PROTEINS: Structure, Function, and Genetics 38:261–272 (2000)

Water Penetration and Escape in Proteins

Angel E. García* and Gerhard Hummer
Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico

ABSTRACT **The kinetics of water penetration and escape in cytochrome c (cyt c) is studied by molecular dynamics (MD) simulations at various temperatures. Water molecules that penetrate the protein interior during the course of an MD simulation are identified by monitoring the number of water molecules in the first coordination shell (within 3.5 Å) of each water molecule in the system. Water molecules in the interior of cyt c have 0–3 water molecules in their first hydration shell and this coordination number persists for extended periods of time. At T = 300 K we identify over 200 events in which water molecules penetrate the protein and reside inside for at least 5 picoseconds (ps) within a 1.5 nanoseconds (ns) time period. Twenty-seven (27) water molecules reside for at least 300 ps, 17 water molecules reside in the protein interior for times longer than 500 ps, and two interior water molecules do not escape; at T = 360 K one water molecule does not escape; at 430 K all water molecules exchange. Some of the internal water molecules show mean square displacements (MSD) of 1 Å^2 characteristic of structural waters. Others show MSD as large as 12 Å^2, suggesting that some of these water molecules occupy transient cavities and diffuse extensively within the protein. Motions of protein-bound water molecules are rotationally hindered, but show large librations. Analysis of the kinetics of water escape in terms of a survival time correlation function shows a power law behavior in time that can be interpreted in terms of a broad distribution of energy barriers, relative to κ_BT, for water exchange. At T = 300 K estimates of the roughness of the activation energy distribution is 4–10 kJ/mol (2–4 κ_BT). Activation enthalpies for water escape are 6–23 kJ/mol. The difference in activation entropies between fast exchanging (0.01 ns) and slow exchanging (0.1–1 ns) water molecules is −27 J/K/mol. Dunitz (Science 1997;264:670.) has estimated the maximum entropy loss of a water molecule due to binding to be 28 J/K/mol. Therefore, our results suggest that the entropy of interior water molecules is similar to entropy of bulk water. Proteins 2000;38:261–272.**
Published 2000 Wiley-Liss, Inc.†

Key words: hydration; water pentration; protein dynamics; cytochrome c

INTRODUCTION

The interaction of water molecules with proteins and DNA plays an important role in biomolecular structure, dynamics, and function. Binding of water molecules at specific sites on the protein surface or in its interior have a determinant influence on the specificity and function of proteins and protein-DNA complexes.[2–4] Water penetration is believed to be a major factor in protein denaturation, in particular, for pressure induced unfolding.[5] The hydration structure of biomolecules has been extensively reviewed in the literature.[6–10] The structure and kinetics of structural water has been studied by NMR,[8,11,12] and from water ^{17}O and ^{2}H nuclear magnetic relaxation dispersion (NMRD).[9,13–15] In recent experiments, Denisov et al. determined the mean residence times of water molecules on the surface and interior of biomolecules and biomolecular complexes.[9,13–15] The time scale at which water penetrates or exchanges with other solvent molecules (at T = 300 K, P = 1 atm) has been observed to be in the range of 10–50 picoseconds (ps) for waters bound in surface crevices,[7] to 0.1–1 nanoseconds (ns) for strongly bound waters,[8,11] and ns to milliseconds (ms) for interior waters (e.g., structural water molecules in proteins).[13,14] Denisov et al.[16] have used ^{17}O and ^{2}H magnetic relaxation dispersion to measure the exchange rate of water buried in BPTI at temperatures varying from 4–80°C. The strong temperature dependence of this rate has been interpreted in terms of large scale conformational fluctuations on an energy landscape with a statistical roughness of 10 kJ/mol.[16]

Ernst et al.[17] have found that hydrophobic cavities in human interleukin-1β contained two to four water molecules that reside within the protein for times longer than 1 ns. These water molecules were not observed in high resolution crystal structures, since water with large mean square displacements (MSD) fluctuations (>1 Å) makes negligible contributions to high resolution X-ray diffraction spectra.[18] However, careful analysis of the low resolution diffraction data by Yu et al.[18] shows that disordered water molecules observed by Ernst et al.[17] are indeed present. Ernst et al. have suggested that hydrophobic cavities commonly observed in protein structures may be filled with disordered water. Woodward and co-workers[19,20] have proposed a *penetration model* for hydrogen exchange in which small fluctuations in the protein pro-

Grant sponsor: United States Department of Energy, Los Alamos National Laboratory (LDRD); Grant number: 98006.

G. Hummer's present address is Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Building 5, Room 132, Bethesda, MD 20852-0520.

*Correspondence to: Angel E. García, Theoretical Biology and Biophysics Group, T10, MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545. E-mail: angel@+10.lanl.gov

Received 29 July 1999; Accepted 23 September 1999

vide transient access of solvent to buried NHs. The penetration model contributes to hydrogen exchange in the protein folded state, in addition to the global unfolding mechanism of Bai et al.[21] Disordered water may also play an important role in protein function and dynamics, and may occupy transiently opened cavities inside proteins.[22]

Structural water in the interior of both oxidized and reduced horse-heart cyt c in solution has been studied by nuclear magnetic resonance (NMR)[12] and X-ray crystallography[23] show deeply buried water molecules that are apparently conserved across the cyt c family. In NMR studies by Qi et al.,[12] six water molecules with residence times greater than 300 ps were located in ferrocytochrome c and five in ferricytochrome c. Two of these water molecules are located near the Heme. One of these water molecules undergoes a large change in position with a change of oxidation state, suggesting that buried structural waters in cyt c may have a role in the solvent reorganization energy associated with electron transfer.[24]

In this work we describe the kinetics of water penetration and escape from the protein surface, crevices, and the protein interior. For this purpose we study nanosecond (ns) molecular dynamics (MD) simulations of oxidized cyt c at three temperatures: 300 K, 360 K, and 430 K. High temperature simulations are used to promote relatively large-scale fluctuations to occur in a time scale amenable to computer simulations. At 300 K all but two of the interior water molecules exchange with water molecules in the solvent. At 360 K all but one of the interior water molecules exchange. At T = 430 K the protein remains folded during the ns time scale of the simulations, and all interior water molecules exchange. MD simulations have shown that cyt c unfolds, in the ns time scale, at T = 550 K.[25] During the MD simulations we observe a few hundred events of water penetration in the ns time scale. Water molecules in the interior of a protein will have 0–3 water molecules in their first hydration shell, and this coordination number persists for extended periods of time. In contrast, in the absence of a protein (i.e., in bulk water), the probability distribution of the water coordination number has a maximum for $N_c = 5$. By combining the instantaneous coordination number of all water molecules in the system, and requiring that this coordination number persists for a period of over 1–2 ps we can distinguish water molecules in the interior and in the crevices (referred to as *bound* waters) from bulk. The choices of a threshold coordination number ($N_c = 3$) and of the minimum residence time ($t_{wait} = 2$ ps), are optimized based on our simulation data. The rates at which water molecules escape from a protein cavity are coupled to large-scale fluctuations that occur as rare events in protein dynamics.[25]

The identification of protein-bound water molecules allows us to study their dynamics in detail. Here we describe the rotational and translational dynamics of bound water molecules, as well as the kinetics of water binding and escape.

MODEL AND METHODS

Description of the System and Simulations

Details of the MD simulations have been described previously.[25] Horse heart cytochrome c in its oxidized (Fe(III)) state is simulated in the presence of water and excess salt. We use the structure of horse heart cyt c in solution and in its reduced state (pdb code: 1frc)[26] as the initial configuration of oxidized cyt c near its folded state. A covalent bond between Met-80 and the heme iron is maintained in all the simulations. His-33 is modeled in its protonated state;[27,28] His-18 and His-26 are modeled as neutral with hydrogens at the δ position. Six internal water molecules found in the solution NMR structure are included in the initial configuration.[26] The total charge of the protein in the oxidized state is +9e. We also include 20 Cl^- and 11 Na^+ ions in the system. Ions are located at random within the aqueous solution. Water molecules are added around the protein to fill a cubic box with ≈57 Å on the side. The walls of this box are at least 11 Å away from any protein atom to allow the protein system to sample non-compact (partially unfolded) conformations at high temperature. The resulting system contains 5,760 water molecules, 31 ions, 1,746 atoms in the protein and a total of 19,057 atoms. The resulting concentrations for the system are [Cl^-] = 0.18 M, [Na^+] = 0.10 M, and [cyt c] = 0.8 mM.

