

# Distinguishing Thermodynamic and Kinetic Views of the Preferential Hydration of Protein Surfaces

M. Hamsa Priya, J. K. Shah, D. Asthagiri, and M. E. Paulaitis

Department of Chemical and Biomolecular Engineering, Ohio State University, Columbus, Ohio; and Department of Chemical and Biomolecular Engineering and the Institute for Multiscale Modeling of Biological Interactions, Johns Hopkins University, Baltimore, Maryland

**ABSTRACT** Motivated by a quasi-chemical view of protein hydration, we define specific hydration sites on the surface of globular proteins in terms of the local water density at each site relative to bulk water density. The corresponding kinetic definition invokes the average residence time for a water molecule at each site and the average time that site remains unoccupied. Bound waters are identified by high site occupancies using either definition. In agreement with previous molecular dynamics simulation studies, we find only a weak correlation between local water densities and water residence times for hydration sites on the surface of two globular proteins, lysozyme and staphylococcal nuclease. However, a strong correlation is obtained when both the average residence and vacancy times are appropriately taken into account. In addition, two distinct kinetic regimes are observed for hydration sites with high occupancies: long residence times relative to vacancy times for a single water molecule, and short residence times with high turnover involving multiple water molecules. We also correlate water dynamics, characterized by average occupancy and vacancy times, with local heterogeneities in surface charge and surface roughness, and show that both features are necessary to obtain sites corresponding to kinetically bound waters.

## INTRODUCTION

Configurational complementarity in protein-protein interactions is a hallmark of molecular recognition, and leads naturally to a consideration of the molecular nature of protein hydration (1,2). The essential features of complementarity embodied in protein structures can be captured to some extent using continuum solvent models. However, water molecules strongly associated with the protein make important contributions by occupying space or providing hydrogen-bond donors and acceptors at the protein-water interface that constrain conformational space, and these effects are lost if the bound water molecules are not explicitly taken into account.

In the quasi-chemical (QC) view of protein hydration, it is natural to consider strongly associated water molecules as part of the protein (3–5). Protein solution thermodynamics is then modeled in terms of quasi-components comprised of the protein and associated water molecules immersed in a statistical field due to the remaining solvent medium. We implemented this QC view in previous work by defining specific hydration sites near the protein surface characterized by high water occupancies (6). An important finding of our earlier study was to show that the spatial distribution of these hydration sites plays an essential role in determining the configurational complementarity of protein-protein interactions.

Specific hydration sites in that study were characterized by their local water densities, which we expressed in terms of the logarithm of the chemical equilibrium constant,  $\eta$ , for water partitioning at a specific site relative to bulk water,

$$\eta \equiv \ln(\rho/\rho_b) = -\beta\Delta\mu_w^{\text{ex}}. \quad (1)$$

Here  $\Delta\mu_w^{\text{ex}}$  is the excess chemical potential of water at the hydration site relative to bulk water,  $\rho/\rho_b$  is the corresponding ratio of water densities, and  $\beta^{-1} = kT$ , the thermal energy. Equation 1 provides a thermodynamic framework for selecting hydration sites representing bound water molecules based on explicit-water molecular dynamics (MD) simulations that supply the required densities. Our criterion for strongly associated or bound water was  $\eta > 2$ , which corresponds to a local density more than seven times that of bulk water, or roughly that found for the maximum density in the first hydration shell of simple monovalent or divalent ions (7,8). Relaxing this criterion to include a larger number of more weakly associated water molecules was found to have a minimal effect on the osmotic second virial coefficient for protein-protein interactions in dilute aqueous solution (6). Thus,  $\eta > 2$  defined a lower bound on the number of explicit water molecules that must be considered to obtain the full effect of water association on these protein-protein interactions.

Protein hydration can also be characterized by water dynamics near the protein surface. Experimental techniques such as magnetic resonance dispersion (9), nuclear magnetic resonance (10), and quasi-elastic neutron scattering (11) measure relaxation times for waters buried in the interior of proteins and residence times for waters within the first hydration shell. These times are typically on the order of 10 ns–1 ms and 10 ps–1 ns, respectively, for small globular proteins. Much faster water relaxation times (0.5–100 ps) in the vicinity of a surface tryptophan residue are probed by fluorescence spectroscopy (12). Based on these timescales, water molecules have been broadly categorized as: 1), internal water (residence time  $\tau \sim 1$  ns to 1 ms); 2), water

Submitted March 14, 2008, and accepted for publication May 15, 2008.

Address reprint requests to D. Asthagiri, E-mail: dilipa@jhu.edu.

J. K. Shah's present address is Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556.

Editor: Steven D. Schwartz.

molecules that interact with the protein surface ( $\tau \sim 10\text{--}100$  ps); and 3), bulk water ( $\tau \sim 1$  ps) (9).

