JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

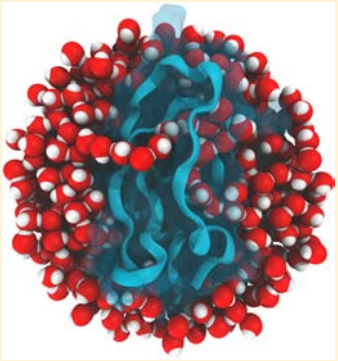
Transient Access to the Protein Interior: Simulation versus NMR

Filip Persson and Bertil Halle*

Biophysical Chemistry, Lund University, P.O. Box 124, SE-22100 Lund, Sweden

Supporting Information

ABSTRACT: Many proteins rely on rare structural fluctuations for their function, whereby solvent and other small molecules gain transient access to internal cavities. In magnetic relaxation dispersion (MRD) experiments, water molecules buried in such cavities are used as intrinsic probes of the intermittent protein motions that govern their exchange with external solvent. While this has allowed a detailed characterization of exchange kinetics for several proteins, little is known about the exchange mechanism. Here, we use a millisecond all-atom MD trajectory produced by Shaw et al. (*Science* **2010**, *330*, 341) to characterize water exchange from the four internal hydration sites in the protein bovine pancreatic trypsin inhibitor. Using a recently developed stochastic point process approach, we compute the survival correlation function probed by MRD experiments as well as other quantities designed to validate the exchange-mediated orientational randomization (EMOR) model used to interpret the MRD data. The EMOR model is found to be quantitatively accurate, and the simulation reproduces the experimental mean survival times for all four sites with activation energy discrepancies in the range 0−3 k_BT. On the other hand, the simulated hydration sites are somewhat too flexible, and the water flip barrier is underestimated by up to 6 k_BT. The simulation reveals that water molecules gain access to the internal sites by a transient aqueduct mechanism, migrating as single-file water chains through transient (<5 ns) tunnels or pores. The present study illustrates the power of state-of-the-art molecular dynamics simulations in validating and extending experimental results.



1. INTRODUCTION

In view of their high atomic packing density[1,2] and low intrinsic isothermal compressibility,[3,4] globular proteins appear to be more solid-like than liquid-like. But even if the onslaught of structural biology has tended to reinforce such a static view, it has always been clear that proteins are dynamic entities. Hydrogen exchange experiments revealed that all backbone amides occasionally make contact with the external solvent, even if the time scales and exchange mechanisms have remained elusive.[5−7] Even more strikingly, but perhaps controversially, tryptophan luminescence quenching experiments suggested that small molecules, like dioxygen, migrate through the protein interior as if it were a liquid.[8−10]

Yet another window on the inner workings of proteins, one that does not require extrinsic agents like exchange catalysts or luminescence quenchers, is provided by water molecules buried in small cavities inside the protein. Originally regarded as a quasi-permanent integral part of the protein structure,[11−13] these internal water molecules have subsequently been shown, with the aid of ^{2}H and ^{17}O magnetic relaxation dispersion (MRD),[14,15] to exchange with external water molecules on time scales ranging from tens of nanoseconds to hundreds of microseconds, depending on the nature of the internal hydration site.

These seemingly incompatible views of globular proteins can be reconciled by recognizing that solvent access to the protein interior is a rare and transient phenomenon. The protein spends long periods in the compact solid-like conformational 'ground state', intermittently interrupted by brief visits to highly excited conformations where the external solvent can access the protein interior. The minute equilibrium population and brief lifetime of such excited states pose technical challenges for experimental characterization of intermittent protein dynamics. This experimental limitation is particularly troubling since biological function often relies on intermittent protein dynamics, as in conformationally gated ligand binding and release.[16−19]

In principle, molecular dynamics (MD) computer simulations could come to the rescue here, by characterizing intermittent protein dynamics in full atomic detail, far beyond what might be inferred from experiment. Statistically adequate sampling of rare events requires what used to be prohibitively long simulation times, but the recent production of a millisecond MD trajectory of the 58-residue protein bovine pancreatic trypsin inhibitor (BPTI), solvated by 4215 explicit water molecules, has decisively removed this limitation.[20]

The ultralong MD simulation of BPTI can be used to address a variety of unresolved issues in protein biophysics. Here, our objective is three-fold. First, we use the MD simulation to quantitatively assess three approximations in the so-called exchange-mediated orientational randomization (EMOR) model[21] used to extract internal-water exchange kinetics from MRD experiments on immobilized proteins.[15] Second, by confronting the MD simulation with experimental results on the exchange kinetics and flexibility of the four internal hydration sites in BPTI, we validate the molecular mechanics force field used in the simulation. Third, we characterize the

Received: April 5, 2013
Published: May 15, 2013

   dx.doi.org/10.1021/ja403405d | *J. Am. Chem. Soc.* 2013, 135, 8735−8748

mechanisms whereby water molecules gain transient access to the protein interior.

In the process of analyzing the MD trajectory, we identified and addressed several nontrivial computational issues. Several strategies have been proposed for locating protein hydration sites in MD simulations.[20,22−26] Here, our priority was to identify all persistent hydration sites (occupied by long-lived water molecules) in a protein that, during the course of the 1 ms simulation, undergoes substantial structural fluctuations. These fluctuations give rise to dynamical disorder,[18] where the water exchange rate at a given site depends on the conformational state of the protein. Of particular interest is the isomeric state of the C14−C38 disulfide bond, which connects the two extensive loops that line the four internal hydration sites[27,28] and mediate the inhibitory binding of BPTI to β-trypsin.[29,30] These disulfide states have been studied by NMR,[31,32] and it has been shown that MD simulations do not reproduce the conformational equilibrium quantitatively.[33−35] When comparing with experimental internal-water exchange rates, we must therefore take the protein conformation into account explicitly.

Our computational analysis is based on the insight that internal-water exchange can, to an excellent approximation, be modeled as a continuous-time process with discrete state space. The time series of exchange events or, equivalently, of residence (interexchange) times, can therefore be described mathematically as a stationary point process.[36−38] The main virtues of the stochastic point process formalism, described in detail elsewhere,[39] are its generality and efficiency. In particular, the formalism allows us to characterize dynamical disorder resulting from conformational heterogeneity and/or from multiple exchange mechanisms. It is then necessary to distinguish the residence time (RT) from the survival time (ST). The RT is the time interval from one exchange to the next, and the ST is the time interval from an arbitrary time point to the next exchange. The RT depends sensitively on the site definition, but experiments, which are not synchronized with the exchange, report on the ST. Because all available kinetic information is contained in the RT distribution (or histogram), it can be used to compute also the experimentally relevant survival correlation function and its time integral, the mean ST, in a way that is vastly more efficient[39] than conventional algorithms.[40]

Our validation of the EMOR model includes a quantitative analysis of site correlations, featuring total residence and survival correlation functions, and the probability that a newly exchanged water molecule returns to the same site. The theoretical framework for this analysis is developed here and presented in full in the Supporting Information.

Protein force field validation and optimization studies typically focus on the most populated conformations,[41,42] corresponding to the deepest free-energy basins. In contrast, the intermittent dynamics examined here involve rarely accessed regions of the energy landscape and may therefore furnish a more demanding test of the force field. The simulation succeeds remarkably well in reproducing the MRD-derived mean STs for the four internal sites. For the three sites W111−W113, the activation energy discrepancy is 1.5 k_BT or less, whereas a two-fold larger discrepancy is obtained for the site (W122) adjacent to the C14−C38 disulfide bond (the conformational preferences of which are not accurately described by the force field). On the other hand, the force field underestimates the large water C_2 flip barrier in the sites by as much as 6 k_BT. Enhanced flexibility in the simulated hydration sites is also manifested via the libration amplitudes.

Frame-by-frame examination of 66 water exchange events in the three deepest sites (W112, W113, and W122) revealed, in all cases, a short-lived (<5 ns) transition state where the buried site is accessed via a single-file water chain migrating through a transient tunnel or pore. We observe two main variants of this transient aqueduct mechanism. In one variant, the C14−C38 disulfide (and the associated loops) is in the 'ground state' (M1), and the water chain is mainly constituted by the water molecules in the three adjacent sites W111−W113, sometimes with a fourth transient site being involved as well. In the second variant, which dominates for site W122, the transition state has the disulfide in the M2 state and one or more new tunnels or pores open up, leading to a higher water content in the interloop region.

2. METHODS

2.1. Analysis of MD Trajectory. Our computational analysis is based on a previously reported 1.031 ms all-atom MD simulation at 300 K of the protein BPTI, solvated by 4215 water molecules, and 6 chloride ions.[20] The analyzed trajectory comprises 4 125 000 frames with sampling resolution $\Delta\tau$ = 0.25 ns.

For the internal hydration sites in BPTI, water exchange can, to a very good approximation, be described as a discrete-state process (a tagged water molecule either resides or does not reside in the site) since the mean residence time of a water molecule in the site is much longer than the duration of the actual exchange event. In analyzing the quasi-continuous MD trajectory, we can therefore model internal-water exchange as a stationary point process.[36−38] A comprehensive account of this rather general stochastic approach can be found elsewhere;[39] here we merely sketch the key points.

For each hydration site α, we construct the RT vector $\mathbf{V}_{R,\alpha}$ and the RT histogram $F_{R,\alpha}(n)$. Because conformational-state transitions do not in general coincide with water exchange events, a convention must be adopted for extracting a state-specific subtrajectory for a given site. The simplest option would be to concatenate all frames belonging to the selected state. However, the resulting truncation of RTs would introduce a bias that shifts the RT distribution to shorter values. We therefore use a 'democratic' approach, where, for a given site, each RT is assigned to the state that is represented in the largest number of frames in that RT.

The RT vector contains $N_{R,\alpha}$ RTs and $N_{F,\alpha}$ frames. The mean RT is simply:[39]

$$\tau_{R,\alpha} = \Delta\tau \frac{N_{F,\alpha}}{N_{R,\alpha}} \tag{1}$$

where $\Delta\tau$ = 0.25 ns is the sampling resolution. The experimentally relevant mean survival time (ST) $\tau_{S,\alpha}$ is usually computed by averaging the initial time over the trajectory.[40] However, because all kinetic information is contained in the RT histogram, it is vastly more efficient to compute the mean ST as[39]

$$\tau_{S,\alpha} = \frac{\Delta\tau}{2N_{F,\alpha}} \sum_{n=1}^{\infty} n^2 F_{R,\alpha}(n) \tag{2}$$

The residence correlation function (RCF) and survival correlation function (SCF) are obtained from the RT histogram as[39]

$$Q_{R,\alpha}(n) = \frac{1}{N_{R,\alpha}} \sum_{p=n+1}^{\infty} F_{R,\alpha}(p) \tag{3}$$

$$Q_{S,\alpha}(n) = \frac{1}{N_{F,\alpha}} \sum_{p=n}^{\infty} \sum_{q=p+1}^{\infty} F_{R,\alpha}(q) \tag{4}$$

with the discrete time interval $n = \tau/\Delta\tau$.