We use the all-atom force field of Cornell et al.[29] Electrostatic interactions are modeled with the particle mesh Ewald (PME)[30] algorithm implemented in Amber 4.1.[31] A cut-off of 9 Å is used for non-bonded and direct electrostatic interactions. The Fourier space part of the Ewald potential is calculated on a cubic grid of 48 points on the side and interpolated over all space with a cubic spline. For the T = 300 K simulation we simulate the system at constant (N,T,P) (P = 1 atm). For T = 360 K and 430 K we scale the density proportionally to the density of water at the liquid-vapor coexistence curve for the given temperature.[32] These calculations are done at constant (N,T,V). The solvated system is subjected to 1,000 steepest-descent energy minimization cycles. The resulting configuration is simulated at T = 100 K for 100 picoseconds (ps). The temperature is then gradually increased to 300 K over a 50 ps time interval. This simulation is extended to 2.5 ns. The configuration at t = 1 ns of the T = 300 K simulation is used to start the T = 360 K calculation. At this time, the volume of the box is increased by 1.02%. The system is slowly heated to T = 360 K over 100 ps and further equilibrated for another 200 ps. This simulation is extended to 1.75 ns. The configuration at t = 100 ps of the T = 360 K simulation is used to continue heating the system to 430 K over a period of 100 ps, followed by a further equilibration for another 100 ps. The volume of the system is increased by 3.1% over the volume at T = 300 K. The simulation is extended to 1.6 ns at T = 430 K. The last 1.5 ns of each trajectory are used for analyzing the kinetics of structural water. Configurations are saved for further analysis at a rate of two per ps. The instantaneous coordination number of each water molecule is obtained for every saved configuration by counting all water molecules within 3.5 Å from the water oxygen atom, which

defines the first hydration shell of water. The distribution of coordination numbers in the protein-solvent system will be compared with similar plots for bulk water.

Water Mean Residence Time

We study the time relaxation of bound water molecules by means of a survival time correlation function, $S(t)$.[33–35] This function is written in terms of a binary function, $p_j(t, t + t'; t_0)$ that adopts a value of *one* if the coordination number of a water molecule labeled j has been less than or equal to a threshold value ($N_c = 3$) from time t to $t + t'$, without escaping (except for a short interval of time, t_0) during this time interval, and zero otherwise. t_0 is taken to be 0.5 ps (i.e., one configuration) to smooth out fast oscillations that are not relevant to our analysis. S(t) is defined by

$$S(t) = \sum_{j=1}^{N_{\text{water}}} \frac{1}{t_{\text{run}} - t} \sum_{t'=0}^{t_{\text{run}} - t} p_j(t', t + t'; t_0), \qquad (1)$$

where N_{water} is the number of water molecules in the system. This function is a time correlation function. $S(0)$ equals the average number of structural water molecules either on the surface or the interior of the protein (i.e., with $N_c \leq 3$), and $S(t)$ gives the average number of water molecules that still remain "bound" after a time t. Averages are calculated over configurations sampled during MD simulations. $S(t)$ shows non-exponential (i.e., power law) time behavior. The $S(t)$ curves are analyzed in terms of water molecules escaping from the hydration shells and cavities as a kinetic process described by a distribution of barrier heights.[36,37] This distribution of barrier heights results from the inhomogeneities of the protein surface and its interior due to amino acid sequence, and from the ensemble of states that the protein samples during the simulations.[36,38,39] We use the Maximum Entropy Method (MEM) described by Steinbach et al.[37] to decompose the calculated survival time correlation functions in terms of a rate distribution distribution, $f(\log \tau)$, such that

$$S(t) = \sum_{i=1}^{N} \Delta(\log \tau) f(\log \tau_i)\, exp(-t/10^{\log \tau_i}), \qquad (2)$$

where τ is a logarithmically gridded rate, N is the total number of logarithmically gridded points of τ, $\Delta(\log \tau)$ is the grid size, and $f(\log \tau)$ is the distribution of rates. This transformation is equivalent to a discrete Laplace transformation of the survival time correlation function.

Dynamics of Protein-Bound Water

To obtain a simple measure of the rotational freedom and diffusional motion of water molecules in the interior of the protein we determine the water dipole moment unit vector, $\hat{m}$, and its autocorrelation function,

$$c(t) = \langle \hat{m}(0) \cdot \hat{m}(t) \rangle, \qquad (3)$$

for selected water molecules. We calculate the rotational and translational motions of protein-bound water molecule relative to an average protein conformation frame. The overall rotations and translations of the system are minimized by aligning the protein non-hydrogen atoms, in the least-square sense. We monitor the time dependence of the unit vectors along the water dipole moment (i.e., the water C2 symmetry direction), $\hat{m}$, a unit vector, $\hat{\omega}$, parallel to H-H and orthogonal to $\hat{m}$, and a unit vector, $\hat{\omega}_{\perp\ \text{perp;}}$, perpendicular to the water plane. Rotations around $\hat{m}$, , and $\hat{\omega}_{\perp\ \text{perp;}}$ measure a water molecule twisting, wagging, and rocking, respectively.[40]

RESULTS AND DISCUSSION

Water Coordination Number Distribution

On average, water molecules have 4–6 water molecules in their first hydration shell (four molecules arranged in a tetrahedron and one or two waters filling interstitial sites on the tetrahedron's faces). However, the instantaneous coordination number fluctuates from 0–12 per water molecule. In bulk water the probability distribution of the coordination number has a maximum for $N_c = 5$. The probability of having zero to three water molecules in the first hydration shell is very small ($1{:}10^7$ *to* $1{:}10^3$). In bulk water, water molecules without coordinated water molecules in the first hydration shell occur as a high energy, short lived, fluctuation event. The probability of such events is negligible if we add the additional condition that the coordination number remains below a threshold value for a few consecutive time periods sampled ($t_{wait} \sim 2.0$ ps). The probability that a bulk water molecule has a small coordination number for an extended period of time (2–5 ps) is thus negligible.

Water molecules solvating a protein have drastically different coordination properties. Interior water molecules have zero to three water molecules in their first hydration shell, and this coordination number persists for extended periods of time. (An exception to this will be proteins containing large clusters of interior water molecules for which these criteria do not apply). Figure 1 shows the distribution of the coordination number for water in a cyt c solution at three temperatures (300 K, 360 K, and 430 K), and in bulk water. The distributions of coordination numbers show a maximum for $N_c = 5$, with fluctuations ranging from 0 to 12 water molecules. For $N_c \geq 4$ all probability distributions of the coordination numbers are very similar. However, for $N_c \leq 3$ the distributions for the protein-solvent systems are quite different than observed for bulk water. This suggests that interior or surface water molecules have $N_c \leq 3$. The distributions for the protein-solvent systems change with temperature, with the number of *bound* waters increasing with temperature. However, the main features observed at 300 K remain the same.

Coordination Number Time Series

To identify water molecules that penetrate the protein, we monitor the time history of the coordination number of each water molecule in the system. From these time series we can identify water penetration (changes in coordination number from $N_c \geq 5$ to $N_c \leq 3$) and water escape events (changes in coordination number from $N_c \leq 3$ to $N_c \geq 5$).

Figure 2 shows the time series of the number of water molecules in the system, with coordination numbers between 0 and 3 water molecules in their first coordination shell, when averaged over a moving 5 ps time window, during the 1.5 nanoseconds MD simulation at 300 K. On average, we find that at each time instant 2–3 water molecules have a coordination number of zero, 6–10 have a coordination number of at most one; 21–29 have a coordination number of at most two; and 69–83 have a coordination number of at most three water molecules. The steady values in the number of water molecules with coordination number below or equal to 3 indicate that the water exchange rate is near a steady state in the ns time scale sampled by our simulations. After the 1.0 ns equilibration from the initial NMR structure, the first configuration of the production part of our MD simulation at 300 K contains six interior water molecules (not connected to other water molecules leading to the bulk solvent) and 19 water molecules in protein surface crevices which are connected directly or indirectly to bulk solvent molecules, with coordination number less or equal to two persisting for over 5 ps. The position of the interior water molecules in the protein are shown in Figure 3, where the six interior water molecules are shown. Four of the six internal water molecules form a single file of hydrogen-bonded molecules, with one end near the Ω_1 loop, and the other end near the heme propionate. Four of these six water molecules were identified by NMR (W1, W2, W5, and W6). [We label water molecules with indices 1–5,760 and refer to each water molecule by this index using the notation W1, W225, etc. W1–W6 are the water molecules identified by NMR in reduced cyt c.] W2 is coordinated to the side chains of Tyr 67, Asn 52, Thr 78, and a heme propionate and it has been linked to the setting of the redox potential of the heme.[12,24] During the 1 ns equilibration stage of the MD simulations, two additional water molecules entered the protein and formed part of the single file water structure. This single file water chain has not been observed by either NMR or X-ray crystallography. However, the number of water molecules assigned by Qi et al. is the minimum number of water molecules required to satisfy the observed NOE signals.[12] Given the proximity and high flexibility of water molecules forming this chain, individual water molecules might not be distinguished experimentally.

