JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 2407−2422

pubs.acs.org/jcim

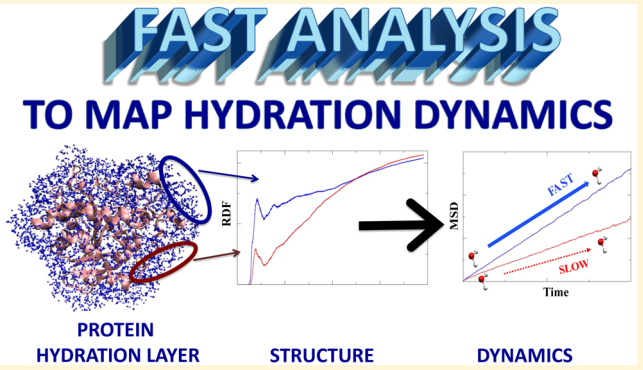
Protein Solvent Shell Structure Provides Rapid Analysis of Hydration Dynamics

Jayangika N. Dahanayake, Elaheh Shahryari, Kirsten M. Roberts, Micah E. Heikes, Chandana Kasireddy,[‡] and Katie R. Mitchell-Koch*

Department of Chemistry, Wichita State University, 1845 Fairmount Street, Wichita, Kansas 67260-0051, United States

Supporting Information

ABSTRACT: The solvation layer surrounding a protein is clearly an intrinsic part of protein structure−dynamics−function, and our understanding of how the hydration dynamics influences protein function is emerging. We have recently reported simulations indicating a correlation between regional hydration dynamics and the structure of the solvation layer around different regions of the enzyme *Candida antarctica* lipase B, wherein the radial distribution function (RDF) was used to calculate the pairwise entropy, providing a link between dynamics (diffusion) and thermodynamics (excess entropy) known as Rosenfeld scaling. Regions with higher RDF values/peaks in the hydration layer (the first peak, within 6 Å of the protein surface) have faster diffusion in the hydration layer. The finding thus hinted at a handle for rapid evaluation of hydration dynamics at different regions on the protein surface in molecular dynamics simulations. Such an approach may move the analysis of hydration dynamics from a specialized venture to routine analysis, enabling an informatics approach to evaluate the role of hydration dynamics in biomolecular function. This paper first confirms that the correlation between regional diffusive dynamics and hydration layer structure (via water center of mass around protein side-chain atom RDF) is observed as a general relationship across a set of proteins. Second, it seeks to devise an approach for rapid analysis of hydration dynamics, determining the minimum amount of information and computational effort required to get a reliable value of hydration dynamics from structural data in MD simulations based on the protein−water RDF. A linear regression model using the integral of the hydration layer in the water−protein RDF was found to provide statistically equivalent apparent diffusion coefficients at the 95% confidence level for a set of 92 regions within five different proteins. In summary, RDF analysis of 10 ns of data after simulation convergence is sufficient to accurately map regions of fast and slow hydration dynamics around a protein surface. Additionally, it is anticipated that a quick look at protein−water RDFs, comparing peak heights, will be useful to provide a qualitative ranking of regions of faster and slower hydration dynamics at the protein surface for rapid analysis when investigating the role of solvent dynamics in protein function.



INTRODUCTION

The protein surface, surrounded by its hydration layer, determines and mediates a diverse array of functions, including ligand binding, protein−protein interactions, and protein structure, stability, and dynamics.[1−6] Spectroscopic measurements of hydration dynamics around proteins indicate that the dynamics of the solvation layer is heterogeneous,[1,7−19] with regional hydration shell properties uniquely contributing to protein function. For example, findings indicate that slower-moving water molecules at the protein surface, also called bound waters (with nanosecond to microsecond time scale dynamics[20,21]), affect molecular recognition processes in protein−protein and protein−ligand interactions, though they are few in number.[22,23] Buried waters, which also move significantly slower, have been shown to lend stability to protein assemblies.[6] Meanwhile, the majority of water molecules in the hydration layer, often called diffusive waters (with picosecond scale dynamics that may be 2−5 times slower than bulk water on average),[24−28] modulate protein structure and dynamics[23,29,30] and are also involved in molecular recognition events. The connection between a protein and its hydration layer is intimate; the two are inextricably linked. The free energy surface at the protein interface determines the structure and dynamics of the hydration layer, which vary by region. Meanwhile, the thermodynamic and dynamic properties of the resulting hydration layer in turn influence the protein structure, dynamics, and function. The scientific community is currently lacking a large and diverse data set analyzing the site-specific dynamics of hydration shells that

Special Issue: Women in Computational Chemistry

Received: January 1, 2019
Published: March 13, 2019

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

would enable a comprehensive understanding of the connections among protein structure−dynamics−function and hydration layer structure−dynamics−thermodynamics. Franck and Han underscore the critical knowledge that the mapping of hydration dynamics around proteins can provide, declaring, "a map of the properties of the hydration water on the surface of a protein would illustrate a distinct fingerprint. This fingerprint may help us decode the function and the binding specificities of the protein, and guide binding or folding processes. The exact meaning of these fingerprints will become clear once the connection between specific types of hydration dynamics and solvation thermodynamics is fully understood".[31]

Up to this point, characterization of regional hydration dynamics of proteins has been a fairly specialized field of analysis, whether by modeling or spectroscopy. Molecular dynamics (MD) simulations are routinely used to complement and interpret spectroscopic measurements of biomolecular solvation, often providing site-specific information for spectroscopic measurements that average over the protein surface.[8,9,16,23,29,32−39] Moreover, simulations have provided insight into the physical principles underlying hydration shell properties and their effects on biomolecules. These efforts, both spectroscopic and theoretical, are labor- and data-intensive efforts.[40] This paper serves to provide a rapid approach for mapping hydration dynamics on protein exteriors. More routine analysis of hydration dynamics will broaden our knowledge of the connection between protein and solvation layer dynamics[41,42] and provide a tool for analysis of the role of mutations on protein structure−function−dynamics.[43] In this paper we ask, "Can an informatics approach be used to provide rapid analysis of hydration dynamics surrounding a protein?"

In order to characterize regional differences in hydration dynamics, our approach involves dividing the external surface of a protein into regions according to secondary structure (i.e., α-helix, β-sheet, loop regions). Further details are provided in Methods. Such an approach typically provides regions of size similar to those locally probed by Overhauser dynamic nuclear polarization spectroscopy[31,44] nuclear Overhauser effects,[10,45] and time-dependent fluorescence measurements,[46−48] describing hydration dynamics in patches or regions of the protein surface on the order of 5−10 Å in each direction.[49,50] This regional characterization of protein hydration dynamics has provided insight into the relationship between protein dynamics and solvation layer dynamics[42] and is poised to enhance understanding of other structure−function relationships in proteins, such as protein−protein binding affinities and molecular recognition of ligands.[7]

While translational water dynamics is typically considered to be "diffusion", several studies indicate that water motion in the first and second shell of a protein is in fact subdiffusive.[51,52] Three-dimensional diffusion in homogeneous liquids is described by both the Einstein relation between the mean square displacement $\langle r^2 \rangle$ and the diffusion coefficient D,

$$\langle r^2 \rangle = \lim_{t \to \infty} 6Dt \tag{1}$$

and the Green−Kubo relation based on the velocity $(v(t))$ autocorrelation function averaged in time and over all particles,

$$D = \frac{1}{3} \int_{t=0}^{\infty} \langle v(0) \cdot v(t) \rangle \, dt \tag{2}$$

Neither the Einstein expression nor the Kubo expression is formally valid for an inhomogeneous system such as water at an interface. Rather, the Smoluchowski equation is appropriate to describe diffusion, and Berne and co-workers have provided an approach to rigorously calculate diffusion at a liquid−vapor interface.[53] However, neither approach is currently tractable for such a complex system as regional dynamics at the protein−water interface. Nevertheless, the translational dynamics of a fluid near an interface can be described via the mean-square displacement (MSD) over time[29,50,54−57] using boundary conditions to evaluate the dynamics within a certain region.[53,58] In this paper, MSD data are reported as *apparent diffusion* using the Einstein relation in three dimensions, providing ease of comparison with values for bulk water and literature reports of apparent diffusion rates within the protein hydration layer.[40,59−61]

The MSD of water near a protein surface is significantly impacted by the available free volume,[20,60,62−64] and studies indicate that the curvature/topology of a protein is a significant source of retardation for water dynamics at an interface.[20,50,65,66] It may be no surprise, then, that the protein−water radial distribution function (RDF or $g(r)$), which provides a measure of water density/packing at the surface, is correlated with the hydration dynamics.[67,65] A lower RDF value for water surrounding a protein side-chain atom indicates less available volume, which thus correlates with lower MSD values. Another source of correlation between the RDF and MSD is tetrahedral structure in the hydration layer and excess (pairwise) translational entropy, S_2,[68−70] which can be calculated from the solvent RDF and is correlated with diffusive dynamics (a dimensionless diffusion coefficient, D^*) via the Rosenfeld relationship,[71] which shows that $\ln D^* \propto S_2$.[65,72] Rosenfeld scaling is an example of a broader set of relationships observed in liquids between excess entropy, a thermodynamic quantity, and transport properties of liquids, including diffusion, viscosity, and thermal conductivity. Such excess entropy scaling was discovered through theoretical work on simple fluid models.[71,72] Since then, it has been observed in diverse systems ranging from various hydrocarbons,[73−76] silicate melts,[77] liquid metals,[78−81] refrigerants,[82] and electrolytes[83] to confined liquids,[84−86] colloidal monolayers,[87] and gases in MOFs[88,89] and nanopores.[90,91] A recent comprehensive review of excess entropy scaling is recommended to readers interested in further detail.[92]

We recently examined the Rosenfeld relationship for the hydration layer of water surrounding the enzyme *Candida antarctica* lipase B (CALB). It was seen that S_2[93−98] calculated from the solvent RDF[99,100] is correlated with D^* of water in the first hydration layer according to the Rosenfeld scaling relationship. For both bulk water[101,102] and water in the CALB hydration layer,[65] the density−diffusion correlation is the reverse of that for a simple liquid (where diffusion increases as density decreases): interfacial regions of lower-density water show higher tetrahedrality and lower entropy along with slower diffusive dynamics (analogous to ice). In summary, the apparent density from the radial distribution function, as indicated by peak height and integration of the hydration layer, has contributions from both differences in the water packing *and* excluded volume from the surrounding protein. Both lower-density structure (typically arising from water structure near hydrophobic interfaces[103−105]) *and* excluded volume (from protein surface curvature) reduce the height of the RDF

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

and contribute to slower translational dynamics in the water layer.

The relationship between structure (RDF) and diffusive dynamics thus hinted at a potential method for more rapid analysis of hydration shell dynamics. While hydration dynamics requires frequent recording of all atomic coordinates in MD simulations (protein and solvent written to disk at least every 0.2 ps), the RDF is a thermodynamic function that can be evaluated from atomic coordinate data saved much less frequently. Furthermore, hydration dynamics must be calculated over finite intervals that are frequently sampled (here, we redefine the water molecules in the hydration layer every 200 ps) using custom scripts. In contrast, the RDF is readily calculated with tools available in common MD packages. Since our report of the correlation between RDF and dynamics was for a single protein, CALB, we first sought to determine whether the hydration layer structure−dynamics correlation is general for a set of proteins. From that, a linear equation describing the relationship between RDF and diffusive dynamics was determined, and its validity with several force fields and at different temperatures was examined. Finally, we assessed the minimum amount of structural information required to provide reliable values of diffusive dynamics in the hydration layer.

METHODS

Simulation Protocol. Five different proteins were considered in this study: CALB, hen egg-white lysozyme (HEWL), subtilisin Carlsberg (SC), ribonuclease A (RNase A), and chemotaxis protein CheY. The starting coordinates for each protein were taken from X-ray crystallographic structures: PDB IDs 1TCA,[106] 4LZT,[107] 1SCA,[108] 2E3W,[109] and 6EKG,[110] for CALB, HEWL, SC, RNase A, and CheY proteins, respectively. These five proteins were selected in order to have variations in protein size and flexibility/rigidity. CALB and SC enzymes are comparatively large proteins, with 317 and 275 amino acid residues, respectively, whereas HEWL, RNase A, and CheY proteins are comparatively smaller. HEWL and RNase A proteins are composed of 129 and 124 amino acid residues, respectively, and CheY protein consists of 123 amino acid residues and a Mg^{2+} ion.

