dynamics. *Phys. Rev. Lett.* 95:038101.
12. Roh, J. H., J. E. Curtis, …, A. P. Sokolov. 2006. Influence of hydration on the dynamics of lysozyme. *Biophys. J.* 91:2573–2588.
13. Magazù, S., G. Maisano, …, A. Benedetto. 2008. Elastic incoherent neutron scattering on systems of biophysical interest: mean square displacement evaluation from self-distribution function. *J. Phys. Chem. B.* 112:8936–8942.
14. Roh, J. H., R. M. Briber, …, A. P. Sokolov. 2009. Dynamics of tRNA at different levels of hydration. *Biophys. J.* 96:2755–2762.
15. Jasnin, M., L. van Eijck, …, G. Zaccai. 2010. Dynamics of heparan sulfate explored by neutron scattering. *Phys. Chem. Chem. Phys.* 12:3360–3362.
16. Nakagawa, H., H. Kamikubo, and M. Kataoka. 2010. Effect of conformational states on protein dynamical transition. *Biochim. Biophys. Acta.* 1804:27–33.
17. Sakai, V. G., and A. Arbe. 2009. Quasielastic neutron scattering in soft matter. *Curr. Opin. Colloid Interface Sci.* 14:381–390.
18. Chen, S.-H., M. Lagi, …, P. Baglioni. 2010. Dynamics of a globular protein and its hydration water studied by neutron scattering and MD simulations. *Spectroscopy.* 24:1–24.
19. Kneller, G. R., and K. Hinsen. 2009. Quantitative model for the heterogeneity of atomic position fluctuations in proteins: a simulation study. *J. Chem. Phys.* 131:045104.
20. Peters, J., and G. R. Kneller. 2013. Motional heterogeneity in human acetylcholinesterase revealed by a non-Gaussian model for elastic incoherent neutron scattering. *J. Chem. Phys.* 139:165102.
21. Stadler, A. M., F. Demmel, …, T. Seydel. 2016. Picosecond to nanosecond dynamics provide a source of conformational entropy for protein folding. *Phys. Chem. Chem. Phys.* 18:21527–21538.
22. Margheritis, E., B. Castellani, …, G. Garau. 2016. Bile acid recognition by NAPE-PLD. *ACS Chem. Biol.* 11:2908–2914.
23. Lenton, S., M. Grimaldo, …, S. C. Marujo Teixeira. 2017. Effect of phosphorylation on a human-like osteopontin peptide. *Biophys. J.* 112:1586–1596.
24. Magazù, S., F. Mezei, …, F. Migliardo. 2017. Protein dynamics as seen by (quasi) elastic neutron scattering. *Biochim. Biophys. Acta.* 1861:3504–3512.
25. Vural, D., X. Hu, …, J. C. Smith. 2017. Quasielastic neutron scattering in biology: theory and applications. *Biochim. Biophys. Acta.* 1861:3638–3650.
26. Marques, M. P., A. L. Batista de Carvalho, …, L. A. Batista de Carvalho. 2017. Intracellular water - an overlooked drug target? Cisplatin impact in cancer cells probed by neutrons. *Phys. Chem. Chem. Phys.* 19:2702–2713.
27. Katava, M., G. Stirnemann, …, A. Paciaroni. 2017. Critical structural fluctuations of proteins upon thermal unfolding challenge the Lindemann criterion. *Proc. Natl. Acad. Sci. USA.* 114:9361–9366.
28. Bicout, D. J., and G. Zaccai. 2001. Protein flexibility from the dynamical transition: a force constant analysis. *Biophys. J.* 80:1115–1123.
29. Doster, W., and M. Settles. 2005. Protein-water displacement distributions. *Biochim. Biophys. Acta.* 1749:173–186.
30. Squires, G. L. 1978. Introduction to the Theory of Thermal Neutron Scattering. Cambridge University Press, Cambridge, UK.
31. Bee, M. 1988. Quasielastic Neutron Scattering. Adam Hilger, Bristol.
32. Vural, D., L. Hong, …, H. R. Glyde. 2013. Long-time mean-square displacements in proteins. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 88:052706.
33. Tokuhisa, A., Y. Joti, …, M. Kataoka. 2007. Non-Gaussian behavior of elastic incoherent neutron scattering profiles of proteins studied by molecular dynamics simulation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 75:041912.
34. Vural, D., L. Hong, …, H. R. Glyde. 2015. Motional displacements in proteins: the origin of wave-vector-dependent values. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 91:052705.
35. Nakagawa, H., H. Kamikubo, …, M. Kataoka. 2004. Protein dynamical heterogeneity derived from neutron incoherent elastic scattering. *J. Phys. Soc. Jpn.* 73:491–495.
36. Vural, D., and H. R. Glyde. 2012. Intrinsic mean-square displacements in proteins. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 86:011926.
37. Lehnert, U., V. Réat, …, C. Pfister. 1998. Thermal motions in bacteriorhodopsin at different hydration levels studied by neutron scattering: correlation with kinetics and light-induced conformational changes. *Biophys. J.* 75:1945–1952.
38. Daniel, R. M., J. L. Finney, …, J. C. Smith. 1999. Enzyme dynamics and activity: time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. J.* 77:2184–2190.
39. Paciaroni, A., S. Cinelli, and G. Onori. 2002. Effect of the environment on the protein dynamical transition: a neutron scattering study. *Biophys. J.* 83:1157–1164.
40. Calandrini, V., V. Hamon, …, G. R. Kneller. 2008. Relaxation dynamics of lysozyme in solution under pressure: combining molecular dynamics simulations and quasielastic neutron scattering. *Chem. Phys.* 345:289–297.
41. Smith, J. C. 1991. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Q. Rev. Biophys.* 24:227–291.
42. Hayward, J. A., and J. C. Smith. 2002. Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophys. J.* 82:1216–1225.
43. Becker, T., and J. C. Smith. 2003. Energy resolution and dynamical heterogeneity effects on elastic incoherent neutron scattering from molecular systems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 67:021904.
44. Yi, Z., Y. Miao, …, J. C. Smith. 2012. Derivation of mean-square displacements for protein dynamics from elastic incoherent neutron scattering. *J. Phys. Chem. B.* 116:5028–5036.
45. Meinhold, L., D. Clement, …, J. C. Smith. 2008. Protein dynamics and stability: the distribution of atomic fluctuations in thermophilic and mesophilic dihydrofolate reductase derived using elastic incoherent neutron scattering. *Biophys. J.* 94:4812–4818.
46. Artymiuk, P. J., C. C. F. Blake, …, K. S. Wilson. 1982. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme of 6 Å resolution. *Acta Crystallogr. B.* 38:778–783.
47. Hess, B., C. Kutzner, …, E. Lindahl. 2008. GROMACS 4: algorithm for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
48. Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
49. Horn, H. W., W. C. Swope, …, T. Head-Gordon. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 120:9665–9678.
50. Essmann, U., L. Perera, …, L. G. Pedersen. 1995. A smooth particle mesh ewald method. *J. Chem. Phys.* 103:8577–8593.
51. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
52. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A Gen. Phys.* 31:1695–1697.
53. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
54. Miao, Y., Z. Yi, …, J. C. Smith. 2012. Temperature-dependent dynamical transitions of different classes of amino acid residue in a globular protein. *J. Am. Chem. Soc.* 134:19576–19579.