How amide hydrogens exchange in native proteins

Filip Persson and Bertil Halle

Department of Biophysical Chemistry, Center for Molecular Protein Science, Lund University, SE-22100 Lund, Sweden

Edited by Irwin D. Kuntz, University of California, San Francisco, CA, and accepted by the Editorial Board June 10, 2015 (received for review March 27, 2015)

Amide hydrogen exchange (HX) is widely used in protein biophysics even though our ignorance about the HX mechanism makes data interpretation imprecise. Notably, the open-exchange-competent conformational state has not been identified. Based on analysis of an ultralong molecular dynamics trajectory of the protein BPTI, we propose that the open (O) states for amides that exchange by subglobal fluctuations are locally distorted conformations with two water molecules directly coordinated to the N–H group. The HX protection factors computed from the relative O-state populations agree well with experiment. The O states of different amides show little or no temporal correlation, even if adjacent residues unfold cooperatively. The mean residence time of the O state is ~100 ps for all examined amides, so the large variation in measured HX rate must be attributed to the opening frequency. A few amides gain solvent access via tunnels or pores penetrated by water chains including native internal water molecules, but most amides access solvent by more local structural distortions. In either case, we argue that an overcoordinated N–H group is necessary for efficient proton transfer by Grotthuss-type structural diffusion.

Before the tightly packed and densely H-bonded structure of globular proteins had been established, Hvidt and Lindnerstrøm-Lang (1) showed that all backbone amide hydrogens of insulin exchange with water hydrogens, implying that all parts of the polypeptide backbone are, at least transiently, exposed to solvent. In the following 60 y, hydrogen exchange (HX), usually monitored by NMR spectroscopy (2) or mass spectrometry (3), has been widely used to study protein folding and stability (4–10), structure (11, 12), flexibility and dynamics (13–15), and solvent accessibility and binding (16, 17), often with single-residue resolution. However, because the exchange mechanism is unclear, HX data from proteins can, at best, be interpreted qualitatively (18–25).

Under most conditions, amide HX is catalyzed by hydroxide ions (26, 27) at a rate that is influenced by inductive and steric effects from adjacent side chains (28). For unstructured peptides, HX is a slow process simply because the hydroxide concentration is low. For example, at 25 °C and pH 4, HX occurs on a time scale of minutes. Under similar conditions, amides buried in globular proteins exchange on a wide range of time scales, extending up to centuries. HX can only occur if the amide is exposed to solvent, so conformational fluctuations must be an integral part of the HX mechanism (18).

Under sufficiently destabilizing conditions HX occurs from the denatured-state ensemble, but under native conditions few amides exchange by such global unfolding (9, 29–31). For example, in bovine pancreatic trypsin inhibitor (BPTI), 8 amides in the core β-sheet exchange by global unfolding under native conditions (7, 32), whereas the remaining 45 amides require less extensive conformational fluctuations. Much of the debate in the protein HX field over the past half-century has centered the nature of these subglobal fluctuations and their frequency, duration, amplitude, and cooperativity (18–25).

According to the standard HX model (18), each amide can exist in a closed (C) state, where exchange cannot occur, or in an open (O) state, where exchange proceeds at a rate $k_{\text{int}}$. The kinetic scheme for H exchange into $D_2O$ then reads as

$$\begin{align*}
N & \overset{k_{\text{on}}}{\underset{k_{\text{off}}}{\rightleftharpoons}} X, \\
N & \overset{k_{\text{on}}}{\underset{k_{\text{off}}}{\rightleftharpoons}} D, \\
O & \overset{k_{\text{on}}}{\underset{k_{\text{off}}}{\rightleftharpoons}} D + H, \\
O & \overset{k_{\text{on}}}{\underset{k_{\text{off}}}{\rightleftharpoons}} D + H,
\end{align*}$$

and the measured steady-state HX rate is $k_{\text{HX}} = k_{\text{on}}k_{\text{off}}/(k_{\text{on}} + k_{\text{off}} + k_{\text{int}})$. To make this phenomenological model practically useful, two auxiliary assumptions are needed to disentangle the conformational and intrinsic parts of the process: (i) The conformational fluctuations ($k_{\text{on}}$ and $k_{\text{off}}$) are independent of pH, and (ii) HX from the O state proceeds at the same rate as in model peptides with the same neighboring side chains, so that $k_{\text{int}} = k_{\text{HX}}$.

Two HX regimes are distinguished with reference to the pH dependence of $k_{\text{HX}}$ (18). If $k_{\text{HX}}$ is constant in some pH range, it follows that $k_{\text{int}} \gg k_{\text{off}} + k_{\text{on}}$ so that $k_{\text{HX}} = k_{\text{on}}$. In this so-called EX1 limit, the HX experiment measures the opening rate, or the mean residence time (MRT), of the C state, $\tau_C = 1/k_{\text{on}}$. For BPTI, such pH invariance has only been observed for the eight core amides, and then only in a narrow pH interval (32).

More commonly, HX experiments are performed in the EX2 limit, where $k_{\text{int}} \ll k_{\text{off}} + k_{\text{on}}$. Then $k_{\text{HX}} = k_{\text{on}}/(k_{\text{off}} + k_{\text{on}})$, where $\tau_C = k_{\text{off}}/k_{\text{on}}$. At equilibrium, the fractional populations, $f_C$ and $f_O$, and the rates are linked by detailed balance, $k_{\text{off}} f_C = k_{\text{on}} f_O$, so the PMF may also be expressed as $\kappa = f_C/f_O$. Clearly, $1/(k_{\text{off}} + k_{\text{on}})$ is the probability of finding the amide in the O state. $1/k_{\text{on}}$ is the C = $\text{O}$ equilibrium constant, and $\beta$ is the free energy difference between the O and C states in units of $k_B$ $T \equiv 1/k_{\text{on}}$. The C state can be thus deduced from the HX rates measured (under EX2 conditions) for the amide in the protein and in a model peptide as $\kappa = k_{\text{HX}}(\text{model})/k_{\text{HX}}(\text{protein})$.

The vast majority of the available protein HX data pertains to the EX2 regime and thus provides no information about the time scales, $\tau_C$ and $\tau_O$, of the conformational fluctuations, except for the EX2 bound: $1/(k_{\text{off}} + k_{\text{on}}) \gg \kappa \approx k_{\text{HX}}(\text{model})$. In the typical case where $k_{\text{HX}}(\text{model}) < k_{\text{HX}}(\text{protein})$, so that $\tau_O \gg \tau_C$, we therefore only know that $\tau_O \approx 1/k_{\text{HX}}(\text{model})$, which is in the millisecond range at pH 9 (EX2 HX data are usually measured at lower pH, where $1/k_{\text{HX}}(\text{protein})$ is even longer). Our analysis indicates that $\tau_O$ is seven orders of magnitude shorter than this upper bound estimate.
The HX experiment is unique in probing sparsely populated conformational states with single-residue resolution. However, the physical significance of the PF is obscured by our ignorance about the structure and dynamics of the O state. Several attempts have been made to correlate experimental PFs with physical attributes of the amides, such as solvent contact (33–37), burial depth (38), intramolecular H-bonds (35, 38–40), packing density (38, 41), or electric field (42). Where significant correlations have been found, they suggest that the chosen attribute can serve as a proxy for the propensity for C→O fluctuations. However, whether based on crystal structures or molecular dynamics (MD) trajectories, these studies examined the time-averaged protein structure, which is dominated by the C state and therefore provides little or no information about the nature of the C→O fluctuations.

In principle, the O state can be identified from molecular simulations, but this requires extensive conformational sampling because most C→O transitions are exceedingly rare. To date, this approach has been tried only with coarse-grained and/or empirical protein models without explicit solvent (43–45), or for HX from the denatured-state ensemble (46). The recent availability of ultralong MD simulations with realistic force fields opens up new opportunities in the search for the elusive O state. BPTI is also among the proteins that have been most thoroughly studied by HX experiments.

**Results**

**Structure of the O State.** A classic MD simulation cannot reveal the O state directly because it does not describe the intrinsic HX step. Our strategy is therefore to postulate a generic (same for all amides) structural criterion that must be satisfied for proton transfer to take place (in the real protein). This criterion is then justified by mechanistic considerations and by its ability to reproduce experimental PFs. The O-state criterion adopted here is that the N–H hydrogen has at least two water oxygens within \( R_{\text{HO}} = 2.6 \) Å. No angular constraint is imposed. This \( R_{\text{HO}} \) cutoff distance closely matches the first minimum in the \( H - O_W \) pair correlation function computed from the MD trajectory (Fig. S1). The precise \( R_{\text{HO}} \) value is not critical because the third water molecule, when present, is significantly more remote (Fig. S2).

