Water is crucial for protein function. There is a minimal hydration level required for full protein functionality. Molecular dynamics (MD) simulations reveal how intimately protein fluctuations couple to solvent mobility. Thus, proteins rely on thermal access to “active” conformations. This is achieved by a continuous rearrangement of hydrogen bonds (H-bonds) with protein residues, a process involving water entering or leaving hydration sites located inside the protein.

Water molecule exchange has a broad time scale depending on its location within the protein and the nature of its hydration site. Surface or hydration water molecules have residence times of tens to hundreds of picoseconds and up to 1 ns for cavities on the protein surface. Internal waters, times (in internal and surface sites), diffuse biexponentially, supporting this three-site scenario. For a favorable orientation of the Thr203 side-chain, the hole often passes through the 21st amino acid and has much longer residence times, which play an important role in satisfying the H-bonding requirements of the internal protein residues, are regarded as the 21st amino acid and have much longer residence times, from nanoseconds to tens of microseconds.

Various experimental techniques, such as NMR, X-ray, neutron diffraction, femtosecond fluorescence, and Raman spectroscopy are used to study water exchange in proteins. The results from these various methods are sometimes contradictory, perhaps because they report on processes occurring on different time scales. MD simulations follow protein and water dynamics in atomistic detail, allowing the calculation of various water properties such as residence times (in internal and surface sites), diffusion coefficients, radial distribution functions, H-bond networks, and free energies. Much MD work was devoted to water exchange kinetics in proteins, mostly surface waters, but some studies have explored internal and cavity water exchange. For example, in P450 there is a large number of water molecules in the active site cavity, which is connected to the bulk by up to seven frequently traveled channels. In the current work we use classical MD simulations to study water exchange at the green fluorescent protein (GFP) active site, which hosts a single water molecule and connects to the bulk through a single transiently opened channel not easily discerned in the X-ray structures.

GFP is a soluble protein widely used in biological and medical research as a fluorescence marker. It has a β-barrel structure, consisting of 11 aligned β-strands and a chromophore (Figure 1) residing on an internal α-helix traversing the barrel axis. Chromophore biosynthesis, which occurs after folding to the barrel structure, involves autocatalytic cyclization of three consecutive amino acids: Ser65, Tyr66, and Gly67 (see Figure 1). From the X-ray structure it was inferred that the chromophore is well protected inside the β-barrel. It was therefore thought that the main function of the barrel is to shield it from external agents, especially oxygen, which could quench its fluorescence.
Several MD simulations of GFP were conducted,\textsuperscript{24,25,35,37} of which only one explicitly focused on water exchange in-and-out the protein β-barrel.\textsuperscript{24} This 50 ns simulation identifies water exchange in mature wild-type GFP (wt-GFP) along the barrel axis, with an alternate pore in the chromophore vicinity apparently only in the premature protein (i.e., prior to the cyclization reaction). It was proposed\textsuperscript{24} that the latter allows oxygen and water entry, which are required for chromophore formation. The number of exchange events ranges from 3 to 50 (Table 1 in ref 24) for various GFP mutants and conditions. Another study\textsuperscript{25} finds that 16 water molecules exchange between the chromophore and the bulk during a 60 ns simulation of the S205V GFP mutant. However, both works did not analyze these exchange events in much detail.

GFP is the only known protein in which photon absorption initiates an excited state proton transfer (ESPT) reaction,\textsuperscript{38} releasing a proton in about 10 ps\textsuperscript{39} and leaving behind a green fluorescing anion. The released proton travels along a preformed H-bond chain,\textsuperscript{40} from the OH atom of the chromophore phenolic ring (originating from Tyr66, subsequently referred to as C66-OH) via a water molecule that we term the “privileged water molecule” (PWM), to Ser205-OH, finally protonating the anionic Glu222. This has been corroborated by time-resolved IR from the Glu222 carboxylate moiety.\textsuperscript{41} Interestingly, on a longer (ns) time scale, the fluorescence from the protonated chromophore (corrected for its excited state lifetime) develops a power-law tail,\textsuperscript{32} which behaves like $t^{-1/2}$ (t is time) below 230 K, changing to $t^{-3/2}$ at higher temperatures.\textsuperscript{43} This was interpreted as proton diffusion along an effectively one-dimensional “proton wire” identified\textsuperscript{44,45} inside the barrel beyond Glu222, with proton escape from the PWM to external solution giving rise to the switch in asymptotic power above 230 K.\textsuperscript{46} In the X-ray structure, an exit pathway from the PWM was identified\textsuperscript{44} via the backbone carbonyl of His148, although this might not be the most efficient way for a proton to travel.

In the present work we do not simulate proton migration, which requires quantum mechanical methodologies, only water migration that can be tracked with a classical force-field. We focus on the PWM, because if it can exit, so might the proton. This would be facilitated if water exit or entry involves the formation of a “water wire” connecting the chromophore with the bulk. Such a “transient aqueduct mechanism” has been observed\textsuperscript{10} in a millisecond trajectory from bovine pancreatic trypsin inhibitor (BPTI).\textsuperscript{11} Our study of the PWM exchange thus achieves a dual goal. It presents a detailed analysis of internal (“crystallographic”) water exchange from within a rigid protein, found to be unexpectedly facile, providing at the same time an indirect test for the proton escape hypothesis.

More specifically, we first identify the “hole in the barrel” through which the PWM exchanges with bulk water (Section 3.1.1), showing that its width determines the frequency of water exchange. We locate the pathways along which the PWM travels, showing that they sometimes involve water wires (Section 3.2). We show that the PWM crosses the barrel walls by a “flip-flop” motion that leads to an interesting time evolution of its mean-squared displacement (MSD), changing from subdiffusive at short times to superdiffusive at longer times and finally to normal diffusion in the bulk (Section 3.3).

2. METHODS

2.1. MD Simulations. The MD simulations for GFP in water were performed with the NAMD 2.8 program,\textsuperscript{47} using the Charmm27 force field\textsuperscript{58} and predominantly with the TIP3P water model.\textsuperscript{49} Force field parameters for the neutral chromophore were taken from Reuter et al.\textsuperscript{50} (see Tables S1–S5 in the Supporting Information (SI)). Data visualization was performed using VMD 1.9.\textsuperscript{51} Analysis scripts were written in the Tcl 8.4.1 scripting language embedded into VMD and extended to include VMD commands.

