Gradual Crossover from Subdiffusion to Normal Diffusion: A Many-Body Effect in Protein Surface Water

Pan Tan,1,2† Yihao Liang,1,2,3 Qin Xu,3 Eugene Mamontov,4 Jinglai Li,2,5 Xiangjun Xing,1,6,* and Liang Hong1,2,*

1School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China
2Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
3State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
4Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
5School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
6Collaborative Innovation Center of Advanced Microstructures, Nanjing 210093, China

(Received 24 November 2017; revised manuscript received 12 March 2018; published 11 June 2018)

Dynamics of hydration water is essential for the function of biomacromolecules. Previous studies have demonstrated that water molecules exhibit subdiffusion on the surface of biomacromolecules; yet the microscopic mechanism remains vague. Here, by performing neutron scattering, molecular dynamics simulations, and analytic modeling on hydrated perdeuterated protein powders, we found water molecules jump randomly between trapping sites on protein surfaces, whose waiting times obey a broad distribution, resulting in subdiffusion. Moreover, the subdiffusive exponent gradually increases with observation time towards normal diffusion due to a many-body volume-exclusion effect.

DOI: 10.1103/PhysRevLett.120.248101

Introduction.—Water is the solvent of life, playing a crucial role in determining the native structure, dynamics, and function of biological macromolecules [1–5]. The diffusive motions of water molecules not only aid the transportation of the function-required essential ingredients, e.g., protons, ions, and substrates across membranes and into the catalytic site of enzymes [1,6,7], but also render internal flexibility to the biomacromolecules. This flexibility is likely to be crucial for the biofunction, as it is absent when the biomacromolecules are dehydrated [1–3,8]. On the other hand, as an active solute, the biomacromolecule will significantly alter the structure and dynamics of the hydration water molecules surrounding it [9–13]. Both experiments and simulations showed that the diffusive motions of the hydration water on the surface of various biomacromolecules: DNA [10], RNA [14], proteins [11], and lipid membranes [12], are significantly retarded as compared to bulk water, and present an anomalous subdiffusion [11,12,15,16]. This subdiffusive motion is characterized by a fractional power-law dependence on time of the mean square atomic displacement (MSD), i.e., \( \langle X^2(\Delta t) \rangle \sim t^\beta \), with \( \beta < 1 \) [11,12,15,16]. Two plausible physical pictures have been proposed to explain subdiffusion of hydration water [17–19]: spatial disorder, i.e., the rough surface of the biomacromolecules forms a fractal, percolated network to hinder the diffusion of water molecules; or temporal disorder, i.e., water molecules jump between traps on the surface of the biomacromolecules with a broad distribution of trapping times. Because of a lack of microscopic evidence, however, the precise physical mechanism remains largely unclear [17,18].

In this work, we studied the diffusive dynamics of surface water on hydrated perdeuterated green fluorescent protein (GFP) and cytochrome P450 (CYP) at physiological conditions by combining neutron scattering experiments with molecular dynamics (MD) simulations and analytic modeling. We showed that the dynamics of hydration water is subdiffusive, and the subdiffusion exponent gradually increases with observation time. By analyzing the MD trajectories of individual water molecules, we directly observed the discrete trapping events of the water molecules on the protein surface and found that the associated trapping times obey a broad distribution, leading to the subdiffusion. Moreover, the deep trapping sites are mostly occupied, and thus water molecules prefer to jump among shallow traps, rendering a gradual increase of subdiffusive power law with the observation time. Finally, a lattice toy model, many-body continuous time random walk, is developed to provide a complete, quantitative description of all the relevant features of diffusive dynamics of the hydration water.