Water dynamics near protein surfaces have also been extensively investigated by MD simulations, and characterized by mean residence times derived from various correlation functions. The survival probability correlation function (13–17) is one such correlation function that has been widely used. The survival probability is defined as the probability of finding a water molecule within a region of interest; e.g., the first hydration shell, for a specific period of time. The hydrogen-bond correlation function (18) and the solvation energy correlation function (19) have also been used. The hydrogen-bond correlation function is defined in terms of the probability of finding a water molecule hydrogen-bonded to a protein atom for a specific period of time, with the hydrogen bond typically defined by the donor-acceptor distance and angle. The solvation correlation function is defined in terms of fluctuations in the potential energy of solute-solvent interactions.

In each case, the correlation function derived from a MD simulation is fit to either a single exponential or a series of exponential functions, or a stretched exponential function (13–19). A single exponential fit gives the water residence time directly. For a series of exponential functions, the mean water residence time is calculated as the weighted average of the characteristic time constants, whereas, for a stretched exponential function, the mean residence time is defined in terms of both the characteristic time and the stretched exponent.

It is widely accepted that molecular features of the protein surface influence water dynamics. However, previous attempts to derive a correlation between water dynamics and local chemical heterogeneities of the protein surface have produced results that are inconclusive or even contradictory (13,15–18,20). For example, the MD simulation study of water dynamics near copper plastocyanin (13) and crambin (15) found that the mean residence times computed from the survival probability correlation function depend on the chemical nature of proximal amino acids with  $\tau_{\text{charged}} \geq \tau_{\text{polar}} > \tau_{\text{nonpolar}} \approx \tau_{\text{bulk}}$ . A similar ordering of water residence times was observed for all 20 amino acids in the end-capped AXA tripeptide motif (20). However, an entirely different dependence of the water survival time was reported in MD simulation study of bovine pancreatic trypsin inhibitor (17):  $\tau_{\text{polar}} > \tau_{\text{nonpolar}} > \tau_{\text{charged}}$ .

MD simulations have also shown that both the survival probability time for waters around negatively charged residues (15,16) and the time that water molecules remain hydrogen-bonded with negatively charged residues (13) are significantly longer than that for waters near positively charged residues. In contrast, the hydrogen-bond correlation time for water near positively charged residues was found to be higher than that near negatively charged residues in the MD simulation study of HP-36 (18).

The correlation between local water densities and the chemical nature of proximal amino acids appears to be even

weaker than that between water residence times and the local chemical environment (20), suggesting at best a weak correlation between local water densities and residence times. Indeed, the lack of a correlation was observed between water densities and water residence times at hydration sites around myoglobin in an MD simulation study of that protein (14). In this study, water was found to reside longer in clefts and notches of the protein surface irrespective of the local chemical environment, suggesting that local topological features of the protein surface may be the dominant factor influencing water dynamics on the protein surface.

Here we analyze protein hydration described by the dynamics of water association with the protein surface, and compare this description to a thermodynamic description based on the QC view of protein hydration. Our interest in the spatial distribution of protein hydration naturally leads us to consider the relationship between the spatial heterogeneity of water dynamics near the protein surface and the local topology and the chemical composition of the protein surface.

## THEORY

Our analysis of water dynamics is based on specific hydration sites near the protein surface, and uses a master equation to describe the chemical reaction dynamics for transitions between occupied and unoccupied states of these sites (21,22). The rate constants in this two-state model are related to quantities that can be extracted directly from MD simulations: the average time,  $\tau_1$ , a site is occupied by a water molecule, and the average time,  $\tau_0$ , that site remains unoccupied. The resulting probability of finding  $n$  occupied states out of a total number of  $N$  realizations is given by the binomial distribution,

$$P_n = \frac{N!}{n!(N-n)!} q^n (1-q)^{N-n}, \quad (2)$$

with  $q = \tau_1/(\tau_1 + \tau_0)$ . The mean of this distribution,  $\langle n \rangle$ , gives the local water density at a site of unit volume,

$$\rho \equiv \frac{\langle n \rangle}{N} = \frac{\tau_1}{\tau_1 + \tau_0}. \quad (3)$$

Defining the average cycle time,  $\tau_{\text{cyc}} = \tau_1 + \tau_0$ , and substituting for  $\eta$  in Eq. 1, provides the desired relationship between the local water density and the average occupancy and vacancy times at each hydration site,

$$\eta = \ln\left(\frac{\tau_1}{\tau_{\text{cyc}}}\right) - \ln\left(\frac{\tau_1}{\tau_{\text{cyc}}}\right)_{\text{bulk}}. \quad (4)$$

