Non-Gaussian behavior of elastic incoherent neutron scattering profiles of proteins studied by molecular dynamics simulation

Atsushi Tokuhisa,[1,2] Yasumasa Joti,[3,4] Hiroshi Nakagawa,[5] Akio Kitao,[3,4] and Mikio Kataoka[1,5,*]

[1]*Graduate School of Materials Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan*

[2]*Computational Biology Group, Neutron Biology Research Center, Quantum Beam Science Direction, Japan Atomic Energy Agency, 8-1 Umemidai, Kizu-cho, Soraku-gun, Kyoto 619-0215, Japan*

[3]*Laboratory of Molecular Design, Institute of Molecular and Cellular Biosciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan*

[4]*Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan*

[5]*Neutron Biophysics Group, Neutron Biology Research Center, Quantum Beam Science Direction, Japan Atomic Energy Agency, Tokai, Ibaraki 319-1195, Japan*

(Received 23 October 2006; published 24 April 2007)

Elastic incoherent neutron scattering (EINS) data can be approximated with a Gaussian function of q in a low q region. However, in a higher q region the deviation from a Gaussian function becomes non-negligible. Protein dynamic properties can be derived from the analyses of the non-Gaussian behavior, which has been experimentally investigated. To evaluate the origins of the non-Gaussian behavior of protein dynamics, we conducted a molecular dynamics (MD) simulation of staphylococcal nuclease. Instead of the ordinary cumulant expansion, we decomposed the non-Gaussian terms into three components: (i) the component originating from the heterogeneity of the mean-square fluctuation, (ii) that from the anisotropy, and (iii) that from higher-order terms such as anharmonicity. The MD simulation revealed various dynamics for each atom. The atomic motions are classified into three types: (i) "harmonic," (ii) "anisotropic," and (iii) "anharmonic." However, each atom has a different degree of anisotropy. The contribution of the anisotropy to the total scattering function averages out due to these differences. Anharmonic motion is described as the jump among multiple minima. The jump distance and the probability of the residence at one site vary from atom to atom. Each anharmonic component oscillates between positive and negative values. Thus, the contribution of the anharmonicity to the total scattering is canceled due to the variations in the anharmonicity. Consequently, the non-Gaussian behavior of the total EINS from a protein can be analyzed by the dynamical heterogeneity.

DOI: 10.1103/PhysRevE.75.041912 PACS number(s): 87.15.He, 61.12.Ex, 87.14.Ee

I. INTRODUCTION

Proteins are fascinating molecular machines in living systems. Their three-dimensional structures are the consequence of evolution to perform various biological functions. As seen in ordinary machines of our daily lives, the relationship between structure and function is linked through dynamics. Therefore, it is essential to understand the physical nature of protein dynamics in order to capture the nature of protein function.

Incoherent neutron scattering is suitable for studying protein dynamics between femto-second (fs) to nano-second (ns) time scales [1,2]. When elastic incoherent neutron scattering (EINS) data is measured as a function of the scattering vector q, it is approximated with a Gaussian function (Gaussian approximation) [3]. The EINS from a single harmonic oscillator is given with a Gaussian function. In a multidimensional system, the Gaussian approximation is valid for a homogeneous harmonic system in the entire q range or for any system with small q values. In the case of a protein, the EINS is mainly caused by the hydrogen atoms due to their large cross section of incoherent scattering. Each hydrogen atom shows a different dynamical motion. The mean-square fluctuation averaged over all the hydrogen atoms in protein, $\langle \Delta r^2 \rangle$, can be reasonably derived from the q dependence of the EINS data because the Gaussian approximation is always valid in the low q range [2].

The temperature dependence of $\langle \Delta r^2 \rangle$ provides information on the dynamical transition, which is an important property of protein dynamics. Various experimental techniques such as x-ray crystallography [4] and Mössbauer spectroscopy [5], as well as EINS [6,7], have been used to detect the inflection of $\langle \Delta r^2 \rangle$ at ~200 K. It has been argued that protein dynamics at lower temperatures can be nearly regarded as harmonic motions within one of the conformational substates and makes a transition to anharmonic diffusive motions that involve jumping among different substates at higher temperatures. Such diffusive anharmonic motions have been shown to be crucial to the function of bacteriorhodopsin in a purple membrane [7] and crystalline ribonuclease A [4].

At higher q regions, deviations of the EINS profile from the Gaussian approximation, namely the non-Gaussian behavior or the non-Gaussianity [8], is not negligible in a protein. The non-Gaussian behavior originates from either the anharmonicity, dynamical heterogeneity, or anisotropic fluctuation of atoms. In protein dynamics, all are plausible ori-

*Author to whom correspondence should be addressed. Mailing address: Graduate School of Materials Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan. Email address: kataoka@ms.naist.jp

gins. Doster *et al.* have analyzed the non-Gaussianity in terms of the anharmonicity in protein dynamics with a two-site jump model [6]. The estimated distance and enthalpy difference between the two sites were 1.5 Å and 12 kJmol^{-1}, respectively. The dynamical transition of a protein clearly indicates an activation of the anharmonic motions. Hence, the anharmonicity would be an origin of the non-Gaussianity, at least above the transition temperature. Recently, dynamical heterogeneity has been used to characterize the non-Gaussianity [8]. In the analysis, the distribution of the mean-square displacement (MSD) was explicitly introduced. Among the possible distributions of the MSD, a simple bimodal distribution was sufficient to explain the non-Gaussian behavior and the dynamical heterogeneity of a protein. The dynamical transition can be explained with a discontinuous increase of the average value of the MSD. Therefore, both analyses can satisfactorily explain the non-Gaussian behavior. However, each analysis has weaknesses. The former analysis assumes that all the hydrogen atoms have identical motions, which is unlikely in a real system, while the latter analysis assumes that all the hydrogen atoms are harmonic oscillators, although anharmonic motions are activated above the dynamical transition temperature.