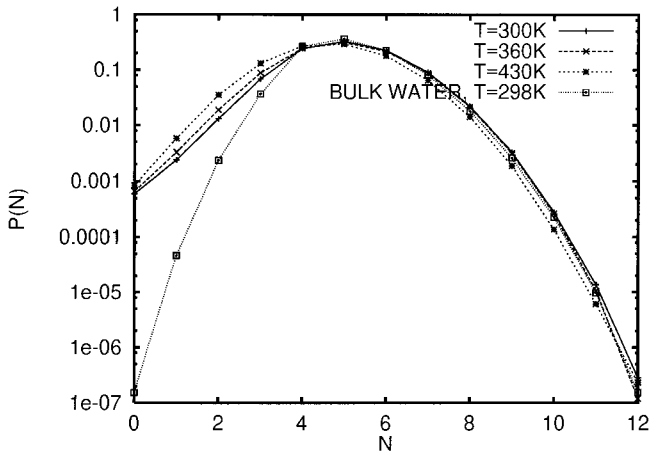

Fig. 1. Probability distribution of the coordination number for waters in a cyt c-water solution at three temperatures (300 K, 360 K, and 430 K), and in bulk water. Differences in these distributions for $N_c \leq 3$ in the presence and absence of the protein are due to water coordination to the protein.

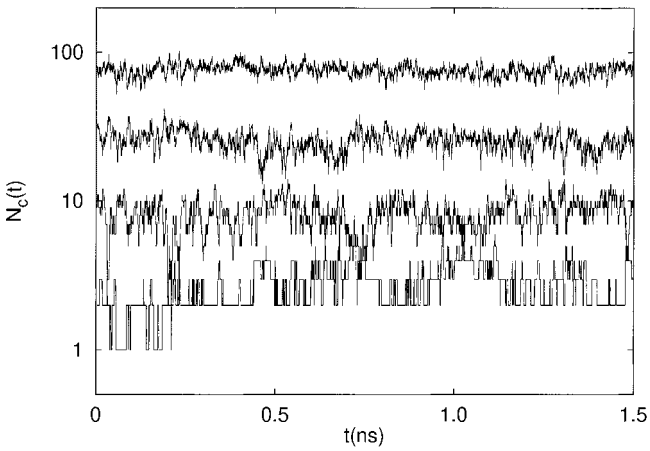

Fig. 2. Time series of the number of water molecules (averaged over a moving 5 ps time window) with coordination number less than or equal to 0, 1, 2, and 3, from bottom to top, respectively, sampled during the 1.5 ns production stage of the T = 300 K MD simulation. The average number of water molecules with up to 0–3 coordinated water molecules are 3 ± 1, 8 ± 2, 25 ± 4, and 75 ± 6, respectively.

Water Residence Times

The residence times of water molecules in the protein interior (or surface crevices) are calculated from the time series of the coordination number of individual water molecules using the survival time correlation function defined in Eq. (1). Figure 4 shows the time series of the instantaneous coordination number for W1 and W2, which do not exchange in the 300 K simulation, but exchange at higher temperatures. Figure 5 shows the time series of the instantaneous coordination number for water molecules that have residence times inside the protein in the 150–1,300-ps range. Analysis of MD simulations on cyt c at 300 K identifies nearly 200 events of water penetration into the protein that last for at least 5 ps, within the 1.5 ns production stage of the MD simulation.

Figure 6 shows the histograms of residence times for water molecules that coordinate to the interior and crevices of cyt c at T = 300 K, 360 K, and 430 K. The majority of the water molecules with low coordination number have residence times in the 1–10 ps time scale. Most of these water molecules bind to the surface of the protein. MD studies of myoglobin[41] and crambin[35] have observed water molecules coordinating polar and non-polar side chains with residence times in the 10-ps time scale. A significant number of water molecules have residence times in the 100–500-ps time scale, previously observed for water molecules coordinating to charged groups in crambin.[35] Overall, at T = 300 K, 17 water molecules are found to

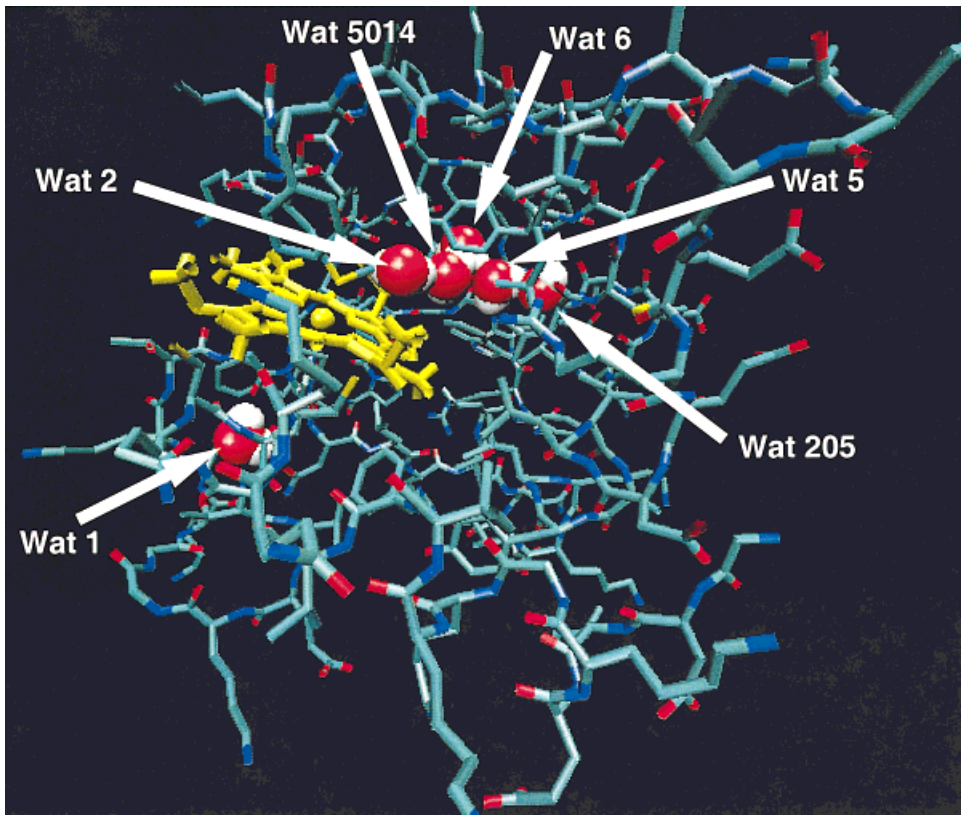

Fig. 3. Positions of six water molecules in the interior of cyt c after a 1.0 ns equilibration time, starting from the NMR structure. Four of the six water molecule form a single-file chain of water molecules extending from the Heme propionate O1A oxygen (W2) to the carboxy oxygen of Thr 63 (W205). W5014 and W5 make hydrogen bonds only to other water molecules and not to polar groups in the protein. W1–W6 were identified by NMR experiments. W205 and W5014 were located initially in the bulk solvent. W3 and W4, located in the interior of the NMR structure, escape during the equilibration process. W2 is coordinated to the side chains of Tyr 67, Asn 52, Thr 78, and a heme propionate.

reside inside the protein for time periods longer than 0.5 ns. The maximum residence time, time of entry, time of escape, MSD in the protein interior and rotational correlation functions for these 17 water molecules are described in Table I. The positions of these water molecules, relative to the protein, are shown in Figure 7 for two time frames. Not all water molecules are in the protein interior simultaneously. It is interesting to notice that, with the exception of W1, water molecules tend to form clusters or linear chains of two or more water molecules. This is consistent with the observations of Ernst et al.,[17] who suggested the formation of clusters of two or more water molecules lower the free energy of interior waters, relative to isolated water molecules.

