# Time-Averaged Distributions of Solute and Solvent Motions: Exploring Proton Wires of GFP and PfM2DH

Camilo Velez-Vega,* Daniel J. J. McKay,* Vibhas Aravamuthan, Robert Pearlstein, and José S. Duca

Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 100 Technology Square, Cambridge, Massachusetts 02139, United States

S *Supporting Information*

ABSTRACT: Proton translocation pathways of selected variants of the green fluorescent protein (GFP) and *Pseudomonas fluorescens* mannitol 2-dehydrogenase (PfM2DH) were investigated via an explicit solvent molecular dynamics-based analysis protocol that allows for direct quantitative relationship between a crystal structure and its time-averaged solute−solvent structure obtained from simulation. Our study of GFP is in good agreement with previous research suggesting that the proton released from the chromophore upon photoexcitation can diffuse through an extended internal hydrogen bonding network that allows for the proton to exit to bulk or be recaptured by the anionic chromophore. Conversely for PfM2DH, we identified the most probable ionization states of key residues along the proton escape channel from the catalytic site to bulk solvent, wherein the solute and high-density solvent crystal structures of binary and ternary complexes were properly reproduced. Furthermore, we proposed a plausible mechanism for this proton translocation process that is consistent with the state-dependent structural shifts observed in our analysis. The time-averaged structures generated from our analyses facilitate validation of MD simulation results and provide a comprehensive profile of the dynamic all-occupancy solvation network within and around a flexible solute, from which detailed hydrogen-bonding networks can be inferred. In this way, potential drawbacks arising from the elucidation of these networks by examination of static crystal structures or via alternate rigid-protein solvation analysis procedures can be overcome. Complementary studies aimed at the effective use of our methodology for alternate implementations (e.g., ligand design) are currently underway.
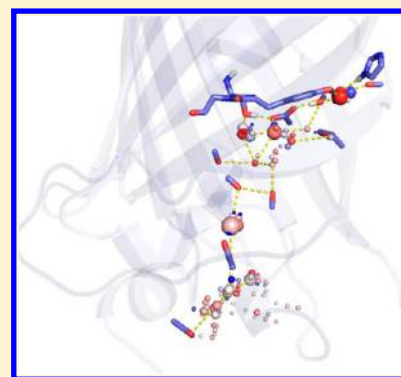
INTRODUCTION

Preclinical drug discovery remains a highly challenging endeavor, despite many advancements in biology, chemistry, and technology over the last several years. This is due in part to the ever broadening scope and complexity of disease indications of interest, and greater expectations for better and safer molecules, as well as a presently incomplete understanding of binding interactions. Thus, a stronger theory of binding and its reduction to practice in the form of in silico tools is necessary to improve the accuracy and efficiency of drug design. In particular, we and others have postulated that these developments can be greatly aided by capitalizing on the hypothesis that structure−function relationships of aqueous solutes are governed largely by the energetic cost of transferring water molecules to and from the solvation layer and bulk water.[1−3] As a valuable step toward substantiating this concept, and in general, promoting a more efficacious drug design process, our group and others within the pharmaceutical industry have, in the past few years, given greater precedence to structure-based explicit solvent dynamics analyses that can offer a more realistic picture of interatomic interactions. Although this has been a customary practice in academic laboratories for quite some time, only until recently has it been fully embraced within an industrial drug discovery setting, mostly due to recent breakthroughs resulting in a substantial reduction of the

computational time and cost required for such analyses. Our efforts in this direction are currently focused on the use of time-averaged solute and solvent structures to further our fundamental understanding of structure-free energy relationships in aqueous systems, and to guide de novo ligand design.

Time-averaged structures have the notable advantage of being amenable to straightforward validation through direct superposition onto corresponding high-resolution crystal structures (where available). This is not the case, in our experience, with more common MD postprocessing techniques, given that adequate validation of the simulation results may be limited by the reduction of information due to clustering by means of low-dimensional order parameters. The latter drawback naturally deters the robustness of any prediction. Alternatively, an extended number of structural markers and/or order parameters can be assessed in light of available experimental/crystallographic data, resulting in a more reliable, yet more laborious postprocessing protocol. However, these types of investigations currently suffer from impractical computational requirements.

Our postprocessing protocol computes the time-averaged distribution of atomic solvation (TADAS) and time-averaged

solute structure (TASS) and outputs a single, ensemble-averaged snapshot depicting the structural dynamics around a particular state of interest. Specifically, the TADAS is calculated by means of a frame-by-frame high-resolution lattice-based water analysis, and the TASS, by alignment and subsequent frame-by-frame averaging of solute atom positions. The TADAS has accurately identified high-density water sites for a number of systems that we have examined (based on comparison of predicted and crystallographically observed positions). Arguably of greater import, the TADAS also provides detailed structural information on low-density solvation, which is typically unobserved in crystal structures and may not be properly described using alternate solvation dynamics analysis procedures.[4−9] In general, such related methods[4−9] are currently limited to exploring the solvation dynamics of fixed solute conformations, and most of them at present allow only the study of high-density solvation.[5−9] In contrast, the all-density solvation statistics underpinning the TADAS are extracted from essentially unrestrained simulations (only protein anchoring restraints are used to facilitate frame-by-frame structural alignment; see the Methods section) whose solute dynamics is represented by the TASS. Note that we are currently unaware of alternate methodologies that carry out structural analysis by direct scrutiny of time-averaged solute structures.

The proposed approach has been implemented in an in-house software package called Water Analysis Tools for Molecular Dynamics (WATMD) which, in addition to TADAS and TASS, computes solvent dipoles and charge distributions, time-averaged solvent accessible surfaces (TASAS) and time-averaged solvent dipole distributions (TASDD). WATMD's visual output can be easily interrogated for differences in the solute−solvent structures from simulations of dissimilar states (e.g., unbound versus bound complexes) or involving different ligands. Thus, far, we have successfully used the different capabilities of WATMD to accurately model several solvated protein systems, as well as to aid the design of potent ligand molecules (to be reported in future publications). The latter includes identification of binding pockets, replacement/displacement of favorable/unfavorable solvation, solvation-based pharmacophore modeling and solvation-based de novo design of unconventional scaffolds. Other prospective applications of our methodology comprise the refinement of crystal structures and model-based qualitative examination of processes for which incomplete or low resolution crystal structures (for limitations, see the Methods section and Appendix 1), or even less informative experimental data (e.g, circular dichroism spectra), are available. More broadly, this protocol can be used to improve understanding of the dynamic solute−solvent structure−function relationship of soluble and membrane-bound proteins at the atomistic level, knowledge that can greatly benefit projects in many areas, including biochemistry, medicinal chemistry, and pharmacology.

Here, we describe our methodology, and validate its predictive capabilities vis-à-vis the well-studied problem of proton translocation within hydrogen bonding (H-bonding) networks spanning between active sites and bulk solvent. In general, proton transfer plays a major role in many biochemical processes, acting as a modulator of protein function and enzyme kinetics at various steps along a catalytic pathway. Thus, elucidation of proton-diffusion mechanisms is not only critical for proper understanding of such processes but also

offers a unique and attractive direction to rational drug design aimed at enhancing or disrupting these processes. We chose proton translocation for this first report of our approach because of its reported high sensitivity to subtle structural variations arising from point mutations or changes in protonation states, wherein minor discrepancies in proton donor−acceptor distances between participating residues/waters, or a low probability of occupancy of these states, can utterly disrupt these processes. Therefore, proton transfer serves as a particularly demanding test of the degree to which our time-averaged analysis approach is able to reproduce such slight, yet highly relevant, structural changes. Despite the latter, it is evident that a distance and occupancy-based metric of proton donor−acceptor H-bonding simply suggests certain likelihood of proton transfer, since proton tunneling cannot be accounted for via the level of the calculations presented here. Our approach thus falls short of estimating actual proton translocation rates for the systems studied, something that is out of the scope of this work and requires stepwise sampling of multiple states along the proton transfer process using QM or QM/MM calculations.

Studies on proton transfer to/from active sites of proteins in solution[10−22] and across membrane-bound proteins or channels[23−32] have been previously published by other researchers. Most of these efforts have centered on the study of carbonic anhydrases, cytochromes, green fluorescent protein, bacteriorhodopsin, and transmembrane proton channels. Overall, proton movement is believed to take place through direct shuttling of protons between amino acid side chains, via proton hopping along structured water wires, or by means of a combination of both mechanisms.[26] Evidence of proton shuttling between side chains can be revealed experimentally by analysis of conformational changes (as resolved by crystallography or NMR) and pK_a variations of key side chains along particular transfer pathways. Conversely, proton transfer along water wires requires the identification of stable water configurations that can, in principle, be observed in high-resolution crystal structures. In the absence of appropriate crystal structures, computational methods can be used to determine solvation structures and identify proton wires.

Computational studies aimed at characterizing proton wires have in general employed methods such as conventional molecular dynamics (MD),[13,14,22,25] Q-Hop MD,[23] discretized Feynman path integral MD,[24] empirical valence bond (EVB),[15,20,21,27] QM,[33−36] and QM/MM,[10,11,37,38] among others.[16,17] These studies have greatly contributed to the biophysical description of proton transfer networks and provided further evidence concerning their involvement in protein function. Furthermore, these studies have reinforced the importance of accounting for structural dynamics when addressing challenging problems of this nature.