Molecular dynamics simulations can reveal the motion of each hydrogen atom on a femto-second to a nano-second time scale. The effects of heterogeneity, anharmonicity, and anisotropy to the EINS profile can be directly estimated and quantitatively discussed. For these purposes, we decomposed the EINS profile into four terms instead of the ordinary cumulant expansion: a Gaussian, and the contributions from heterogeneity, anisotropy, and anharmonicity. Anharmonicity originates from the multiple-minima nature of the protein energy landscape [9]. In a nonparabolic energy surface, the fourth- or higher-order terms of atomic fluctuations do not vanish. On the other hand, heterogeneity originates from the atom-dependent nature of protein fluctuation. It should be noted that even a heterogeneous harmonic system can show a non-Gaussian behavior. In general, atomic fluctuations of a protein vary from atom to atom. The second-order moment of an atomic coordinate, i.e., the atomic mean-square fluctuation, strongly depends on the atom position [10,11]. We call this feature "the second-order heterogeneity" throughout this paper, although it is commonly termed "heterogeneity" in this field [8,12,13].

The combination of computer simulations with neutron scattering experiments allows a wide range of dynamical phenomena in condensed-phase biomolecular systems to be characterized [2,14]. Despite numerous efforts using computer simulations [12,13,15,16], the non-Gaussian behavior of the EINS profile has yet to be fully understood. In this paper, we demonstrate that the second-order heterogeneity is the dominant component of the non-Gaussian behavior of an EINS profile based on the results of molecular dynamics (MD) simulations on an enzyme, Staphylococcal nuclease (SNase). We also examine the non-Gaussian behavior of the EINS of each atom to reveal the effects of anharmonicity, heterogeneity, and anisotropy on the total EINS.

II. METHODS

A. Elastic incoherent structure factor (EISF)

The quantity of interest in the elastic incoherent neutron scattering (EINS) experiment is the elastic incoherent structure factor (EISF), $S(\mathbf{q},0)$, where $\mathbf{q}$ is the momentum transfer. Theoretically the EISF is given as [17]

$$S(\mathbf{q},0) = \frac{1}{N}\sum_k s_k(\mathbf{q},0) = \frac{1}{N}\sum_k |\langle e^{-i\mathbf{q}\cdot\mathbf{r}_k}\rangle|^2 = \overline{|\langle e^{-i\mathbf{q}\cdot\mathbf{r}_k}\rangle|^2} = \overline{s_k(\mathbf{q},0)}, \quad (1)$$

where N, $\mathbf{r}_k$, and $s_k(\mathbf{q},0)$ are the number of hydrogen atoms in the protein, the position vector of the kth hydrogen atom, and the atomic scattering function of the kth atom, respectively. The angle bracket indicates an ensemble average and the bar denotes the average over all the hydrogen atoms in the protein. Because the incoherent scattering length of a hydrogen atom is much larger than all the other atoms in a protein, the contribution from nonhydrogen atoms to the observed EISF is negligible and thus not considered in Eq. (1). According to Eq. (1), the EISF is easily calculated from the coordinate trajectory of the MD simulation. When the origin of the coordinate is chosen to be the average of position vector, $\langle \mathbf{r}_k \rangle$, Eq. (1) can be expressed as

$$S(\mathbf{q},0) = \frac{1}{N}\sum_k |\langle e^{-i\mathbf{q}\cdot\Delta\mathbf{r}_k}\rangle|^2 = \overline{|\langle e^{-i\mathbf{q}\cdot\Delta\mathbf{r}_k}\rangle|^2}, \quad (2)$$

where $\Delta\mathbf{r}_k$ is the displacement from the average position.

B. Gaussian approximation and polynomial expansion of the elastic incoherent structure factor

The EISF can be expressed as a one-dimensional Gaussian function of $q=|\mathbf{q}|$ in the Gaussian approximation as

$$S(q,0) = \exp\left(-\frac{1}{3}\overline{\langle \Delta r^2\rangle} q^2\right) \quad (3)$$

or

$$\ln S(q,0) = -\frac{1}{3}\overline{\langle \Delta r^2\rangle} q^2, \quad (4)$$

where $\Delta r=|\Delta\mathbf{r}|$. Here the assumed distribution of each atomic fluctuation is an isotropic Gaussian in three dimensions and $\overline{\langle \Delta r^2\rangle}/3=\overline{\langle \Delta x_k^2+\Delta y_k^2+\Delta z_k^2\rangle}/3$ is derived as the "effective" one-dimensional variance of the distribution. Therefore, plotting the observed $\ln S(q,0)$ as a function of q^2 can estimate the mean-square fluctuation $\overline{\langle \Delta r^2\rangle}$ as the slope of the regression line. This method is commonly employed to analyze the experimentally observed EISF. In order to clarify the non-Gaussian component of the EISF, we need to describe the index of an exponential function with a polynomial expression. Using the cumulant expansion, Eq. (2) can be rewritten as

$$S(\mathbf{q},0) = \overline{s_k(\mathbf{q},0)} = \overline{\{e^{-\gamma_{k,2}(\mathbf{n_q})q^2+\gamma_{k,4}(\mathbf{n_q})q^4-\gamma_{k,6}(\mathbf{n_q})q^6+\cdots}\}}, \quad (5)$$

where $q=|\mathbf{q}|$ and $\mathbf{n_q}$ is a unit vector along the vector $\mathbf{q}$ as $\mathbf{q}=q\mathbf{n_q}$. The coefficients, $\gamma_{k,2m}(\mathbf{n_q})$ $(m=1,2,3,\ldots)$, are de-

fined as the second- or higher-order moments,

$$\gamma_{k,2}(\mathbf{n_q}) = \langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^2\rangle,$$

$$\gamma_{k,4}(\mathbf{n_q}) = \frac{1}{12}\{\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^4\rangle - 3\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^2\rangle^2\},$$

$$\gamma_{k,6}(\mathbf{n_q}) = \frac{1}{360}\left\{\begin{array}{l}\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^6\rangle - 15\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^4\rangle\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^2\rangle \\ -10\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^3\rangle^2 + 30\langle(\mathbf{n_q}\cdot\Delta\mathbf{r}_k)^2\rangle^3\end{array}\right\}$$