Translational and Rotational Diffusion of Water Molecules

The mean square displacements (MSD) of water molecules in the interior of the protein, at T = 300 K, vary from 0.4 Å^2 to over 12 Å^2, for long-lived water molecules. Two water molecules that did not exchange during the T = 300 K simulation, W1 and W2, show MSD of 0.8 Å^2 and 1.6 Å^2, respectively. Small MSD values can be associated with tightly-bound water molecules, while large MSD values indicate extensive diffusion inside the protein, suggesting that cavities are transiently open as part of the protein fluctuations. The trajectories of water molecules that exhibit large MSD values while in the protein interior (relative to the averaged protein configuration) are shown

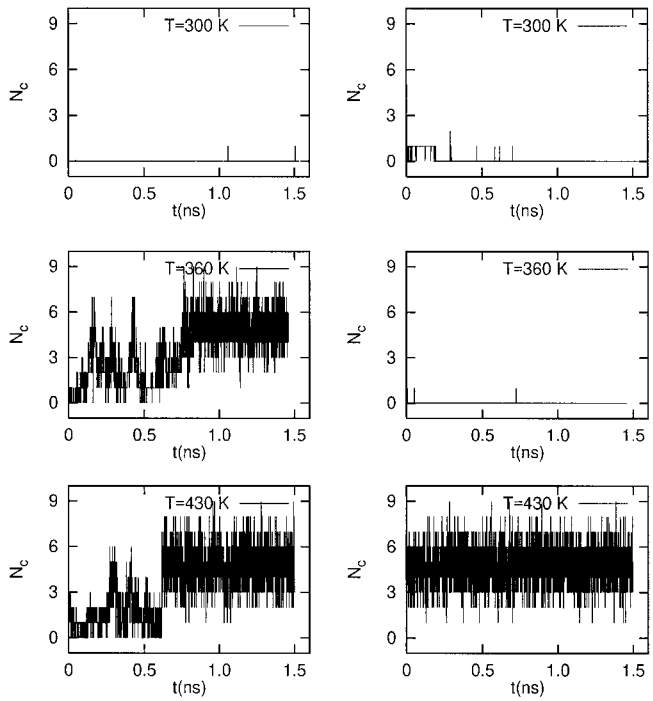

Fig. 4. Time series of the instantaneous coordination number for W1 (left) and W2 (right) at T = 300 K, 360 K and 430 K. These water molecules do not exchange with solvent water molecules during the 1.5 ns production run at 300 K, but exchange at higher temperatures. W2 exchanges during the 200 ps equilibration stage of the T = 430 K simulation.

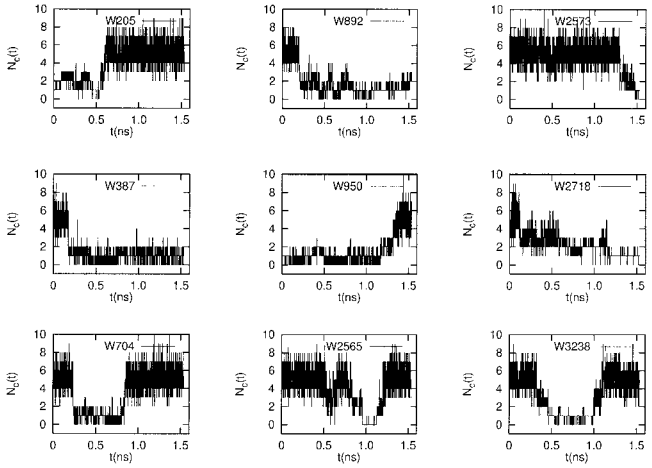

Fig. 5. Time series of the instantaneous coordination number for water molecules that reside in the protein interior for 150–1,360 ps (with the exception of W1 and W2, shown in Fig. 4). The figure shows the coordination number for W205, W387, W704, W892, W950, W2565, W2573, W2718, and W3238, with maximum residence times inside the protein of 556 ps, 1360 ps, 607 ps, 1215 ps, 1321 ps, 258 ps, 152 ps, 529 ps, and 795 ps, respectively.

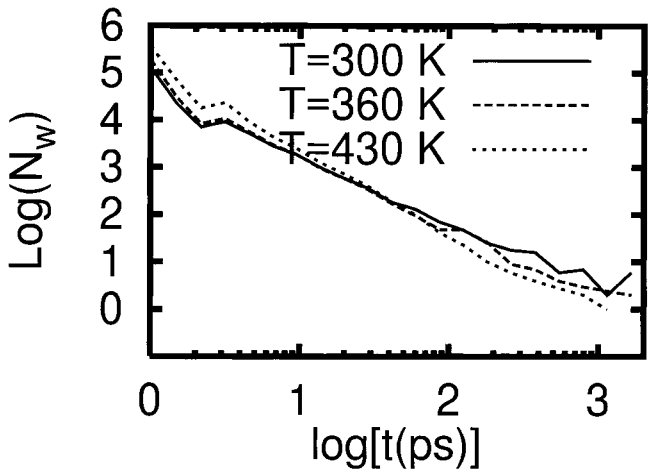

Fig. 6. Histograms of the logarithmic residence time of water molecules in the surface and interior of cyt c during 1.5 ns production stages of MD at 300 K, 360 K, and 430 K. These distributions show a broad range of residence times ranging from ps to ns time scales, consistent with a broad distribution of energy barriers involved in the water exchange kinetics. These distributions show an approximate power-law scaling with exponent ~2.5 for times in the 1–1,000 ps time scale.

in Figure 8. Positions of other water molecules with large residence times but small MSD, are shown in Figure 7.

The degree of rotational flexibility for these water molecules is illustrated in Figure 9 for W1, and in Figure 10 for W2. These figures show the rotational correlation function, $c(t)$, and the time history of the three components of the normalized vector, $\hat{m} = (m_x, m_y, m_z)$, and $\hat{\omega} = (\omega_x, \omega_y, \omega_z)$, defined above. The Rotation of the $\hat{m}$ vector in these water molecules is hindered and the rotational correlation functions, $c(t)$, show that orientational memory persists for periods larger than $t_{run}/2$. However, the time series of the

TABLE I. List of Water Molecules With Maximum Residence Times Inside the Protein Larger than 500 ps Observed During the T = 300 K Simulation†

Water index	t_{res} (ps)	t_{first} (ps)	t_{last} (ps)	MSD Å^2	$c(t)_{min}$ $\tau_{rot}(ns)$
1	1534	0	1534	0.4	0.8
2	1534	0	1534	0.7	0.75
387	1360	173	1534	0.4	0.86
950	1321	0	1321	2.4	0.6
451	1273	260	1534	0.4	0.91
892	1215	319	1534	2.1	0.5
3185	905	628	1534	7.0	0.35
3564	866	660	1526	0.5	0.88
399	791	453	1244	0.9	0.55
1282	736	696	1432	0.9	0.7
5597	695	839	1534	0.6	0.80
4003	627	200	828	12.2	0 (300 ps)
3238	617	397	1014	2.8	0.5
704	607	235	843	1.3	0.7
205	556	0	556	4.8	0 (270 ps)
2718	529	578	1107	3.5	0.5
188	510	269	780	5.9	0.5

†Water molecules are listed in descending order of maximum residence time. t_{res} is the maximum residence time, t_{first} is the time at which it entered the protein, and t_{last} is the time at which it escaped. MSD is the Mean Square Displacement for the water molecules while in the protein interior.

vector $\hat{\omega}$ shows fast flips through 180 deg about the dipole axis, $\hat{m}$. These flips occur infrequently (as seen in ω_z in Fig. 9) for W1, but quite frequently for W2 (as seen in ω_z in Fig. 10). For these water molecules, $c(t)$ shows a fast decay ($t \sim 0.5$ ps), followed by a slow decay (with time scales comparable to $t_{run}/2$) to $c(t) \sim 0.8$. The fast decay reflects the fast librations about an average orientation. The non-oscillatory, fast but infrequent changes in orientation (e.g., changes in m_z at $t \sim 1.1$ ns, for W1, and at $t \sim .25$ ns for W2) are reflected in the long time decay in $c(t)$, with characteristic times $> t_{run}/2$, where the water does not lose memory of its initial orientation. This long-term memory is due to the slowly changing protein interior. The mean square deviations of the librations can be obtained from $c(t \rightarrow t_{run}/2)$, where $c(t \rightarrow t_{run}/2) \sim 1 - \frac{1}{2} \langle \delta\phi^2 \rangle$. For $c(t \rightarrow t_{run}/2) \sim .75$, $\sqrt{\langle \delta\phi^2 \rangle} \sim 40$ deg, where $\langle \delta\phi^2 \rangle$ is the solid angle fluctuation of $\hat{m}$.