Three different GFP crystal structures, Protein Data Bank (PDB) entries 1EMB,\textsuperscript{40} 1EME,\textsuperscript{52} and 2WUR\textsuperscript{45} (resolutions of 2.13 Å, 2.50 Å, and 0.90 Å, respectively) were used as initial structures. Several VMD plugins were utilized to set up the system. Hydrogen atoms were added using the Psfgen plugin (ver. 1.4.7). The Solvate plugin (ver. 1.4) was used to solvate the protein in a cubic water box with 10 Å padding in each dimension (number of added water molecules: 9323, 8818, and 10 520 for PDB structures 1EMB, 1EME, and 2WUR, respectively). The Autoionize plugin (ver. 1.3) was used to neutralize the electric charge of the molecule by adding Cl$^-$ and Na$^+$ ions, reaching ionic strength of 0.05 M.

Histidine protonation states for 1EME and 1EMB were determined using the PROPKA 3.0 software.\textsuperscript{53} In 2WUR, hydrogen atoms are visible, therefore His protonation states were taken directly from the X-ray structure. A list of histidine residues and their protonation states is presented in Table S6 and Figure S2 of the SI. The total number of atoms in the simulation was 31 782, 30 220, and 35 174 for PDB structures 1EMB, 1EME, and 2WUR, respectively.

The short-range electrostatic and van der Waals interactions were gradually switched off between 10–12 Å. The long-range electrostatic interactions were calculated using the particle mesh Ewald method.\textsuperscript{54} In accordance with the use of a rigid water model, hydrogen−oxygen and hydrogen−hydrogen distances of water were constrained using the SHAKE algorithm.\textsuperscript{55} The integration time step was 1 fs, with coordinates saved every 500 fs for analysis. The temperature during the simulation was kept constant using the Langevin thermostat with a damping coefficient of 5 ps$^{-1}$. The pressure was maintained using the Langevin barostat\textsuperscript{46} with a piston period of 200 fs and damping time of 50 fs.

Figure 1. GFP chromophore is synthesized inside the folded protein from three consecutive amino acid residues: Ser65 (green), Tyr66 (blue), and Gly67 (red), following several steps of oxidation and dehydration.\textsuperscript{34} This creates the aromatic imidazolone ring whose π-system couples, via the C66-C66 double bond, to the Tyr66 phenolic ring from which the proton leaves upon photoexcitation. Electron density transferred from the phenolic to the imidazolone ring stabilizes the excited anion, giving rise to the enhanced photoacidity. Chromophore atoms in PDB format are shown in Figure S1 and Table S1 of the SI.
Simulations were performed at 5 different temperatures, 300, 310, 320, 330, and 340 K, for each of the three initial structures. After energy minimization using the conjugate gradient algorithm, each trajectory was propagated under constant volume and temperature for 50 ps for thermal equilibration (to target temperatures). Then it was continued under constant temperature and pressure (1 atm) for 18 ns. Data for analysis was taken after the system had reached equilibrium. The time required for equilibration was 0.5–1.5 ns, as determined from the stabilization times of the RMSD, radius of gyration, and the number of H-bonds between water and protein (Figure S3 in the SI). The total simulation time was 298 ns.

In the simulations reported herein, the production runs were performed under NPT conditions, which better match experimental conditions. In order to verify that the thermostat/barostat do not introduce artifacts, we ran three test simulations in the more traditional NVE ensemble. Likewise, the water model used was the standard 3-site TIP3P water model, but the TIP4P/2005 water model was checked in two test simulations. The additional test simulations were set as follows: (1) NVE at 310 K with TIP3P, (2) NVE at 330 K with TIP3P, (3) NPT at 310 K with TIP4P/2005, and (4) NVE at 310 K with TIP4P/2005 water model. The NVE simulation was run for 18 ns after a 1 ns equilibration in NPT. The rest of the procedures were similar to the original runs.

2.2. Analysis of Simulation Data. The PWM of interest here is the one accepting the H-bond from the Cro66-OH moiety. (Two oxygen atoms were considered H-bonded if the distance between them was under 3.5 Å, one of them carried a hydrogen atom, and the smallest O−H−O angle was larger than 140°). There were, however, times when the chromophore rotated its OH group away from the binding site, and the water occupying it was an H-bond donor rather than acceptor. Such donor waters were included in the analysis whenever they simultaneously H-bonded also to OG-Ser205.

Altogether we have collected 843 PWM exchange events, mostly above 310 K. At the lower temperatures (300 and 310 K), there were only a few exchange events, hence events taking place before the end of the equilibration stage were also included in the analysis.

Root mean square deviations (RMSD) of amino-acid residues were calculated by the following procedure. Every time the coordinates were saved (i.e., every 0.5 ps), the protein was aligned with the respective equilibrated structure using all backbone atoms (N, CA, C, O). Then the backbone atom RMSD was calculated with respect to the equilibrated structure. Water molecule RMSD was calculated for its oxygen atom.

2.3. MolAxis Calculations. MolAxis is a tool for the identification of high clearance pathways or corridors, which represent molecular channels in the complement space of the protein. We have used this program to search for spatial channels emanating from the Cro66-OH moiety, in order to corroborate the pathways found from MD simulations. The program assigns flux scores to the detected channels according to their width and length (high score for shorter and wider channels). This can be used to estimate the plausibility of the passage of small molecules through each individual channel. MolAxis was previously utilized for detecting channels in GFP structures by Li et al.

3. RESULTS AND DISCUSSION

3.1. The Hole in the Barrel Wall. In GFP, 50% of the residues are arranged in 11 anti-parallel β-strands forming the barrel walls (Figure 2). This regularity is achieved by pairs of H-bonds between two residues from adjacent strands, connecting the backbone carbonyl of one to the amide nitrogen of the other, and vice versa. It was noted that “the ideal symmetry of the β-barrel is broken only ... between strands 7 and 8”. Nevertheless, there is also an irregularity between strands 7 and 10. Water conducting pores were observed to form in both of these locations in premature (precyclized) GFP. Here we show that in mature GFP, a pore forming between strands 7 and 10 serves as a focal point for the PWM exchange.