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

The RT and ST statistics presented here are subject to a systematic binning error and a random statistical error.[39] The binning error is due to the replacement of the quasi-continuous RT probability density by a discrete RT histogram and to the possibility that exchange events can escape detection when the trajectory is sampled at a finite resolution $\Delta\tau$.[39] Because the STs of interest here are much longer than $\Delta\tau$, the binning error is unimportant. The statistical error arises because the RT ensemble is incompletely sampled by a trajectory of finite length. All ST statistics presented, such as the SCF $Q_{S,\alpha}(n)$ and the mean ST $\tau_{S,\alpha}$, as well as quantities derived from them, such as the activation energy discrepancy $\Delta\Delta E_{A,\alpha}$, are accompanied by estimates of the statistical error. A comprehensive tretament of the binning and statistical errors in point process analysis of MD trajectories has been presented.[39]

Multiexponential deconvolution of the SCF was performed with the non-negative least-squares (NNLS) algorithm,[43,44] which is more robust than nonlinear optimization methods and does not require an *a priori* assumption about the number of exponentials. So as not to place undue weight on long τ values, we resampled the SCF by selecting 100 points uniformly spaced along the curve in a semilog plot. For the NNLS kernel, we used 200 exponentials with log-spaced decay times between 2.5 ns and 5 μs. NNLS components with decay times differing by <10% were merged.

Site correlation, that is, the time-dependent probability that a water molecule returns to a previously occupied site, was not discussed in our general account of the point process approach to MD trajectory analysis.[39] The theoretical basis of site correlation analysis is therefore presented here in the Supporting Information (Section S6). The mean total RT and ST were computed from eqs S15−S22 using the occupancy vectors $\mathbf{A}_\alpha$ (excluding vacant frames). Our site correlation analysis includes all conformational states, but this has little effect on the experimentally relevant mean (total) ST, which is dominated by the long RTs in the experimentally most populated M1 state.

Hydration site flexibility was characterized through order parameters and flip times, computed from the MD trajectory as described in Supporting Information (Section S7).

2.2. Analysis of ^{2}H Magnetic Relaxation Dispersion. The EMOR model used to analyze the MRD data has been fully described elsewhere.[21] But to appreciate the physical significance of the model parameters and the nature of the simplifying assumptions in the general model, it is sufficient to consider the fast-exchange limit of the model. For the EMOR model, the fast-exchange regime coincides with the motional-narrowing regime, where the conventional second-order perturbation theory of spin relaxation is valid.[45] For a hydration site α with mean survival time $\tau_{S,\alpha}$ and locally averaged water-^{2}H nuclear quadrupole frequency $\omega_{Q,\alpha}$ the fast-exchange regime is defined by $\omega_{Q,\alpha}\,\tau_{S,\alpha} \ll 1$.[21] In the fast-exchange regime, the observed longitudinal relaxation rate is given by

$$R_1(\omega_0) = \frac{1}{N_W}\sum_\alpha \xi_\alpha R_{1,\alpha}(\omega_0) + R_{1,ext} \tag{5}$$

with ω_0 the (angular) Larmor frequency, N_W the water/protein mole ratio in the sample, and $R_{1,ext}$ the frequency-independent relaxation contribution from external (hydration and bulk) water molecules. Furthermore, ξ_α is the water occupancy of internal hydration site α with intrinsic relaxation rate:

$$R_{1,\alpha}(\omega_0) = \frac{2}{3}\omega_{Q,0}^2 S_\alpha^2\left(1+\frac{\eta_\alpha^2}{3}\right)[0.2J_\alpha(\omega_0) + 0.8J_\alpha(2\omega_0)] \tag{6}$$

where $\omega_{Q,0} = (3/2)\,\pi\chi = 1.06\times10^6$ rad s^{-1} is the rigid-lattice ^{2}H quadrupole frequency. The averaging effect of local motions in the site is described by the principal value of the rank-2 orientational order tensor, $S_\alpha = \omega_{Q,\alpha}/\omega_{Q,0}$, and the averaged asymmetry parameter η_α.[21] The isotropic order parameter, as conventionally defined in connection with the model-free approach to spin relaxation,[46] is $S_{iso,\alpha} = S_\alpha(1+\eta_\alpha^2/3)^{1/2}$.

The spectral density function $J_\alpha(\omega)$ in eq 6 is the cosine transform of an orientational time correlation function $C_\alpha(\tau) = \langle F(\Omega_0)F^*(\Omega)\rangle_\alpha$ that decays from 1 to 0. The arguments of the function F are the Euler angles that specify the orientation of the principal frame of the locally averaged EFG tensor at the ^{2}H nucleus at time τ (Ω) or at time 0 (Ω_0). The ensemble average can be expressed in terms of the isotropic orientational distribution $f(\Omega_0) = 1/(8\pi^2)$ and the orientational propagator $f_\alpha(\Omega,\tau \mid \Omega_0)$. If the protein is fixed (as in a cross-linked protein gel), the orientation Ω can only change if the water molecule leaves the site (and is exchanged for another water molecule). In the EMOR model, the orientation is taken to be completely randomized upon exchange. In other words, we assume that, immediately after leaving the site, the water molecule looses all correlation with the site. It then has the same probability of returning to the site as any other water molecule in the sample. We refer to this as the instantaneous randomization approximation.

Furthermore, we assume that the mean ST of a water molecule in the site is much longer than the typical duration of the actual exchange event. We refer to this as the point process approximation. After these two approximations, the orientational propagator can be expressed as

$$f_\alpha(\Omega,\tau \mid \Omega_0) = f(\Omega) + [\delta(\Omega-\Omega_0) - f(\Omega)]Q_{S,\alpha}(\tau) \tag{7}$$

where $Q_{S,\alpha}(\tau)$ is the SCF introduced in Section 2.1.

The third approximation in the EMOR model is the assumption that water exchange can be described as a Poisson process. The SCF then decays exponentially:

$$Q_{S,\alpha}(\tau) = \exp(-\tau/\tau_{S,\alpha}) \tag{8}$$

where $\tau_{S,\alpha}$ is the mean ST. This implies, among other things, that the RT (the time from arrival to departure) of a given water visitor in the site is independent of the RT of the previous visitor. Adopting the three approximations underlying eqs 7 and 8, we obtain $C_\alpha(\tau) = Q_{S,\alpha}(\tau) = \exp(-\tau/\tau_{S,\alpha})$ so the spectral density $J_\alpha(\omega)$ is a Lorentzian function. One of the objectives of this work is to use MD simulation data to quantitatively assess the accuracy of these three approximations.

It follows from the preceding that, in the fast-exchange regime, the MRD profile $R_1(\omega_0)$ is a weighted (by $\xi_\alpha\, S_{iso,\alpha}^2$) sum of Fourier-transformed SCFs for the different long-lived (internal) hydration sites of the protein. The contribution from each site is fully determined by the two parameters $\xi_\alpha\, S_{iso,\alpha}^2$ and $\tau_{S,\alpha}$. However, if the mean survival time $\tau_{S,\alpha}$ is not much smaller than $1/\omega_{Q,\alpha} \approx 1$ μs, then we are no longer in the fast-exchange regime, and we must use the more general version of the EMOR theory, based on the stochastic Liouville equation.[21] The contribution from each site is then described by four parameters: ξ_α, S_α, η_α, and $\tau_{S,\alpha}$.

For the comparison with the MD simulation, we use a MRD data set for immobilized BPTI at 20 °C and pD 6.5, where contributions from nonwater labile deuterons can be neglected.[15] The general EMOR theory can be implemented at various levels of approximation.[21] In particular, the theory is simplified in the so-called dilute regime, where $N_W \gg 1$. For the data analyzed here $N_W = 3259$, so we are squarely in the dilute regime. For the analysis, we used the exact expression, eq (4.7) in ref 21, for the dilute regime, which is slightly more accurate than the expression used in the original publication.[15] Moreover, we performed the nonlinear optimization with the trust-region reflective algorithm (rather than the Levenberg−Marquardt algorithm used before),[15] which allows us to constrain the model parameters to their physically admissible ranges: ≥ 0 for all parameters and ≤ 1 for S_α and η_α. As before,[15] we found that the MRD data can be well described ($\chi^2_{red} \approx 1.5$) by a single slow component (requiring the general EMOR theory) and two components in the fast-exchange regime. The faster of the latter two, with $\xi_\alpha\, S_{iso,\alpha}^2 = 2.4$ and $\tau_{S,\alpha} = 7$ ns attributed to water molecules confined between protein molecules joined by short cross-links,[15] will not concern us here.

3. RESULTS AND DISCUSSION

3.1. Identification of Persistent Hydration Sites. As the first step, we identify the internal hydration sites. Here, we are only interested in persistent hydration sites, such that the