For every protein considered, all of the crystallographic water molecules were kept, as we showed in previous studies that this leads to the fastest equilibration in aqueous simulations.[111] All of the simulations were performed using the GROMACS (version 2016.3)[112] software package. XSEDE computing resources[113] were used for the simulations. As discussed and used in our previous studies,[65] the AMBER03 protein force field[114] with the SPC/E water model[115] was used for the primary data set. This combination of the AMBER03 protein force field and SPC/E water model has been shown to provide very good agreement between water dynamics in simulations and those measured by 2D-IR and magnetic resonance spectroscopy.[116,21] Furthermore, Tarek and Tobias showed that the SPC/E water model gives better agreement with neutron scattering data, representing the short-time motions of protein hydration water with better fidelity than the commonly used TIP3P water model.[54] We also investigated the correlation between RDF and diffusive dynamics with another common all-atom protein force field, OPLS-AA, and the TIP4P water model, which requires more computational effort than the three-site SPC/E model but models bulk water structure and dynamics with high fidelity.[117]

Details of the simulation protocols are provided in the Supporting Information. Results were obtained from sets of 100 ns production runs, and three trajectories were generated for each system using different randomly assigned initial velocities. For statistical sampling, analysis of diffusion was block-averaged with 25 ns time blocks acquired from multiple trajectories, and analysis of the RDF was averaged over three trajectories for the primary data set. Uncertainties are reported at the 95% confidence level using the Student's t test.[118] When calculating data based on the radial distribution function, the standard deviation was calculated, since the RDF was obtained already as an average over 100 ns of simulation time in each trajectory.

Protein−Water Radial Distribution Functions. A radial distribution function provides the probability to find a certain atom or particle at a given distance from another particle. The RDF contains a considerable amount of information. For instance, in a homogeneous solution, the solute−solvent RDF can be integrated to a given distance to provide the number of solvent particles surrounding the solute. The height of the first peak corresponds to how tightly bound the solvent shell is, and the minimum between solvent shells is thus related to the free energy barrier for the exchange of solvent between solvent shells.[119] For nonatomic solvents, the RDFs of individual atoms, for example H or O of water, provide information on the orientation of the solvent around the solute. At a sufficient distance from the central atom, the value of the RDF goes to 1, as the function is normalized relative to an ideal gas (an unstructured system) under the same conditions. Thermodynamic data gleaned from RDFs are typically based on systems with spherical symmetry. Figure S2 shows the RDF for bulk water in the CALB protein simulation box at a distance of 10 Å from the protein. It can be seen that it converges to 1 before 10 Å, indicating that bulk water in our protein simulation has typical bulk water behavior. For the protein hydration shells, the long-distance RDF value does not always converge to 1 (Figures 1 and 3a) because the dimensions of the protein are 3.6 nm × 3.2 nm for the smallest protein (CheY) and 5.2 nm × 4.4 nm for the largest protein (CALB).

For the RDFs considered in these solvated proteins, the system is far from spherical. Here the central atom is a side-chain atom on the protein, and the solvent's (water's) center of mass is used rather than oxygen or hydrogen. This particular RDF was selected because it reproducibly is of similar shape across protein regions and correlates highly with the water dynamics. The side-chain atom−water oxygen RDF was also calculated, and it shows finer structure of the solvent shells. Indeed, comparison of the water center of mass RDF with water oxygen RDF (Figure 1) indicates that the large peak in the center of mass RDF from ∼3−6 Å, which we will call the hydration layer, contains three shells of water. Spectroscopic and theoretical efforts have reached the consensus that the first few shells of water have dynamics most affected by the protein.[39,51,120,121]

In this work, different regions at the protein surface are considered. Therefore, when the protein−water RDF is evaluated for a given region, the resulting function is averaged over all protein side-chain atoms with time-averaged water distributions around each atom. Integration of the function does not provide the total number of waters in the solvation layer at that region, as it would in a typical solute−solvent RDF. Rather, it indicates the average number of waters in the solvation layer around any given protein side-chain atom in

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

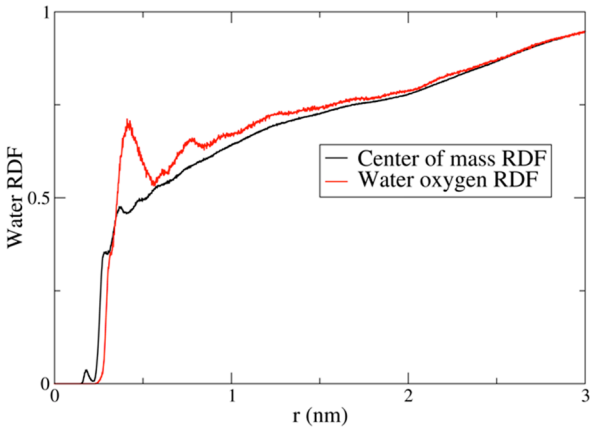


Figure 1. Comparison of radial distribution functions (RDFs) of water center of mass (black line) or water oxygen (red line) around protein side-chain atoms calculated around one region (residues 1−12) of CALB protein.

that region (in units of number probability once integrated by $4\pi r^2 dr$). The number of waters around a protein side-chain atom is affected by the excluded volume occupied by the protein, which precludes the presence of a water molecule at a given site because it is occupied by a protein atom. The excluded volume is a function of the microscopic placement of atoms and the regional curvature of the protein surface. The number of waters around a protein side-chain atom also depends on the packing of water, which is influenced by the identity of the side chains and protein surface curvature.[67,122,123] These latter effects could perhaps could be considered *true* variations in the water density, which are also manifested in the water oxygen−oxygen RDF distances and the tetrahedrality of the hydration layer. Figure S3 shows the density of water in the available volume at different regions around the protein surface, and it can be seen that this truer measure of water density is also correlated with the hydration dynamics. The detailed protocol for obtaining the hydration layer available volume and calculating the density is provided in the Supporting Information.

Analysis of Regional Hydration Layers. In this study, in order to develop an approach for the rapid analysis of hydration dynamics using hydration shell structure, the general relationship between diffusion and the first hydration shell structure around protein *regions* was considered. This approach was carried out using four main protocols: (1) dividing a protein into exterior surface regions and excluding buried/interior surface areas, which are identified according to solvent-accessible surface area; (2) calculating RDFs for these regions and defining the hydration layer; (3) integrating the RDF hydration layer to quantify structural data for correlation with the hydration dynamics around protein regions; and (4) identifying slow-diffusing/buried crystallographic water. Detailed information on these protocols is provided in the Supporting Information.

The division of each protein into separate regions was carried out as the initial step. The secondary structure of each protein was considered on the basis of secondary structure details of the X-ray crystallographic structures. Solvent-accessible surface areas per residue were calculated for the initial crystal structures of each protein using the Chimera software package with a probe radius of 1.4 Å.[124] Individual secondary structures (α-helices and β-strands) and connectors between secondary structures were considered to be separate regions for mapping hydration dynamics to the protein surface. Buried interior regions were excluded, because of the vastly different time scales for buried water dynamics (vide supra), by considering the solvent-accessible surface area details. Detailed information about division of the protein into regions is provided in Protocol 1 in the Supporting Information. An example of the division of a protein surface into regions is illustrated in Figure 2 for RNaseA, and Figure S4 provides the

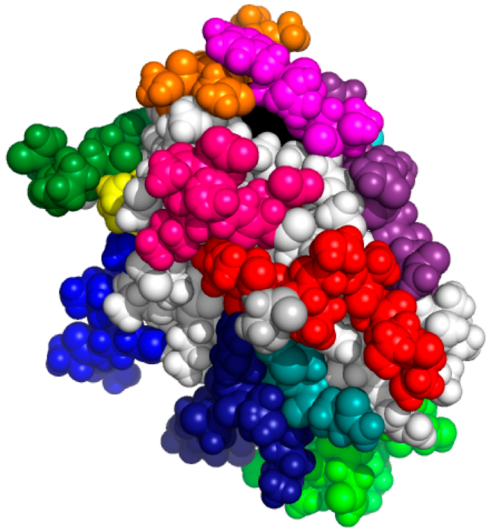


Figure 2. External solvent-exposed regions of RNase A protein divided by secondary structure that were used for analysis of solvent shell structure and regional hydration dynamics: blue, residues 1−5; red, residues 12−22; cyan, residues 27−29; dark blue, residues 31−43; dark pink, residues 49−53; orange, residues 59−63; beige, residues 67−71; magenta, residues 76−78; green, residues 85−96; dark purple, residues 100−105; forest green, residues 113−115; yellow, residues 116−120; aquamarine, residues 122−124.

regions for each of the five proteins studied here, considered in solvent shell structure−dynamics analysis. For the CALB, SC, HEWL, RNase A, and CheY proteins, 23, 26, 17, 13, and 13 protein regions, respectively, were considered, producing a total of 92 data points for analysis of the correlation between diffusion and density of the hydration shell around protein regions.

In order to define the hydration layer, the RDFs of water center of mass around protein side-chain atoms were calculated for each protein region considered, as shown in Figure 3a. It should be noted that AMBER03 is an all-atom force field. It was found that for all of the exterior protein regions, the hydration layer is contained within 6 Å. Detailed information about calculating the RDF and determining the hydration layer cutoff is provided in Protocol 2 in the Supporting Information. During identification of hydration layers, RDFs were visually inspected in order to discover any anomalous hydration shell structures so that such a region could be excluded from the analysis. An example of the RDF of such a region is shown in Figure 3b, which is the RDF of RNase A protein residues 73−75. It contains three peaks within the hydration layer cutoff of 6 Å. Of the 93 regions considered in this work, only this one three-residue sequence had an RDF of irregular shape.

Recently,[65] we observed that the height and area of the hydration layer peaks in the regional RDFs around CALB

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

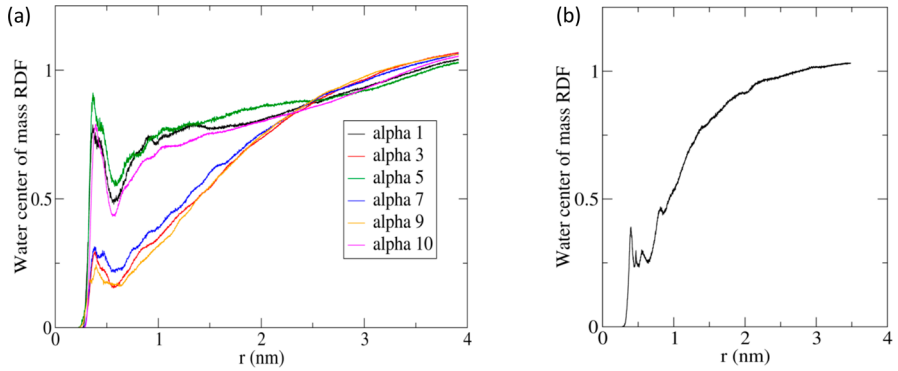


Figure 3. Radial distribution functions of water center of mass around protein side-chain atoms calculated around (a) α-helices in CALB protein and (b) one region of RNase A protein (residues 73−75) with an anomalous hydration shell structure.

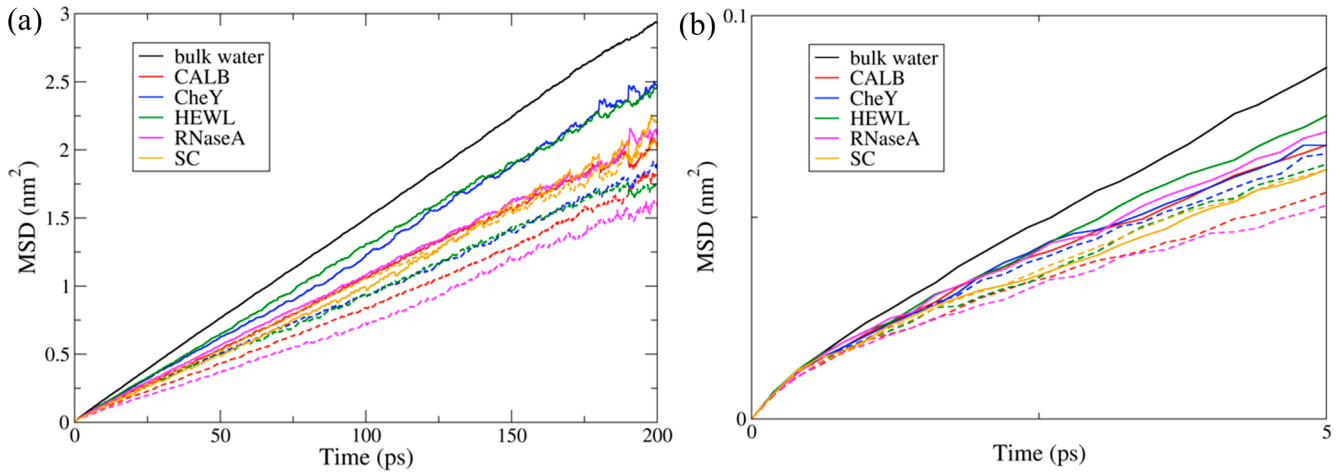


Figure 4. Mean-square displacement graphs for the five proteins (red, CALB; blue, CheY; green, HEWL; magenta, RNase A; orange, SC) and bulk water (black) for (a) 200 ps and (b) the first 5 ps. Two regions (solid line and dotted line) for each protein are shown.

enzyme are correlated with the hydration dynamics. We sought an easily evaluated quantity that could be extracted from the hydration layer RDF in order to devise and evaluate a rapid method to map the regional hydration dynamics. The first method is to numerically integrate $g(r)$ (i.e., the RDF) out to 6 Å, the edge of the hydration layer (henceforth called "RDF hydration layer integration"). Direct integration of the RDF data (without the r^2 Jacobian term) is simple and fast and can be accomplished with a variety of graphical programs or mathematics software. It should be noted, however, that its physical interpretation and units (Å^{-2}) are meaningless, though it is useful for an empirical correlation. Comparison of RDF "linear" integrals works here because all of the RDFs can be described by one shape/function scaled by a proportionality factor characteristic of each region. In the GROMACS software, the hydration layer RDF can be evaluated in one line and subsequently integrated out to r = 0.6 nm in a single line with the following command:

```
g_analyze -f rdf.xvg -integrate -e 0.6
```

Detailed information about RDF hydration layer integration is provided in Protocol 3 in the Supporting Information. The second method considered here to quantify RDF data is to use the maximum peak height of $g(r)$ in the hydration layer, which is located consistently at 0.39 ± 0.01 nm (called "RDF peak height"). Following these two methods, structural data describing hydration layers around protein regions were examined separately using both the RDF hydration layer integration method and peak height, obtained as an average over the 100 ns total simulation time.

