Differences between calcium rich and depleted alpha-lactalbumin investigated by molecular dynamics simulations and incoherent neutron scattering

Dominik Zeller,[1] Pan Tan,[2] Liang Hong,[2] Daniele Di Bari,[1,3] Victoria Garcia Sakai,[4] and Judith Peters[1,*]

[1]*University Grenoble Alpes, LiPhy, CNRS, F-38000 Grenoble, France and Institut Laue Langevin, F-38042 Grenoble Cedex 9, France*

[2]*School of Physics and Astronomy and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China*

[3]*Physics Department, University of Perugia, 06123 Perugia, Italy*

[4]*ISIS Pulsed Neutron and Muon Source, STFC Rutherford Appleton Laboratory, Chilton, Oxfordshire OX11 0QX, United Kingdom*

(Received 8 November 2019; accepted 21 February 2020; published 25 March 2020)

We present a study comparing atomic motional amplitudes in calcium rich and depleted alpha-lactalbumin. The investigations were performed by elastic incoherent neutron scattering (EINS) and molecular dynamics (MD) simulations. As the variations were expected to be very small, three different hydration levels and timescales (instrumental resolutions) were measured. In addition, we used two models to extract the mean square displacements (MSDs) from the EINS data, one taking into account the motional heterogeneity of the MSD. At a timescale of several nanoseconds, small differences in the amplitudes between the calcium enriched and depleted alpha-lactalbumin are visible, whereas at lower timescales no changes can be concluded within the statistics. The results are compared to MD simulations at 280 and 300 K by extracting the MSDs of the trajectories in two separate ways: first by direct calculation, and second by a virtual neutron experiment using the same models as for the experimental data. We show that the simulated data give qualitatively similar results as the experimental data but quantitatively there are differences. Furthermore, the distribution of the MSDs in the simulations suggests that the inclusion of heterogeneity is reasonable for alpha-lactalbumin, but a bi-or trimodal approach may be sufficient.

DOI: 10.1103/PhysRevE.101.032415

I. INTRODUCTION

A number of proteins have the ability of binding ions that may lead to changes in the protein's structure and dynamics at the atomic scale, and subsequently, may affect their functionality. Recent studies have shown that for example, in the case of enzymes the presence of small inhibitors might influence the dynamics in a measurable way compared to the dynamics of the wild type form [1–5] and single point genetic mutation in proteins can affect collective density fluctuations in hydrating water [6]. Another case is alpha-lactalbumin (α-La), the major whey protein found in the milk of all mammals. It is a simple Ca^{2+} binding milk protein and has a significant role in biosynthesis of lactose in the lactating mammary gland. Together with the enzyme β-1,4-galactosyltransferase (β4GalT) it forms a complex which is responsible for the lactose synthase, i.e., transforming galactose and glucose into lactose. It strongly binds the cation Ca^{2+} and results in changes in the tertiary structure of the protein (see Fig. 1). Besides Ca^{2+}, the binding site can also bind Mg^{2+}, Mn^{2+}, Na^{+}, or K^{+}, which induce similar but smaller structural changes than Ca^{2+}. However, the corresponding binding constants are much lower except in the case of Mn^{2+}. In general, the binding of a cation stabilizes α-La and increases its thermal denaturation temperature. Furthermore, recently Shinozaki and Iwaoka [7] showed that Ca^{2+} and Mn^{2+} accelerates folding to the native form of α-La, an effect not seen with the other cations. α-La can also bind Zn^{2+} at several other distinct binding sites, but results in a decrease in the stability of α-La bound to Ca^{2+}. The apo form refers to α-La which is not bound to Ca^{2+}. Owing to the characteristics described above, α-La is often used as a simple model for Ca^{2+} binding proteins.

In addition to structural changes, the binding of α-La to Ca^{2+} may also generate structural rearrangements capable of influencing locally molecular dynamics and therefore varying the functionality of the protein. The task of probing such small effects is not easy and a sophisticated approach is required. Incoherent neutron scattering is a technique used to probe atomic and molecular dynamics on timescales of pico- to nanoseconds, and when combined with molecular dynamics (MD) simulations forms a powerful partnership, and could indeed offer unique insights into changes occurring in the atomic scale (amplitudes of about a few angstroms). Despite the two techniques accessing very similar times and dimensions, they are not always in full quantitative agreement [8]. To better understand the reasons behind such disagreements, we choose to make a detailed study of the dynamics of α-La, which is commercially available, to do exhaustive neutron experiments, and sufficiently small to permit accurate simulations. We study the dynamics in both the Ca rich and depleted forms, which from hereafter will be referred to as α-La_ca and α-La_dep, respectively.

The intensities measured using incoherent neutron scattering experiments are commonly used to extract mean square displacements (MSDs) of the protons within the protein,

*jpeters@ill.fr

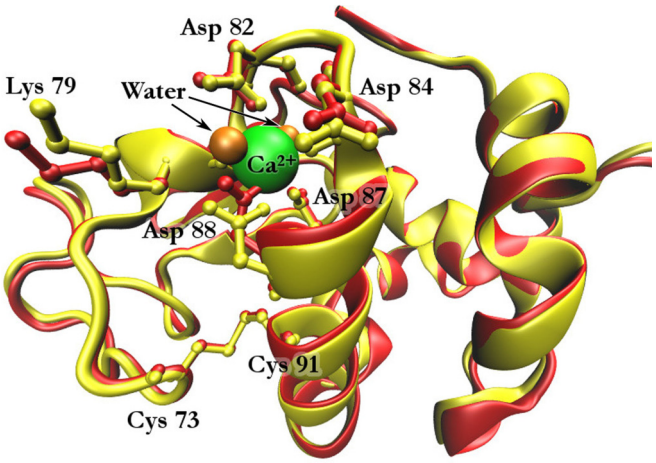

FIG. 1. Calcium rich (yellow, bright) and depleted (red, dark) forms of α-La and the variations induced on their structures. α-La from most mammals consists of 123 amino acid residues [1] and its molecular 4 weight is ≈14.2 kDa.

