THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

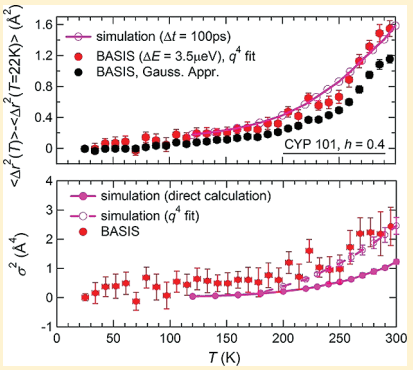
Derivation of Mean-Square Displacements for Protein Dynamics from Elastic Incoherent Neutron Scattering

Zheng Yi,[†,‡] Yinglong Miao,[†,‡] Jerome Baudry,[†,‡] Nitin Jain,[‡] and Jeremy C. Smith*[,†,‡]

[†]University of Tennessee/Oak Ridge National Laboratory Center for Molecular Biophysics, P.O. Box 2008, 1 Bethel Valley Road, Oak Ridge, Tennessee 37831, United States

[‡]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, Tennessee 37996, United States

ABSTRACT: The derivation of mean-square displacements from elastic incoherent neutron scattering (EINS) of proteins is examined, with the aid of experiments on camphor-bound cytochrome P450cam and complementary molecular dynamics simulations. It is shown that a q^4 correction to the elastic incoherent structure factor (where q is the scattering vector) can be simply used to reliably estimate from the experiment both the average mean-square atomic displacement, $\langle \Delta r^2 \rangle$ of the nonexchanged hydrogen atoms in the protein and its variance, σ^2. The molecular dynamics simulation results are in broad agreement with the experimentally derived $\langle \Delta r^2 \rangle$ and σ^2 derived from EINS on instruments at two different energy resolutions, corresponding to dynamics on the ∼100 ps and ∼1 ns time scales. Significant dynamical heterogeneity is found to arise from methyl-group rotations. The easy-to-apply q^4 correction extends the information extracted from elastic incoherent neutron scattering experiments and should be of wide applicability.



1. INTRODUCTION

Incoherent neutron scattering (INS) furnishes a method for the direct measurement of atomic motions on the pico- to nanosecond time scales and thus has been widely used to characterize the molecular dynamics of proteins, membranes, hydration water in biological systems[1−7] and even entire cells.[8,9] The incoherent scattering cross-section of hydrogen atoms is ∼40 times larger than that of other elements, and as hydrogen atoms are distributed nearly homogeneously within proteins, INS provides a global picture of the protein internal dynamics.

Many dynamic neutron scattering experiments have concentrated on the relatively intense elastic scattering component. One of the important quantities derivable from elastic incoherent neutron scattering (EINS) is the average mean-square displacement (MSD) of the hydrogen atoms, $\langle \Delta r^2 \rangle$.[3,10] The temperature dependence of $\langle \Delta r^2 \rangle$ has been of particular interest. For proteins, $\langle \Delta r^2 \rangle$ increases approximately linearly with temperature up to 120 K, consistent with harmonic dynamics. A first transition in the gradient occurs at T_m ∼120−160 K, is solvent independent, and has been attributed to the activation of methyl group rotations.[11−15] A second inflection at T_D ∼ 200−240 K is, in contrast, solvent dependent[3,6,10,16,17] and has also been detected with a variety of other techniques, such as Mössbauer spectroscopy,[18] optical absorption spectroscopy,[19] X-ray diffraction[20] and molecular dynamics (MD) simulation.[21−25] The T_D transition has received broad attention over the past 20 years.

In EINS, $\langle \Delta r^2 \rangle$ has been commonly extracted by assuming the Gaussian approximation to the q-dependence of elastic neutron scattering intensity (where q is the scattering vector), i.e., $S_{el}(q,\omega = 0) = e^{-q^2\langle \Delta r^2 \rangle/6}$.[26] However, strictly speaking the Gaussian approximation is valid only for samples in which the scattering atoms have identical $\langle \Delta r^2 \rangle$ and undergo certain types of motions, such as harmonic oscillations over the entire q range. In the case of hydrated proteins, the Gaussian approximation is valid only at small q values. At larger q, deviation of the EINS from the Gaussian approximation, i.e., non-Gaussianity, is non-negligible.

There are two possible origins of non-Gaussian behavior: the anharmonic dynamics of single atoms and the existence of a distribution of mean-square displacements among the atoms in the structure,[11,27−30] the latter identified as "dynamical heterogeneity".[11] Early work assumed only non-Gaussian single-atom dynamics, characterizing the dynamics with a two-site jump model, and neglected dynamical heterogeneity by assuming all hydrogen atoms to have identical motions.[3] However, studies using MD simulation showed that the dominant contribution to non-Gaussianity comes from dynamical heterogeneity.[11,30−32]

In previous, theoretical work, we proposed a simple method for analyzing experimental data using a correction to the Gaussian approximation by including a fourth-order term of the scattering vector.[27,33] The method allows both the mean $\langle \Delta r^2 \rangle$ and σ^2, the variance of $\langle \Delta r^2 \rangle$, to be extracted from the EINS data. Here, we examine the application to describe dynamics in

Received: October 31, 2011
Revised: March 16, 2012
Published: April 3, 2012

a test case—the temperature dependence of the dynamics of cytochrome P450cam. Cytochrome P450s are essential hemoprotein monooxygenases that catalyze a large variety of biochemical reactions included in carcinogenesis, drug metabolism, lipid and steroid biosynthesis, and degradation of pollutants.[34] Among these chemical reactions are hydroxylation, sulfoxidation, epoxidation, dehalogenation, deformylation, dealkylation, and C−C coupling.[35] Cytochrome P450cam (CYP101), an enzyme from *Pseudomonas putida* that catalyzes the regio- and stereospecific hydroxylation of camphor in its active site, has long served as a prototypical model system for studying P450s.[36,37] CYP101 is selected here as a convenient protein on which to test the methods, but the results are expected to have broad applicability in molecular biophysics.

The experiments were performed on two inelastic scattering instruments, of different energy resolutions, so as to separately determine the elastic scattering on approximately sub-100 ps and sub-ns time scales, and on dry and hydrated (0.4 g D_2O/g protein) CYP101 samples. Theoretical models using the Gaussian approximation or including the correction to $O(q^4)$ are examined and compared with the results from MD simulations. The MD trajectories allow the origin of the dynamical heterogeneity at higher temperatures to be elucidated. $\langle \Delta r^2 \rangle$ and σ^2 are extracted from the q^4-corrected model fitted to the experimental data and both are found to be in broad agreement with MD simulation results.

2. MATERIALS AND METHODS

2.1. Sample Preparation. Expression and purification of camphor-bound cytochrome P450 in hydrogenated form was carried out using previously published protocols.[38] Hydrogenated samples were exchanged extensively with D_2O buffer (50 mM potassium phosphate, 50 mM KCl, 2 mM camphor, 1 mM DTT, pH 7.4) to replace exchangeable hydrogen atoms with deuterium atoms and lyophilized for 2 days to remove the residual water. The buffer was diluted accordingly to maintain its concentration constant relative to the protein. 150 mg of dried protein sample was then hydrated to the target hydration level $h \approx 0.4$ (g water/g protein) at room temperature by exposing the samples to D_2O vapor in a closed desiccator. The sample hydration levels were determined by the mass change during hydration. Sample masses were measured before and after the neutron scattering experiments and no change of mass was detected. The powder samples were placed in the aluminum foil (3 cm × 10 cm) and packed in annular sample cans for the neutron measurements.

