# Intrinsic mean-square displacements in proteins

Derya Vural and Henry R. Glyde

# Intrinsic Mean Square Displacements in Proteins

Derya Vural[1] and Henry R. Glyde[1]

[1]*Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716-2570, USA*

The thermal mean square displacement (MSD) of hydrogen in proteins and its associated hydration water is measured by neutron scattering experiments and used an indicator of protein function. The observed MSD as currently determined depends on the energy resolution width of the neutron scattering instrument employed. We propose a method for obtaining the intrinsic MSD of H in the proteins, one that is independent of the instrument resolution width. The intrinsic MSD is defined as the infinite time value of $\langle r^2 \rangle$ that appears in the Debye-Waller factor. The method consists of fitting a model to the resolution broadened elastic incoherent structure factor or to the resolution dependent MSD. The model contains the intrinsic MSD, the instrument resolution width and a rate constant characterizing the motions of H in the protein. The method is illustrated by obtaining the intrinsic MSD $\langle r^2 \rangle$ of heparan sulphate (HS-0.4), Ribonuclease A and Staphysloccal Nuclase (SNase) from data in the literature.

## I. INTRODUCTION

Following pioneering experiments and subsequent developments, the mean square displacement (MSD) of hydrogen in proteins can now be readily observed in neutron scattering experiments[1–11]. Specifically, the global average MSD of H throughout the protein is typically determined from the elastic component of the incoherent dynamic structure factor (DSF). In proteins at low temperature, the MSD is small. As temperature is increased the MSD increases and often goes through a marked increase at a specific temperature or temperatures, $T_D$, denoted the dynamical transition.[1–3, 12–16] The onset of large values of MSD are associated with the onset of function in proteins.[3, 17–22] Essentially, the large amplitude MSD enables contact between different parts of the protein which promotes chemical activity, function and possible folding. A large MSD is used as an indicator that function in a protein is possible.

In most current methods of data analysis, the MSD and $T_D$ extracted from experiment depend on the energy resolution of the neutron scattering instrument employed. Different MSD $\langle r^2 \rangle_{exp}$ are extracted from data observed on different instruments. The higher the resolution, the larger that apparent $\langle r^2 \rangle_{exp}$ observed and the lower the apparent $T_D$ observed. A high energy resolution instrument is needed to observe all the motions that contribute to the MSD, including the slow, long time motions. Indeed, measurements in the same protein at the same hydration level on different instruments have been made explicitly to demonstrate the dependence of $\langle r^2 \rangle_{exp}$ on the instrument resolution.[23–25]

Specifically, the observed $\langle r^2 \rangle_{exp}$ is typically obtained from the observed, resolution broadened elastic incoherent DSF, $O(\mathbf{Q}, \omega = 0)$, as a function of wave vector transfer, $\mathbf{Q}$, as,

$$\langle r^2 \rangle_{exp} = -3 \frac{d \ln O_{exp}(\mathbf{Q}, \omega = 0)}{dQ^2}. \qquad (1)$$

Examples of this $\langle r^2 \rangle_{exp}$ obtained on different instruments which displays the dependence of $\langle r^2 \rangle_{exp}$ in the

instrument energy resolution width are shown in Fig. 1.

The goal of the present paper is to extract the intrinsic MSD, $\langle r^2 \rangle$, from the neutron scattering data. This is the intrinsic MSD independent of the instrument resolution width, $W$. The intrinsic MSD is defined here as the $\langle r^2(t = \infty) \rangle = \langle r^2 \rangle$ value of the MSD that appears in the full Debye-Waller factor. The intrinsic MSD includes motions up to $t = \infty$. It is the MSD that would be observed on an instrument with zero energy resolution width, $W \to 0$.

To obtain $\langle r^2 \rangle$ from data, a simple model of the observed incoherent DSF, $O(\mathbf{Q}, \omega)$, which includes the $\langle r^2 \rangle$, the resolution width W and a simple description of the motional processes with a single relaxation parameter, $\lambda$, is developed. The simple model for the normalized, resolution broadened $O(\mathbf{Q}, \omega = 0)$ is,

$$O_N(\mathbf{Q}, \omega = 0) = I_\infty + (1 - I_\infty)\frac{W}{(W + \lambda)} + A \qquad (2)$$

where $I_\infty$ is the Debye-Waller factor,

$$I_\infty = \exp(-\frac{1}{3}Q^2 \langle r^2 \rangle). \qquad (3)$$

and $A$ is an additive constant included if the data includes a constant. Eq. (3) is the definition of the intrinsic $\langle r^2 \rangle$. The model is fitted to the observed $O(\mathbf{Q}, \omega = 0)$ for a given W with $\langle r^2 \rangle$ and $\lambda$ treated as free parameters to be determined by best fit. An intrinsic $\langle r^2 \rangle$ independent of W is obtained from fits to existing data in this way.

As in experiment, Eq. (1), a MSD $\langle r^2 \rangle_{slope}$ obtained from the slope of the model $O(\mathbf{Q}, \omega = 0)$ given by,

$$\begin{aligned} \langle r^2 \rangle_{slope} &= -3 \frac{d \ln O_N(\mathbf{Q}, \omega = 0)}{dQ^2} \\ &= \langle r^2 \rangle / [1 + \frac{W}{I_\infty \lambda}]. \end{aligned} \qquad (4)$$

can be introduced. As with $\langle r^2 \rangle_{exp}$, this $\langle r^2 \rangle_{slope}$ depends on the instrument resolution width W, specifically on the ratio $W/\lambda$. The expression for $\langle r^2 \rangle_{slope}$ corresponds to the $\langle r^2 \rangle_{exp}$ and can also be fitted to the
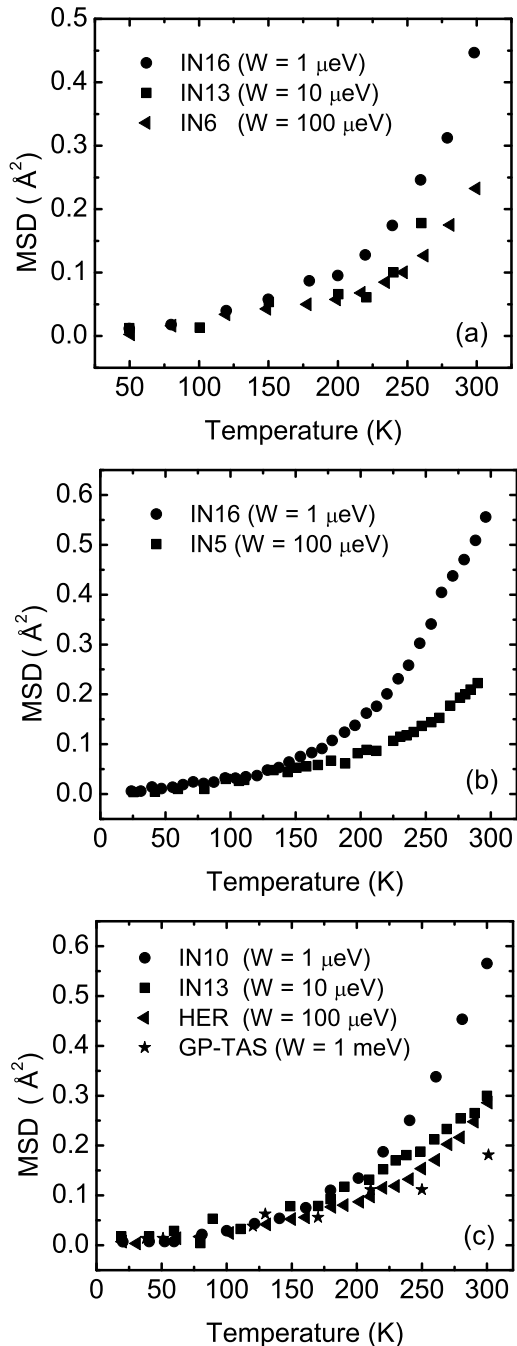
FIG. 1: MSD of $H$ in proteins observed on neutron scattering instruments having different energy resolution widths, W: (a) in hydrated heparan sulphate (HS-0.4) observed by Jasnin et. al[24] on IN16 ($W = 1$ $\mu$eV), IN13 ($W = 10$ $\mu$eV) and IN6 ($W = 100$ $\mu$eV), (b) in hydrated Ribonuclease A observed by Wood et. al[23] on IN16 ($W = 1$ $\mu$eV) and IN5 ($W = 100$ $\mu$eV), and (c) in hydrated Staphysloccal Nuclase (SNase) observed by Nakagawa et. al[25] on IN10 ($W = 1$ $\mu$eV), IN13 ($W = 10$ $\mu$eV), HER ($W = 100$ $\mu$eV) and GP-TAS ($W = 1$ meV).

observed values of $\langle r^2 \rangle_{exp}$ such as shown in Fig. 1 to obtain the intrinsic $\langle r^2 \rangle$. The intrinsic $\langle r^2 \rangle$ is always greater than $\langle r^2 \rangle_{slope}$. The intrinsic $T_D$ is defined as the temperature at which the intrinsic $\langle r^2 \rangle$ shows a marked increase with increasing temperature. There can be more than one $T_D$[13, 15].

In the following section, we develop the model incoherent DSF. In section 3, the model is fitted to observed values of the elastic incoherent DSF found in the literature to obtain the intrinsic $\langle r^2 \rangle$ and $T_D$ in three proteins. The intrinsic $\langle r^2 \rangle$ and other parameters are discussed in section 4, where suggestions for making the present simple, illustrative model more sophisticated are made.