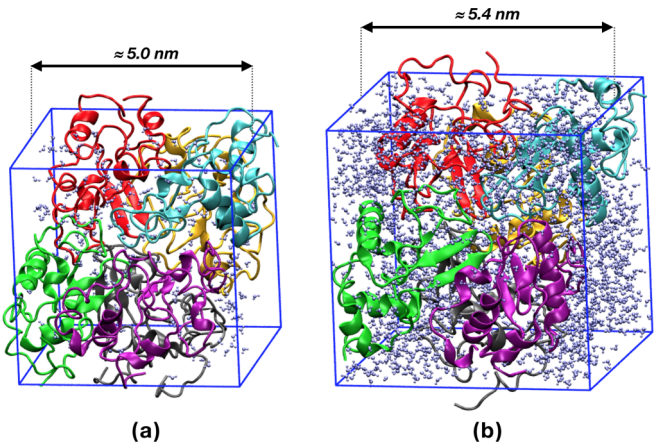

FIG. 2. Visualization of the dry and hydrated simulations of α-La_dep. (a) Dry environment (0.05h). (b) Hydrated environment (0.4h). The lines show the simulation box size and inside the six chains of α-La_dep at 300 K are visualized. The little (blue) molecules indicate water. For the sake of better visibility, we represented the proteins as connected molecules, therefore going beyond the limits of the box, and not as distributed within the same box due to the boundary conditions.

almost exclusively by assuming a harmonic approximation of all possible dynamical contributions, i.e., the Gaussian approximation [9,10]. Furthermore, it is common to combine results from different neutron spectrometers, since they cover different timescales and length scales. Recently, we applied models that go beyond this approximation and include dynamic heterogeneity, to be able to fully exploit a wider instrumental spatial window [11]. In this work, we combine data from three neutron spectrometers which access different timescales (have different energy resolutions) and length scales (have different momentum transfer coverage). We apply a few models, including the commonly used Gaussian approximation, to the data to investigate to what extent they help to disentangle small effects on the dynamics and make a comparison to results from MD simulations.

II. EXPERIMENTAL SECTION

A. Sample preparation for neutron experiments

All experiments described use bovine alpha-lactalbumin (α-La), either in its natural form with Ca^{2+} (α-La_ca) or Ca^{2+} depleted α-La_dep. The protein was purchased from Sigma-Aldrich in lyophilized powder form. Three different hydration levels were prepared for each batch and protein type, hydrated with heavy water, D_2O. This is so the neutron signal is dominated by the incoherent scattering from the protons in the protein (owing to the large incoherent neutron cross section of hydrogen compared to deuterium or other atoms constituting the protein structure [12]). The hydration level was determined from the difference in mass with and without D_2O and is defined as $h =$ grams D_2O/ grams dry protein. The different levels of hydration were $h \approx 0$ (dry), $h \approx 0.4$, and $h \approx 0.8$. The dry lyophilized sample represents the case where only harmonic motions are present up to room temperature and $0.4h$ corresponds to around a hydration level of one or two layers of water on the protein surface [13], which is sufficient to allow for localized dynamical motions. Finally, $0.8h$ represents a gel state close to full hydration. The purchased lyophilized protein powder was dried for at least 24 h, after which it was weighed and then loaded into flat aluminum sample holders (standard for neutron spectroscopy experiments) and vacuum sealed with indium wire. The $0.4h$ and $0.8h$ samples were also dried before hydration, and then left in a D_2O rich environment to uptake the water. The samples were weighed periodically until they achieved the desired uptake of D_2O and then sealed with indium in similar flat aluminum holders. Masses of around 100 mg were used to obtain around 10% scattering and ensure there was no significant multiple scattering.

B. Simulation setup

MD simulations of hydrated protein powder were used representing the interactions of proteins with a small amount of water, to be able to compare them directly with experimental data. The simulations were started from two different protein structures which can be found in the Protein Data Bank (PDB) [14]: (1) bovine α-La with calcium (α-La_ca), PDB ID: 1F6S and (2) bovine α-La without calcium (α-La_dep), PDB ID: 1F6R. Both structures were published by Chrysina *et al.* [15]. Each PDB structure consists of six distinct α-La proteins (= chains), allowing to calculate an average dynamics of a single α-La chain. As a matter of fact, Tarek and Tobias [16] pointed out that a single protein covered by a shell of water is not sufficient to describe a powder protein by simulations. Instead, a crystal composed by two proteins or more resulted in a realistic model to reproduce neutron scattering data. This is the reason why, in the present case, we used six chains of proteins placed in each box (see Fig. 2).

The protein molecule was centered in a cubic box of size 8.39 nm at first, with the CHARMM 27 force field [17,18], and the TIP4P-EW water model [19], using GROMACS 5.0.7 (GPU version) as the MD engine [20,21]. The boxes were filled with water molecules to start with, which were then deleted

(starting from the outside) until the number of water molecules around the protein met the desired hydration level h. The box for a hydration level of $0.4h$ contained 1824 (dep)/1834 (Ca) water molecules and 232 (both) for the dry system ($0.05h$). All systems were electrically neutralized by adding NaCl. Van der Waals interaction was truncated at 1.2 nm with the Lennard-Jones potential switched to zero gradually at 1.0 nm. A particle mesh Ewald [22] with a Coulomb cutoff of 1.2 nm was used to calculate electrostatic interaction. All bonds involving hydrogen atoms were constrained with the LINCS [23] algorithm. The systems were firstly energetically minimized using steepest descent steps with a maximum force of 10.0 kJ mol^{-1} nm^{-1} and a maximum of 50 000 steps. Then they were equilibrated in the *NVT* ensemble at $T = 280$ K (and 300 K) for 300 ps and in the *NPT* ensemble at $p = 1$ bar for 50 ns, with a 0.5-fs time step to slowly release the unreasonable atom contact and suppress vacuum. The temperature coupling was performed using the velocity-rescale algorithm with a coupling time of $\tau = 1$ps [24]. The pressure coupling was performed using the Parrinello-Rahman algorithm with a coupling time of $\tau = 1$ps [25]. The production MD simulations for hydration level $0.4h$ were conducted in the *NPT* ensemble for 100 ns, with a 2-fs time step, while those of the dry systems were conducted for 500 ns. Only the last 20-ns trajectories recorded at every 2 ps were used for the analysis. For such a dense system, the global translation and rotation of the protein molecules was strongly suppressed [26,27]. A visualization of the difference in box size and hydration level is shown in Fig. 2.

III. NEUTRON SCATTERING EXPERIMENTS

Data were collected on three neutron spectrometers, all so-called inverted geometry spectrometers, covering a wide temporal range, namely, OSIRIS [28] at the ISIS Neutron and Muon Facility, UK; IN13 [29] at the Institut Laue Langevin, (ILL), Grenoble, France; and SPHERES [30] at the MLZ Munich reactor in Germany. The data collected at the ILL can be found under the DOIs in Refs. [31] and [32]. OSIRIS and SPHERES use crystal analyzers that reflect cold neutrons (λ of 6.27 and 6.66 Å, respectively) allowing access to a momentum transfer range, Q range, up to 1.8 Å^{-1}, whereas IN13 uses a thermal neutron crystal analyzer (λ of 2.23 Å) which opens up the accessible Q range to 4.9 Å^{-1}. This permits us to probe dynamics occurring in a variety of length scales, where distance $d = 2\pi/Q$. In addition, the three instruments differ in energy resolutions allowing access to motions from a few picoseconds to a few nanoseconds. Specifically, they are 25 8, and 0.7 μeV, for OSIRIS, IN13, and SPHERES, respectively. Transmission values for all samples were measured on IN13 to be above 90% so that multiple scattering effects were not taken into consideration for the data treatment. The initial data reduction was done with LAMP [33] for IN13, SLAW [34] for SPHERES, and MANTID [35] for OSIRIS. Slab can corrections for a flat sample holder and normalizations providing the relative detector efficiency and the instrumental resolution were done with LAMP for the samples measured on IN13 and SPHERES. The measurements on OSIRIS were corrected using the empty sample holder and normalized in MANTID. All intensity normalizations were done with the lowest available temperature data of each scan. Therefore, the difference between the slab correction algorithm and the subtraction of the empty sample holder alone are negligible.