GROMACS software was used to calculate relative differences in the apparent diffusion coefficients of water molecules in the hydration layer of each protein region considered. Apparent diffusion coefficients were calculated using MSD ($\langle r^2 \rangle$) data via eq 1. MSDs were calculated for hydration layer water molecules within time blocks evaluated every 200 ps. At time t = 0, water molecules within the hydration layer are identified. Their dynamics are then evaluated by MSD, determining the least-squares fitting of a straight line over the t = 20 to 60 ps time interval, during which $\langle r^2 \rangle$ is observed to be consistently linear and the majority of waters remain within the hydration layer.[42,65] Figure 4a shows MSD graphs for the five proteins, and deviations from linearity can be seen after ∼100 ps. For these hydration layer dynamics, nonlinearity after 100 ps corresponds to contributions from diverging populations, where a fraction of water molecules have escaped the hydration layer (achieving bulklike behavior) while others remain in the hydration layer with more hindered translation. Figure 4b shows the MSD graph for the first 5 ps, where nonlinear behavior is seen at short time scales. In Brownian motion, the MSD at short time scales represents ballistic motion prior to a particle's diffusive motion, in which it undergoes collisions and follows a random walk (with linear MSD as $t \rightarrow \infty$ in a homogeneous environment).[125] Figure S6

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

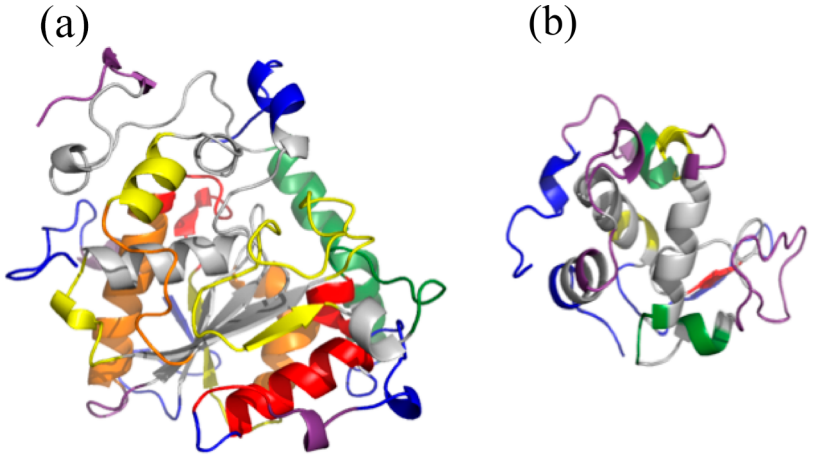


Figure 5. Diffusion maps for (a) CALB protein and (b) HEWL protein. Protein regions are color-coded according to the water diffusion coefficients: purple, >1.90 × 10^{-5} cm^2/s; blue, (1.80−1.89) × 10^{-5} cm^2/s; green, (1.70−1.79) × 10^{-5} cm^2/s; yellow, (1.60−1.69) × 10^{-5} cm^2/s; orange, (1.50−1.59) × 10^{-5} cm^2/s; red, <1.50 × 10^{-5} cm^2/s.

shows linear fitting data for different time intervals, where it can be seen that the 20−60 ps time interval (Figure S6b) has the best linear fit with highest correlation coefficient.

In addition to analysis of the average translational dynamics within different regions of the protein hydration layers, slow-diffusing water molecules were identified for the HEWL, SC, RNase A, and CheY proteins using the protocol in our previously published work that identified slow water molecules for CALB and horse heart cytochrome *c*.[111] Slow-diffusing water molecules have been shown to have functional importance in biomolecular recognition processes by altering the entropy of binding. These tightly bound waters have slower dynamics (up to several orders of magnitude slower) than the majority of water molecules (diffusive waters) within a regional hydration shell. Detailed information about identifying slow-diffusing/buried crystallographic water is provided in Protocol 4 in the Supporting Information. These identified slow-diffusing water molecules were compared with experimentally determined and previously published regions of tightly bound waters.

RESULTS

Heterogeneous Hydration Dynamics. The regional hydration dynamics around the different protein regions were characterized by apparent diffusion coefficients (provided in Tables S1−S5), and heterogeneous hydration dynamics were observed at the exterior regions of all five proteins studied here. The calculated apparent diffusion coefficients for the considered exterior protein regions of these five proteins are all within a similar scale across the five proteins. As can be seen in Tables S1−S5, the apparent diffusion coefficients are in the ranges of (1.44−1.96) × 10^{-5} cm^2/s for CALB protein, (1.38−2.03) × 10^{-5} cm^2/s for HEWL protein, (1.51−1.92) × 10^{-5} cm^2/s for SC protein, (1.40−2.01) × 10^{-5} cm^2/s for RNase A protein, and (1.76−1.95) × 10^{-5} cm^2/s for CheY protein, indicating that the studied protein interfacial regions show similar ranges of diffusive heterogeneity, regardless of differences in protein size and protein flexibility/rigidity. Figure 5 shows a map of the surface hydration dynamics overlaid onto the structures of CALB and HEWL. The dynamics are color-coded from fastest (purple) to slowest (red). Maps for the other three proteins are provided in Figure S5. In our previous studies using CALB, it was observed that regions with fast dynamics by one measure of hydration dynamics (residence times) generally exhibit fast dynamics by other measures (i.e., diffusion, reorientational dynamics),[65] consistent with literature reports (see the Discussion for further details). Therefore, the above-mentioned heterogeneity of diffusive dynamics at these protein interfaces should be an indicator of the heterogeneity of other hydration dynamics parameters.

RDF Peak Height and Integration: Correlations with Diffusive Dynamics. Correlations between apparent diffusion coefficients at each region and structural data given by both RDF peak height and hydration layer integration were examined. In our previous work on CALB enzyme,[65] integration of the regional water oxygen−protein atom RDF at different regions showed a linear relationship with the apparent diffusion coefficients of water molecules in the vicinity, with a correlation coefficient of 0.85 (Figure S7a).[65] In the present work, it was found that when the RDF of the water center of mass around CALB side-chain atoms was used, there was a higher correlation (R^2 = 0.91) between diffusion coefficients of regional solvent shell water molecules and integration of the hydration layer peak of the RDF (Figure S7). For this reason, the water center of mass around side-chain atoms RDF is used in this present work for rapid analysis of the local hydration dynamics. It was investigated whether the RDF peak height and hydration layer integration methods can provide statistically equivalent measures of solvent shell structure. For each protein, the correlation between local solvent shell water diffusion and water structure was analyzed considering both RDF peak heights and hydration layer integrals (Figures S8−S12). It was observed that for all five proteins studied, the method of using RDF hydration layer integration (out to 6 Å, the edge of the solvation layer) provides better correlations with regional solvent shell water diffusion rates (R^2 = 0.91 for CALB, 0.82 for HEWL, 0.80 for SC, 0.90 for RNase A, and 0.81 for CheY protein) compared with the RDF peak height method (R^2 = 0.74 for CALB, 0.64 for HEWL, 0.58 for SC, 0.58 for RNase A, and 0.66 for CheY protein). Thus, it was concluded that RDF hydration layer integration is a more reliable approach for the rapid analysis of local hydration dynamics, and this is the focus of analysis from this point on.

Solvent Shell Structure as an Approach for the Rapid Analysis of Hydration Dynamics. As discussed above, a

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

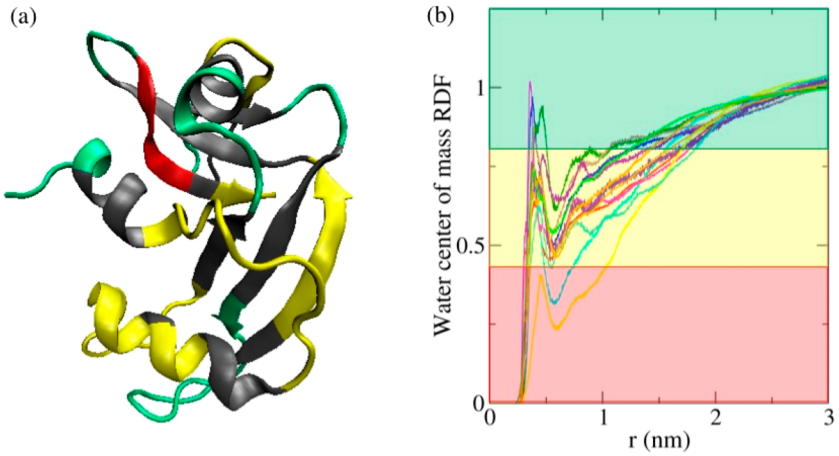


Figure 6. Color-coded maps for RNase A protein hydration layer dynamics and structure. (a) Protein hydration shell water diffusion coefficients mapped to the RNase A structure. Water diffusion coefficient values: green, >1.9 × 10^{-5} cm^2/s; yellow, (1.4−1.9) × 10^{-5} cm^2/s; red, <1.4 × 10^{-5} cm^2/s. (b) Regional RDFs: water center of mass around protein side-chain atoms for RNase A protein regions. Shaded regions indicate hydration dynamics according to the regional protein hydration shell water apparent diffusion coefficients (green, >1.9 × 10^{-5} cm^2/s; yellow, (1.4−1.9) × 10^{-5} cm^2/s; red, <1.4 × 10−5 cm^2/s). The color code for the RDF traces corresponding to different regions is the same as the color code used in Figure 2.

good correlation was observed between the local solvent shell structure, characterized by the integral of the regional hydration layer RDF, and the local hydration dynamics, characterized by regional solvation layer water diffusion coefficients. Figure 6 provides a color-coded map illustrating the structure−dynamics correlation in the regional hydration layers of RNase A protein. Figure 6a maps the regional diffusion coefficients to the protein structure. Figure 6b shows the regional water−protein RDFs, color-coded by region and shaded for fast (green), intermediate (yellow), and slow (red) hydration dynamics ranges. Figure 2 provides the color-coded regions that correspond to the colored RDF traces. It can be seen that RNase A protein regions with comparatively higher RDF peaks exhibit faster water dynamics (color-coded green), whereas RNase A protein regions with comparatively lower RDF peaks for water exhibit slower water dynamics (color-coded red).