With this criterion, 41 out of 53 amides sample the O state in the trajectory, with a mean water coordination within 2.6 Å of \( N_W = 2.001 \) (Fig. 1A). For the 53 amides in the C state, \( N_W = 0.364 \pm 0.393 \), ranging from \( \leq 0.001 \) (for the eight core amides for amides 29, 35, 51, and 52) to \( \geq 0.8 \) (for the exposed surface amides 3, 11, 15, 19, 30, 32, 34, 39, 48, 57, and 58, and for amides 10, 38, and 41 that donate H-bonds to one of the four internal waters). For fully solvent-exposed model amides, such as N-methylacetamide, \( N_W \approx 1 \) (48). In the O state, the N–H group is thus overcoordinated.

For nearly all of the examined amides, the requirement that \( N_W \geq 2 \) automatically guarantees that the intramolecular H-bond of the N–H group, which is present in the C state for more than 70% of the amides, is disrupted in the O state (Fig. 1B). (As discussed below, the converse is not true.) To eliminate the few remaining intramolecular H-bonds, we used a stronger O-state criterion requiring at least two water oxygens within 2.6 Å from the amide H as well as no protein N, O, or S atom within 2.6 Å of the amide H (except for backbone O or N atoms in the same or adjacent residues). The effect of this modification is significant only for Phe33 and Thr32 (Fig. 1B). The amide of Phe33 accesses...
the weak O state in only one frame, where the intramolecular H-bond partner (Arg20.O) and two water molecules are all just within the 2.6-Å cutoff. With the strong O-state criterion, this single frame is assigned to the C state, thereby excluding Phe33 from analysis. The remaining 12 amides, all located within the 2.6-Å cutoff at Thr32, where the side-chain hydroxyl oxygen is just inside the 2.6-Å cutoff in half of the weak O-state frames.

To assess structural differences between the O and C states, we compute, for each amide that samples the O state, the local rmsd σloc = √[(σi/O)2 − (σi/C)2]/2, where (σi/O) is the position of a particular atom averaged over all frames where the considered amide is in the O state, and the outer angular brackets denote an average over the 59 ± 14 nonhydrogen atoms located within 7 Å of the amide nitrogen (in the first frame of the trajectory). Fig. 1C compares σloc with the root-mean-square fluctuation σB = [3 (B2)/(8 π2)]1/2, where B is the crystallographic B factor from the room-temperature BPTI structure set (49), averaged over the same set of atoms as for σloc. The systematic O/C structural difference measured by σloc is not significantly correlated with the local flexibility σB (r2 = 0.06); the former shows much more variation along the backbone, from 0.05 to 3.1 Å. Moreover, σloc does not correlate with secondary structure (or lack thereof) in the C state.

One expects that weakly protected amides require smaller structural adjustments to become exchange-competent. This is indeed the case; σloc shows a significant correlation (r2 = 0.44) with βΔGexp = ln[FO/T] (Fig. S3). This correlation is virtually independent of the cutoff radius in the examined 5- to 8-Å range. In contrast to a previous suggestion (41), we do not find a significant correlation (r2 = 0.03) between local rigidity, as measured by σN2, and ΔGexp (Fig. S3).

A more detailed view of the O state for two amides is provided by the snapshots in Fig. 1 D and E, to be discussed in the following. Additional (interactive) O-state structures can be found in Fig. S4.

Simulated Versus Experimental Protection Factors. As a test of our O-state definition, we use it to compute PFs that can be compared with experiment. For each frame in the trajectory, we apply the structural criterion to assign each of the 53 backbone amides in BPTI to the O or C state. The PF is then computed as κsim = NFO/NFO where NFO and NFC are the number of frames where the amide is in state O or C.

The FX rate, kFX, for H exchange into D2O has been measured by NMR for each of the 53 amides in BPTI (7, 50–53). For our analysis, we use experimental PFs (at 300 K) for a subset of 41 amides in BPTI (Fig. S3). This correlation is virtually independent of the cutoff radius in the examined 5- to 8-Å range (28). The 41 experimental PFs are listed in Table S1 and a detailed description of how they were deduced can be found in SI Materials and Methods.

In the simulation, 41 amides access the O state, whereas 12 amides remain in the C state throughout the trajectory. The former set of 41 amides includes 11 of the 12 surface amides for which we lack reliable experimental PFs. Consequently, 30 amides are available for a quantitative comparison between simulation and experiment. For all but three of these amides, the simulation-based prediction of the O/C free energy difference ΔG agrees to better than 2.5 kBT with the corresponding experimental result (Fig. 2 A and B and Table S1). In terms of the deviation parameter ΔΔG = ΔGsim − ΔGexp, the overall agreement for these 30 amides can be expressed as the average absolute deviation, β′(ΔΔG) = 1.57, or the average signed deviation, β(ΔΔG) = 0.44, showing that, on average, the simulation slightly overestimates the protection. The error bars in Fig. 2 A and B represent the statistical uncertainty in κsim due to the finite length of the trajectory (SI Materials and Methods).

Our O-state definition is further supported by the complete absence in the trajectory of C → O transitions for the eight core amides (residues 20–24, 31, 33, and 45, with κsim ≥ 106). Because these amides exchange by global unfolding (7, 32), they should not access the O state in the analyzed native-state trajectory. Assuming Poisson statistics, the probability of observing at least one C → O transition in a trajectory of length T for an amide with PF κ and MRT τO in the O state is P(T) = 1−exp(−T/κτO). Even if τO were as short as 1 μs for the globally unfolded protein, P would be merely ~10−4 for T = 0.26 ns and κ = 106. By the same token, the trajectory length required to observe at least one opening event with probability P* = 1−1/ε ~ 0.63 is T = τO = κτO. For amides exchanging by subglobal fluctuations with τO ~ 100 ps (see below) and κ ~ 105, 106, and 107, the required trajectory length is 10 ns, 1 μs, and 100 μs, respectively.

In summary, among the 41 available experimental PFs, 35 are fully consistent with the simulation, either quantitatively (27 amides) or qualitatively (8 amides). Two of the remaining six amides, in residues Cys14 and Cys38, are sensitive to the conformation of the 14–38 disulfide bond. If these two amides are allowed to access also the experimentally unresolved (54) minor disulfide conformations M2 and M3 in the O state, we obtain good agreement with experiment for both Cys14 (βΔDG = −0.78) and Cys38 (βΔDG = 1.43). Apart from the eight amides in the slow-exchange core and the amide of Cys14, three more amides do not undergo any C → O transitions: Ala27 (κsim = 102–104) located in a turn, Asp50 (not included in the experimental data set) in the C-terminal α helix, and Cys51 (κsim = 102–103) involved in a disulfide bond. The discrepancies for these residues remain unexplained.

The PF comparison in Fig. 2 A and B is reassuring, considering the known sources of systematic error. Foremost among these is the assumption, used to extract the experimental PFs, that the intrinsic HX rate in the O state is the same as in model peptides (42). The PF comparison may also have been affected by differences in solvent conditions between simulation and experiment (stronger H-bonds and hydrophobicity in D2O than in H2O; about half of the five carbonyl groups protonated at pH 3.5 but none at pH ~7). As regards the simulation, the main concern is the quality of the empirical force field. However, the force field used in this simulation (47) has features (55) well beyond the reach of NMR data related to conformational flexibility of native proteins (55–57). Moreover, the MD trajectory used here yields excellent agreement with the NMR-determined exchange times of the four internal water molecules in BPTI (58).

Cooperativity and Kinetics. An important characteristic of subglobal C → O fluctuations is their degree of cooperativity. Are they truly local or do several nearby amides access the O state simultaneously? To address this question, we compute the O-state correlation matrix C(n, n′) where access to O of amides n and n′ are uncorrelated and 1 if they are perfectly correlated (SI Materials and Methods). We find that the vast majority of the 41 amides that access the O state during the trajectory are uncorrelated. Only five amide pairs have C(n, n′) > 0.03 (Fig. 2C). The largest correlation, C(37, 38) = 0.17, involves amides near the 14–38 disulfide bond. Other significantly correlated amide pairs are C(4, 5) = 0.07 in the cooperatively unfolded N-terminal 310 helix and C(27, 28) = 0.06 in the turn between the two β-strands. The only significant correlation between residues that are far apart in the sequence is C(5, 52) = 0.04 in polypeptide segments linked by the 5–55 disulfide bond.