Results and discussions.—Neutron scattering directly probes the fluctuation of the nuclear position and is highly sensitive to hydrogen atoms. Hence, it is a powerful tool for studies of diffusion of water in a variety of environments, from porous silica to carbon nanotube or the surface of biomacromolecules [11,16,20,21]. Here, the neutron scattering experiments were performed at 280 K on both perdeuterated GFP and CYP [see Figs. 1(a) and 1(b)] at a hydration level of 0.4 g water/g protein, using the back-scattering spectrometer (BASIS) at Oak Ridge National Laboratory.
laboratory [22]. This hydration level corresponds roughly to one single layer of water molecules covering the protein surface [1]. The perdeuterated proteins hydrated by H\textsubscript{2}O are the experimental key here to suppress the contribution to the neutron signal from hydrogen atoms in proteins, and the measured neutron data thus are determined by the dynamics of hydration water. For quantitative comparison, all-atom MD simulations were performed at the same hydration level and temperature as the experiments. Details of the experimental and MD protocol are provided in the Supplemental Material [23].

The neutron scattering spectra are presented as the imaginary part of the dynamic susceptibility, \( \chi''(q, \nu) = S(q, \nu)/n_B(\nu) \), where \( S(q, \nu) \) is the dynamic structure factor, presenting the distribution of the dynamic modes in the sample over frequency at a given wave vector, \( q \), and \( n_B(\nu) \) is the Bose factor \( n_B(\nu) = [\exp(h\nu/kT) - 1]^{-1} \). \( \chi'' \) features relaxation processes on different time scales as distinct peaks with associated characteristic relaxation times as \( t_c = 1/(2\pi \nu_{\text{peak}}) \) [15,43,44].

Figures 2(a) and 2(b) present experimental and MD-derived susceptibility spectra \( \chi'' \) for hydration water on CYP and GFP at various \( q \), respectively, which are in good mutual agreement. The Cole-Cole distribution function [15,43] [Eq. (S2) in the Supplemental Material [23]] is applied to model both the experimental and MD-derived \( \chi'' \), providing the value of characteristic relaxation time, \( t_c \), at a given \( q \), i.e., roughly the time for water molecules to diffuse a distance of \( \sim 2\pi/q \). The \( q \) dependence of the resulting \( t_c \) is presented in Fig. 2(c) on a double logarithmic scale for both experimental and simulation systems, which are again in quantitative agreement, validating the simulation systems on the time scales from 10 ps to 1 ns. The wave vectors measured experimentally [Fig. 2(c)] range from 0.7 to 1.7 Å\(^{-1} \), corresponding to a spatial range of 4 Å–1 nm. Different power-law dependences, \( t_c \propto q^{-n} \), are observed for hydration and bulk water. For hydration water on either CYP or GFP, \( n \) is found to be 2.5 ± 0.1, indicating a
subdiffusive motion \(X^2(\Delta t) \sim t^\beta\) with \(\beta = 2/n = 0.8 \pm 0.03\) in the time window probed (10 ps–1 ns). In contrast, a normal diffusion is observed in bulk water, which corresponds to \(n = 2, \beta = 1\) [Fig. 2(c)]. Figure 2(d) presents the MD-derived \(X^2(\Delta t)\) for both hydration water and bulk water in the same time window as probed experimentally, an effective power-law fit provides, \(\beta = 0.8\) and \(\beta = 1\), respectively, confirming the values derived from the neutron susceptibility spectra. The value of \(\beta\) for the protein hydration water obtained in the present work is in quantitative agreement with the values reported in early neutron experiments and MD simulations [11,15,17–19].

Such subdiffusive motions might be quite general for hydration water on proteins, since the two proteins studied here differ considerably in both secondary and tertiary structures [Figs. 1(a) and 1(b)]; whereas GFP consists mostly of \(\beta\) sheets wrapped into a barrel-like structure, CYP consists of comparable amounts of \(\beta\) sheets and \(\alpha\) helices and contain three closely packed domains [45].