To illustrate the difference in rotational behavior between various water molecules, Figure 11 and Figure 12 show the rotational correlation function, $c(t)$, and the time history of the three components of the unit vector, $\hat{m}$, for two water molecules (W205 and W387) that exchange and also show different MSD in the protein interior [MSD(W205) = 4.8 Å^2, and MSD(W387) = 0.4 Å^2]. Rotational correlation functions of bound and unbound waters are calculated separately. Figure 11 shows that the rotation of W205 is hindered in the protein interior, but that fast and infrequent flips occur. $c(t)$ shows a memory of 0.25 ns. This estimate of the internal water relaxation time is unreliable since the infrequent flips dominate the relaxation behavior. The rotation in bulk water shows frequent rotations with a characteristic time of ~10 ps (similar to

the Debye relaxation time). Figure 12 shows that the rotation of W387 is hindered in the protein interior, with $c(t)$ exhibiting a memory similar to $t_{run}/2$, as was the case for W1 and W2. Water molecules in the protein-bound state show large librations. The rotation in the free state (bulk) also show frequent rotations with a characteristic time of ~10 ps. The contrasting rotational behavior of these water molecules suggests that molecules exhibiting large MSD are more likely to show fast rotational relaxation times (relative to t_{run}), although slower than the relaxation time in bulk water (~10 ps). This indicates that the rotation of protein-bound waters is hindered. Table I summarizes MSD and $c_{min}(t)$ for all water molecules with residence times longer than 500 ps. These water molecules show dynamics similar to the four molecules described above.

Protein-Bound Water Survival Time Correlation Function

Figure 13 shows the mean survival time correlation function of water molecules in the interior of cyt c at 300 K, 360 K and 430 K. The information in $S(t)$ is similar to that in Figure 6 where the distribution of water residence times is shown. The $S(t)$ curves show power-law behavior in time, over various time decades. For short times, $S(t) \sim t^{-2}$, for $t \leq 2$ ps, and $S(t) \sim t^{-.5}$, for $2 \leq t \leq 500$ ps. This power-law behavior is indicative of multiple energy barriers being involved in the exchange kinetics.[36,37] As temperature is increased, the residence time of water molecules is shortened and only one water does not exchange with the solvent during the 1.5 ns simulation at T = 360 K. At T = 430 K all interior water molecules exchange at least once, and mean residence times are shorter than 500 ps.

Water Relaxation Rates Distribution

The water survival time correlation functions shown in Figure 13 are decomposed in terms of a logarithmic distribution of rates using the maximum entropy method, as described by Steinbach et al.[37] and shown in Eq. (2). Figure 14 shows the rate distributions, $f(\log \tau)$, obtained at temperatures T = 300 K, 360 K, and 430 K. In the MEM we grid $\log \tau$ into 200 points logarithmically distributed in the time range $10^{-2} - 10^{4}$ ps, with an assumed uniform distribution of errors of 0.001, and $\Delta(\log \tau) = 0.03$. We use the values of $S(t)$ from $t = t_{wait} = 2.0$ ps to $t = t_{run}/2 = 750$ ps, where t_{run} is the length of the production stage of the MD simulations. At T = 300 K we observe that the rate distribution has three peaks centered at 5 ps, 26 ps, and 164 ps. These distributions can be interpreted in terms of water molecules that bind to the protein in regions nearby the protein surface and are easily exchanged, water molecules that penetrate near the surface and coordinate to charged side chains, and water molecules that penetrate deeply into the protein interior and exchange slowly. The upper bounds of these peaks are determined by the limited sampling. The distributions in the 1–10 ps and 10–100 ps are well characterized by our nanoseconds MD simulations. The tail of the distribution in the 100–1,000 ps is limited by the time scale of the simulations. We assign energy scales to the rate distribution peaks by transforming rates to enthalpies using, $H = RT \ln(10)[\log(AT/T_0) - \log(\tau))]$, where A is an Arrhenius prefactor, not needed for our estimates, that can be determined from the temperature dependence of the rate distributions.[37] Taking $T_0 = 300$ K, we obtain that the energy differences corresponding to the peaks at 5 ps and 164 ps, $\Delta H = RT \ln(10) \cdot \log(164ps/5ps) \sim 10$ kJ/mol, for the largest energy difference sampled by water molecules bound to cyt c, and $\Delta H \sim 4$ kJ/mol for water molecules bound to cyt c and exhibiting rates in the 1–10 ps and 10–100 ps. These values are comparable to those measured by Denisov and Halle, who measure an energy roughness of 10 kJ/mol for BPTI.[16]

The temperature dependence of the distribution of rates is shown in Figure 14. We find that the position of the peak for the fastest rate distributions shift toward faster rates as T increases. The interpretation of this T dependence can be simplified if we assume that components of the three peaks in the rate distributions at 300 K also shift toward shorter times as T increases. Within this assumption, we can assign the rates corresponding to the maxima of the distributions for each peak. The maximum of the second peak is taken as the top of the shoulder of the first peak at T = 360 K, and as the maximum of the first peak at T = 430 K. The maximum peak positions are shown in Table II. By fitting the distribution maximum for each peak as a function of temperature to

$$\tau = \tau_0 \left(\frac{T}{T_0}\right) exp\,(\Delta G^{\neq}/\kappa_B T)$$

$$= \tau_0 \left(\frac{T}{T_0}\right) exp\,(\Delta S^{\neq}/\kappa_B)\, exp\,(\Delta H^{\neq}/\kappa_B T), \qquad (4)$$

where τ_0 is an exponential prefactor and κ_B it the Boltzmann constant, T_0 is a reference temperature (taken as 300 K in our fittings) we obtain the relative entropic and enthalpic components of the water escape activation energy, within the assumption of no temperature dependence of $\Delta\Delta S^{\ddagger}$ and $\Delta H^{\dagger}$. [τ_0 is given by transition state theory as $\kappa_B/h \sim 6.3$ ps, at $T = T_0$.] We find that the activation enthalpies for water escape are 6 kJ/mol for fast exchanging water molecules ($\tau \sim 1$–10 ps at T = 300 K), 18 kJ/mol ($\tau \sim 10$–100 ps at T = 300 K), and 23 kJ/mol for slowly exchanging water molecules ($\tau > 100$ ps). The differences in activation entropies, $\Delta\Delta S^{\ddagger}$, are 0 J/K, −27 J/K, and −27 J/K, respectively. We calculate the entropies relative to the first peak contribution. The negative sign implies that the entropy of the escape transition state is lower than the bound state. Dunitz[1] has argued that the largest entropy lost by a water molecule upon binding is 28 J/K/mol, which compares well with the differences in activation entropies of −27 J/K/mol. Also, the large mobilities of most of the bound water molecules suggest that their entropy is not much lower than that in the bulk. Zhang and Hermans[42] have also concluded that buried waters have entropy comparable to that of liquid water or ice. By calculating activation entropies relative to the fast-exchanging state entropy, our results are independent