The ratio for bulk water is fixed by the density of water at the conditions of interest (Eq. 3). For a hydration site  $1 \text{ \AA}^3$  in volume and water at 300 K, we have

$$\eta = 3.4 + \ln\left(\frac{\tau_1}{\tau_{\text{cyc}}}\right). \quad (5)$$

The maximum value of  $\eta = 3.4$  is obtained when

$$\tau_1 \rightarrow \tau_{\text{cyc}} = \tau_1 + \tau_0,$$

and can be realized by two different kinetic pathways. A site can be occupied by a single water molecule for long periods of time relative to the time it is unoccupied—i.e.,  $\tau_1 \gg \tau_0$ —or it can be occupied for relatively short periods of time with high turnover—i.e.,  $\tau_0 \rightarrow 0$  at small  $\tau_1$ . These two pathways distinguish sites with bound waters from those sites that are simply highly accessible to water. Of course, a practical definition of kinetically bound water sites will depend on the extent to which  $\tau_1$  is taken to be greater than  $\tau_0$ . However, the least restrictive kinetic criterion for water association in either case is  $\tau_1 > \tau_0$ , which corresponds to  $\eta > 2.7$ .

We also note that the second moment of the binomial distribution, Eq. 2, relates local water density fluctuations to the fractional vacancy time at each hydration site,

$$\frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle} = \frac{\tau_0}{\tau_1 + \tau_0}. \quad (6)$$

Thus, the average occupancy and vacancy times of specific hydration sites in this kinetic description of preferential hydration includes information on both water densities and fluctuations in water densities locally near the protein surface.

## METHODS

Lysozyme (PDB ID: 1LYZ) (23) was solvated in a cubic box 62 Å on a side containing 7107 TIP3P water molecules (24) and staphylococcal nuclease (PDB ID: 1JOO) (25) was solvated in a cubic box of 82 Å containing 18,480 TIP3P water molecules. MD simulations of these proteins were carried out at 300 K and 1 bar using NAMD 2.6 (26) with the CHARMM27 force field (27). The average water residence time in the first hydration shell of proteins is on the order of a few picoseconds. We can consider a protein to be essentially rigid on this timescale; therefore, the protein atoms were held fixed throughout the simulation. Bulk water properties were determined from an independent MD simulation of 512 TIP3P water molecules at the same temperature and pressure.

Temperature was held constant in these simulations by applying Langevin dynamics to all heavy atoms using a damping coefficient of  $1 \text{ ps}^{-1}$ . Constant pressure was maintained using a Nosé-Hoover Langevin piston with a period of 200 fs and a decay of 100 fs. Periodic boundary conditions were imposed, and the particle-mesh Ewald method with a real-space cutoff of 12 Å was used in computing the electrostatic interactions. The same cutoff was applied to nonbonded nonelectrostatic interactions. The TIP3P water geometry was constrained by the SHAKE algorithm (28). The system was initially minimized for 20,000 steps and then equilibrated for 200 ps. Configurations were saved every 0.1 ps over a production run of 2 ns with a time step of 2 fs.

Specific hydration sites were defined as before (6) by constructing a network of grid points separated by 1 Å to fill the proximal volume within 3.5 Å of the heavy atoms on the protein surface. Water occupancy and vacancy times for each site were recorded over the course of the MD simulation, and arithmetic averages computed for both characteristic times. Different studies adopt different approaches to compute the average water occupancy or residence times. In most cases, the average residence time is estimated by fitting a series of exponential functions or a stretched exponential function to a time correlations function. The average residence times obtained from such fits are biased by the few infrequent long times a water molecule resides in the region of interest. By adopting the arithmetic averaging here, we obtain

a more realistic representation of the frequency of water exchanges at a hydration site.

Using these methods, the number of hydration sites with  $\eta > 2$  obtained from the MD simulation of lysozyme was 150 out of 8290 sites. This number is slightly higher than that reported previously—135 out of 7855 sites (6)—and is attributed to the different reference frames that were used in constructing the network of grid points around the protein in the two studies. The pattern of high-occupancy hydration sites ( $\eta > 2$ ) around the protein was found by visual inspection to be the essentially same, however, independent of the reference frame. For staphylococcal nuclease, 224 out of 12,936 hydration sites were found with  $\eta > 2$ .

## RESULTS AND DISCUSSION

Fig. 1 confirms the relationship given by Eq. 4 between the water density at a specific hydration site and the average site occupancy and cycle times. The results show that the local water density is only weakly correlated with the average occupancy time, although a slightly better correlation is obtained for the occupancy time compared to the cycle time. This observation is in agreement with the findings of the MD simulation study of myoglobin in which no correlation was found between local water densities and residence times at specific hydration sites around this protein (14).