## II. FORMULATION OF THE MODEL

### A. Dynamical Structure Factor

Neutrons incident on proteins interact and scatter predominantly from the hydrogen nuclei in the proteins and in the associated hydration water. The hydrogen nucleus, the proton, has a large, incoherent scattering cross-section for neutrons, 82 barns, which dominates all others. Neglecting the scattering from the electrons and other nuclei, the observed scattering intensity is proportional to the incoherent dynamical structure factor (DSF), $S_{inc}(\mathbf{Q}, \omega)$ of the H,

$$S_{inc}(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I_{inc}(\mathbf{Q}, t), \quad (5)$$

where

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{i=1}^{N} \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}_i(t)) \exp(i\mathbf{Q} \cdot \mathbf{r}_i(0)) \rangle \quad (6)$$

is the intermediate incoherent DSF. In Eq. (5), $\hbar\mathbf{Q}$ and $\hbar\omega$ are the momentum and energy, respectively, transferred from the neutron to the protein in the scattering.

The neutrons scatter from the $H$ at points $r_i(t)$ in the protein. The incoherent DSF is proportional to an average over the N self correlation functions of the $r_i(t)$ of each individual H. The H is distributed in very different environments throughout the protein and these correlation functions may be quite different with different time scales[26–28]. Our first approximation is to represent this average over all H by a single representative H so that the $I_{inc}(\mathbf{Q}, t)$ reduces to

$$I(\mathbf{Q}, t) = \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(t)) \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle. \quad (7)$$

This self correlation function of $\mathbf{r}(t)$ represents a weighted distribution over correlation functions and will contain all time scales, short and long, of H throughout the protein. We also drop the incoherent "inc" from $S(\mathbf{Q}, \omega)$ and $I(\mathbf{Q}, t)$ from now on. Eq. (7) is the usual approximation made in the analysis of experimental data.

The neutron instrument has an energy resolution function, denoted here by $R(\omega)$. A perfect energy resolution would be $R(\omega) = \delta(\omega)$, no resolution width. A representation of an $R(\omega)$ of finite width $W$ is

$$R(\omega) = \frac{1}{\pi} \frac{W}{W^2 + \omega^2}, \qquad (8)$$

a Lorentzian function, which has Fourier transform $R(t) = \exp(-Wt)$. We will also use a Gaussian $R(\mathbf{Q}, \omega)$. The observed function is a convolution of $S(\mathbf{Q}, \omega)$ and $R(\omega)$,

$$O(\mathbf{Q}, \omega) = \int_{-\infty}^{\infty} d\omega' S(\mathbf{Q}, \omega') R(\omega - \omega') \qquad (9)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I(\mathbf{Q}, t) R(t) \qquad (10)$$

$$O(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \exp(i\omega t) I(\mathbf{Q}, t) \exp(-Wt) \qquad (11)$$

Commonly measured is the elastic (zero energy transfer, $\omega = 0$) component of the incoherent $O(\mathbf{Q}, \omega)$,

$$O(\mathbf{Q}, \omega = 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, t) \exp(-Wt). \qquad (12)$$

TABLE I: The resolution width and the resolution time of neutron scattering instruments.

| Instrument | Energy Resolution ($\mu$eV) | $W$ (THz) | Time Scale (ps) |
|---|---|---|---|
| IN16 | $\sim$1 | 0.00025 | $\sim$4000 |
| IN10 | $\sim$1 | 0.00025 | $\sim$4000 |
| IN13 | $\sim$10 | 0.0025 | $\sim$400 |
| IN6 | $\sim$100 | 0.025 | $\sim$40 |
| IN5 | $\sim$100 | 0.025 | $\sim$40 |
| HER | $\sim$100 | 0.025 | $\sim$40 |
| GP-TAS | $\sim$1000 | 0.25 | $\sim$4 |

In $O(\mathbf{Q}, \omega = 0)$, we see that the role of a finite resolution width is to cut off the integrand after at time $\tau \sim W^{-1}$ so that long time processes in $I(\mathbf{Q}, t)$ are not observed in $O(\mathbf{Q}, \omega = 0)$. The higher the instrument resolution, the longer the time in $I(\mathbf{Q}, t)$ that can be observed in $O(\mathbf{Q}, \omega = 0)$. This is completely general independent of the form of $I(\mathbf{Q}, t)$ and $R(t)$. Only for infinitely sharp resolution, $W = 0$, $R(t) = 1$ and

$R(\omega) = \delta(\omega)$, are all long time motions in $I(\mathbf{Q}, t)$ observed in $O(\mathbf{Q}, \omega = 0)$.

In experiments, it is often assumed that the observed $O(\mathbf{Q}, \omega = 0)$ is given by $O_{exp}(\mathbf{Q}, \omega = 0) = \exp[-\frac{1}{3}Q^2 \langle r^2 \rangle]$. The experimental $\langle r^2 \rangle_{exp}$ is then obtained as $\langle r^2 \rangle_{exp} = -3d \ln O_{exp}(\mathbf{Q}, \omega = 0)/dQ^2$ as in Eq. (1).

Since from Eq. (12) the observed $O(\mathbf{Q}, \omega = 0)$ depends on the instrument resolution width $W$, this $\langle r^2 \rangle_{exp}$ will depend on the instrument resolution.

## B. The Model

The aim is to construct a simple model for $I(\mathbf{Q}, t)$ which we can fit to data to obtain the intrinsic value of $\langle r^2 \rangle$ that includes the contributions from all motions, including long time motions.

We begin by separating $I(\mathbf{Q}, t)$ into a $t = \infty$ ($I_\infty = I(\mathbf{Q}, t = \infty)$) and a time dependent part,

$$I(\mathbf{Q}, t) = I_\infty + (I(\mathbf{Q}, t) - I_\infty). \qquad (13)$$

where, from Eq. (7),

$$
\begin{aligned}
I_\infty = I(\mathbf{Q}, t = \infty) &= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(\infty)) \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\
&= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(\infty)) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\
&= \langle \exp(-i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \langle \exp(i\mathbf{Q} \cdot \mathbf{r}(0)) \rangle \\
&= \exp(-\frac{1}{3}Q^2 \langle r^2 \rangle + O(Q^4)) \qquad (14)
\end{aligned}
$$

is the infinite time limit. To obtain the last expression we assume: (1) that $\mathbf{r}(\infty)$ and $\mathbf{r}(0)$ are completely uncorrelated so that the averages of them are independent, (2) that the system is translationally invariant in time (no CM motion) so that $\mathbf{r}(\infty) = \mathbf{r}(0)$ and (3) that in a cumulant expansion of $\langle \exp(-i\mathbf{Q} \cdot \mathbf{r}) \rangle$, cumulants beyond the second cumulant are negligible. The latter is valid if $\mathbf{Q}$ is small or if the distribution over $\mathbf{r}$ is approximately a Gaussian distribution. The cumulants beyond the second vanish exactly for all $\mathbf{Q}$ if the distribution over $\mathbf{r}$ is exactly Gaussian.

We take $\langle r^2 \rangle$ in $I(\mathbf{Q}, \infty)$ in Eq. (14) as the definition of the intrinsic, long time value of $\langle r^2 \rangle$ in the protein that includes all motional processes. It is the $\langle r^2 \rangle$ that would be observed with an infinitely high resolution instrument, $R(\omega) = \delta(\omega)$. That is, $I_\infty = \exp(-Q^2 \langle r^2 \rangle / 3)$ is the definition of the intrinsic $\langle r^2 \rangle$ as in Eq. (3).