$$\cdots. \tag{6}$$

To derive Eq. (5) and Eq. (6), the condition $\langle\Delta\mathbf{r}_k\rangle=\mathbf{0}$ is used. In the treatment by Doster *et al.* [6], all the hydrogen atoms in a protein are assumed to have the same $\gamma_{k,2m}(\mathbf{n_q})$ values, i.e., $\gamma_{k,2m}(\mathbf{n_q})$ does not depend on k. In this paper, we employ a rigorous treatment of the k dependence of $\gamma_{k,2m}(\mathbf{n_q})$. From Eq. (5), the natural logarithm of the EISF can be expanded as

$$\begin{aligned}\ln S(\mathbf{q},0) &= -\overline{\gamma_{k,2}(\mathbf{n_q})}q^2 \\ &+ \left[\frac{1}{2}\{\overline{\gamma_{k,2}^2(\mathbf{n_q})} - (\overline{\gamma_{k,2}(\mathbf{n_q})})^2\} + \overline{\gamma_{k,4}(\mathbf{n_q})}\right]q^4 \\ &- \left[\frac{1}{6}\{\overline{\gamma_{k,2}^3(\mathbf{n_q})} - 3(\overline{\gamma_{k,2}^2(\mathbf{n_q})})(\overline{\gamma_{k,2}(\mathbf{n_q})}) + 2(\overline{\gamma_{k,2}(\mathbf{n_q})})^3\}\right. \\ &\left. + \overline{\gamma_{k,2}(\mathbf{n_q})\gamma_{k,4}(\mathbf{n_q})} - (\overline{\gamma_{k,2}(\mathbf{n_q})})(\overline{\gamma_{k,4}(\mathbf{n_q})}) + \overline{\gamma_{k,6}(\mathbf{n_q})}\right]q^6 \\ &+ \cdots. \end{aligned} \tag{7}$$

In the low q range, the terms of q^4 and higher-order terms are expected to be negligible.

C. Decomposition of the elastic incoherent structure factor

In the high q range, $\ln S(q,0)$ deviates from the linear function of q^2 [6,8], indicating that the q^4 term and higher-order terms in Eq. (7) are not negligible. Although Eq. (7) is mathematically correct, the physical origins for the non-Gaussianity are not explicitly described. In order to clarify the physical origins of the non-Gaussianity, the contributions from the anharmonicity, heterogeneity, and anisotropy to the higher-order terms should be evaluated individually and explicitly. Thus, we decomposed Eq. (5) into four terms instead of the cumulant expansion as follows:

$$\begin{aligned}S(\mathbf{q},0) &= S^{\text{gauss}}(q,0) + S^{\text{2-hetero}}(q,0) + S^{\text{2-aniso}}(\mathbf{q},0) \\ &+ S^{\text{higher}}(\mathbf{q},0) = S^{\text{gauss}}(q,0) + \overline{s_k^{\text{2-hetero}}(q,0)} \\ &+ \overline{s_k^{\text{2-aniso}}(\mathbf{q},0)} + \overline{s_k^{\text{higher}}(\mathbf{q},0)},\end{aligned} \tag{8}$$

where

$$S^{\text{gauss}}(q,0) = \exp\left(-\frac{1}{3}\overline{\langle\Delta r^2\rangle}q^2\right), \tag{9}$$

$$s_k^{\text{2-hetero}}(q,0) = \exp\left(-\frac{1}{3}\langle\Delta r_k^2\rangle q^2\right) - \exp\left(-\frac{1}{3}\overline{\langle\Delta r^2\rangle}q^2\right), \tag{10}$$

$$s_k^{\text{2-aniso}}(\mathbf{q},0) = e^{-\gamma_{k,2}(\mathbf{n_q})q^2} - \exp\left(-\frac{1}{3}\langle\Delta r_k^2\rangle q^2\right), \tag{11}$$

$$s_k^{\text{higher}}(\mathbf{q},0) = s_k(\mathbf{q},0) - e^{-\gamma_{k,2}(\mathbf{n_q})q^2}. \tag{12}$$

$\langle\Delta r_k^2\rangle$ is the mean-square fluctuation of the kth hydrogen atom. The Gaussian approximation only considers the first term in Eq. (8). The second and the third terms represent contributions from isotropic and anisotropic heterogeneity, respectively. $S^{\text{2-hetero}}(q,0)$ corresponds to the "second-order heterogeneity" and $S^{\text{2-aniso}}(q,0)$ is the "second-order anisotropy." The fourth term represents other contributions, especially anharmonicity. In the previous work [8], only $S^{\text{gauss}}(q,0)$ and $S^{\text{2-hetero}}(q,0)$ were explicitly considered.

EINS experiments are typically performed with powder or solution samples. The observed EISF is, thus, the spatial average of $S(q,0)$, which leads to the expression of the EISF with $q=|\mathbf{q}|$ instead of $\mathbf{q}$.

D. Simulation procedure

We performed molecular dynamics (MD) simulations of Staphylococcal nuclease (SNase) in water using the program package AMBER7 [18] with the AMBER *ff94* force field [20] and the TIP3P water model [21]. A periodic boundary condition was imposed and the particle mesh Ewald method [19] was used (the Lennard-Jones interaction and the real space Ewald sum smoothly switched to 0 at 10 Å). The system was weakly coupled to a heat bath by the Berendsen method [22] with a relaxation time of 0.2 ps in order to maintain a constant temperature of 300 K. The Berendsen method was also used to maintain a constant pressure of 1 bar with a relaxation time of 0.2 ps. Hydrogen atoms were constrained using SHAKE [23].

The crystal structure of the SNase protein data bank (PDB [24] code: 1STN [25]) was employed as the initial coordinate of the simulation. Missing residues 1–5 and 142–149 were constructed using the NMR structures of SNase (PDB code: 1JOQ and 1JOR [26]). The system consisted of 149 residues (2395 atoms), 13 630 water molecules, and 8 chloride ions in a box ($79.1\times81.3\times81.3$ Å^3).