Plots of the average occupancy time,  $\tau_1$ , versus the average vacancy time,  $\tau_0$ , for all hydration sites on the surface of lysozyme and for the high occupancy sites ( $\eta > 2.0$ ) on the surface of staphylococcal nuclease, are shown in Figs. 2 and 3, respectively. Only 23 of the 150 high occupancy sites on lysozyme and 25 of the 224 high occupancy sites on

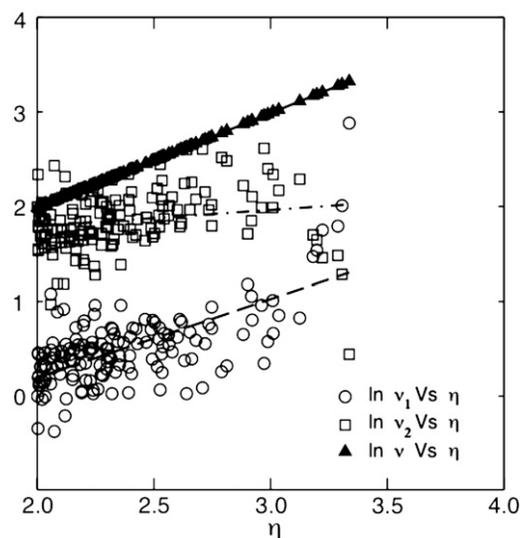


FIGURE 1 Natural logarithm of the ratio of average occupancy times ( $v_1 = \tau_1/\tau_{1,b}$ ), average cycle times ( $v_2 = \tau_{\text{cyc},b}/\tau_{\text{cyc}}$ ), and the combination of the two times ( $v = v_1 v_2$ ), defined in Eq. 4, as a function of  $\eta$  (Eq. 1) for hydration sites within 3.5 Å of the surface of lysozyme. The subscript  $b$  here denotes the characteristic time for bulk water. Only those sites corresponding to  $\eta > 2.0$  are shown. These characteristic times were obtained from MD simulations of lysozyme in TIP3P water at 300 K and 1 bar. Correlation coefficients for the linear fits of the data are 0.654, 0.176, and 0.999, respectively.

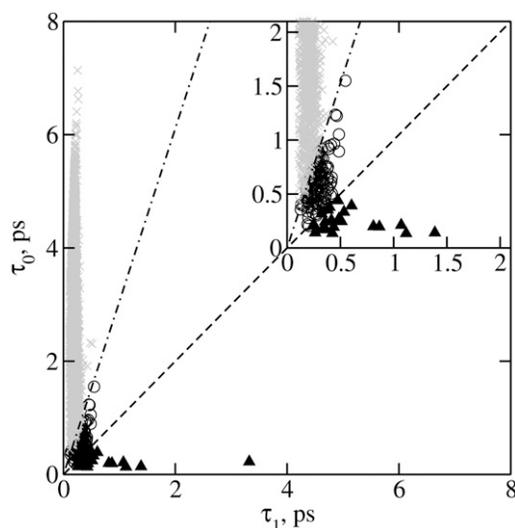


FIGURE 2 Average occupancy time,  $\tau_1$ , and average vacancy time,  $\tau_0$ , for each hydration site in a grid filling the proximal volume within 3.5 Å of the heavy atoms on the surface of lysozyme, obtained from MD simulation of this protein in TIP3P water at 300 K and 1 bar. (Solid triangles,  $\eta > 2.7$ ; open circles,  $2.7 > \eta > 2.0$ ; and shaded crosses,  $\eta < 2.0$ .) The dashed  $\tau_1 = \tau_0$  line separates sites that correspond to kinetically bound waters (solid triangles) from the sites corresponding to high water occupancies ( $\eta > 2.0$ ) (open circles), while the dash-dot line separates the high occupancy sites from all other hydration sites (shaded crosses). The inset shows only those data for  $\tau_1, \tau_0 \leq 2$  ps.

staphylococcal nuclease are found below the  $\tau_1 = \tau_0$  diagonal line, and as such, satisfy the kinetic criterion for bound water. These sites are also isolated from one another for the most part, as determined by computing clusters of sites that are within 4 Å of one another. For lysozyme, we obtained 18 clusters for the 23 sites with only a single cluster containing a maximum of three sites, while for staphylococcal nuclease, we obtained 23 clusters for the 25 sites with no cluster having more than two sites. The weak correlation between site water densities and residence times follows directly from the observation that a large fraction of the high occupancy sites are characterized by high turnover of water occupancies with occupancy times that span a relatively narrow range:  $\tau_1 \sim 0.3$ – $0.4$  ps, or roughly twice that for a site in bulk water (0.18 ps).