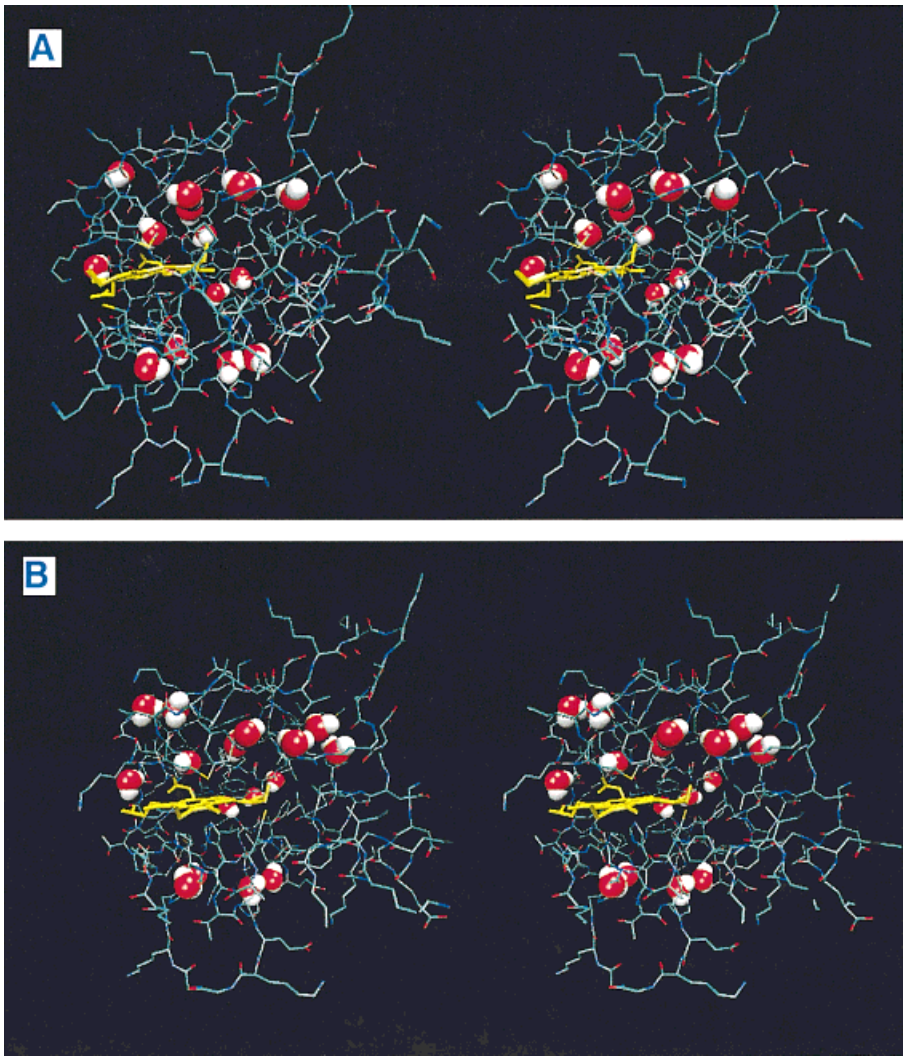

Fig. 7. Stereo plots of cyt c configurations at (**A**) t = 500 ps and (**B**) t = 850 ps of the production stage of the MD at T = 300 K. This figure shows positions of 17 water molecules residing in the interior of the protein for time periods longer than 500 ps. Not all water molecules are in the protein interior simultaneously. However, configurations at t = 500 (13 water molecules) or 850 ps (15 water molecules) together show all the water molecules that reside in the protein interior for times longer than 500 ps.

of the choice of τ_0, assuming τ_0 does not change from one water escape process to another.

CONCLUSIONS

We have studied the dynamics and kinetics of protein-bound water molecules in cytochrome c by ns time scale MD simulations at various temperatures. Water molecules that penetrate the protein and escape into the bulk solution are monitored through their coordination number. This approach is different from what is most commonly used when studying protein hydration, where the coordination number of protein atoms is monitored. Experimentally, it has been known that water molecules and other small ligands penetrate and escape the protein interior. Here, we have shown that, for water, this occurs easily within the ns time scale. Water penetration and exchange occur over a wide distribution of time scales, consistent with a process that can occur in a multitude of ways and different protein sites. The binding and escape of water molecules can also be coupled to protein fluctuations which occur over a broad distribution of time scales, consistent with the rough energy landscape picture of Frauenfelder et al.[36] In our simulations at 300 K, approximately 200 water molecules reside inside the protein for periods of 5 ps or more, and 17 water molecules have residence times longer than 0.5 ns. A large number of water molecules relax in a shorter time scale, and two water molecules exchange within times longer than 1.5 ns. One of these two water molecules exchanges at 360 K, but the other exchanges only at 430 K. At 430 K all water molecules show residence times shorter than 500 ps. We cannot give any estimates for the mean residence times of these two water molecules at 300 K, except that they are of the order of ns or longer. Qi et al.[26,12] found at least six water molecules with residence times longer than 300 ps.

The identification of protein-bound water molecules allows us to study their translational and rotational diffusion. We found that rotations of protein-bound water molecules are hindered. Most of the water fluctuations are described by large librations (with solid angle fluctuations of 20–40 deg), and by fast rotations about the water molecule C2 symmetry axis. These results are in qualita-

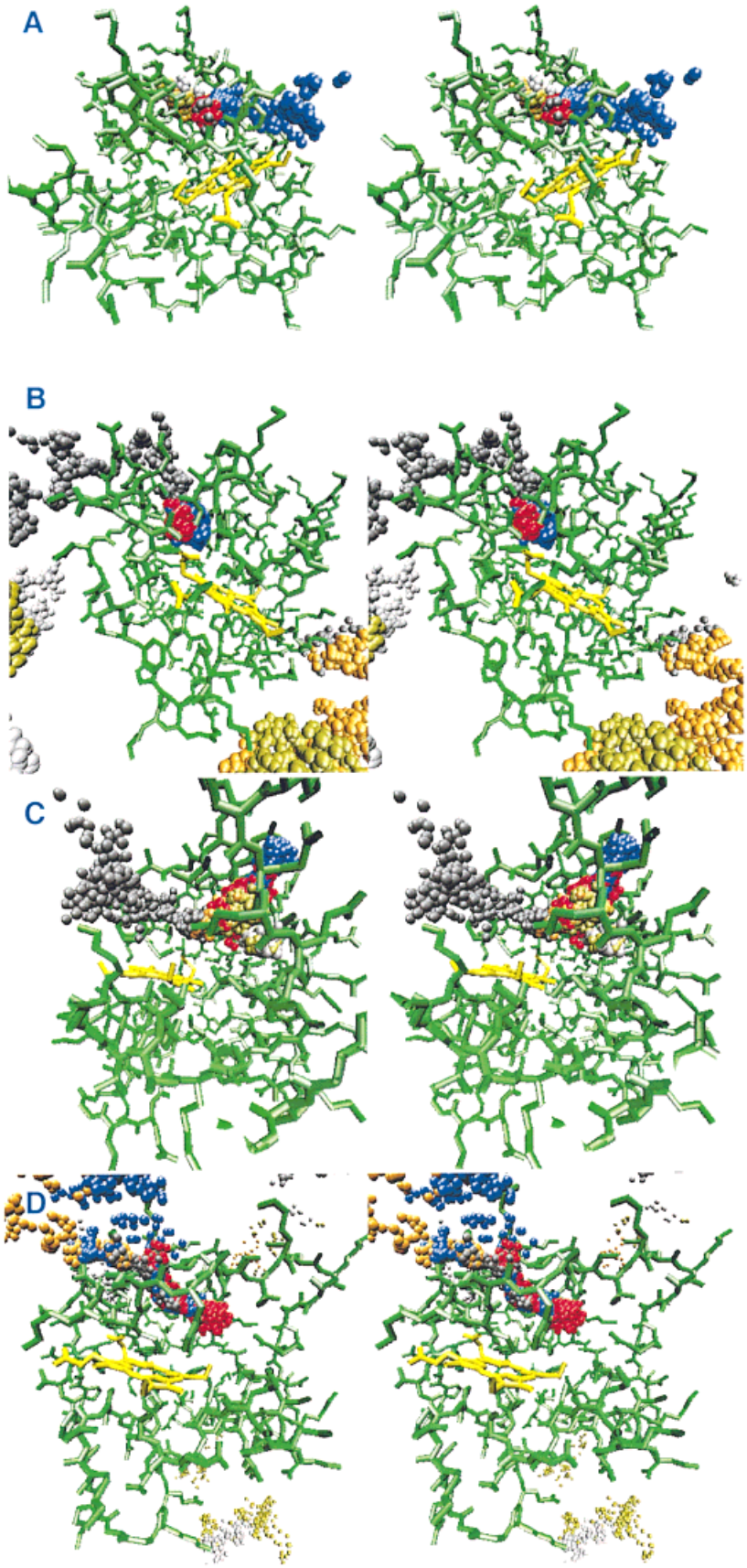

Fig. 8. Stereo plots of the trajectories of selected water molecules relative to the average protein configuration. These water molecules reside for times of 500–900 ps inside the protein during the production stage of the 300 K MD simulation. The water molecules shown are (**A**) W188, (**B**) W205, (**C**) W3185, and (**D**) W4003. The averaged mean square displacements (MSD) of these water molecules while bound to cyt c, are 5.9 Å^2, 4.8 Å^2, 7.0 Å^2 and 12.3 Å^2, respectively. The positions of other water molecules with large residence times but small MSD relative to the protein are shown in Fig. 7. The positions of water molecules relative to the time-averaged structure of cyt c are shown with solid spheres. The color of the spheres is changed with time (blue: $0 < t \leq 0.25$ ns; red: $0.25 < t \leq 0.5$ ns; gray: $0.5 < t \leq 0.75$ ns; orange: $0.75 < t \leq 1.0$ ns; yellow: $1.0 < t \leq 1.25$ ns; silver: $1.25 < t \leq 1.5$ ns).

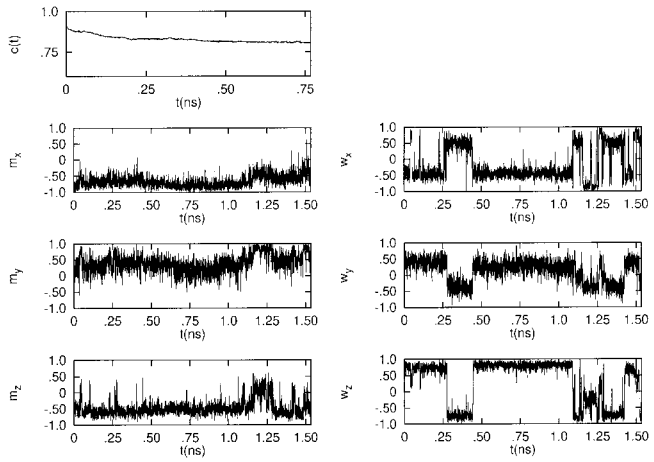

Fig. 9. Rotational correlation function, $c(t)$, and the time history of the three components of the unit vector, $\hat{m}$, which lies along the water dipole (left), and the three components of the unit vector, $\hat{\omega}$, which lies parallel to the water H-H segment (right), for the unexchanged water molecule, W1.