To characterize the kinetics of C→O fluctuations, we compute the MRTs τO = 1/kO and τC = 1/kC as the average number of consecutive frames in each state multiplied by the sampling resolution Δt = 0.25 ns. Most O-state visits are only a single frame and none lasts more than six frames (Table S2). Many visits must therefore be shorter than Δt and only a fraction of these will be recorded at the given sampling resolution. By modeling the C→O fluctuations as an alternating Poisson process, we can correct these amides remain uncorrelated for. As shown in SI Materials and Methods, the corrected MRTs are given by τO = −Δτ/ln[1/NFO/NFC] and τC = NFO/FO ln[1−NFO/NFC], where NFO and NFC were defined in connection with the PF and N = NFO = NC = 1 is the number of visits to the O state during the trajectory.
three orders of magnitude (from 1 ns to 2 μs). The large variation in PF among the amides is therefore due almost entirely to variation in τC (or the opening rate, kO). The mean and SD of τO for the 34 amides is 81 ± 18 ps. For these amides, the O state is thus highly unstable, that is, the free energy barrier for the O → C transition is small.

The remarkably short value of τO led us to examine the possibility that the amide water coordination fluctuates rapidly between 2 and 1 while the protein configuration remains open. O-state visits would then appear in clusters, rather than being randomly distributed along the trajectory. If this were the case, we would also observe a large number of short (one or a few frames) visits to the C state. However, this is not the case; the C-state residence time distributions for most of the 34 amides are close to exponential (Fig. S5), as expected for an alternating (C/O) Poisson process. We therefore conclude that the short τO is a robust property of the O-state definition introduced here, which also yields PFs in good agreement with experiment.

Discussion
Relation to Previous Work. The simulation-based HX analysis presented here differs in two essential ways from previous computational HX studies. First, the MD trajectory used here is based on a realistic physical model with explicit water and state-of-the-art force field (47). An earlier study (45) with similar objectives used a coarse-grained implicit-solvent model and defined the O state as having no Cβ atom within 6.5 Å of the reference Cγ atom. This O-state definition is much more disruptive than ours (Fig. 1C); for example, the O state of Arg53 (Fig. 1E) has five other Cβ atoms within 6.5 Å.

The second key feature of our analysis is the length of the MD trajectory, 262 μs. The longest atomistic MD trajectory previously used for HX analysis was three orders of magnitude shorter (90 ns) (36). Rather than attempting to identify the O state, that study correlated the trajectory-averaged (essentially C-state) amide solvent access with the number of exchanged hydrogens (detected by mass spectrometry) for a set of peptide fragments (36). However, the strong linear correlation thus obtained seems to be spurious, largely resulting from the trivial increase of both variables with peptide size.

Nature of the O State. It is generally agreed that the two necessary conditions for exchange competence are direct access to external solvent and disruption of any intramolecular H-bond with the N-H group (18–25, 40). Our O-state definition incorporates these conditions by stipulating that the amide hydrogen has at least two water oxygens within 2.6 Å and that the amide hydrogen has no other polar protein atom (except in neighboring residues) within 2.6 Å. For all but a few amides (Fig. 1B), the second condition is automatically satisfied if the first one is obeyed.

We have explored several other O-state criteria, such as solvent-accessible surface area and intramolecular H-bonding, but the agreement with the experimental PFs is invariably worse than in Fig. 2A and B. For example, if we retain the condition of disrupted intramolecular H-bond but require at least one (rather than two) waters within 2.6 Å of the amide hydrogen, then the O state is somewhat more long-lived (τO = 0.05–5 ns) but the PF is severely underestimated (β ∆ΔG = −4.85). This more permissive O-state criterion may capture more of the water-penetrated conformations that are implicated in internal-water exchange (58).

According to our results, the O state is not only improbable (fO < 1), it is also highly labile: τO = 41–113 ps for the examined amides (Fig. 2D). In the few cases where τO has been experimentally inferred for amides that exchange by subglobal fluctuations, values in the microsecond range have been reported (59). However, these inferences depend critically on the two auxiliary assumptions in the standard HX model, namely, that the relevant conformational fluctuations are pH-independent and that the intrinsic HX rate in the O state is the same as for model peptides.
These assumptions are not likely to be quantitatively accurate for amides that exchange by subglobal fluctuations.

The τO values of order 100 ps inferred here are seven orders of magnitude shorter than the MRT of the globally unfolded protein, which is the O state for the eight core amides (32). At least for BPTI, few, if any, amides seem to have a τO value in the wide interval between these extremes. For an alternating Poisson process, there is a 90% probability that an amide with τC = κτO = T/ln10 undergoes at least one opening event in a trajectory of length T (discussed above). Therefore, even if τO were as long as 1 μs, we would have observed the O state in the 0.26-ms trajectory for at least 10 of the 30 amides in Fig. 2A (i.e., those with κ < 100).

The striking disparity in τO values of amides that exchange by subglobal versus global fluctuations is the result of a highly cooperative global unfolding. At least for a small single-domain protein such as BPTI (with three disulfide bonds), we do not observe a continuous spectrum of conformational fluctuations on all length and time scales. Instead, it seems that only highly localized and short-lived fluctuations can occur as long as the β-sheet core is intact.

Solvent Penetration Versus Local Unfolding. Much of the debate about the HX mechanism in proteins has been framed as a dichotomy between “solvent penetration” and “local fluctuation” scenarios (18–25). These imprecisely defined scenarios are not necessarily mutually exclusive; to some extent they may be different sides of the same coin. Using the same MD trajectory, we have previously shown that the internal water molecules in BPTI exchange by way of H-bonded water chains that penetrate the protein through transient tunnels or pores (58). Such chains are not a general feature of the amide O state that we identify here, even though chains formed by some of the “native” internal water molecules are seen for a few amides, for example Gly36 (Fig. 1D) and Lys41, but not in all O-state configurations (Fig. S4). The open states for internal-water-exchange are also short-lived (a few nanoseconds) (58), but the opening frequency is one to two orders of magnitude lower (60) than for the HX O states of the amides that H-bond to internal waters.

For most of the examined amides, the conformation of the O state is probably best described as locally distorted or “unfolded.” The almost complete lack of temporal correlation of C → O transitions in different amides (Fig. 2C) does not exclude local unfolding cooperativity. Especially in the terminal helices, we do see that several adjacent residues are unfolded even if only one amide at a time is in the O state (Fig. 1E). This observation is consistent with the finding (Fig. 1B) that overcoordinated (Nw ≥ 2) amides rarely have intramolecular H-bonds, whereas a disrupted intramolecular H-bond is not a sufficient O-state criterion (discussed above). In larger proteins, the O state may well be reached by cooperative unfolding of larger structural units (9). In any case, the intermittent nature of the C → O fluctuations suggests that “coughing” may be a more accurate metaphor than “breathing.”

The modest extent of structural distortion in the O states identified here (Fig. 1C) is consistent with the small (less than an order of magnitude) effect of 8 M urea on kHX for the BPTI amides that exchange by subglobal fluctuations (29). Because urea has a similarly small effect on κk (61, 62), this observation indicates that the C≡O equilibrium is much less sensitive to urea than the global N≡U equilibrium, consistent with a modest increase of solvent exposure in the O state.

Among experimental probes, amide HX and internal-water exchange (60) are unique in their ability to monitor rare and transient conformational fluctuations. The C≡O transitions discussed here are not observable by conventional NMR relaxation methods, whether relaxation is induced by anisotropic nuclear couplings (JCH is too small) or by chemical shifts (the effective correlation time, ~ τO, is too short).

Proton Transfer Mechanism. Without “smoking gun” type of evidence, the exchange-competent state wherein the proton is transferred cannot be unambiguously established. Nevertheless, we believe that the O state identified here, with its overcoordinated amide hydrogen, is a plausible candidate also from a mechanistic standpoint.

The classic kinetic scheme for intermolecular proton transfer (PT) posits a diffusion-controlled formation of a H-bonded encounter complex, followed by a fast PT equilibrium within the complex (27). However, this phenomenological framework does not address the crucial participation of water molecules in the PT mechanism. In bulk aqueous solution, the hydroxide ion interacts strongly with three water molecules (63) during its pico-second lifetime (64). The excess negative charge migrates by structural diffusion, involving more or less concerted proton jumps through the H-bond network (65, 66). In keeping with these notions, we envision the following PT scenario in the overcoordinated O state (Fig. S6).