Two distinct physical pictures have been previously proposed to heuristically understand the subdiffusion of hydration water: spatial disorder and temporal disorder [17–19]. In the spatial disorder picture, the protein surfaces are thought to be rough and fractal, which slow down the dynamics of hydration water and lead to sub diffusion. Mathematically, \(X^2(\Delta t)\) scales as \(t^\mu\) to the power of \(d_f/d_s\), where \(d_f\) is the fractal dimension of the network of the jump motions, and \(d_s\) is the spectral dimension and is related to the connectivity of the network [17,18]. In the alternative scenario of temporal disorder, water molecules are assumed to jump stochastically between many trapping sites with a broad distribution of trapping times, which is defined as the time a molecule spends at a trapping site before its next jump trial. In the simple toy model called continuous time random walk (CTRW), the distribution of trapping time is assumed to be power-law-like with the divergent expectation of the mean value: \(P(\tau) \sim \tau^{-1/(1+\mu)}\), \(0 < \mu < 1\). The exponent of the subdiffusion is thus fully determined by \(\mu\) as \(X^2(\Delta t) \sim t^\mu\) [17,18,46]. We note, however, these two scenarios are not necessarily mutually exclusive. Hybrid theories involving both mechanisms have been proposed in literature, e.g., to explain the subdiffusive protein internal dynamics [47].

An important difference between these two physical pictures is the existence of correlation between the walking steps for a single particle in the scenario of spatial disorder, which are expected to be strongly correlated over the entire jumping process because of the intrinsic long range correlation in the percolated, fractal network, while no such correlation is expected in the scenario of temporal disorder [46,48].

To explore the microscopic mechanism of the subdiffusive dynamics (Fig. 2), we analyze the MD trajectories of each water molecule on the protein surface. Figure 3(a) is a scatter plot obtained by projecting the MD trajectory of the oxygen atom of one selected water molecule on CYP recorded at every 100 ps continuously over 100 ns, and a video displaying this trajectory is presented in the Supplemental Material [23]. More examples of such scatter plots of individual water molecule on CYP and GFP are presented in Fig. S2 [23]. Upon observation of these trajectories on the ps-to-100-ns timescales, it becomes clear that the water dynamics consists of two modes: rattling within one trapping site (basin) at short time scales and jumping over to neighboring traps at longer time scales. The typical jump distance is about 2–3 Å, even though larger jumps do exist. (More statistical analysis on jump distances is displayed in Table ST1 in the Supplemental Material [23].) This provides strong support for the scenario of temporal disorder. Moreover, the typical size of trapping basins is about 1–2 Å. It is very unlikely that two water molecules occupy the same basin at the same time, as the interwater molecule distance is about 3 Å (see the radial distribution function in Fig. S3), larger than the size of the traps.

For statistical analysis, the MD trajectories of all hydration water molecules were projected similarly as Fig. 3(a), all trapping events were identified, and all jumping displacements \(\vec{s}_i\) were determined. The correlation function between distinct jump displacements for a given water molecule is defined as [46,48]

\[
C(k) = \langle \vec{s}_i \ast \vec{s}_{i+k} \rangle, \tag{1}
\]

where \(\vec{s}_i\) and \(\vec{s}_{i+k}\) are, respectively, the displacements for the \(i\)th and \((i+k)\)th jumps of the molecule, and the bracket indicates an ensemble average over all the hydration water molecules, and the resulting \(C(k)\) are presented in Fig. 3(b) for water on CYP and in Fig. S4B of Ref. [23] for water on GFP. Virtually zero correlation is observed, which clearly
excludes the scenario of spatial disorder. Moreover, as shown in Fig. S3 [23], the radial distribution function of hydration water molecules on protein surface is qualitatively similar to that of bulk water, no sign of fractal structure, further disfavoring the scenario of spatial disorder.