Incoherent neutron scattering measurements give access to the elastic incoherent structure factor (EISF), S, which is a function of the momentum transfer Q at the elastic line, where the energy transfer $\hbar\omega$ that occurs between the neutrons and the scattering atoms (mostly hydrogen), as a result of the scattering event, is approximately zero. The most commonly used approach to analyze this intensity is to assume that the atomic nuclei undergo harmonic motions around their equilibrium positions [9] and thus fit the data to the so-called Gaussian approximation (GA). The intensity can then be expressed as

$$S(Q, 0 \pm \Delta E; \langle r^2 \rangle) \approx S_0 \exp\left(\frac{-Q^2 \langle r^2 \rangle_{\mathrm{GA}}}{3}\right), \tag{1}$$

where ΔE corresponds to the instrumental energy resolution. From this expression, values for the static mean square displacements of the atoms, $\langle r^2 \rangle_{\mathrm{GA}}$, are obtained at each temperature point measured, by fitting the slope of the logarithm of the scattered intensities plotted vs Q^2 according to

$$\langle r^2 \rangle_{\mathrm{GA}} \approx -3 \frac{\partial \ln S(Q, 0 \pm \Delta E; \langle r^2 \rangle)}{\partial Q^2}. \tag{2}$$

The Gaussian approximation is strictly valid for $Q \to 0$, and it holds up to $\langle r^2 \rangle_{\mathrm{GA}} Q^2_{\mathrm{max}} \approx 1$, restricting the Q range that can be used for this type of analysis considerably.

A model that imposes no constraints on the Q range is that developed by Kneller and Hinsen [36] and applied to experimental data by Peters and Kneller [37]. It differs from the GA in that it takes into account motional heterogeneity of the amino acid side chains and their environment, compared to the Gaussian approximation where only one atomic motion is representative for all hydrogens. The motional heterogeneity of the hydrogen atoms is described by a Gamma distribution and the corresponding elastic intensity can be calculated analytically as

$$S(Q; \langle r^2 \rangle, \beta) = \frac{1}{\left(1 + \frac{Q^2 r^2_{\mathrm{PK}}}{3\beta}\right)^{\beta}}, \tag{3}$$

where β is a measure of the homogeneity in the atomic motions; e.g., when $\beta \to \infty$ the Gaussian form is retrieved. Fits of the data give, then, access to the corresponding static mean square displacement, $\langle r^2 \rangle_{\mathrm{PK}}$, where PK stands for the Peters-Kneller model hereafter.

An earlier attempt to account for motional heterogeneity in modeling the EISF was suggested by Meinhold *et al.* [38] by describing the mean square motional amplitudes by a Weibull distribution. However, this approach is not investigated here.

IV. ANALYSIS OF SIMULATED DATA

A. Direct calculation of the MSD

The α-La_ca (α-La_dep) proteins used in the simulations consist of a total of 11 512 (11 457) protein atoms. The number of atoms is different in the two forms, because some chains are missing some amino acids (residues) at the end of the α-La chain since they were not resolved in the PDB structure. In order to compare the simulation data to the experiment, we

analyze the H atoms in the protein, which account for the majority of the scattering signal in the neutron experiments. Furthermore, in order to be consistent in the evaluation of the MSD, only H atoms which are in all chains are considered: Every single α-La consists of at least 922 H atoms which are of the same type for all α-La protein chains. Therefore, with six single α-La chains in each simulation, in total $6 \times 922 = 5532$ H atoms have been evaluated to calculate the MSD and thus to analyze the averaged atomic movements of the protein. The MSD of a single atom α at location $r_\alpha(t)$ at time step t in the simulation is calculated via

$$\mathrm{MSD}_\alpha(t) = \langle [r_\alpha(t_0) - r_\alpha(t_0 + t)]^2 \rangle_{t_0}, \tag{4}$$

with $\langle \cdots \rangle_{t_0}$ being the average over all t_0 defined by the time steps of the simulation. From these individual atoms, a mean μ(t) of the MSD can be calculated.

We first calculated the time average of the MSD according to Eq. (4) using the complete 20-ns trajectories. Further, to estimate the error of the mean of the MSD due to different conformations, the 20-ns simulations were truncated in four equally time spaced parts of 5 ns. The result of the four independent parts was then averaged to obtain a mean MSD $\bar{\mu}_i(t)$ and its sample standard deviation $s(t)$ taken as an estimation of the error:

$$s(t) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N=4} [\mu_i(t) - \bar{\mu}_i(t)]^2}. \tag{5}$$

Finally, we compared the MSDs obtained from a direct calculation with the ones using the fast correlation algorithm proposed by Kneller *et al.* [39]. See Figs. S1 and S2 in the Supplemental Material [40] for the results of the different checks.

In order to compare the dynamic MSD $\bar{\mu}(t)$ of the simulations with the static MSD $\langle r^2 \rangle$ calculated by the models, it has to be divided by 2 since the static MSD $\langle r^2 \rangle$ is defined as a time independent quantity due to the confined motion resulting in [41,42]

$$2\langle r^2 \rangle = \mathrm{MSD}(t \to \infty). \tag{6}$$

For convenience, in the following, the MSDs obtained from the simulations will be labeled as direct MSD $\Delta_{\mathrm{dir}}(t)$ with the following definition:

$$\Delta_{\mathrm{dir}}(t) = \tfrac{1}{2}\bar{\mu}(t). \tag{7}$$

The $\Delta_{\mathrm{dir}}(t)$ can be directly compared to the MSDs extracted from the fits [see Eqs. (2) and (3)].