Different regimes of kinetic behavior on the  $\tau_1$ – $\tau_0$  plots of Figs. 2 and 3 can be related to local chemical and topological features of protein surfaces by considering the four idealized models of protein surfaces depicted in Fig. 4. These models were chosen collectively to include an overall composition of surface charges (Fig. 4, *b* and *d*) and surface roughness (Fig. 4, *c* and *d*) representative of small globular proteins, in general. The corresponding  $\tau_1$ – $\tau_0$  plot for the 10 highest occupancy sites (highest  $\eta$ -values) in each case is shown in Fig. 5. The overall range of  $\tau_1$ – $\tau_0$  values in this plot is strikingly similar to that obtained for lysozyme and staphylococcal nuclease (Fig. 2, *inset*, and Fig. 3), although the range of  $\tau_1$ – $\tau_0$  values for the individual models are much more restricted. The similarity in the overall range suggests

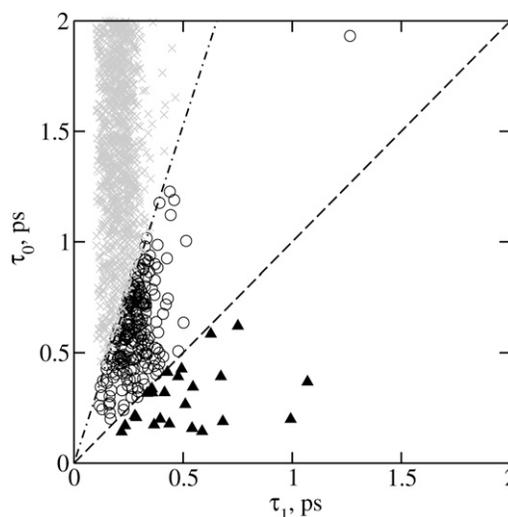


FIGURE 3 Average occupancy time,  $\tau_1$ , and average vacancy time,  $\tau_0$ , for each hydration site in a grid filling the proximal volume within 3.5 Å of the heavy atoms on the surface of staphylococcal nuclease, obtained from MD simulation of this protein in TIP3P water at 300 K and 1 bar. (Solid triangles,  $\eta > 2.7$ ; open circles,  $2.7 > \eta > 2.0$ ; and shaded crosses,  $\eta < 2.0$ .) The dashed  $\tau_1 = \tau_0$  line separates sites that correspond to kinetically bound waters (solid triangles) from the sites corresponding to high water occupancies ( $\eta > 2.0$ ) (open circles), while the dash-dot line separates the high occupancy sites from all other hydration sites (shaded crosses).

that these four idealized protein surfaces collectively encompass the different kinetic regimes of hydration behavior observed for lysozyme and staphylococcal nuclease.

For the smooth dipolar surface in Fig. 4 *a* (no surface roughness), local heterogeneities exist only in the equatorial region separating the hemispheres of neutral and negatively charged atoms. It is striking that the highest occupancy sites are found only at this interface, rather than within the hemisphere of charged atoms. The same behavior is observed when the two hemispheres are neutral and positively charged or positively and negatively charged (not shown), indicating that the local water density is sensitive to local heterogeneities in surface charge, rather than the magnitude of the surface charge density. However, none of the high occupancy sites correspond to kinetically bound waters; i.e.,  $\tau_0 \gg \tau_1 \sim 0.2$ – $0.3$  ps (Fig. 5). These sites also fall outside the range for high occupancy sites ( $\eta > 2.0$ ) for lysozyme and staphylococcal nuclease due to the high circumferential mobility of water molecules in the equatorial region.

When the charges are dispersed over the surface (Fig. 4 *b*), the average vacancy time is reduced significantly without much impact on the average occupancy times for the highest occupancy sites (Fig. 5). Nonetheless,  $\tau_1 < \tau_0$  for all these sites, indicating that heterogeneities in the surface charge alone do not produce sites corresponding to kinetically bound waters. Virtually identical kinetic behavior is observed when local heterogeneities in surface roughness are introduced (Fig. 4 *c*). However, the highest occupancy sites are now found in clefts and grooves on the rough surface.