The next step was to determine whether the linear correlation between regional RDFs and diffusive dynamics can be observed as a general relationship across the set of all protein hydration layer regions studied here (five proteins). Figure 7 shows that a single linear equation between RDF hydration layer integration and regional water diffusive dynamics was observed across all of the proteins studied in the set with a correlation coefficient of $R^2 = 0.84$. As described in Methods, 92 data points were considered, and it can be seen that data points corresponding to all five points fall on or around a single line. Furthermore, as described with the calculated diffusion ranges, the data in Figure 7 also reveal that all five proteins considered in this study show heterogeneous hydration dynamics at exterior regions. This correlation between local solvent shell structure, characterized by RDF hydration layer integral values, and local hydration dynamics, characterized by regional solvent shell water diffusion coefficients, produces a model equation for rapid analysis of regional hydration dynamics given by apparent diffusion coefficients: $D = (3.82x + 1.22) \times 10^{-5}$ cm^2/s, where x is the integral of the water−protein atom RDF out to 6 Å. Figure S13 shows the same data as in Figure 7 but also includes the structure and dynamics data for SPC/E bulk water, showing

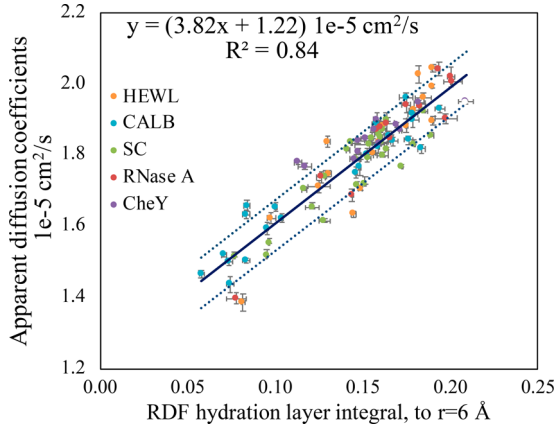


Figure 7. Linear correlation of local water structure/density, obtained using integration of the RDF, with the regional hydration layer diffusive dynamics across a set of five proteins, with error bars determined at the 95% confidence level according to Student's t test values for the diffusion coefficients.

that bulk SPC/E water's structure−dynamics correlation falls fairly close to the line established for regional hydration layers (also modeled with SPC/E water).

Figure 7 shows the plot that represents the general relationship between RDF hydration layer integrals and regional hydration layer diffusion coefficients across the set of proteins, with error bars. As mentioned in Methods, the uncertainties for the diffusion coefficients are reported at the 95% confidence level using Student's t test. The uncertainty in the RDF hydration layer integral is reported as the standard deviation (from three 100 ns trajectories).

The 92 regions examined across these five proteins consist of α-helices (26%), β-sheets (15%), and loop/connector regions (59%). We examined whether the model equation held better with one type of secondary structure than another. All three types of regions were found to have rather high correlations between RDF hydration layer integrals and apparent diffusion coefficients, ranging from $R^2 = 0.77$ for α-helices to $R^2 = 0.89$ for loop/connector regions (Figure S14), and gave slightly

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

different linear fits. The correlation coefficients trend with the size of the data set, and the small discrepancies between the linear fits for each secondary structure type and the model equation given by the aggregate data suggest that the RDF−diffusion relationship is general for all types of protein regions on the exterior surface.

Evaluation of the Validity of the Linear Regression Model Using Statistical Analysis. According to the results mentioned above, a general linear relationship between regional RDF hydration layer integrals and regional hydration dynamics was observed across the set of considered proteins, with a correlation coefficient of R^2 = 0.84, revealing a good linear correlation between these two variables. Furthermore, it can be considered as a model for the rapid analysis of diffusion coefficients of water molecules at regional solvent shells around exterior protein regions using hydration layer peak integration of the regional RDF. Therefore, it is important to evaluate the validity of this linear regression model using statistical analysis, as described below.

The margin of error at the 95% confidence level for the apparent diffusion coefficients was calculated according to Student's t test.[126] Figure 7 shows the plot of the general relationship between regional hydration layer RDF peak integration values and solvent shell water diffusion coefficients across the set of proteins, with dotted lines indicating the 95% confidence interval (margin of error = $\pm 0.072 \times 10^{-5}$ cm^2/s) according to t test values for the diffusion coefficients. It can be seen that the majority of the calculated data points for diffusion coefficients are within the 95% confidence interval, indicating the validity of the structure−dynamics linear regression model.

Furthermore, since the sample size (92 data points) is relatively large within this study, the z test was used to test the hypothesis that the diffusion coefficients produced by the developed model (Table S6) are not significantly different from calculated diffusion coefficients from the MD simulations. For the 95% confidence level, the critical value for a two-tailed test is z_c = 1.96. For the two means of calculating diffusion coefficients, the results for diffusion coefficients calculated directly from simulation ($\overline{X}_1$) and apparent diffusion coefficients produced using the developed model based on RDF integration ($\overline{X}_2$), the z statistic was computed as

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (3)$$

where σ is the standard deviation and n is the sample size. Since it is observed that the calculated $|z|$ value of 0.212 is less than z_c, it is concluded that the considered hypothesis is not rejected, indicating that the diffusion coefficients produced by the developed model are not significantly different at the 95% confidence level from the diffusion coefficients calculated directly from MD trajectories (which requires significantly more effort).

As the next approach to evaluate the validity of the linear regression model, residual errors were examined, as shown in Figure 8. It can be seen that the residual errors are randomly distributed about an average residual error of zero, with no apparent trend toward increasing or decreasing residual errors, indicating that data can be modeled with a straight-line relationship.

Model Validation with Different Water Models and Different Force Fields. In order to determine the effect of the water model on the correlation between regional translational dynamics and RDF hydration layer integrals, results were obtained using the TIP4P water model for the CALB and HEWL proteins (Figure 9). It can be seen that with the TIP4P water model, the linear correlation between the RDF hydration layer integral and water diffusion coefficient is valid (R^2 = 0.87). Furthermore, the correlation has a slope similar to that obtained with the SPC/E water model (3.89 for TIP4P vs 3.82 for SPC/E) but a different intercept because of slight differences in the models' bulk diffusion coefficients.

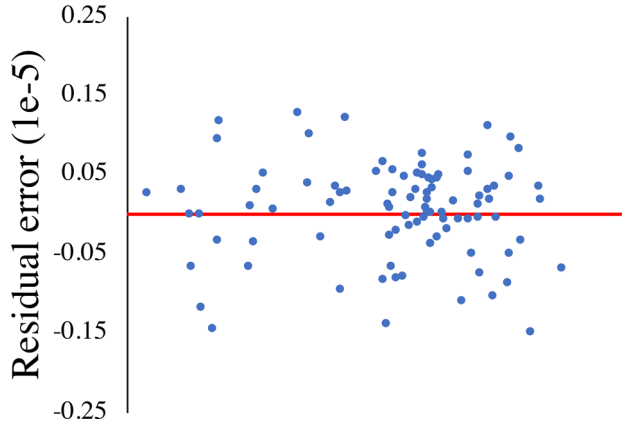


Figure 8. Plot of the residual error in the diffusion coefficients as a function of the value of the RDF hydration layer integral (to r = 6 Å) using the linear regression model. The red line shows a residual error of zero.

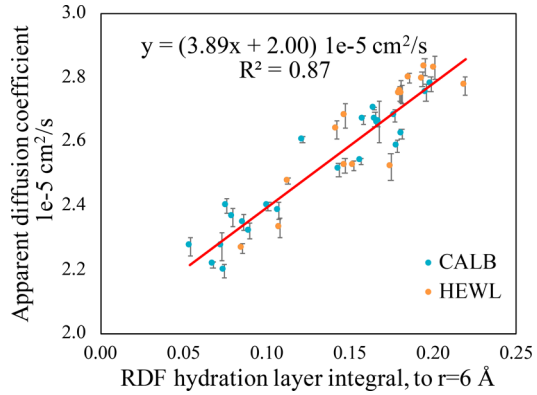


Figure 9. Linear correlation of the local water structure/density obtained using integration of the RDF with the local hydration layer apparent diffusion coefficients for CALB and HEWL protein with the TIP4P water model.

In order to determine the validity across protein force fields for the correlation between regional RDF hydration layer integrals and water diffusion coefficients, results were obtained using OPLSAA force field to model the CALB and HEWL enzymes. OPLSAA is another all-atom protein force field, and it can be seen in Figure 10 that the hydration layer structure−dynamics correlation is valid (R^2 = 0.88). Furthermore, the model produced with the OPLSAA force field [$y = (3.61x + 1.22) \times 10^{-5}$ cm^2/s] is similar to the model obtained with the AMBER03 force field [$y = (3.82x + 1.22) \times 10^{-5}$ cm^2/s], suggesting that the developed model is valid for other all-atom protein force fields under similar simulation conditions.

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

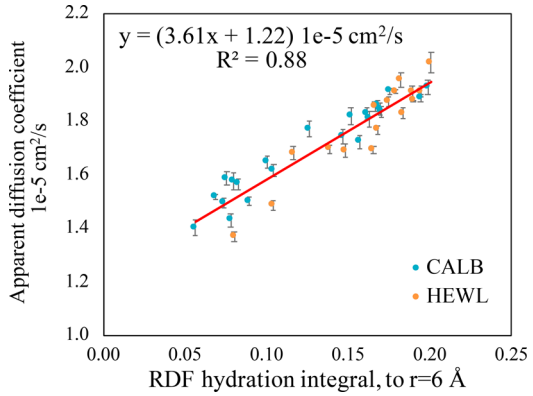


Figure 10. Linear correlation of local water structure obtained using integration of the RDF with local solvent shell water diffusion coefficients for the CALB and HEWL protein with OPLSAA force field.

Minimum Time Required for Accurate Determinations of Density Using RDFs. As mentioned earlier, one purpose of this study is to develop an approach for rapid analysis of hydration dynamics using minimum computational effort. Therefore, we examined whether 10 ns of RDF data is sufficient to predict hydration dynamics. Of course, it is important to evaluate data from converged simulations. The time required for each protein to obtain convergence based on RMSD values was evaluated, as shown in Table S7 and Figure S15. The minimum time to convergence was 2 ns for SC, and the maximum was 8 ns for CALB. Next, the RDF data were averaged over the subsequent 10 ns of data after convergence. Figure 11 provides the correlation between the diffusive dynamics and the RDF hydration layer integral averaged over a

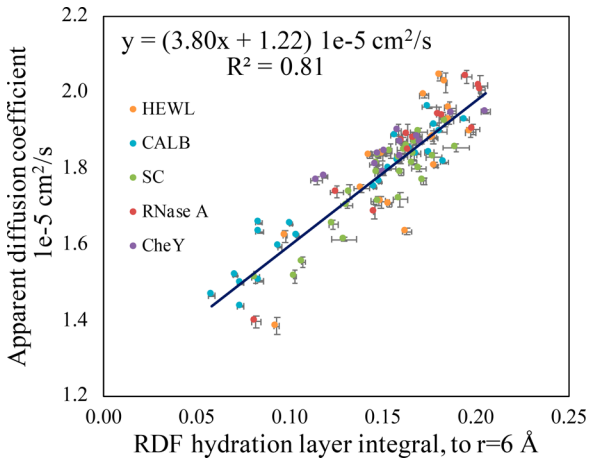


Figure 11. Linear correlation of local water structure/density obtained using integration of the RDF of the first 10 ns after reaching equilibrium with local solvent shell water diffusion coefficients.

10 ns trajectory after convergence. With these RDF data, the linear relationship was observed to have a correlation coefficient of 0.81 and essentially the same model equation as obtained from multiple 100 ns trajectories (as shown in Figure 7). Table 1 shows the percent errors in the RDF hydration layer integral values at different time intervals relative to the function averaged over 100 ns. These results give further confidence in using short-length MD trajectories

Table 1. Percent Errors of RDF Hydration Layer Integrals at Different Simulation Time Points Relative to RDF Values for an Entire 100 ns Trajectory

	average percent error			
protein	10 ns	7 ns	5 ns	2 ns
CALB	1.03	0.50	0.43	0.93
HEWL	0.93	0.60	0.66	1.39
SC	1.39	0.74	0.66	1.24
RNase A	3.30	1.68	0.68	1.65
CheY	2.64	2.28	3.89	4.60

to acquire structural data from which to predict and map regional hydration dynamics.

Finally, the z test was used to test the hypothesis that the diffusion coefficients produced by the above-developed rapid analysis model (using the RDF of the first 10 ns after convergence) are not significantly different from calculated diffusion coefficients from the MD simulations. Since it is observed that the calculated $|z|$ of 0.516 is less than $z_c = 1.96$, it is concluded that the considered hypothesis is not rejected, indicating that diffusion coefficients produced by the developed model (using the RDF of first 10 ns after convergence) are not significantly different at the 95% confidence level from the diffusion coefficients calculated directly from MD trajectories. Overall, it can be stated that short MD trajectories (10 ns after equilibration) can be used to acquire reliable RDF data from which to calculate translational dynamics in regional hydration layers using the linear correlation model. This enables researchers to revisit archived protein trajectories and rapidly evaluate and map regional hydration dynamics.

Model Validation at Lower Temperatures. In order to determine the validity of the correlation between the regional hydration shell RDF integral and the water diffusion coefficient across a range of laboratory-relevant temperatures, results were obtained at 278 and 288 K for the CALB and HEWL proteins. It was observed that the RDF hydration layer integrals at different temperatures were very similar, but the hydration layer diffusive dynamics of course slows down as the temperature decreases. The linear correlation between structure and dynamics holds quite well at 288 K, as can be seen in Figure 12b, with decreased slope and intercept values. It appears that the generality of the equation across proteins may begin to diverge at 5 °C, with a lower correlation. Figure S16 shows linear fits to the individual proteins rather than the aggregate set, as shown in Figure 12a. Therefore, it can be stated that linear correlations are observed between regional hydration shell RDF integrals and apparent water diffusion rates at temperatures near room temperature (288 and 298 K) but that the established model performs more poorly as the temperature approaches water's freezing point.