B. Indirect calculation of the MSD: A virtual neutron experiment

An alternative way to compare the simulation results with the experimental data is to extract the MSD from the convolution of the instrumental resolution $R(\omega)$ with the theoretical dynamic incoherent structure factor (DISF) $S_{inc}(Q, \omega)$ calculated with the help of the simulation data. The DISF was calculated with the program MDANSE [43] (v.1.1). The resolution function $R(\omega)$ for each instrument was approximated by a normalized Gaussian function with a full width at half maximum (FWHM) equivalent to the resolution of the instrument:

$$G(\omega, \sigma_{\mathrm{res}}) = \frac{1}{\sigma_{\mathrm{res}}\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\omega}{\sigma_{\mathrm{res}}}\right)^2\right\}, \tag{8}$$

where

$$\mathrm{FWHM} = \sigma_{\mathrm{res}}\sqrt{8\ln(2)} \approx 2.35\sigma_{\mathrm{res}}. \tag{9}$$

The FWHM of each instrument was obtained by matching the above-defined Gaussian function to data from vanadium which is used to measure, experimentally, the resolution of neutron spectrometers since it is an isotropic incoherent scatterer. For IN13 and OSIRIS data from a vanadium standard summed over all momentum transfers, Q was used; for SPHERES, the resolution function found in the literature for the large angle detectors was used [44] (see Fig. 2) (Voigt profile with $\sigma_{\mathrm{res}} = 0.244\,\mu$eV; $\gamma_{\mathrm{res}} = 0.052\,\mu$eV). The resolution functions are then convoluted with the DISF which is obtained from the simulation. For each DISF calculated with an absolute momentum transfer Q_m, $N_q = 50Q$-vectors Q_i with a randomized direction and an absolute length of $Q_i = Q_m + \Delta Q$, with $\Delta Q \leqslant 0.05\,\text{Å}^{-1}$, are averaged. In total, the DISF is then calculated in MDANSE as

$$S_{inc}(Q_m, \omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} I_{inc}(Q_m, t)\exp(i\omega t)dt, \tag{10}$$

$$I_{inc}(Q_m, t) = \frac{1}{N_\alpha}\sum_\alpha \left\langle \frac{1}{N_q}\sum_i \exp\{iQ_i r_\alpha(t_0)\}\exp\{iQ_i r_\alpha(t_0 + t)\}\right\rangle_{t_0}, \tag{11}$$

where N_α is the number of H atoms in the simulation and r_α their location. t and t_0 are defined by the time steps of the trajectory.

From the resolution broadened DISF, $S^R_{inc}(Q, \omega)$, the elastic incoherent structure factor EISF(Q_m) is computed by summing up the intensities in the range $\omega = \pm\mathrm{FWHM}/2$ and the resulting EISF(Q) is normalized by EISF($Q_m = 0$). The obtained EISF(Q) can be fitted in the same way as the experimental data to calculate the MSD. It is important to mention that for the experimental data the lowest temperature scan was used for the normalization, whereas here the value obtained at $Q_m = 0$ was taken due to the lack of a simulation at very low temperature. The models chosen to analyze the EISF(Q) are the same as for the experimental data, namely, the Gaussian approximation [Eq. (1)] and the PK model [Eq. (3)], over Q ranges of $0 - 1\,\text{Å}^{-1}$ and $0 - 4\,\text{Å}^{-1}$, respectively.

V. NEUTRON SCATTERING RESULTS

The MSDs were extracted as described above for the three instruments and according to the two models. The PK model fits data over a much wider range of Q values, and thus is expected to yield additional information about the motional amplitudes. As, for instance, methyl group rotations are small motions and become particularly visible only at

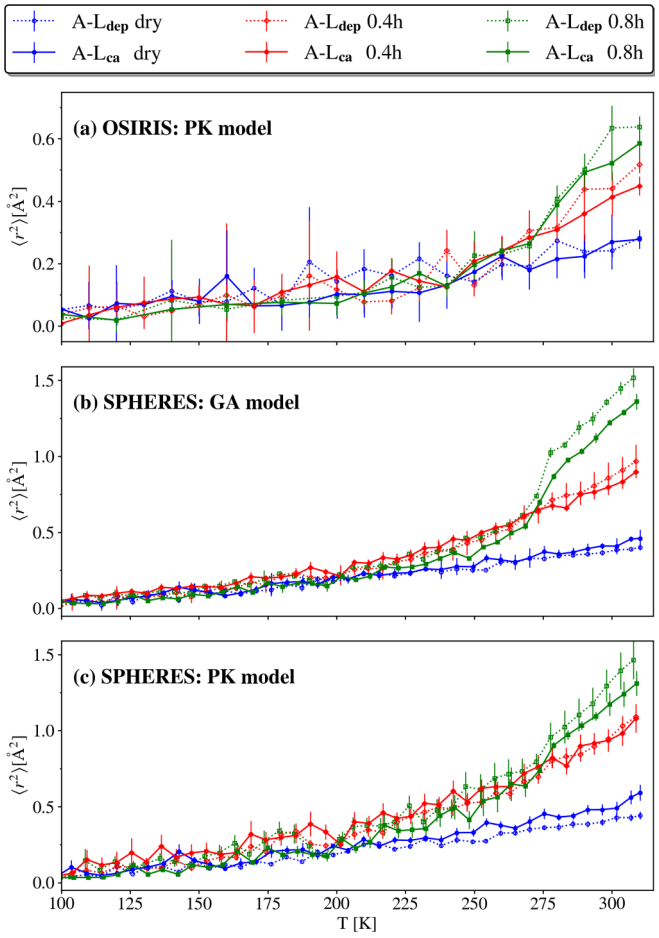

FIG. 3. Difference of MSD between α-La_dep and α-La_ca for OSIRIS (a) and SPHERES (b,c). Data are only shown from 100 to 310 K to enhance the differences, since at temperature <100 K there are no differences within statistical error. The MSD values of the α-La_dep are shown as dotted lines and empty symbols, and those for α-La_ca as solid lines and filled symbols. The dry samples are blue (lower curves), the $0.4h$ samples are red (middle curves), and the $0.8h$ samples are green (upper curves).

higher Q values (>2 Å^{-1}) [45,46], such treatment is expected to give a more precise description. However, it includes one more fit parameter and gives thus higher error bars for the fitting parameters. No significant differences were observed within statistical error between the dynamics (and MSD) of α-La_dep and α-La_ca on the timescales of OSIRIS or IN13 at any of the three hydrations. An example of this is shown for the OSIRIS data fit using the PK model, in Fig. 3 (top). Furthermore, both models show similar trends. Absolutely no differences are appreciable in the dry proteins (blue curves). At $h = 0.4$ (red curves) and $h = 0.8$ (green curves), the dynamics are almost identical except at the higher temperatures ($290 - 310$ K), where small differences are visible and both models suggest that the α-La_ca has a slightly smaller MSD, indicating less dynamics. Given the large error bars, the effect is not conclusive; however, the trend would confirm the findings of Chrysina *et al.* [15], which suggest that the binding of a protein to a cation stabilizes the protein, irrespective of the hydration level.

Differences between the dynamics of the two samples are visible on the timescale of the SPHERES instrument (Figs. 3(b) and 3(c), slower dynamics up to a couple of nanoseconds). Already in the dry state, the α-La_ca sample has a slightly higher MSD than the α-La_dep sample above 250 K. This difference is emphasized when using the model that uses a larger Q range, the PK model, suggesting that also smaller amplitudes (corresponding to higher Q values) have to be included in the analysis to permit such a subtle differentiation.