2.2. Neutron Scattering Experiments. Neutron scattering spectra were collected at the NG2 high-flux backscattering spectrometer (HFBS) in the Center for Neutron Research at National Institute of Standard and Technology with a fixed energy resolution of ~0.9 μeV (corresponding to an elastic time resolution of ~1 ns) and the backscattering spectrometer BASIS at the Spallation Neutron Source (SNS) with a resolution of ~3.5 μeV (corresponding to an elastic time resolution of ~100 ps).The accessible ranges of the scattering wave vector q were 0.6−1.4 Å^{-1} (HFBS) and 0.3−1.9 Å^{-1} (BASIS). Elastic scans were performed upon heating from 6 to 300 K (HFBS) and from 22 to 300 K (BASIS) at a heating rate of 1.0 K/min. No correction for multiple scattering needed to be applied since the neutron transmission was over 0.9.[39] The elastic neutron scattering spectra were corrected for cell scattering and normalized to the lowest temperature data point for each q value and then to the value at q = 0.

2.3. Molecular Dynamics Simulations. For the MD simulations, the recent X-ray crystal structure of camphor-bound CYP101 (PDB ID: 3L63) was used[40] (Figure 1a). The

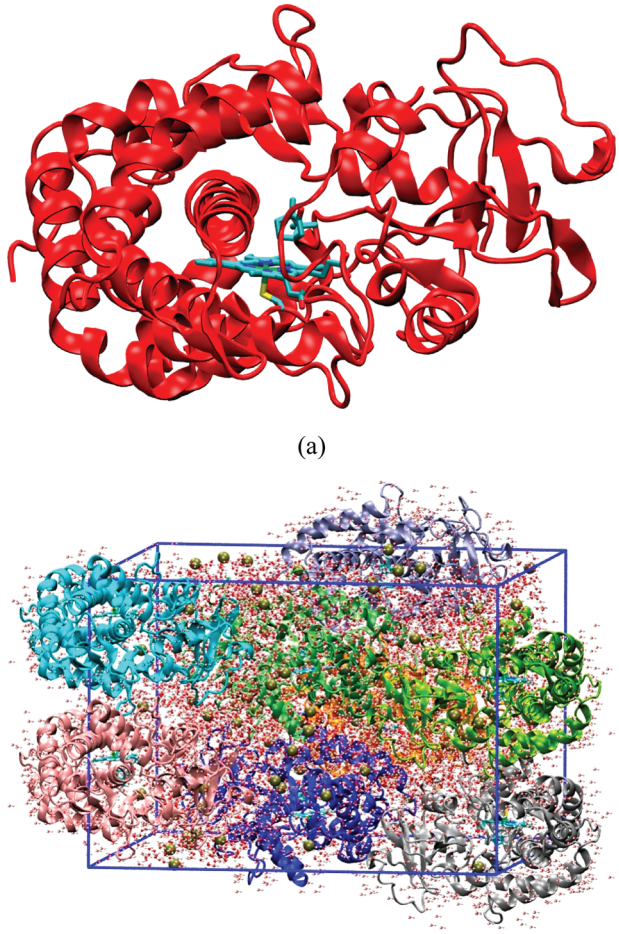


Figure 1. (a) X-ray crystal structure of 3L63 camphor-bound CYP101 (PDB ID: 3L63), in which the missing nine N-terminal residues were added. The heme and camphor are shown in cyan color. (b) Schematic representation of model of hydrated powder (h = 0.4 g D_2O/g protein) used for the MD simulation.

missing N-terminus (residues 1−9) was added as a random coil extending above Phe26 and Leu45 using program MOE[41] with energy minimization.[42] The *psfgen* plugin in VMD[43] was used to build the protein structural topology with protein residues set to the standard CHARMM protonation states at neutral pH with the following exceptions according to pK_a calculations of ref 44: Asp297 and Glu366 are protonated, His62 is protonated on δ-nitrogen, and seven other histidines (17, 21, 80, 176, 308, 337, and 355) are doubly protonated. The heme iron was coordinated to the sulfur atom of Cys357, providing a five-coordinate high-spin state. A unit cell containing four proteins was created by applying transformation matrices and was replicated in the z direction to form an "axbx2c" lattice. A water box was then overlaid on the protein lattice, and all but the required number of water molecules closest to the protein molecules were removed to give a hydration level h = 0.4 g water/g protein. The hydration comprised 3616 crystallographically identified and 4014 additional water molecules. The protein charge was neutralized with 0.1 M KCl. A schematic

dx.doi.org/10.1021/jp2102868 | J. Phys. Chem. B 2012, 116, 5028−5036

representation of the final simulation model is shown in Figure 1b.

NAMD[45] was used to perform MD simulation on the above system using the CHARMM22 force field[46,47] for the protein, CHARMM parameters for camphor,[48] and the TIP3P model for the water molecules.[49] Standard CHARMM force field parameters[47] were used for the heme group (toppar_all22_prot_heme.str) including an explicit Fe−S bond to Cys357, except that atomic partial charge of SG atom in Cys357 was adjusted to −0.07e to account for the resting state of CYP101.[44] The system was first equilibrated for 100 ns in the NPT ensemble (300 K and 1 atm), followed by NPT MD simulations from 300 to 220 K at 10 K intervals and then 220 to 120 K at 20 K intervals by cooling the equilibrated 300 K configurations at −0.1 K/ps rate to each of the desired temperatures. At each temperature, the system was equilibrated for 1 ns and then a 2 ns production run was performed for analysis. The MSD of the nonexchangeable protein hydrogen atoms was derived using the g_msd module in the GROMACS package.[50]

3. THEORETICAL APPROACH

3.1. Dynamic Structure Factors. The experimentally determined quantity is the incoherent neutron scattering dynamic structure factor $S(\vec{q},\omega)$, where $\hbar\vec{q}$ is the momentum transfer and $\hbar\omega$ is the energy transfer of the scattered neutrons. $S(\vec{q},\omega)$ is the time Fourier-transform of the intermediate scattering function $I(\vec{q},t)$

$$S(\vec{q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{d}t\, \mathrm{e}^{-i\omega t} I(\vec{q},t) \tag{1}$$

with, assuming scattering only from hydrogen atoms

$$I(\vec{q},t) = \frac{1}{N} b \sum_{\alpha} \langle \mathrm{e}^{-i\vec{q}\cdot\vec{r}_\alpha(0)} \mathrm{e}^{i\vec{q}\cdot\vec{r}_\alpha(t)} \rangle \tag{2}$$

where N is the number of hydrogen atoms in the system, b is the incoherent scattering length of hydrogen (and is omitted in the following discussion) and $\vec{r}_\alpha(t)$ denotes the time-dependent position vector of hydrogen atom α. The brackets $\langle ... \rangle$ express the ensemble average.