Diffusion from Velocity Autocorrelation Function. Sharp recently suggested that for the hydration layer, diffusion coefficients calculated by the Kubo relation (integration of velocity autocorrelation functions) provide more local measures of diffusive dynamics, since MSD analysis necessarily involves movement away from the site of origin and perhaps outside of the region of interest.[40] Indeed, velocity ACFs decay within ~10 ps, while our MSD analysis evaluates water translation from 20 to 60 ps after identification at its original position in the hydration layer (at time $t = 0$), a time frame on the order of regional hydration layer residence times.[42]

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

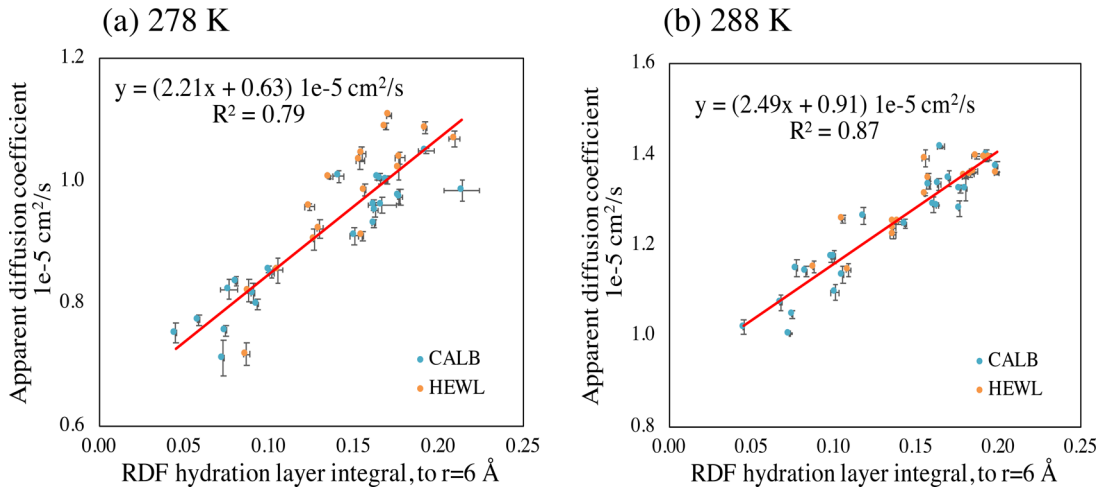


Figure 12. Linear correlation of local water structure/density obtained using integration of the RDF with local solvent shell water diffusion coefficients at (a) 278 K and (b) 288 K.

Table 2. Comparison of Slow-Diffusing Crystallographic Water Identified in the Present Study with Experimentally Identified Buried/Long-Residing Crystallographic Water Molecules

protein	locations of experimentally identified buried/long-residing crystallographic water molecules	locations of slow-diffusing crystallographic water molecules identified in this study	crystallographic water residue numbers of slow-diffusing water molecules identified in this study
HEWL	Ala31, Thr40, Ile55, Leu56, Arg61, Cys64, Leu84, Ser86, Ala90[127]	Thr40, Gly54, Ile55, Arg61, Cys64, Leu83, Leu84, Ser85, Ser86, Asp87	133, 137, 143, 213, 250
SC	His67, Thr71, Val205, Met222, Pro225, His226[128]	Thr33, His67, Thr71, Val205, Tyr206, Pro225, His226	453, 471, 483, 486, 557, 656, 679, 761, 785
CheY	around residues 17, 37, 71 and 121[21]	Asn18, Lys21, Lys70	343, 407
RNase A	reported to have no buried waters[32]	none identified	

Apparent diffusion coefficients were calculated using the Kubo relation (eq 2) and were found to be consistent with values calculated by MSD, with a high correlation between the diffusive dynamics (via the Kubo relation) and RDF hydration layer integral values (R^2 = 0.86; Figure S17).

Identifying Buried/Bound Waters or Water Molecules Having Long Residence Times. Slow-diffusing water molecules were identified for the HEWL, SC, RNase A, and CheY proteins using the protocol in our previously published work.[111] Detailed information regarding identification of slow-diffusing/buried crystallographic water is provided in Protocol 4 in the Supporting Information. These identified slow-diffusing water molecules were compared with experimentally determined and previously published regions of tightly bound or long-residing water molecules. Table 2 provides this comparison along with the locations and water residue numbers of identified slow-diffusing water molecules.

Marchi and co-workers identified the locations of five buried water molecules at the HEWL protein.[127] Our protocol for identifying slow-diffusing crystallographic water molecules successfully captured the buried waters at these locations, with the exception of the crystallographic water near Ala90, which Marchi and co-workers reported as joining the surrounding bulk water after 4 ns.[127] Pedersen et al. identified buried water molecules in the interface between the two half-domains in SC.[128] Our method was successful in capturing slow-diffusing water molecules present in these locations. Early crystallographic analysis of RNase A speculated that a region of conserved crystallographic water molecules near residues 64–66 may have long residence times,[129] but the simulation data suggest that these waters diffuse at rates within the range of external hydration layer dynamics found to be common across all of the studied proteins in this set [$(1.4–2.0) \times 10^{-5}$ cm^2/s]. In fact, RNase A is now considered to be among the few proteins that do not have buried water molecules,[32] and our protocol for identifying slow-diffusing crystallographic water molecules did not capture any slow-diffusing crystallographic water molecules on the RNase A protein surface. For CheY protein, Han and co-workers identified bound surface water within the vicinity of residues 17, 37, 71, and 121 using Overhauser dynamic nuclear polarization measurements.[21] Our method captured slow-diffusing water molecules for the CheY protein at nearby residue sites 18, 21, and 70.

DISCUSSION

At different regions of proteins' surfaces, a high correlation is seen between the protein–water (water center of mass around protein side-chain atoms) radial distribution function and the mobility of water molecules in the hydration layer. The integral of the RDF hydration layer peak was found to be the structural data point most highly correlated with apparent diffusion coefficients calculated at different regions around each protein. Five proteins were examined, and the diffusion coefficient and RDF integral values were seen to obey a general relationship, providing a correlation with R^2 = 0.84. Furthermore, this correlation was seen to hold over a range of laboratory-relevant temperatures and with a different all-atom protein force field (OPLSAA) and water model (TIP4P). The model linear equation changes modestly with force field and, as expected, more significantly with temperature (as the kinetic energy is reduced).

The protein–water RDF reflects differences in the local water structure,[117,130] including variation in tetrahedrality and local excess entropy arising from surface curvature and local

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407–2422

arrangements of protein hydrogen-bond partners. The protein−water RDF also reflects local solvent-excluded volume (occupied by the protein), which itself reduces the translational mobility of water relative to the bulk.[20,63,64,66] In this survey of proteins, it is seen that there is a general, predictive relationship between the protein−water RDF and the solvation shell water diffusion rates for multiple proteins. Halle and coworkers observed from NMR spectroscopic results and simulations data that water dynamics at exterior surfaces (i.e., diffusive hydration layer waters, such as those investigated herein) are similar across a set of proteins.[20] This agrees with our results. In contrast to external, diffusive hydration layer water molecules with a common distribution of dynamics, protein-specific effects are seen to arise in the dynamics of water within confined sites (i.e., tightly bound or buried waters). For such reasons, it is recommended that buried waters be identified separately using the protocol outlined in Methods (detailed in Protocol 4 in the Supporting Information). It should be noted that the procedure for identifying bound waters is somewhat arbitrary, but the simulation results presented here were sufficient to capture the majority of experimentally identified buried or bound waters for these proteins. A further complication is that experimental determinations of "bound waters" can suffer from bias,[122] as concluded by Zhang and Matthews, who analyzed crystallographic identification of solvent-binding sites.[132] Once buried or tightly bound water molecules are identified, statistical analysis of their dynamics can be carried out over multiple blocks of simulation data. Identification of sites of slow-moving waters can serve as markers for regions that potentially have roles in protein−protein interactions[22] or form parts of the substrate access pathway.[7,22,23] Solvent entropy (i.e., from tightly bound water molecules) has also been implicated in allosteric regulation.[131]

The structural informatics approach taken here is for characterization of hydration dynamics for exterior regions of the protein, where the majority of diffusive solvent shell waters are found. Although results from structural data are indicated to be statistically equivalent to diffusion coefficients calculated directly from MD trajectories, the approach may find its niche in qualitatively ranking the regions of a protein as having fast, intermediate, or slow-moving hydration shell water molecules. This information can be applied to the evaluation of protein function, for instance, by identifying likely sites for protein−ligand and protein−protein interactions, which are often marked by retarded hydration dynamics that plays a role in molecular recognition.[7,22] In addition to molecular recognition, it has long been recognized that hydration dynamics is intimately connected to protein dynamics.[33,34,36,37,41,133−149] Our recent work on the solvation layer dynamics of CALB suggests that differences in regional solvation layer dynamics give rise to different friction environments that modulate the flexibility of protein regions.[42] Acquiring protein hydration dynamics via structural information will facilitate evaluation of a large set of data across many proteins so that hypotheses regarding the connection between protein dynamics and solvent dynamics can be further tested.

Although this work focuses on apparent diffusion coefficients, excess entropy scaling (such as Rosenfeld scaling) with S_2 calculated from structural data is valid for other transport properties described by time correlation functions, including viscosity and reorientation times.[92] Moreover, we have seen that for the CALB enzyme, hydration layer residence times and water reorientation times are also correlated with the RDF hydration layer integral, albeit with slightly lower values of R^2 correlation values. It should also be noted that these measures of dynamics correlate well with diffusion within the CALB data set.[65] Thus, evaluation of solvation shell structure may also yield linear regression models describing other measures of water dynamics and transport properties in the hydration layer if desired. The mapping provided here may not be as quantitatively correlated for other measures of dynamics, such as reorientation and residence times, but it should be indicative of regions of fast and slow dynamics. Correlated translational and reorientational dynamics have been measured experimentally[51,150] (by neutron scattering) and observed and described in simulation studies.[151,152] Not only are translational and reorientational dynamics correlated when averaged over all of the water molecules in a region,[65,153] but also the translational and rotational dynamics of individual water molecules are correlated.[154,155] Such findings are in line with those of Laage and Hynes, who described correlated hydrogen-bond lifetimes and reorientation in bulk water.[156]

Furthermore, for a rapid estimate of hydration dynamics around a certain protein region, comparison of protein solvation layer RDF peak heights can be suggested to give a good approximation. As can be seen in the RDFs given in Figures 3a and 6b, the alpha 1, 5, and 10 regions of CALB (Figure 3a) and regions 1−5, 113−115, 67−71, and 85−96 of RNase A (Figure 6b) have comparatively higher RDF peaks. According to our developed protocol, we should be able to guess that water molecules present in the hydration layer around these regions should have comparatively fast hydration dynamics. This is indeed so, with the calculated apparent diffusion coefficients (Tables S1 and S4) indicating that solvent shell water molecules at these regions have faster diffusive dynamics. Therefore, it serves as a rule of thumb that regions having higher (lower) peak values in the hydration layer RDFs have faster (slower)-moving waters. Furthermore, the model could be evaluated for application in rapidly analyzing hydration dynamics near other interfaces, such as polymeric materials, and other biomolecular structures, such as lipid bilayers.

CONCLUSIONS

An informatics approach for rapid evaluation of regional hydration dynamics around a protein has been identified, by which integration of the hydration layer peak in the water−protein RDF, acquired from a 10 ns MD trajectory after simulation convergence (typically, the system is equilibrated within 2−5 ns), can be used to calculate apparent diffusion coefficients in the hydration layer. This can then be used to identify regions of fast-moving and slow-moving water around different parts of the protein, mapping variations in water dynamics to the protein surface. The solvent RDF data can be easily extracted from archived MD trajectories to map the hydration dynamics, since this method does not require the frequent saving of solvent coordinates that is required for direct calculation of diffusive dynamics. Thus, old data can yield new information. This will open up new avenues for evaluation of the role of hydration dynamics in protein structure, dynamics, and function.

Apparent diffusion coefficients of water molecules in protein hydration layers were calculated from MSD data around 92 different regions of five proteins using MD simulations. The water dynamics at different regions of the proteins' surfaces

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

(i.e., each external α-helix, β-sheet, and loop region) were shown to vary by location, echoing spectroscopic measurements of protein solvation in previous literature reports showing heterogeneous hydration dynamics. A robust correlation between water dynamics and hydration layer structure, as indicated by integration of the hydration layer peak in the water center of mass around protein side-chain atom RDF (out to 6 Å), was found for the entire data set. We previously demonstrated in studies of the CALB enzyme that this structure−dynamics correlation arises from an excess entropy relationship, common in transport coefficients,[92] that is described by Rosenfeld scaling.[65] Excluded volume from surrounding protein atoms also contributes to slowed dynamics with concomitant reduced RDF integral values.