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

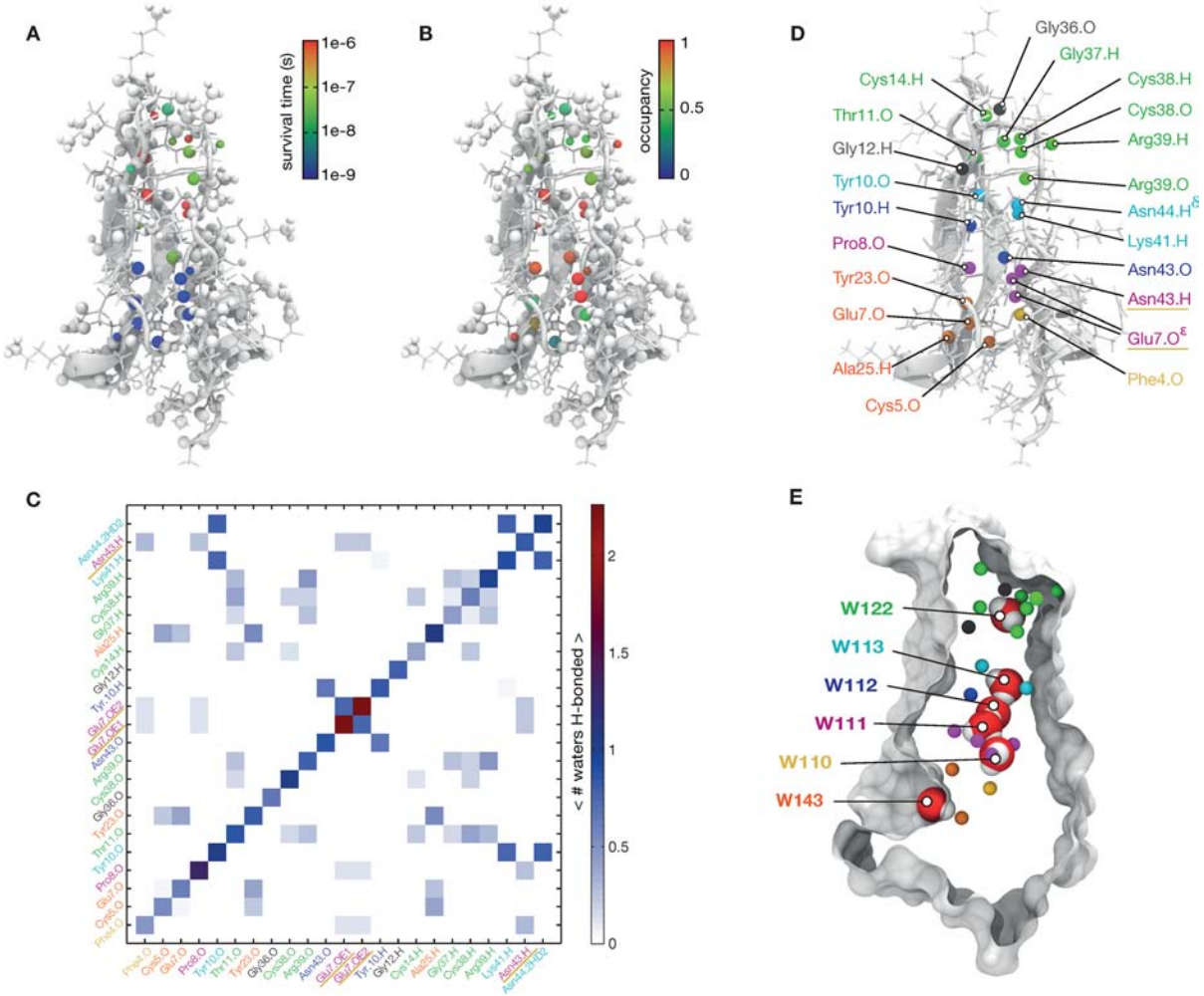


Figure 1. Identification of persistent hydration sites in BPTI. (A,B) Subset of 23 protein atoms (A) H-bonded to water molecules with mean survival time >1 ns and (B) with mean hydration number >0.1. (C) Hydration matrix for the subset of 23 protein atoms. (D) Identification of the 23 atoms in (C), color-coded according to hydration site. Underlined atoms are shared by sites W110 and W111. (E) Water molecules residing in the six persistent hydration sites.

resident water molecule makes at least 2 H-bonds with protein polar atoms and remains in the site for at least 1 ns. To find these sites, we first identify, in each frame of the trajectory, all H-bonds D−H···A where either D or A is a water oxygen and the other partner is one of the 170 polar (O, N or S) atoms in BPTI. A H-bond was taken to exist if $R(\mathrm{H}\cdots\mathrm{A}) \leq 3.0$ Å and $\theta(\mathrm{DHA}) \geq 130°$. For each protein atom μ engaged in at least one H-bond to water, we computed the mean survival time $\tau_{S,\mu}$ (Section 2.1) for the H-bonded water molecules and the mean hydration number $H_{\mu\mu}$, defined as the trajectory-averaged number of water molecules H-bonded to atom μ. We then selected the protein atoms with $\tau_S > 1$ ns (Figure 1A) and those with $H_{\mu\mu} > 0.1$ (Figure 1B).

For the set of 23 protein atoms satisfying both of these criteria, we constructed a hydration matrix[26] **H** with diagonal elements $H_{\mu\mu}$ as defined above and off-diagonal elements $H_{\mu\nu}$ giving the trajectory-averaged number of water molecules simultaneously H-bonded to protein atoms μ and ν (Figure 1C). The protein atoms constituting the persistent hydration sites are then identified with elements >0.1 in any given row (or column) of the hydration matrix. The requirement of at least 2 H-bonds excludes 2 of the 23 protein atoms.

The remaining 21 protein atoms define 6 hydration sites (Figure 1D). These six sites include the four well-known internal hydration sites of BPTI, traditionally labeled W111−W113 and W122 (Figure 1E). The protein atoms defining these sites (Table S1) are also found to be within H-bonding distance of an internal water molecule in the room-temperature crystal structure 5pti.[27] However, for site W122, we find three additional H-bond partners (Table S1, Figure S1), not seen in the crystal structure, which belong to this hydration site when the nearby C14−C38 disulfide bond is in the m_{C14} isomeric state (Sec. 3.2).

The two remaining sites, W110 and W143, are the two most persistent external hydration sites in BPTI. In the crystal structure 5pti, site W143 is distinguished by a water molecule confined to a moderately deep surface pocket. The external site W110 shares 3 water H-bond partners with the nearby internal site W111. Unlike the conformation-dependent versions of site W122 (Figure S1), the two neighboring sites W110 and W111 may be occupied simultaneously.

In the following, we focus on the four internal hydration sites for which quantitative experimental data are available.[14,15] Among these, the three most deeply buried sites W112, W113, and W122, with the most long-lived water molecules, will be

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

analyzed most thoroughly. These three sites are henceforth referred to as the 'deep' sites.

For every trajectory frame, each of the six persistent hydration sites is assigned to either of two internal states: 'occupied' or 'vacant'. The site is said to be occupied if a water molecule is H-bonded to at least 2 (or 3 for site W122) of the site-defining protein atoms (Table S1). If this is not the case, the site is designated as vacant. A site that is vacant in this sense is not necessarily devoid of water. Rather, the site is either disrupted (with the site-defining protein atoms no longer in spatial proximity) or multiply occupied (but with no water molecule satisfying the H-bond criteria, as during the actual exchange event). For each site, we then construct an occupancy vector $\mathbf{A}_{\alpha,0}$, where the element $A_{\alpha,0}(k)$ is the index (a number between 1 and 4215) of the water molecule occupying site α in frame k, or 0 if the site is vacant.

3.2. Conformational Heterogeneity. The four internal hydration sites are lined by two extensive loops, L1 (residues 7−17) and L2 (residues 36−47), that are involved in the tight inhibitory binding of BPTI to β-trypsin.[30] These loops are bridged by the C14−C38 disulfide bond and by H-bonds mediated by the water molecules in the four internal sites (Table S1). Since the isomeric state of the disulfide bond affects the multimodal distribution of backbone dihedrals in the loops,[33] it should also affect the rate of exchange of the internal water molecules H-bonded to these backbone atoms. Indeed, cleavage of the disulfide bond accelerates water exchange from site W122 by 2−4 orders of magnitude.[47] To understand the internal-water exchange kinetics, it is therefore necessary to characterize the conformational state of the protein in the neighborhood of the hydration sites. Furthermore, the failure of the simulation to quantitatively reproduce the true conformational distribution[20,34] must be taken into account when comparing simulated and experimental water exchange rates.

Experimentally, the C14−C38 disulfide bond has been shown to interconvert among a major isomeric state (M) and two minor states (m_{C14} and m_{C38}).[31,32] Following Xue et al.,[34] who used a less densely sampled version of the same MD trajectory,[20] we performed a cluster analysis of the isomeric states of the C14−C38 disulfide bond, including a Gaussian decomposition of the M state into three substates (Figure S2). Based on the dihedral angle cutoffs thus established (Table S2), we assign each frame of the trajectory to either of six conformational states: M1, M2, M3, m_{C14}, m_{C38} or 'other'.

Whereas (the water oxygen in) site W122 is only 4.0 Å from (the center of) the C14−C38 disulfide bond, the most remote internal site W111 is 14.5 Å away. Indeed, the water exchange rate in site W111 depends less on the disulfide isomeric state than on the rotameric state of the nearby Glu-7 side-chain (Figure S3). In the room-temperature crystal structure 5pti, this side-chain (except the C^{β} atom) was modeled in two partially occupied conformations:[27] an open state (E7op) with the side-chain protruding from the protein surface and a closed state (E7cl) with the side-chain acting like a lid on the pore-like entrance to site W111. In the simulation, the E7 side-chain makes frequent transitions between these two rotameric states (Figure S3), as also observed in an earlier MD simulation.[48]

The principal conformational states seen in the simulation have also been observed experimentally, albeit with significantly different populations (Table 1). With regard to the C14−C38 disulfide states, this discrepancy has already been discussed.[20,34,35] Converting the population ratios (simulation/experiment) to second-order free energy differences via the Boltzmann factor, one finds $\Delta\Delta G \approx 1.4\ k_BT$ for the M1 state and 4.0 k_BT for the m_{C14} state. Because the M1 state dominates strongly both in solution[32] and in the crystal,[27,28] we shall compare experimental water exchange rates with simulated rates pertaining to state M1.

Table 1. Conformational State Populations (%)

state	simulation[a]	experiment
M1	25.4	100 (crystal)[27,28]
M2	2.8	
M3	6.4	
M	34.6	95 (solution[a])[32]
m_{C14}	50.0	1 (solution[a])[32]
m_{C38}	6.4	4 (solution[a])[32]
other	9.0	
E7cl	70.2, 79.9 (M1)[b]	70 (RT),[27] 53 (cryo)[28]
E7op	29.8, 20.1 (M1)[b]	30 (RT),[27] 47 (cryo)[28]

[a]At 300 K. [b]E7 state populations within C14−C38 disulfide state M1.

For the E7 states, room-temperature and cryogenic (125 K) crystallography yield different results (Table 1). Presumably, this discrepancy is a cryo-artifact, where flash-cooling remodels the conformational distribution of side chains.[49,50] Comparing the simulation with the room-temperature crystal structure, we find quantitative agreement (Table 1). However, since the crystal structure is in the M1 disulfide state, we should consider the E7 conformational equilibrium within that state only. Even though the population of the E7cl state then increases from 70 to 80%, the agreement with experiment is still excellent ($\Delta\Delta G \approx 0.1\ k_BT$).

To shed further light on the dependence of water exchange on protein conformation, we focus on a spatial region containing the four internal hydration sites. We define this region as the convex hull spanning the backbone atoms of two segments of the polypeptide chain (Figure 2): residues 7−14 (within loop L1) and residues 32−43 (including part of a β strand and part of loop L2). These segments contain all the 10 residues whose backbone atoms are H-bonded to internal water molecules (Table S1).

Figure 2 shows the variation, during the 1 ms MD trajectory, of the volume of the interloop region, of the RMSD of the selected backbone atoms, and of the number of water molecules within the region (Table 2). Both the mean and fluctuation of these variables correlate strongly with the C14−C38 isomeric state (Figure 2). In the experimentally dominant M1 state, the region is compact and rigid with 4.0 ± 0.3 enclosed water molecules. In the highly populated (in the simulation) m_{C14} state, the volume has expanded by ∼10%, the backbone is more flexible, and the water content is slightly larger. In the rare and transient M2 state, the volume is ∼20% larger, the backbone RMSD is 3-fold larger, and the region now contains 6.3 ± 2.2 water molecules, suggesting that this state is on the kinetic pathway of water exchange (Section 3.7).

3.3. Water Exchange Kinetics. The C14−C38 and E7 conformational states in each frame were encoded in two state vectors, which, together with the occupancy vectors $\mathbf{A}_{\alpha,0}$ for the six persistent hydration sites, constitute the basis for the subsequent analysis (Sec. 2.1). Relevant statistics for the six sites are compiled in Tables S3−S9.