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

$$I'(\mathbf{Q}, t) = (1 - I_\infty)C(t) \qquad (15)$$

where $C(t)$ has the limits $C(t = 0) = 1$, $C(t = \infty) = 0$. An example is
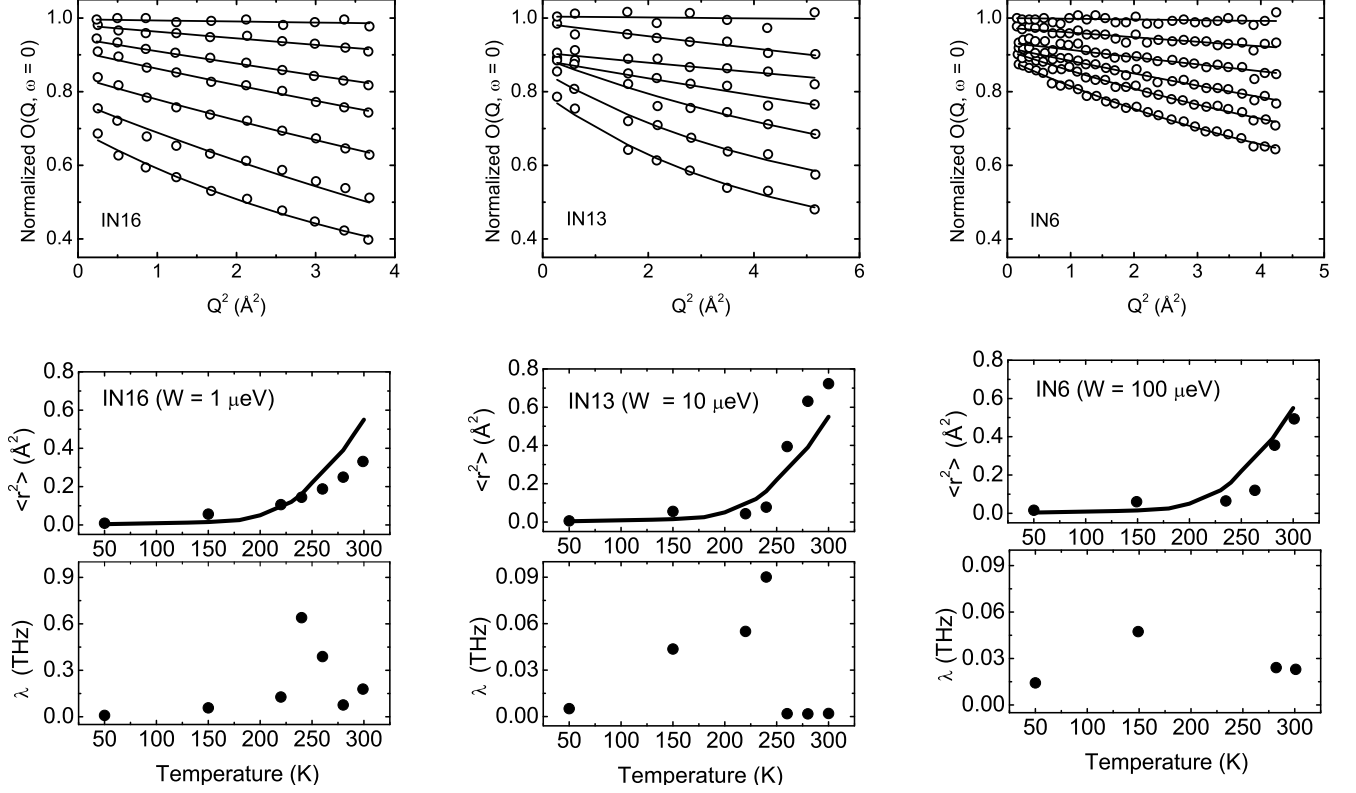
$$C(t) = \exp(-\lambda t), \qquad (16)$$

FIG. 2: Upper frame: Elastic DSF, $O_N(\mathbf{Q}, \omega = 0)$, of $H$ in heparan sulphate observed by Jasnin et.al[24] (open circles) and fit of model Eq. (2) to the observed $O_N(\mathbf{Q}, \omega = 0)$. Lower frames: the best fit values of the intrinsic $\langle r^2 \rangle$ and $\lambda$. The solid line is a guide to the eye for $\langle r^2 \rangle$, the same line for all three instruments.

where $\lambda$ represents the decay constant of correlations in the protein. The model is

$$I(\mathbf{Q}, t) = I_\infty(\mathbf{Q}) + (1 - I_\infty(\mathbf{Q}))C(t). \qquad (17)$$

which is constructed to have the correct limits at $t = 0$ and $t = \infty$ and to have a plausible representation of a single motional decay process, for example diffusion, at intermediate times.

Much more accurate description of motional processes that incorporate a spectrum of relaxation times have been developed and implemented.[29–31] This leads to more complete expressions for $C(t)$, especially at long times. Specific examples are a stretched exponential and particularly the fractional Ornstein-Uhlenbeck motional processes which provides relaxation functions the include long time motions. However, as discussed further below, in this application of extracting the intrinsic $\langle r^2 \rangle$ from fits to experimental data we found that the intrinsic $\langle r^2 \rangle$ obtained was not sensitive to the form of $C(t)$ used in the model Eq. (17).

Substituting the model $I(\mathbf{Q}, t)$ in $O(\mathbf{Q}, \omega = 0)$ the observed elastic function is

$$
\begin{aligned}
O(\mathbf{Q}, \omega = 0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt I(\mathbf{Q}, t) \exp(-Wt) \\
&= I_\infty \frac{1}{\pi W} + (1 - I_\infty)\frac{1}{\pi(W + \lambda)}. \quad (18)
\end{aligned}
$$

The data is usually presented as the "normalized" $O(\mathbf{Q}, \omega = 0)$, the $O(\mathbf{Q}, \omega = 0)$ above divided by $O(\mathbf{Q} = 0, \omega = 0)$ at $\mathbf{Q} = 0$. From Eq. (18), the model $O(\mathbf{Q} = 0, \omega = 0) = (\pi W)^{-1}$ and $O_N(\mathbf{Q}, \omega = 0)$ is given by Eq. (2). A constant $A$ is added to Eq. (2) since the data for different $\mathbf{Q}$ values are sometimes separated from one another in a figure by a constant for clarity.

The $O_N(\mathbf{Q}, \omega = 0)$ is the simple model that we fit to data for a given $W$. The $\langle r^2 \rangle$, $\lambda$ and $A$ are treated as free fitting parameters to be determined by the best fit to data. The parameter $A$ plays no role. In this way we determine the intrinsic long time value of $\langle r^2 \rangle$ and $\lambda$. The intrinsic $\langle r^2 \rangle$ should be independent of $W$. It can be compared with the experimental values $\langle r^2 \rangle_{exp}$ which are obtained from the slope of the observed $O_N(\mathbf{Q}, \omega = 0)$ using (12) and which are resolution dependent. Eq. (2) is the initial simple model of $O_N(\mathbf{Q}, \omega = 0)$ we use in section 3 to test the method.

We can also obtain a value of the MSD using the slope expression, Eq. (1), and our model $O_N(\mathbf{Q}, \omega = 0)$ of Eq. (2). This value, as with $\langle r^2 \rangle_{exp}$, will depend on the instrument resolution $W$ since $O_N(\mathbf{Q}, \omega = 0)$ depends on $W$. The model MSD obtained from the slope, $\langle r^2 \rangle_{slope} = -3d \ln O_N(\mathbf{Q}, \omega = 0)/dQ^2$ and differentiating Eq. (2) is given by Eq. (4). We see immediately that $\langle r^2 \rangle_{slope}$ depends on the ratio of $W/\lambda$. If $W \ll \lambda$, then the instrument resolution is high enough to catch all the decay of $I(\mathbf{Q}, t)$ to $I_\infty$ and $\langle r^2 \rangle_{slope} \longrightarrow \langle r^2 \rangle$, the intrinsic and full $\langle r^2 \rangle$. If the fit of $O_N(\mathbf{Q}, \omega = 0)$ to data is good, so that $\lambda$ and $\langle r^2 \rangle$ in Eq. (4) are well determined, then we expect the $\langle r^2 \rangle_{slope}$ to reproduce the observed $\langle r^2 \rangle_{exp}$ well.

## III. RESULTS

In this section we fit our model of the observed, resolution broadened elastic incoherent DSF $O_N(\mathbf{Q}, \omega = 0)$ given by Eq. (2) to data in the literature. The goal is to determine the intrinsic MSD $\langle r^2 \rangle$ of $H$ in specific proteins and obtain a value for the relaxation parameter $\lambda$ which describes the approach of $\langle r^2 \rangle$ to its long time, intrinsic value.

## A. The MSD

The top frame of Fig. 2 shows a fit of the model $O_N(\mathbf{Q}, \omega = 0)$ to the normalized $O(\mathbf{Q}, \omega = 0)$ observed in heparan sulphate (HS-0.4) by Jasnin et. al [24]. $O(\mathbf{Q}, \omega = 0)$ is observed on three instruments which have different energy resolutions, $W = 1$, 10 and 100 $\mu$eV. There is data for $O(\mathbf{Q}, \omega = 0)$ at seven temperatures, the lowest temperature at the top. The fits are generally good but some data points lie off the fitted line. The middle and bottom frames in Fig. 2 show the values of $\langle r^2 \rangle$ and $\lambda$ in the model $O_N(\mathbf{Q}, \omega = 0)$ that give the best fit. The best fit $\langle r^2 \rangle$ increases with temperature, markedly for temperatures above $T \simeq 230$ K. There is some scatter in the $\langle r^2 \rangle$ which arises from the uncertainty in the fit. The scatter or uncertainty of $\lambda$ is particularly large showing that the data is relatively insensitive to the value of $\lambda$. Conversely, we can say that the data is not very discriminating or sensitive to the model employed for the time dependence of $C(t)$. However, importantly, the $\langle r^2 \rangle$ emerging from the fit is independent of the resolution, $W$. The solid black line in the middle frame of Fig. 2 is a guide to the eye through all the $\langle r^2 \rangle$, the same line for all instruments. This shows that while there is substantial fluctuations in the individual values of $\langle r^2 \rangle$, the average $\langle r^2 \rangle$ obtained is independent of the instrument resolution $W$.