A MD simulation was performed by the following procedure: (1) The energy was minimized to optimize the positions of the hydrogen atoms; (2) 260-ps MD was performed to equilibrate the system to 300 K and 1 bar by gradually relaxing the restraints; (3) 1-ns MD was carried out with an integration time step of 2 fs. The coordinate trajectory was stored every 0.1 ps and was used in the analysis. It should be noted that the external motions of the protein were removed in the calculation. In this paper, we focus on the behavior of the EINS profile, which originates from the internal dynamics of the protein.

III. RESULTS AND DISCUSSION

A. Contribution of non-Gaussian components to the total elastic incoherent structure factor

Figure 1 shows the elastic incoherent neutron scattering profile calculated by a 1-ns MD simulation of SNase at

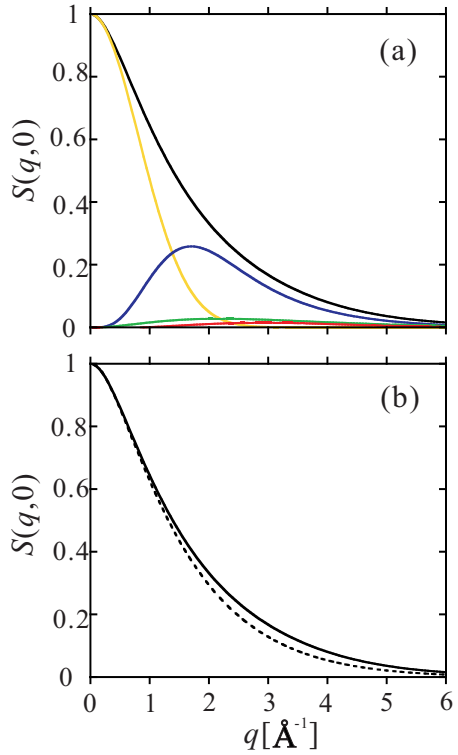

FIG. 1. (Color) (a) EISF as a function of q calculated using the result of 1-ns MD simulation at 300 K. Total $S(q,0)$ (black line) and its components $S^{\text{gauss}}(q,0)$ (yellow line), $S^{\text{2-hetero}}(q,0)$ (blue line), $S^{\text{2-aniso}}(q,0)$ (green line), and $S^{\text{higher}}(q,0)$ (red line) are shown. (b) Total $S(q,0)$ (solid line) and the sum of $S^{\text{gauss}}(q,0)$ and $S^{\text{2-hetero}}(q,0)$ (broken line) are shown.

300 K with their components, $S^{\text{gauss}}(q,0)$, $S^{\text{2-hetero}}(q,0)$, $S^{\text{2-aniso}}(q,0)$, and $S^{\text{higher}}(q,0)$. The contribution ratios of these components to the total EINS in area basis are 57.4%, 34.6%, 5.5%, and 2.5%, respectively. $S^{\text{2-aniso}}(q,0)$ is a broad maximum centered at $q=2.5$ Å^{-1} and gives a small flat base to the total $S(q,0)$ between $q=1$ and $q=5$ Å^{-1}. $S^{\text{higher}}(q,0)$ is negligible up to $q=1$ Å^{-1} and is a broad maximum centered at $q=3$ Å^{-1}. It also gives a small, flat base. The sum of the first two terms, $S^{\text{gauss}}(q,0)$ and $S^{\text{2-hetero}}(q,0)$, is dominant in the total $S(q,0)$ as shown in Fig. 1(b). The MD simulation clearly shows that the non-Gaussianity of $S(q,0)$ originates mainly from the second-order heterogeneity. The contributions from anisotropy and anharmonicity are negligible. Therefore, we conclude that the analysis of the non-Gaussianity with the dynamical heterogeneity [8] is valid even at room temperature.

B. Relation between the second-order heterogeneity and the atomic mean-square fluctuation

MD simulations provide a detailed trajectory for each atom. Therefore, the neutron scattering function from an individual atom can be rigorously examined. As shown in the previous section, $S^{\text{2-hetero}}(q,0)$ contributes to the total EINS by 34.6% in area basis. The contribution of each atomic component to $S^{\text{2-hetero}}(q,0)$ is considered by calculating the integral of $s_k^{\text{2-hetero}}(q,0)$ over q from $q=0$ to infinity. As defined in Eq. (10), $s_k^{\text{2-hetero}}(q,0)$ is closely related with the atomic mean-square fluctuation $\langle \Delta r_k^2 \rangle$. The integral ς_k is given as

$$\varsigma_k = \int_0^\infty s_k^{\text{2-hetero}}(q,0)dq. \tag{13}$$

This integral can be analytically calculated using Eq. (10) as

$$\varsigma_k = \frac{\sqrt{3\pi}}{2}[\{\langle \Delta r_k^2 \rangle\}^{-1/2} - \{\overline{\langle \Delta r^2 \rangle}\}^{-1/2}]. \tag{14}$$

Equation (14) shows that ς_k should be proportional to $\{\langle \Delta r_k^2 \rangle\}^{-1/2}$. In fact, the numerically calculated ς_k values by Eq. (13) are clearly proportional to $\{\langle \Delta r_k^2 \rangle\}^{-1/2}$ as shown in Fig. 2(a). Note that the abscissa is the reverse of the square root of $\langle \Delta r_k^2 \rangle$. The smaller the atomic mean-square fluctuation is, the larger the contribution to the second-order heterogeneity becomes. From Eq. (8) and Eq. (13), the average value of ς_k over hydrogen atoms, $\overline{\varsigma_k}$, indicates the area of $S^{\text{2-hetero}}(q,0)$. Even though ς_k of a largely fluctuating atom with $\langle \Delta r_k^2 \rangle > \overline{\langle \Delta r^2 \rangle}$ is negative, the average value, $\overline{\varsigma_k}$, can be theoretically proved to be non-negative. ς_k equals to the integral of $\overline{s_k^{\text{2-hetero}}(q,0)}$ from 0 to infinity and $\overline{s_k^{\text{2-hetero}}(q,0)}$ is non-negative because of the relationship between the arithmetic average and the geometric average as shown in Eq. (10).