At $h = 0.4$ no difference between the samples is observed in the PK model, and using the GA model gives a small difference where the α-La_dep is more mobile than the α-La_ca. This could indeed be the case, as for the highly hydrated samples ($h = 0.8$) the same trend is observed, and, more apparent, the MSD is larger for the α-La_dep above 270 K. It is in fact the opposite behavior as for the dry, but more in line with the expected scenario of stabilization of the α-La upon binding calcium. A higher resilience of a protein upon binding of a cation indicates an increased free energy including a higher enthalpy arising from bonded interactions [47]. Entropy is likely rather unchanged at the same hydration level.

It appears that the sample hydrated at $0.4h$ presents higher dynamics between 200 and 270 K than the one hydrated at $0.8h$. Similar effects were already observed for the green fluorescent protein (GFP) [48] and interpreted by the authors as a suppression of protein dynamics at lower temperatures by hydration water and an enhancement of it at higher temperatures. Moreover, in the $0.4h$ sample, the water is in a confined or glassy state so that secondary relaxations set in upon heating, whereas in the $0.8h$ sample where water is primarily bulk water, it is in a frozen state. The steplike increase of the MSD around 270 K for the highest hydrated sample corresponds thus to the melting of the surrounding water.

The GA model shows a more pronounced feature at the melting of ice in the higher hydrated sample compared to the PK model. The GA model covers indeed only larger length scales representing more likely the melting of the ice, whereas in the PK model, which also covers local length scales, an average of the larger and smaller length scales slightly smears out such effects.

VI. MD SIMULATION RESULTS

The MSD obtained directly and indirectly from the simulations were compared. To relate the time dependent direct MSD $\Delta_{\rm dir}(t)$ to the results of the time independent indirect MSD $\langle r^2 \rangle_{\rm ind}$, Heisenberg's uncertainty principle was used:

$$\Delta_{\rm dir}(\tau_{\rm FWHM}) \approx \langle r^2 \rangle_{\rm ind}, \tag{12}$$

with $\tau_{\rm FWHM} = \hbar/{\rm FWHM}$. The corresponding times for each instrument are summarized in Table I.

Figure 4 shows the results of the MSDs of the two different fitting models GA and PK (blue and red, respectively) and the directly calculated MSDs (black), for the two simulated samples (dry and hydrated) at all three instrumental resolutions. The results obtained with the various methods to calculate the MSDs directly were so close (differences below 0.5%)

TABLE I. Instrument resolution FWHM vs time. Relation of the FWHM of Gaussian instrument resolution in energy space to the time τ in time space.

Instrument	FWHM (μeV)	τ_{FWHM} (ps)
OSIRIS	24.8	30
IN13	10.8	60
SPHERES	0.62	1060

that it was not possible to represent them individually in Fig. 4. For the dry protein, the MSD calculated via the direct method is always larger than the MSD calculated from the models, but the behavior between the simulations is the same. Assuming that the direct calculation represents a result as close as possible to the true MSD, the difference between the values of the direct method and those from the models could indicate the order of magnitude of the error introduced by using models. As anticipated, the MSD increases with increasing temperature and in fact the effect is larger on smaller timescales, i.e., at lower instrumental resolutions (IN13 and OSIRIS). The α-La_ca simulations also have a slightly higher MSD for all instrumental resolutions. When comparing the different models, the GA evaluates to a higher MSD than the PK model for IN13 and OSIRIS. For SPHERES this behavior is inverted.

Similar trends are observed for the hydrated protein, except for three main differences. First, the MSD of the models is

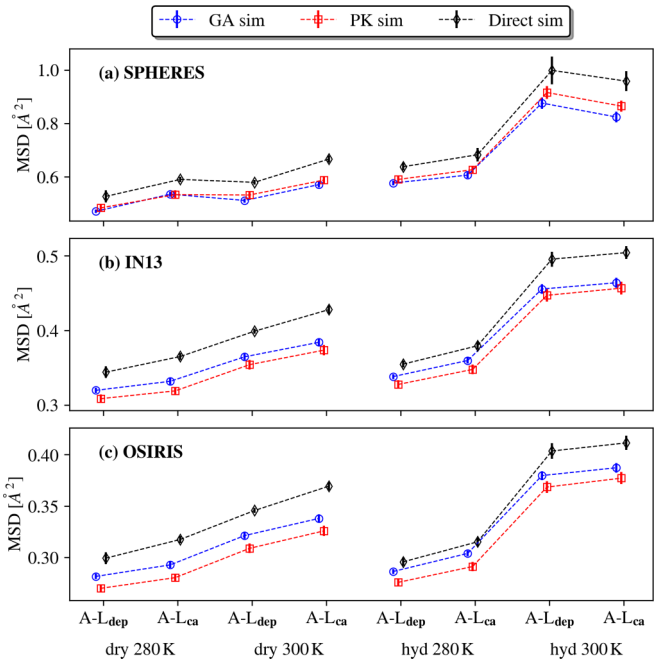

FIG. 4. MSDs from analyzing data from MD simulations using both indirect (GA approximation: lower curves, PK model: middle curves) and direct calculations (upper curves) for the Gaussian resolution function of SPHERES (a), IN13 (b), and OSIRIS (c). On the left side the dry α-La is shown and on the right side the hydrated protein at $0.4h$. As indicated, each side is ordered in the same way by increasing temperature (280 and 300 K) and α-La_dep is next to α-La_ca.

much closer to the direct MSD, albeit still smaller. Secondly, the difference between the MSDs at 280 and 300 K is much larger. Thirdly, for SPHERES the MSD for α-La_ca at 300 K is slightly lower than for α-La_dep, which is the case for all methods considered.

The simulation also allows us to calculate the distribution of the MSDs for the protons in the protein. This is calculated following the method used by Yi *et al.* [49] which enables an evaluation of the main contributions to the heterogeneity and of how many populations with different motions are present. Figure 5 shows the distribution at $t = 30$ ps (OSIRIS), $t = 60$ ps (IN13), and $t = 1$ns (SPHERES) for all simulations. The curve for each time was obtained by binning the individual direct MSD values in steps of 0.02 Å^2 together and normalized by the total number of H atoms. The individual direct MSD values were obtained by averaging the value of the four independent slices of 5 ns for each simulation in the same way as for the direct MSD evaluation. In all simulations and for all three times t, one large peak at around 0.13 Å^2 is visible (dashed vertical line). Only for the distribution at 1 ns (green) a small second peak around 1.35 Å^2 is visible. The latter peak was identified in the simulations to correspond to methyl group rotations, which exist also within the IN13 data [37,45], but are then retrieved at much smaller MSD values below 0.5 Å^2 and cannot be separated from the motions in the main peak. For the hydrated samples the first peak is shifted slightly to higher MSD values and its peak is significantly smaller than for 30 and 60 ps. This effect is emphasized at 300 K. No significant variation is observed between the α-La_dep and α-La_ca samples. This seems understandable as the components forming the two samples are extremely similar.