Whereas crystallography[27,28] and MRD[14,15] indicate that the four internal hydration sites are fully occupied, the simulation indicates that these sites are 'vacant' (with no water molecule satisfying the site-defining H-bond criteria) as much as 13.3−

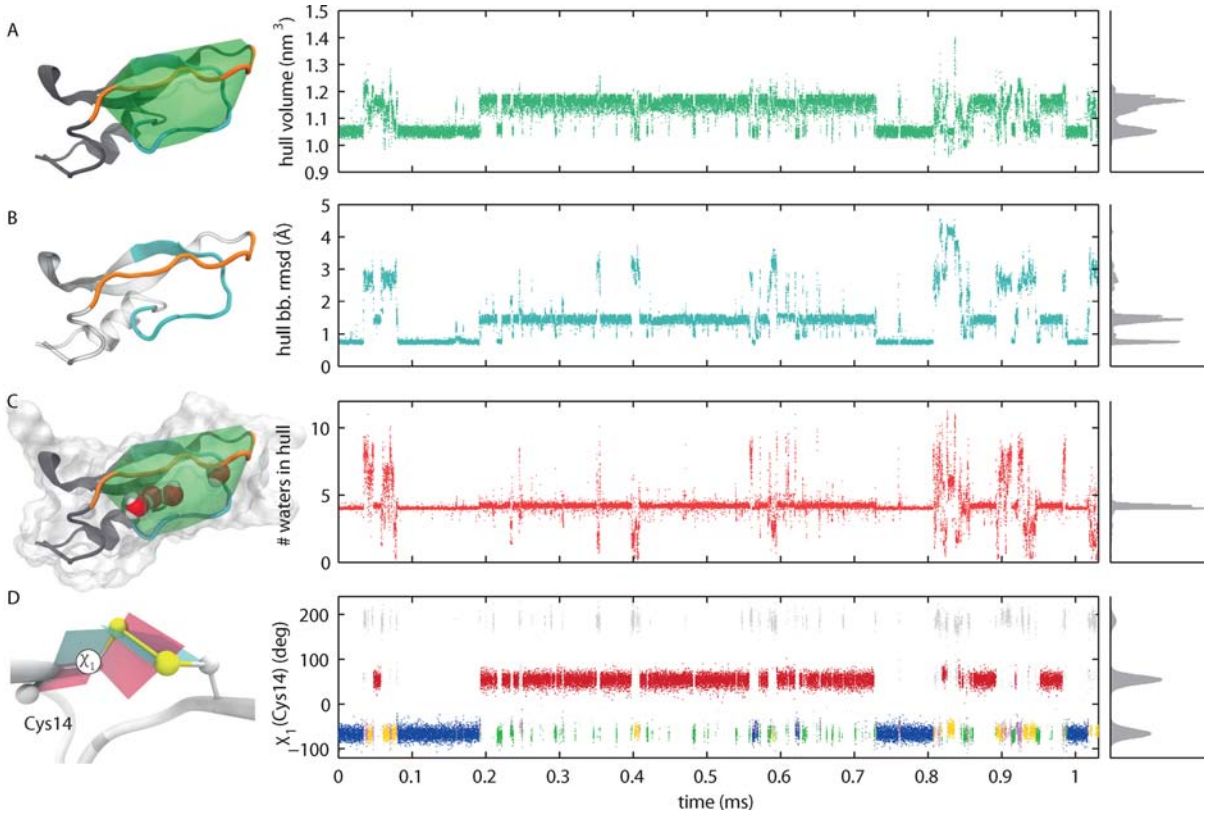


Figure 2. Fluctuations in the interloop region during 1 ms. (A) Volume of the convex hull spanning the four internal hydration sites. (B) Backbone RMSD, relative to crystal structure 5pti, for the residues defining the hull. (C) Number of water molecules inside the hull. (D) Dihedral angle χ_1(C14), with disulfide isomeric states color-coded: M1 (blue), M2 (magenta), M3 (yellow), m_{C14} (red), m_{C38} (green), and other (gray). The time series values plotted here are either averaged over consecutive 25 ns windows (A–C) or sampled with 25 ns resolution (D). The full-trajectory distributions are projected on the right-hand axis.

Table 2. Water Penetration of Interloop Region

state	no. waters
M1	4.0 ± 0.3
M2	6.3 ± 2.2
M3	4.0 ± 2.3
m_{C14}	4.2 ± 0.6
m_{C38}	4.2 ± 1.0
other	4.6 ± 2.7

22.5% of the time. This discrepancy is largely a consequence of the conformational bias of the force field. In the experimentally dominant M1 state, the vacancy fraction is <1% for the deep sites and 7% for site W111 (Tables S5–S8). For the external sites W110 and W143, the vacancy fraction is much higher: 60 and 67%, respectively. This is partly due to the higher flexibility of these sites, allowing for frequent violations of the site H-bond criteria. Moreover, since the typical water residence time in these sites is only ∼1 ns, a substantial fraction of the frames represent transition states where the site accommodates two water molecules, neither of which satisfies the H-bond criteria for the site. A quantitative study of water exchange from external hydration sites would require a denser sampling (than 0.25 ns) of the trajectory.

To analyze water exchange kinetics, we delete the vacant frames from the $\mathbf{A}_{\alpha,0}$ vector. This deletion has no effect on the residence time statistics. From the resulting occupancy vector $\mathbf{A}_{\alpha}$, containing more than 3 million frames, we obtain a time series of exchange events that constitutes a stationary point process.[36−39] An equivalent representation of the point process is the sequence of RTs, that is, the time intervals between successive exchange events. We thus construct the RT vector $\mathbf{V}_{R,\alpha}$, a time series of $N_{R,\alpha}$ RTs and $N_{F,\alpha}$ frames, from which we compute the RT histogram $F_{R,\alpha}(n)$, giving the number of RTs of length n frames.[39]

Because experiments are not synchronized with the exchange events, the experimentally relevant quantity is not the RT but the ST, that is, the time, from an arbitrary starting point, to the next exchange. But the RT serves as a sensitive indicator of dynamical disorder in the computational analysis. Moreover, since the RT and ST distributions are not independent,[36,38,39] the RT histogram can be used to compute the ST statistics at a fraction of the computational cost of conventional algorithms (Section 2.1).[39]

From the RT histogram, we compute (the discretized versions of) the RCF $Q_{R,\alpha}(\tau)$ and the SCF $Q_{S,\alpha}(\tau)$ as well as their time integrals, the mean RT $\tau_{R,\alpha}$, and the mean ST $\tau_{S,\alpha}$ (Section 2.1).[39] The RCF is the fraction of RTs that are longer than τ, and the SCF is the probability that a water molecule residing in the site at a randomly chosen time point does not leave the site in the subsequent time interval τ. If the RTs are independent (uncorrelated) and exponentially distributed, the point process reduces to a Poisson process,[36−38] for which $\tau_{S,\alpha} = \tau_{R,\alpha}$ and $Q_{S,\alpha}(\tau) = Q_{R,\alpha}(\tau) = \exp(-\tau/\tau_{S,\alpha})$.

Figure 3 (top row) shows the RCF and SCF for the four internal hydration sites. When all conformational states are included, the RCF decays mostly on a much shorter time scale than the SCF and neither decays exponentially. This difference

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

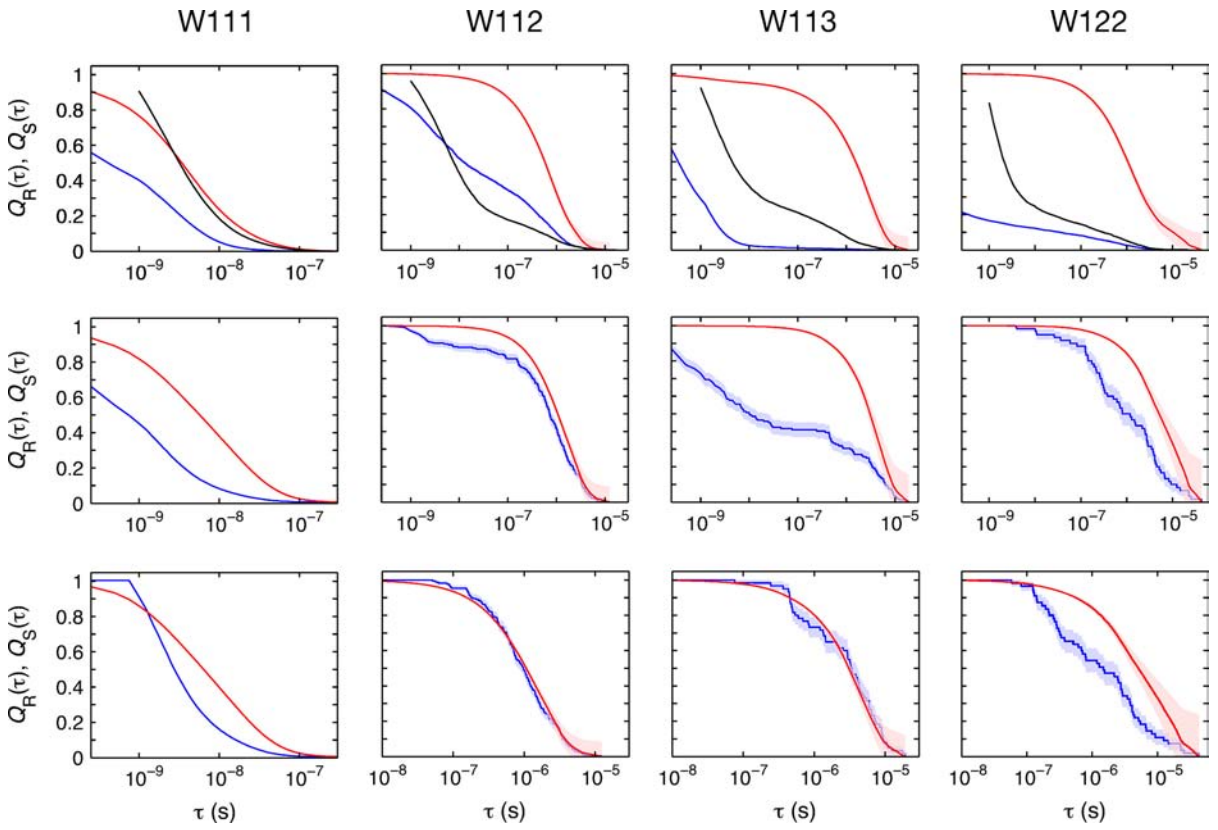


Figure 3. SCF (red) and RCF (blue) for the four internal hydration sites in BPTI. Top row: all conformational states. The black curve is the RCF reported by Shaw et al.[20] Middle row: M1 state (M1/E7cl for W111). Bottom row: states as middle row but with all RTs shorter than 50 ns (1 ns for W111) excluded. The shaded areas represent the estimated statistical error due to the finite length of the trajectory.