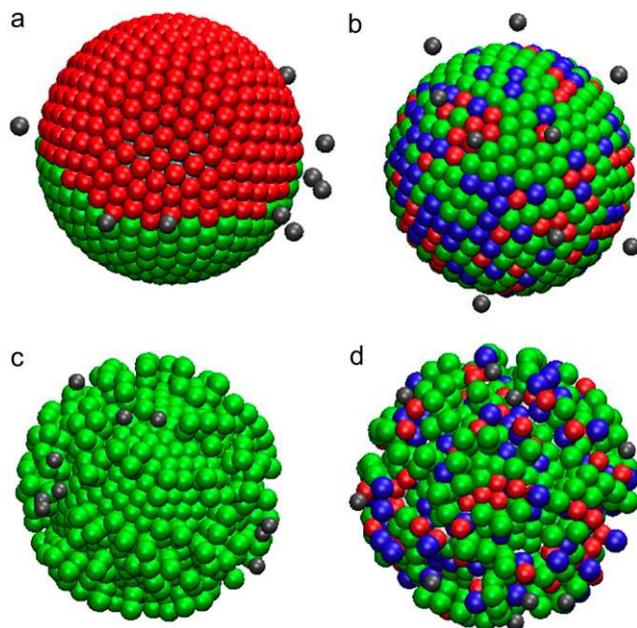


FIGURE 4 Ten highest occupancy sites on the surfaces of spheres with the following chemical and/or topological features: (a) smooth dipolar surface; (b) smooth random surface; (c) rough apolar surface; and (d) rough random surface. The four spheres have a radius of 10 Å, and are filled with neutral atoms in the interior. Atoms on the surface are: neutral (green); negatively charged (−0.51 e, red); or positively charged (+0.51 e, blue). The models are constructed to be electrostatically neutral by placing an atom of opposite charge below the surface for every charged atom on the surface. The gray spheres represent water oxygens at these high-occupancy sites. Lennard-Jones parameters for all atoms are  $\sigma = 3.4$  Å and  $\epsilon = -0.12$  kcal/mol.

It is only when the surface is rough and the surface charge is dispersed (Fig. 4 d) that we find the highest occupancy sites corresponding to  $\tau_1 > \tau_0$  (Fig. 5). In this case,  $\tau_0 \sim 0.2$ –0.3 ps and  $\tau_1$  spans the same range of values obtained for lysozyme and staphylococcal nuclease (Fig. 2, inset, and Fig. 3). We conclude, therefore, that local heterogeneities in both the surface charge and roughness are necessary to obtain specific hydration sites on the protein surface that correspond to kinetically bound waters.

We also calculated osmotic second virial coefficients for protein-protein interactions using different characterizations of preferential hydration defined by the kinetically bound water sites alone or by all the high occupancy sites. The results are shown in Table 1. The interaction part of the second virial coefficient,  $\beta_{22}$ , calculated here accounts for nonideal contributions due to protein-protein interactions, and is obtained by subtracting the Donnan contribution. Details of this calculation using a molecular thermodynamic model of protein solutions are given elsewhere (6,29).

In this previous work, we showed that a number of highly complementary protein-protein contact configurations are eliminated by including a spatially heterogeneous distribution of explicit water molecules strongly associated with the protein surface at specific hydration sites. Short-ranged protein-protein interactions thus become less favorable. This

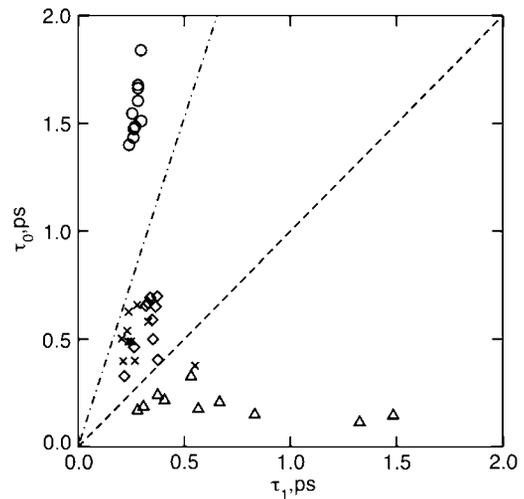


FIGURE 5 Average occupancy time,  $\tau_1$ , and average vacancy time,  $\tau_0$ , for the 10 highest occupancy sites on the surface of four idealized (spherical) models of globular proteins obtained from MD simulations in TIP3P water at 300 K. (Circles) Smooth dipolar surface; (diamonds) smooth random surface; (crosses) rough apolar surface; and (triangles) rough random surface. The dashed line and the dash-dot line are defined in Figs. 2 and 3.

effect is seen in Table 1 by comparing  $\beta_{22}$  calculated for a continuum solvent (no explicit water molecules) to that calculated with all high occupancy hydration sites taken into account. For both lysozyme and staphylococcal nuclease,  $\beta_{22}$  becomes less negative—less favorable protein-protein interactions—when explicit waters at the high occupancy hydration sites are taken into account. In contrast, the contribution from the kinetically bound water sites alone is negligible; the calculated  $\beta_{22}$  is essentially the same as that for the continuum solvent. The observation reflects our finding that explicit water molecules at the kinetically bound hydration sites are located in regions of high surface roughness and charge heterogeneity, which tend to be buried; thus, they have less impact on reducing the complementarity of protein-protein contact configurations compared to explicit water molecules located at the other high occupancy hydration sites, which are more solvent-accessible.