3.1.1. Structural Evidence. Figure 2 shows the (simulated) average distances between residues on opposing strands 7 and 8 (black) versus 7 and 10 (red), at low versus high temperatures (T). Both strand pairs are separated to larger distances than all other opposing strand pairs, which average 3.20 Å at 300 K, increasing slightly to 3.33 Å at 340 K. However, while the average distances between opposing residues on strands 7 and 8 increase only slightly with T, for residues on strands 7 and 10 (specifically the S205−N146 and T203−H148 pairs), there is a substantial increase with T.

This “barrel cracking” transition is depicted with more detail in Figure 3, which shows the distribution of distances between three pairs of atoms from the above-mentioned strand 7–10 residues. The distribution has 2 or 3 wide peaks, reflecting discrete “states” that the distance can adopt. For
these 3 pairs, the longest distances are only accessed at $T = 320$ K or higher. To check whether this transition extends further inside the protein, we have monitored the chromophore geometry. The chromophore OH group is situated directly between strands 7 and 10, but we did not find any analogous transition for the RMSD of its constituting atoms or for the dihedral angle between its rings. Such a dramatic change is not seen even for the nearby residues along these strands, so we locate the “hole in the barrel” between residue pairs 146–205 and 148–203.

The instability of strand 7 can be partly attributed to the long loop (14 residues) separating it from strand 6. Although adjacent by sequence, strand 6 is almost on the opposite face of the barrel. The interconnecting 129–142 loop moves rather freely in solution, conveying thermal energy to strand 7. This is reflected in the gradual decrease of the 7–10 gap as we go from the N-terminus of strand 7 (the end connected to this loop) to its C-terminus.

3.1.2. Dynamic Evidence. GFP is a remarkably rigid protein, with backbone atoms deviating, on average, only 0.9 Å from their crystal structure positions (1 ns trajectory from ref 35). Here we find that the backbone-atom RMSD (averaged over all GFP residues and all PDB structures) is 1.17 Å ($\pm 0.88$) at 300 K. This is somewhat larger than the previously reported value$^{35}$ due to our longer trajectories. Excluding the loops, the average $\beta$-barrel backbone RMSD is 0.85 Å ($\pm 0.32$) at 300 K, increasing to 1.19 Å ($\pm 0.44$) at 340 K. In comparison, the averaged RMSD’s for residues on strands 7 and 10 are substantially higher, 0.91 Å ($\pm 0.34$) at 300 K and 1.32 Å ($\pm 0.37$) at 340 K.

Backbone RMSD’s for all 229 GFP residues at the 5 temperatures are listed in Table S7 of the SI. Figure S4 shows the trends graphically (cf. Figure 4 in ref 35). Noticeable motions occur in the loop segments, and these correspond to the maxima in the RMSD plot (Figure S4). The longest loop is the 129–142 loop connecting strands 6 and 7, and its RMSD increases most substantially with $T$.

Figure 4A shows the RMSD’s of the residues on strands 7 and 10, in the low and high temperature regimes. There is a noticeable increase with $T$, but it is not equal for both residues of an H-bonded pair. Rather, strand 7 residues 143–145, which are adjacent to the long loop (129–142), have the largest RMSD’s, particularly at the higher temperatures. This suggests that thermal energy can be transferred from the long loop to adjacent residues on strand 7. On strand 10, only residues 203–205 have large RMSD’s at high $T$, which could be explained from their interaction with the Cro66-OH moiety. Thus, the motion of the long loop and the chromophore can explain the large increase in the 7–10 gap noted above. The increased gap may explain the enhanced H/D exchange noted for these strands in some NMR experiments.$^{60-62}$

Figure 4B shows the correlation between the RMSD of a given residue and the distance within the pair it participates in. It is seen that the increase of residue RMSD with increasing pair distance does not occur equally for both strands. Whereas for strand 7 we see a positive correlation between interstrand distance and residue RMSD, in strand 10 we do not see such a correlation. We can attribute the correlation in strand 7 to the proximity of its N-terminus to the long 129–142 loop, which transfers thermal energy from solution. The residues at strand 10, held by quite short loops, have their RMSD values affected by proximity to the
chromophore rather than by proximity to a long loop. Thus, the maximal RMSD occurs for residues 203 and 205, which can be considered as the focal point for the exit/entry path (cf. Figure 3). This conclusion will be corroborated below from the statistics of water exchange events.

Figure 5 shows the β-sheet residues on opposite sides of GFP colored according to their 320 K RMSD values from Table S7. It is clearly seen that RMSD’s for the water exchange side are higher than those for the opposite side. The only “weak spot” (white) in the barrel walls is indeed in the vicinity of the 7–10 gap.

3.1.3. Correlation with Water Exchange Events. The chromophore OH group is situated between strands 7 and 10, where we locate the “hole in the barrel” through which the PWM can exit, whereas we hardly find PWM exchange through the 7–8 gap (3% of the total). As Figure 6A shows, the 7–10 gap opening correlates with the rate of PWM exchange. This figure records the time evolution of the distances between three opposing residue pairs on strands 7 and 10 (shown is just one of the bonds between the pairs; the other behaves similarly). In this trajectory segment, they tend to increase with time. The L207–N144 distance starts increasing at 4 ns, followed at 10 ns by the T203–H148 distance. Water exchange events are marked by vertical red lines. It is seen that the increase in the T203–H148 distance at 10 ns is also the onset of rapid water exchange events.

This correlation between interstrand distances and PWM exchange extends to individual exchange events. Figure 6B shows a blow-up of the T203–H148 distance trajectory, in which exchanges tend to occur as this distance momentarily increases above its average value.

Figure 6C shows these distances averaged over time slots of 20 ps before and after an exchange event (the “open” state) minus the average over all other times (the “closed” state). All six distances increase at times close to exchange events, by almost 1 Å for some pairs. The increase is largest at the T203–H148 and S205–N146 pairs, in agreement with their role in forming the “hole in the barrel” (cf. Figure 3).