In Fig. 2(b), the magnitude of ς_k is shown by the colors mapped onto the hydrogen atoms in SNase. The atoms in the inner core region have large ς_k values (red), while the atoms in the outer region have small ς_k values (blue). This indicates that the former atoms have smaller mean-square fluctuations while the latter atoms show larger mean-square fluctuations. Therefore, the atomic fluctuation gradually increases from the inner core region to the outer region. As seen in Fig. 2(b), most of the hydrogen atoms in a protein core positively contribute to $S^{\text{2-hetero}}(q,0)$, especially atoms in the so-called "β-core region" [25].

Figure 2(c) shows the distribution of $\langle \Delta r_k^2 \rangle$ over hydrogen atoms in the protein. Two peaks are observed around 0.4 and 1.2 Å^2. These values are significantly smaller than the average value $\overline{\langle \Delta r^2 \rangle}=2.25$ Å^2. Atoms around these peaks positively contribute to $\overline{\varsigma_k}$. The frequency of the distribution is not a maximum at zero, because all atoms in the protein fluctuate at least $\sim$0.1 Å^2 by the thermal energy at 300 K. The distribution of $\langle \Delta r_k^2 \rangle$ has a long tail toward higher values from the two peaks (up to 35.1 Å^2). This distribution is similar to the distribution of the crystallographic B factor [8]. The distribution of $\langle \Delta r_k^2 \rangle$ can be deduced from the non-Gaussianity of $S(q,0)$ because $S^{\text{2-hetero}}(q,0)$ is the major origin of the non-Gaussianity as shown in Sec. III A. An experimental attempt to obtain a distribution function of $\langle \Delta r_k^2 \rangle$ was performed by Nakagawa *et al.* [8]. They indicated that a bimodal distribution function model was more appropriate than either a Gaussian distribution function model or an exponential distribution function model. The distribution of $\langle \Delta r_k^2 \rangle$ shown in Fig. 2(c) suggests the bimodal distribution model is valid as a first approximation.

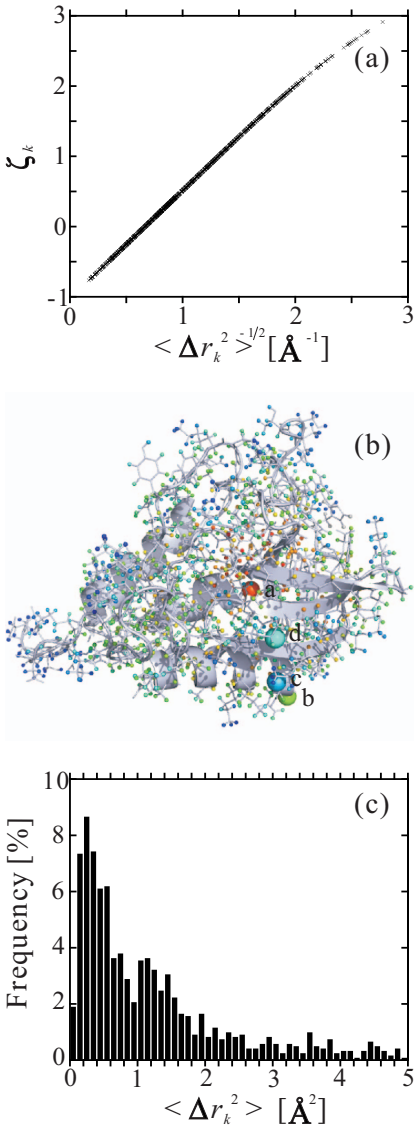

FIG. 2. (Color) Relationship between the second-order heterogeneity and the mean-square fluctuations of the hydrogen atoms. (a) ς_k given by Eq. (13) as a function of $\{\langle \Delta r_k^2 \rangle\}^{-1/2}$. (b) Magnitudes of ς_k are shown by colors mapped onto the hydrogen atoms along with the ribbon model of SNase using PYMOL [30]. The ς_k values decrease in the order of red, orange, yellow, green, light blue, and blue. The atoms selected for Fig. 3 are highlighted with large spheres. (c) Distribution of $\langle \Delta r_k^2 \rangle$.