Upon approach of the hydroxide ion, the “first” water molecule, which accepts a H-bond from the N → O group, extracts the amide hydrogen, thereby converting the hydroxide ion into a water molecule, possibly via one or more intermediate H-bonded water molecules. Concomitantly, the “second” water molecule reorients to interact strongly (as H-bond donor) with the transient imidate ion, which now accepts H-bonds from both water molecules. Without the second water molecule, the incompletely solvated imidate ion would quickly revert to its original N → O form without having exchanged its proton. When two water molecules are present, there is at least a 50% chance (more if the process is concerted) for the amide to acquire a new proton (or deuteron) by extracting it from the second water molecule, which thereby regenerates the catalytic hydroxide ion. If the second water molecule is linked via H-bonds to additional water molecules, the regenerated hydroxide ion may appear at some distance from the N → O group. Indeed, either the entry or the exit of the excess charge might proceed along a water wire in a pore connecting the amide with bulk solvent. Although involvement of such water wires may be rare, Gly36 is the only case observed here (Fig. 1D), theoretical studies support this possibility (67).

If an overcoordinated N → O group is required for amide PT in the protein, then the PT mechanism must differ somewhat in model amides, where Nw ≈ 1 (48). This would violate the assumption in the standard two-state model that the intrinsic HX rate, kinc, is the same in these two situations. At first sight, this assumption seems unlikely to be valid because the microsolvation of the N → O group must differ. However, the proton transfer process is diffusion-controlled (27) and is therefore not affected by the rate of equilibration between the encounter complex and the N → O state. kinc depends only linearly on the target size, which, in any case, is ill-defined owing to the “delocalization” of the hydroxide charge (65, 66). The imidate ion is likely to be coordinated by two water molecules also in the encounter complex of model peptides, but in that case a second water molecule is never far away and may move into position as the hydroxide ion approaches. In contrast, in the more confined O state of the protein, the N → O group must be “presolvated” with two water molecules. In conclusion, we believe that the evidence favors the overcoordinated N → O group as the salient feature of the O state for amides that exchange by subglobal fluctuations.

Materials and Methods

Our computational analysis is based on a previously reported all-atom MD simulation at 300 K of the protein BPTI, solvated by 4,215 water molecules (47). The protonation state of ionizable groups corresponds to neutral pH, with the net protein charge neutralized by six chloride ions. A subset of 1,048,349 frames with 0.25-ns spacing, corresponding to a 0.262-ms-long trajectory, was extracted by requiring that the 14–38 disulfide bond is in the experimentally dominant M1 conformation (54). Further details can be found in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank D. E. Shaw Research for sharing the BPTI trajectory, Pär Söderhjelm for useful comments, and the Swedish Research Council for financial support.


**Supporting Information**

**Persson and Halle 10.1073/pnas.1506079112**

**SI Materials and Methods**

**Simulation Data.** The analysis was based on a previously reported 1.031-ns MD simulation at 300 K of BPTI solvated by 4,215 water molecules (1). The trajectory was sampled with a resolution of Δt = 0.25 ns, yielding 4,125,000 frames. Using VMD (2), we aligned all protein atoms in each frame with the first frame in the trajectory. For the Amber f99SB-ILDN force field used in the simulation, the rotational isomer populations for the C14–C38 disulfide bond are known (3–6) to differ substantially from the experimentally deduced ones (7, 8). The analysis was therefore restricted to the subset of Np = 1,048,349 frames (corresponding to a 262.1-μs-long trajectory) where the 14–38 disulfide bond is in the experimentally dominant (95%) M1 conformation. We refer to this subset as “the trajectory.” If this constraint is applied only to the amide C state, we find that the disulfide remains in the M1 conformation also in the amide O state, except for two amides (Gly37 and Cys38) near the disulfide bond (Fig. S7). For these two amides, 60–70% of the O-state frames coincide with the disulfide M2 conformation, which also dominates for the transient water tunnel that constitutes the open state for exchange of the internal water molecule (W122) buried under the 14–38 disulfide bond (6).

**Experimental Data.** To minimize errors introduced by temperature corrections, we gave preference to k_{H2O} data measured as close as possible to the simulation temperature, 300 K. First we included data for 19 amides measured at 30°C (9) and for the amide of Gly37 reported at 10°C but back-correlated to the measurement temperature 30°C using an Arrhenius activation energy E_A = 30 kcal-mol^{-1} (10). Then we added data for nine amides that were measured at 36°C and extrapolated to 30°C with E_A = 35 kcal-mol^{-1} (9) and for the three most slowly exchanging amides (in residues 21–23) measured at high temperature and extrapolated to 30°C with E_A = 78 kcal-mol^{-1} (9). (The latter three amides were not used for the quantitative comparison because they do not access the O state in the simulation.) For these 32 amides were then corrected from 30°C to 300 K using the activation energies quoted above and, for the first set of 19 amides, E_A(kcal-mol^{-1}) = -1.17 -9.306 log(k_{H2O}(30°C, pH=3.5)), a relationship obtained from a linear fit for 11 amides that were measured at both 10 and 30°C (9, 10). Finally, we included data for nine amides that were only measured at 10°C (10), extrapolated to 300 K with E_A(kcal-mol^{-1}) = -0.73 - 8.072 log(k_{H2O}(10°C, pH=4.6)). The HX rate constants for these 41 amides, corrected to 300 K, are given in Table S1.

The PF was obtained from the experimental HX rate constant, k_{H2O}, as k_{exp} = k_{cat}/k_{H2O}, where k_{cat} = k_0 + k_1 [D3O] + k_2 [OD] = k_0 + k_1 10^{-30} + k_2 10^{-30} or k_{cat}, where k_0, k_1, and k_2 are, respectively, the rate constants for uncatalyzed, acid-catalyzed, and base-catalyzed HX. Furthermore, pD = pH - 0.4 (11) and pK_{D3O} is the negative logarithm of the ionization constant for D3O (12). Reference rate constants for poly-DL-alanine in D2O were used with values at 20°C of k_02 = 10^{-15} min^{-1}, k_12 = 10^{62} min^{-1}, and k_22 = 10^{10.05} min^{-1} (13), and with activation energies E_A,D2O = 13 kcal-mol^{-1}, E_A,L = 15 kcal-mol^{-1}, and E_A,L = 2.6 kcal-mol^{-1} taken from Roder’s Sphere program (www.fccc.edu/research/labs/roder/sphere/sphere.html). Nearest-neighbor effects were corrected for according to k_0 = B_1 B_2 k_00, k_1 = A_1 L k_10, and k_2 = B_1 B_2 k_20, where, for example, B_1 refers to the side chain that has the considered amide group to its left. The correction factors were taken from ref. 13, except for the basic forms of Asp and Glu, where the factors were taken from ref. 14. The C-terminal carboxyl group (which is involved in a salt bridge) was taken to be fully deprotonated at the experimental pH* values, whereas for the other carboxyl groups we used experimentally determined pK_{D3O} values: 3.9 (Asp3), 3.5 (Asp50), 4.2 (Glu7), and 4.3 (Glu49) (15). The temperature dependence of these pK_{D3O} values is negligibly small.

In this way, we obtained k_{cat} at 300 K and at the experimental pH* values of 3.5 or 4.6. Even at the lower pH* value, the basic mechanism dominates strongly over the acid mechanism. The uncatalyzed mechanism makes a minor but significant contribution. The 41 PFs obtained by combining these k_{cat} values with the experimental k_{H2O} values are listed in Table S1. The experimental uncertainties in k_{cat} and k_{H2O} are not known, but we note that ±10% error propagates to ±0.14 in β ΔG_{exp} shown in Fig. 24.

**O-State Correlation Matrix.** To investigate the temporal correlation of the O state of different amides, we consider the joint probability, P(n, n′), that amides n and n′ are in the O state at the same time. Defining an indicator function, h(n,k), such that h = 1 if amide n is in the O state in frame k and h = 0 otherwise, we compute P(n, n′) = (h(n,k) h(n′,k)) by averaging over all frames in the trajectory. We then compute the O-state correlation matrix C(n, n′) = |P(n, n′) – P(n)P(n′)|/[P(n)[1 – P(n)]|1 – P(n′)]^{1/2}, where P(n) = (h(n,k)) is the a priori probability that amide n is in state O. For uncorrelated amides, P(n, n′) = P(n)P(n′) so that C(n, n′) = 0; for perfectly correlated amides, P(n, n′) = P(n)P(n′) so that C(n, n′) = 1; and for perfectly anticorrelated amides C(n, n′) = −1.