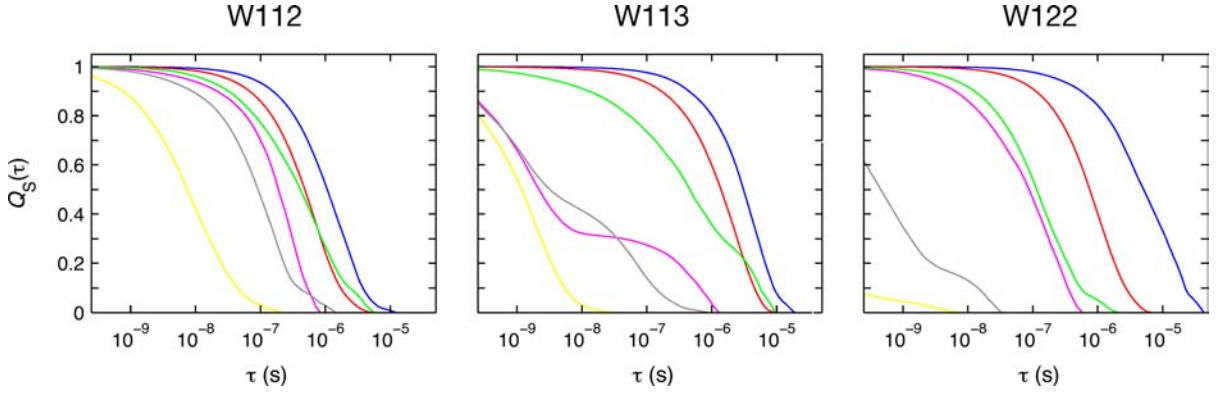


Figure 4. Survival correlation function for deep hydration sites in BPTI in each of the six C14−C38 disulfide isomeric states: M1 (blue), M2 (magenta), M3 (yellow), m_{C14} (red), m_{C38} (green), and other (gray).

is most pronounced for sites W113 and W122, where the ratio $\tau_{S,\alpha}/\tau_{R,\alpha}$ is 104 and 35, respectively (Tables S6 and S7). This strong deviation from Poisson statistics, which we refer to as dynamical disorder,[18] is caused partly by conformational heterogeneity with state-dependent water exchange rates and partly by multiple exchange mechanisms in a given state (Section 3.7).

To identify the physical basis of the dynamical disorder, we constructed state-specific subtrajectories by assigning each RT to the most abundant state in that RT (Section 2.1). Because the water exchange rate depends on the conformational state, the distribution of RTs over states (Table S3) can differ greatly from the distribution of frames over states (Table 1), that is, the state populations. For example, while state M1 accounts for ∼25% of the frames, it accounts for <1% of the RTs in sites W113 and W122.

Computing the SCF for each site in each state, we find that the water exchange kinetics for the deep sites depend strongly on the isomeric state of the C14−C38 disulfide bond (Figure 4). In the experimentally relevant M1 state, much of the difference between the RCF and SCF has disappeared (Figure 3, middle row), and the ratio $\tau_{S,\alpha}/\tau_{R,\alpha}$ is now only 2.2 for W113 and W122 (Tables S7 and S8). Non-negative least-squares (NNLS) analysis (Section 2.1) shows that the SCF can be accurately reconstructed with one (W113) or two (W112 and W122) exponential components (Table S10, Figures S4−S6), and the NNLS-based mean ST $\langle\tau_{S,n}\rangle = \sum_n f_n \tau_{S,n}$ (Table S10) is merely 4−5% larger than the mean ST $\tau_{S,\alpha}$ computed directly

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

from the RT histogram (Tables S6–S8), well within the statistical error of the latter. In both the biexponential cases, the slower component dominates, with 90 and 75% relative weight for W112 and W122, respectively (Table S10). The SCF components in these sites can be linked to kinetically distinct exchange mechanisms (Section 3.7).

Even though the SCF for site W113 is essentially single exponential in state M1 (Figure S5), it differs markedly from the RCF (Figure 3). The origin of this difference is revealed by examining the RT histogram, which is clearly bimodal for site W113 (Figure S8). The three deep sites have similar numbers (within a factor 3) of RTs in state M1, but site W113 deviates by having a large fraction short RTs. Thus, the fraction RTs < 50 ns is 58% for W113, compared to 15% for W112 and 8% for W122. Furthermore, W113 has 13% single-frame (0.25 ns) RTs, while the other two sites have none (Tables S6–S8). If RTs shorter than 50 ns are excluded from the analysis, the RCF and SCF are identical within the statistical error for W112 and W113 (Figure 3, bottom row), as are the mean RT and ST. For site W122, there is still a significant difference between the RCF and SCF (Figure 3, bottom row), and $\tau_{S,\alpha}$ is a factor 2 longer than $\tau_{R,\alpha}$ (9.0 versus 4.4 μs), presumably a result of multiple exchange mechanisms (Section 3.7).

The short RTs have virtually no effect on the SCF or on $\tau_{S,\alpha}$ for any of the deep sites, because the ST statistics are strongly biased toward the longer RTs due to the overwhelming probability that a randomly chosen initial time falls in a long RT. Since experiments probe the SCF rather than the RCF, it follows that dynamical disorder will escape detection if it is associated with a subset of short RTs.

Figure 3 (top row) also includes the 'survival probability' reported by Shaw et al.,[20] which, in fact, is the RCF. The substantial deviation from our RCF can be attributed to the different criteria used to define hydration site occupancy. Specifically, Shaw et al. used two different occupancy cutoffs for each site, thereby eliminating many short RTs. As a result, their RCF decays more slowly than ours. If the experimentally relevant SCF had been computed with the site definition of Shaw et al., it would presumably be very similar to our SCF. In other words, for long-lived hydration sites like W112, W113, and W122, the details of the site definition essentially only affect the short-RT part of the RT histogram, with little or no effect on the experimentally relevant ST statistics.

3.4. EMOR Model Validation. Before confronting the results of our simulation analysis with experiment, we shall use the MD data to quantitatively assess the three approximations inherent in the EMOR model[21] used to interpret the experimental MRD data: the Poisson, the instantaneous randomization, and the point process approximations.

The Poisson approximation assumes that the SCF decays exponentially, as in eq 8. As we have seen, this is the case for site W113, while for sites W112 and W122, there is a second component with relative weight 10 or 25% and shorter decay time (Table S10). For site W111, the four dominant (>5%) NNLS components in the SCF have decay times in the range 2–26 ns. The multiexponential form of the SCF, if real, would not have justified introducing further components in the analysis of the experimental MRD data, since, for site W111, the shorter decay times correspond to dispersions in the less densely sampled high-frequency tail of the MRD profile (Section 3.5). The correlation time for W111 deduced from the MRD fit should thus be regarded as an effective mean ST, averaged over the several exchange mechanisms that appear to operate for this least buried site.

For the deep sites, the Poisson approximation is essentially validated, although the minor (25%) W122 component with a decay time of 2 μs (Table S10) would, if real, 'contaminate' the combined W112 + W113 component (Section 3.5). However, the experimentally derived[14,15] occupancies and order parameters for the deep sites are consistent with exponential SCFs. In other words, the MRD data do not indicate any 'missing amplitude', as expected if the SCF were markedly multiexponential and if there were an overlap between the decay time spectra for site W122 and sites W112 + W113.

The instantaneous randomization approximation, inherent in eq 7, assumes that, once a water molecule has left a site, it is no longer correlated with that site. In other words, it has the same probability of returning to the site as any of the other 4214 water molecules in the simulated system. In reality, the probability that a recently exchanged water molecule returns to the site exceeds 1/4215 but remains below 0.02 at all times and for all 4 internal hydration sites (Figure 5). At short times

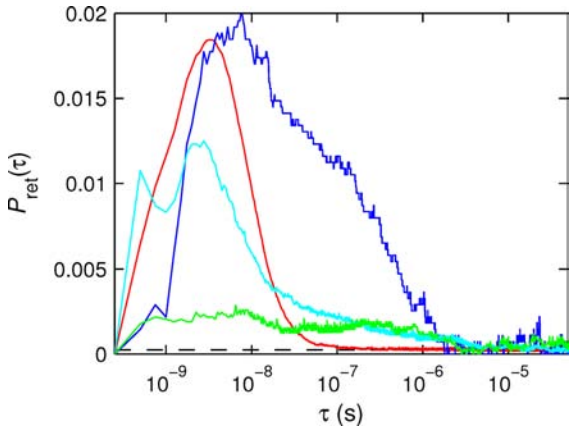


Figure 5. Return probability for the internal hydration sites in BPTI (all conformational states): W111 (red), W112 (blue), W113 (cyan), and W122 (green). The dashed line is the asymptotic value $P_{ret}(\tau \rightarrow \infty) = 1/N_W$.

(say $<10^{-8}$ s), the enhanced return probability can be attributed simply to spatial proximity: the water molecule has not yet 'mixed' completely with the other water molecule. At long times, the enhanced return probability reflects exchanges from one internal site to another and back again. This phenomenon is most pronounced for site W112 (Figure 5), where the water molecule can 'jump' to either of the flanking sites W111 or W113 before returning to W112. However, this is still a rare occurrence (Section 3.7).

Site correlation may be taken into account by replacing, in eq 7, the SCF $Q_{S,\alpha}(\tau)$ by the total survival correlation function (tSCF) $Q_{tS,\alpha}(\tau)$, which is the probability that the same water molecule resides in site α at the two time points t and $t + \tau$, regardless of its whereabouts in the meantime. To assess the importance of site correlation, we compare the mean total ST $\tau_{tS,\alpha}$, the time integral of the normalized tSCF (Section S6, Supporting Information), with the mean ST $\tau_{S,\alpha}$ (for a single visit to the site). As seen from Tables S5–S8, $\tau_{tS,\alpha}$ exceeds $\tau_{S,\alpha}$ by at most 0.5% for the 4 internal sites. The instantaneous randomization approximation is thus highly accurate for these sites.

The analogous total RT $\tau_{tR,\alpha}$ exceeds $\tau_{R,\alpha}$ by <1% for sites W111 and W112, by 9% for W113, and by 25% for W122 (Tables S5–S8). The relatively large difference for sites W113 and W122 can be attributed to the large fraction of RTs associated with minor conformational states (M3 and 'other') for these sites (Table S3). (Our site correlation analysis includes all conformational states.) However, only $Q_{tS,\alpha}(\tau)$ and $\tau_{tS,\alpha}$ are relevant for assessing the accuracy of the EMOR model.

The point process approximation, also inherent in eq 7, amounts to assuming that water exchange is instantaneous. This approximation is valid if the typical duration of an exchange event is much shorter than the mean ST. For the 66 exchange events examined in detail, the lifetime of the exchange-competent transition state never exceeds 5 ns for any of the 4 internal sites (Section 3.7). For the deep sites, with $\tau_{S,\alpha} > 1$ μs, the point process approximation is thus highly accurate. For site W111, with $\tau_{S,\alpha} \approx 14$ ns in state M1, this approximation may not be quantitatively accurate, but the experimental $\tau_{S,\alpha}$ is considerably longer (Section 3.5).

3.5. Simulation versus Experiment. Having validated the three main approximations in the EMOR model, we can confidently use that model to extract mean survival times (and other parameters) for the internal hydration sites from MRD experiments on immobilized BPTI.[15] By comparing these experimentally derived mean STs with the corresponding values computed directly from the MD trajectory, we can assess the accuracy of the molecular mechanics force field used in the simulation.