In this paper, we make use of our time-averaged solute−solvent analysis methodology to study the following: (1) wild type (WT) green fluorescent protein from *Aequorea victoria* (GFP) and two of its mutants, for which high-resolution crystal structures are available; GFP is a landmark protein for the study of proton wires, given the direct link between biological and biomedical imaging and proton transfer from/to its chromophore upon photoexcitation;[39] the study of this protein provides validation of our MD time-averaged modeling approach; (2) binary and ternary complexes of *Pseudomonas fluorescens* mannitol 2-dehydrogenase,[40] as a more prognostic
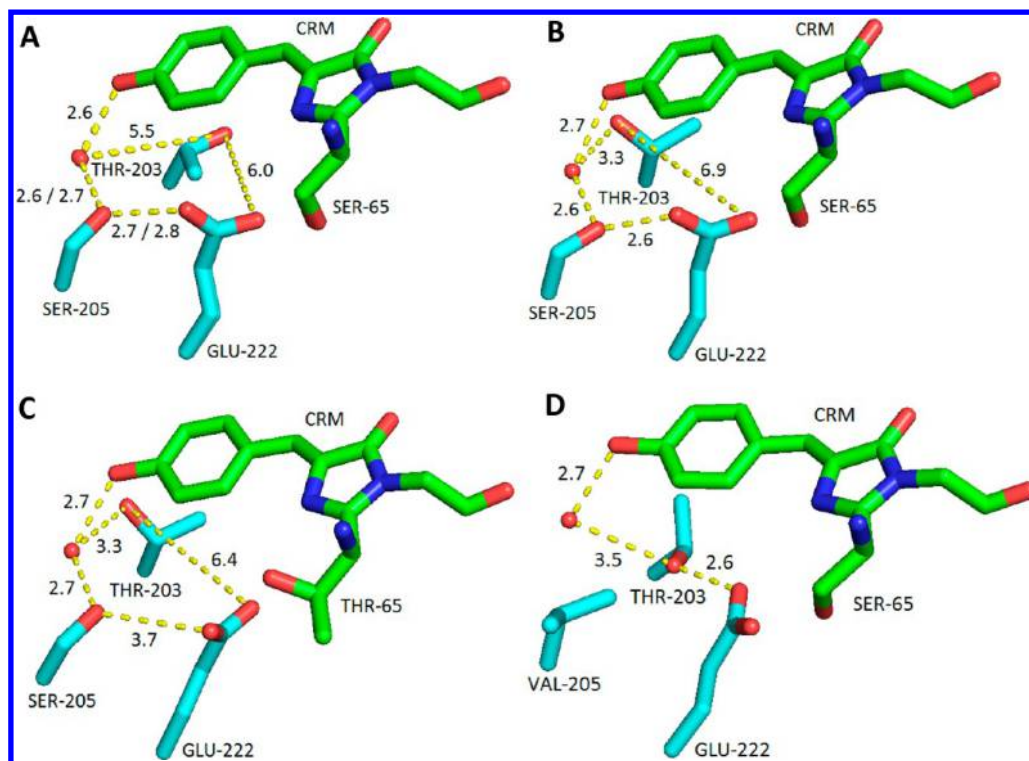
Figure 1. Relative positions of the chromophore (CRM), a key structural water, and side chains T203, S/V205, and E222, for GFP structures (A) 1gfl (chains A and B are superposed), (B) 2wur, (C) 1ema, and (D) 2qle. Nitrogens are colored in blue, oxygens in red, CRM carbons in green, and side chain carbons in cyan. The red sphere corresponds to the oxygen of the structural water, and the dashed lines represent the distances (in black) between the different groups. Distances for all figures are in angstroms (Å).

investigation allowing the postulation of a detailed mechanism for proton transfer upon mannitol binding and oxidation.

BIOLOGICAL BACKGROUND

Green Fluorescent Protein (GFP). GFP absorbs at approximately 395 nm in the protonated (A state) of its chromophore and at approximately 475 nm in the deprotonated (B state).[39] These states are believed to coexist in WT GFP at a ratio of 6:1 (A:B).[41] Chattoraj et al.[42] reported that upon photoexcitation (denoted by *) at room temperature, B* emits instantaneously at 508 nm with no visible decay during the length of the experiment. Conversely, the characteristic emission of A* (at 460 nm) rapidly decays upon photoexcitation at room temperature, followed by an increase in emission at 508 nm. These results suggested that A* transforms rapidly into a deprotonated intermediate species I* (see Figure 4 in ref 39), which is responsible for the observed fluorescence at 508 nm upon absorption at 398 nm.[42]

The mechanism of proton transfer leading to the formation of an intermediate state upon photoexcitation of A has been the focus of numerous research efforts.[39,43−47] Overall, a widely accepted hypothesis suggests the occurrence of a concerted proton transfer from the chromophore's phenol to E222 through a structural water and the hydroxyl of S205. The viability of this mechanism is supported by the position of the groups involved in the wire, as observed in crystal structures representative of state A (PDB codes 1gfl and 2wur). Figure 1A (1gfl-A and 1gfl-B) and B (2wur) show that the chromophore's hydroxyl, the structural water, and the side chains of S205 and E222 are appropriately positioned to allow proton transfer, noting that the mature chromophore is formed by residues 65−67 (colored green in Figure 1).

Numerous GFP mutants with varying spectra have also been reported,[48−52] but only a few high-resolution crystal structures of these variants are currently available. Of particular interest for this work are the single point mutations S205V and S65T (believed to be representative of state B), which lead to a 30-fold reduction in the proton transfer rate from the chromophore[53] and complete disruption,[41] respectively. The structural basis for the altered proton transfer capability of these two mutants becomes apparent upon analysis of their corresponding crystal structures (PDB codes 1ema and 2qle for S65T and S205V, respectively). Figure 1C (1ema) and D (2qle, chain A) show a rearrangement of the groups involved in the proton wire of the WT structures (Figure 1A and B). Specifically, the E222 side chain in S65T twists away from S205 (oxygen−oxygen [O−O] distance 3.7 Å), precluding proton transfer between these residues. Conversely, the absence of a polar side chain in S205V at position 205 results in the rearrangement of T203, which now acts as a bridging residue for proton transfer between the structural water and E222. Note that the T203 → structural water → E222 proton wire is unlikely to exist in WT and S65T, based on the long O−O distances observed in the crystal structures (see Figure 1A−C).

Capitalizing on the high resolution of the aforementioned crystal structures (1.9 Å for 1gfl and 1ema, 0.9 Å for 2wur, and 1.6 for 2qle), as well as on the ample amount of literature (including various computational studies[17,54−60]) on the structure of GFP and its proton wires, we studied the solute and solvent structural dynamics of the WT, S65T, and S205V forms, as a necessary step toward validation of our time-averaged analysis protocol. Our observations are initially centered on the study of proton translocation from the chromophore to E222, followed by a broader discussion of

proton transfer from the chromophore to solvent for the WT protein.

***Pseudomonas fluorescens* Mannitol 2-Dehydrogenase (PfM2DH).** PfM2DH belongs to the long-chain mannitol dehydrogenase (MDH) family of enzymes, which are responsible for the reversible oxidation of D-mannitol to D-fructose, D-sorbitol to L-sorbose, and D-arabinitol to D-xylulose. The catalytic process encompasses the transfer of a hydride from the alcohol substrate to nicotinamide adenine dinucleotide (NAD$^+$) without the involvement of metal cations, followed by the translocation of a proton from the catalytic site to bulk solvent.[61] Previous work suggested that proton transfer might be the rate-limiting step of the enzymatic catalysis in alcohol dehydrogenases.[62,63] Thus, a detailed understanding of the proton translocation networks in these proteins may be highly advantageous for further optimization of this reaction.

High-resolution crystal structures of the binary (enzyme-NAD$^+$; PDB code 1lj8, 1.7 Å resolution) and tertiary (enzyme-NAD$^+$-mannitol; PDB code 1m2w, 1.8 Å resolution) complexes of PfM2DH were previously resolved by Kavanagh et al.[40] Therein, the authors proposed two proton transfer mechanisms from/to bulk to/from the catalytic site: In the first mechanism, E292 serves directly as the proton shuttle, wherein the side chain of this residue swings out of the active site toward bulk solvent, carrying the excess proton along. In the second mechanism, E292 plays an indirect role as a gate that opens and closes a water channel through which the proton moves out of the active site. Both proton transfer mechanisms were further studied by Klimacek et al.[64] In particular, the authors exploited their finding that E292 and K295 appear to exhibit a significant change in pK_a upon formation of the tertiary structure, to perform an MD study of the structural variations resulting from changes in the ionization state of these two residues. The results of these simulations, analyzed in the context of kinetic and mutagenesis information, disfavor the first mechanism, i.e., the proton appears to be transported to bulk along a water wire that is gated by E292. Moreover, by means of transient proton release/uptake experiments,[63,64] Klimacek et al. also found that only one proton is presumably released during the presteady state phase of the reaction, postulating that K295 is protonated and E292 is deprotonated in the binary complex. Yet, for these experiments, the enzyme equivalents turned over in the presteady state were considerably lower than unity, a complication that increases the uncertainty in these measurements (possibly suggestive of a low concentration of active protein in solution that could result in reduced proton release relative to the actual stoichiometry). Moreover, the requirement that K295 and E292 are respectively protonated and deprotonated in the binary structure necessarily leads to the assignment of a suboptimal H-bonding network (as observed in Figure S1A and in Figure S1A of ref 64 for crystal structure 1lj8) that contrasts with the more consistent H-bonding allocation resulting when both K295 and E292 are considered protonated in the binary crystal (see Figure S1B and Figure S1B of ref 64).