Such analysis helps us to understand if a complete distribution of Gaussian motions is required to describe the MSD or if a bi- or trimodal approach is sufficient. According to our results, a bimodal description seems to be very reasonable, in agreement with recent works of Vural *et al.* [41] or Doster [45].

VII. COMPARISON OF EXPERIMENTAL AND SIMULATED RESULTS

In order to compare the MD simulations with the experimental data, the results of the fitting models from the previous section are plotted together with the results of the experiments.

The experimental data were collected at 5–10 K intervals, which unfortunately are not always in coincidence with the two simulation temperatures. To reduce the effects on the results, experimental MSD values were averaged over three temperature values (smoothing average) and then the MSDs between two smoothed temperature data points have been linearly interpolated. This procedure ensured that the simulated and experimental data were at the same temperature as the simulations.

As can be seen in Fig. 6, the MSDs extracted from the simulated data agree well with the experimental data and indicate that the different models hardly allow differentiation. Furthermore, the variations between simulated and experimental results may arise mainly from the instrumental limitations.

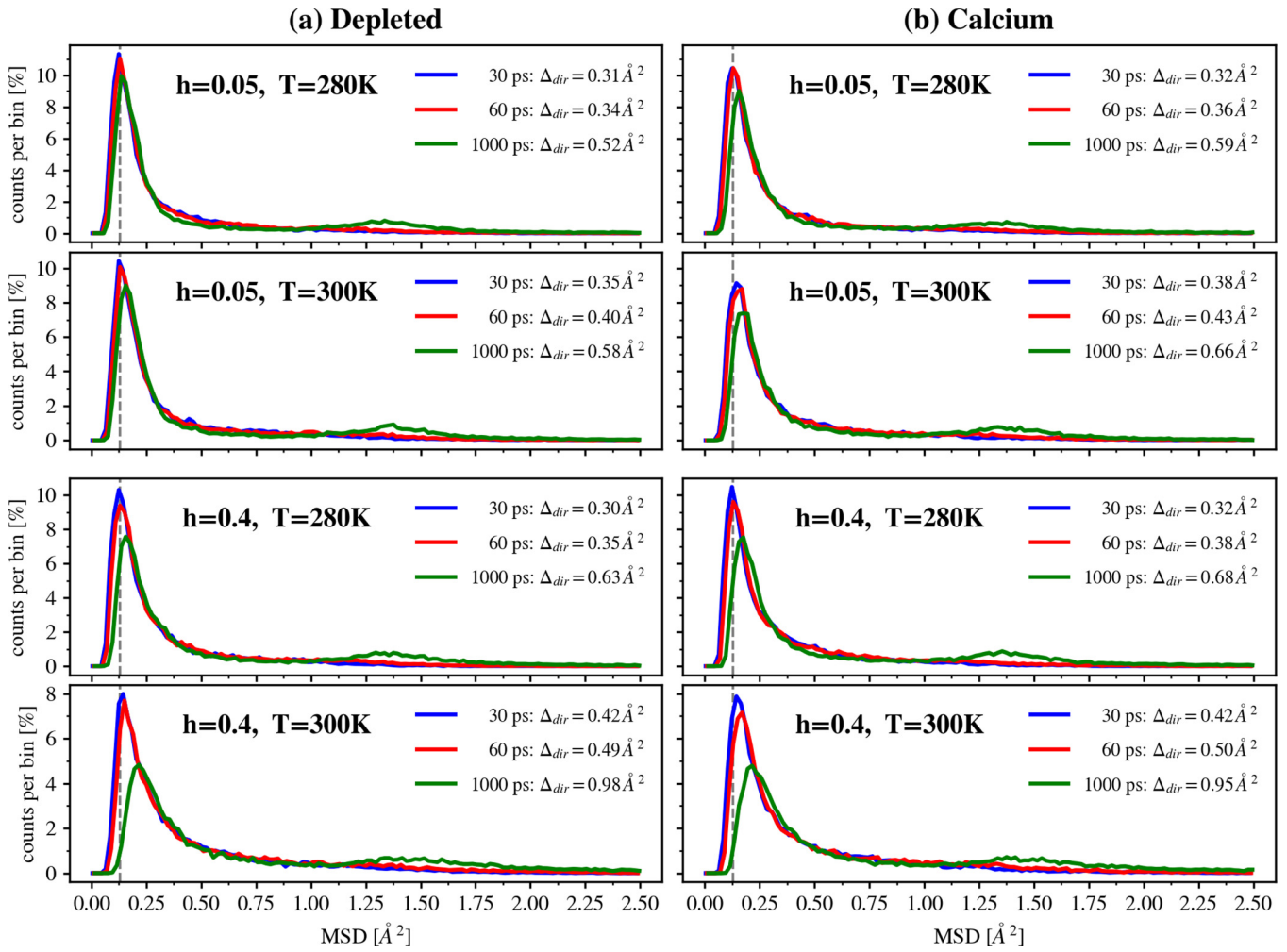

FIG. 5. Distributions of the direct MSDs. Comparison of distributions of the MSDs at $t = 30$ ps (OSIRIS, highest curve in the peak), 60 ps (IN13, middle curve in the peak), and 1000 ps (SPHERES, lowest curve in the peak) for α-La_ca (b) and α-La_dep (a). The upper two figures are the dry samples at 280 and 300 K; the lower two figures are the hydrated samples at 280 and 300 K. The MSD values were obtained by the average value of the four independent slices of 5 ns for each simulation. Δ_{dir} in the legend shows the mean value of the distribution as defined in Eq. (7).

Finally, the differences between the models are larger for the experimental data reflecting the worse statistics.

The experimental MSDs of the depleted hydrated sample seem systematically higher than those of the α-La_ca sample, which is hardly visible within the statistics in the simulation results at 280 K and below 1 ns. It indicates slightly enhanced dynamics for the depleted sample in such conditions, which could be expected as calcium has a stabilizing effect [50]. The higher mobility becomes visible only in the simulations at higher temperatures and longer timescales, as the variations in the sample are certainly small. An interesting point is the difference between the models. For the simulations, the PK model mainly evaluates a very slightly smaller MSD values whereas for the experiments they were larger. One has to note that each spectrometer has not only its specific time resolution, but also a characteristic Q range. Both dimensions are important and are related. Therefore, the evaluated results do not only depend on the time resolution, but also on the accessible spatial domain, which permits us to see various behaviors of the samples.