The Shaw et al. 1 ms BPTI simulation[20] used the isoleucine-modified version[51] of the all-atom AMBER ff99SB force field,[52] which compares favorably among the common protein force fields,[41,42] even though it does not reproduce the C14–C38 disulfide isomeric equilibria in BPTI quantitatively (Section 3.2).[20,34] Water interactions were described with the TIP4P-Ew water model, which, in contrast to the commonly used TIP3P model, yields bulk-water dynamics in excellent agreement with experiment.[53,54] The combination AMBER ff99SB plus TIP4P-Ew has been validated with respect to conformational bias in tripeptides[55] and unfolding stability of the Trp-cage fold.[56]

The previously reported[15] water ^{2}H MRD profile from immobilized BPTI in D_2O at pD 6.5 and 20 °C is shown in Figure 6. The protein was immobilized with glutaraldehyde, which forms stable cross-links of variable length with (some of) the five amino groups on the protein surface. The EMOR model[21] with three components (Section 2.2) provides an excellent fit to the data, with parameter values given in Table 3. Minor differences from the parameter values reported previously[15] result from the slightly more accurate numerical implementation of the EMOR theory and the more robust optimization algorithm used here (Section 2.2).

According to the EMOR model fit, the relaxation dispersion is dominated by a kinetic component with $\tau_{S,1} = 5.4 \pm 1.1$ μs. The occupancy, $\xi_1 = 1.8 \pm 0.4$, of this component suggests that it represents two sites, and the magnitude of $\tau_{S,1}$ identifies these sites as W112 and W113.[15] Before comparing experiment with simulation, we note that the MRD experiments pertain to 293 K and D_2O solvent, whereas the simulation refers to 300 K and H_2O solvent. To correct for the temperature difference, we use the experimentally determined activation energy,[15] $\Delta E_{A,1} = 43$ kJ mol^{-1}. To (partially) correct for the solvent H/D isotope effect, we divide the experimental $\tau_{S,1}$ by the bulk-water viscosity ratio $\eta(D_2O)/\eta(H_2O) = 1.225$ (at 300 K). We thus obtain the experimental result $\tau_{S,1} = 2.9 \pm 0.6$ μs, corrected to 300 K and H_2O.

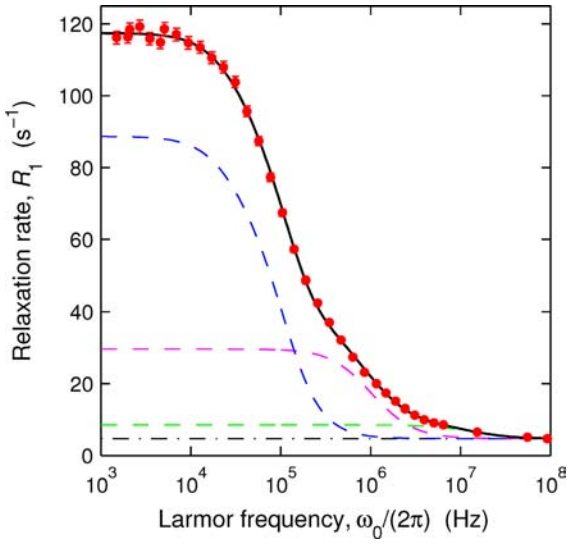


Figure 6. Water ^{2}H MRD profile from immobilized BPTI at 20 °C and pD 6.5. The red circles are the data, the black solid curve is the EMOR fit, the dashed curves are the three EMOR components (1 = blue, 2 = magenta, 3 = green), and the dash-dotted line is the external-water contribution (also included in the components).

Table 3. Results of EMOR Fits to ^{2}H MRD Profile[a]

parameter (unit)	unconstrained	$\xi_1 = 2$ fixed
ξ_1	1.8 ± 0.4	2
$\tau_{S,1}$ (μs)	5.4 ± 1.1	5.8 ± 0.4
S_1	0.93 ± 0.06	0.93 ± 0.06
η_1	0.33 ± 0.26	0.24 ± 0.06
$\xi_2 S^2_{iso,2}$	1.23 ± 0.04	1.23 ± 0.04
$\tau_{S,2}$ (ns)	88 ± 4	87 ± 4
$\xi_3 S^2_{iso,3}$	2.4 ± 0.2	2.4 ± 0.2
$\tau_{S,3}$ (ns)	6.8 ± 0.8	6.8 ± 0.8
$10^{-3} \times \nu_{dyn}$[b]	2.6 ± 0.1	2.6 ± 0.1
χ^2_{red}	1.50	1.47

[a]Quoted errors correspond to one standard deviation, propagated from the uniform 1.5% uncertainty in R_1. [b]Dynamic hydration number is obtained from the fitted parameter $R_{1,ext}$ as $\nu_{dyn} = N_W(R_{1,ext}/R_{1,bulk} - 1)$.

For the simulation, we consider only the M1 disulfide isomeric state, which accounts for ~95% of the experimental protein population.[32] For this state, we find $\tau_{S,W112} = 1.7 \pm 0.2$ μs and $\tau_{S,W113} = 4.1 \pm 0.5$ μs (Tables S6 and S7). The difference between these values is too small to be resolved experimentally, consistent with the occupancy, $\xi_1 = 1.8 \pm 0.4$, deduced from the fit. The appropriate quantity to compare with is therefore the arithmetic average $\tau_{S,1} = (\tau_{S,W112} + \tau_{S,W113})/2 = 2.9 \pm 0.3$ μs. The exact agreement with the (corrected) experimental value may be fortuitous to some extent, but it does suggest that the force field is not far off the mark. Taking the experimental and computational uncertainties into account, we can express the agreement in terms of an activation energy discrepancy: $\Delta\Delta E_{A,1} = 0.0 \pm 0.2$ k_BT.

The second component obtained from the EMOR model fit has $\tau_{S,2} = 88 \pm 4$ ns and an amplitude parameter consistent with single occupancy. This component was assigned to the least deeply buried internal hydration site W111.[15] Correcting for the temperature difference (with $\Delta E_{A,2} = 16$ kJ mol^{-1})[15] and

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735–8748

the H/D solvent isotope effect as before, we obtain $\tau_{S,2} = 61 \pm 3$ ns. The simulation yields $\tau_{S,2} = 13.7 \pm 0.6$ ns for the M1 state (Table S5), so the activation energy discrepancy becomes $\Delta\Delta E_{A,2} = 1.5 \pm 0.1\ k_BT$. Using instead the mean ST $\tau_{S,2} = 50 \pm 3$ ns obtained from ^{17}O MRD data for immobilized BPTI in H_2O at 293 K,[15] we obtain (after temperature correction) $\Delta\Delta E_{A,2} = 1.1 \pm 0.1\ k_BT$. NNLS analysis reveals that the simulated SCF $Q_{S,W111}(\tau)$ for state M1 does not decay exponentially (as assumed in the EMOR model) but can be decomposed into 5 significant components with decay times in the range 2−73 ns. (This is true also for substate M1/E7cl.) If this dynamical disorder is real, it might account for part of the (modest) discrepancy between experiment and simulation for site W111.

For site W122, a difference-MRD experiment on wild-type BPTI and the G36S mutant (where site W122 is occupied by the hydroxyl group of Ser-36, rather than by a water molecule), and involving both the ^2H and ^{17}O nuclides, yielded $\tau_{S,W122} = 170 \pm 20\ \mu s$ at 300 K and in D_2O[14] or $139 \pm 16\ \mu s$ after H/D viscosity correction. (At 293 K, $\tau_{S,W122} = 400\ \mu s$, implying that the contribution of site W122 to the ^2H MRD profile in Figure 6 would be less than half of the experimental uncertainty in R_1.) The simulation yields for state M1 $\tau_{S,W122} = 9.0 \pm 2.3\ \mu s$ (Table S8). (We ignore the slight deviation from exponential SCF; see Table S10.) The activation energy discrepancy for site W122 thus becomes $\Delta\Delta E_{A,W122} = 2.7 \pm 0.3\ k_BT$.

In summary, we find that the force field performs remarkably well, with activation energy discrepancies of 1.5 k_BT or less for 3 of the 4 internal sites. The nearly 2-fold larger discrepancy for W122 is probably related to the known inability of the force field to quantitatively model the isomer populations of the nearby C14−C38 disulfide bond (Section 3.2).[20,34] The ability of the force field to accurately predict the exchange rates of deeply buried water molecules may appear surprising since the TIP4P-Ew model (like most other water models) has been parametrized to reproduce bulk water properties. In particular, the dipole moment of TIP4P-Ew water is augmented by ∼25% to mimic the effect of polarizability in the bulk liquid. However, the four internal sites considered here are highly polar, with each buried water engaged in four H-bonds so the polarization effects may be similar to those in bulk water. Furthermore, since water exchange is rate limited by protein conformational fluctuations (Section 3.7), the mean survival time is not governed solely by water−protein interactions.

3.6. Hydration Site Flexibility. The polarity of the H-bonds that typically link a buried water molecule to the surrounding protein atoms effectively prevents water rotation in the site and only allows small-amplitude librations and symmetric 180° flips about the water dipole (C_2) axis. The partial averaging of the nuclear quadrupole coupling by such anisotropic motions is manifested as a reduced orientational order parameter S_α for the relaxation contributions induced by slower isotropic motions (Section 2.2). For the ubiquitous subpicosecond librations, this is the only significant effect.[57] But the C_2 flip, which in inorganic crystal hydrates can occur on time scales ranging from picoseconds to milliseconds,[58] may also give rise to a direct spin relaxation contribution. In solution MRD studies, C_2 flips slower than the tumbling time of the protein (∼ 3 ns for BPTI) have no effect.[57] But when the protein is immobilized, the order parameter reflects all internal motions during the much longer survival time of the internal water molecule, and C_2 flips on the nanosecond−microsecond time scale can give rise to distinct steps in the MRD profile.[21]

If an internal water molecule undergoes C_2 flips during its residence in the site, we expect that $|S| < 0.55$ (Supporting Information, Section S7).[21] The much larger order parameters deduced for sites W111−W113 (Table 3) therefore rule out the presence of C_2 flips on time scales shorter than or comparable to τ_S (∼ 90 ns for W111, and 5 μs for W112 and W113). For W122, which exchanges too slowly to be observed in the immobilized protein, we can rule out C_2 flips faster than ∼1 μs, since flips on the time scale $\sim 10^{-8}$ to 10^{-6} s would have been manifested as an additional dispersion step, whereas flips on a nanosecond or faster time scale would have reduced the solution-MRD order parameter.[14]

The experimentally inferred absence of C_2 flips in the internal hydration sites of BPTI provides an additional opportunity to test the force field of the MD simulation. In stark contrast to the experimental results, analysis of the MD trajectory (Supporting Information, Section S7) reveals frequent C_2 flips in all four sites, as illustrated for site W122 in Figure S10. For the least deeply buried site W111, the mean flip survival time $\tau_{S,flip} = 9$ ns is comparable to the mean ST $\tau_S = 14$ ns, so the flip motion cannot be meaningfully distinguished from the actual exchange event. But for the 3 deep sites, $\tau_{S,flip}$ is 1−2 orders of magnitude shorter than τ_S (Table 4). For site W112, the discrepancy between simulation and experiment ($\tau_{S,flip} = 14$ ns versus $\gg 5\ \mu s$) corresponds to a activation energy difference of more than 6 k_BT. This failure should probably not be attributed to the TIP4P-Ew water model, which accurately reproduces rotational and translational dynamics in bulk water.[54] More likely, the discrepancy is related to the protein force field's ability (or lack thereof) to describe the collective fluctuations of the hydration site cavity that govern the large flip barrier.