The limited confidence in the ionization states of E292 and K295 provides an excellent opportunity to test whether our approach is able to identify a set of protonation states whose TASS and TADAS structures reproduce the conformations observed in the high-resolution crystal structures of each complex. For this purpose, we modeled the binary enzyme—NAD$^+$ and tertiary enzyme—NAD$^+$—mannitol complex struc-

tures at different protonation states of E292 and K295 and compared our results against the reported findings.[40,63,64] We also performed and analyzed simulations on the enzyme—NADH—fructose complex in an effort to better understand the proton transfer mechanism involved in the oxidation of mannitol via PfM2DH.

■ METHODS

For each of the cases studied, the corresponding PDB files were prepared by building missing atoms using Leap[65] and assigning rotameric and protonation states leading to a visually consistent H-bonding network. [We raise awareness on the importance of a correct assignment of all protonation states, the assessment of which may require several preliminary runs. For instance, the H-bonding network that can be construed from the 1gfl crystal (which is devoid of hydrogens) is ambiguous with respect to the most probable protonation state of H169. It was therefore imperative to model the structures using both ε and δ histidine. Remarkably, the δ-H169 simulations are in excellent agreement with the corresponding crystal structures (as discussed in the Results and Discussion section), whereas substantial deviations from the experimental conformations were observed in the ε-H169 simulations.] After removing crystal waters existing beyond ~1 Å from the protein surface, we solvated the system in a box of TIP3P waters while complying with the minimum image convention. The overall charge of the system was then set to zero by neutralizing selected charged amino acid side chains that are fully solvent-exposed in the crystal structure and located far from the binding site or from an H-bonding network potentially involved in proton transfer. Care was taken to perturb the centroid of the overall charge distribution as little as possible relative to that of the original system, in order to avoid altering the dipole fields. We have employed this neutralization method in numerous *in-house* MD studies (as support for medicinal chemistry projects), finding that it introduces no artifacts to the simulation results. Furthermore, it offers an advantage over the conventional practice of adding counterions to the system, given that this can drastically perturb the water structure and lead to a significantly longer equilibration time. We verified the suitability of this practice for the systems studied in this work by carrying out two test runs without side chain neutralization (1gfl-A for GFP and 1lj8 with protonated K295 and E292 for PfM2DH) and superposing the resulting time-averaged solute—solvent structure onto those obtained from multiple replicate runs using our neutralization approach. The cases without side chain neutralization led only to structural differences arising from statistical uncertainty, akin to those observed in Figures S2 and S4.

Conformationally and topologically averaged AM1-BCC[66,67] charges were used for the small molecules and the GFP chromophore. Namely, we first generated a crude set of AM1-BCC charges and employed the Omega[68] software package to produce a large ensemble of conformers for each molecule. This was followed by an electrostatic energy calculation on each conformer, carried out using the absolute values of the partial charges. AM1-BCC calculations were then performed for each of the ten conformers with the lowest electrostatic energy, and the resulting charge sets were averaged to obtain the final charges for each molecule. In preliminary runs (not shown) for this work using the 1gfl crystal, improved accuracy was achieved with this charge set over RESP[69] in reproducing the crystal conformation. More generally, our group and others[70] have

found that this charge assignment procedure is the most reliable one currently available; yet, this topic remains controversial.

Finally, we generated the respective topology and coordinate files via Leap using the Amber99sb[71] force field with the parm@Frosst[72] small molecule extension. Parm@Frosst is the result of an 18 y effort to improve the performance of MD simulations with Amber9x or ffxx parameter files, in the presence of small molecules of pharmaceutical interest. Its use in numerous drug discovery projects suggests that it might be a better alternative to the use of GAFF.[73] The reader is encouraged to read ref 72 for further information on parm@Frosst.

Our MD simulations were carried out within the *NPT* ensemble via the Amber 11[74] software package, using the following procedure: We initially minimized the waters and annealed them by heating the system from 0 to 40 K and then cooling it back to 0 K, over a period of 8 ps, while restraining the remaining atoms with a stiff force constant of 20 kcal/mol·$Å^2$. The system was subsequently warmed to 300 K over a period of 20 ps, using a time-step of 1 fs, and restraining all C_α atoms with a force constant of 5.0 kcal/mol·$Å^2$. This was followed by a 500 ps equilibration run at 300 K, in which all C_αs were restrained with a 2.0 kcal/mol·$Å^2$ force constant. Finally, we carried out 10 ns production runs with a 2 fs time-step, saving frames every 0.25 ps. During the production phase, we used a soft 2.0 kcal/mol·$Å^2$ restraint on designated residue C_αs (15 atoms for WT GFP and its mutants, and 27 atoms for PfM2DH) of two secondary motifs that are neither solvent exposed nor near the binding site. The latter are mainly anchoring constraints that have the sole purpose of keeping the overall system from drifting across the simulation box, allowing proper alignment of the frames prior to the WATMD analysis.

The Ptraj[75] package, which is distributed as part of AmberTools[74] (see manual for more information), was used to process each production MD trajectory. Specifically, we obtained an ensemble-averaged solute structure (i.e., a TASS) and used it as reference to align all trajectory frames. We then stripped the nonwater atoms from all frames to obtain the set of aligned solvent structures. Next, we produced a grid-averaged structure from this solvent-only set to generate the time-averaged distribution of atomic solvation (TADAS). Briefly, the TADAS was obtained by overlaying two grids onto the simulation box, one for water oxygen atoms and one for water hydrogen atoms, each subdivided into 1 $Å^3$ voxels. Each solvent atom position from the MD production run is then assigned to a corresponding voxel, and the average occupation is calculated for each voxel over the total number of frames. The TADAS is finally merged with the TASS to produce a time-averaged structure of the entire system.

We conducted 10 replicate MD simulations for each of the main cases and 5 for each of the supporting cases, all of which produced distinct trajectories due to the use of a random seed for initial velocity assignments. In this way, broader sampling of phase space was promoted in searching for alternate stable states. Each trajectory was processed separately to obtain a solute–solvent time-averaged structure. All time-averaged structures corresponding to the same case were then visually clustered based on apparent differences in the solute and/or solvent structures around the regions of interest. For each case, a time-averaged structure that is representative of the dominant state (most frequently occurring) among all replicates was chosen for further analysis.

We remark that our procedure can be used to study crystal or NMR structures of any resolution, or homology models. However, the confidence on the results decreases with the quality of the structure/model. The outcome is particularly affected when a key segment of the protein (e.g., one close to the site of interest) is missing and there is no reliable template that can be used to build a model for such region. In general, MD simulations can in principle correct minor inconsistencies in the initial model; still, if a structure has major imprecisions, these will likely not be corrected during the time scales of our simulations, thus requiring longer runs and/or more elaborate sampling techniques.

All figures were produced using the Pymol[76] software. The visualization of solute–solvent time-averaged structures is nonstandard. Therefore, we recommend reading Appendix 1 of the Supporting Information, which provides information that is important to understand the visual output presented in this work, as well as other relevant details of our methodology.

■ RESULTS AND DISCUSSION

GFP. Wildtype crystals 1gfl-A, 1gfl-B and 2wur have congruent solute structures with all-atom RMSDs of 0.13 Å between 1gfl-A and 1gfl-B and 1.30 Å between 1gfl-A and 2wur. Nonetheless, there are differences in the number/location of buried crystal waters, which can be crucial for proper elucidation of proton transfer networks. Figure 2 shows good
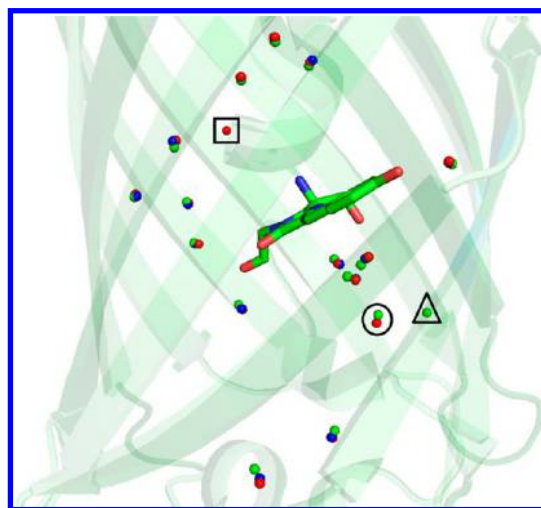
Figure 2. Buried waters resolved in crystal structures 1gfl-A (red), 1gfl-B (blue), and 2wur (green). For 2wur, two positions of W2108 and W2307 are reported, but only the B coordinates are shown/modeled here. The β-barrel and the chromophore are also depicted for reference.