Similarly, Fig. 3 compares the intrinsic $\langle r^2 \rangle$, the model $\langle r^2 \rangle_{slope}$ and the observed $\langle r^2 \rangle_{exp}$ in HS-0.4. The solid points are the intrinsic $\langle r^2 \rangle$ emerging from the fits in Fig. 2 and the solid line is again a guide to the eye through
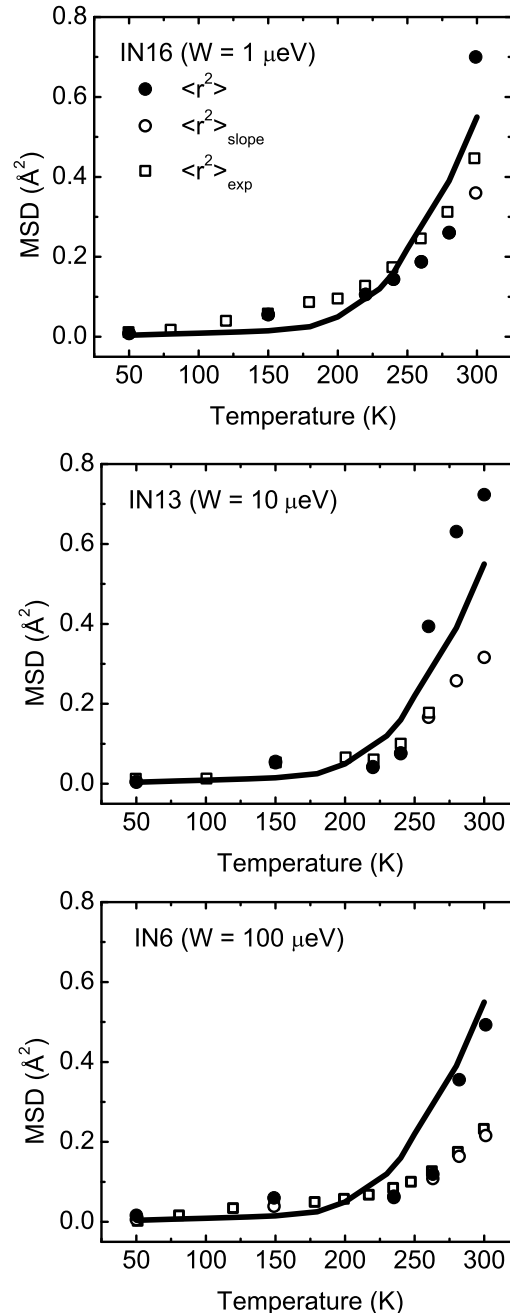


FIG. 3: MSD, $\langle r^2 \rangle_{exp}$, in hydrated HS-0.4 observed Jasnin et. al[24] on IN16, IN13 and IN6 (open squares). The intrinsic $\langle r^2 \rangle$ (solid circles) are the intrinsic MSD obtained by fitting the model Eq. (2) to data shown in Fig. 2. The solid line is a guide to the eye through the intrinsic $\langle r^2 \rangle$, the same for all instruments. The $\langle r^2 \rangle_{slope}$ (open circles) are the MSD calculated from Eq. (4) which should be similar to $\langle r^2 \rangle_{exp}$.

these $\langle r^2 \rangle$, the same line for all three instruments. The $\langle r^2 \rangle_{exp}$ is the MSD obtained by Jasnin et. al from the slope of their data using Eq. (1). The $\langle r^2 \rangle_{slope}$ is obtained from the slope of the model $O(\mathbf{Q}, \omega = 0)$ given by
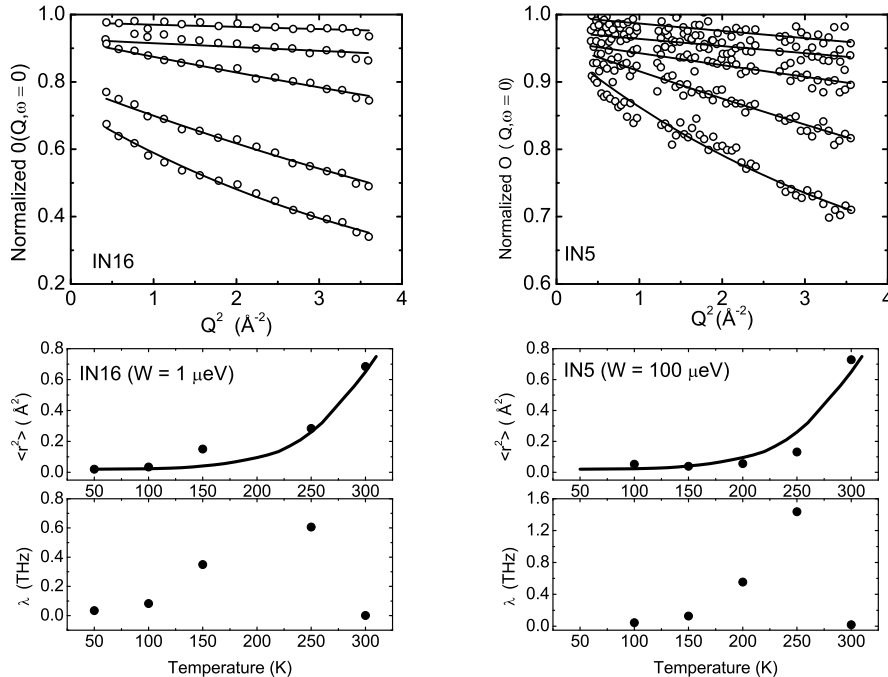
FIG. 4: Upper frame: Elastic DSF, $O_N(\mathbf{Q}, \omega = 0)$, observed by Wood et. al[23] on IN16 and IN5 (open circles) and fit of model Eq. (2) to the observed $O_N(\mathbf{Q}, \omega = 0)$. Lower frames: the best fit values of $\langle r^2 \rangle$ and $\lambda$. The solid line is a guide to the eye to the best fit $\langle r^2 \rangle$, the same line for both instruments.

Eq. (4). If the fit to the data is precise, the $\langle r^2 \rangle_{slope}$ and $\langle r^2 \rangle_{exp}$ should agree. This is the case for IN13 and IN6 but less so for the IN16. When the instrument resolution $W$ is small, we expect $\langle r^2 \rangle_{slope}$ to coincide with the intrinsic $\langle r^2 \rangle$. This is the case for IN16 where $\langle r^2 \rangle_{slope}$ (open circles) lie on top of the $\langle r^2 \rangle$ (black dots). For IN16, the $\langle r^2 \rangle_{slope}$ and $\langle r^2 \rangle_{exp}$ differ somewhat for $T > 250$ K. This indicates that there is not a good fit at higher temperatures, as can be seen in the upper frame of Fig. 2. The essential point of Fig. 3 is that the intrinsic $\langle r^2 \rangle$ is independent of $W$ and that $\langle r^2 \rangle$ shows a marked increase at temperature, $T_D \simeq 230$ K, the intrinsic dynamical transition temperature of HS-0.4. Similarly, the upper frame of Fig. 4 shows the fit of the model $O_N(\mathbf{Q}, \omega = 0)$ of Eq. (2) to the measured $O_N(\mathbf{Q}, \omega = 0)$ in Ribonuclease A observed by Wood et. al[23]. The fits are good except that there is significant scatter in the data taken on IN5. The best fit values of $\langle r^2 \rangle$ and $\lambda$ are shown in the lower frames of Fig. 4. Again the best fit values of the intrinsic $\langle r^2 \rangle$ is independent of the instrument energy resolution $W$, as seen from the solid line, a guide to the eye to the intrinsic $\langle r^2 \rangle$, the same for both instruments. In Fig. 5, the intrinsic $\langle r^2 \rangle$, the solid line, is compared with the resolution dependent observed $\langle r^2 \rangle_{exp}$ obtained from the slope of the data with $Q^2$. While the $\langle r^2 \rangle_{exp}$ is very different for the two instruments displaying the dependence of $\langle r^2 \rangle_{exp}$ on the resolution width $W$, the $\langle r^2 \rangle_{exp}$ observed on IN16 agrees well with the intrinsic $\langle r^2 \rangle$. This indicates that $\langle r^2 \rangle$ in Ribonuclease A reaches its final, equilibrium value

within a time period observable on IN16 ($\tau \sim 4$ nanoseconds). In a similar way, we fitted the model Eq. (2) to values of $O_N(\mathbf{Q}, \omega = 0)$ observed by Nakagawa et. al[25] in Staphysloccal Nuclase (SNase) on four different instruments. The resulting intrinsic $\langle r^2 \rangle$ obtained from the fits are shown in Fig. 6 as solid dots for each instrument with a solid line (guide) through the dots, the same line for all four instruments. Although there is substantial scatter in the $\langle r^2 \rangle$, an intrinsic $\langle r^2 \rangle$ independent of the instrument resolution width $W$ can be obtained by fitting the model to the observed $O_N(Q, \omega = 0)$. To test the sensitivity of the results to the model of $C(t)$ and to the shape of the instrument resolution (IR), we changed $C(t)$ to Gaussian, $C(t) = \exp(-\lambda^2 t^2/2)$, and the IR to Gaussian, $R(\omega) = \exp(-\omega^2/2W^2)/\sqrt{2\pi W^2}$. With both $C(t)$ and $R(\omega)$ a Gaussian, an analytic expression for $O_N(\mathbf{Q}, \omega = 0)$ can again be obtained. Fig. 7 shows the resulting intrinsic $\langle r^2 \rangle$ obtained from fitting the Gaussian model to the data of Nakagawa et. al. The $\langle r^2 \rangle$ in Figs. (6) and (7) are barely distinguishable. This shows that intrinsic $\langle r^2 \rangle$ obtained is not sensitive to the form of $C(t)$ and $R(\omega)$ used in the model at the present level of precision of the data.