We also analyzed the distribution of waiting times, $\tau_w$, defined as the time a water molecule is observed in MD to be trapped at a site before jumping out. As shown in Figs. 4(a) (CYP) and S5A (GFP in the Supplemental Material [23]), the distribution $P(\tau_w)$ is very broad. Analyses of protein surface structures indicate that residues with charges and local concave geometry tend to trap water molecules for a longer time. Details are discussed in the Supplemental Material [23] (Fig. S6). It is also evident from Figs. 4(a) and S5A that the log-log plot of $P(\tau_w)$ continuously bends downwards as $\tau_w$ increases, and the effective slope, $\mu$, steadily increases from 0.3 to 1.3 with the waiting time [see insets of Figs. 4(a) and S5A [23]]. Concomitantly, MD-derived $\langle X^2(\Delta t) \rangle$ exhibits a crossover from subdiffusion to normal diffusion [Figs. 4(b) and S5B [23]] with the observation time, and the effective power law $\beta$ changes gradually from $\sim 0.75$ to $\sim 0.95$ over the time window studied [Figs. 4(c) and S5C [23]]. A simple CTRW model is clearly incapable of explaining these behaviors. It is of importance to note that there is no significant time window where the exponent stays constant, and the exponent, 0.8, obtained in Fig. 2(d) is an effective fit averaged over the experimental time window from 10 ps to 1 ns.

The model CTRW treats all water molecules on the protein surface independently; i.e., it ignores interaction between neighboring water molecules. At the hydration level of $h = 0.4$ studied here, about 48% of the trapping sites on CYP surface and 46% on GFP surface are covered by water. (This occupancy rate is estimated as the ratio between the number of water molecules and that of trapping sites on the protein surface discovered in MD.) Consequently, there would be a substantial probability that a molecule jumps into an occupied site if it follows CTRW model. This is clearly unphysical, since the trapping sites [Fig. 3(a)] are not large enough to hold two water molecules. On the other hand, a deep trap is most likely occupied for long time by one given molecule, and therefore inaccessible to jumping molecules. Hence, water molecules should preferentially jump to shallow traps, and therefore effectively diffuse faster.

Model and analyses.—To capture these nontrivial consequences of volume-exclusion interaction, we modify and

FIG. 4. Statistical results derived for hydration water on CYP. (a) Black dots denote the MD-derived distribution of waiting time, $P(\tau_w)$. The red curve represents a fit using Model MB-CTRW [Eq. (2)], yielding $\mu = 0.24$, $\tau_0 = 10$ ps, and $\eta = 0.52$. The arrow points out a small population of frozen water molecules, which are stuck in deep traps on the protein surface for the entire 100 ns MD simulation. The inset presents the dependence of the effective power law, $\mu$, of $P(\tau_w) = \tau_w^{-(1+\mu)}$ on the waiting time, which is obtained by power-law fitting in each one-decay-long time window. (b) Mean square atomic displacement (MSD), $\langle X^2(\Delta t) \rangle$, derived from MD (black solid square). Green and red lines represent power-law fits in a time window from 10 to 100 ps and from 10 to 100 ns, respectively. (c) Effective subdiffusive exponent $\beta$ obtained by performing power law fits to MSD derived from MD (black dots) and from Monte Carlo simulation based on MB-CTRW (red dots) in different time windows, with each window being one-decay long. (d) MSD calculated directly from MD (black dots) and from Monte Carlo simulation based on Model MB-CTRW (red dots, see more details in the Supplemental Material [23]). Similar analysis for hydration water on GFP is presented in Fig. S5 in Ref. [23].

248101-4
generalize the CTRW model. We require that each site can hold at most one molecule and assume that a fraction \( \eta \) of sites are occupied. As in the CTRW model, each trapping site is characterized by an intrinsic trapping time (ITT), denoted as \( \tau \), whose physical significance is the average time that a particle stays on this site, if all other sites are empty. The probability distribution of \( \tau \) follows a power law \( f(\tau) = \mu \tau^{\mu - 1} e^{-\frac{\tau}{\tau_w}} \). A water molecule at a site with \( \tau \) tries to jump to a randomly selected site with probability \( 1/\tau \) per unit time, and the jump is successful only if the target site is empty. If the target site is already occupied, the jumping molecule is bounced back to the original site. We call the modified model the many-body continuous time random walk (MB-CTRW) model.