overall agreement between the buried waters of the three crystals, except that some are only present in 1gfl-A, 2wur, or 1gfl-A and 2wur (Figure 2, enclosed within a black square, triangle, and circle, respectively). These discrepancies strongly suggest the need to model all three WT structures in order to assess the stability of the unmatched solvation sites and, more importantly, to improve the reliability of our analyses. We also carried out simulations of crystal structures 2qle and 1ema, corresponding to mutant forms S205V and S65T. We remark that 1ema is believed to correspond to state B[39] (see the Biological Background section), so for this case the
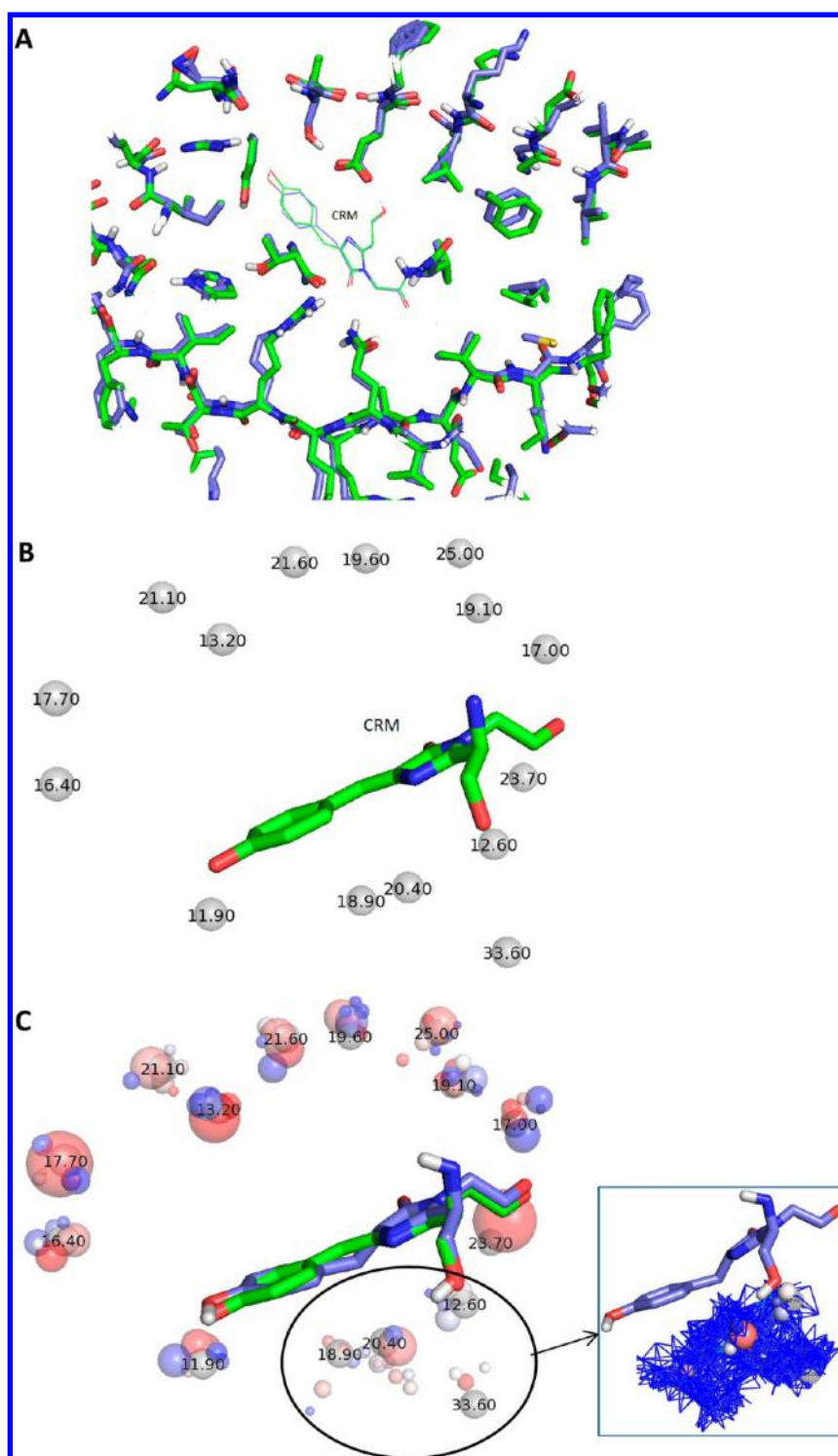
Figure 3. Preliminary validation of simulations of WT 1gfl-A crystal structure. (A) Overlay of the crystal (green carbons) and representative time-averaged solute (slate carbons) structures around CRM (line representation). (B) High density crystal waters (gray spheres; labeled by β-factor) around CRM of the crystal structure. (C) Superposition of the representative time-averaged structure of CRM and the high-occupancy solvation around CRM, onto part B. The sphere volume and color (red-white-blue translucent spectrum) of the high-occupancy solvation correlate respectively with the occupancy and the charge of each hydration site. The inset of part C shows the all-occupancy solvation surface within the region delimited by the black oval, colored by occupancy; the crystal waters and sites of high-occupancy solvation (red-white-blue spheres) are also shown, for reference.

chromophore's phenol and E222 side chain were modeled as deprotonated and protonated, respectively.

Ten replicas were run for each case, all of which resulted in time-averaged solute−solvent structures with low all-atom

RMSDs from the corresponding crystal: 1.38 ± 0.13 Å for 1gfl-A, 1.38 ± 0.15 for 1gfl-B, 1.61 ± 0.22 Å for 2wur, 1.28 ± 0.07 Å for 1ema, and 1.60 ± 0.31 for 2qle. Accordingly, only moderate conformational diversity was observed in the region
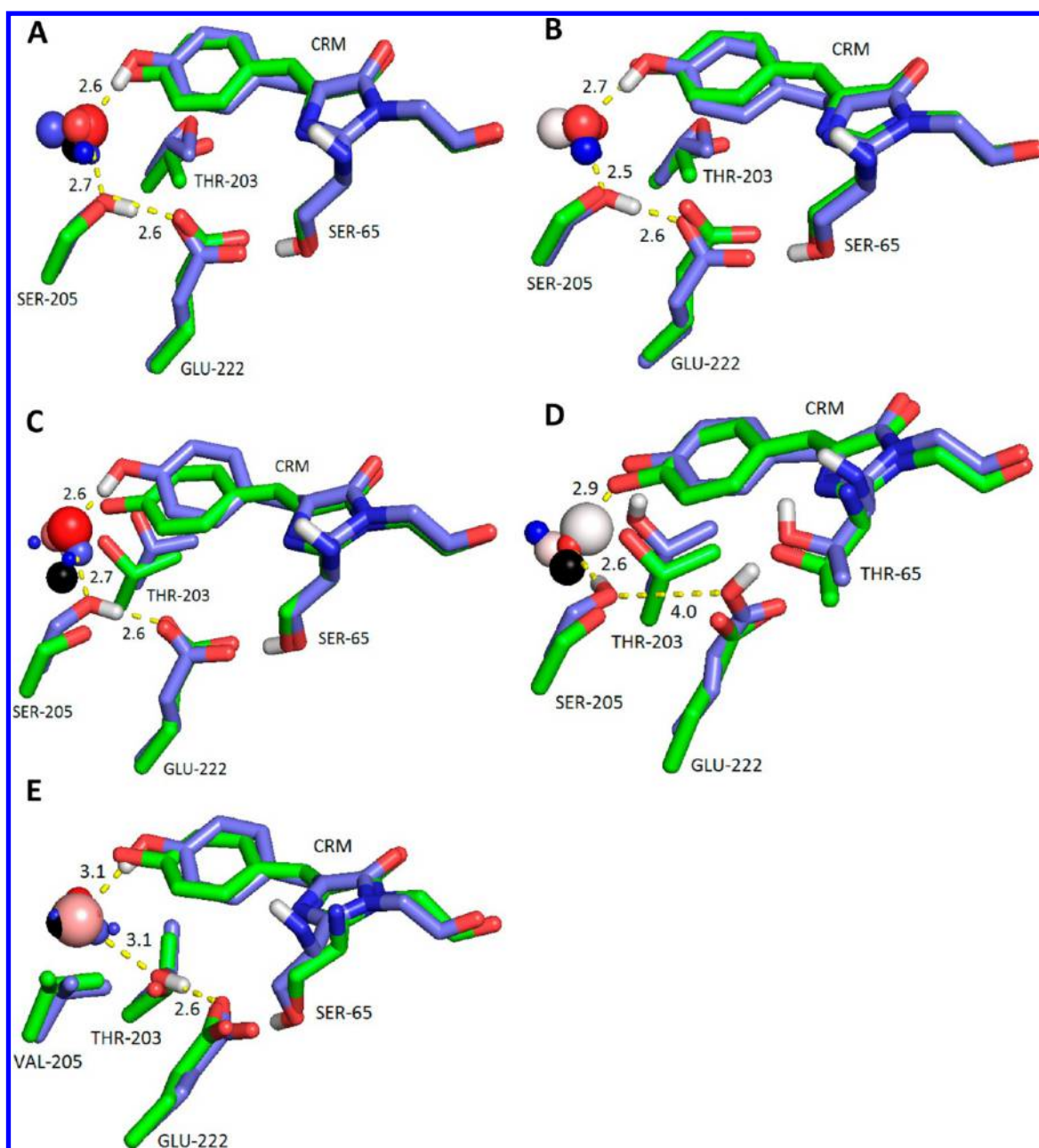
Figure 4. Overlay of crystal and representative time-averaged structures for several GFP cases, showing relevant distances along the CRM ↔ E222 proton translocation route. CRM and the side chains and hydration site involved in the proton wire are shown for (A) WT 1gfl-A, (B) WT 1gfl-B, (C) WT 2wur, (D) S65T, and (E) S205V. Crystal waters are represented as black spheres; see Figures 1 and 2 for other color conventions.

around the chromophore. In particular, representative structures of the varying states observed for the region around the chromophore (shown in Figure S2) illustrate that high-occupancy solvation is well conserved, except in the vicinity of S/V205, T203, and E222, where solvent fluctuations are coupled to the rotameric state of these three residues. Yet, a dominant state could be clearly isolated for each case (based on the number of runs with matching time-averaged solute−solvent structures), allowing for direct comparison between the crystal and the representative time-averaged structures of the corresponding set of simulations.