At finite instrumental energy resolution, $\Delta\omega$, the experimentally accessible scattering function in ω-space is the convolution of $S(\vec{q},\omega)$ with the instrumental resolution function $R(\omega;\Delta\omega)$:

$$\begin{aligned} S_R(\vec{q},\omega;\Delta\omega) &= S(\vec{q},\omega) \otimes R(\omega;\Delta\omega) \\ &= \int_{-\infty}^{+\infty} S(\vec{q},\omega-\omega') R(\omega';\Delta\omega)\, \mathrm{d}\omega' \end{aligned} \tag{3}$$

In EINS, the measured elastic intensity $S_{el}(\vec{q},\omega = 0;\Delta\omega)$ is the convolution product of the theoretical $S_{el}(\vec{q},\omega = 0)$ with $R(\omega;\Delta\omega)$:

$$\begin{aligned} S_{el}(\vec{q},\omega = 0;\Delta\omega) &= S(\vec{q},0) \otimes R(\omega;\Delta\omega) \\ &= \int_{-\infty}^{+\infty} S(\vec{q},\omega') R(\omega';\Delta\omega)\, \mathrm{d}\omega' \end{aligned} \tag{4}$$

The dynamic structure factor for elastic scattering $S_{el}(\vec{q},\omega = 0;\Delta\omega)$ is roughly equivalent to the intermediate scattering function $I(\vec{q},t_R)$ at t_R:[51]

$$S_{el}(\vec{q},\omega = 0;\Delta\omega) \approx I(\vec{q},t_R) = \frac{1}{N}\sum_{\alpha} \langle \mathrm{e}^{i\vec{q}\cdot[\vec{r}_\alpha(t_R)-\vec{r}_\alpha(0)]} \rangle \tag{5}$$

where $\Delta\omega$ is the instrumental energy resolution and $t_R = (1/\Delta\omega)$ determines the time scale of the experimentally observable motion.

3.2. Gaussian Approximation. The method that has been most commonly used for deriving the MSD from EINS measurements involves two approximations.[6,10] First, it is assumed that all hydrogen atoms have identical MSDs such that eq 5 can be rewritten as

$$S_{el}(\vec{q},\omega = 0) = \frac{1}{N}\sum_{\alpha} \langle \mathrm{e}^{i\vec{q}\cdot\Delta\vec{r}_\alpha(t_R)} \rangle \approx \langle \mathrm{e}^{i\vec{q}\cdot\Delta\vec{r}(t_R)} \rangle \tag{6}$$

where the mean-square displacements of individual atoms, $\langle \Delta\vec{r}_\alpha{}^2 \rangle$, are replaced by the average mean square displacement $\langle \Delta\vec{r}^2 \rangle$.[52] Then, using the cumulant expansion,[26,53] we have

$$\langle \mathrm{e}^{i\vec{q}\cdot\Delta\vec{r}(t_R)} \rangle = \mathrm{e}^{-\vec{q}^2\rho_2(t_R)+\vec{q}^4\rho_4(t_R)\mp\ldots} \tag{7}$$

where the cumulants are defined as follows:

$$\begin{aligned} \rho_2(t) &= \frac{1}{2!}\left\langle \Delta\vec{r}^2 \right\rangle \\ \rho_4(t) &= \frac{1}{4!}[\langle \Delta\vec{r}^4 \rangle - 3\langle \Delta\vec{r}^2 \rangle^2] \\ &\vdots \end{aligned} \tag{8}$$

The odd terms in q in the expansion vanish. In the Gaussian approximation, the above expansion is truncated after the q^2 term.

$$S_{el}(\vec{q},\omega = 0) = \mathrm{e}^{-1/2\vec{q}^2\langle \Delta\vec{r}^2 \rangle}$$

and, if the MSD of each atom is assumed to be isotropic:

$$S_{el}(q,\omega = 0) = \mathrm{e}^{-1/6q^2\langle \Delta r^2 \rangle} \tag{9}$$

Some publications have used $^1/_3$ as the factor in the exponent of eq 9,[6] and in this case, assuming harmonic dynamics, $\langle \Delta r^2 \rangle$ represents the mean-square fluctuation (MSF) from the average position $\langle [r(t) - \langle r \rangle]^2 \rangle$.[52,54] The MSF is commonly used in the Debye−Waller factor in the harmonic approximation and it is then half of the MSD, i.e., $\langle [r(t) - \langle r \rangle]^2 \rangle = (\langle [r(t) - r(0)]^2 \rangle / 2)$.

The Gaussian approximation is strictly valid for certain model systems, such as ideal gases and the harmonic oscillator. For hydrated protein systems, it is valid at short times or small q. The fitting of eq 9 to $S_{el}(q,\omega = 0)$ versus q^2 on a semilog scale has been frequently employed to estimate the MSDs of proteins and determine the temperature dependence thereof.

3.3. Non-Gaussianity. We now briefly recapitulate the method proposed previously[27] for treating non-Gaussian scattering. At high q, deviations of the EINS from the Gaussian approximation are non-negligible. As mentioned in the Introduction, there are two possible reasons for non-Gaussian behavior: single-atom non-Gaussian scattering and dynamic heterogeneity.

Non-Gaussian single-atom dynamics leads to the contribution of higher-order terms being non-negligible. Equation 7 can then be expanded as

dx.doi.org/10.1021/jp2102868 | J. Phys. Chem. B 2012, 116, 5028−5036

$$S_{el}(q, \omega = 0) = e^{-1/6q^2\langle\Delta r^2\rangle}[1 + \sum_{m=2}^{\infty} \rho_m(-q^2)^m] \quad (10)$$

Retaining only the $m = 2$ term $S_{el}(q,\omega = 0)$ can be written as

$$S_{el}(q, \omega = 0) = e^{-1/6q^2\langle\Delta r^2\rangle}\left[1 + \frac{1}{72}(\langle\Delta r^4\rangle - \langle\Delta r^2\rangle^2)q^4\right] \quad (11)$$

Dynamical heterogeneity also leads to non-Gaussian behavior of the EINS. The correction can be made as follows

$$\begin{aligned} S(q, \omega = 0) &= e^{-1/6q^2\langle\Delta r^2\rangle}\left[\frac{1}{N}\sum_{\alpha=1}^{N} e^{-1/6q^2(\langle\Delta r_\alpha{}^2\rangle - \langle\Delta r^2\rangle)}\right] \\ &= e^{-1/6q^2\langle\Delta r^2\rangle}\left[\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{-q^2}{6}\right)^m \mu(m)\right] \\ &= e^{-1/6q^2\langle\Delta r^2\rangle}\left(\frac{1}{N}\sum_{\alpha=1}^{N} 1 - \frac{q^2}{6}\frac{1}{N}\right. \\ &\quad \sum_{\alpha=1}^{N}(\langle\Delta r_\alpha{}^2\rangle - \langle\Delta r^2\rangle) + \frac{1}{2}\left(\frac{-q^2}{6}\right)^2 \\ &\quad \frac{1}{N}\sum_{\alpha=1}^{N}(\langle\Delta r_\alpha{}^2\rangle - \langle\Delta r^2\rangle)^2 - \frac{1}{3}\left(\frac{-q^2}{6}\right)^3 \\ &\quad \left.\frac{1}{N}\sum_{\alpha=1}^{N}(\langle\Delta r_\alpha{}^2\rangle - \langle\Delta r^2\rangle)^3 + \ldots\right) \\ &\approx e^{-1/6q^2\langle\Delta r^2\rangle}\left(1 + \frac{q^4}{72}\sigma^2\right) \end{aligned} \quad (12)$$

where μ_m is the mth central moment of the distribution of $\langle\Delta r^2\rangle$, and σ^2 is the variance, $(1/N)\sum_{\alpha=1}^{N}(\langle\Delta r_\alpha{}^2\rangle - \langle\Delta r^2\rangle)^2$. Equation 12 is valid if $(-q^2/6)^m\,\mu(m) \ll 1$.