3.2. The Exchange Pathways. Having identified the “hole in the barrel”, we continue by mapping the complete exchange pathway for the PWM using both MD trajectories and the MolAxis software. Subsequently, we show that the hole sometimes fills with water, which might readily conduct protons away from the GFP chromophore.

3.2.1. Water Path Analysis. We define the path of a water molecule as the ordered list of all O, N, and S protein atoms that it encounters (creates an H-bond to) along its trajectory, up to a distance of 20 Å from the Cro66-OH site. Since we focus here on water exchange between the chromophore and the bulk, we only consider water molecules that do not spend more than 10% of their time further inside the protein than the Cro66-OH site. Atoms considered as “protein interior” are listed in Table S8 of the SI.

In order to trace an average path, a normalized order of the m protein atoms visited along the path was defined. For a given water molecule trajectory, let ni be the index of first encounter with atom i, then its normalized order oi is defined by

\[ o_i = n_i / m \]

For example, if a water molecule binds to atoms a, b, c, and d in the order a → b → c → d, then their normalized orders will be \( o_a = 1/4 \), \( o_b = 2/4 \), \( o_c = 3/4 \), and \( o_d = 1 \), so that atoms to which the water binds at early stages have low normalized orders, and vice versa. When several atoms are encountered simultaneously, their relative order is selected arbitrarily.

For each exit/entry path, each encountered protein atom i (except for the Cro66-OH site itself) was assigned a normalized order oi. The \( o_i \) were averaged over all paths in which the atom participates (at all trajectories over all temperatures); this average is referred to by us as the “sequence index”. Due to the averaging process, atoms encountered simultaneously will tend to have similar sequence indices. Indices for atoms appearing in more than 10% of the paths are presented in Table S9 of the SI. All the residues on this list belong to strands 7 or 10, none to strand 8. This shows the dominance of the 7–10 gap over the 7–8 gap, which must be utilized less than 10% of the time. Indeed, analysis of the full contact list suggests that, near room temperature, the weight of the 7–8 path is only 3%.

There are 27 different atoms in Table S9, of which 17 are oxygens, 10 nitrogens, and no sulfur. The average value of the path length, \( m_i \), is around 9. We notice that the sequence indices cluster around three values for both entrance and exit paths, as shown in Figure S5 of the SI. Such behavior indicates a stepwise (rather than continuous) progress along the pathway, where the atoms aligning the water path can be divided into three layers (ca. 3 in each), according to their sequence index. A water entering the protein will first bind to atoms in the external (surface) layer, perhaps diffuse on it for a while, then move to the middle layer, and finally to the innermost layer that surrounds the Cro66-OH binding site.

Figure 7 shows these atoms colored according to their sequence index, from red (early encounters) to blue (late encounters). The left (right) panel depicts entry (exit) pathways. The figure thus shows that the PWM uses similar paths on its way in and out. Figure 8 shows a schematic depiction of the entrance path with these three groups colored red, gray, and blue, where red is closest to the bulk and blue is near the chromophore.

3.2.2. MolAxis Paths. MolAxis employs computational geometry techniques to compute channels connecting a given point within a protein to the bulk solvent.9 We have utilized it for the Cro66-OH site (within a sphere of radius 3 Å,
calculation including H atoms) for three types of input data: (i) the X-ray structures from PDB files 2WUR, 1EME, and 1EMB; (ii) their minimized structures, and (iii) trajectory snapshots at 1, 5, and 18 ns. A total of 59 structures were analyzed, in which 225 paths were detected (2−8 paths per structure). In the MD structures, there are typically more paths at later times: on average (over all temperatures) 2.7, 4.5, and 5.5 paths per structure for 1, 5, and 18 ns, respectively. Perhaps this is connected with the slight increase of backbone RMSD with time seen in Figure S3.

Four main exit routes were identified from the chromophore area (Figure 9A): two (comprising 77% of the paths) from the side, and two from the top and bottom caps. Additionally, two minor paths from the opposite side of the barrel appear in MD snapshots from late times (18 ns) or high temperatures (340 K). In particular, MolAxis path 1 (which got the highest score) is between strands 7 and 10, as in our MD trajectories, while path 2 is between strands 7 and 8, as suggested in ref 35. The weights of paths 1 and 2 are 57% and 43%, respectively.

Further analysis was limited to 300 and 340 K (29 structures). For these we have assigned a sequence index (see Section 3.2.1) to all O, N, S atoms along the MolAxis path, leaving only the 10 most frequently observed ones (S appears only once, in Met153).

Figure 9B compares MolAxis paths 1 and 2 with the PWM path from our MD simulations by dividing the atoms found along these paths into three groups: (i) those appearing only in a MolAxis path (magenta), (ii) those only in the MD path (green), and (iii) 18 atoms that are common to both paths (white). The common atoms further establish the equivalence between the PWM path and MolAxis path 1. It is interesting that the path can be identified already in the X-ray structure (Figure 9A), although it is likely too narrow to transfer a water molecule, which starts exiting only around 310 K as thermal energy enhances the pore size fluctuations.

3.2.3. Proton Wire along the Water Exchange Path. The route that serves for water exchange between the chromophore and the bulk may also be used for proton transport. Particularly, rapid proton transport is expected if a single file of water molecules forms,63 creating a "water wire".
through the β-barrel wall. Formation of water wires was shown to couple to water exchange in BPTI in the so-called "transient aqueduct mechanism", and we now show that a similar mechanism is operative in GFP.

The algorithm we have used here searches for all water wires originating from the OH moiety of the chromophore that are up to 4 water molecule long. This cutoff was chosen because a water chain consisting of 4 water molecules has an average length of 9.3 Å, which is already outside the β-barrel wall. Length 3 corresponds to an average distance of 6.5 Å, which is occasionally on the protein surface.