In summary, a linear regression model for rapid analysis of regional hydration dynamics was developed and evaluated. Analysis indicates that apparent diffusion coefficients predicted from integration of the hydration layer in the RDF using the linear model are statististically equivalent at the 95% confidence interval to those calculated directly from simulation data (block-averaged over three 100 ns trajectories). Further analysis determined that 10 ns of MD data acquired after simulation convergence is sufficient to provide valid dynamics data and that RDF hydration layer integration values after 10 ns have 1−3% error compared to RDF data averaged over 100 ns. The model provided herein is based on room-temperature simulations in the GROMACS software package using the AMBER03 protein force field and SPC/E water model, a force field/water model combination shown to have very good agreement with spectroscopically determined water dynamics values. Furthermore, a linear correlation is valid at lower temperature and for the OPLSAA protein force field and TIP4P water model as well. We suggest that the linear equation model provided here can be used directly on unknown proteins simulated using similar simulation conditions and anticipate that linear models can be developed for other simulation protocols. A protocol for quick identification of buried water molecules is also provided. The results pave the way for rapid analysis of protein hydration dynamics. It is known that solvation layer characteristics are intrinsic to protein structure, dynamics, and function, and more facile access to characterization of the hydration layer will enable advancement of our understanding of solvation's role in biomolecular function.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00009.

Details of simulation protocols and protocols for rapid analysis of hydration dynamics; protocol for the calculation of hydration layer water density using excluded volume; RDF for bulk water in the CALB protein simulation box (Figure S2); linear correlation of local water density calculated using hydration shell excluded volume with local solvent shell water apparent diffusion coefficients (Figure S3); external, solvent-exposed regions of proteins used for analysis of solvent shell structure and regional hydration dynamics (Figure S4); diffusion map for each protein (Figure S5); mean-square displacement linear fitting for different time intervals (Figure S6); comparison of relationships between water diffusion and water structure approximated using the water oxygen RDF and the water center of mass RDF (Figure S7); comparison of relationships between water diffusion and water structure approximated using RDF hydration layer integration and peak height (Figures S8−S12); data point corresponding to SPC/E bulk water with linear correlation of local water density with local solvent shell water apparent diffusion coefficients (Figure S13); linear correlation of local water density with local solvent shell water apparent diffusion coefficients for α-helices, β-sheets, and connectors (Figure S14); root-mean-square deviation for each protein (Figure S15); linear correlation of local water density with local solvent shell water apparent diffusion coefficients at 278 K (Figure S16); linear correlation of local water density calculated using RDF peak integral with local solvent shell water diffusion coefficients calculated using velocity autocorrelation function (Figure S17); diffusion coefficients of protein regional solvent shell water obtained using MD (Tables S1−S5); comparison of protein regional solvent shell water diffusion coefficients produced by the developed model with diffusion coefficients obtained from MD simulations (Table S6); protein simulation details for each protein (Table S7) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: katie.mitchell-koch@wichita.edu.

ORCID

Katie R. Mitchell-Koch: 0000-0002-9173-3677

Present Address

‡C.K.: National Institute of Environmental Health Sciences, Bioinformatics, 111 T. W. Alexander Drive, Durham, NC 27709, USA.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant CHE-1665157 and the Kansas INBRE, P20 GM103418. Acknowledgement is made to the donors of the American Chemical Society Petroleum Research Fund for support of this research. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation CHE1800022. XSEDE is supported by National Science Foundation Grant ACI-1548562.

■ REFERENCES

(1) Pal, S. K.; Zewail, A. H. Dynamics of Water in Biological Recognition. *Chem. Rev.* **2004**, *104*, 2099−123.

(2) Levy, Y.; Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389−415.

(3) Helms, V. Protein Dynamics Tightly Connected to the Dynamics of Surrounding and Internal Water Molecules. *ChemPhysChem* **2007**, *8*, 23−33.

(4) Chalikian, T. V. Volumetric Properties of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 207−35.

(5) Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108*, 74−108.

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

(6) Mukherjee, S.; Mondal, S.; Deshmukh, A. A.; Gopal, B.; Bagchi, B. What Gives an Insulin Hexamer Its Unique Shape and Stability? Role of Ten Confined Water Molecules. *J. Phys. Chem. B* **2018**, *122*, 1631−1637.
(7) Grossman, M.; Born, B.; Heyden, M.; Tworowski, D.; Fields, G. B.; Sagi, I.; Havenith, M. Correlated Structural Kinetics and Retarded Solvent Dynamics at the Metalloprotease Active Site. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1102−8.
(8) Meister, K.; Ebbinghaus, S.; Xu, Y.; Duman, J. G.; DeVries, A.; Gruebele, M.; Leitner, D. M.; Havenith, M. Long-Range Protein-Water Dynamics in Hyperactive Insect Antifreeze Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1617−22.
(9) Dielmann-Gessner, J.; Grossman, M.; Conti Nibali, V.; Born, B.; Solomonov, I.; Fields, G. B.; Havenith, M.; Sagi, I. Enzymatic Turnover of Macromolecules Generates Long-Lasting Protein-Water-Coupled Motions beyond Reaction Steady State. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17857−62.
(10) Nucci, N. V.; Pometun, M. S.; Wand, A. J. Mapping the Hydration Dynamics of Ubiquitin. *J. Am. Chem. Soc.* **2011**, *133*, 12326−9.
(11) Armstrong, B. D.; Choi, J.; Lopez, C.; Wesener, D. A.; Hubbell, W.; Cavagnero, S.; Han, S. Site-Specific Hydration Dynamics in the Nonpolar Core of a Molten Globule by Dynamic Nuclear Polarization of Water. *J. Am. Chem. Soc.* **2011**, *133*, 5987−95.
(12) Nucci, N. V.; Pometun, M. S.; Wand, A. J. Site-Resolved Measurement of Water-Protein Interactions by Solution Nmr. *Nat. Struct. Mol. Biol.* **2011**, *18*, 245−9.
(13) Mattea, C.; Qvist, J.; Halle, B. Dynamics at the Protein-Water Interface from 17 O Spin Relaxation in Deeply Supercooled Solutions. *Biophys. J.* **2008**, *95*, 2951−2963.
(14) Jha, S. K.; Ji, M. B.; Gaffney, K. J.; Boxer, S. G. Site-Specific Measurement of Water Dynamics in the Substrate Pocket of Ketosteroid Isomerase Using Time-Resolved Vibrational Spectroscopy. *J. Phys. Chem. B* **2012**, *116*, 11414−11421.
(15) King, J. T.; Kubarych, K. J. Site-Specific Coupling of Hydration Water and Protein Flexibility Studied in Solution with Ultrafast 2d-Ir Spectroscopy. *J. Am. Chem. Soc.* **2012**, *134*, 18705−12.
(16) Malardier-Jugroot, C.; Johnson, M. E.; Murarka, R. K.; Head-Gordon, T. Aqueous Peptides as Experimental Models for Hydration Water Dynamics near Protein Surfaces. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4903−4908.
(17) Qiu, W.; Zhang, L.; Okobiah, O.; Yang, Y.; Wang, L.; Zhong, D.; Zewail, A. H. Ultrafast Solvation Dynamics of Human Serum Albumin: Correlations with Conformational Transitions and Site-Selected Recognition. *J. Phys. Chem. B* **2006**, *110*, 10540−9.
(18) Qiu, W.; Kao, Y. T.; Zhang, L.; Yang, Y.; Wang, L.; Stites, W. E.; Zhong, D.; Zewail, A. H. Protein Surface Hydration Mapped by Site-Specific Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13979−84.
(19) Jia, M.; Yang, J.; Qin, Y.; Wang, D.; Pan, H.; Wang, L.; Xu, J.; Zhong, D. Determination of Protein Surface Hydration by Systematic Charge Mutations. *J. Phys. Chem. Lett.* **2015**, *6*, 5100−5.
(20) Persson, F.; Soderhjelm, P.; Halle, B. How Proteins Modify Water Dynamics. *J. Chem. Phys.* **2018**, *148*, 215103.
(21) Barnes, R.; Sun, S.; Fichou, Y.; Dahlquist, F. W.; Heyden, M.; Han, S. I. Spatially Heterogeneous Surface Water Diffusivity around Structured Protein Surfaces at Equilibrium. *J. Am. Chem. Soc.* **2017**, *139*, 17890−17901.
(22) Caro, J. A.; Harpole, K. W.; Kasinath, V.; Lim, J.; Granja, J.; Valentine, K. G.; Sharp, K. A.; Wand, A. J. Entropy in Molecular Recognition by Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 6563−6568.
(23) Dielmann-Gessner, J.; Grossman, M.; Conti Nibali, V.; Born, B.; Solomonov, I.; Fields, G. B.; Havenith, M.; Sagi, I. Enzymatic Turnover of Macromolecules Generates Long-Lasting Protein-Water-Coupled Motions Beyond Reaction Steady State. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17857−62.
(24) Halle, B. Protein Hydration Dynamics in Solution: A Critical Survey. *Philos. Trans R. Soc., B* **2004**, *359*, 1207−23 (discussion 1223−4 and 1323−8) .
(25) Mondal, S.; Mukherjee, S.; Bagchi, B. Origin of Diverse Time Scales in the Protein Hydration Layer Solvation Dynamics: A Simulation Study. *J. Chem. Phys.* **2017**, *147*, 154901.
(26) Yang, J.; Wang, Y.; Wang, L.; Zhong, D. Mapping Hydration Dynamics around a Beta-Barrel Protein. *J. Am. Chem. Soc.* **2017**, *139*, 4399−4408.
(27) Sterpone, F.; Stirnemann, G.; Laage, D. Magnitude and Molecular Origin of Water Slowdown Next to a Protein. *J. Am. Chem. Soc.* **2012**, *134*, 4116−4119.
(28) Mukherjee, S.; Mondal, S.; Bagchi, B. Distinguishing Dynamical Features of Water inside Protein Hydration Layer: Distribution Reveals What Is Hidden Behind the Average. *J. Chem. Phys.* **2017**, *147*, 024901.
(29) Schiro, G.; Fichou, Y.; Gallat, F. X.; Wood, K.; Gabel, F.; Moulin, M.; Hartlein, M.; Heyden, M.; Colletier, J. P.; Orecchini, A.; Paciaroni, A.; Wuttke, J.; Tobias, D. J.; Weik, M. Translational Diffusion of Hydration Water Correlates with Functional Motions in Folded and Intrinsically Disordered Proteins. *Nat. Commun.* **2015**, *6*, 6490.
(30) Bellissent-Funel, M. C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673−97.
(31) Franck, J. M.; Han, S. Overhauser Dynamic Nuclear Polarization for the Study of Hydration Dynamics, Explained. *Methods Enzymol.* **2019**, *615*, 131−175.
(32) Denisov, V. P.; Halle, B. Thermal Denaturation of Ribonuclease a Characterized by Water 17o and 2h Magnetic Relaxation Dispersion. *Biochemistry* **1998**, *37*, 9595−9604.
(33) King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J. Site-Specific Hydration Dynamics of Globular Proteins and the Role of Constrained Water in Solvent Exchange with Amphiphilic Cosolvents. *J. Phys. Chem. B* **2012**, *116*, 5604−5611.
(34) King, J. T.; Kubarych, K. J. Site-Specific Coupling of Hydration Water and Protein Flexibility Studied in Solution with Ultrafast 2d-Ir Spectroscopy. *J. Am. Chem. Soc.* **2012**, *134*, 18705−18712.
(35) Fisette, O.; Paslack, C.; Barnes, R.; Isas, J. M.; Langen, R.; Heyden, M.; Han, S.; Schafer, L. V. Hydration Dynamics of a Peripheral Membrane Protein. *J. Am. Chem. Soc.* **2016**, *138*, 11526−11535.
(36) Bagchi, B. Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem. Rev.* **2005**, *105*, 3197−3219.
(37) Fogarty, A. C.; Duboué-Dijon, E.; Sterpone, F.; Hynes, J. T.; Laage, D. Biomolecular Hydration Dynamics: A Jump Model Perspective. *Chem. Soc. Rev.* **2013**, *42*, 5672−5683.
(38) George, D. K.; Charkhesht, A.; Hull, O. A.; Mishra, A.; Capelluto, D. G.; Mitchell-Koch, K. R.; Vinh, N. Q. New Insights into the Dynamics of Zwitterionic Micelles and Their Hydration Waters by Gigahertz-to-Terahertz Dielectric Spectroscopy. *J. Phys. Chem. B* **2016**, *120*, 10757−10767.
(39) Charkhesht, A.; Regmi, C. K.; Mitchell-Koch, K. R.; Cheng, S.; Vinh, N. Q. High-Precision Megahertz-to-Terahertz Dielectric Spectroscopy of Protein Collective Motions and Hydration Dynamics. *J. Phys. Chem. B* **2018**, *122*, 6341−6350.
(40) Sharp, K. A. Companion Simulations and Modeling to NMR-Based Dynamical Studies of Proteins. *Methods Enzymol.* **2019**, *615*, 1−41.
(41) Khodadadi, S.; Sokolov, A. P. Atomistic Details of Protein Dynamics and the Role of Hydration Water. *Biochim. Biophys. Acta, Gen. Subj.* **2017**, *1861*, 3546−3552.
(42) Dahanayake, J. N.; Mitchell-Koch, K. R. How Does Solvation Layer Mobility Affect Protein Structural Dynamics? *Front Mol. Biosci.* **2018**, *5*, 65.
(43) Meinhardt, S.; Manley, M. W.; Parente, D. J.; Swint-Kruse, L. Rheostats and Toggle Switches for Modulating Protein Function. *PLoS One* **2013**, *8*, e83502.