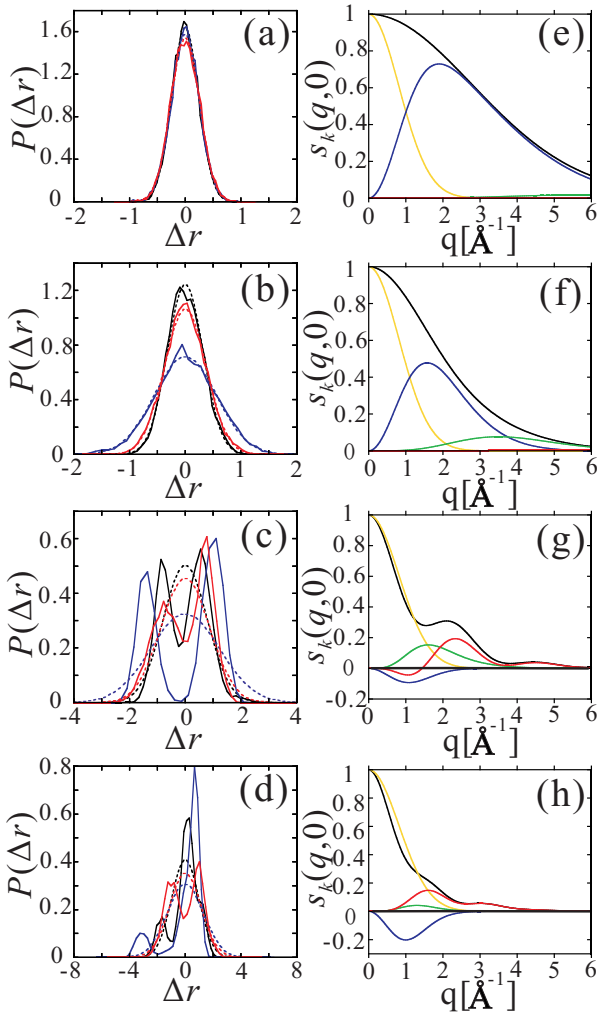

FIG. 3. (Color) Probability distributions of the atomic fluctuations (a–d) and the atomic scattering function with its components (e–h) for typical hydrogen atoms. Selected atoms are as follows: for (a) and (e) the hydrogen atom bound to Ca of Met 98 as an example of a "harmonic" atom; for (b) and (f) the hydrogen atom bound to Ca of Asn 68 as an example of an "anisotropic" atom; for (c) and (g) the hydrogen atom bound to Cg of Glu67 as one example of an "anharmonic" atom; and for (d) and (h) the hydrogen atom bound to Cg of Ile18 as another example of an "anharmonic" atom. From (a) to (d), solid lines indicate the distributions of atomic fluctuations and broken lines indicate the approximated Gaussian function with the variances $\langle \Delta x_k^2 \rangle$, $\langle \Delta y_k^2 \rangle$, and $\langle \Delta z_k^2 \rangle$. Blue, black, and red indicate the x, y, and z direction, respectively. From (e) to (h), $s_k(q,0)$, $S^{gauss}(q,0)$, $s_k^{2\text{-hetero}}(q,0)$, $s_k^{2\text{-aniso}}(q,0)$, and $s_k^{higher}(q,0)$ are depicted with black, yellow, blue, green, and red lines, respectively.

C. Atomic dependence of anisotropy and anharmonicity

Protein dynamics is highly anisotropic and anharmonic. In fact, the present MD simulation on SNase reveal that numerous hydrogen atoms show anharmonic motions. The number of atoms that show anisotropy is also not small. Nevertheless, the contributions from anharmonicity and anisotropy to the total $S(q,0)$ are negligible as shown in Fig. 1. The contribution of each atomic scattering function to the total scattering function will be examined in detail hereafter.

MD simulations reveal that hydrogen atoms can be categorized into the following four types in terms of dynamics: "harmonic," "anisotropic," and two types of "anharmonic" atoms. Figure 3 shows a typical example of the probability of atomic fluctuation and the atomic scattering function for each category.

The distribution of the atomic fluctuation of the "harmonic" atom is nearly an isotropic Gaussian as shown in Fig. 3(a) for the hydrogen atom bound to Cα of Met98 [atom a in Fig. 2(b)]. The atomic scattering function is composed of $S^{\text{gauss}}(q,0)$ and $s_k^{2\text{-hetero}}(q,0)$ [Fig. 3(e)]. The other components are essentially negligible as these atoms contribute only to the second-order heterogeneity. There are 394 "har-

monic" atoms. It should be noted that $\langle \Delta r_k^2 \rangle \ll \overline{\langle \Delta r^2 \rangle}$ in these atoms.

The hydrogen atom of Cα of Asn68 is a good example of the "anisotropic" atoms [atom *b* in Fig. 2(b)]. The distribution of the atomic fluctuation is Gaussian, but shows a distinct anisotropy. When the probability is plotted on a tertiary structure, the shape is a rotational ellipsoid (in this case oblate). The contribution of $s_k^{2\text{-aniso}}(q,0)$ becomes remarkable as shown in Fig. 3(f). The axial ratio and rotation radius of the ellipsoid determine the position and height of the broad maximum of $s_k^{2\text{-aniso}}(q,0)$. The "anisotropic" atom does not substantially contribute to $s_k^{\text{higher}}(q,0)$. $\langle \Delta r_k^2 \rangle$ of the hydrogen atom (0.55 Å^2) is much smaller than the average $\overline{\langle \Delta r^2 \rangle}$ =2.25 Å^2. There are 337 "anisotropic" atoms.

One type of the "anharmonic" atoms is characterized by two clear maxima for the distribution probability [Fig. 3(g)]. This reflects a two-site jump motion, which is the same motion introduced by Doster *et al.* A typical example is the hydrogen atom of Cγ of Glu67 [atom *c* in Fig. 2(b)]. "Anharmonic" atoms are highly mobile ($\langle \Delta r_k^2 \rangle$=2.89 Å^2 for the atom *c*). The shape of the probability on a tertiary structure is a hollow ellipsoid. The atomic scattering function is composed of four components [Fig. 3(g)]. The contribution to $s_k^{2\text{-hetero}}(q,0)$ is small, but is negative in the lower *q* region. The contributions of $s_k^{2\text{-aniso}}(q,0)$ and $s_k^{\text{higher}}(q,0)$ are larger than $s_k^{2\text{-hetero}}(q,0)$. $s_k^{\text{higher}}(q,0)$ oscillates around zero. The peak height of $s_k^{2\text{-aniso}}(q,0)$ is almost the same as that of $s_k^{\text{higher}}(q,0)$, but the position is located on the lower *q* side. In this case, the negative contributions of $s_k^{2\text{-hetero}}(q,0)$ and $s_k^{\text{higher}}(q,0)$ are canceled by the positive contribution of $s_k^{2\text{-aniso}}(q,0)$. As a whole, the atomic scattering function is a Gaussian function on the lower *q* side and $s_k^{\text{higher}}(q,0)$ determines the scattering function of the higher *q* side. The jump distance, that is, the distance between two maxima, and the ratio of the heights determine the peak position and the crossing point of $s_k^{\text{higher}}(q,0)$. When the jump distances are broadly distributed, $s_k^{\text{higher}}(q,0)$ is canceled and $S^{\text{higher}}(q,0)$ becomes negligible. There are 253 "anharmonic" atoms.

The other type of "anharmonic" atoms is characterized as a complex shape of the probability of atomic fluctuation [Fig. 3(h)]. These types of atoms are also highly mobile ($\langle \Delta r_k^2 \rangle$=3.92 Å^2). Figure 3(d) shows a typical example of the probability of the atomic fluctuation for the hydrogen atom bound to Cγ of Ile18 [atom *d* in Fig. 2(b)]. A two-site jump is dominant for the *z* direction. The probability function for the *x* direction is composed of a major peak at the origin and a small subsidiary peak, suggesting two-site jumps rarely occur. The probability function for the *y* direction is similar to that for the *x* direction, but the main peak has a shoulder, suggesting a jump among multiple minima. Figure 3(h) shows the atomic scattering function and its component for the atom. $s_k^{2\text{-hetero}}(q,0)$ is negative between q=0 and q =2 Å^{-1}, but the magnitude is rather small. There are 228 of these types of atoms. Therefore, the contribution to $s_k^{2\text{-hetero}}(q,0)$ is averaged out. Although the motions are quite anisotropic, $s_k^{2\text{-aniso}}(q,0)$ does not contribute to the total atomic scattering. $s_k^{\text{higher}}(q,0)$ shows two maxima, which are reflected in the atomic scattering function. The oscillation period of $s_k^{\text{higher}}(q,0)$ depends on the distance among the minima. Therefore, their contributions to the total scattering function also cancel.