Figure 10A shows the fraction of frames in which water wires of length ≥X were found (e.g., X = 2 includes wires consisting of 2, 3, and 4 water molecules). These values were averaged separately for low (300 and 310 K) and high (320, 330, and 340 K) temperatures. As we choose to focus on water exchange between the chromophore and the bulk, we only leave water wires ending at the protein exterior. This is achieved by deleting all wires for which the last atom is found at a distance of less than 3.5 Å from a "protein interior" atom (as defined in Table S8 of the SI). The figure shows the average value of the water wire probability for the whole simulation time (black), as well as near water exchange events (red) versus all other times (blue).

The high value for X = 1 suggests that, for most of the time, a water molecule is H-bonded to the Cro66-OH site. It decreases near a water exchange event, as the PWM leaves, but not very strongly. Hence the leaving PWM is replaced quite rapidly by another water molecule. Considering X = 3 and/or 4 as cases where a chromophore-bulk wire exists, we conclude that (at the average temperature of 320 K) such a connection exists more than 25% of the time. Moreover, the wire probability is greatly enhanced (about a factor of 2) near water exchange events.

Figure 10B shows a van’t-Hoff plot for the fraction of the simulation time during which water wires exist (X = 3 or 4). It is not quite linear, but the points near room temperature give an enthalpy of ΔH = 23–26 kJ mol⁻¹ for the process of wire formation. This is remarkably similar to the activation enthalpy obtained (Figure 6 in ref 46) from the switching times of the neutral chromophore fluorescence from r⁻¹/₂ to r⁻³/₂ behavior. The comparison of reaction enthalpy with activation enthalpy is justified for a highly endothermic reaction, since by well-known structure–reactivity correlations (see, e.g., refs 64–66) they coincide as ΔH → ∞.

Thus, the process that allows for occasional proton escape from the chromophore vicinity could be a through-barrel water wire formation. This may occur instead of, or in addition to, the proton transport through His148 (a strand 7 residue) advocated in ref 46. Indeed, recent ¹³C NMR
experiments suggest a vital role for strand 7 residues in ground-state proton transport in GFP.62

We have also checked the dependence of the probability for water-wire formation on the side-chain orientation of the adjacent Thr203 residue. In agreement with ref 67, we find that of the three possible rotamers of a threonine residue (Figure S6 in the SI), only two (dihedral angles of $-60^\circ$ and $180^\circ$) have appreciable population for Thr203 in wt-GFP at room-temperature (Figure S7 in the SI). The $180^\circ$ rotamer has its hydrophobic methyl group pointing toward the Cro66-OH site, whereas the $-60^\circ$ rotamer does not. Interestingly, during periods when a water wire exists, the population of the $-60^\circ$ rotamer is greatly enhanced (Figure S8 in the SI). Thus, hydrophobic interactions impede through-pore water-wire formation, and Thr203 rotation (Figure 5 of ref 44) is a prerequisite for its formation. At times near water exchange (20 ps before and after water binds to the chromophore) we see the OG1 atom of the $-60^\circ$ rotamer of Thr203 bind to the first water in the water wire 25% of the time. For the anionic form of the chromophore, the $-60^\circ$ rotamer becomes dominant, and thus one may expect enhanced water-wire formation and proton escape following excited-state proton transfer in GFP.

Exchanging water molecules utilize the water wires to proceed toward their destination (the chromophore or the bulk). Figure 11 shows a distance profile of a single water molecule entering the protein and binding to the chromophore. Its location within the water wire (if it exists) is recorded as blue dots, where ‘1’ indicates binding to the chromophore, ‘2’ means that it is second in line from the chromophore (i.e., bonded to the water which is bonded to the chromopore), etc. It is seen that as the water makes its way toward the chromophore, it also proceeds in a stepwise fashion along the water wire modifying its location from 4 to 1. Thus, the PWM takes advantage of the water wire, progressing by switching places with the other wire molecules.

3.3. Kinetics. So far the discussion has focused on the “hole in the barrel” through which the PWM is transported. We now switch to analyze its dynamics and kinetics along the entry/exit pathways.

3.3.1. Binding Profile for the Exchanging Water. Figure 12A shows three representative examples for the time dependence of the distance between the PWM oxygen atom and the OH moiety of the GFP chromophore. Three additional examples for each $T$ are shown in Figure S9 of the SI. The bound state is characterized by a short H-bonding distance, ca. 2.8 Å. It lives for a considerable amount of time, during which distance fluctuations are minimal. The other extreme is bulk water, characterized by large distances and large variations thereof, reflecting its large self-diffusion coefficient in the bulk. Intermediate between these two extremes is the water diffusing on the protein surface or its interior. We can identify these states more quantitatively by correlating the distance with the H-bonding pattern of the PWM along its path.
randomly selected water molecules that were at a distance >5 Å from any protein atom during the calculation time (500 ps per water molecule).

3.3.2. The Flip-Flop Mechanism. The distance profile in Figure 12B exhibits time periods in which the water is H-bonded to some protein site, mediated by periods of large “jumps”. This is in line with the water pathways discussed in Section 3.2.1, which traverse three layers, or pockets. We now describe how the water moves between these “binding pockets”.

Jumps along the water pathway are more clearly identified from the (absolute value of the) difference in the water–chromophore distance, \( r \), for sequential frames (times \( t \)):

\[
\Delta r(t) = |r(t) - r(t-1)|
\]

At each temperature, we have averaged \( \Delta r(t) \) for times when \( r(t) \) has the same value. These are shown in Figure 13A.

The minimum in \( \Delta r \) at \( r = 2.75 \) Å represents the average H-bond distance between the water and the chromophore. A second minimum, at 5.25 Å, corresponds with the distance to the middle (gray) layer in Figure 8. A shallower minimum around 7 Å likely represents the PWM on the surface, the outer (red) layer in Figure 8. The intervening maxima indicate that the water rapidly transits between these sites. We note that these results do not depend on the ensemble or water model, as similar behavior is observed for test simulations in the NVE ensemble and/or with the TIP4P/2005 water model (Figure S10 in the SI). Yet both TIP3P and TIP4P are rigid water models. We have not investigated here the effect of flexible or polarizable water models.