We note that \( \tau \) is an intrinsic property of a trapping site, and is different from the waiting time \( \tau_w \) observed in MD at a given hydration level. The latter depends not only on the potential well of the trapping site, but also on the occupancy rate of neighboring traps. Given the distribution of \( \tau \), the probability distribution of waiting time \( \tau_w \) can be calculated using the mean-field approximation. With technical details relegated to the Supplemental Material [23], the final result is

\[
P(\tau_w) = \mu \tau_w^\mu \left( 1 - \eta \right) \int_{\tau_0}^{\infty} dt \frac{t^{-\mu}}{1 - \eta + C t} e^{-\frac{\tau_w (1-\eta)}{t}}, \quad (2)
\]

where \( C \) is a function of \( \tau_0, \mu, \) and \( \eta \), defined in Eq. (S7) in Ref. [23]. The values of these three parameters are determined by fitting Eq. (2) to the MD-derived distribution of \( \tau_w \). The key result of MB-CTRW is the gradual change of \( P(\tau_w) \), because the many-body volume-exclusion effect will enhance the sampling rates of shallow trapping sites over the deep ones, which will gradually modify \( P(\tau_w) \) to bend continuously downwards. This is the essential feature of \( P(\tau_w) \) found in MD. As shown in Figs. 4(a) and S5A, a remarkable agreement of \( P(\tau_w) \) between MB-CTRW and MD simulation for hydration waters on both proteins is observed in MD at \( \tau_w \). As shown in Figs. 4(c), 4(d) (S5C and S5D), the probability distribution of \( \tau_w \) per unit time, and the jump is successful only if the target site is empty. If the target site is already occupied, the jumping molecule is bounced back to the original site. We call the modified model the many-body continuous time random walk (MB-CTRW) model.

Conclusion.—We have developed a comprehensive and compelling physical picture for the diffusion of hydration water on protein surfaces. We have demonstrated the existence of trapping basins for hydration water and have shown that the subdiffusive motion arises from the broad distribution of trapping times. The deep trapping sites are, however, mostly occupied, and thus water molecules preferentially jump to shallow sites. This many-body volume-exclusion interaction leads to biased sampling of trapping times, and results in a continuous increase of the effective diffusion exponent \( \beta \) with the observation time, i.e., a gradual crossover from subdiffusion to normal diffusion. All these features are accurately captured by our mean field lattice toy model (many-body continuous time random walk) with remarkable precision.

It has been widely demonstrated that dynamics of water is strongly coupled to that of the enclosed protein molecule, e.g., through hydrogen bonds [44,49]. The mobility of water can thus be passed on to the protein through such coupling, influencing or even controlling the dynamical behaviors of functional importance, such as the fluctuation rate of the protein among different enzymatic states and the migration rate of ligands in and out of the catalytic pocket of the protein molecule, etc. [50]. The present work shows that the many-body volume-exclusion effect makes water molecules jump preferentially among shallow sites, and thus effectively diffuse faster. The resulting greater mobility in water can be eventually delivered to the enclosed protein molecule to gain sufficient flexibility required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for the protein molecule to gain sufficient flexibility required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function. This might provide a mechanism to explain why certain hydration (about 20% in weight) is required for its function.

The authors acknowledge NSF China 11674217, 11504231, 31400704, 31770772, 11771289, and 31630002 and Shanghai Municipal Education Commission and Shanghai Education Development Foundation via “Shu Guang” Project for financial support, and the Center for High Performance Computing at Shanghai Jiao Tong University for computing resources. The authors also thank for Miss Keyi Wu for discussion about the CTRW model and Mr. Zhuo Liu for assistance on the neutron scattering experiment at Oak Ridge National Laboratory (ORNL). The neutron scattering experiment on BASIS (SNS, ORNL) was supported by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under Contract No. DE-AC05-00OR22725.

Note to whom all correspondence should be addressed. xxing@sjtu.edu.cn; hongliang@sjtu.edu.cn.

These authors contributed equally to this work.