**Binning Error.** Fig. S8 shows a small part of a trajectory sampled in two different ways. On the upper continuous time line, O → C and C → O transitions are indicated by tick marks. On the lower discretely sampled time line, the observation time points, that is, the frames saved for analysis, are indicated by dots separated by the time interval Δt, which is the sampling resolution. For each frame, the observed state is indicated.

The distribution of state residence times (RTs) in the continuous-time trajectory, that is, the separation of adjacent tick marks on the upper time line in Fig. S8, is described by the continuous probability densities ψ_00(τ) and ψ_01(τ), such that ψ_00(τ) dτ is the fraction of all O-state RTs that are within dτ in length. Here, and in the following, quantities pertaining to the continuous-time trajectory are indicated by a zero superscript.

The total length T of the trajectory is the sum of the total time T_{O} spent in state O and the total time T_{C} spent in state C,

\[ T = T_{O} + T_{C}. \]  

These parts may be written as

\[ T_{O} = N_{O}^{0} \rho_{O}, \]  

\[ T_{C} = N_{C}^{0} \rho_{C}, \]  

where \( N_{O}^{0} \) is the total number of O-state residences in the continuous-time trajectory and \( \rho_{O} \) is the mean RT in the O state, that is,

\[ \rho_{O} = \int_{0}^{\infty} \tau \psi_{O}(\tau) d\tau. \]  

and similarly for state C.
Analogous relations hold for the discretely sampled trajectory, which is of the same total length $T$ as the continuous trajectory, namely,

$$T = T_O + T_C,$$  \hspace{1cm} [S4]

with

$$T_O = N_O \tau_O = N_{FO} \Delta \tau,$$  \hspace{1cm} [S5a]

$$T_C = N_C \tau_C = N_{FC} \Delta \tau,$$  \hspace{1cm} [S5b]

where $N_{FO}$ and $N_{FC}$ are the total number of frames in states O and C, respectively. The mean RT in state O, as judged from the discrete trajectory, is

$$\tau_O = \Delta \tau(n_O) = \Delta \tau \frac{1}{N_O} \sum_{n=1}^{N_O} n_{O,n} = \Delta \tau \frac{N_{FO}}{N_O}, \hspace{1cm} [S6]$$

and similarly for $\tau_C$. Here, $n_{O,n}$ is the number of frames in the $n$th visit to state O.

The PF is estimated from the discrete trajectory as $\kappa = N_{FC}/N_{FO}$, which, in view of Eq. S5, can also be expressed as

$$\kappa = \frac{T_C}{T_O} = \frac{N_C \tau_C}{N_O \tau_O} \hspace{1cm} [S7]$$

Because of the finite sampling resolution $\Delta \tau$, this PF estimate is subject to a systematic binning error. Ideally, we would like to compare the experimental PF with a theoretical PF based on the continuous trajectory (in the limit $\Delta \tau \to 0$), that is,

$$\kappa^0 = \frac{T_O^0}{T_C^0} = \frac{N^0_{FO}}{N^0_{FC}} \hspace{1cm} [S8]$$

We would also like to know the true mean RTs $r_O^0$ and $r_C^0$ rather than the estimates $\tau_O$ and $\tau_C$ based on the discrete trajectory.

The origin of the binning error is evident from Fig. S8, which shows two short O-state residences, with $\tau_O < \Delta \tau$. Some of these short residences will be assigned a too-short RT, namely $\tau_{O,ass} = \Delta \tau$ (this applies to the RT detected in frame 12), whereas others will not be detected at all (this applies to the RT between frames 14 and 15). However, the binning error is also due to the random lengthening and shortening of RTs cancels out to first order, but the omission of some RTs gives rise to a systematic error of order $\Delta \tau/r_O^0$. The net effect of these binning errors is therefore to make $\tau_O$ longer than $\tau_O^0$. However, the omission of some RTs makes $N_O$ smaller than $N_O^0$. As we shall see, for a Poisson process, these effects precisely cancel out, so that $T_O = N_O \tau_O$ is equal to $T_O^0 = N_O^0 \tau_O^0$, unaffected by the binning error. Similar considerations apply to the C state. Omissions of short O-state RTs makes $N_C$ smaller than $N_C^0$, but the merging of the flanking C-state RTs as well as the “conversion” of the omitted O-state RT to the C state makes $\tau_C$ longer than $\tau_C^0$. Again, for a Poisson process, these effects precisely cancel out, making $T_C = T_C^0$. Therefore, for a Poisson process, the PF is unaffected by the binning error, that is, $\kappa = \kappa^0$. We shall now demonstrate explicitly and show how the mean RTs are affected by the binning error.

To correct quantitatively for the binning error, we assume that the conformational fluctuations that interconvert a given amide between the C and O states can be modeled as an alternating Poisson process (sometimes called a dual Poisson process). In other words, the RTs are independently and exponentially distributed with normalized probability densities

$$\psi^0_O(\tau) = \frac{1}{\tau_O^0} \exp\left(-\frac{\tau}{\tau_O^0}\right), \hspace{1cm} [S9a]$$

$$\psi^0_C(\tau) = \frac{1}{\tau_C^0} \exp\left(-\frac{\tau}{\tau_C^0}\right). \hspace{1cm} [S9b]$$

If all frames in the trajectory are used, the first and last RTs will in general be truncated. It would thus seem that the terminal RTs differ from the internal RTs, being shorter on average. However, for a Poisson process, the sampling bias (that a randomly selected time point is likely to fall in a long RT) cancels the truncation effect. Therefore, we can describe all RTs, both internal and terminal, with the same probability density $S9$. Because it does not matter whether or not we prune the trajectory by removing the terminal RTs, they should be included to improve the statistical accuracy.

The RT histogram $F_O(n)$ is obtained by substituting $\psi^0_O(\tau)$ from Eq. S9a into equation S5 of ref. 16 and performing the integrations. The result is, for $n \geq 1$,

$$F_O(n) = N^0_O \tau_O^0 (1 - e^{-\tau_O^0})^{n-1} (1 - e^{-\tau_O^0})^2, \hspace{1cm} [S10]$$

where we have defined

$$x = \frac{\Delta \tau}{\tau_O^0}. \hspace{1cm} [S11]$$

The number of (detected) O-state RTs can then be obtained as

$$N_O = \sum_{n=1}^{\infty} F_O(n) = N^0_O \frac{(1 - e^{-\tau_O^0})}{x} \hspace{1cm} [S12]$$

where, in the second step, we have summed the geometric series. Similarly, the number of (detected) O-state frames is obtained as

$$N_{FO} = \sum_{n=1}^{\infty} n F_O(n) = N^0_O \frac{1}{x} \hspace{1cm} [S13]$$

Combining Eqs. S2a, S11, and S13, we obtain

$$T_O = N_{FO} \Delta \tau = N^0_O \tau^0_O = T^0_O. \hspace{1cm} [S14]$$

According to Eqs. S1 and S4, $T_O + T_C = T^0_O + T^0_C$. It therefore follows from Eq. S14 that $T_C = T^0_C$, and, in view of Eqs. S7 and S8, that

$$\kappa = \kappa^0. \hspace{1cm} [S15]$$

The PF is thus immune to the binning error if the O/C transitions can be described as an alternating Poisson process.

We now consider the effect of the binning error on the mean RTs. Using Eqs. S11–S13, we can write the O-state mean RT in the following ways:

$$\tau_O = \Delta \tau \frac{N^0_O}{N_O} = \tau_O^0 \frac{N^0_O}{N_O} = \tau_O^0 \frac{x}{(1 - e^{-\tau_O^0})} \frac{1 - e^{-\tau_O^0}}{\tau_O^0} \frac{\Delta \tau}{(1 - e^{-\tau_O^0})} \hspace{1cm} [S16]$$

Combining the second and fifth members of Eq. S16 with Eq. S11, we can express $\tau_O^0$ in terms of simulation-derived quantities as

$$\tau_O^0 = -\frac{\Delta \tau}{\ln(1 - N_O/N_{FO})}. \hspace{1cm} [S17]$$

This result shows that for an amide where all O-state RTs are a single frame, so that $N_O = N_O^0$, we cannot obtain $\tau_O^0$. However, $\kappa^0$ (which is equal to $\kappa$) can still be obtained.

Persson and Hallé www.pnas.org/cgi/content/short/1506079112
A Taylor expansion of the fourth member of Eq. S16 yields

$$\frac{\tau_O}{\tau_C} = 1 + \frac{1}{2} x + \frac{1}{12} x^2 - \frac{1}{720} x^4 + \frac{1}{30240} x^6 \ldots,$$

[S18]

showing that the binning error on the mean RT is of first order in $x = \Delta \tau / \tau_C$, as we have shown more generally (without assuming a Poisson process) in another context (16).