Water translation along the path is coupled to angular reorientation, as can be visualized in movie file S1 in the SI. Let us define \( \theta \) as the angle between the PWM dipole moment at time \( t \) and its orientation for the bound PWM in the minimized conformation. Figure 13B shows this angle as a function of the water–chromophore distance. In the bulk (large distances), where water does not interact with the protein, \( \theta \) is expected to be random, therefore averaging to 0°. At shorter distances, where the water interacts with the protein, its orientation is not random. At the H-bonding distance, the angle fluctuates around 40°. The deviation from a strict 0° expected from a bound state stems from a contribution of orientations in which the water is not H-bonded to the chromophore or, even if H-bonded, slightly deviates from the X-ray angle. As the distance increases, \( \theta \) increases through a local maximum, to a local minimum that again occurs at 5.25 Å (cf. panel A). At 300 K, the maximum is nearly 140°, representing a flip in dipole direction as the PWM transits to the next binding site. The PWM hop between sites is thus assisted by a flip-flop motion, in which the water molecule uses its hydrogen atoms as “hands” to perform a “backflip” of its oxygen atom, thus inverting its dipole direction.

3.3.3. Mean Squared Displacement (MSD) along the Exit Pathway. Water diffusion in the bulk shows a linear time dependence of the MSD. For three-dimensional diffusion with diffusion coefficient \( D \), the Einstein relation reads

\[
\text{MSD}(t) \equiv \langle (r(t + \tau) - r(t))^2 \rangle = 6Dt
\]

Here, \( \langle ... \rangle \) denotes ensemble average. This behavior, termed “normal” diffusion, is characterized by a constant value for MSD(\( t \))/\( t = 6D \).
Water diffusion slows down as it approaches a protein. In addition, it might become anomalous. Abnormal diffusion is often depicted as a power law,

$$\text{MSD}(t) = 6D t^\alpha$$

where $0 < \alpha < 1$ is called "subdiffusion" and $1 < \alpha < 2$ is "superdiffusion" ($\alpha = 1$ is, of course, normal diffusion). In crowded environments, one most often observes subdiffusion when the diffusion path is on the length-scale of the obstacles, which on longer length/time scales might become normal again ("transient subdiffusion").

When divided by $t$, one obtains,

$$\frac{\text{MSD}(t)}{t} = 6D t^{\alpha - 1}$$

Here, a negative slope (of the log–log plot) corresponds to subdiffusion, whereas positive slope is superdiffusion.

Figures 11 and 12 show water–chromophore distance profiles exhibiting sequential jumps with different velocities at different regions of space. Such "jumps" could lead to abnormal diffusion. Hence we have calculated the averaged MSD of water molecules that at some time $t_0$ (reset as the time origin) were bound to the chromophore. Figure 14 presents $\frac{\text{MSD}(t)}{t}$ for the leaving water molecule at the five temperatures investigated (points). As a reference, we have also calculated it for water molecules in the bulk phase of our simulations (dashed lines). The data are shown on a log–log scale, in which a power law becomes a straight line, revealing the $\alpha$ dependence more clearly. Diffusion of bulk water molecules is nearly normal (linear MSD/$t$), with a slight negative slope suggesting a minor subdiffusion effect that arises from the proximity of the protein. The exiting water behaves differently: it exhibits short time subdiffusion ($\alpha - 1 < 0$) that switches to superdiffusion at longer times ($\alpha - 1 > 0$).

At low $T$, this switch occurs around 100 ps, whereas at higher temperatures it occurs earlier, between 20 and 50 ps. At times longer than ca. 1000 ps, $\frac{\text{MSD}(t)}{t}$ approaches the

---

**Figure 11.** Black solid line: distance between the water oxygen and the chromophore OH group near a binding event, from a 320 K simulation (PDB file 2WUR). Red solid line with blue dots: location of the PWM within the water wire, if it exists. Blue dots mark times at which a wire exists. The water location along the wire is 1 when it is directly H-bonded to the chromophore. Insets: visualization of snapshots at various times showing the water molecules participating in the water wire at the indicated time. The PWM is colored blue.

**Figure 12.** (A) Distance between the PWM and Cro66-OH oxygen atoms for trajectory segments at 300, 320, and 340 K. (B) Top: the distance profile at 300 K (same as left figure in panel A), indicating the time window for each of the four states. Bottom: number of H-bonds the PWM forms with other water molecules (red) and with protein atoms (green).
(nearly constant) dashed curves calculated for bulk water. Thus, both sub- and superdiffusion are transient here. This transition from sub- to superdiffusion is independent of the NPT ensemble and/or the TIP3P water model, because qualitatively similar results were obtained in test runs with the NVE ensemble and/or the TIP4P/2005 water model (see Figure S11 in the SI).

The straight lines in Figure 14 are linear fits from which we have obtained the $\alpha$ values in the two regimes (see Table 1). With increasing temperature, both the sub- and superdiffusion $\alpha$ values tend toward $\alpha = 1$, implying that diffusion becomes less abnormal as protein fluctuations increase.

Reports on sub- to superdiffusion transitions are rather rare in the literature. There is a recent work showing such sub- to superdiffusion transition of protons on the surface of a DMPC membrane. Also, García and Hummer observed a transition between sub- and superdiffusion of protein modes on a time scale of 100 ps, similar to the transition time that we find here. This suggests that the motion of the exiting water molecule is enslaved to that of the protein, explaining why it becomes less abnormal with increasing temperature.

The square root of the MSD (RMSD) is indicative of the layer in which the water diffuses in the sub- and superdiffusion phases. Thus, at 300 K, the subdiffusion regime extends to 100 ps, where the RMSD $\approx 1.6$ Å (see Figure 14), corresponding to water rattling within its original binding site. Superdiffusion extends to about 1000 ps, whereas the RMSD of the fitted line $\approx 17$ Å, corresponding to water in the bulk. Thus, water transport across the protein walls is superdiffusive, and this could arise from flip-flop hops between interior and surface sites.