As part of our validation procedure, we overlaid each representative time-averaged structure onto its conforming crystal and visually assessed the agreement between solute structures and high-occupancy solvent/water sites. Particular

attention was given to discrepancies in residues/waters that are not solvent exposed and exhibit low β-factors, since these might be symptomatic of errors/inaccuracies in the simulation protocol. This practice offers a more rigorous validation criterion than the sole use of RMSDs or other one-dimensional order parameters. Concerning the solute, this process is exemplified for 1gfl-A in Figure 3A, which shows a clipped overlay of the crystal structure (green sticks) and the time-averaged (slate sticks) protein structure around the chromophore (green and slate lines, respectively). A similar result was observed for the solute time-averaged structures of the remaining four cases, which show detailed consistency with the corresponding crystals. We also found good overlap between the buried crystal waters and high-occupancy solvation sites in all cases. This is illustrated for 1gfl-A in Figure 3B,

which shows the locations of high occupancy waters (i.e., those with fairly low β-factor; gray, labeled by β-factor) around the chromophore (green) in the WT crystal structure, and in Figure 3C, which depicts the superposition, onto Figure 3B, of the high-occupancy solvation sites (red−white−blue translucent spectrum, see Appendix 1 and the figure caption for more information) around the time-averaged structure of the chromophore (slate).

Of special interest is the region delimited by the black oval in Figure 3C which, despite displaying high-occupancy solvation around the four crystal waters therein, stands out as having a more dynamic solvation structure. [The dynamic nature of these waters is suitably represented by the all-occupancy solvation surface shown as a dense network of blue lines connecting solvent occupation sites (inset of Figure 3C), which reveals constant solvent diffusion within this region.] In particular, only a few rather small high-occupancy solvation spheres are observed around crystal water W357, in agreement with its reported β-factor of 33.60 (see Figure 3C), which is the highest among all 1gfl-A buried crystal waters. An increased motion of the waters around this region is also observed for the WT 1gfl-B and 2wur simulations (for reference, see Figure S2B and C). Such dynamic behavior is consistent with the absence of a crystal water in this region for 1gfl-B (see encircled waters in Figure 2), and in the degenerate positions of crystal waters W2108 and W2307 in 2wur.

Our analysis also suggests that the location of the crystal water enclosed in a black triangle in Figure 2, which is unique to 2wur, does not correspond to a high-occupancy solvation site; although water does sample this region in about half of our WT runs, high-occupancy spheres were only observed in 1 of the 30 simulations. This result is consistent with the fact that the β-factor for this crystal water is the highest among all the buried waters reported in 2wur. On the other hand, the crystal water enclosed in the black square in Figure 2, unique to 1gfl-A, remained trapped in this region throughout the 1gfl-A simulations; yet, water was unable to diffuse into this region in any of the 1gfl-B and 2wur simulations. The latter indicates the presence of a high-energy barrier that prevents this water from escaping or diffusing into the protein, noting that this limitation could be difficult to overcome even by means of enhanced sampling techniques due to the intrinsic stability of the barrel. Nevertheless, given that this water lies within a rather small hydrophobic enclosure, we surmise that solvation of this region is highly enthalpically and entropically unfavorable and thus unlikely to occur.

We subsequently carried out a more stringent assessment of the accuracy of our models. Specifically, somewhat modest structural changes (relative to WT) are observed in the crystal structures of mutants S65T and S205V, along the chromophore → E222 proton transfer route (see Figure 1). Nevertheless, these changes lead to dramatic differences in the proton transfer rate along that pathway. We therefore sought to test whether these subtle, but key, structural variations are manifested in the representative time-averaged solute−solvent structures.

Figure 4 shows an overlay of the crystal and representative time-averaged positions of the three residues and the solvation site directly involved in the chromophore ↔ E222 proton wire, for WT 1gfl-A (Figure 4A), WT 1gfl-B (Figure 4B), WT 2wur (Figure 4C), S65T (Figure 4D), and S205V (Figure 4E). Altogether, our time-averaged representations do a remarkable job of capturing the solute conformational changes observed in

the crystal. For WT and S65T, T203 remains uninvolved in the chromophore ↔ E222 proton translocation route, in the first case displaying flexibility due to its inability to establish strong H-bonds with its neighboring residues or with other solvent sites, and in the second case, H-bonding with the chromophore's anionic phenol group. Conversely the relevant shift in the position of T203 for S205V, which acts as a bridge between E222 and the high-occupancy solvation site, is well reproduced. Moreover, the distances between the O−O groups involved in the wire for the time-averaged conformations of WT and S65T are consistent with those of the corresponding crystals (see Figures 4A−C and 1A and B for WT and Figures 4D and 1C for S65T), whereas a somewhat more significant difference is seen for S205V. Specifically for S205V, the time-averaged O−O distances along the wire are 3.1−3.1−2.6 Å (Figure 4E), while the corresponding distances in the crystal are 2.7−3.5−2.6 Å (Figure 1D). Based on the particularly long 3.5 Å distance between the structural water and T203 observed in the S205V crystal structure (which may be too long to support proton transfer), it is reasonable to argue that the S205V simulation refines the crystal by further relaxing the relative position of the three "stepping stones" to fairly even, albeit quite long distances. Such distances are likely to allow proton translocation to E222, while still limiting the transfer rate relative to the WT. It is also worth noting that, in addition to longer "hopping" distances, the slower rate observed for S205V could also be coupled to the seemingly inferior proton transfer capability of threonine, relative to that of serine.[77]

Further examination of the high-density solvation sites produced by the TADAS gives additional details on the practicality of our approach. For the four cases in which the chromophore's phenol is protonated (Figure 4A−C and E), the phenolic hydroxyl points directly to a sphere whose intense red coloring color indicates almost exclusive population by water oxygen. This oxygen-only sphere is part of a stable solvation site (i.e., all spheres in Figure 4) that remains within only eight voxels for 70−90% percent of the simulation time for these four cases. This contrasts with the white coloring of the corresponding sphere in Figure 4D, which indicates a 2:1 hydrogen:oxygen (i.e., waterlike) occupancy ratio, suggesting a weaker interaction between this water and the anionic phenol group. In turn, the water in Figure 4D now accepts an H-bond from the hydroxyl of S205, as readily deduced from the rotation of this group toward a red-colored sphere within the solvation site. The most probable locations of water hydrogens are similarly specified in the TADAS. Figure 4A and C−E show spheres of deep blue color (i.e., regions of hydrogen-only occupancy) at the left side of each solvation site, corresponding to long-lived interactions between water hydrogen and the carbonyl of N146 (not shown). Conversely, Figures 4A−C and E display a second region that is mostly populated by hydrogen (deep blue spheres at the bottom-right of each solvation site), indicative of the recurrent interaction between the hydrogen of the water and the oxygen of T203's hydroxyl. Overall, the statistics embedded in the high-occupancy TADAS representation (i.e., spheres colored by red−white−blue spectrum) give valuable insights about the properties of each water. These include (i) the mobility of a structural water, which can be derived from the size (or absence) of the sphere(s) and/or from the all-occupancy surface, and (ii) the specificity of water oxygen and hydrogen interactions with the solute or with other waters, as inferred by the coloring of the spheres (which can be very useful for identifying the most probable hydrogen bonding
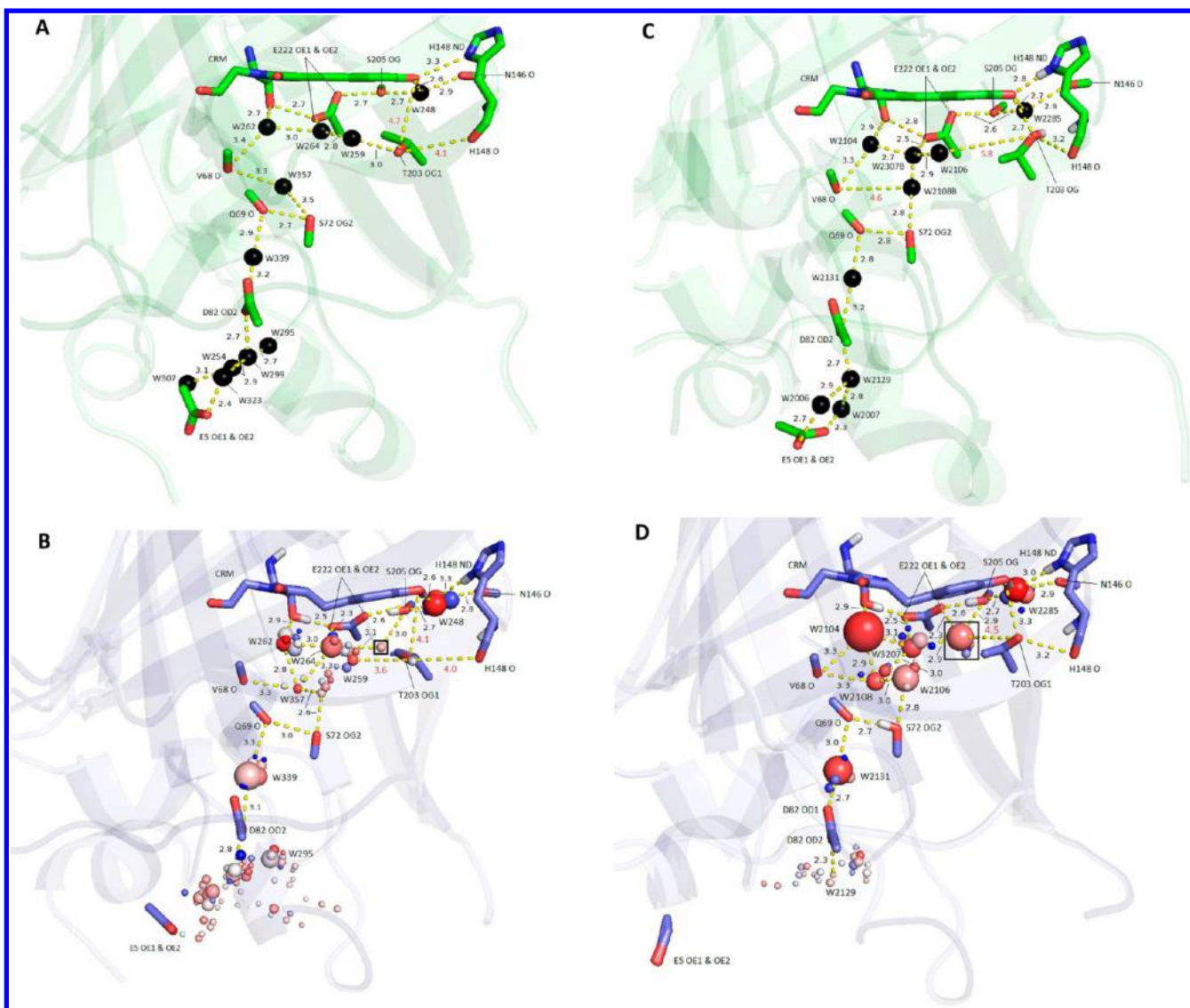
Figure 5. H-bonding network around the WT GFP chromophore for (A) crystal structure 1gfl-A; (B) representative time-averaged solute−solvent structure from 1gfl-A simulations; (B) crystal structure 2wur; and (D) representative time-averaged solute−solvent structure from 2wur simulations. Only the chromophore, relevant side chain groups, and high occupancy hydration sites/crystal waters along the route are shown. For each case, the numbering of the waters is consistent with the corresponding crystal structure. The values in red correspond to measurements that are notably longer than H-bonding distance. The β-barrel is also shown for reference.