## B. The Dynamical Transition Temperature, $T_D$

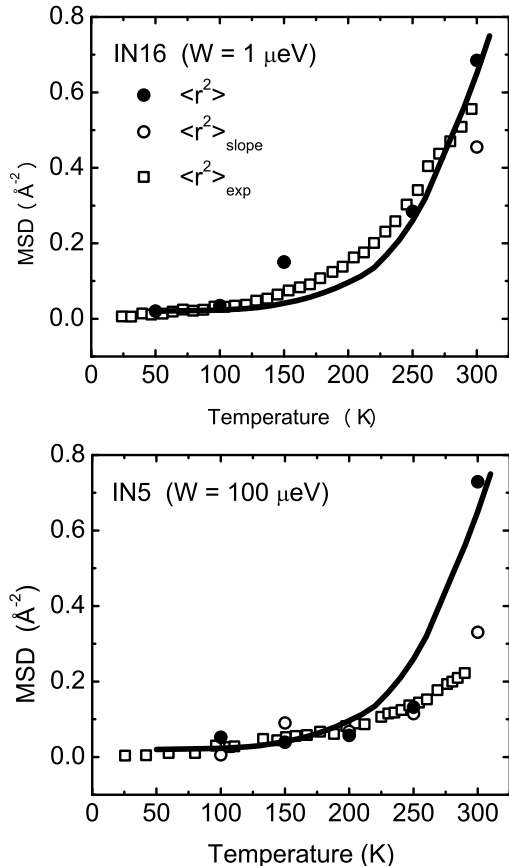In this subsection, we consider the apparent dependence of the dynamical transition temperature on the

FIG. 5: MSD, $\langle r^2 \rangle_{exp}$, in hydrated Ribonuclease A observed by Wood et. al [23] on IN16 and IN5 (open squares). The intrinsic MSD $\langle r^2 \rangle$ (solid circles) are obtained by fitting the model Eq. (2) to the data shown in Fig. 4. The solid line is a guide to the eye to the intrinsic MSD, the same for both instruments. The MSD $\langle r^2 \rangle_{slope}$ (open circles) is calculated from Eq. (4).

instrument energy resolution $W$. The upper frame of Fig. 8 shows the $\langle r^2 \rangle_{exp}$ of H in glutamate dehydrogenase observed on IN16 and IN6. The $\langle r^2 \rangle_{exp}$ is obtained as usual from the slope of the observed $O_{exp}(\mathbf{Q}, \omega = 0)$ with $Q^2$ as given by Eq. (1). In this example, the dynamical transition temperature, $T_D$, the temperature at which $\langle r^2 \rangle_{exp}$ increases markedly with temperature, appears to depend on the neutron instrument used. On IN16, the apparent $T_D$ is $T_D \simeq 150$K while that on IN6 is $T_D \simeq 230$K. To determine the intrinsic $\langle r^2 \rangle$ and $T_D$ for this protein, we fitted the model $\langle r^2 \rangle_{slope}$ given by Eq. (4) to the observed $\langle r^2 \rangle_{exp}$ in Fig. 8. Eq. (4) is the model equivalent of Eq. (1). Specially, at each temperature we determined the two parameters $\langle r^2 \rangle$ and $\lambda$ by setting $\langle r^2 \rangle_{slope}$ in Eq. (4) equal to the $\langle r^2 \rangle_{exp}$ of Fig. 8 for each instrument. The values of $\langle r^2 \rangle_{slope}$ actually fitted are shown as the solid lines through the data points in the upper frame. The resulting values of $\langle r^2 \rangle$ and $\lambda$

are shown as the lower frames of Fig. 8.

The intrinsic $\langle r^2 \rangle$ shows two interesting features. Firstly, the intrinsic $\langle r^2 \rangle$ is very similar to the $\langle r^2 \rangle_{exp}$ observed on IN16. The intrinsic $T_D$ is $T_D \simeq 150$ K. Secondly, and unexpectedly, the $\langle r^2 \rangle$ is found to decrease for $T \gtrsim 200$ K and is lower at $T \simeq 250$ K than at $T = 200$ K. This seems unphysical. To explore this effect, we arbitrarily kept the intrinsic $\langle r^2 \rangle$ constant at its 200 K value for temperatures $T \gtrsim 200$ K. This $\langle r^2 \rangle$ is shown in Fig. 9. The resulting $\langle r^2 \rangle_{slope}$ obtained when $\langle r^2 \rangle$ is held constant for $T > 200$ K is shown as the solid line in Fig. 9. The $\langle r^2 \rangle_{slope}$ for $W = 1$ $\mu$eV (IN16) continues to increase above $T = 200$ K but eventually reaches a plateau above 250 K. This plateau of $\langle r^2 \rangle$ at higher temperature appears to be unique for H in glutamate dehydrogenase or there is an issue with the sample or data on IN16 for $T > 200$ K. Further study of this effect would be interesting.

The chief result is that an intrinsic $\langle r^2 \rangle$ and $T_D$ independent of instrument resolution can be obtained. The intrinsic $T_D$ is the temperature at which the intrinsic $\langle r^2 \rangle$ begins to increase markedly. In glutamate dehydrogenase, the intrinsic $T_D$ is at $T_D \simeq 150$ K in Fig. 8. This is very close to the $T_D$ observed on IN16 where $W = 1$ $\mu$eV. When $W$ is larger, the apparent $T_D$ seen in $\langle r^2 \rangle_{exp}$ is moved to higher temperature. The larger the $W$, the higher the apparent $T_D$. This finding is consistent with the results of Becker et al.[16]. Returning to the previous subsection, we note that the data in Fig. 6 shows that the larger the $W$, the higher the temperature that is needed to observe a marked increase in $\langle r^2 \rangle_{exp}$. That is, the apparent $T_D$ of $\langle r^2 \rangle_{exp}$ in Fig. 6 is shifted to a higher temperature for larger W. In this way, the data shown in Figs. (6) and (8) are manifestations of the same phenomena.

## IV.    DISCUSSION

The goal of many measurements of quasielastic neutron scattering from proteins is to determine the thermal mean square displacement (MSD) of hydrogen in the protein and in its hydration water. The observed MSD, $\langle r^2 \rangle_{exp}$, is generally obtained from the slope of the observed elastic incoherent DSF $O(\mathbf{Q}, \omega = 0)$ using Eq. (1). The $O(\mathbf{Q}, \omega = 0)$ is the resolution broadened DSF $S(\mathbf{Q}, \omega)$ at $\omega = 0$. However, the $\langle r^2 \rangle_{exp}$ given by Eq. (1) is equal to the actual $\langle r^2 \rangle$ of the protein only if the scattered intensity exactly at $\omega = 0$ is observed. This is the case only if the energy resolution width, $W$, of the instrument is zero so that $O(\mathbf{Q}, \omega = 0)$ reduces to $S(\mathbf{Q}, \omega = 0)$. In this limit $O(\mathbf{Q}, \omega = 0)$ is well approximated by its time independent part $I_\infty$ given by Eq. (3), as seen from Eq. (4), and $\langle r^2 \rangle_{exp}$ equals $\langle r^2 \rangle$. When $W$ is finite the scattered intensity at finite $\omega$ around $\omega = 0$ is also observed in $O(\mathbf{Q}, \omega = 0)$. In this case the $\langle r^2 \rangle_{exp}$ extracted from Eq. (1) depends on the resolution width $W$ and is smaller than $\langle r^2 \rangle$. More precisely, it depends on the ratio of $W$ and to the rate constants that govern
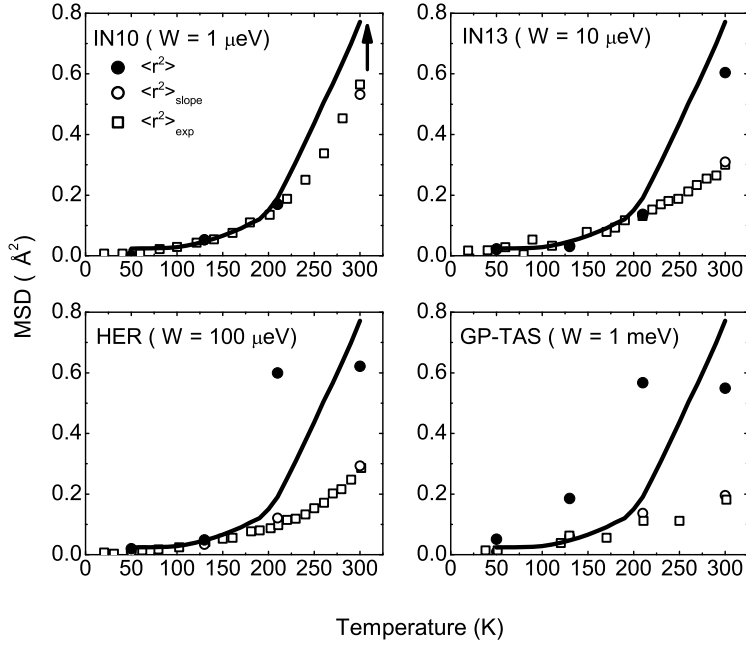
FIG. 6: MSD, $\langle r^2 \rangle_{exp}$, of the hydrated SNase observed by Nakagawa et. al [25] on IN10 ($W = 1$ $\mu$eV), IN13 ($W = 10$ $\mu$eV), HER ($W = 100$ $\mu$eV) and GP-TAS ($W = 1000$ $\mu$eV)(open squares). The intrinsic MSD $\langle r^2 \rangle$(solid circle) is obtained by fitting the Eq. (2) to the elastic intensity data observed Nakagawa et. al [25]. The solid line is a guide to the eye to the intrinsic MSD, the same for all instruments. The MSD $\langle r^2 \rangle_{slope}$(open circles) is calculated from Eq. (4).
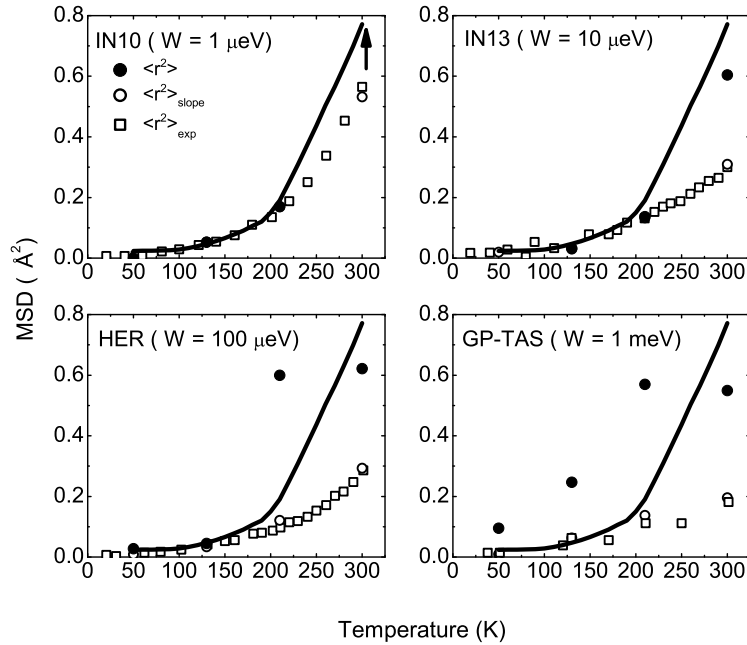