VIII. DISCUSSION AND CONCLUSION

MD simulations are a very powerful tool to understand, in more detail, the dynamics of individual atoms that are measured for a sample in a neutron scattering experiment, as both techniques give access to comparable temporal and spatial scales. Unlike the common simulations run in solution, comparison to elastic incoherent neutron scattering (EINS) measurements, frequently done with hydrated powders, has required the development of approaches to simulate hydrated powders [49,51] by adapting the setup accordingly.

A direct comparison of neutron data and simulated signals is not always trivial as the absolute values depend significantly on, one hand, data corrections and normalization, and on the other, on the accuracy of force fields and starting structures. It is also common to find that simulations cannot reproduce results extracted from neutron scattering data quantitatively (see [8] or Fig. 6), hence, the decision to compare MSDs by extracting them in a very similar way from both experiment and simulation. In addition, we checked that different

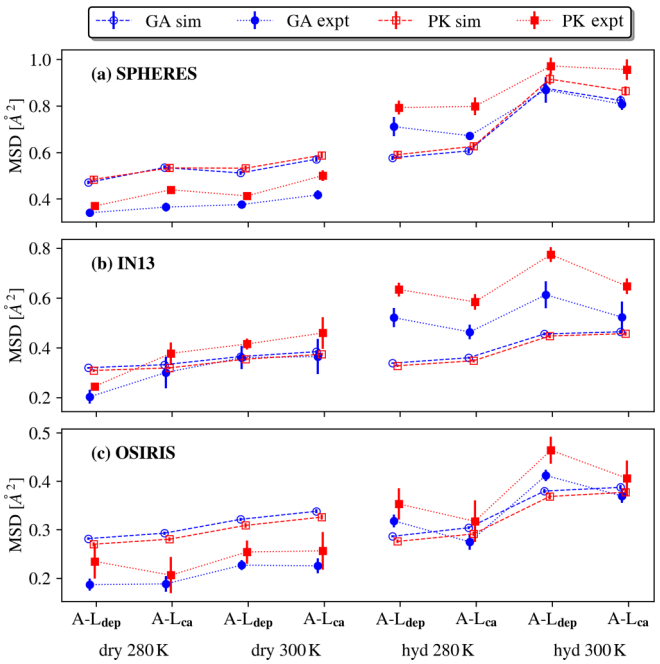

FIG. 6. MSD: Experimental data vs MD simulated data. Comparison between the three models obtained by fitting the experimental data (open symbols, dashed lines) and the MD simulations (filled symbols, dotted lines). The circles designate the GA approximation and the squares the PK model.

approaches to calculate the static MSDs from the simulated trajectories gave identical results.

The order of magnitude of the values of experimental MSDs are well reproduced by the simulations, with MSDs of hydrated samples being larger than those of the dry samples. The results indicate that the models describing the simulated EISF (obtained from the DISF) underestimate the simulated directly calculated MSDs (see Fig. 4). One might therefore speculate that no model is able to take the whole dynamics into account and that effects due to the limited space and time windows are not negligible. For the hydrated protein, the differences are not as large as for the dry protein. In addition, the difference between the models is not negligible, but the trends are always the same and in agreement with the direct MSDs; i.e., all curves obtained through the different models are mainly parallel. Interestingly, the hierarchy between the models is always the same for an identical instrument resolution (with the exception of SPHERES in a dry environment). One can therefore conclude that the GA gives equally good results as the other models, since the absolute values of the MSDs are unknown.

In comparison to the experimental data, the simulation cannot provide reliable quantitative results (see Fig. 6). In general, the experimental MSDs of the PK model are higher than those from the GA model, whereas for the simulations this trend is inverted in most cases. Here, it has to be stressed again, this behavior is highly dependent on the chosen Q range and thus no definitive trend can be concluded. The experimental curves show larger differences and in particular the GA model gives MSDs which are more strikingly different from the MSDs obtained through the PK model. Nevertheless, none of these results favors any one model over another, as the statistics are probably not good enough to discriminate small effects, eventually due to the different Q ranges used.

As shown by Fig. 5 the distribution of the MSDs can be mainly described by two different peaks which are independent of hydration. The second peak is most visible above 1 ns, whereas below 60 ps it is not well distinguished. It is mainly the H atoms of the methyl groups (not shown here) that are contributing to this peak, which is in accordance to the findings of Yi *et al.* [49]. Methyl group rotations indeed contribute to the elastic neutron spectra and the findings here support that they are a major contributor to heterogeneity originating from these motions, which becomes more visible at longer timescales. Yi *et al.* [49] simulated the camphor-bound cytochrome P450 at $h = 0.4$ in a way comparable to the simulations here. They also showed that this peak is more dominant at higher temperatures. Furthermore, the second peak at larger amplitudes is also more pronounced at 1 ns. At 100 ps it is closer to the first peak and much broader. In addition, Tokuhisa *et al.* [52] simulated staphylococcal nuclease (SNase) in a water box at 300 K and also found two distinguishable peaks. The time was not documented but the evaluated simulation time was 1 ns, indicating that the investigated time window was likely smaller than 100 ps.

Overall this leads to the conclusion that the two models give reasonable results in comparison to the direct MSDs from the MD simulations. For a precise data set, the differences between the models are not significant concerning the trends, but the quantitative values are, depending on the evaluated Q range. The PK model gives further insight into the standard deviation of the MSD, but with respect to the MSD it does not give more accurate results. Furthermore, it is also important to state again that in contrast to the experimental data, the simulated EISF was not normalized to the lowest temperature data due to the lack of such simulation data, which could also partly explain the quantitative differences. Doing that, one would more consistently treat experimental and simulated data and eliminate more uncertainties, which might arise.

ACKNOWLEDGMENTS

D.Z. was supported by a Ph.D. scholarship cofunded by the Communauté Université Grenoble Alpes, the STFC Rutherford Appleton Laboratory, and the Institut Laue Langevin. D.D.B. acknowledges support via a scholarship from the da Vinci program of the French-Italian University. J.P. and V.G.S. gratefully acknowledge the support by M. Johnson to partly acquire the Ph.D. grant. The authors gratefully acknowledge the financial support provided by JCNS to perform the neutron scattering measurements at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany. Experiments at the ISIS Neutron and Muon Source were supported by a beam time allocation from the Science and Technology Facilities Council. We thank the ILL for beam time allocation. D.Z. thankfully appreciates the assistance of M. Zamponi during the experiment at the MLZ and the MDANSE software support of E. Pellegrini and R. Perenon.

[1] N. Katsutoshi and S. Shintaro, Eur. J. Biochem. **182**, 111 (1989).

[2] M. Trapp, M. Tehei, M. Trovaslet, F. Nachon, N. Martinez, M. M. Koza, M. Weik, P. Masson, and J. Peters, J. R. Soc. Interface **11**, 20140372 (2014).

[3] J. Peters, N. Martinez, M. Trovaslet, K. Scannapieco, M. M. Koza, P. Masson, and F. Nachon, Phys. Chem. Chem. Phys. **18**, 12992 (2016).