network in case of ambiguity in the crystal structure). It is also worth mentioning that the complete TADAS surface (see for instance the inset of Figure 3C) includes statistics on low-density solvation that is not observed in the crystal, information that can be highly relevant for structure−function interpretation and de novo drug design.

Despite the current consensus regarding the mechanism of proton shuttling from the chromophore leading to the effective protonation of E222,[43,47] it is uncertain whether the proton is further translocated within the protein or out to bulk solvent, and if so, the most likely pathways for this process. An interesting hypothesis that finds support in various experimental studies has been put forward by Agmon and co-workers.[45,78] Namely, these authors postulated that the proton released upon photoexcitation can undergo a one-dimensional random walk, consistent with a geminate recombination model (which would allow for reprotonation of the chromophore), along an internal proton network that is considerably more

extensive than the short chromophore ↔ E222 wire discussed above; this model is in quantitative agreement with the $t^{-1/2}$ decay observed at temperatures lower than 230 K.[19] Moreover, Agmon et al. proposed that the proton released from the phenol of the chromophore can eventually exit the protein through two short proton wires: the first one comprised of the high-occupancy water/solvation site observed in Figure 4 and the backbone O of N146, and the second one formed upon rotation of the T203 side chain, which now acts as a bridge between the chromophore's phenol and the backbone O of H148. The existence of an irreversible proton exit network is quantitatively consistent with the $t^{-3/2}$ decay observed at temperatures above 245 K,[79] whose rationalization would otherwise require the presence of a three-dimensional internal H-bonding network which has not yet been observed for WT GFP. We now describe the extended H-bonding pathways from/to the chromophore's phenol (both internal and to bulk solvent) identified by means of a broader examination of our

time-averaged structures, and discuss our findings in light of the above-mentioned hypotheses.

Figure 5A and C illustrate static H-bonding networks construed from the phenol of the chromophore (CRM), as derived from WT crystal structures 1gfl-A and 2wur. Both diagrams show primary pathways to bulk solvent involving various key waters and residues S205, E222, V68, Q69, S72, D82, and E5, whose H-bonding distances appear to be suitable for proton transfer along this route. Our representative time-averaged structures, shown in Figure 5B and D for the 1gfl-A and 2wur simulations, respectively, similarly support the likelihood of proton transfer to/from bulk along this route. Moreover, Figure 5B and D exhibit distinctive features arising from solute–solvent dynamics, which improve the accuracy of the structural analysis. For instance, examination of crystal 1gfl-A (Figure 5A) in the absence of crystal 2wur might have led to the conclusion that the distances between W262–V68 O (3.4 Å) and W357–S72 OG2 (3.5 Å) are too long to allow for efficient proton transfer. Conversely, the time-averaged structure of 1gfl-A (Figure 5B) readily highlights the mobile nature of the local solvation in this region, whose high-occupancy centers equilibrate at positions that preserve H-bonding distances in this region.

Similarly, our dynamic structures reveal that D82 OD2 is surrounded by multiple high-occupancy solvation spheres (see Figure 5B and D) that are directly connected to bulk solvent, suggesting that proton transfer to solvent is conceivable even without the involvement of E5. Moreover, the shape of the E5 carboxylic acid in Figure 5B and D (where the ensemble-averaged position of both OE atoms has collapsed) indicates that this group is rotating continuously out in solvent, maximizing its entropy at little enthalpic cost. The latter result might be difficult to infer from the crystal structures alone, given that the D82 side chain is part of a pocket that appears to be shielded from bulk solvent in 1gfl-A and 2wur (i.e., the D82 side chain is surrounded by crystal waters of low β-factor), which would otherwise suggest that proton transfer is only possible when the somewhat mobile side chain of E5 (as hinted by its moderately high β-factor in the crystals) is able to establish a position similar to that observed in the crystals (see Figure 5A and C), where it H-bonds with the cluster of low β-factor crystal waters around the side chain of D82. Note that our findings do not negate the hypothesis that E5 plays a crucial role in WT GFP as a proton collector.[78]

Investigation of the primary H-bonding network from the CRM to bulk by means of our time-averaged structures also led to the identification of a high occupancy solvation site that does not correspond to any of the crystal waters in 1gfl-A or 2wur. This unlabeled site is enclosed in the black square in Figure 5B and D. Interestingly, it is directly H-bonded to S205 and to the cluster of dynamic solvation that eventually leads to bulk solvent through D28, implying that E222 could possibly be bypassed en route to bulk solvent along this pathway.

An alternate proton escape pathway that was previously proposed[45] is also apparent from our simulations. Namely, the CRM → T203 OH → H148 O → bulk route shown in Figure 5C and D for the crystal and representative time-averaged structures of 2wur, respectively, displays H-bonding distances that are appropriate for proton transfer. In addition, given that our simulations support a high stability of the CRM phenol and H148 O groups and a permanent exposure of H148 O to bulk solvent, our analysis agrees with previous studies[45,78] in that this pathway is essentially dependent upon the rotameric state

of the T203 side chain. The ON/OFF likelihood of proton transfer along this path was qualitatively assessed from our simulations; that is, the ON state (Figure 5C and D) was observed in all of the time-averaged structures corresponding to the dominant state of 2wur (6 out of 10 runs) and in 3 out of 10 time-averaged structures for the 1gfl-A simulations. We remark that even before the 2wur structure had been solved, this switch was ingeniously proposed by Agmon,[45] who used the crystal structure of mutant S65T (see Figure 4D) to hypothesize that the rotation of T203 was also plausible in the WT protein. Despite the latter, a dynamic analysis, such as the one presented here, provides a straightforward way to elucidate probable states that might otherwise remain hidden within static views such as that in the 1gfl-A crystal structure (see Figure 5A).

A third proton exit route, corresponding to the CRM → W248/W2285 → N146 O → bulk pathway, was also evaluated in light of our time-averaged structures. Figure 5A–D displays a conserved H-bond between W248/W2285 and N146 O, implying that this step is facile. However, our simulations also suggest that there is no rotameric shift of the N146 backbone carbonyl toward solvent. Even more, bulk water only sporadically diffuses into the vicinity of N146 O, as illustrated by Figure S3A and B, which highlights the absence of high-occupancy solvation within direct H-bonding distance from N146 O, and by Figure S3C and D, showing how the all-occupancy network still samples regions that are close enough to N146 O to allow for a proton transfer event. Given the latter, we expect this route to have low, yet nonzero, likelihood of proton transfer to bulk.

Lastly, our time-averaged analysis shows no plausible proton transfer route between the network depicted in Figure 5 and the so-called biosynthesis wire[78] spanning the region between the chromophore and the end of the β-barrel opposite to E5. Specifically, we see no bridging solvation between both networks, nor a clear rotation of key groups (e.g., carbonyls of V68 or Q69), such that the proton released from the chromophore's phenol can diffuse to the biosynthesis wire.

Altogether, our analysis provides a fully dynamic picture of GFP's H-bonding network that can lead to proton translocation upon photoexcitation of the chromophore. We remark that a stepwise dynamics analysis, in which all distinct states along a particular proton translocation route are studied individually, may offer additional mechanistic details relevant to understanding GFP's photoexcitation process.