Both eq 11 and eq 12 have similar expressions, indicating the general validity of the proposed q^4 model. In systems in which dynamical heterogeneity is the dominant contribution to non-Gaussian behavior, the elastic scattering can be in principle used to obtain experimentally the variance and higher statistical moments of the distribution of mean-square displacements of individual atoms. Previous studies have shown that, at small q, non-Gaussian behavior in a protein indeed arises mainly from dynamical heterogeneity.[27,32] Therefore, in the present work eq 12 is fitted to the experimental data to derive $\langle\Delta r^2\rangle$ and σ^2. It should be noted that the q^4 correction to Gaussian approximation does not presume any particular form of the distribution of MSDs. The variance of the MSD of all hydrogen atoms σ^2, is a general measure of the width of the MSD distribution. Alternative approaches have used predefined distribution models to interpret the experimental EINS data, and require a number of variables to be fitted.[32,55]

4. RESULTS AND DISCUSSION

4.1. $\langle\Delta r^2\rangle$ Derived Making the Gaussian Approximation. The experiments were performed on two inelastic scattering instruments, of different energy resolution—the backscattering instruments BASIS ($\Delta E \sim 3.5$ μeV) and HFBS ($\Delta E \sim 0.9$ μeV)—so as to separately determine the elastic scattering on the approximately sub-100 ps and sub-ns time scales. Figure 2a shows the elastic intensities $S_{el}(q,\omega = 0)$ of hydrated camphor-bound CYP101 ($h = 0.4$) obtained on BASIS as a function of q^2 at selected temperatures between 100 and 300 K. The data were first normalized by the scattering at the lowest temperature, 22 K, and then further normalized by a constant factor to ensure $S_{el}(q = 0, \omega = 0) = 1$. The dashed lines are the results of fitting eq 9, which uses the Gaussian approximation. Since the approximation is valid only for $q \to 0$, the fit was confined to the q^2 range 0.1−1.7 Å^{-2}. The deviations of the experimental profiles from the fitted lines at higher q^2 (∼1.7−2.9 Å^{-2}) illustrate the limitations of the Gaussian approximation. The inset shows the elastic scattering as a function of temperature.

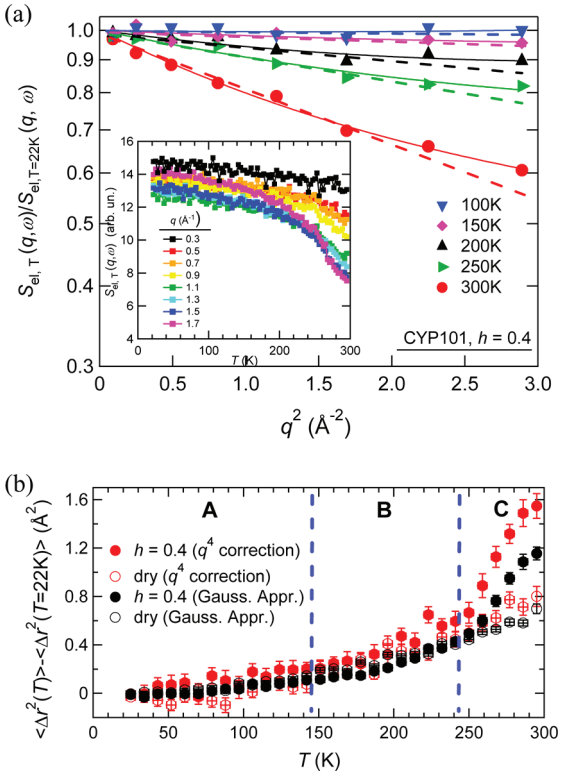


Figure 2. (a) Elastic intensity $S_{el}(q,\omega = 0)$ of hydrated camphor-bound CYP101 ($h = 0.4$) obtained on the BASIS spectrometer at SNS with a resolution of 3.5 μeV (fwhm, ∼ 100 ps). The data were first normalized by the scattering at the lowest temperature, 22 K, and then further normalized by a constant factor to ensure $S_{el}(q = 0, \omega = 0) = 1$. The dashed lines are the fits using the Gaussian approximation (eq 9) in the q range up to $q^2 \sim 1.7$ Å^{-2} and the solid lines are the fits using q^4 correction (eq 12). The inset shows the elastic scattering as a function of temperature. (b) $\langle\Delta r^2(T)\rangle$ of dry and hydrated proteins as a function of temperature from the two fitting methods.

$\langle\Delta r^2\rangle$ of dry and hydrated camphor-bound CYP101 resulting from the above Gaussian-approximation fits are plotted versus T in Figure 2b (in black). $\langle\Delta r^2\rangle$ here represents the average displacement of all hydrogen atoms arising from motions on time scales faster than ∼100 ps, the effective time resolution of BASIS. $\langle\Delta r^2\rangle$ of both samples increases roughly linearly as a function of T at low T (<∼140 K). The first small but significant deviation from linear behavior appears at $T_m \sim 140$ K. This transition has been observed in other proteins and attributed to methyl rotations.[11−13] A second transition appears at $T_D \sim 240$ K and has also been observed in other proteins.[3,56,57] The ∼240 K dynamic transition is observed

dx.doi.org/10.1021/jp2102868 | J. Phys. Chem. B 2012, 116, 5028−5036

Table 1. Resilience Force Constants $\langle k \rangle$ Obtained in Three Temperature Regimes

		$\langle k \rangle$ (22 K < T < 140K) [N/m]	$\langle k' \rangle$ (140 K < T < 240 K) [N/m]	$\langle k'' \rangle$ (240 K < T < 300 K) [N/m]
Gauss. approx.	h = 0.4	1.27 ± 0.30	0.37 ± 0.05	0.11 ± 0.01
	dry	1.27 ± 0.17	0.46 ± 0.02	
q^4 correction	h = 0.4	1.03 ± 0.33	0.29 ± 0.04	0.08 ± 0.01
	dry	1.49 ± 0.60	0.33 ± 0.03	

only in the hydrated protein and not in the dry sample, again consistent with previous results which have shown this transition to be solvent driven.[6,58]

4.2. $\langle \Delta r^2 \rangle$ Derived Using the q^4 Correction to $S_{el}(q,\omega = 0)$. In this section, we apply the q^4 correction using eq 12 to fit $S_{el}(q,\omega = 0)$. The solid lines in Figure 2a show the results of applying the q^4 correction using eq 12 to fit $S_{el}(q,\omega = 0)$. Fits with higher-order terms (up to $O(q^8)$) were also tried but resulted in negligible improvement in the fit quality. Equation 12 fits the data well and the resulting $\langle \Delta r^2 \rangle$ is plotted as a function of T in Figure 2b (in red). At low T $\langle \Delta r^2 \rangle$ is somewhat noisier than the data derived using the Gaussian approximation. For T > 200 K, $\langle \Delta r^2 \rangle$ obtained using the q^4 correction is significantly larger than that obtained from the Gaussian approximation, reaching ~1.55 Å^2 at 300 K, compared with ~1.2 Å^2 with the Gaussian approximation.