Table 4. Hydration Site Flexibility[a]

property (unit)	W112	W113	W122
S or S_{iso}[b]	−0.491	−0.501	0.539
S_{lib} or $S_{iso,lib}$[c]	0.885	0.903	0.911
$\tau_{R,flip}$ (ns)	12	56	103
$\tau_{S,flip}$ (ns)	14	73	380
τ_S (μs)	1.7	4.1	9.0

[a]MD results for state M1. [b]S quoted for W112 and W113, S_{iso} quoted for W122. [c]Librational order parameter estimated as $S_{lib} = S/S_{flip}$ with $S_{flip} = -0.555$ for W112 and W113, and as $S_{iso,lib} = S_{iso}/S_{iso,flip}$ with $S_{iso,flip} = (0.350)^{1/2}$ for W122 (Supporting Information, Section S7).

While the force field substantially underestimates the C_2 flip barrier, it fares somewhat better in the description of librational disorder. The water molecules in the deep hydration sites are highly ordered, so we can correct for the artifactual flip reduction of the order parameter (Supporting Information, Section S7) to obtain purely librational order parameters. For sites W112 and W113, the MRD experiment yields $S = 0.93 \pm 0.06$ (Table 3), which compares rather well with the MD average, $S_{lib} = 0.894$ for these sites (Table 4). For site W122, solution difference-MRD studies[14] yield $S_{iso} = 0.94 \pm 0.01$, not far from the MD result $S_{iso,lib} = 0.911$. Arguably, the relevant quantity to compare here is not S but $1 - S^2$, which increases with the libration amplitude.[59] When this is done, we find that the simulation overestimates $1 - S^2$ by 50% in both cases. In other words, the sites are somewhat too flexible, in line with the too low flip barrier. We note that all MD results discussed here pertain exclusively to the experimentally relevant M1 state. We

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

note also that any additional disorder associated with flexibility of the interprotein cross-links would make the sites appear more flexible in the experiment than in the simulation, contrary to what is found here.

3.7. Water Exchange Mechanisms. By what mechanisms do water molecules exchange from the deep sites in BPTI? Do they migrate through the protein matrix as isolated water molecules or does the internal site become connected to the exterior? Two versions of the latter scenario can be envisaged: solvent might penetrate to the internal site without significant distortion of the backbone fold or the protein might unfold to the extent that the internal site effectively becomes an external hydration site.

To shed light on this issue, we examined in detail the exchange events that preceded the 25 longest RTs in each site. As before, we only consider state M1. More precisely, we examined exchange events RT(k − 1) − X − RT(k), where both RTs k − 1 and k were assigned to state M1, but no restriction was placed on the disulfide conformation in the transition state X. Depending on the site, we characterized between 11 and 33% of the exchange events in state M1 (Table S11). With a resolution of 0.25 ns, a complete picture of the exchange process cannot be expected. Indeed, in some cases, the exchange occurred between two frames and could then not be characterized. Nonetheless, several significant observations were made.

First, the lifetime of the transition state X did not exceed 5 ns in any case. Since the mean ST is in the μs range for the deep sites, this observation validates the point process approximation in the EMOR model (Section 3.4).

Second, all examined exchange events involve the transient formation of a single-file water chain penetrating the protein via a narrow pore from the surface to the site or via a tunnel from the surface to the site and back to the surface at a different point (Figures 7 and S13). We refer to this as a transient aqueduct mechanism.

Third, the observed exchange mechanisms can be assigned to either of two types, depending on the C14−C38 isomer populated in the transition state X. In ∼ 90% of the exchanges from sites W112 and W113, the disulfide remained in isomeric state M1 in the exchange transition state X. The water chains are then largely composed of the water molecules occupying sites W111−W113, often with an additional site W112* formed above site W112 (see below). In 70% of the exchanges from site W122, the disulfide adopted the M2 isomer in the transition state X. Disulfide state M2, which is only populated in 2.8% of the frames in the entire trajectory, differs from state M1 in the mean configuration of the L1 and L2 loops (Figure S11) and in their flexibility (Figure 2). Moreover, the volume of the interloop region is ∼20% larger than in state M1 (Figure 2), and there are 6.3 ± 2.2 water molecules in this region, as compared to 4.0 ± 0.3 in state M1 (Table 2). This water influx does not represent new internal sites in addition to the ones present in state M1. Rather, state M2 features a qualitatively different hydration motif, where the interloop region is penetrated by single-file water chains in transient tunnels or pores. These penetrating water molecules rarely satisfy the occupation H-bond criteria of the original deep sites, which therefore are recorded as 'vacant' in most frames in state M2 (Tables S6−S8).

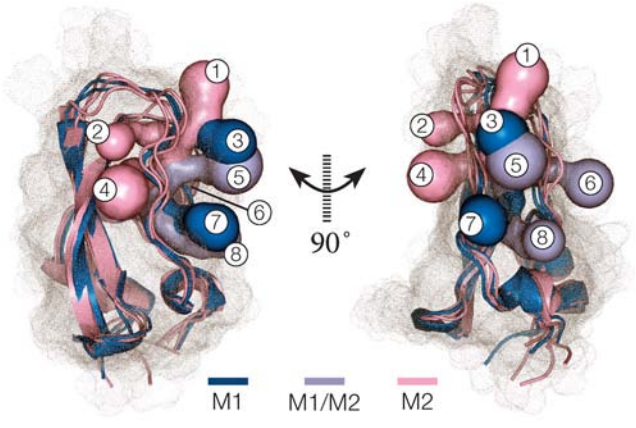


Figure 7. Transition states for the major water exchange mechanisms from the deep sites W112, W113 and W122 in states M1 or M2. Tunnels and pores, identified with Caver 3.0 and rendered with PyMol 1.5,[67] are superimposed in this dual representation (the two views are rotated by 90° about the vertical axis). The numbers label the mouths of tunnels (t) and pores (p) used by site W112 [t(5,7), t(6,7), p7]; site W113 [t(5,6), t(6,7), t(5,7/8)]; and site W122 [t(1,5,6,8), t(2,3), t(2,8), p3, p5].

Since the water molecules occupying the three adjacent sites W111−W113 are mutually H-bonded, one expects their exchange to be correlated to some extent. This is in fact observed, even though exchange from W111 is 2 orders of magnitude faster than for W112, which in turn is a factor 2.4 faster than for W113. The dominant mechanism for site W112 is a two-step interchange with site W111 (mechanism A in Figure S13). In the first step, an external water molecule enters site W111, displacing its former tenant to a transient site W112* just outside site W112 (Figure S12), where it H-bonds to the water molecules in all 3 sites W111−W113. The second step is a concerted water displacement W112* → W112 → W111 → external. As a result, the water molecules originally occupying sites W111 and W112 have now swapped places. Once in the W111 site, the former W112 tenant typically exchanges on a nanosecond time scale. This mechanism accounts for ∼90% of the observed exchanges from site W112, and it is tempting to associate it with the dominant (weight 0.90) NNLS component of the SCF for this site (Table S10). In the minor mechanism for site W112, the transient W112* site is instead populated via a tunnel leading from the external site W143.

For site W113, most of the exchanges involve concerted jumps of all the water molecules in sites W111−W113, initiated by an external water molecule that enters from either end. In a variant of this mechanism, site W112 is bypassed, and the water originally in site W113 ends up in site W111 through two pairwise interchanges: W113 ↔ W112* and W112* ↔ W111. The observation that exchange in site W113 often is coupled to exchange in site W112, whereas the reverse rarely happens, is consistent with the somewhat longer mean ST for site W113. Site W113 differs from sites W112 and W122 in having a large fraction of short RTs in state M1 (Table S7 and Figure S8). While these short RTs hardly affect the SCF probed by the MRD experiment, the effect may be real and deserves further study (with a denser sampling of the trajectory).

For site W122, 70% of the exchanges occur via the transient isomeric state M2 of the nearby C14−C38 disulfide bond, possibly corresponding to the dominant (weight 0.75) NNLS component for this site (Table S10). The exchange mechanisms involve water chains in tunnels or pores penetrating the interloop region (Figures 7 and S13). The

dx.doi.org/10.1021/ja403405d | J. Am. Chem. Soc. 2013, 135, 8735−8748

M1 → M2 transition accompanying water exchange from site W122 in the dominant mechanism was referred to as a 'revolving door' mechanism in a recent accelerated MD study.[35]

Although we observe concerted exchange processes, such as W113 also involving W112 or either of these also involving W111, repeated water exchanges back and forth among the internal sites are a rare occurrence. If such back-and-forth shuttling were more prevalent, the number of unique visitors to these sites would drop markedly below what is expected by chance. If the N_{RT} visits to a given site are uncorrelated in the sense that the next visitor is equally likely to be any of the N_W = 4215 water molecules in the system, then the expected number of unique visitors is $N_W[1 - (1 - 1/N_W)^{N_{RT}}]$. For site W122, the 59 unique visitors (in state M1) closely matches the statistical prediction 59.6 (Table S11), consistent with the very small return probability $P_{ret}(\tau)$ for this site (Figure 5). For sites W112 and W113, the number of unique visitors is reduced by 4 and 7%, respectively, from the chance expectation (Table S11), consistent with the somewhat larger return probability for these sites (Figure 5).

The aqueduct mechanism is consistent with the current understanding of the thermodynamics of water partitioning between the protein interior and the bulk solvent.[60] As a rule of thumb, water molecules occupy internal sites where three or four H-bonds (per water molecule) can be maintained, but not where they can engage in less than two H-bonds. A single-file water chain, with two H-bonds per water molecule, represents a borderline case, where the delicate balance may be controlled by weaker interactions with the local environment. Single-file water chains can penetrate carbon nanotubes[61] and may also form transiently in lipd bilayers.[62,63] In proteins, water chains are found in membrane proteins, notably aquaporins,[64] and as proton conduction pathways in enzymes.[65]

The water chains observed here in BPTI are rare and transient configurations of relatively high free energy. In the M1−M1−M1 mechanism that dominates for W111−W113, the backbone fold is not significantly perturbed during the exchange so the pore or tunnel must be formed by cooperative side-chain rearrangements. In the M1−M2−M1 mechanism that dominates for W122, the backbone of the 'upper' part (residues 11−14) of the L1 loop bulges out in the M2 transition state (Figure S11), but this can hardly be described as even a local unfolding. The phenyl ring of Tyr-35, which is sandwiched between the W113 and W122 sites, does not flip during the water exchange. Indeed, no ring flip is observed for Tyr-35 during the entire 1 ms trajectory,[20] consistent with the experimentally determined ∼20 ms flip time.[66]

Like internal-water exchange, amide hydrogen exchange also requires transient solvent access to interior sites.[5−7] However, in amide hydrogen exchange, not only water molecules but also a catalytic ion (usually OH^-) must access the site. The high-free-energy conformations sampled by amide hydrogen exchange are therefore expected to be more extensively perturbed than the transition state for internal-water exchange. Whether some variant of the transient aqueduct mechanism plays a role also in amide hydrogen exchange remains to be seen.