PfM2DH. The crystal structures of the binary (enzyme–NAD+) and tertiary (enzyme–NAD+–mannitol) complexes of PfM2DH are highly homologous (the all-atom RMSDs between 1lj8 and the A and B monomers of 1m2w are 0.97 and 0.89 Å, respectively). However, differences between both complexes along the proton transfer channel are apparent (from the catalytic site to bulk solvent), discrepancies that have been previously studied in detail to infer proton translocation mechanisms for this channel.[40] We sought to accurately reproduce these structural differences in the MD time-averaged solute–solvent structures. Specifically, considering the experimental pK_a values of 9.2 and 6.8 for K295 in the binary and ternary complexes,[64] respectively, we modeled the enzyme–NAD+ complex with a protonated K295 and either protonated or deprotonated E292, and the ternary enzyme–NAD+–mannitol complex in all four possible ionization states of the K295/E292 pair (the anionic and cationic states of these
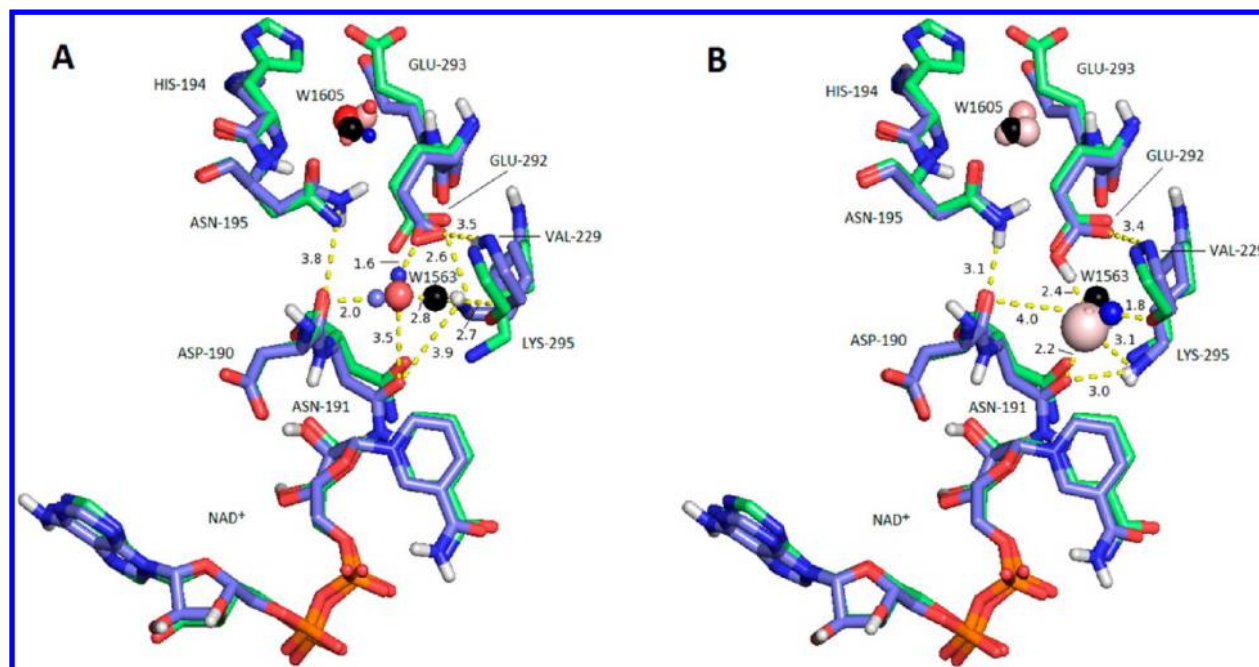
Figure 6. Overlay of the crystal and representative time-averaged structures of the PfM2DH−NAD$^+$ complex along the proton transfer channel from the catalytic site to bulk solvent, for (A) E292$^-$/K295$^+$ and (B) E292/K295$^+$. Only those distances and residue atoms of the enzyme that are relevant for the proton wire are shown, and nonpolar hydrogens have been removed.

residues are labeled with a $^+$ or a $^-$ sign in the following discussion).

Overall, superposition of the time-averaged protein structures with varying ionization states onto the corresponding crystal structure reveals good agreement (all-atom RMSD < 1.2 Å) with all binary and ternary simulations. This behavior provides preliminary support for the consistency of our simulations, suggesting that changes in the ionization state of K295 and/or E292 lead only to minor changes in protein structure. This is expected from the low all-atom RMSD between the binary and ternary crystal complexes whose structures are believed to denote different K295/E292 protonation states. In general, we consider that the incidence of sizable deviations in the RMSD between a simulated enzyme structure and its corresponding *high-resolution* crystal (e.g., see Figure S2 of ref 64) is sufficient motivation for reexamining the protonation and rotameric state assignment of the starting MD structure(s).

Aside from displaying a high similarity with the overall crystal conformations, the time-averaged structure of the dominant state (after analysis of all replicate simulations) should also properly reproduce key discrepancies along the proton channel arising from changes in the protonation of K295/E292 and/or the presence of a ligand.

Concerning the binary complex, we find that the corresponding replicate simulations for both E292$^-$/K295$^+$ and E292/K295$^+$ sample mostly a single solute−solvent state in each case (see Figure S4). Figure 6 shows an overlay between the crystal structure (green carbons) and a representative MD time-averaged structure of the dominant state (slate carbons) for E292$^-$/K295$^+$ (Figure 6A) and E292/K295$^+$ (Figure 6B); only NAD$^+$ and relevant residues and waters/solvation sites along the catalytic proton channel are displayed. In both cases, the locations of several residues and NAD$^+$ and W1605 (numbering based on 1lj8) are very similar to those of the crystal. However, a major difference stands out; namely, that in Figure 6A the K295$^+$ side chain rotates sharply

toward E292$^-$, thereby establishing a local H-bonding network that is notably different from that observed in Figure 6B in both the crystal and the time-averaged structures. In particular, Figure 6B suggests that the hydrogens of W1563 recurrently H-bond with the carbonyl oxygen of V229 (time-averaged H−O distance =1.8 Å; see figure caption for more information) and the δ-oxygen of N191 (H−O = 2.2 Å), whereas the W1563 oxygen interacts with the hydroxyl hydrogen of E292 (O−O = 2.4 Å) and the ε−amino group of K295$^+$ (O−N = 3.1 Å). This network is further stabilized by interactions between K295$^+$ and N191 (N−O = 3.0 Å), N195 and D190 (N−O = 3.1 Å), K295$^+$ and N300 (N−O = 2.7 Å; not shown in Figure 6B), and possibly between E292 and V229 (O−N = 3.4 Å). Conversely, the time-averaged structure in Figure 6A suggests that the shifted W1563 hydrogens now H-bond with an oxygen of E292$^-$ (H−O = 1.6 Å) and the carbonyl oxygen of D190 (H−O = 2.0 Å), whereas the W1563 oxygen appears to H-bond mainly with the NH$_3$ group of K295$^+$ (O−N = 2.8 Å), thereby losing its interaction with the carbonyl of N191 (O−O = 3.5 Å). Additional stabilization is provided by H-bonds between K295$^+$ and E292$^-$ (N−O = 2.6 Å), K295$^+$ and V229 (O−O = 2.7 Å), and conceivably between E292$^-$ and V229 (O−N = 3.5 Å). We further note that in the E292$^-$ simulation, the H-bonds between K295$^+$ and N300 (N−O = 4.0 Å; not shown in Figure 6A) and between N195 and D190 (N−O = 3.8 Å) are effectively lost (the former due to the shift of K295$^+$, and the latter due to the rotation of N195's NH$_2$ group away from D190 to avoid proximity to the high-occupancy solvation hydrogens of W1563). Overall, from Figure 6, it is clear that the enzyme−NAD$^+$ solute−water crystal structure is properly reproduced only when E292 is protonated. Moreover, the local H-bonding network around W1563 appears to be more unfavorable for E292$^-$, suggesting that the protonated state may be preferred under physiological conditions.