We now turn to the mean RT $\tau_C$ of the C state. Because the problem is symmetric in the two states, it may be thought that we could proceed in complete analogy with the preceding treatment of the O state. However, this is not so, because equation S5 of Ref. 16 is not strictly valid for an alternating (Poisson) process. That equation assigns with a certain probability RTs in the interval $(n-1)\Delta \tau < \tau \leq (n+1)\Delta \tau$ to the $n^{th}$ bin of the RT histogram $F(n)$. This takes care of the discretization error, which can make the discrete RT shorter or longer than the continuous RT, and it also allows for the possibility that some RTs (with $\tau < \Delta \tau$) escape detection. However, equation S5 of ref. 16 does not describe the other consequence of an omitted RT, namely, that the two flanking RTs are merged. In the present HX context, the O-state RTs are always short whereas the C-state RTs are always long. Therefore, the merging effect is only relevant for the C state. We can therefore use equation S5 of ref. 16 for the O state, but not for the C state. Instead, we shall exploit the complementary nature of the two states in the alternating process.

We established below Eq. S14 that $T_C = T_C^O$. Combining this result with Eqs. S2b and S5b, we obtain

$$\tau_C^O = \tau_C N_C^O / N_C = \Delta \tau \tau_C N_F C / N_C,$$  

[S19]

where, in the second step, we have used the second equality in Eq. S5b. Next, note that for every O-state RT that is lost due to the finite sampling resolution, a C-state RT is also lost (when the two flanking C-state RTs are merged into one, together with the lost O-state RT). The number $\Delta N$ of undetected (or lost) RT pairs can therefore be written as

$$\Delta N = N_C^0 - N_C = N_O^0 - N_O.$$  

[S20]

From Eqs. S13 and S17, we find that

$$N_O^0 = -N_O \ln (1 - N_O/N_F O),$$  

[S21]

which is combined with Eq. S20 to yield

$$N_C^0 = N_C - N_O \left[ 1 + N_F O / N_O \ln (1 - N_O/N_F O) \right].$$  

[S22]

By inserting this result into Eq. S19 we obtain an expression for the true mean RT for state C in terms of quantities derived from the MD trajectory. It is seen that for an amide where all O-state RTs are a single frame, so that $N_C = N_O$, we cannot obtain $\tau_C^O$. As noted above, this is also true for $\tau_C^O$. In both cases, the reason is that, when all O-state RTs are a single frame, we have no quantitative information about the number $\Delta N$ of lost RT pairs (except that they are many). However, when $N_O << N_F O$, then Eq. S22 shows that $N_C^O = N_C$, as expected. If $N_F C >> N_O$ (meaning that $\kappa \approx 1$), as is the case for the amides examined here, then both of the terminal (truncated) RTs are with high probability associated with the C state. Then $N_C = N_O + 1$ so Eq. S22 reduces to

$$N_C^0 = 1 - N_F O \ln (1 - N_O/N_F O).$$  

[S23]

Of course, the distinction between $N_C$ and $N_O$ only matters if they are small. For most amides, $N_O \gg 1$ and then we can set $N_C = N_O$ irrespective of the state associated with the terminal RTs. If $N_C = N_O$ it follows from Eq. S20 that $N_C^O = N_C$. Comparing Eqs. S16 and S19, we then find that showing that the binning error lengths $\tau_C$ and $\tau_O$ by the same factor. The Taylor expansion in Eq. S18 is thus valid also for $\tau_C^O / \tau_C$.  

**Statistical Error.** We now consider the statistical uncertainties (or SEs) in the PF and mean RTs, resulting from the finite length of the MD trajectory. These errors will be significant if the number of detected RTs ($N_F O$) is small. As in our treatment of the binning error, we model the O/C fluctuations by an alternating Poisson process, with the RT probability densities given by Eq. S9. Because the individual RTs are then independent, we can treat the O-state RTs separately from the C-state RTs. For a Poisson process, the variance is the square of the mean so the SEM RT is

$$\sigma(\tau_O) = \frac{\tau_O}{N_F O^{1/2}}$$  

[S25]

with an analogous expression for $\sigma(\tau_C)$. The SE of the PF

$$\sigma(\kappa) = \left[ \left( \frac{\tau_C \sigma(\tau_O)}{\tau_O} \right)^2 + \left( \frac{\sigma(\tau_C)}{\tau_C} \right)^2 \right]^{1/2},$$  

[S26]

which is combined with Eq. S25 to give

$$\sigma(\kappa) = \kappa \left[ \frac{1}{N_O} + \frac{1}{N_C} \right]^{1/2}.$$  

[S27]