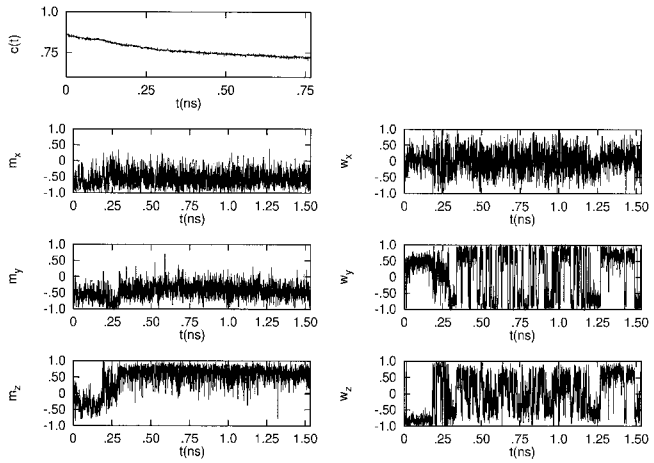

Fig. 10. Rotational correlation function, $c(t)$, and the time history of the three components of the unit vector, $\hat{m}$, which lies along the water dipole (left), and the three components of the unit vector, $\hat{\omega}$, which lies parallel to the water H-H segment (right), for the unexchanged water molecule, W2.

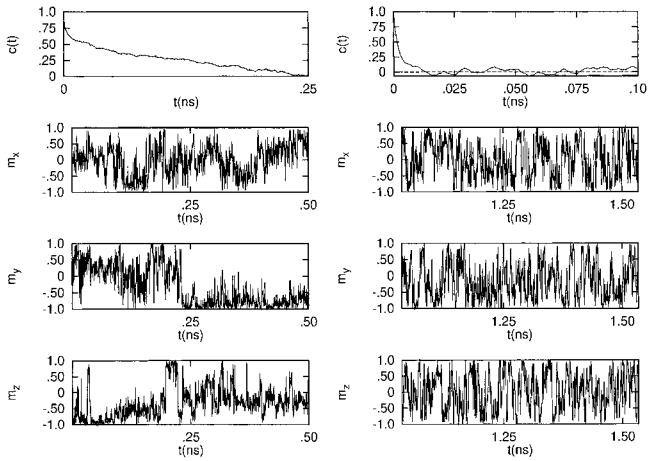

Fig. 11. Rotational correlation function, $c(t)$, and time history of the three components of the unit vector, $\hat{m}$, which lies along the water dipole for a water labeled W341, in the protein-bound (left) and free (right) states.

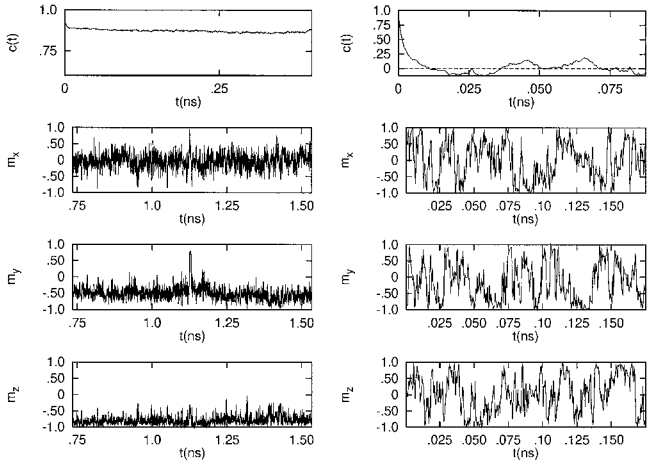

Fig. 12. Rotational correlation function, $c(t)$, and time history of the three components of the normalized vector, $\hat{m}$, along the water dipole for a water molecule labeled W523, in the protein-bound (left) and free (right) states.

tive agreement with the NMRD measurements on BPTI by Denisov et al.[40] However, Denisov found much smaller fluctuations in the libration amplitudes for BPTI, where the bound water molecules exchange with bulk water in the 15 ns to 1 μs time scale.

Protein-bound water molecules also show a broad distribution of translational displacements. Among water molecules that reside within the protein interior for times longer than 0.5 ns, some showed small MSD (0.4 Å^2) within the protein interior, while others show extensive diffusion within the protein interior, with MSD as large as 12 Å^2. Water molecules with large MSD also show greater rotational freedom, and rotational relaxation times that are shorter than for tightly bound water molecules, but much larger than in bulk water. Water molecules with large MSD may not be easily observable by NMR or X-ray diffraction techniques.

Rotationally-hindered water molecules affect the GHz dielectric response of the protein-water solution. Estimates of the number of water molecules solvating cyt c have been obtained from microwave dielectric measurements[43] on ferricytochrome c to be 180 ± 40. The survival time correlation function shows that, on average 243 water molecules are bound to cyt c, and 112 ± 7 are bound over a time window of at least 5 ps. Microwave measurements in the spectral region above 10^9 Hz (1 ns) are sensitive to the number of rotationally hindered bound water molecules in the ps time-scale. The measured Debye relaxation time, τ_D, is 8.8 ps, which compares to bulk water (8.5 ps) and is similar to the rotational correlation time of free water in our calculations (∼10 ps). Water molecules with long residence times in the protein-bound state show hindered rotations, in agreement with the measurements of Wei et al.[43]

The mean residence time of interior water molecules

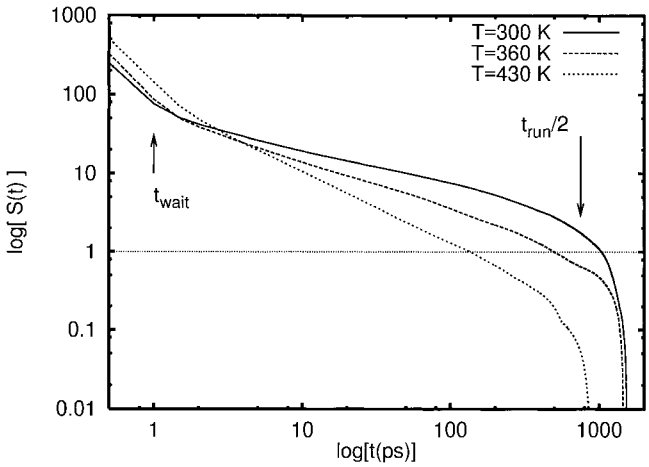

Fig. 13. Mean survival time correlation function, $S(t)$, of water molecules in the interior of cyt c at 300 K, 360 K and 430 K. $S(t = 0)$ measures the average number of water molecules in the interior of cyt c (i.e., $N_c \leq 3$), and $S(t)$ gives the average number of water molecules that remain in the interior after a period of time, t, given that they were present at t = 0. These curves show power-law behavior in time, with different exponents at short (<2 ps) and long times (>2 ps). This corresponds to broad residence time distributions shown in Figure 6.

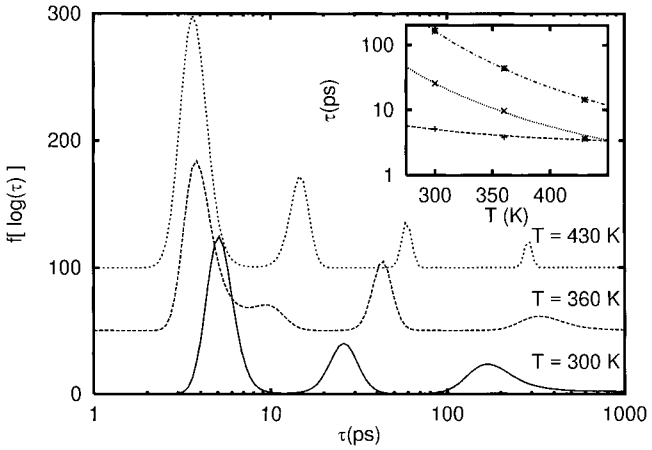

Fig. 14. Distribution of escape times, $f(\log \tau)$, obtained by the Maximum Entropy Method, for the survival time correlation functions, $S(t)$, shown in Eq. (2) and obtained at temperatures 300 K, 360 K, and 430 K. The inset shows a semi-log plot of the rates at the maximum of the rate distribution peaks as a function of T, where T is in K. The solid lines are fits of the data to Eq. (4) with the activation entropies and enthalpies for water escape as fitting parameters. Differences in activation entropies and enthalpies are listed in Table II.