FIG. 7: MSD, $\langle r^2 \rangle_{exp}$, of the hydrated SNase observed by Nakagawa et. al [25] on IN10 ($W = 1$ $\mu$eV), IN13 ($W = 10$ $\mu$eV), HER ($W = 100$ $\mu$eV) and GP-TAS ($W = 1000$ $\mu$eV)(open squares). The intrinsic MSD $\langle r^2 \rangle$(solid circle) is obtained by fitting the observed DSF equation, where the Gaussian resolution and $C(t)$ functions are used, to the elastic intensity data observed Nakagawa et. al [25]. The solid line is a guide to the eye to the intrinsic MSD, the same for all instruments. The MSD $\langle r^2 \rangle_{slope}$(open circles) is calculated from the $Q^2$ dependence of the observed DSF.
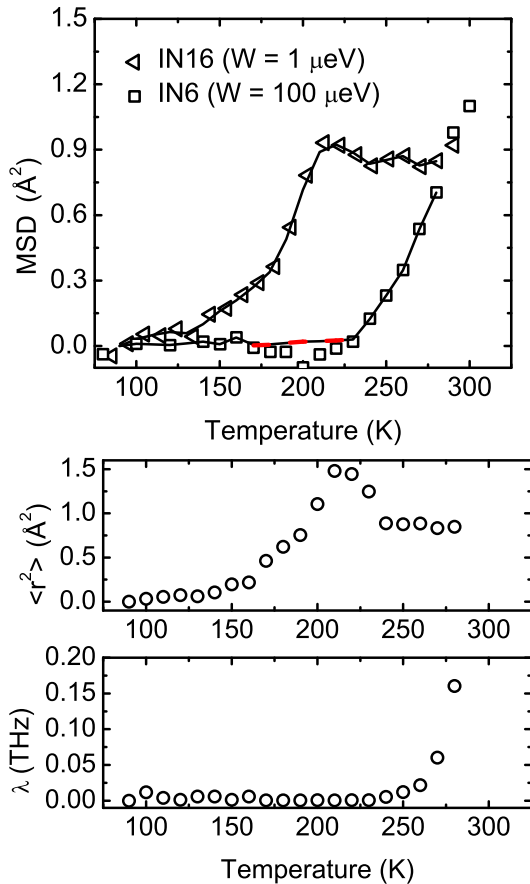
FIG. 8: MSD, $\langle r^2 \rangle_{exp}$, of glutamate dehydrogenase in $CD_3OD/D_2O$ observed by Daniel et. al [32] on IN16 ($W = 1\,\mu eV$) (triangles) and IN6 ($W = 100\,\mu eV$) (squares). Eq. (4) for $\langle r^2 \rangle_{slope}$ is fitted to the observed $\langle r^2 \rangle_{exp}$ and the fitted $\langle r^2 \rangle_{slope}$ is shown as a solid line. The lower frames show the intrinsic $\langle r^2 \rangle$ and $\lambda$ obtained from the fit. The extracted intrinsic $\langle r^2 \rangle$ decreases above $T = 220$ K.

FIG. 9: The MSD as in Fig. 8. In this case the intrinsic MSD $\langle r^2 \rangle$ is arbitrarily held constant at temperatures $T > 200K$. The solid lines show the values of $\langle r^2 \rangle_{slope}$ obtained for IN16 and IN6 when the $\langle r^2 \rangle$ is held constant for $T > 200K$. The expected $\langle r^2 \rangle_{slope}$ continues to increase above 200 K.

the motions in the protein. For a given protein, different $\langle r^2 \rangle_{exp}$ are observed on instruments having different resolution widths $W$. The impact of a finite $W$ is different in different proteins.

The aim of the present paper is to provide a method in which the intrinsic $\langle r^2 \rangle$ can be extracted from data taken on instruments that have different $W$. The essence of the method is to create a model of $O(\mathbf{Q}, \omega = 0)$ in which the intrinsic, infinite time $\langle r^2 \rangle$ appears explicitly and to fit the model to the observed $O(\mathbf{Q}, \omega = 0)$ to obtain $\langle r^2 \rangle$. The model must also contain a description of the motional processes and the resolution width $W$. We began with a simple model of the motions and a single, global Debye-Waller factor which led to Eq. (2). By fitting the model to data taken on instruments that have different energy resolutions we showed that an instrument independent $\langle r^2 \rangle$ can be obtained. The $\langle r^2 \rangle_{slope}$ given by Eq. (4) is the model equivalent
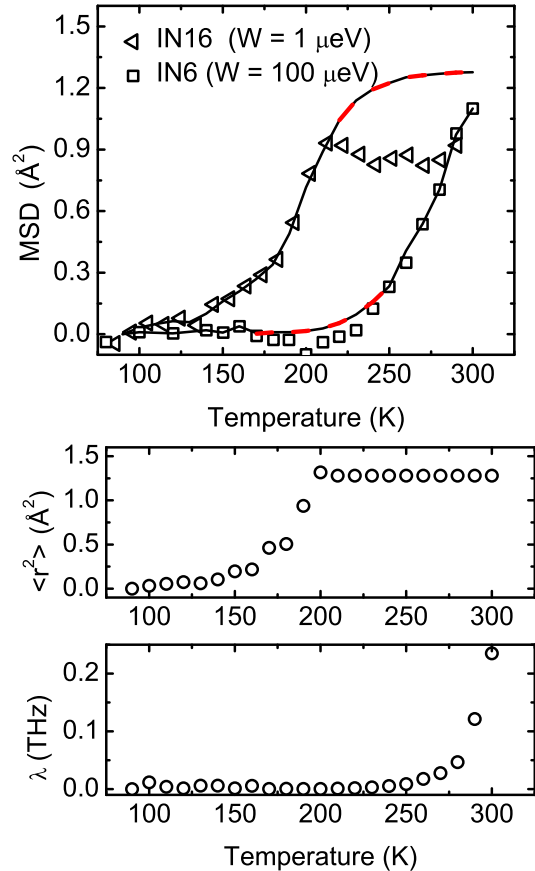
of $\langle r^2 \rangle_{exp}$. The model $\langle r^2 \rangle_{slope}$ can also be fitted to observed values of $\langle r^2 \rangle_{exp}$ to obtain the intrinsic $\langle r^2 \rangle$ if the measured $O(\mathbf{Q}, \omega = 0)$ are not available.