4. CONCLUSIONS

Molecular recognition and enzymatic catalysis often take place at sites that are buried inside the protein. A key step in the functional cycle of such proteins is the transient access of water and other small molecules to the protein interior. Because this is a rare event, it is challenging to study experimentally. Water ^2H and ^{17}O MRD experiments have furnished quantitative information about the rates at which water molecules access internal hydration sites in many proteins. For BPTI, the mean survival times of water molecules have been determined at all four internal hydration sites. Yet, MRD measurements cannot reveal the mechanism of water exchange nor can they, without laborious difference experiments, identify the hydration sites. Furthermore, the EMOR model used to interpret the MRD data is based on several simplifying assumptions, which, if violated, could compromise the interpretation. Fortunately, these concerns can all be addressed with MD simulations.

The relationship between experiment and computer simulation is a symbiotic one, where each stands to gain from the other. Simulations can be used to validate and, if necessary, refine the theoretical model used to interpret the data and they can greatly expand the information content by providing a wealth of atomic detail. On the other hand, even if run on powerful hardware that allows exhaustive sampling of conformational space, a simulation is never better than the compromises that have gone into the semiempirical force field on which it is based. Bench-marking against reliable experimental data is therefore vital to the future well-being of the simulation enterprise.

The present study illustrates the power of state-of-the-art MD simulations in validating and extending experimental results. At the same time, it identifies certain short-comings in the current generation of force fields. The principal conclusions of this work are as follows:

(1) Water exchange from the deep internal sites of BPTI can be accurately described as a point process with exponential survival correlation function and negligible site correlation. The EMOR model is thus quantitatively justified in this case.
(2) The mean survival times, in the range 2−5 μs, of hydration sites W112 and W113 are quantitatively reproduced by the simulation. For the other two sites, the discrepancy, expressed as an activation energy offset, is 1.5−3 k_BT.
(3) Even when the analysis is restricted to the experimentally relevant disulfide isomer M1, the simulated hydration sites are somewhat too flexible, and the barrier governing the symmetric 180° water flip is too low by as much as 6 k_BT.
(4) Water molecules gain access to the internal sites by a transient aqueduct mechanism, migrating as single-file water chains through transient (<5 ns) tunnels or pores.

■ ASSOCIATED CONTENT

Supporting Information

List of protein atoms constituting persistent hydration sites (Table S1); hydration matrices for site W122 in 6 different disulfide states (Figure S1); dihedral angle ranges for disulfide isomeric states (Table S2); scatter plots of C14−C38 dihedrals and Gaussian decomposition of the M states (Figure S2); time series plot of E7 conformation relative to site W111 (Figure S3); hydration site statistics (Tables S3 − S9); NNLS decomposition of SCF for the deep sites (Table S10, Figures S4 − S6); RT vector and histogram for the deep sites (Figures S7 − S9); theoretical analysis of site correlation and return probability (Section S6); theoretical results and computational algorithm for site disorder and C_2 flips (Section S7); time series

plot of water C_2 flips in site W122 (Figure S10); exchange mechanism characteristics for the deep sites (Table S11); backbone superposition for states M1 and M2 (Figure S11); snapshot of exchange event in site W112 (Figure S12); representative selection of transition states for water exchange in the deep sites (Figure S13). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

bertil.halle@bpc.lu.se

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank D. E. Shaw Research for sharing the BPTI trajectory, Paul Maragakis for helpful communication, and the Swedish Research Council for financial support.

REFERENCES

(1) Liang, J.; Dill, K. A. *Biophys. J.* **2001**, *81*, 751−766.
(2) Seeliger, D.; de Groot, B. L. *Proteins* **2007**, *68*, 595−601.
(3) Kharakoz, D. P. *Biophys. J.* **2000**, *79*, 511−525.
(4) Mori, K.; Seki, Y.; Yamada, Y.; Matsumoto, H.; Soda, K. *J. Chem. Phys.* **2006**, *125* (054903), 1−13.
(5) Hvidt, A.; Nielsen, S. O. *Adv. Protein Sci.* **1966**, *21*, 287−386.
(6) Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1983**, *16*, 521−655.
(7) Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W. *Protein Sci.* **2012**, *21*, 987−995.
(8) Lakowicz, J. R.; Weber, G. *Biochemistry* **1973**, *12*, 4171−4179.
(9) Calhoun, D. B.; Vanderkooi, J. M.; Woodrow, G. V.; Englander, S. W. *Biochemistry* **1983**, *22*, 1526−1532.
(10) Strambini, G. B.; Gonnelli, M. *Biochemistry* **2011**, *50*, 970−980.
(11) Weber, B. H.; Storm, M. C.; Boyer, P. D. *Arch. Biochem. Biophys.* **1974**, *163*, 1−6.
(12) Edsall, J. T.; McKenzie, H. A. *Advan. Biophys.* **1983**, *16*, 53−183.
(13) Tüchsen, E.; Hayes, J. M.; Ramaprasad, S.; Copie, V.; Woodward, C. *Biochemistry* **1987**, *26*, 5163−5172.
(14) Denisov, V. P.; Peters, J.; Hörlein, H. D.; Halle, B. *Nat. Struct. Biol.* **1996**, *3*, 505−509.
(15) Persson, E.; Halle, B. *J. Am. Chem. Soc.* **2008**, *130*, 1774−1787.
(16) Szabo, A.; Shoup, D.; Northrup, S. H.; McCammon, J. A. *J. Chem. Phys.* **1982**, *77*, 4484−4493.
(17) Agmon, N.; Hopfield, J. J. *J. Chem. Phys.* **1983**, *78*, 6947−6959.
(18) Zwanzig, R. *Acc. Chem. Res.* **1990**, *23*, 148−152.
(19) Zhou, H.-X.; Wlodek, S. T.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9280−9283.
(20) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−346.
(21) Nilsson, T.; Halle, B. *J. Chem. Phys.* **2012**, *137* (054503), 1−15.
(22) Lounnas, V.; Pettitt, B. M. *Proteins* **1994**, *18*, 133−147.
(23) Bakowies, D.; van Gunsteren, W. F. *Proteins* **2002**, *47*, 534−147.
(24) Henchman, R. H.; McCammon, J. A. *J. Comput. Chem.* **2002**, *23*, 861−869.
(25) Sanjeev, B. S.; Vishveshwara, S. *Proteins* **2004**, *55*, 915−923.
(26) Soda, K.; Shimbo, Y.; Seki, Y.; Taiji, M. *Biophys. Chem.* **2011**, *156*, 31−42.
(27) Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. *J. Mol. Biol.* **1984**, *180*, 301−329.
(28) Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr. D* **1996**, *52*, 18−29.
(29) Ascenzi, P.; Bocedi, A.; Bolognesi, M.; Spallarossa, A.; Coletta, M.; De Cristofaro, R.; Menegatti, E. *Curr. Protein Pept. Sci.* **2003**, *4*, 231−251.
(30) Hanson, W. M.; Domek, G. J.; Horvarth, M. P.; Goldenberg, D. P. *J. Mol. Biol.* **2007**, *366*, 230−243.
(31) Otting, G.; Liepinsh, E.; Wüthrich, K. *Biochemistry* **1993**, *32*, 3571−3582.
(32) Grey, M. J.; Wang, C.; Palmer, A. G. *J. Am. Chem. Soc.* **2003**, *125*, 14324−14335.
(33) Long, D.; Brüschweiler, R. *J. Am. Chem. Soc.* **2011**, *133*, 18999−19005.
(34) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2012**, *134*, 2555−2562.
(35) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 2997−3002.
(36) Cox, D. R.; Miller, H. D. *The Theory of Stochastic Processes*; Chapman & Hall: London, 1965.
(37) Cox, D. R.; Isham, V. *Point Processes*; Chapman & Hall: London, 1980.
(38) Daley, D. J.; Vere-Jones, D. *An Introduction to the Theory of Point Processes*, 2nd ed.; Springer: New York, 2003; Vol. *I*.
(39) Halle, B.; Persson, F. *J. Chem. Theory Comput.* 2013, in press (DOI: 10.1021/ct400161u).
(40) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Phys. Chem.* **1983**, *87*, 5071−5083.
(41) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E *Plos One* **2012**, *7* (e32131), 1−6.
(42) Cino, E. A.; Choy, W.-Y.; Karttunen, M. *J. Chem. Theory Comput.* **2012**, *8*, 2725−2740.
(43) Lawson, C. L.; Hanson, R. J. *Solving Least Squares Problems*; Prentice-Hall: Englewood Cliffs, NJ, 1974.
(44) Whittall, K. P.; MacKay, A. L. *J. Magn. Reson.* **1989**, *84*, 134−152.
(45) Abragam, A. *The Principles of Nuclear Magnetism*; Clarendon: Oxford, 1961.
(46) Halle, B. *J. Chem. Phys.* **2009**, *131* (224507), 1−22.
(47) Denisov, V. P.; Peters, J.; Hörlein, H. D.; Halle, B. *Biochemistry* **2004**, *43*, 12020−12027.
(48) Schiffer, C. A.; van Gunsteren, W. F. *Proteins* **1999**, *36*, 501−511.
(49) Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4793−4798.
(50) Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16247−16252.
(51) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E *Proteins* **2010**, *78*, 1950−1958.
(52) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(53) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(54) English, N. J. *Mol. Phys.* **2005**, *103*, 1945−1960.
(55) Nerenberg, P. S.; Head-Gordon, T. *J. Chem. Theory Comput.* **2011**, *7*, 1220−1230.
(56) Paschek, D.; Day, R.; Garcia, A. E. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19840−19847.
(57) Denisov, V. P.; Venu, K.; Peters, J.; Hörlein, H. D.; Halle, B. *J. Phys. Chem. B* **1997**, *101*, 9380−9389.
(58) Larsson, K.; Tegenfeldt, J.; Hermansson, K. *J. Chem. Soc. Faraday Trans.* **1991**, *87*, 1193−1200.
(59) Brüschweiler, R.; Wright, P. E. *J. Am. Chem. Soc.* **1994**, *116*, 8426−8427.
(60) Yu, H.; Rick, S. W. *J. Phys. Chem. B* **2010**, *114*, 11552−11560.
(61) Rasaiah, J. C.; Garde, S.; Hummer, G. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713−740.
(62) Jansen, M.; Blume, A. *Biophys. J.* **1995**, *68*, 997−1008.
(63) Vernier, P. T.; Levine, Z. A.; Gundersen, M. A. *Proc. IEEE* **2013**, *101*, 494−504.
(64) de Groot, B. L.; Grubmüller, H. *Curr. Opin. Struct. Biol.* **2005**, *15*, 176−183.

(65) Kaila, V. R. I.; Hummer, G. *Phys. Chem. Chem. Phys. 13*, 13207 − 13215.
(66) Wagner, G.; Brühwiler, D.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 227−231.
(67) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. *PLoS Comput. Biol.* **2012**, *8* (e1002708), 1−12.