(i.e., excluding surface and crevices) are in the 100 ps time scale. In the analysis of the dynamical properties of cyt c (for the same simulations used here[25]) we noticed that the diffusion of cyt c in configurational space exhibited subdiffusion for time scale faster than 100 ps, and superdiffusion for slower time scales. Subdiffusion refers to a dependence of the mean square displacements in time as a power law with exponent less than 1 (i.e., $\langle x^2 \rangle \sim t^{\beta}$, with $\beta < 1$). Superdiffusion refers to $\beta > 1$. Subdiffusion was explained in terms of a system that gets trapped in local minima. Superdiffusion, however, may result from sudden changes in the local potential energy surface of the protein as a result of water penetration or escape. Within this framework, a protein oscillates about a local minimum until a water molecule either escapes or penetrates, forcing the molecule to follow a rapid drift to a new local equilibrium. Thus, water molecules play a key role in the non-linear dynamics of proteins, where, due to their finite size, coupling to the protein energy surface is not merely electrostatic screening or providing a viscous coupling.

TABLE II. Location of the Maxima for the Various Peaks in the Distribution of Relaxation Rates†

T (K)	$\langle \tau_1(ps) \rangle$	$\langle \tau_2(ps) \rangle$	$\langle \tau_3(ps) \rangle$
300	5.1	25.8	164.4
360	3.8	9.7	43.5
430	3.6	3.6	14.5
$\Delta H^{\ddagger}$ (kJ/mol)	6	18	23
$\Delta\Delta S^{\ddagger}$ (J/K/mol)	0	−27	−27

†The free energy, ΔG are calculated from the fitting of log $\langle \tau_i \rangle$ to Eq. (4).

ACKNOWLEDGMENTS

We would like to thank B. McMahon for his assistance in the performance and interpretation of the MEM analysis. We thank H. Frauenfelder, S. Garde, D. LeMaster, R. Pomés and C. K. Woodward for enlightening discussions.

REFERENCES

1. Dunitz, JD. The entropic cost of bound water in crystals and biomolecules. Science 1994;264:670.
2. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature 1991;352:495–505.
3. Robinson CR, Sligar SG. Molecular recognition mediated by bound water: a mechanism for star activity of the restriction endonuclease *eco* RI. J Mol Biol 1993;234:302–306.
4. Robinson CR, Sligar SG. Changes in solvation during DNA-binding and cleavage are critical to altered specificity of the EcoRI endonuclease. Proc Natl Acad Sci 1998;95:2186–2191.
5. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. Proc Natl Acad Sci USA 1998;95:1552–1555.
6. Rupley JA, Careri G. Protein hydration and function. Adv Protein Chem 1991;41:37–172.
7. Levitt M, Park BH. Water: Now you see it, now you don't. Structure 1993;1:223–226.
8. Otting G, Liepinsh E. Protein hydration viewed by high-resolution NMR spectroscopy: implications for magnetic resonance image contrast. Acc Chem Res 1995;28:171–177.
9. Denisov VP, Halle B. Protein hydration dynamics in aqueous solution. Farad Disc 1996;103:227–244.
10. Schoenborn BP, García AE, Knott R. Hydration in protein crystallography. Prog Biophys Molec Biol 1995;64:105–119.
11. Otting G, Liepinsh E, Wüthrich K. Protein hydration in aqueous solution. Science 1991;254:974–980.
12. Qi PX, Urbauer JL, Fuentes EJ, Leopold MF, Wand AJ. Structural water in oxydized and reduced horse heart cytochrome c. Nat Struct Biol 1994;1:378–382.
13. Denisov VP, Halle B. Hydrogen exchange and protein hydration: the deuteron spin relaxation dispersions of bovine pancreatic trypsin inhibitor and ubiquitin. J Mol Biol 1995;245:698–709.
14. Denisov VP, Halle B. Protein hydration dynamics in aqueous solution: a comparison of bovine pancreatic trypsin inhibitor and ubiquitin by oxygen-17 spin relaxation dispersion. J Mol Biol 1995;245:682–697.
15. Denisov VP, Johnson BH, Halle B. Hydration of denatured and molten globule proteins. Nat Struct Biol 1999;6:253–260.
16. Denisov VP, Peters J, Horlein HD, Halle B. Using buried water-

molecules to explore the energy landscape of proteins. Nat Struct Biol 1996;3:505–509.
17. Ernst JA, Clubb RT, Zhou HX, Gronenborn AM, Clore GM. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science 1995;267:1813–1817.
18. Yu B, Blaber M, Gronenborn AM, Clore GM, Caspar DLD. Disordered water within a hydrophobic protein cavity visualized by X-ray crystallography. Proc Natl Acad Sci USA 1999;96:103–108.
19. Li R, Woodward C. The hydrogen exchange core and protein folding. Protein Sci 1999;8:1571–1591.
20. Woodward CK, Simon I. Hydrogen exchange rates and the dynamics structure of proteins. Mol Cell Biochem 1982;48:135–160.
21. Bai YW, Sosnick TR, Mayne L, Englander SW. Protein folding intermediates: native-state hydrogen exchange. Science 1995; 269:192–197.
22. Riistama S, Hummer G, Puustinen A, Dyer RB, Woodruff WH, Wikstrom M. Bound water in the proton translocation mechanism of the heme-copper oxidases. FEBS Lett 1997;414:275–280.
23. Bushnell GW, Louie GV, Brayer GD. High-resolution 3-dimensional structure of horse heart cytochrome c. J Mol Biol 1990;214: 585–595.
24. Muegge I, Qi PX, Wand AJ, Chu ZT, Warshel A. The reorganization energy of cytochrome c revisited. J Phys Chem B 1997;101:825–836.
25. García AE, Hummer G. Sampling of conformational substates by cytochrome c: correlation to hydrogen exchange. Prot Genet 1999;36:175–191.
26. Qi PX, DiStefano DL, Wand AJ. Solution structure of horse heart ferrocytochrome c determined by high-resolution nmr and restrained simulated annealing. Biochemistry 1994;33:6408–6417.
27. Simonson T, Perahia D. Microscopic dielectric properties of cytochrome c from molecular dynamics simulations in aqueous solution. J Am Chem Soc 1995;117:7987–8000.
28. Hartshorn RT, Moore GR. A denaturation-induced proton-uptake study of horse ferricytochrome-c. Biochem J 1989;258:595–598.
29. Cornell WD, Cieplak P, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
30. Darden T, York D, Pedersen L. Particle mesh Ewald: An *N* log (*N*) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
31. Pearlman DA, Case DA, Caldwell JW. AMBER 1995;Vol. 4.1.
32. Garde S, Hummer G, García AE, Paulaitis ME, Pratt LR. Origin of entropy convergence in hydrophobic hydration and protein folding. Phys Rev Lett 1996;77:4966–4968.
33. Sciortino F, Fornili SL. Hydrogen-bond cooperativity in simulated water: Time-dependence analysis of pair interactions. J Phys Chem 1989;90:2786–2792.
34. Sciortino F, Poole PH, Stanley HE, Havlin S. Lifetime of the bond network and gel-like anomalies in supercooled water. Phys Rev Lett 1990;64:1686–1689.
35. García AE, Stiller L. Computation of the mean residence time of water in the hydration shells of biomolecules. J Comput Chem 1993;14:1396–1406.
36. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
37. Steinbach PJ, Chu K, Frauenfelder H, Johnson JB, Lamb DC, Nienhaus GU, Sauke TB, Young RD. Determination of rate distributions from kinetic experiments. Biop J 1992;61:235–245.
38. García AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett 1992;68:2696–2699.
39. García AE, Hummer G, Blumfield R, Krumhansl JA. Multi-basin dynamics of a protein in a crystal environment. Physica D 1997;107:225–239.
40. Denisov VP, Venu K, Peters J, Horlein HD, Halle B. Orientational disorder and entropy of water in protein cavities. J Phys Chem B 1997;101:9380–9389.
41. Gu W, Schoenborn BP. Molecular dynamics simulation of hydration in myoglobin. Proteins 1995;22:20–26.
42. Zhang L, Hermans J. Hydrophilicity of cavities in proteins. Proteins 1996;24:433–438.
43. Wei Y-Z, Kumbharkhane AC, Sadeghi M, et al. Protein hydration investigations with high-frequency dielectric spectroscopy. J Phys Chem 1994;98:6644–6651.