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

(44) Franck, J. M.; Pavlova, A.; Scott, J. A.; Han, S. Quantitative Overhauser Effect Dynamic Nuclear Polarization for the Analysis of Local Water Dynamics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *74*, 33−56.
(45) Jorge, C.; Marques, B. S.; Valentine, K. G.; Wand, A. J. Characterizing Protein Hydration Dynamics Using Solution NMR Spectroscopy. *Methods Enzymol.* **2019**, *615*, 77−101.
(46) Nilsson, L.; Halle, B. Molecular Origin of Time-Dependent Fluorescence Shifts in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13867−72.
(47) Kamal, J. K.; Zhao, L.; Zewail, A. H. Ultrafast Hydration Dynamics in Protein Unfolding: Human Serum Albumin. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 13411−6.
(48) Zhang, L.; Wang, L.; Kao, Y. T.; Qiu, W.; Yang, Y.; Okobiah, O.; Zhong, D. Mapping Hydration Dynamics around a Protein Surface. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18461−6.
(49) Lingwood, M. D.; Han, S. Solution-State Dynamic Nuclear Polarization. *Annu. Rep. NMR Spectrosc.* **2011**, *73*, 83−126.
(50) Barnes, R.; Sun, S.; Fichou, Y.; Dahlquist, F. W.; Heyden, M.; Han, S. Spatially Heterogeneous Surface Water Diffusivity around Structured Protein Surfaces at Equilibrium. *J. Am. Chem. Soc.* **2017**, *139*, 17890−17901.
(51) Perticaroli, S.; Ehlers, G.; Stanley, C. B.; Mamontov, E.; O'Neill, H.; Zhang, Q.; Cheng, X.; Myles, D. A.; Katsaras, J.; Nickels, J. D. Description of Hydration Water in Protein (Green Fluorescent Protein) Solution. *J. Am. Chem. Soc.* **2017**, *139*, 1098−1105.
(52) Abseher, R.; Schreiber, H.; Steinhauser, O. The Influence of a Protein on Water Dynamics in Its Vicinity Investigated by Molecular Dynamics Simulation. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 366−78.
(53) Liu, P.; Harder, E.; Berne, B. On the Calculation of Diffusion Coefficients in Confined Fluids and Interfaces with an Application to the Liquid- Vapor Interface of Water. *J. Phys. Chem. B* **2004**, *108*, 6595−6602.
(54) Tarek, M.; Tobias, D. J. The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-Scattering Experiments. *Biophys. J.* **2000**, *79*, 3244−57.
(55) Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics Simulation Evidence of Anomalous Diffusion of Protein Hydration Water. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1996**, *53*, R3040−R3043.
(56) Wood, K.; Frölich, A.; Paciaroni, A.; Moulin, M.; Härtlein, M.; Zaccai, G.; Tobias, D. J.; Weik, M. Coincidence of Dynamical Transitions in a Soluble Protein and Its Hydration Water: Direct Measurements by Neutron Scattering and Md Simulations. *J. Am. Chem. Soc.* **2008**, *130*, 4586−4587.
(57) Lagi, M.; Chu, X.; Kim, C.; Mallamace, F.; Baglioni, P.; Chen, S.-H. The Low-Temperature Dynamic Crossover Phenomenon in Protein Hydration Water: Simulations Vs Experiments. *J. Phys. Chem. B* **2008**, *112*, 1571−1575.
(58) Norton, C. D.; Thompson, W. H. On the Diffusion of Acetonitrile in Nanoscale Amorphous Silica Pores. Understanding Anisotropy and the Effects of Hydrogen Bonding. *J. Phys. Chem. C* **2013**, *117*, 19107−19114.
(59) Tournier, A. L.; Xu, J.; Smith, J. C. Translational Hydration Water Dynamics Drives the Protein Glass Transition. *Biophys. J.* **2003**, *85*, 1871−5.
(60) Chiavazzo, E.; Fasano, M.; Asinari, P.; Decuzzi, P. Scaling Behaviour for the Water Transport in Nanoconfined Geometries. *Nat. Commun.* **2014**, *5*, 3565.
(61) Russo, D.; Hura, G.; Head-Gordon, T. Hydration Dynamics near a Model Protein Surface. *Biophys. J.* **2004**, *86*, 1852−62.
(62) Wong, C. F.; McCammon, J. A. Computer Simulation and the Design of New Biological Molecules. *Isr. J. Chem.* **1986**, *27*, 211−215.
(63) Luise, A.; Falconi, M.; Desideri, A. Molecular Dynamics Simulation of Solvated Azurin: Correlation between Surface Solvent Accessibility and Water Residence Times. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 56−67.
(64) Henchman, R. H.; McCammon, J. A. Structural and Dynamic Properties of Water around Acetylcholinesterase. *Protein Sci.* **2002**, *11*, 2080−90.
(65) Dahanayake, J. N.; Mitchell-Koch, K. R. Entropy Connects Water Structure and Dynamics in Protein Hydration Layer. *Phys. Chem. Chem. Phys.* **2018**, *20*, 14765−14777.
(66) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **2000**, *79*, 2966−74.
(67) Makarov, V. A.; Feig, M.; Andrews, B. K.; Pettitt, B. M. Diffusion of Solvent around Biomolecular Solutes: A Molecular Dynamics Simulation Study. *Biophys. J.* **1998**, *75*, 150−8.
(68) Czapiewski, D.; Zielkiewicz, J. Structural Properties of Hydration Shell around Various Conformations of Simple Polypeptides. *J. Phys. Chem. B* **2010**, *114*, 4536−4550.
(69) Wade, A. D.; Wang, L. P.; Huggins, D. J. Assimilating Radial Distribution Functions to Build Water Models with Improved Structural Properties. *J. Chem. Inf. Model.* **2018**, *58*, 1766−1778.
(70) Sharma, R.; Agarwal, M.; Chakravarty, C. Estimating the Entropy of Liquids from Atom-Atom Radial Distribution Functions: Silica, Beryllium Fluoride and Water. *Mol. Phys.* **2008**, *106*, 1925−1938.
(71) Rosenfeld, Y. Relation between the Transport Coefficients and the Internal Entropy of Simple Systems. *Phys. Rev. A: At., Mol., Opt. Phys.* **1977**, *15*, 2545.
(72) Dzugutov, M. A Universal Scaling Law for Atomic Diffusion in Condensed Matter. *Nature* **1996**, *381*, 137.
(73) Chopra, R.; Truskett, T. M.; Errington, J. R. On the Use of Excess Entropy Scaling to Describe Single-Molecule and Collective Dynamic Properties of Hydrocarbon Isomer Fluids. *J. Phys. Chem. B* **2010**, *114*, 16487−16493.
(74) Vaz, R. V.; Magalhães, A. L.; Fernandes, D. L.; Silva, C. M. Universal Correlation of Self-Diffusion Coefficients of Model and Real Fluids Based on Residual Entropy Scaling Law. *Chem. Eng. Sci.* **2012**, *79*, 153−162.
(75) Novak, L. T. Predictive Corresponding-States Viscosity Model for the Entire Fluid Region: N-Alkanes. *Ind. Eng. Chem. Res.* **2013**, *52*, 6841−6847.
(76) Baled, H. O.; Gamwo, I. K.; Enick, R. M.; McHugh, M. A. Viscosity Models for Pure Hydrocarbons at Extreme Conditions: A Review and Comparative Study. *Fuel* **2018**, *218*, 89−111.
(77) Goel, G.; Lacks, D. J.; Van Orman, J. A. Transport Coefficients in Silicate Melts from Structural Data Via a Structure-Thermodynamics-Dynamics Relationship. *Phy Rev. E* **2011**, *84*, 051506.
(78) Jakse, N.; Pasturel, A. Excess Entropy Scaling Law for Diffusivity in Liquid Metals. *Sci. Rep.* **2016**, *6*, 20689.
(79) Singh, M.; Liu, H.; Kumar, S. K.; Ganguly, A.; Chakravarty, C. Excess Entropy and Structural Transitions in a Two-Dimensional Square-Shoulder Fluid. *J. Chem. Phys.* **2010**, *132*, 074503.
(80) Sharma, R.; Chakraborty, S. N.; Chakravarty, C. Entropy, Diffusivity, and Structural Order in Liquids with Waterlike Anomalies. *J. Chem. Phys.* **2006**, *125*, 204501.
(81) Cao, Q.-L.; Wang, P.-P.; Shao, J.-X.; Wang, F.-H. Transport Properties and Entropy-Scaling Laws for Diffusion Coefficients in Liquid Fe 0.9 Ni 0.1 up to 350 GPa. *RSC Adv.* **2016**, *6*, 84420−84425.
(82) Fouad, W. A.; Vega, L. F. Transport Properties of Hfc and Hfo Based Refrigerants Using an Excess Entropy Scaling Approach. *J. Supercrit. Fluids* **2018**, *131*, 106−116.
(83) Bastea, S. Thermodynamics and Diffusion in Size-Symmetric and Asymmetric Dense Electrolytes. *J. Chem. Phys.* **2011**, *135*, 084515.
(84) Mittal, J.; Errington, J. R.; Truskett, T. M. Thermodynamics Predicts How Confinement Modifies the Dynamics of the Equilibrium Hard-Sphere Fluid. *Phys. Rev. Lett.* **2006**, *96*, 177804.
(85) Ingebrigtsen, T. S.; Errington, J. R.; Truskett, T. M.; Dyre, J. C. Predicting How Nanoconfinement Changes the Relaxation Time of a Supercooled Liquid. *Phys. Rev. Lett.* **2013**, *111*, 235901.
(86) Mittal, J.; Errington, J. R.; Truskett, T. M. Relationships between Self-Diffusivity, Packing Fraction, and Excess Entropy in