To further quantify the difference between the two analytical methods, a resilience force constant, $\langle k \rangle$, i.e., the slope of $\langle \Delta r^2 \rangle$ versus T, was calculated as

$$\langle k \rangle = k_B/(\mathrm{d}\langle \Delta r^2 \rangle/\mathrm{d}T) \quad (13)$$

where k_B is the Boltzmann constant. The resilience constant has been used to characterize the "softness" of biological systems:[10] the smaller the force constant, the less resilient (i.e., the softer) the environment is. $\langle \Delta r^2 \rangle$ can be divided into three different temperature regions: region A ($T < T_m$), region B ($T_m < T < T_D$), and region C ($T > T_D$). $\langle k \rangle$ of the three regions derived from the two fitting methods are listed in Table 1. Using the q^4 correction, $\langle k \rangle$ of hydrated camphor-bound CYP101 in regions A, B, and C are 1.03 N/m, 0.29 N/m, and 0.08 N/m, respectively. Hence, the protein is relatively rigid at low T in the harmonic region, while the appearance of anharmonicities at higher temperatures softens the system. The same trend was found in the results from using the Gaussian approximation but with higher values in all regions, consistent with the smaller $\langle \Delta r^2 \rangle$.

4.3. $\langle \Delta r^2 \rangle$ at Different Energy Resolutions. To further examine the use of the q^4 correction for analyzing EINS data, spectra from camphor-bound CYP101 were also measured on HFBS at NIST, an instrument with a higher energy resolution (ΔE ~ 0.9 μeV) than BASIS. Figure 3a shows normalized $S_{el}(q,\omega = 0)$ as a function of q^2 at selected temperatures and the corresponding fits using eq 12. The agreement of the fitted curves with the experimental scattering data from both HFBS and BASIS shows that the scattering of camphor-bound CYP101 on both instruments can be appropriately described by applying the q^4 correction to the Gaussian approximation.

$\langle \Delta r^2 \rangle$ of dry and hydrated camphor-bound CYP101 obtained from both BASIS and HFBS are compared in Figure 3b. $\langle \Delta r^2 \rangle$ calculated from the MD simulations on the equivalent time scales of t = 100 ps and t = 1 ns are also plotted. In Figure 4, the time evolution of $\langle \Delta r^2(T) \rangle$ from the MD simulations at selected temperatures is plotted.

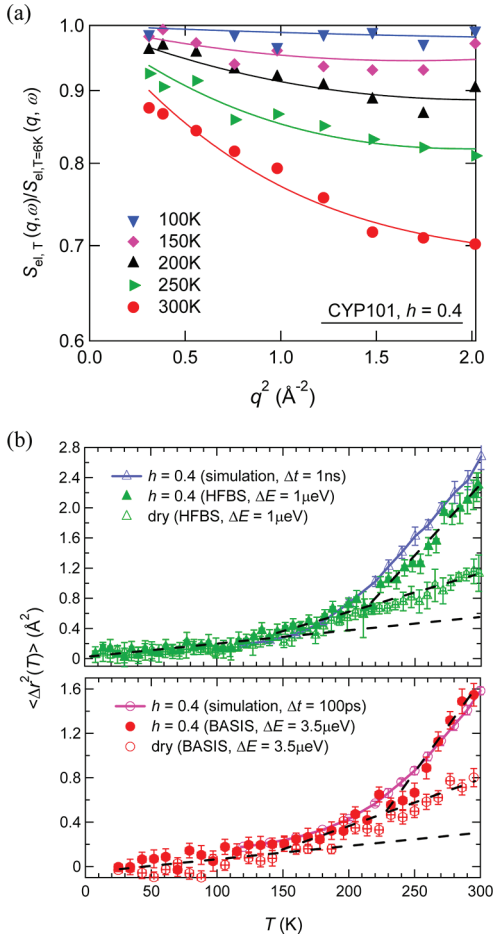


Figure 3. (a) $S_{el}(q,\omega = 0)$ of hydrated camphor-bound CYP101 (h = 0.4) using HFBS at NIST with a resolution of 0.9 μeV (fwhm, ~ 1 ns). Data are normalized by the data at the lowest temperature, 6 K. (b) $\langle \Delta r^2(T) \rangle$ of hydrated camphor-bound CYP101 from the MD simulations at t = 100 ps and 1 ns, from BASIS with a resolution of 3.5 μeV (fwhm, ~ 100 ps) and from HFBS with a resolution of 0.9 μeV (fwhm, ~1 ns). Dashed lines are linear fits in the suitable temperature regions.

$\langle \Delta r^2 \rangle$ obtained from the MD simulations closely match the results derived from BASIS and above T_m are about 10% higher than the experimental values of HFBS. An alternative method for calculating $\langle \Delta r^2 \rangle$ from the MD involves calculating $S_{el}(q,\omega = 0)$ from the MD trajectories and then fitting eq 12 to the resulting profile. This resulted in $\langle \Delta r^2 \rangle$ being ~10% lower than the direct calculation for HFBS, and ~5% for BASIS (results not shown). The difference is likely to arise in part from the fact that a Fourier transform with finite resolution in frequency space only approximately corresponds to a time cutoff, and also to the q^4 approximation inherent in eq 12.

In the temperature range from 6 to 140 K (T_m), the experimental $\langle \Delta r^2 \rangle$ obtained from both instruments are

dx.doi.org/10.1021/jp2102868 | J. Phys. Chem. B 2012, 116, 5028−5036

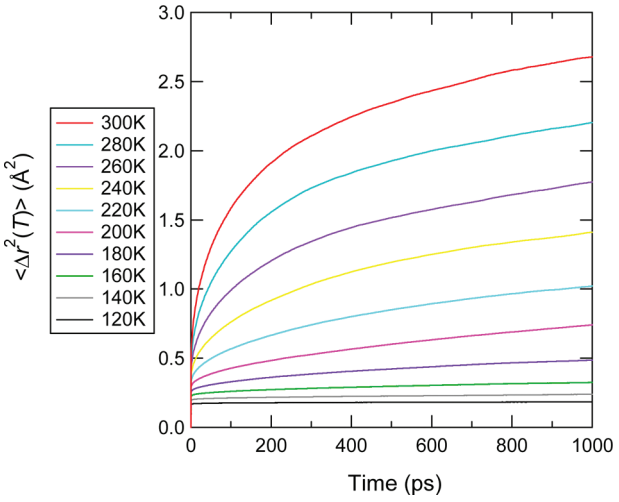


Figure 4. Time evolution of $\langle \Delta r^2(T) \rangle$ from the MD simulations at selected temperatures.

identical within error, and this is consistent with Figure 4, where for $T \leq 140$ K $\langle \Delta r^2 \rangle$ is almost time scale independent from 100 ps to 1 ns, suggesting that the dynamics is mostly described by fast, vibrational motions in the protein. For $T > T_m$, the displacements obtained from the higher-resolution instrument are higher than that obtained from BASIS, consistent with significant displacement occurring on the 100 ps to 1 ns time scale. As the temperature increases beyond 140 K so does the difference between $\langle \Delta r^2 \rangle$ on the two time scales in Figure 4, reaching $\langle \Delta r^2 \rangle_{1ns} \approx 2\langle \Delta r^2 \rangle_{100ps}$ at 300 K. The agreement between the MD simulation and the experiment confirms that the fitting of $S_{el}(q,\omega = 0)$ using eq 12 adequately represents the non-Gaussian behavior and can be used to estimate realistic approximate MSDs from EINS data.