<table>
<thead>
<tr>
<th>Residue</th>
<th>No.</th>
<th>$^{10} \log(k_{HX}/s^{-1})$</th>
<th>$^{10} \log k_{exp}$</th>
<th>$^{10} \log k_{sim}$</th>
<th>$\beta \Delta \Delta G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>3</td>
<td>—</td>
<td>—</td>
<td>1.047</td>
<td>—</td>
</tr>
<tr>
<td>Phe</td>
<td>4</td>
<td>—</td>
<td>—</td>
<td>3.337</td>
<td>—</td>
</tr>
<tr>
<td>Cys</td>
<td>5</td>
<td>−5.556</td>
<td>3.704</td>
<td>3.844</td>
<td>0.324</td>
</tr>
<tr>
<td>Leu</td>
<td>6</td>
<td>−5.810</td>
<td>3.229</td>
<td>3.149</td>
<td>−0.184</td>
</tr>
<tr>
<td>Glu</td>
<td>7</td>
<td>−6.092</td>
<td>3.574</td>
<td>3.561</td>
<td>−0.030</td>
</tr>
<tr>
<td>Tyr</td>
<td>10</td>
<td>−5.457</td>
<td>2.500</td>
<td>2.437</td>
<td>−0.146</td>
</tr>
<tr>
<td>Thr</td>
<td>11</td>
<td>—</td>
<td>—</td>
<td>1.046</td>
<td>—</td>
</tr>
<tr>
<td>Gly</td>
<td>12</td>
<td>−3.659</td>
<td>1.668</td>
<td>2.168</td>
<td>1.152</td>
</tr>
<tr>
<td>Cys</td>
<td>14</td>
<td>−4.551</td>
<td>2.400</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lys</td>
<td>15</td>
<td>—</td>
<td>—</td>
<td>1.114</td>
<td>—</td>
</tr>
<tr>
<td>Ala</td>
<td>16</td>
<td>−5.359</td>
<td>3.024</td>
<td>2.623</td>
<td>−0.925</td>
</tr>
<tr>
<td>Arg</td>
<td>17</td>
<td>−3.139</td>
<td>0.761</td>
<td>1.670</td>
<td>2.094</td>
</tr>
<tr>
<td>Ile</td>
<td>18</td>
<td>−8.344</td>
<td>5.375</td>
<td>4.573</td>
<td>−1.847</td>
</tr>
<tr>
<td>Ile</td>
<td>19</td>
<td>−4.396</td>
<td>0.980</td>
<td>2.002</td>
<td>2.355</td>
</tr>
<tr>
<td>Arg</td>
<td>20</td>
<td>−9.089</td>
<td>6.478</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Tyr</td>
<td>21</td>
<td>−11.068</td>
<td>8.559</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phe</td>
<td>22</td>
<td>−11.068</td>
<td>8.419</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Tyr</td>
<td>23</td>
<td>−11.068</td>
<td>8.400</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Asn</td>
<td>24</td>
<td>−8.198</td>
<td>6.276</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ala</td>
<td>25</td>
<td>−3.458</td>
<td>1.323</td>
<td>0.925</td>
<td>−0.916</td>
</tr>
<tr>
<td>Lys</td>
<td>26</td>
<td>—</td>
<td>—</td>
<td>3.396</td>
<td>—</td>
</tr>
<tr>
<td>Ala</td>
<td>27</td>
<td>−5.021</td>
<td>2.686</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gly</td>
<td>28</td>
<td>−5.520</td>
<td>3.334</td>
<td>4.367</td>
<td>2.378</td>
</tr>
<tr>
<td>Leu</td>
<td>29</td>
<td>−7.566</td>
<td>4.716</td>
<td>5.322</td>
<td>1.395</td>
</tr>
<tr>
<td>Cys</td>
<td>30</td>
<td>−3.718</td>
<td>1.597</td>
<td>2.015</td>
<td>0.962</td>
</tr>
<tr>
<td>Gin</td>
<td>31</td>
<td>−8.344</td>
<td>6.402</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Thr</td>
<td>32</td>
<td>−4.808</td>
<td>2.476</td>
<td>1.763</td>
<td>−1.643</td>
</tr>
<tr>
<td>Phe</td>
<td>33</td>
<td>−8.265</td>
<td>5.765</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Val</td>
<td>34</td>
<td>−4.719</td>
<td>1.622</td>
<td>1.817</td>
<td>0.450</td>
</tr>
<tr>
<td>Tyr</td>
<td>35</td>
<td>−7.742</td>
<td>4.879</td>
<td>6.020</td>
<td>2.627</td>
</tr>
<tr>
<td>Gly</td>
<td>36</td>
<td>−6.298</td>
<td>4.158</td>
<td>4.029</td>
<td>−0.296</td>
</tr>
<tr>
<td>Gly</td>
<td>37</td>
<td>−6.984</td>
<td>4.968</td>
<td>4.112</td>
<td>−1.972</td>
</tr>
<tr>
<td>Cys</td>
<td>38</td>
<td>−3.697</td>
<td>1.956</td>
<td>3.979</td>
<td>4.658</td>
</tr>
<tr>
<td>Arg</td>
<td>39</td>
<td>—</td>
<td>—</td>
<td>0.968</td>
<td>—</td>
</tr>
<tr>
<td>Ala</td>
<td>40</td>
<td>—</td>
<td>—</td>
<td>2.098</td>
<td>—</td>
</tr>
<tr>
<td>Lys</td>
<td>41</td>
<td>−5.532</td>
<td>3.035</td>
<td>3.510</td>
<td>1.092</td>
</tr>
<tr>
<td>Arg</td>
<td>42</td>
<td>−3.191</td>
<td>0.930</td>
<td>1.617</td>
<td>1.581</td>
</tr>
<tr>
<td>Asn</td>
<td>43</td>
<td>−3.978</td>
<td>2.226</td>
<td>1.687</td>
<td>−1.240</td>
</tr>
<tr>
<td>Asn</td>
<td>44</td>
<td>−7.742</td>
<td>6.090</td>
<td>4.659</td>
<td>−3.296</td>
</tr>
<tr>
<td>Phe</td>
<td>45</td>
<td>−8.265</td>
<td>5.886</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lys</td>
<td>46</td>
<td>—</td>
<td>—</td>
<td>1.681</td>
<td>—</td>
</tr>
<tr>
<td>Ser</td>
<td>47</td>
<td>−3.927</td>
<td>1.956</td>
<td>4.429</td>
<td>5.695</td>
</tr>
<tr>
<td>Ala</td>
<td>48</td>
<td>−3.202</td>
<td>1.044</td>
<td>1.494</td>
<td>1.037</td>
</tr>
<tr>
<td>Glu</td>
<td>49</td>
<td>—</td>
<td>—</td>
<td>3.535</td>
<td>—</td>
</tr>
<tr>
<td>Asp</td>
<td>50</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cys</td>
<td>51</td>
<td>−6.489</td>
<td>4.788</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Met</td>
<td>52</td>
<td>−6.803</td>
<td>4.790</td>
<td>5.322</td>
<td>1.223</td>
</tr>
<tr>
<td>Arg</td>
<td>53</td>
<td>−6.537</td>
<td>4.267</td>
<td>4.339</td>
<td>0.167</td>
</tr>
<tr>
<td>Thr</td>
<td>54</td>
<td>−5.769</td>
<td>3.458</td>
<td>3.874</td>
<td>0.959</td>
</tr>
<tr>
<td>Cys</td>
<td>55</td>
<td>−6.974</td>
<td>5.261</td>
<td>4.235</td>
<td>−2.363</td>
</tr>
<tr>
<td>Gly</td>
<td>56</td>
<td>−5.210</td>
<td>3.478</td>
<td>2.621</td>
<td>−1.974</td>
</tr>
<tr>
<td>Gly</td>
<td>57</td>
<td>—</td>
<td>—</td>
<td>1.044</td>
<td>—</td>
</tr>
<tr>
<td>Ala</td>
<td>58</td>
<td>—</td>
<td>—</td>
<td>1.121</td>
<td>—</td>
</tr>
</tbody>
</table>
Table S2. Amide O/C state statistics from simulation

<table>
<thead>
<tr>
<th>Residue</th>
<th>No.</th>
<th>N_O</th>
<th>N_FO</th>
<th>max(n_O)</th>
<th>max(n_FO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>3</td>
<td>78,768</td>
<td>86,307</td>
<td>5</td>
<td>576</td>
</tr>
<tr>
<td>Phe</td>
<td>4</td>
<td>459</td>
<td>482</td>
<td>3</td>
<td>2,212</td>
</tr>
<tr>
<td>Cys</td>
<td>5</td>
<td>145</td>
<td>150</td>
<td>2</td>
<td>104,293</td>
</tr>
<tr>
<td>Leu</td>
<td>6</td>
<td>729</td>
<td>743</td>
<td>3</td>
<td>13,312</td>
</tr>
<tr>
<td>Glu</td>
<td>7</td>
<td>280</td>
<td>288</td>
<td>2</td>
<td>117,517</td>
</tr>
<tr>
<td>Tyr</td>
<td>10</td>
<td>3,754</td>
<td>3,821</td>
<td>3</td>
<td>2,921</td>
</tr>
<tr>
<td>Thr</td>
<td>11</td>
<td>78,782</td>
<td>86,593</td>
<td>6</td>
<td>294</td>
</tr>
<tr>
<td>Gly</td>
<td>12</td>
<td>7,013</td>
<td>7,063</td>
<td>2</td>
<td>2,003</td>
</tr>
<tr>
<td>Cys</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Lys</td>
<td>15</td>
<td>69,107</td>
<td>74,871</td>
<td>4</td>
<td>410</td>
</tr>
<tr>
<td>Ala</td>
<td>16</td>
<td>2,425</td>
<td>2,493</td>
<td>3</td>
<td>14,459</td>
</tr>
<tr>
<td>Arg</td>
<td>17</td>
<td>21,075</td>
<td>21,933</td>
<td>6</td>
<td>10,498</td>
</tr>
<tr>
<td>Ile</td>
<td>18</td>
<td>28</td>
<td>28</td>
<td>1</td>
<td>121,241</td>
</tr>
<tr>
<td>Ile</td>
<td>19</td>
<td>10,217</td>
<td>10,326</td>
<td>3</td>
<td>1,119</td>
</tr>
<tr>
<td>Arg</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Tyr</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Phe</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Tyr</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Asn</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Ala</td>
<td>25</td>
<td>99,171</td>
<td>111,430</td>
<td>6</td>
<td>194</td>
</tr>
<tr>
<td>Lys</td>
<td>26</td>
<td>420</td>
<td>421</td>
<td>2</td>
<td>20,898</td>
</tr>
<tr>
<td>Ala</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Gly</td>
<td>28</td>
<td>42</td>
<td>45</td>
<td>2</td>
<td>291,667</td>
</tr>
<tr>
<td>Leu</td>
<td>29</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>776,924</td>
</tr>
<tr>
<td>Cys</td>
<td>30</td>
<td>9,841</td>
<td>10,032</td>
<td>4</td>
<td>1,366</td>
</tr>
<tr>
<td>Gln</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Thr</td>
<td>32</td>
<td>17,462</td>
<td>17,799</td>
<td>3</td>
<td>540</td>
</tr>
<tr>
<td>Phe</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Val</td>
<td>34</td>
<td>15,410</td>
<td>15,722</td>
<td>3</td>
<td>672</td>
</tr>
<tr>
<td>Tyr</td>
<td>35</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>937,750</td>
</tr>
<tr>
<td>Gly</td>
<td>36</td>
<td>90</td>
<td>98</td>
<td>3</td>
<td>463,949</td>
</tr>
<tr>
<td>Gly</td>
<td>37</td>
<td>76</td>
<td>81</td>
<td>2</td>
<td>439,529</td>
</tr>
<tr>
<td>Cys</td>
<td>38</td>
<td>104</td>
<td>110</td>
<td>2</td>
<td>439,154</td>
</tr>
<tr>
<td>Arg</td>
<td>39</td>
<td>91,524</td>
<td>101,809</td>
<td>5</td>
<td>4,756</td>
</tr>
<tr>
<td>Ala</td>
<td>40</td>
<td>8,151</td>
<td>8,302</td>
<td>4</td>
<td>1,433</td>
</tr>
<tr>
<td>Lys</td>
<td>41</td>
<td>301</td>
<td>324</td>
<td>2</td>
<td>37,624</td>
</tr>
<tr>
<td>Arg</td>
<td>42</td>
<td>23,414</td>
<td>24,733</td>
<td>5</td>
<td>1,343</td>
</tr>
<tr>
<td>Asn</td>
<td>43</td>
<td>19,973</td>
<td>21,120</td>
<td>5</td>
<td>2,498</td>
</tr>
<tr>
<td>Asn</td>
<td>44</td>
<td>23</td>
<td>23</td>
<td>1</td>
<td>143,438</td>
</tr>
<tr>
<td>Phe</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Lys</td>
<td>46</td>
<td>20,456</td>
<td>21,394</td>
<td>5</td>
<td>1,119</td>
</tr>
<tr>
<td>Ser</td>
<td>47</td>
<td>39</td>
<td>39</td>
<td>1</td>
<td>107,407</td>
</tr>
<tr>
<td>Ala</td>
<td>48</td>
<td>31,451</td>
<td>32,537</td>
<td>3</td>
<td>378</td>
</tr>
<tr>
<td>Glu</td>
<td>49</td>
<td>306</td>
<td>306</td>
<td>1</td>
<td>16,036</td>
</tr>
<tr>
<td>Asp</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Cys</td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1,048,349</td>
</tr>
<tr>
<td>Met</td>
<td>52</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>440,280</td>
</tr>
<tr>
<td>Arg</td>
<td>53</td>
<td>45</td>
<td>48</td>
<td>3</td>
<td>339,677</td>
</tr>
<tr>
<td>Thr</td>
<td>54</td>
<td>129</td>
<td>140</td>
<td>3</td>
<td>249,296</td>
</tr>
<tr>
<td>Cys</td>
<td>55</td>
<td>60</td>
<td>61</td>
<td>2</td>
<td>634,880</td>
</tr>
<tr>
<td>Gly</td>
<td>56</td>
<td>2,333</td>
<td>2,505</td>
<td>4</td>
<td>4,954</td>
</tr>
<tr>
<td>Gly</td>
<td>57</td>
<td>78,840</td>
<td>86,797</td>
<td>6</td>
<td>340</td>
</tr>
<tr>
<td>Ala</td>
<td>58</td>
<td>68,407</td>
<td>73,791</td>
<td>5</td>
<td>341</td>
</tr>
</tbody>
</table>