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

Simple Bulk and Confined Fluids. *J. Phys. Chem. B* **2007**, *111*, 10054−10063.
(87) Ma, X.; Chen, W.; Wang, Z.; Peng, Y.; Han, Y.; Tong, P. Test of the Universal Scaling Law of Diffusion in Colloidal Monolayers. *Phys. Rev. Lett.* **2013**, *110*, 078302.
(88) Liu, Y.; Guo, F.; Hu, J.; Zhao, S.; Liu, H.; Hu, Y. Entropy Prediction for H 2 Adsorption in Metal−Organic Frameworks. *Phys. Chem. Chem. Phys.* **2016**, *18*, 23998−24005.
(89) Fu, J.; Tian, Y.; Wu, J. Classical Density Functional Theory for Methane Adsorption in Metal-Organic Framework Materials. *AIChE J.* **2015**, *61*, 3012−3021.
(90) Liu, Y.; Fu, J.; Wu, J. Excess-Entropy Scaling for Gas Diffusivity in Nanoporous Materials. *Langmuir* **2013**, *29*, 12997−13002.
(91) He, P.; Li, H.; Hou, X. Excess-Entropy Scaling of Dynamics for Methane in Various Nanoporous Materials. *Chem. Phys. Lett.* **2014**, *593*, 83−88.
(92) Dyre, J. C. Perspective: Excess-Entropy Scaling. *J. Chem. Phys.* **2018**, *149*, 210901.
(93) Nayar, D.; Chakravarty, C. Water and Water-Like Liquids: Relationships between Structure, Entropy and Mobility. *Phys. Chem. Chem. Phys.* **2013**, *15*, 14162−77.
(94) Errington, J. R.; Truskett, T. M.; Mittal, J. Excess-Entropy-Based Anomalies for a Waterlike Fluid. *J. Chem. Phys.* **2006**, *125*, 244502.
(95) Baranyai, A.; Evans, D. J. Direct Entropy Calculation from Computer Simulation of Liquids. *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *40*, 3817.
(96) Baranyai, A.; Evans, D. J. Three-Particle Contribution to the Configurational Entropy of Simple Fluids. *Phys. Rev. A: At., Mol., Opt. Phys.* **1990**, *42*, 849.
(97) Chakraborty, S. N.; Chakravarty, C. Diffusivity, Excess Entropy, and the Potential-Energy Landscape of Monatomic Liquids. *J. Chem. Phys.* **2006**, *124*, 014507.
(98) Giuffre, E.; Prestipino, S.; Saija, F.; Saitta, A. M.; Giaquinta, P. V. Entropy from Correlations in Tip4p Water. *J. Chem. Theory Comput.* **2010**, *6*, 625−36.
(99) Czapiewski, D.; Zielkiewicz, J. Structural Properties of Hydration Shell around Various Conformations of Simple Polypeptides. *J. Phys. Chem. B* **2010**, *114*, 4536−50.
(100) Laird, B. B.; Haymet, A. Calculation of the Entropy from Multiparticle Correlation Functions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1992**, *45*, 5680.
(101) Agarwal, M.; Singh, M.; Sharma, R.; Parvez Alam, M.; Chakravarty, C. Relationship between Structure, Entropy, and Diffusivity in Water and Water-Like Liquids. *J. Phys. Chem. B* **2010**, *114*, 6995−7001.
(102) Fomin, Y. D.; Ryzhov, V. N. Water-like Anomalies of Core-Softened Fluids: Dependence on the Trajectories in ($P\rho T$) Space. 2010, arXiv:1010.4169 [cond-mat.soft]. arXiv.org e-Print archive. https://arxiv.org/abs/1010.4169 (accessed Jan 1, 2019).
(103) Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *J. Phys. Chem. B* **2010**, *114*, 1632−7.
(104) Sansom, M. S.; Kerr, I. D.; Breed, J.; Sankararamakrishnan, R. Water in Channel-Like Cavities: Structure and Dynamics. *Biophys. J.* **1996**, *70*, 693−702.
(105) Wu, X.; Lu, W.; Streacker, L. M.; Ashbaugh, H. S.; Ben-Amotz, D. Temperature-Dependent Hydrophobic Crossover Length Scale and Water Tetrahedral Order. *J. Phys. Chem. Lett.* **2018**, *9*, 1012−1017.
(106) Uppenberg, J.; Hansen, M. T.; Patkar, S.; Jones, T. A. The Sequence, Crystal Structure Determination and Refinement of Two Crystal Forms of Lipase B from Candida Antarctica. *Structure* **1994**, *2*, 293−308.
(107) Walsh, M. A.; Schneider, T. R.; Sieker, L. C.; Dauter, Z.; Lamzin, V. S.; Wilson, K. S. Refinement of Triclinic Hen Egg-White Lysozyme at Atomic Resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 522−546.
(108) Fitzpatrick, P. A.; Steinmetz, A.; Ringe, D.; Klibanov, A. M. Enzyme Crystal Structure in a Neat Organic Solvent. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 8653−8657.
(109) Boerema, D. J.; Tereshko, V. A.; Kent, S. B. Total Synthesis by Modern Chemical Ligation Methods and High Resolution (1.1 Å) X-Ray Structure of Ribonuclease A. *Biopolymers* **2008**, *90*, 278−286.
(110) Quax, T. E.; Altegoer, F.; Rossi, F.; Li, Z.; Rodriguez-Franco, M.; Kraus, F.; Bange, G.; Albers, S.-V. Structure and Function of the Archaeal Response Regulator Chey. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1259−E1268.
(111) Dahanayake, J. N.; Gautam, D. N.; Verma, R.; Mitchell-Koch, K. R. To Keep or Not to Keep? The Question of Crystallographic Waters for Enzyme Simulations in Organic Solvent. *Mol. Simul.* **2016**, *42*, 1001−1013.
(112) Abraham, M.; van der Spoel, D.; Lindahl, E.; Hess, B. *GROMACS User Manual*, version 2016.3; GROMACS Development Team, 2017.
(113) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62−74.
(114) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(115) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(116) King, J. T.; Arthur, E. J.; Brooks, C. L.; Kubarych, K. J. Crowding Induced Collective Hydration of Biological Macromolecules over Extended Distances. *J. Am. Chem. Soc.* **2014**, *136*, 188−94.
(117) Head-Gordon, T.; Hura, G. Water Structure from Scattering Experiments and Simulation. *Chem. Rev.* **2002**, *102*, 2651−70.
(118) Shoemaker, D. P.; Garland, C. W.; Nibler, J. W. *Experiments in Physical Chemistry*; WCB McGraw-Hill: Boston, 1996.
(119) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.* **1951**, *19*, 774−777.
(120) Jungwirth, P. Biological Water or Rather Water in Biology? *J. Phys. Chem. Lett.* **2015**, *6*, 2449−2451.
(121) Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017**, *117*, 10694−10725.
(122) Kovacs, H.; Mark, A. E.; van Gunsteren, W. F. Solvent Structure at a Hydrophobic Protein Surface. *Proteins: Struct., Funct., Genet.* **1997**, *27*, 395−404.
(123) Scoppola, E.; Sodo, A.; McLain, S. E.; Ricci, M. A.; Bruni, F. Water-Peptide Site-Specific Interactions: A Structural Study on the Hydration of Glutathione. *Biophys. J.* **2014**, *106*, 1701−9.
(124) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera−a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(125) Li, T.; Raizen, M. G. Brownian Motion at Short Time Scales. *Ann. Phys. (Berlin, Ger.)* **2013**, *525*, 281.
(126) Taylor, J. *Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*; University Science Books: Sausalito, CA, 1997.
(127) Sterpone, F.; Ceccarelli, M.; Marchi, M. Dynamics of Hydration in Hen Egg White Lysozyme. *J. Mol. Biol.* **2001**, *311*, 409−19.
(128) Pedersen, J. T.; Olsen, O. H.; Betzel, C.; Eschenburg, S.; Branner, S.; Hastrup, S. Cavity Mutants of Savinase: Crystal Structures and Differential Scanning Calorimetry Experiments Give Hints of the Function of the Buried Water Molecules in Subtilisins. *J. Mol. Biol.* **1994**, *242*, 193−202.

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422

(129) Gilliland, G. L.; Dill, J.; Pechik, I.; Svensson, L. A.; Sjolin, L. The Active Site of Bovine Pancreatic Ribonuclease: An Example of Solvent Modulated Specificity. *Protein Pept. Lett.* **1994**, *1*, 60−65.
(130) Merzel, F.; Smith, J. C. Is the First Hydration Shell of Lysozyme of Higher Density Than Bulk Water? *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5378−5383.
(131) Capdevila, D. A.; Edmonds, K. A.; Campanello, G. C.; Wu, H.; Gonzalez-Gutierrez, G.; Giedroc, D. P. Functional Role of Solvent Entropy and Conformational Entropy of Metal Binding in a Dynamically Driven Allosteric System. *J. Am. Chem. Soc.* **2018**, *140*, 9108−9119.
(132) Zhang, X. J.; Matthews, B. Conservation of Solvent-Binding Sites in 10 Crystal Forms of T4 Lysozyme. *Protein Sci.* **1994**, *3*, 1031−1039.
(133) Bagchi, B. *Water in Biological and Chemical Processes: From Structure and Dynamics to Function*; Cambridge University Press, 2013.
(134) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An Extended Dynamical Hydration Shell around Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20749−20752.
(135) Beece, D.; Eisenstein, L.; Frauenfelder, H.; Good, D.; Marden, M.; Reinisch, L.; Reynolds, A.; Sorensen, L.; Yue, K. Solvent Viscosity and Protein Dynamics. *Biochemistry* **1980**, *19*, 5147−5157.
(136) Imry, Y.; Gavish, B. Correlation Functions and Structure Factors for a Lattice in a Viscous Medium. *J. Chem. Phys.* **1974**, *61*, 1554−1558.
(137) Grunwald, J.; Wirz, B.; Scollar, M. P.; Klibanov, A. M. Asymmetric Oxidoreductions Catalyzed by Alcohol Dehydrogenase in Organic Solvents. *J. Am. Chem. Soc.* **1986**, *108*, 6732−6734.
(138) Dordick, J. S.; Marletta, M. A.; Klibanov, A. M. Peroxidases Depolymerize Lignin in Organic Media but Not in Water. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6255−6257.
(139) Zaks, A.; Klibanov, A. M. The Effect of Water on Enzyme Action in Organic Media. *J. Biol. Chem.* **1988**, *263*, 8017−8021.
(140) Trodler, P.; Pleiss, J. Modeling Structure and Flexibility of Candida Antarctica Lipase B in Organic Solvents. *BMC Struct. Biol.* **2008**, *8*, 9.
(141) Klibanov, A. M. Improving Enzymes by Using Them in Organic Solvents. *Nature* **2001**, *409*, 241.
(142) Clark, D. S. Characteristics of Nearly Dry Enzymes in Organic Solvents: Implications for Biocatalysis in the Absence of Water. *Philos. Trans R Soc. Lond B Biol. Sci.* **2004**, *359*, 1299−1307.
(143) Toba, S.; Merz, K. M. The Concept of Solvent Compatibility and Its Impact on Protein Stability and Activity Enhancement in Nonaqueous Solvents. *J. Am. Chem. Soc.* **1997**, *119*, 9939−9948.
(144) Eppler, R. K.; Komor, R. S.; Huynh, J.; Dordick, J. S.; Reimer, J. A.; Clark, D. S. Water Dynamics and Salt-Activation of Enzymes in Organic Media: Mechanistic Implications Revealed by Nmr Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5706−5710.
(145) Ru, M. T.; Dordick, J. S.; Reimer, J. A.; Clark, D. S. Optimizing the Salt-Induced Activation of Enzymes in Organic Solvents: Effects of Lyophilization Time and Water Content. *Biotechnol. Bioeng.* **1999**, *63*, 233−241.
(146) Khmelnitsky, Y. L.; Welch, S. H.; Clark, D. S.; Dordick, J. S. Salts Dramatically Enhance Activity of Enzymes Suspended in Organic Solvents. *J. Am. Chem. Soc.* **1994**, *116*, 2647−2648.
(147) Schirò, G.; Fichou, Y.; Gallat, F.-X.; Wood, K.; Gabel, F.; Moulin, M.; Härtlein, M.; Heyden, M.; Colletier, J.-P.; Orecchini, A.; Paciaroni, A.; Wuttke, J.; Tobias, D. J.; Weik, M. Translational Diffusion of Hydration Water Correlates with Functional Motions in Folded and Intrinsically Disordered Proteins. *Nat. Commun.* **2015**, *6*, 6490.
(148) Duboué-Dijon, E.; Fogarty, A. C.; Hynes, J. T.; Laage, D. Dynamical Disorder in the DNA Hydration Shell. *J. Am. Chem. Soc.* **2016**, *138*, 7610−7620.
(149) Oroguchi, T.; Nakasako, M. Changes in Hydration Structure Are Necessary for Collective Motions of a Multi-Domain Protein. *Sci. Rep.* **2016**, *6*, 26302.
(150) Liu, L.; Faraone, A.; Mou, C.; Yen, C.; Chen, S. Slow Dynamics of Supercooled Water Confined in Nanoporous Silica Materials. *J. Phys.: Condens. Matter* **2004**, *16*, S5403.
(151) Liu, C.; Zhang, Y.; Zhang, J.; Wang, J.; Li, W.; Wang, W. Interplay between Translational Diffusion and Large-Amplitude Angular Jumps of Water Molecules. *J. Chem. Phys.* **2018**, *148*, 184502.
(152) Di Cola, D.; Deriu, A.; Sampoli, M.; Torcini, A. Proton Dynamics in Supercooled Water by Molecular Dynamics Simulations and Quasielastic Neutron Scattering. *J. Chem. Phys.* **1996**, *104*, 4223−4232.
(153) Mazza, M. G.; Giovambattista, N.; Starr, F. W.; Stanley, H. E. Relation between Rotational and Translational Dynamic Heterogeneities in Water. *Phys. Rev. Lett.* **2006**, *96*, 057803.
(154) Sciortino, F.; Geiger, A.; Stanley, H. E. Effect of Defects on Molecular Mobility in Liquid Water. *Nature* **1991**, *354*, 218.
(155) Mukherjee, B. Microscopic Origin of Temporal Heterogeneities in Translational Dynamics of Liquid Water. *J. Chem. Phys.* **2015**, *143*, 054503.
(156) Laage, D.; Hynes, J. T. A Molecular Jump Mechanism of Water Reorientation. *Science* **2006**, *311*, 832−835.

DOI: 10.1021/acs.jcim.9b00009
J. Chem. Inf. Model. 2019, 59, 2407−2422