TABLE 1 Interaction part of second virial coefficient,  $\beta_{22}$ , as a function of hydration conditions for lysozyme at pH 7 and ionic strength 0.007 mol/L and staphylococcal nuclease at pH 6.5 and ionic strength 0.01 mol/L

Hydration condition	No. of sites	$\beta_{22} \times 10^4$ mol ml/g <sup>2</sup>
<i>Lysozyme</i>		
Continuum solvent	0	−87.44
Kinetically bound waters sites ( $\eta > 2.7$ )	23	−86.81
High occupancy sites ( $\eta > 2.0$ )	150	−82.92
<i>Staphylococcal nuclease</i>		
Continuum solvent	0	−80.58
Kinetically bound waters sites ( $\eta > 2.7$ )	25	−80.87
High occupancy sites ( $\eta > 2.0$ )	242	−64.54

## CONCLUSIONS

Preferential hydration of protein surfaces was analyzed by computing average water residence times and vacancy times at specific hydration sites on the protein surface. This analysis revealed two distinct kinetic regimes for those hydration sites defined thermodynamically to have high local water densities: long residence times relative to vacancy times for a single water molecule, corresponding to kinetically bound water molecules, and short residence times with high turnover involving multiple water molecules. Those sites corresponding to kinetically bound water molecules comprise only a small fraction of the total number of high occupancy sites, and are correlated with local heterogeneities in both surface charge and roughness. Moreover, these sites have little impact on calculated osmotic second virial coefficients for protein-protein interactions. The impact of preferential hydration on these weak protein-protein interactions is due primarily to the preferential hydration of sites characterized by high occupancy and high turnover—i.e., those sites on the protein surface that are accessible to water.

In deriving a relationship between kinetic and thermodynamic views of preferential hydration (Eq. 4), we found that the thermodynamic characterization in terms of the local water density at specific hydration sites and the kinetic characterization in terms of water occupancy and vacancy times at these sites are not equally informative. Specifically, while it is possible to obtain the local water density from a knowledge of site occupancy and vacancy times, it is not possible to derive the average occupancy and vacancy times knowing just the local densities for the hydration sites. Indeed, the weak correlation that we found between local water densities and average residence times follows directly from this analysis and the observation that most high occupancy sites on the two protein surfaces we studied have a narrower range of occupancy times compared to the range of vacancy times.

The thermodynamic and kinetic perspectives of preferential hydration would be equally informative if just the ratio of water occupancy and vacancy times at each site was sufficient for an accurate description of hydration. We find, however, two distinct regimes of kinetic behavior for lysozyme and staphylococcal nuclease, and collectively, for the four spherical models of a protein surface—one characterized by specific hydration sites with high turnover in occupancies ( $\tau_1 \sim \text{constant} < \tau_0$ ), and the other characterized by specific hydration sites with strongly associated or kinetically bound waters ( $\tau_0 \sim \text{constant} < \tau_1$ ). We conclude, therefore, that a more complete description of the preferential hydration of protein surfaces is achieved when occupancy and vacancy times are taken to be independent of one another.

Of course, the corollary is that an additional parameter in the thermodynamic analysis is required to obtain an equivalent description of preferential hydration. Recognizing that the average site occupancy and vacancy times in the kinetic model characterize both water densities (Eq. 3) and fluctua-

tions in water densities (Eq. 6) locally near protein surfaces, we submit that the logical, although not necessarily practical choice for an additional thermodynamic parameter is the water-water pair distance distribution function. Extracting this parameter from MD simulations in the heterogeneous environment of the protein-water interface with the spatial resolution demonstrated here for  $\tau_0$  and  $\tau_1$  is a formidable, if not impossible task. An advantage of the kinetic model is that water occupancy and vacancy times characteristic of specific hydration locally at sites on the protein surface are indeed readily accessible from MD simulations.

Finally, the kinetic analysis of preferential hydration presented here does not take into account any coupling of water dynamics to the protein dynamics, since the protein was held fixed in our MD simulations. This coupling undoubtedly would be important in an analysis of water dynamics near protein surfaces. In the context of our focus on protein hydration, though, we note that the water occupancy times corresponding to the high occupancy/high accessibility sites are all less than 0.5 ps, which is more than an order-of-magnitude smaller than the characteristic time for side-chain rotations of the amino acids on a protein surface (30). We conclude, therefore, that our kinetic characterization of preferential hydration is unaffected by protein dynamics on these longer timescales, other than introducing the need to consider an ensemble of protein configurations that would be accessible on timescales for the protein-protein interactions of interest.

Financial support from the National Science Foundation (grant No. BES-0555281) and the Department of Energy (grant No. DE-FG02-04ER25626) is gratefully acknowledged.