Fig. S1. Pair correlation function, \(g(r)\), for water oxygens relative to the amide hydrogen, averaged over all amides and over all O-state frames (blue solid curve), over all C-state frames (red solid curve), or over all frames (gray dotted curve). (Inset) The corresponding running coordination number, \(N(r)\). The primary coordination number, \(N_{W}\), including water oxygens within the 2.6-Å cutoff (vertical gray line), is given for the two states.
Fig. S2. Pair correlation function, \( g(r) \) (Left) and running coordination number, \( N(r) \) (Right) for water oxygens relative to the indicated amide hydrogen, averaged over all O-state frames (blue curve) or all C-state frames (red curve). The primary coordination number, \( N_W \), including water oxygens within the 2.6-Å cutoff (vertical gray line), is given for the two states.

Fig. S3. Correlation of O/C RMS deviation \( \sigma_{\text{loc}} \) from simulation (Left) and inverse RMS fluctuation \( 1/\sigma_B \) from crystal structure (Right), in both cases averaged over atoms within 7 Å from amide nitrogen, with O/C free energy difference \( \beta \Delta G_{\text{exp}} \) from experimental HX rates.

Fig. S4. Snapshots of O states for nine selected amides with the N–H group and the two (or three) primary waters in space-filling and other waters within a 7-Å sphere (and in tunnels or pores) in stick representation. To the left, the backbone conformation is shown for the selected O-state frame (dark gray) and for the first C-state frame in the trajectory (light gray). To the right, the molecular surface (1.4 Å probe radius, standard vdW radii) is shown. For Gly36 and Lys41, several snapshots (with and without water tunnels) are shown. Click on a structure to activate the interactive mode.

Fig. S5. Normalized C-state residence time distribution, \( \psi_C(\tau) \), for the indicated amides (gray bars) and single-exponential fit (red curve). The number, \( N_C \), of visits to the C state during the trajectory is indicated for each amide as is the C-state MRT, \( \tau_C \), obtained from the fit. For 16 of the 20 amides shown, the fitted \( \tau_C \) is within 10% of the value calculated (without binning-error correction) from the trajectory statistics as \( \tau_C = \Delta \tau N_C/N_C \).

Fig. S6. Schematic representation of a proton transfer within an encounter complex with two water molecules directly coordinated to the N–H group. At the endpoints of the proton transfer process, at least one of these waters accepts a H-bond (blue dashed lines) from the N–H or N–D group. In the intermediate structure, both water molecules donate H-bonds (red dashed lines) to the imidate nitrogen. More than two water molecules may participate in the more or less concerted proton transfer (a third water molecule is shown to the right in the configuration at the top).

Fig. S7. Fractional populations of C14–C38 disulfide rotamers (M1 blue, M2 magenta, other gray) in the O state for the amides of BPTI when the C state is constrained to the experimentally dominant M1 conformation.

Fig. S8. Continuous (Top) and discretely sampled (Bottom) trajectory of alternating residences of an amide in the O and C states.
$g(r)$

$N_{w,o} = 2.00$

$N_{w,c} = 0.36$
\begin{align*}
g(r) &\quad N(r) \\
3 &\quad 2.00 \quad 0.90 \\
4 &\quad 2.00 \quad 0.02 \\
5 &\quad 2.00 \quad 0.04 \\
6 &\quad 2.00 \quad 0.16 \\
7 &\quad 2.00 \quad 0.88 \\
10 &\quad 2.00 \quad 0.90 \\
\end{align*}
g(r)

N(r)

2.00 0.03

2.00 0.71

2.00 0.62

2.00 0.74

2.00 0.01

0.00

2.00 0.71

2.00 0.03
residue 18 — frame 64345

residue 29 — frame 3073482

residue 36 — frame 67279
residue 36 — frame 151272

residue 36 — frame 286220

residue 41 — frame 286406
Figure S6. Snapshots of O states for 9 selected amides with the NH group and the 2 (or 3) primary waters in space-filling and other waters within a 7-Å sphere (and in tunnels or pores) in stick representation. To the left, the backbone conformation is shown for the selected O-state frame (dark gray) and for the first C-state frame in the trajectory (light gray). To the right, the molecular surface (1.4 Å probe radius, standard vdW radii) is shown. For Gly36 and Lys41, several snapshots (with and without water tunnels) are shown. Click on a structure to activate the interactive mode!
\( \psi_C(\tau) \)

\( N_C = 78769, \tau_C = 2.86 \text{ (ns)} \)

\( N_C = 3755, \tau_C = 62.41 \text{ (ns)} \)

\( N_C = 78783, \tau_C = 2.88 \text{ (ns)} \)

\( N_C = 7014, \tau_C = 36.30 \text{ (ns)} \)

\( N_C = 69108, \tau_C = 3.25 \text{ (ns)} \)

\( N_C = 21076, \tau_C = 7.45 \text{ (ns)} \)

\( N_C = 10218, \tau_C = 24.61 \text{ (ns)} \)

\( N_C = 99172, \tau_C = 2.18 \text{ (ns)} \)

\( N_C = 421, \tau_C = 617.03 \text{ (ns)} \)

\( N_C = 9842, \tau_C = 25.74 \text{ (ns)} \)

\( \tau (s) \)
\[ N_C = 17463, \ \tau_C = 14.49 \text{ (ns)} \]

\[ N_C = 15411, \ \tau_C = 16.86 \text{ (ns)} \]

\[ N_C = 91525, \ \tau_C = 2.33 \text{ (ns)} \]

\[ N_C = 8152, \ \tau_C = 30.20 \text{ (ns)} \]

\[ N_C = 23415, \ \tau_C = 8.26 \text{ (ns)} \]

\[ N_C = 19974, \ \tau_C = 8.84 \text{ (ns)} \]

\[ N_C = 20457, \ \tau_C = 9.86 \text{ (ns)} \]

\[ N_C = 31452, \ \tau_C = 7.95 \text{ (ns)} \]

\[ N_C = 78841, \ \tau_C = 2.84 \text{ (ns)} \]

\[ N_C = 68408, \ \tau_C = 3.29 \text{ (ns)} \]
C14-C38 state populations

residue

0
0.2
0.4
0.6
0.8
1