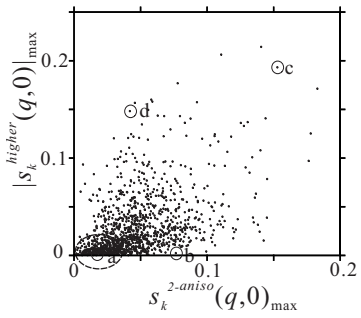

FIG. 4. Relation of the maximum value of $|s_k^{\text{higher}}(q,0)|$ ($|s_k^{\text{higher}}(q,0)|_{\text{max}}$) with the maximum value of $s_k^{2\text{-aniso}}(q,0)$ [$s_k^{2\text{-aniso}}(q,0)_{\text{max}}$]. A circle (drawn with a broken line) indicates the densely distributed region. Atoms shown in Figs. 3(a)–3(d) are marked with ◦.

The $|s_k^{\text{higher}}(q,0)|_{\text{max}}$ (the maximum value of $|s_k^{\text{higher}}(q,0)|$) are plotted against $s_k^{2\text{-aniso}}(q,0)_{\text{max}}$ [the maximum value of $s_k^{2\text{-aniso}}(q,0)$] in Fig. 4, which shows the relation between $s_k^{\text{higher}}(q,0)$ and $s_k^{2\text{-aniso}}(q,0)$. Numerous hydrogen atoms are distributed around the origin. Most of these atoms belong to the "harmonic" atoms located in the inner core region of SNase. In Fig. 4, the location of a typical atom [atom *a* in Fig. 2(b)] is the circle labeled *a*. The "anisotropic" atoms show $|s_k^{\text{higher}}(q,0)|_{\text{max}} \sim 0$ and $s_k^{2\text{-aniso}}(q,0)_{\text{max}} \gg 0$. The circle labeled *b* is the $s_k^{\text{higher}}(q,0)$ and $s_k^{2\text{-aniso}}(q,0)$ values for the typical example [atom *b* in Fig. 2(b)]. The "anharmonic" atoms show $|s_k^{\text{higher}}(q,0)|_{\text{max}} \gg 0$ and $s_k^{2\text{-aniso}}(q,0)_{\text{max}} \gg 0$ (circle *c* for atom *c*), or show $|s_k^{\text{higher}}(q,0)|_{\text{max}} \gg 0$ and $s_k^{2\text{-aniso}}(q,0)_{\text{max}} \sim 0$ (circle *d* for the atom *d*).

As shown above, the $s_k^{2\text{-aniso}}(q,0)$ and $s_k^{\text{higher}}(q,0)$ values for some atoms are relatively large. However, their average over all the atoms, $S^{2\text{-aniso}}(q,0)$ and $S^{\text{higher}}(q,0)$, do not significantly contribute to the total $S(q,0)$. Because the degree of anisotropy and the properties of anharmonic motions vary from atom to atom, thus, both $s_k^{2\text{-aniso}}(q,0)$ and $s_k^{\text{higher}}(q,0)$ do not enhance or weaken each other. To confirm this conclusion, a histogram of the $s_k^{2\text{-aniso}}(q,0)$ and $s_k^{\text{higher}}(q,0)$ values at five *q* values, ranging from 1 Å^{-1} to 5 Å^{-1}, are shown in Fig. 5. The values of $S^{2\text{-aniso}}$ and $S^{\text{higher}}(q,0)$ [average values of $s_k^{2\text{-aniso}}(q,0)$ and $s_k^{\text{higher}}(q,0)$] shown by arrows are small in the entire *q* range as shown in Fig. 1(a). Clear peaks are observed near $s_k(q,0)$=0 in all distributions. These peaks are due mainly to the "harmonic" atoms. The values of $s_k^{2\text{-aniso}}q,0$ are distributed in the range, $0 < s_k^{2\text{-aniso}}(q,0) \leq 0.183$. It should be noted that $s_k^{2\text{-aniso}}(q,0)$ can be theoretically proved to be non-negative [Eq. (11)]. Although the $s_k^{2\text{-aniso}}(q,0)$ values are all positive, their distribution sharply peaks near zero and thus, the average is nearly zero. The $s_k^{\text{higher}}(q,0)$ values are distributed from the negative region to the positive region [$-0.039 \leq s_k^{\text{higher}}(q,0) \leq 0.214$] and cancel on average.

In this paper, we investigated the non-Gaussian behavior of the EINS data using a MD simulation of SNase in water at

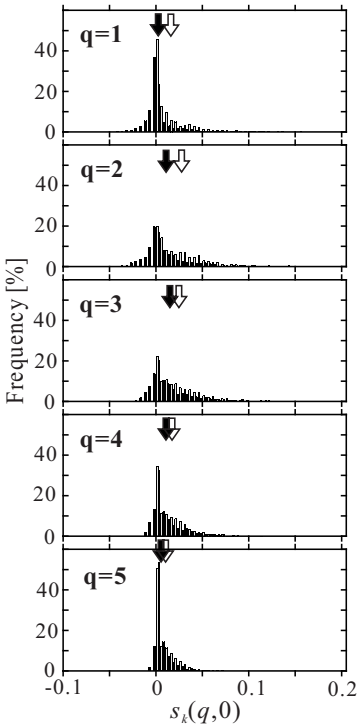

FIG. 5. Histograms of $s_k^{2\text{-aniso}}(q,0)$ (grey bars) and $s_k^{\text{higher}}(q,0)$ (black bars) values at $q=1$, 2, 3, 4, and 5 Å^{-1}. Open arrows and solid arrows indicate the average values of $s_k^{2\text{-aniso}}(q,0)$ and $s_k^{\text{higher}}(q,0)$ at each q value, which equal $S^{2\text{-aniso}}(q,0)$ and $S^{\text{higher}}(q,0)$ at given q.