The method used here to analyze the elastic scattering data yields not only $\langle \Delta r^2 \rangle$, but also the corresponding variance, σ^2. In Figure 5, we plot the variance σ^2, obtained from the fit of eq

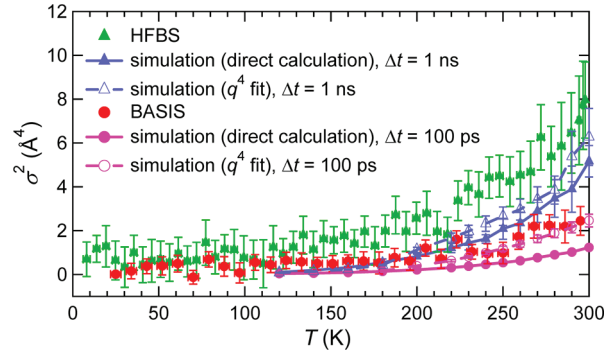


Figure 5. Variance σ^2 obtained from the experiments and the MD simulations versus T. The triangles and spheres are the data from HFBS and BASIS, respectively. The solid line is the simulation result directly calculated from the MD trajectories. The dashed line is calculated by fitting $S_{el}(q,\omega = 0)$ from the MD trajectories to eq 12.

12 to $S_{el}(q,\omega = 0)$ from both instruments, together with σ^2 calculated from the MD simulations.

σ^2 for the MD simulations was calculated in two different ways: one by direct calculation of the variance of $\langle \Delta r^2 \rangle$ from the MD trajectories (dashed line in Figure 5), and the other by calculating $S_{el}(q,\omega = 0)$ from the trajectories then, analogously to the experimental treatment, performing the q^4 fit (eq 12) to the result (solid line in Figure 5).

At $T < 140$ K, σ^2 from both instruments is low and noisy. However, for $T > 140$ K, σ^2 is significantly larger on HFBS, the higher-resolution (longer time scale) instrument. This change between different energy resolutions is also observed in σ^2 obtained from the MD simulations. The general agreement between simulation and experiment shows that the σ^2 extracted experimentally are physically meaningful.

As was seen for the MSD, there is a significant difference between σ^2 calculated directly from the MD and from the q^4 fit to the MD-derived $S_{el}(q,\omega = 0)$, arising in part from the arbitrary time cutoff and the q^4 approximation. However, the q^4 fit method clearly provides a useful quantitative estimate of σ^2. When comparing simulation with experiment further sources of error occur, such as in the simulation force fields and phase-space coupling, in the statistical accuracy of the experimental data, and in the normalization procedure.

A significant increase with T in the experimental σ^2 as seen in Figure 5 indicates increased dynamical heterogeneity in the system. However, as mentioned above, the q^4 correction to Gaussian approximation does not presume any particular form of the distribution of MSDs, such as a single-peak or multipeak distribution or Weibull distribution function.[32] Determining the shape of the distribution requires fitting predefined distribution models to the experimental EINS data and/or complementary MD simulations.

4.4. Distribution of $\langle \Delta r^2 \rangle_\alpha$ of Hydrogen Atoms. The MD simulations allow the individual mean-square displacements to be computed, and thus an analysis to be made of the distributions underlying the experimentally derived σ^2. Figure 6

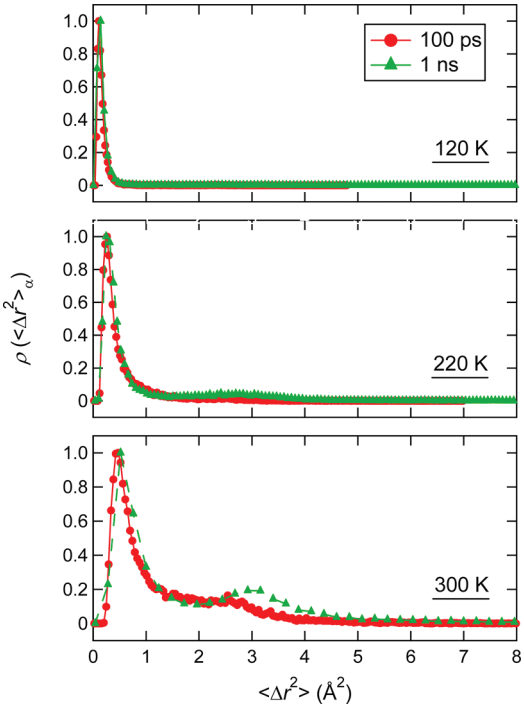


Figure 6. Distributions of $\langle \Delta r^2(T) \rangle_\alpha$ of nonexchangeable hydrogen atoms in camphor-bound CYP101 calculated from the 2 ns product MD runs at 120, 220, and 300 K temperatures with $\Delta t = 100$ ps corresponding to BASIS energy/time resolution and with $\Delta t = 1$ ns corresponding to HFBS energy/time resolution. The maxima of the distributions have been normalized to 1.

dx.doi.org/10.1021/jp2102868 | J. Phys. Chem. B 2012, 116, 5028–5036

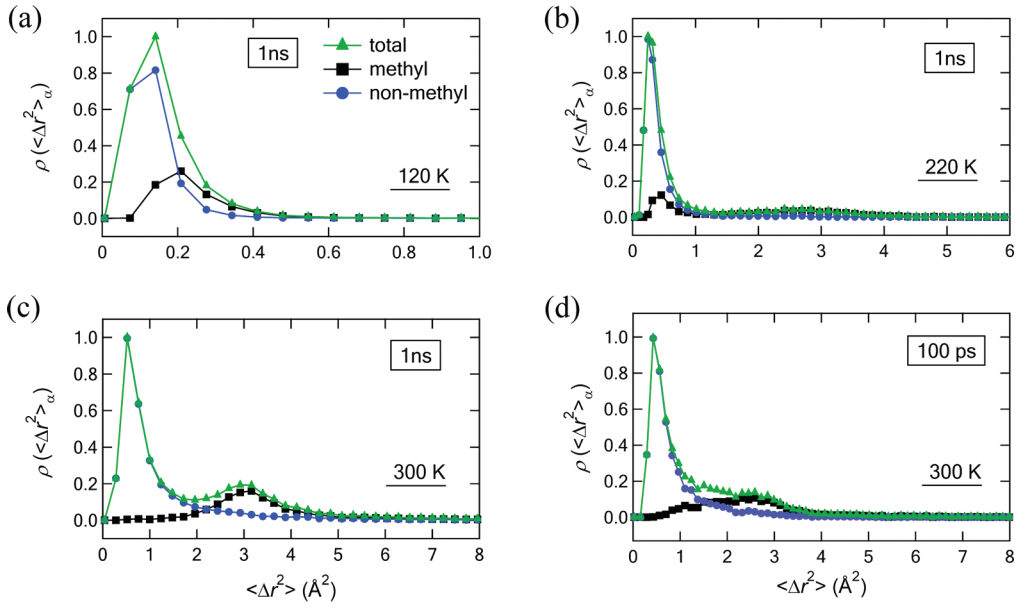


Figure 7. Distributions of $\langle \Delta r^2(T) \rangle_\alpha$ of (total, methyl, and nonmethyl) hydrogen atoms calculated from the MD simulation with Δt = 1 ns (120, 220, and 300 K) and Δt = 100 ps (300 K).