The dependence of the $\langle r^2 \rangle_{exp}$ obtained from the slope of $O(\mathbf{Q}, \omega = 0)$ with $Q^2$ using Eq. (1), and its associated dynamical transition temperature, $T_D$, on the instrument resolution width, $W$ may be illustrated and clarified using Eq. (4) for $\langle r^2 \rangle_{slope}$. The $\langle r^2 \rangle_{slope}$ is the model equivalent of $\langle r^2 \rangle_{exp}$ given by Eq. (1). Firstly, the intrinsic $\langle r^2 \rangle$ is always greater than the observed $\langle r^2 \rangle_{slope}$ at a given temperature. The larger is $W$, the further the observed $\langle r^2 \rangle_{slope}$ is suppressed below the intrinsic $\langle r^2 \rangle$. Similarly, the intrinsic $T_D$ obtained from the intrinsic $\langle r^2 \rangle$ is always at a lower temperature than the observed $T_D$. The larger is $W$, the more the apparent $T_D$ is shifted to higher temperature above the intrinsic $T_D$. The dependence of $\langle r^2 \rangle_{slope}$ and $T_D$ on $W$ can be illustrated by plotting $\langle r^2 \rangle_{slope}$ for increasing values of $W$ using by Eq. (4) as shown in Fig. 10. As input, we use the intrinsic $\langle r^2 \rangle$ of glutamate dehydrogenase shown in Fig. 8
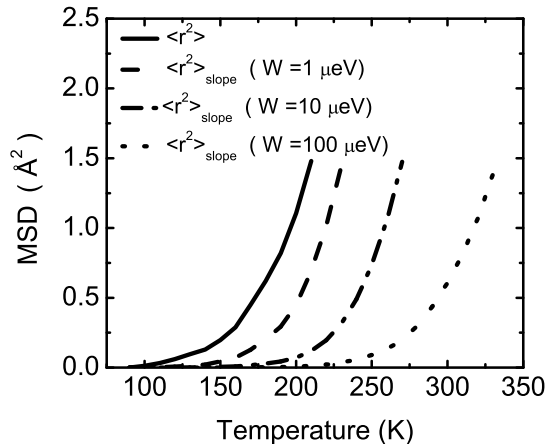
FIG. 10: The resolution broadened MSD, $\langle r^2 \rangle_{slope}$, for different resolution widths, $W$ obtained from Eq.(4) illustrating the dependence of the apparent $T_D$ on W. The intrinsic $\langle r^2 \rangle$ and $\lambda$ of glutamate dehydrogenase, obtained from the fits to data by Daniel et al. [32] shown in Fig. 8, are used in Eq. (4).

plus an extrapolation of this $\langle r^2 \rangle$ to higher temperatures. A reasonable $\lambda$ similar to that Fig. 8 is also used. This intrinsic $\langle r^2 \rangle$ has an intrinsic $T_D \simeq 150$ K. From Fig. 10, we see that at a given temperature, $\langle r^2 \rangle_{slope}$ calculated from Eq. (4) always lies below $\langle r^2 \rangle$. The larger is $W$, the smaller is $\langle r^2 \rangle_{slope}$. Equivalently, we could say that $\langle r^2 \rangle_{slope}$ in Fig. 10 is shifted to higher temperature. The larger the $W$, the further the $\langle r^2 \rangle_{slope}$ is shifted to higher temperature. These shifts may be described as a $\langle r^2 \rangle_{slope}$ that lies below $\langle r^2 \rangle$ at a given temperature or a $T_D$ that is shifted to higher temperature than the intrinsic or actual $T_D$. The reduction in $\langle r^2 \rangle_{slope}$ and the apparent increase in $T_D$ with increasing $W$ are one and the same effect.

The impact of a finite $W$ can also be seen in $I(\mathbf{Q}, t)$ given by Eqs. (12), (18) and (4). In Eq. (12), the integral over time is cut off after a time $t_W = W^{-1}$. If $W$ is large, $t_W$ is short and long time motions/correlations in $C(t)$ cannot contribute to the integral in Eq. (12). The $O(\mathbf{Q}, \omega = 0)$ then contains only the shorter time motions and $O(\mathbf{Q}, \omega = 0)$ depends on $W$. What is important is the ratio of $W$ to $\lambda$. Kneller and Calandrini [30] have also shown that the magnitude of corrections to the observed $\langle r^2 \rangle_{slope}$ for resolution effects depend on the ratio $W/\lambda$. If the intrinsic correlations in the protein (in $C(t)$) decay rapidly, $\lambda >> W$, then all correlations in $C(t)$ are included in Eq. (12) and the integral again becomes independent of $W$. Thus we expect the impact of finite resolution width $W$ to vary from protein to protein.

In any model used to extract the intrinsic $\langle r^2 \rangle$, some characterization of the rate of decay of correlations relative to $W$ is needed. For example, in a White Noise model of correlations, $C(t) = \delta(t)$, there are no correlations and any $W$ will capture all correlations and yield

the intrinsic $\langle r^2 \rangle$. In section 3, we saw that the $\langle r^2 \rangle_{exp}$ of HS-0.4 observed on IN16 ($W = 1$ $\mu$eV) was close to the intrinsic $\langle r^2 \rangle$. This means that all the correlations in $C(t)$ of HS-0.4 are largely captured within a time scale $W^{-1} = 4$ ps. Whenever the observed $\langle r^2 \rangle_{exp}$ is close to the intrinsic $\langle r^2 \rangle$, we may infer that the instrument resolution is small enough that all or most of the correlations contributing to $C(t)$ have been captured in Eq. (12).

The present model is very simple and is intended to be illustrative only. The model has several limitations. Firstly, we used a simple representation $C(t) = \exp(-\lambda t)$ to describe the relaxation of correlations in the protein. This $C(t)$ is appropriate for a single diffusion process. There is also no representation of ballistic propagation or vibration in $C(t)$. The $\lambda$ could also be $\mathbf{Q}$ dependent. The model could be improved by using a stretched exponential which represents several diffusion mechanisms contributing to $C(t)$. However, we found that the data was not significantly precise to distinguished between a simple exponential and a stretched exponential $C(t) = \exp(-(\lambda t)^\beta)$. The two $C(t)$ lead to the same intrinsic $\langle r^2 \rangle$. Even better, an Ornstein-Uhlenbeck (O-U) function $C(t) = E_\alpha(-(\lambda t)^\alpha)$ could be used. Calandrini et al.[31] have found good fits to simulation data and to QENS data using the O-U function $E_\alpha$ with $\alpha = 1/2$. For $\alpha = 1/2$, this has an analytic form $E_{1/2}(-(\lambda t)^{1/2}) = \exp(\lambda t) erfc((\lambda t)^{1/2})$. We also used this form in Eq. (17) ($C(t) = \exp(\lambda t) erfc((\lambda t)^{1/2})$) and found the same value of intrinsic $\langle r^2 \rangle$ within precision from fits to data. Current data does not appear to be sufficiently precise to distinguish between simple and sophisticated models of $C(t)$ for the purpose of determining $\langle r^2 \rangle$. In future applications a more sophisticated $C(t)$ could be incorporated.

In contrast, we found that a stretched exponential with $\beta = 0.2$ and O-U function with $\alpha = 1/2$ provide much better fits to simulations than a single exponential. The simulations have much more precision especially at larger times. We also found that a stretched exponential with $\beta = 0.2$ and $E_\alpha$ for $\alpha = 1/2$ are quite similar functions.

The present model uses only a single, global Debye-Waller factor $I_\infty$ in Eq. (3). This simplification is made by approximating Eq. (6) by Eq. (7). In a real protein, the H occupies a variety of positions that have different values of $\langle r^2 \rangle$. If a distribution of $\langle r^2 \rangle$ values is included[26–28], there are terms in $I_\infty$ beyond a simple Gaussian represented by a single parameter $\langle r^2 \rangle$. In future applications a generalization to include a distribution of $\langle r^2 \rangle$ values based on Eq. (6) should be incorporated.

Improvements to the resolution function used could also be made. Also, recent simulations suggest that some motions contribute to $\langle r^2 \rangle$ only after long (picosecond) times. In applications of the present method to these proteins, the data from reasonably high resolution instruments may be needed so that information on these longer time scale motions is included in the observed $O(\mathbf{Q}, \omega = 0)$.

## V. SUMMARY

We have proposed a method for obtaining the intrinsic MSD of H in the proteins independent of the instrument resolution width, $W$. The method consists of fitting a model to the observed resolution dependent data or resolution dependent MSD. The model contains the intrinsic MSD, the instrument resolution width W and a rate constant characterizing the motions of H in the protein. The method is applied to existing data in the literature to obtain the intrinsic MSD $\langle r^2 \rangle$ of heparan sulphate (HS-0.4), Ribonuclease A and Staphysloccal Nuclase (SNase).

The intrinsic MSD is defined as the infinite time $\langle r^2 \rangle = \langle r^2(t = \infty) \rangle$ that appears in the Debye-Waller factor. The intrinsic $\langle r^2 \rangle$ is always greater than the apparent, resolution dependent $\langle r^2 \rangle_{exp}$. From fits to data at different temperatures, an intrinsic $\langle r^2 \rangle$ as a function of temperature is obtained. This intrinsic $\langle r^2 \rangle$ shows a dynamical transition. The intrinsic dynamical transition temperature, $T_D$, is defined as temperature at which the intrinsic MSD begins to increase markedly with temperature. The intrinsic $T_D$ is always less than the apparent, resolution dependent $T_D$.

The essential ingredients of a model needed to obtain $\langle r^2 \rangle$ are (1) a definition of $\langle r^2 \rangle$, (2) a characteristic time (or times) of the motions that contribute to $\langle r^2 \rangle$ and (3) the instrument resolution width. The present model is simple and illustrative only and can be much improved. At the same time, much current data in the literature may not be sufficiently precise to distinguish between simple and more sophisticated models.

## VI. ACKNOWLEDGEMENTS

## VII. APPENDIX

The purpose of this appendix is to derive the expression

$$I_\infty = \exp(-\frac{1}{3}Q^2\langle r^2 \rangle), \qquad (19)$$

to identify the approximations made to obtain it and to clarify the meaning of $\langle r^2 \rangle$. $I_\infty$ is the t $\rightarrow \infty$ limit of the intermediate incoherent DSF, $I(\mathbf{Q},t)$ defined in Eq. (7). $I(\mathbf{Q},\infty)$ is independent of time t. The Fourier transform of $I_\infty$ is therefore purely elastic (zero except at $\omega = 0$) and given by $S_{el}(\mathbf{Q},\omega) = I_\infty\delta(\omega)$. This shows that $I_\infty$ arises from structure in the protein since translationally invariant systems such as gases and liquids that have no structure have no elastic scattering. Thus the motions

that contribute to $I_\infty$ and $\langle r^2 \rangle$ are therefore vibrations, hindered rotations, restricted diffusion and all motions that in some way are restricted by or reflect a structure. $I_\infty$ is exactly the Debye-Waller factor that appears in the scattering of X-rays or neutrons from crystals[33–35]. In crystalline solids, $I_\infty$ is the reduction in intensity of a Bragg peak arising from atomic vibration in the solid. For H in proteins, $I_\infty$ is the reduction in intensity of the incoherent elastic scattering at any $\mathbf{Q}$ value arising from all restricted motions of the H.