300 K. The non-Gaussian term can be classified into three components: (i) the second-order heterogeneity, $S^{2\text{-hetero}}(q,0)$, (ii) the second-order anisotropy, $S^{2\text{-aniso}}(q,0)$, and (iii) the higher-order term, $S^{\text{higher}}(q,0)$. We elucidate that the contributions of $S^{2\text{-aniso}}(q,0)$ and $S^{\text{higher}}(q,0)$ to the total EINS are negligibly smaller than that of $S^{2\text{-hetero}}(q,0)$. We also clarified the reason why $S^{2\text{-aniso}}(q,0)$ and $S^{\text{higher}}(q,0)$ do not substantially contribute to the total EINS, even though some atoms behave quite anharmonic and/or anisotropic. We performed the MD simulation analyses for bovine pancreatic trypsin inhibitor (BPTI), hen egg white lysozyme, and T4 lysozyme as well. The analyses of these proteins confirmed that the second-order heterogeneity is the major origin for the non-Gaussianity (data not shown). We conclude that the analysis of the non-Gaussian behavior of the experimental EINS data in terms of the second-order heterogeneity is reliable and reasonable for a globular protein. The total EISF can be approximated as

$$S(q,0) \approx S^{\text{gauss}}(q,0) + S^{2\text{-hetero}}(q,0) = \frac{1}{N}\sum_k \exp\left(-\frac{1}{3}\langle \Delta r_k^2\rangle q^2\right)$$

$$= \int_0^\infty P(x_k)e^{-\chi_k q^2} d(\chi_k), \tag{15}$$

where $\chi_k \equiv \frac{1}{3}\langle \Delta r_k^2\rangle$ and $P(\chi_k)$ is the distribution function of χ_k. This equation indicates that the q dependence of $S(q,0)$ depends mostly on the distribution of $\langle \Delta r_k^2\rangle$. In the other words, the distribution of $\langle \Delta r_k^2\rangle$ can be deduced from the q dependence of the experimental $S(q,0)$.

The analysis of the non-Gaussianity in terms of anharmonicity was performed by Doster *et al.* [6]. They derived a two-site jump model. The present analysis indicated that the two-site jump model is valid for the anharmonic behavior of each atom. However, the contribution of anharmonicity to the total EINS is averaged out by the distribution of the distances between two sites. Therefore, we consider that it is quite difficult to obtain the anharmonicity from the analysis of the non-Gaussianity of the experimental EINS. It has been noted that the anharmonic property of protein dynamics [4,7,27–29] is important for its function. The temperature dependence of protein dynamics [4–7] clearly indicates that the anharmonicity contributes to the dynamical transition. The MD simulation at various temperatures should provide useful information to analyze the anharmonicity with the EINS.

ACKNOWLEDGMENTS

We thank Professor Nobuhiro Go and Dr. Hironari Kamikubo for valuable comments and helpful discussions. This work was supported by the Presidential Special Fund of NAIST to A.T., and a Grant-in-Aid for Scientific Research on Priority Area "Water and Biomolecules" to Y.J., A.K., and M.K. from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

[1] F. Gabel *et al.*, Q. Rev. Biophys. **35**, 327 (2002).
[2] J. C. Smith, Q. Rev. Biophys. **24**, 227 (1991).
[3] A. Rahman, K. S. Singwi, and A. Sjolander, Phys. Rev. **126**, 986 (1962).
[4] B. F. Rasmussen *et al.*, Nature (London) **357**, 423 (1992).
[5] F. Parak, E. W. Knapp, and D. Kucheida, J. Mol. Biol. **161**, 177 (1982).
[6] W. Doster, S. Cusack, and W. Petry, Nature (London) **337**, 754 (1989).
[7] M. Ferrand *et al.*, Proc. Natl. Acad. Sci. U.S.A. **90**, 9668 (1993).
[8] H. Nakagawa *et al.*, J. Phys. Soc. Jpn. **73**, 491 (2004).
[9] A. Kitao, S. Hayward, and N. Go, Proteins **33**, 496 (1998).
[10] N. Go, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. U.S.A. **80**, 3696 (1983).
[11] J. A. McCammon, B. R. Gelin, and M. Karplus, Nature (London) **267**, 585 (1977).
[12] J. A. Hayward and J. C. Smith, Biophys. J. **82**, 1216 (2002).
[13] T. Becker and J. C. Smith, Phys. Rev. E **67**, 021904 (2003).
[14] J. Smith, K. Kuczera, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **87**, 1601 (1990).
[15] J. A. Hayward *et al.*, Chem. Phys. **292**, 389 (2003).

[16] T. Becker *et al.*, Biophys. J. **87**, 1436 (2004).
[17] T. Rog *et al.*, J. Comput. Chem. **24**, 657 (2003).
[18] D. A. Case *et al.*, Computer program package AMBER 7 (University of California, San Francisco, CA, 2002).
[19] U. Essmann *et al.*, J. Chem. Phys. **103**, 8577 (1995).
[20] W. D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).
[21] W. L. Jogensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).
[22] H. Berendsen *et al.*, J. Chem. Phys. **81**, 3684 (1984).
[23] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[24] H. M. Berman *et al.*, Acta Crystallogr., Sect. D: Biol. Crystallogr. **58**, 899 (2002).
[25] T. R. Hynes and R. O. Fox, Proteins **10**, 92 (1991).
[26] J. F. Wang *et al.*, J. Biomol. NMR **10**, 143 (1997).
[27] A. Kitao and N. Go, Curr. Opin. Struct. Biol. **9**, 164 (1999).
[28] H. J. Berendsen and S. Hayward, Curr. Opin. Struct. Biol. **10**, 165 (2000).
[29] F. G. Parak, Curr. Opin. Struct. Biol. **13**, 552 (2003).
[30] W. L. DeLano, Computer program PYMOL (Delano Scientific, San Carlos, CA, 2002).