plots the distributions of the nonexchangeable hydrogen atoms $\langle \Delta r^2(T) \rangle_\alpha$ at 120, 220, and 300 K obtained from the MD simulations at $\Delta t \sim 100$ ps and $\Delta t \sim 1$ ns (corresponding to the energy resolutions of BASIS and HFBS, respectively). At T = 120 K, the distributions for $\Delta t \sim 100$ ps and $\Delta t \sim 1$ ns are almost identical, both displaying a single peak at ∼0.11 Å^2. At 220 K, this peak broadens and a further, very low and broad peak appears in the range of 1−4 Å^2 for only the $\Delta t \sim 1$ ns data. At 300 K, the new peak increases in intensity for $\Delta t \sim 1$ ns and also appears for $\Delta t \sim 100$ ps. These data indicate that, above T_D, motions on the time scale between ∼100 ps and ∼1 ns are significant and heterogeneous.

In previous work, the onset of anharmonic behavior at T_m was attributed to methyl group rotations.[11−15] Therefore, in Figure 7 we divide the protein nonexchangeable hydrogen atoms into those in the methyl groups and the remainder (i.e., nonmethyl), and plot the corresponding distributions of $\langle \Delta r^2(T) \rangle_\alpha$, calculated from the MD simulation with Δt = 1 ns.

At 120 K, the distribution of methyl hydrogen atoms has a single peak at ∼0.18 Å^2, a slightly higher value than that of the nonmethyl hydrogen atoms. At 220 K, the peak for the distribution of nonmethyl hydrogen atoms shifts to a higher value (∼0.25 Å^2) and becomes slightly broader. The distribution for the methyl hydrogen atoms splits into two parts: one peak <1.0 Å^2 from methyl groups not undergoing rotational jumps on the 1 ns time scale, and a broad shoulder in the range of 2−4 Å^2 arising from the activated methyl group rotations. When the temperature is increased to 300 K, the peak at <1 Å^2 disappears and the rotational motion leads to a pronounced peak at ∼3 Å^2 for methyl groups. In contrast to the methyl groups, the distribution of the nonmethyl hydrogen atoms remains centered at $\langle \Delta r^2(T) \rangle_\alpha$ < 1 Å^2, but becomes increasingly skewed with increasing T. This increased asymmetricity, together with the contribution of methyl hydrogen atoms, gives rise to the high dynamical heterogeneity and variance at 300 K.

To further examine the influence of the energy resolution on the dynamical heterogeneity and the MSD derived from EINS, we also plot the distribution of $\langle \Delta r^2(T) \rangle_\alpha$ of hydrogen atoms from methyl groups and the remainder at 300 K on the shorter time scale, i.e., Δt = 100 ps in Figure 7d. At 300 K, the distributions of the nonmethyl hydrogen atoms with Δt = 100 ps and Δt = 1 ns are closely similar. In contrast, for the methyl hydrogen atoms, the distribution at Δt = 100 ps is much broader, with significant probability at ∼1 Å^2, corresponding to the methyl groups with relatively high energy barriers that do not rotate. The distribution at Δt = 1 ns has only one peak, at ∼3.0 Å^2, indicating that most of methyl groups undergo rotational transitions within 1 ns. This heterogeneity in methyl rotational correlation times and corresponding energy barriers agree quantitatively with MD works on lysozyme and myoglobin.[6,15]

5. CONCLUSION

The interpretation of the elastic incoherent neutron scattering from biological systems has hitherto mostly invoked the Gaussian approximation to derive the average mean-square displacement, $\langle \Delta r^2 \rangle$ of the nonexchangeable hydrogen atoms. However, neglect of non-Gaussian behavior is valid only at $q \rightarrow 0$ and is found here to lead to an underestimation of $\langle \Delta r^2 \rangle$ at temperatures above T_m. The present results examine an alternative method of analyzing $S_{el}(q,\omega = 0)$ involving a simple correction of the Gaussian approximation to $O(q^4)$ that not only leads to a better fit to the data but also demonstrates that σ^2, the variance of $\langle \Delta r^2 \rangle$ can be conveniently extracted together with $\langle \Delta r^2 \rangle$ from the experiment.

The q^4 correction on camphor-bound CYP101 was examined here using the data from instruments with different energy resolutions: HFBS and BASIS. $S_{el}(q,\omega = 0)$ from both instruments are well described with the fit to $O(q^4)$. The MD simulations are used here to aid in interpretation of the experimental data and to independently calculate the derived quantities. Methyl-group rotational jumps contribute strongly to the dynamical heterogeneity, which is particularly marked in results from the higher-resolution instrument.

The temperature-dependent $\langle \Delta r^2 \rangle$ and σ^2 obtained from both instruments agree reasonably well with the results from the MD simulations on the corresponding 100 ps and 1 ns time

dx.doi.org/10.1021/jp2102868 | J. Phys. Chem. B 2012, 116, 5028−5036

scales. It is possible to incorporate dynamical heterogeneity effects into EISF analysis by assuming specific functional forms for the distribution of mean-square displacements, and this approach has recently been examined using both simulation[59] and experiment,[32] although not both together. However, the present results demonstrate that, without assuming a functional form of the distribution, the determination of both $\langle \Delta r^2 \rangle$ and σ^2 from elastic incoherent neutron scattering experiments is feasible and meaningful. The method extends the usefulness of elastic incoherent neutron scattering with a simple correction and as such should be of broad applicability.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project was supported by the National Science Foundation (NSF) award (MCB-0842871). We thank Eugene Mamontov at SNS on for assistance with BASIS experiments, and Liang Hong for valuable discussions. Computing time was provided by an NSF TeraGrid award (grant TG-MCA08 × 032) on the Kraken supercomputer at the University of Tennessee and a National Energy Research Scientific Computing Center (NERSC) award (project m906) on the Franklin and Hopper supercomputers.