To obtain $I_\infty$, we begin with $I_{inc}(\mathbf{Q},t)$ given by Eq. (6). The first approximation is to reduce Eq. (6) to Eq. (7). If Eq. (6) were retained then terms beyond the Gaussian in Eq. (19) will be obtained[26–28]. Following the arguments below Eq. (14), the t $\rightarrow \infty$ limit of Eq. (7) is,

$$\begin{aligned} I_\infty &= I(\mathbf{Q}, t = \infty) = \langle \exp(-i\mathbf{Q}\cdot\mathbf{r}(\infty))\exp(i\mathbf{Q}\cdot\mathbf{r}(0))\rangle \\ &= \langle\exp(-i\mathbf{Q}\cdot\mathbf{r}(0))\rangle\langle\exp(i\mathbf{Q}\cdot\mathbf{r}(0))\rangle. \end{aligned} \quad (20)$$

To arrive at the final line of Eq. (20) we have assumed: (1) that the position, $\mathbf{r}(\infty)$ at t $\rightarrow \infty$ of each H is uncorrelated with its position $\mathbf{r}(0)$ at $t = 0$ so the expectation values of $\mathbf{r}(\infty)$ and $\mathbf{r}(0)$ are independent and (2) that the protein properties are independent of the time when they are observed so that $\langle r^n(\infty)\rangle = \langle r^n(0)\rangle = \langle r^n\rangle$. $I_\infty$ is the product of two identical exponentials, one containing $i$ and the other $-i$.

To proceed we make a cumulant expansion of each expectation value in $I_\infty$,

$$\langle\exp(ix)\rangle = \exp[\sum_{n=1}^\infty \frac{(i)^n}{n!}\mu_n] \qquad (21)$$

where the $\mu_n$ are cumulants and $x = \pm(\mathbf{Q}\cdot\mathbf{r})$. The cumulants $\mu_n$ are a combination of moments,

$$\begin{aligned} \mu_1 &= \langle x \rangle \\ \mu_2 &= \langle x^2\rangle - \langle x\rangle^2 \\ \mu_3 &= \langle x^3\rangle - 3\langle x^2\rangle\langle x\rangle + 2\langle x\rangle^3 \\ \mu_4 &= \langle x^4\rangle - 3\langle x^2\rangle^2 - 4\langle x^3\rangle\langle x\rangle + 12\langle x^2\rangle\langle x\rangle^2 - 6\langle x\rangle^4. \end{aligned} \quad (22)$$

Because of the $\pm i$ in the exponentials in Eq. (20), the odd cumulants cancel. In $I_\infty$ only even cumulants appear. Thus,

$$\begin{aligned} I_\infty &= \langle\exp(-i\mathbf{Q}\cdot\mathbf{r})\rangle\langle\exp(i\mathbf{Q}\cdot\mathbf{r})\rangle = \exp[2\sum_{n\ even}^\infty \frac{(i)^n}{n!}\mu_n] \\ &= \exp[-\mu_2 + \frac{1}{12}\mu_4 - \frac{2}{6!}\mu_6 + ...] \end{aligned} \quad (23)$$

We also take $\langle\mathbf{r}\rangle = 0$. Thus up to fourth order,

$$I_\infty = \exp[-\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle + \frac{1}{12}[\langle[\mathbf{Q}\cdot\mathbf{r}]^4\rangle - 3\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle^2]] \quad (24)$$

On neglecting the fourth order cumulant and assuming cubic symmetry so that $\langle[\mathbf{Q}\cdot\mathbf{r}]^2\rangle = Q^2\langle r_z^2\rangle = \frac{1}{3}Q^2\langle r^2\rangle$,

where the z axis is chosen parallel to $\mathbf{Q}$, we obtain $I_\infty$ in Eq. (19). The fourth order cumulant will be small if $\mathbf{Q}$ is small or if the motional distribution of the H atom is approximately Gaussian.

Eq. (19) defines the intrinsic $\langle r^2 \rangle$. This $\langle r^2 \rangle$ is the same as that defined in simulations as the $t \to \infty$ limit of $\frac{1}{2}\langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle$ since $\langle \mathbf{r}(t) \cdot \mathbf{r}(0) \rangle = 0$ at $t \to \infty$ and $\langle r^2(\infty) \rangle = \langle r^2(0) \rangle = \langle r^2 \rangle$. Explicitly, the intrinsic value of $\langle r^2 \rangle$ is defined as the value that appears in the Debye-Waller factor $I_\infty$.

[1] W. Doster, S. Cusack, and W. Petry, Nature (London) **337**, 754 (1989).
[2] W. Doster, S. Cusack, and W. Petry, Phys. Rev. Lett. **65**, 1080 (1990).
[3] G. Zaccai, Science **288**, 1604 (2000).
[4] S. Dellerue, A. Petrescu, J. C. Smith, S. Longeville, and M. C. Bellissent-Funel, Physica B **276**, 514 (2000).
[5] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik, and G. Zaccai, Quart. Rev. Biophys. **35**, 327 (2002).
[6] V. G. Sakai and A. Arbe, Current Opinion in Colloid & Interface Science **14**, 381 (2009).
[7] W. Doster, Biochim. Biophys. Acta **1804**, 3 (2010).
[8] A. Paciaroni, S. Cinelli, and G. Onori, Biophys. Journ. **83**, 1157 (2002).
[9] G. Chen, P. W. Fenimore, H. Frauenfelder, F. Mezei, J. Swenson, and R. D. Young, Phil. Mag. **88**, 3877 (2008).
[10] S. Magazù, F. Migliardo, and A. Benedetto, J. Phys. Chem. B **114**, 9268 (2010).
[11] S. Magazù, F. Migliardo, and A. Benedetto, J. Phys. Chem. B **115**, 7736 (2011).
[12] D. J. Bicout and G. Zaccai, Biophys. Journ. **80**, 1115 (2001).
[13] J. H. Roh, V. N. Novikov, R. B. Gregory, J. E. Curtis, Z. Chowdhuri, and A. P. Sokolov, Phys. Rev. Lett. **95**, 038101 (2005).
[14] S. Khodadadi, S. Pawlus, J. H. Roh, V. G. Sakai, E. Marnontov, and A. P. Sokolov, J. Chem. Phys. **128**, 195106 (2008).
[15] L. Hong, N. Smolin, B. Lindner, A. P. Sokolov, and J. C. Smith, Phys. Rev. Lett. **107**, 148102 (2011).
[16] T. Becker, J. A. Hayward, J. L. Finney, R. M. Daniel, and J. C. Smith, Biophys. Journ. **87**, 1436 (2004).
[17] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[18] M. Ferrand, A. J. Dianoux, W. Petry, and G. Zaccai, Proc. Nat. Acad. Sci. U.S.A. **90**, 9668 (1993).
[19] C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[20] J. Pieper, A. Buchsteiner, N. A. Dencher, R. E. Lechner, and T. Hauβ, Phys. Rev. Lett. **100**, 228103 (2008).
[21] R. Biehl, B. Hoffmann, M. Monkenbusch, P. Falus, S. Préost, R. Merkel, and D. Richter, Phys. Rev. Lett. **101**, 138102 (2008).
[22] L. Zhu, D. Frenkel, and P. G. Bolhuis, Phys. Rev. Lett. **106**, 168103 (2011).
[23] K. Wood, S. Grudinin, B. Kessler, M. Weik, M. Johnson, G. R. Kneller, D. Oesterhelt, and G. Zaccai, J. Mol. Biol. **380**, 581 (2008).
[24] 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).
[25] H. Nakagawa, H. Kamikubo, and M. Kataoka, Biochim. Biophys. Acta **1804**, 27 (2010).
[26] L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney, and J. C. Smith, Biophys. Journ. **94**, 4812 (2008).
[27] G. R. Kneller and K. Hinsen, J. Chem. Phys. **131**, 045104 (2009).
[28] Z. Yi, Y. Miao, J. Baudry, N. Jain, and J. C. Smith, In Press (2012).
[29] G. R. Kneller, Phys. Chem. Chem. Phys. **7**, 2641 2005.
[30] G. R. Kneller, and V. Calandrini, J. Chem. Phys. **126**, 125107 2007.
[31] V. Calandrini, V. Hamon, K. Hinsen, P. Calligari, M.-C. Jain, and G. R. Kneller, Chem. Phys. **345**, 289 2008.
[32] R. M. Daniel, J. L. Finney, V. Reat, R. Dunn, M. Ferrand, and J. C. Smith, Biophys. Journ. **77**, 2184 (1999).
[33] P. Debye, Ann. Phys. **39**, 789 (1912).
[34] I. Waller, Z. Physik **51**, 213 (1928).
[35] A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, 1963).