## REFERENCES

1. Neal, B. L., D. Asthagiri, and A. M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* 75:2469–2477.
2. Asthagiri, D., B. L. Neal, and A. M. Lenhoff. 1999. Calculation of short range interactions between proteins. *Biophys. Chem.* 78:219–231.
3. Paulaitis, M. E., and L. R. Pratt. 2002. Hydration theory for molecular biophysics. *Adv. Protein Chem.* 62:283–310.
4. Pratt, L. R., and R. A. LaViolette. 1998. Quasi-chemical theories of associated liquids. *Mol. Phys.* 94:909–915.
5. Pratt, L. R., and S. B. Rempe. 1999. Quasi-chemical theory and implicit solvent models for simulations. In *Simulation and Theory of Electrostatic Interactions in Solution. Computational Chemistry, Biophysics, and Aqueous Solutions.* AIP Conference Proceedings. L. R. Pratt, and G. Hummer, editors. AIP Press, Melville, NY.
6. Paliwal, A., D. Asthagiri, D. Abras, A. M. Lenhoff, and M. E. Paulaitis. 2005. Light-scattering studies of protein solutions: role of hydration in weak protein-protein interactions. *Biophys. J.* 89:1564–1573.
7. Babu, S., and C. Lim. 1999. Theory of ionic hydration: insights from molecular dynamics simulations and experiment. *J. Phys. Chem. B.* 103:7958–7968.
8. Asthagiri, D., L. R. Pratt, M. E. Paulaitis, and S. B. Rempe. 2004. Hydration structure and free energy of biomolecularly specific aqueous di-cations, including  $\text{Zn}^{2+}$  and first-transition-row metals. *J. Am. Chem. Soc.* 126:1285–1289.
9. Denisov, V. P., and B. Halle. 1996. Protein hydration dynamics in aqueous solution. *Faraday Discuss.* 103:244–274.

10. Otting, G., E. Liepinsh, and K. Wüthrich. 1991. Protein hydration in aqueous solution. *Science*. 254:974–980.
11. Dellerue, S., A. J. Petrescu, J. C. Smith, and M. C. Bellissent-Funel. 2001. Radially softening diffusive motions in a globular protein. *Biophys. J.* 81:1666–1676.
12. Pal, S. K., J. Peon, and A. H. Zewail. 2002. Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. USA*. 99:1763–1768.
13. Rocchi, C., A. R. Bizzarri, and S. Cannistraro. 1997. Water residence times around copper plastocyanin: a molecular dynamics simulation approach. *Chem. Phys.* 214:261–276.
14. Makarov, V. A., B. K. Andrews, P. E. Smith, and B. M. Pettitt. 2000. Residence times of water molecules in the hydration sites of myoglobin. *Biophys. J.* 79:2966–2974.
15. Garcia, A. E., and L. Stillier. 1993. Computation of the mean residence time of water in the hydration shells of biomolecules. *J. Comput. Chem.* 14:1396–1406.
16. Schröder, C., T. Rudas, S. Boresch, and O. Steinhauser. 2006. Simulation studies of the protein water interface. I. Properties at the molecular resolution. *J. Chem. Phys.* 124:1–18.
17. Brunne, R. M., E. Liepinsh, G. Otting, K. Wüthrich, and W. F. van Gunsteren. 1993. Hydration of proteins: a comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J. Mol. Biol.* 231:1040–1048.
18. Bandyopadhyay, S., S. Chakraborty, and B. Bagchi. 2005. Secondary structure sensitivity of hydrogen bond lifetime dynamics in the protein hydration layer. *J. Am. Chem. Soc.* 127:16660–16667.
19. Bandyopadhyay, S., S. Chakraborty, S. Balasubramanian, and B. Bagchi. 2005. Sensitivity of polar solvation dynamics to the secondary structures of aqueous proteins and the role of surface exposure of the probe. *J. Am. Chem. Soc.* 127:4071–4075.
20. Beck, D. A. C., D. O. V. Alonso, and V. Daggett. 2003. A microscopic view of peptide and protein solvation. *Biophys. Chem.* 100:221–237.
21. Zwanzig, R. 2001. Nonequilibrium Statistical Mechanics. Oxford University Press, Oxford, UK.
22. Gillespie, D. T. 2002. The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction. *J. Phys. Chem. A*. 106:5063–5071.
23. Diamond, R., D. C. Phillips, C. C. F. Blake, and A. C. T. North. 1974. Real space refinement of the structure of hen egg white lysozyme. *J. Mol. Biol.* 82:371–391.
24. Jorgensen, W., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
25. Wang, J., D. M. Trucks, F. Abildgaard, Z. Dzakula, Z. Zolnai, and J. Markley. 1997. Solution structures of staphylococcal nuclease from multidimensional, multinuclear NMR: nuclease-H124L and its ternary complex with  $\text{Ca}^{2+}$  and thymidine-3',5'-bisphosphate. *J. Biomol. NMR*. 10:143–164.
26. Kalé, L., R. Skeel, M. Bhandarkar, R. Brunner, N. G. N. Kraweta, J. Phillips, A. Shinzaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.
27. MacKerell, A. D., N. Banavali, and N. Foloppe. 2000. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*. 56:257–265.
28. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23:327–341.
29. Asthagiri, D., A. Paliwal, D. Abras, A. M. Lenhoff, and M. E. Paulaitis. 2005. A Consistent experimental and modeling approach to light-scattering studies of protein-protein interactions in solution. *Biophys. J.* 88:3300–3309.
30. McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, UK.