[4] M. Saouessi, J. Peters, and G. R. Kneller, J. Chem. Phys. **150**, 161104 (2019).

[5] M. Saouessi, J. Peters, and G. R. Kneller, J. Chem. Phys. **151**, 125103 (2019).

[6] D. Russo, G. Rea, M. D. Lambreva, M. Haertlein, M. Moulin, A. De Francesco, and G. Campi, J. Phys. Chem. Lett. **7**, 2429 (2016).

[7] R. Shinozaki and M. Iwaoka, Int. J. Mol. Sci. **18**, 1996 (2017).

[8] B. Aoun, E. Pellegrini, M. Trapp, F. Natali, L. Cantù, P. Brocca, Y. Gerelli, B. Demé, M. M. Koza, M. Johnson, and J. Peters, Eur. Phys. J. E **39**, 48 (2016).

[9] A. Rahman, K. S. Singwi, and A. Sjolander, Phys. Rev. **126**, 986 (1962).

[10] G. Squires, *Introduction to the Theory of Thermal Neutron Scattering* (Dover Publications, New York, 1978).

[11] D. Zeller, M. T. F. Telling, M. Zamponi, V. Garcia Sakai, and J. Peters, J. Chem. Phys. **149**, 234908 (2018).

[12] V. F. Sears, Neutron News **3**, 26 (1992).

[13] S. Perticaroli, G. Ehlers, C. B. Stanley, E. Mamontov, H. O'Neill, Q. Zhang, X. Cheng, D. A. Myles, J. Katsaras, and J. D. Nickels, J. Am. Chem. Soc. **139**, 1098 (2017).

[14] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000); also see www.rcsb.org.

[15] E. D. Chrysina, K. Brew, and K. R. Acharya, J. Biol. Chem. **275**, 37021 (2000).

[16] M. Tarek and D. J. Tobias, Biophys. J. **79**, 3244 (2000).

[17] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, and W. E. Reiher *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[18] A. D. Mackerell, Jr., M. Feig, and C. L. Brooks III, J. Comput. Chem. **25**, 1400 (2004).

[19] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, J. Chem. Phys. **120**, 9665 (2004).

[20] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, J. Comput. Chem. **26**, 1701 (2005).

[21] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1-2**, 19 (2015).

[22] U. Essmann, L. Perera, and M. L. Berkowitz, J. Chem. Phys. **103**, 8577 (1995).

[23] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[24] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

[25] S. Melchionna, G. Ciccotti, and B. L. Holian, Mol. Phys. **78**, 533 (1993).

[26] J. H. Roh, V. N. Novikov, R. B. Gregory, J. E. Curtis, Z. Chowdhuri, and A. P. Sokolov, Phys. Rev. Lett. **95**, 038101 (2005).

[27] P. Tan, Y. Liang, Q. Xu, E. Mamontov, J. Li, X. Xing, and L. Hong, Phys. Rev. Lett. **120**, 248101 (2018).

[28] M. T. F. Telling and K. H. Andersen, Phys. Chem. Chem. Phys. **7**, 1255 (2005).

[29] F. Natali, J. Peters, D. Russo, S. Barbieri, C. Chiapponi, A. Cupane, A. Deriu, M. T. Di Bari, E. Farhi, Y. Gerelli, P. Mariani, A. Paciaroni, C. Rivasseau, G. Schirò, and F. Sonvico, Neutron News **19**, 14 (2008).

[30] J. Wuttke, A. Budwig, M. Drochner, H. Kammerling, F. J. Kayser, H. Kleines, V. Ossovyi, L. C. Pardo, M. Prager, D. Richter, G. J. Schneider, H. Schneider, and S. Staringer, Rev. Sci. Instrum. **83**, 075109 (2012).

[31] D. Zeller and J. Peters, http://dx.doi.org/10.5291/ILL-DATA.INTER-337, 2016.

[32] D. Zeller, A. Cisse, L. Misuraca, F. Natali, and J. Peters, https://doi.ill.fr/10.5291/ILL-DATA.CRG-2485, 2018.

[33] D. Richard, M. Ferrand, and G. J. Kearley, J. Neutron Res. **4**, 33 (1996).

[34] http://apps.jcns.fz-juelich.de/man/slaw.html.

[35] O. Arnold *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **764**, 156 (2014).

[36] G. R. Kneller and K. Hinsen, J. Chem. Phys. **131**, 045104 (2009).

[37] J. Peters and G. R. Kneller, J. Chem. Phys. **139**, 165102 (2013).

[38] L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney, and J. C. Smith, Biophys. J. **94**, 4812 (2008).

[39] G. R. Kneller, V. Keiner, M. Kneller, and M. Schiller, Comput. Phys. Commun. **91**, 191 (1995).

[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.101.032415 for the comparison between experimental and simulated MSDs.

[41] D. Vural, J. C. Smith, and H. R. Glyde, Biophys. J. **114**, 2397 (2018).

[42] D. Vural, L. Hong, J. C. Smith, and H. R. Glyde, Phys. Rev. E **91**, 052705 (2015).

[43] G. Goret, B. Aoun, and E. Pellegrini, J. Chem. Inf. Model **57**, 1 (2017); also see www.mdanse.org.

[44] J. Wuttke and M. Zamponi, Rev. Sci. Instrum. **84**, 115108 (2013).

[45] W. Doster, Int. J. Mol. Theor. Phys. **2**, 1 (2018).

[46] W. Doster, H. Nakagawa, and M. S. Appavou, J. Chem. Phys. **139**, 045105 (2013); W. Doster, and M. Settles, Biochim. Biophys. Acta **1749**, 173 (2005).

[47] M. Tehei, D. Madern, C. Pfister, and G. Zaccai, Proc. Natl. Acad. Sci. USA **98**, 14356 (2001).

[48] L. Hong, D. C. Glass, J. D. Nickels, S. Perticaroli, Z. Yi, M. Tyagi, H. O'Neill, Q. Zhang, A. P. Sokolov, and J. C. Smith, Phys. Rev. Lett. **110**, 028104 (2013); J. D. Nickels, H. O'Neill, L. Hong, M. Tyagi, G. Ehlers, K. L. Weiss, Q. Zhang, Z. Yi, E. Mamontov, J. C. Smith, and A. P. Sokolov, Biophys. J. **103**, 1566 (2012).

[49] Z. Yi, Y. Miao, J. Baudry, N. Jain, and J. C. Smith, J. Phys. Chem. B **116**, 5028 (2012).

[50] E. A. Permyakov and L. J. Berliner, FEBS Lett. **473**, 269 (2000).

[51] T. Mounir and D. J. Tobias, J. Am. Chem. Soc. **121**, 9740 (1999).

[52] A. Tokuhisa, Y. Joti, H. Nakagawa, A. Kitao, and M. Kataoka, Phys. Rev. E **75**, 041912 (2007).