### Table 1. Fitted $\alpha$ Values in the Sub-/Superdiffusion Regimes

<table>
<thead>
<tr>
<th>$T$, K</th>
<th>subdiffusion</th>
<th>superdiffusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>0.379 ± 0.017</td>
<td>1.742 ± 0.006</td>
</tr>
<tr>
<td>310</td>
<td>0.425 ± 0.015</td>
<td>1.732 ± 0.003</td>
</tr>
<tr>
<td>320</td>
<td>0.426 ± 0.011</td>
<td>1.657 ± 0.003</td>
</tr>
<tr>
<td>330</td>
<td>0.586 ± 0.014</td>
<td>1.492 ± 0.003</td>
</tr>
<tr>
<td>340</td>
<td>0.600 ± 0.009</td>
<td>1.541 ± 0.006</td>
</tr>
</tbody>
</table>

A full list of fitted parameters is given in Tables S10 and S11 of the SI.

Data fitted to eq 5 using Origin 8.6. Values at each temperature were averaged over all trajectories.

3.3.4. The Kinetics of Water Exchange. The residence time (RT) of the water molecule in the binding site is indicative of the rate of water exchange. We define RT as the time interval between two consecutive events of different water molecules binding to the chromophore, and we calculate the average RT ($\tau_R$) at each temperature by averaging over the calculated RT of the individual water molecules.

From the same data, we calculate the residence correlation function (RCF) of the ensemble as follows: given that a water binds (for the first time) at $t = 0$, the RCF($t$) is defined as the fraction of water molecules with RT’s larger than $t$. Thus, $\text{RCF}(0) = 1$, decaying to zero as $t \to \infty$. 

Figure 13. (A) Difference of water–chromophore distance from previous frame ($\Delta t = 0.5$ ps), eq 2, as a function of the water–chromophore distance. Water molecules included in the calculation are those H-bonded to the chromophore at some time, which did not go further inside the protein, averaged (at each temperature) over all snapshots exhibiting the same distance. (B) $\theta$, angle between the water dipole vector at time $t$ and its initial value at the bound state of the minimized PDB structure (2WUR, 1EME, or 1EMB). Values for each temperature were averaged over all trajectories.

Figure 14. MSD/t as a function of $t$ for exiting water molecules on a log–log scale. The MSD was calculated for each water molecule with respect to its position when first bound to the chromophore. Data for each temperature were averaged over all trajectories at the indicated temperature (points). Full lines are linear fits to eq 5 utilized to obtain $\alpha$ in the two time-regimes. Dashed lines: MSD/t for water molecules in the bulk, averaged (at each temperature) over 1500 randomly selected water molecules that did not come closer than 5 Å to any protein atom for 500 ps. Fitted parameters for bulk water are given in Table S12 of the SI. Inset: MSD/t as a function of $t$ in a linear scale for the same set of water molecules.
Table 2. Values of Fitted Parameters for RCF(τ) with Stretched and Biexponential Functions\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>T, K</th>
<th>(\tau_{av}), ps</th>
<th>(\tau_{av}^p), ps</th>
<th>(\beta^p)</th>
<th>(\tau_{av}^s), ps</th>
<th>(\tau_{av}^p), ps</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>3778 ± 3833</td>
<td>3672 ± 0.82</td>
<td>823</td>
<td>5159</td>
<td></td>
</tr>
<tr>
<td>310</td>
<td>1930 ± 2909</td>
<td>814 ± 0.51</td>
<td>662</td>
<td>5445</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>181 ± 215</td>
<td>157 ± 0.91</td>
<td>139</td>
<td>464</td>
<td></td>
</tr>
<tr>
<td>330</td>
<td>139 ± 198</td>
<td>102 ± 0.76</td>
<td>84</td>
<td>299</td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>151 ± 239</td>
<td>77 ± 0.56</td>
<td>49</td>
<td>331</td>
<td></td>
</tr>
</tbody>
</table>

\(\text{av}\) = average at temperature \(T\), \(\text{av, p}\) = average at \(T\) with possible escape to solution, \(\text{av, s}\) = average at \(T\) at 300 K.

\(\text{RCF}(\tau) = A \exp(-t/\tau)\)

(6)

In this model, \(\tau\) is the “stretched exponential” residence time, and \(0 < \beta < 1\) depicts the deviation from single-exponential behavior. The latter is recovered, of course, when \(\beta = 1\).

Alternately, the RCF can be represented by a biexponential function:

\(\text{RCF}(\tau) = A \exp(-t/\tau_A) + B \exp(-t/\tau_B)\)

(7)

Where \(A\) and \(B\) represent short and long RT’s, respectively, and \(A + B = 1\).

Figure 15 shows RCF(τ) at the five temperatures, fitted to both stretched and biexponential functions, with parameters collected in Tables S13 and S14 of the SI, respectively, and summarized in Table 2. R values are close to 1, indicating that both functions fit the data equally well. The fact that both \(A\) and \(B\) differ from 0 or 1 indicates that the transient decay here is not single-exponential. While \(\tau_{av}\) from the stretched exponential function is closest to the average RT, \(\tau_{av}^p\) the biexponential times, \(\tau_A\) and \(\tau_B\) are more readily interpreted in terms of a simple kinetic model. Assume that water migration within the protein can be described as hops between the three above-mentioned sites:

bound \(\rightarrow\) cavity \(\rightarrow\) surface

(8)

Water migration between the chromophore and the intermediate layer is reversible, but once on the surface, its probability of re-entry is negligible. As a result, the fourth step, proceeding from the surface to the bulk, is not reflected in the binding probability (e.g., as a third exponential), and the kinetic equations yield biexponential decay. This supports the three-layer migration scenario identified in Figure 8.

Interestingly, the figure shows a large gap between the low (300, 310 K) and high (320, 330, and 340 K) temperature data. This is in accord with the dramatic increase in the strand 7–10 gap above 310 K, as indicated in Figures 3 and 4A. Recall also that the 300 K MDS showed a transition from sub- to superdiffusion at 100 ps (Figure 14), interpreted as the time the water molecule first leaves the chromophore. In agreement with this interpretation, the RCF is nearly 1 up to 100 ps, so that during the subdiffusion regime the PWM does not yet leave its binding site.