REFERENCES

(1) Lovesey, S. *Theory of Neutron Scattering from Condensed Matter*; Oxford University Press: New York, 1987.
(2) Bee, M. *Quasielastic neutron Scattering: Principles and Applications in Solid State Chemistry, Biology and Materials Science*; Adam Hilger: Bristol and Philadelphia, PA, 1988.
(3) Doster, W.; Cusack, S.; Petry, W. *Nature* **1989**, *337*, 754.
(4) Smith, J. C. *Q. Rev. Biophys.* **1991**, *24*, 227.
(5) Magazu, S.; Maisano, G.; Migliardo, F.; Mondelli, C. *Biophys. J.* **2004**, *86*, 3241.
(6) Roh, J. H.; Curtis, J. E.; Azzam, S.; Novikov, V. N.; Peral, I.; Chowdhuri, Z.; Gregory, R. B.; Sokolov, A. P. *Biophys. J.* **2006**, *91*, 2573.
(7) Chen, S. H.; Liu, L.; Fratini, E.; Baglioni, P.; Faraone, A.; Mamontov, E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9012.
(8) Jasnin, M.; Moulin, M.; Haertlein, M.; Zaccai, G.; Tehei, M. *Biophys. J.* **2008**, *95*, 857.
(9) Jasnin, M.; Moulin, M.; Haertlein, M.; Zaccai, G.; Tehei, M. *Embo. Rep.* **2008**, *9*, 543.
(10) Zaccai, G. *Science* **2000**, *288*, 1604.
(11) Hayward, J. A.; Smith, J. C. *Biophys. J.* **2002**, *82*, 1216.
(12) Roh, J. H.; Novikov, V. N.; Gregory, R. B.; Curtis, J. E.; Chowdhuri, Z.; Sokolov, A. P. *Phys. Rev. Lett.* **2005**, 95.
(13) Kneller, G. R.; Doster, W.; Settles, M.; Cusack, S.; Smith, J. C. *J. Chem. Phys.* **1992**, *97*, 8864.
(14) Wood, K.; Tobias, D. J.; Kessler, B.; Gabel, F.; Oesterhelt, D.; Mulder, F. A. A.; Zaccai, G.; Weik, M. *J. Am. Chem. Soc.* **2010**, *132*, 4990.
(15) Krishnan, M.; Kurkal-Siebert, V.; Smith, J. C. *J. Phys. Chem. B* **2008**, *112*, 5522.
(16) Fitter, J.; Lechner, R. E.; Dencher, N. A. *Biophys. J.* **1997**, *73*, 2126.
(17) Gabel, F.; Bicout, D.; Lehnert, U.; Tehei, M.; Weik, M.; Zaccai, G. *Q. Rev. Biophys.* **2002**, *35*, 327.
(18) Parak, F. G. *Rep. Prog. Phys.* **2003**, *66*, 103.
(19) Melchers, B.; Knapp, E. W.; Parak, F.; Cordone, L.; Cupane, A.; Leone, M. *Biophys. J.* **1996**, *70*, 2092.
(20) Rasmussen, B. F.; Stock, A. M.; Ringe, D.; Petsko, G. A. *Nature* **1992**, *357*, 423.
(21) Wong, C. F.; Zheng, C.; Mccammon, J. A. *Chem. Phys. Lett.* **1989**, *154*, 151.
(22) Smith, J.; Kuczera, K.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 1601.
(23) Steinbach, P. J.; Loncharich, R. J.; Brooks, B. R. *Chem. Phys.* **1991**, *158*, 383.
(24) Tarek, M.; Tobias, D. J. *Biophys. J.* **2000**, *79*, 3244.
(25) Tournier, A. L.; Smith, J. C. *Phys. Rev. Lett.* **2003**, *91*, 208106.
(26) Rahman, A.; Singwi, K. S.; Sjolander, A. *Phys. Rev.* **1962**, *126*, 986.
(27) Becker, T.; Smith, J. C. *Phys. Rev. E* **2003**, *67*, 21904.
(28) Nakagawa, H.; T., A.; Kamikubo, H.; Joti, Y.; Kitao, A.; Kataoka, M. *Mater. Sci. Eng., A* **2006**, *442*, 5.
(29) Tokuhisa, A.; Joti, Y.; Nakagawa, H.; Kitao, A.; Kataoka, M. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2007**, *75*, 041912.
(30) Hayward, J. A.; Finney, J. L.; Daniel, R. M.; Smith, J. C. *Biophys. J.* **2003**, *85*, 679.
(31) Nakagawa, H.; Kamikubo, H.; Tsukushi, I.; Kanaya, T.; Kataoka, M. *J. Phys. Soc. Jpn.* **2004**, *73*, 491.
(32) Meinhold, L.; Clement, D.; Tehei, M.; Daniel, R.; Finney, J. L.; Smith, J. C. *Biophys. J.* **2008**, *94*, 4812.
(33) Becker, T.; Hayward, J. A.; Finney, J. L.; Daniel, R. M.; Smith, J. C. *Biophys. J.* **2004**, *87*, 1436.
(34) Lynch, T.; Price, A. *Am. Fam. Physician* **2007**, *76*, 391.
(35) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2010**, *110*, 949.
(36) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed; Ortiz de Montellano, P. R., Ed. Kluwer Academic/Plenum Publishers: New York, 2005.
(37) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, B. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615.
(38) Zhang, W.; Pochapsky, S. S.; Pochapsky, T. C.; Jain, N. U. *J. Mol. Biol.* **2008**, *384*, 349.
(39) Khodadadi, S.; Roh, J. H.; Kisliuk, A.; Mamontov, E.; Tyagi, M.; Woodson, S. A.; Briber, R. M.; Sokolov, A. P. *Biophys. J.* **2010**, *98*, 1321.
(40) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314.
(41) *MOE: Molecular Operating Environment version 2010*; Chemical Computing Group Ltd.: Montréal, Canada, 2010.
(42) Miao, Y.; Baudry, J. *Biophys. J.* **2011**, *101*, 1493.
(43) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
(44) Zheng, J. J.; Altun, A.; Thiel, W. *J. Comput. Chem.* **2007**, *28*, 2147.
(45) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, 1781.
(46) Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; et al. *Faseb. J.* **1992**, *6*, A143.
(47) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(48) Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 8142.
(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(50) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model* **2001**, *7*, 306.
(51) Doster, W.; Diehl, M.; Gebhardt, R.; Lechner, R. E.; Pieper, J. *Chem. Phys.* **2003**, *292*, 487.
(52) Kurkal-Siebert, V.; Daniel, R. M.; Finney, J. L.; Tehei, M.; Dunn, R. V.; Smith, J. C. *J. Non-Cryst. Solids* **2006**, *352*, 4387.
(53) Rog, T.; Murzyn, K.; Hinsen, K.; Kneller, G. R. *J. Comput. Chem.* **2003**, *24*, 657.
(54) Joti, Y.; Nakagawa, H.; Kataoka, M.; Kitao, A. *J. Phys. Chem. B* **2008**, *112*, 3522.

(55) Engler, N.; Ostermann, A.; Niimura, N.; Parak, F. G. *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 10243.
(56) Roh, J. H.; Briber, R. M.; Damjanovic, A.; Thirumalai, D.; Woodson, S. A.; Sokolov, A. P. *Biophys. J.* **2009**, *96*, 2755.
(57) Reat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D.; Zaccai, G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4970.
(58) Curtis, J. E.; Tarek, M.; Tobias, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 15928.
(59) Kneller, G. R.; Hinsen, K. *J. Chem. Phys.* **2009**, *131*, 045104.