4. CONCLUSION

Internal water molecules within enclosed protein cavities are traditionally perceived as “crystallographic” molecules that are an inherent part of the protein.\textsuperscript{33,46} Hence it might seem unexpected if they exchange readily with bulk water (on the single- or sub-nanosecond time scale). Here we focused on the PWM within GFP, so termed because it has the privilege of conducting the proton from the photoexcited chromophore, leading to the green fluorescent signal utilized in numerous bioengineering applications. The proton is believed to hop across the PWM and stick to the Glu222 residue for the duration of the chromophore’s excited-state lifetime.\textsuperscript{30}

Nevertheless, time-resolved fluorescence data from GFP suggested that this proton might exhibit much richer dynamics, which includes drunkard motion along a wire with possible escape to solution.\textsuperscript{33,46} The present MD investigation shows that the PWM exhibits equally rich dynamics.

Our study is divided into two parts: the hole in the barrel and what flows through it. The “hole” is a weak spot between strands 7 and 10 of the GFP \(\beta\)-barrel, which exhibits the highest backbone RMSD of all nonloop segments. The temperature dependence of the interstrand gap (Figure 3) shows a “barrel-cracking” transition that is focused on residue pairs 146–205 and 148–203, whose separation increases most dramatically at 320 K and above. Although there is a weak spot also between strands 7 and 8, we find that it is rarely utilized for PWM exchange. We have mapped out all protein atoms encountered along the pathway, and the top 90 of them all belong to residues on the 7 and 10 strands. These divide into three “layers”: the bound state, an intermediate state, and the protein surface. Concomitantly, widening of the
7–10 gap leads to a rapid succession of water exchange events. This occurs readily for $T \geq 320$ K. For the anionic form of the chromophore (not investigated herein), water exchange events are much more probable (Table 1 of ref 24).

At $320$ K, the RMSD of the "long loop" connecting strands 6 and 7 increase substantially. Possibly, it acts as a "sail in the wind", collecting thermal energy from solution and channeling it to strand 7. Indeed, residues on strand 7 close to this loop also increase their RMSD substantially at 320 K, eventually producing a hole in the hull through which water leaks.

It is not just a single water molecule passing through the hole. In nearly half of the cases, a water wire is threaded through the hole, on which the PWM advances by exchanging places with neighboring water molecules sequentially, one water–water switching event at a time.

This water wire could allow proton leak-out from the barrel, because water wires are exquisite proton conductors. This might be the dominant mechanism of proton escape from GFP. Indeed, from the temperature dependence we find an enthalpy of water-wire formation in the range 23–26 kJ mol$^{-1}$, nearly identical to the activation enthalpy for the "conformational switch", which allegedly leads to proton escape, identified in the time-resolved fluorescence data.46

It was argued that proton escape may be triggered by Thr203 rotation that completes a hydrogen-bond pathway from Cro66-OH to the backbone CO of His 148 (Figure 5 of ref 44). In the alternate scenario, in which the proton escapes through the above water wire, Thr203 rotation gates water-wire formation that is otherwise impeded by its hydrophobic methyl group.

In the second part of this study we followed the tortuous path of the leaving water molecule. The MSD along this path behaves quite remarkably. Starting from the bound PWM, its MSD shows subdiffusive behavior in the initial 100 ps, switching to superdiffusion as it traverses the barrel wall and emerges on the GFP surface. Once released to the bulk, after about 1 ns (depending on $T$), the MSD reverts to the customary normal diffusion behavior.

While subdiffusion has been observed quite frequently in congested environments, a superdiffusion phase is much less expected. It could result from large-amplitude but infrequent hops between protein cavities (i.e., the binding, internal and surface sites). After a long waiting time comes a rapid intersite hop, which tends to occur via a "flip-flop" mechanism. The PWM lets go of two (out of 4) H-bonds, performing 180° rotation, which thus couples to its translation into the adjacent site.

The kinetics of the RT in the binding site was also monitored. It is clearly nonexponential, but can be equally well described as stretched- or biexponential. The two-exponent decay of the binding probability was tentatively ascribed to the three-site kinetics in eq 8, in which the internal migration of the water molecule is reversible, but once on the surface its probability of re-entry becomes negligible.

Considerably greater computational effort is required to also monitor directly, in real time, the proton dynamics in and through the GFP barrel. However, to an extent, the rapid proton motion is "enslaved" to water dynamics that prepares the environment for its migration. In this context, it is amusing to consider possible implications of the water-wire formation connecting the chromophore with the bulk at 320 K and above. In warm water, GFP might become a "proton gun", a prediction that could be accessible experimentally. This may complement the ability of certain GFP mutants, e.g., "killer red", to act as "electron guns".31 One wonders whether such a functionality, if exists, was important for the survival of marine life in the warm oceans of eras gone by.

ASSOCIATED CONTENT

Supporting Information

Supporting tables (S1–S5, force field parameters for GFP chromophore; S6, protonation states of histidine; S7, RMSD of individual residues; S8, list of "internal" H-bonding atoms; S9, sequence indices for entrance/exit paths; S10, S11, parameters for sub/superdiffusion fits to water MSD; S12, fitting parameters to bulk water MSD; S13, S14, parameters for stretched and biexponential fitted RCF(1)) and supporting figures (S1, GFP chromophore scheme; S2, histidine in different protonation states; S3, parameters for assessing MD equilibration times; S4, RMSD of individual residues; S5, sequence indices distribution; S6, rotamers of T203; S7 and S8, T203 dihedral angle distributions; S9, water-chromophore distance profiles; S10, flip-flop mechanism for different MD methodologies; S11, sub/superdiffusion for different MD methodologies). Supporting movie file S1, one PWM exit trajectory showing coupled translation and reorientation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: agmon@fh.huji.ac.il.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by The Israel Science Foundation (Grant Number 766/12). A.S. acknowledges support from the Dalia and Dan Maydan Fellowship Endowment Fund. The Fritz Haber Research Center is supported by the Minerva Gesellschaft für die Forschung, München, FRG.

REFERENCES


■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on February 13, 2015. Additional changes were made throughout the main text. The revised paper was re-posted on February 18, 2015.