The most probable protonation state of the ternary enzyme−NAD$^+$−mannitol complex was also elucidated by analysis of all
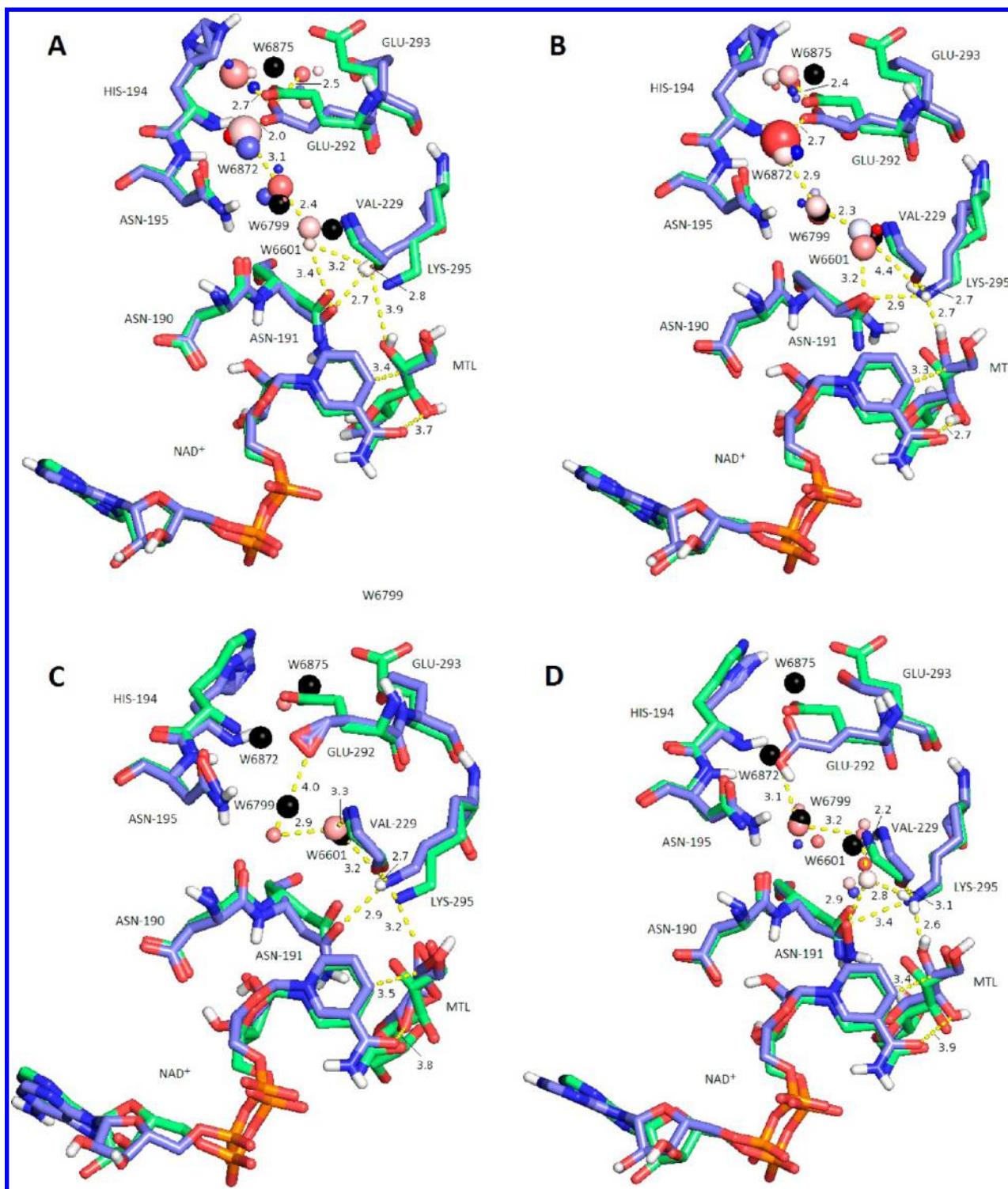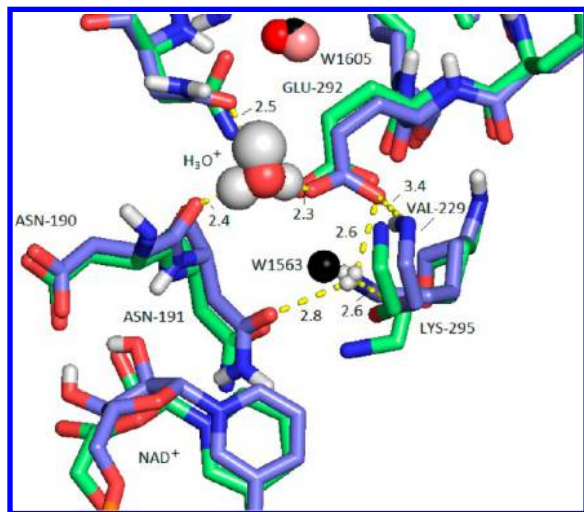
Figure 7. Overlay of the crystal and representative time-averaged structures of the PfM2DH−NAD$^+$−MTL complex along the proton transfer channel from the catalytic site to bulk solvent, for (A) E292$^-$/K295$^+$, (B) E292$^-$/K295, (C) E292/K295$^+$, and (D) E292/K295.

replicate simulations for each case. E292$^-$/K295 is the most stable case, as evidenced by the number of runs (9 out of 10) whose time-average structures represent the same state (see Figure S5). For each of the remaining cases the dominant state is still straightforward to identify, given that the time-averaged structures of 6 or 7 out of the 10 replicates correspond to the same state. Figure 7 shows a superposition of the crystal and representative time-averaged structures of the ternary complex

along the proton channel. Key discrepancies become apparent in each case upon detailed inspection, resulting in unambiguous identification of the ionization states for K295 and E292 that most likely correspond to those of the crystal structure. The most evident deviations are observed in Figure 7C and D (states E292/K295$^+$ and E292/K295, respectively), which display a clear downward shift of the E292 COOH group, leading to the displacement of hydration site W6872

(numbering based on 1m2w) and consequent disruption of the water wire to bulk solvent. Note that three high-occupancy solvation sites between K295 and E292 are still observed in Figure 7D, but their location differs substantially relative to those in the crystal. Thus, it is improbable that the crystal conformation represents a state in which both E292 and K295 are neutral. Another clear mismatch is observed in Figure 7A, where the side chain of K295$^+$ swings away from the mannitol's (MTL) proton donor (N−O = 3.9 Å) relative to the crystal position (O−N = 2.8 Å). Interestingly, an analogous, although less visually distinct, outcome is also observed in Figure 7C. In contrast to the previous cases, Figure 7B shows that the distance between the ε-amino group of K295 and the MTL O2 in the E292$^-$/K295 time-averaged structure (2.7 Å, see Figure 7B) is well matched to that of the crystal structure. Furthermore, the three waters involved in the proton wire, as well as the position and orientation of the E292 COO$^-$ group, are most accurately reproduced by means of this protonation state. Figure 7B also exhibits the most consistent relative orientation between NAD$^+$ and MTL around the region involved in the hydride transfer, as measured by the distance between the reactive carbons (see Figures 7A−D; C−C crystal = 2.9 Å) and the stabilizing amide-hydroxyl interaction (O−O crystal = 2.9 Å). Based on the above arguments, we believe that the crystal structure represents a state in which both E292 and K295 are deprotonated.

Additional simulations were carried out to further examine the reliability of the above analysis. Namely, we tested whether our methodology could reproduce a more sensible variation of the E292$^-$/K295$^+$ network proposed in ref 64 (see Figure S1A in this document) for the binary complex, in which the water therein (i.e., W1563) is replaced by a hydronium. We ran five replicas of this state, all of which display time-averaged structures that are almost coincident with the representative state shown in Figure 8. Thus, this case leads to an arrangement that is somewhat similar to that observed in Figure 6A, where the hydronium is displaced from the location of crystal W1563 owing to the upward shift of K295$^+$; the H$_3$O$^+$ is stabilized via interactions with the carbonyl oxygen of N190 (O−O = 2.4 Å), the E292$^-$ side chain (O−O = 2.3 Å), and the N195 side chain

Figure 8. Superposition of the crystal and representative time-averaged structures of the PfM2DH−NAD$^+$ complex around W1563, for a set of simulations in which W1563 is modeled as an H$_3$O$^+$.

amide oxygen (O−N = 2.5 Å). The latter results strengthen our hypothesis that Figure S1B, corresponding to the doubly protonated E292/K295$^+$ case, is the likely ionization state in the 1lj8 crystal structure. Nonetheless, this assertion implies that two protons should be released during presteady state kinetics, granted that the most probable state of 1m2w is that in which both E292 and K295 are deprotonated. Therefore, we explored an alternative scenario that would still be consistent with the kinetic measurements by Klimacek et al.,[63,64] in which H194 (the only candidate residue along the proton channel that could in principle trap one of the protons released before it reaches bulk solvent) is doubly protonated in the enzyme−NAD$^+$−mannitol complex. Interestingly, the time-averaged structures along the proton escape channel for the five replicates that we ran are in good agreement with the ternary crystal structure (see Figure S6), except that H194$^+$ now establishes a long-lived H-bond with the COO$^-$ group of E292$^-$, as depicted in Figure 9. This outcome supports the viability of this state; yet, its
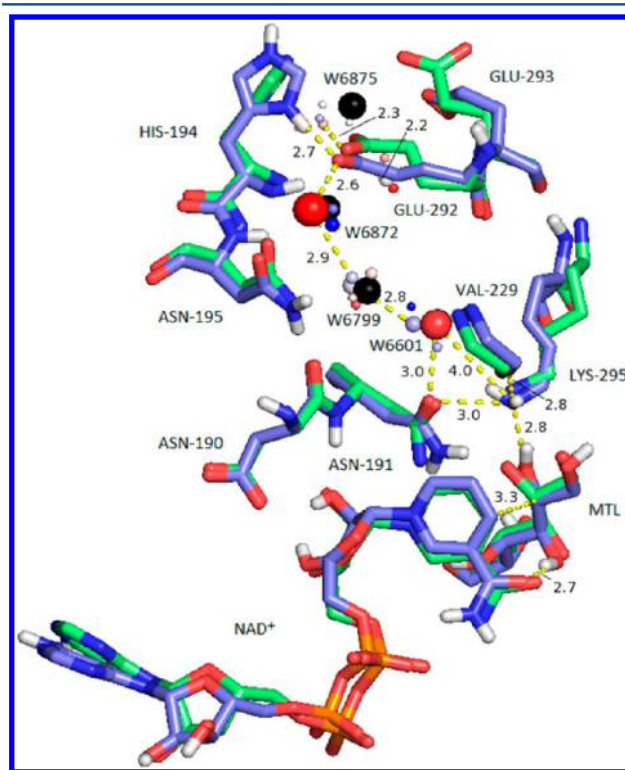

Figure 9. Overlay of the crystal and representative time-averaged structures of the PfM2DH−NAD$^+$−MTL complex along the proton transfer channel from the catalytic site to bulk solvent, for with E292$^-$/K295 and H194$^+$. Relevant H-bonding distances for the representative time-averaged structure are shown, to facilitate comparison with those in Figure 7B.

actual biological likelihood is an open question that could be better elucidated by means of QM/MM calculations, and/or more quantitative kinetic experiments.

Plausible insights about the proton transfer mechanism can be derived from our time-averaged analysis of the enzymatic oxidation of mannitol via PfM2DH. To this end, we also carried out five replicate simulations of the enzyme−NADH−fructose complex for each of the four possible protonation states of the E292/K295 pair. Time-averaged structures illustrating the major states sampled along the proton channel, for each case, are shown in Figure S7. The E292$^-$/K295$^+$ case shows good
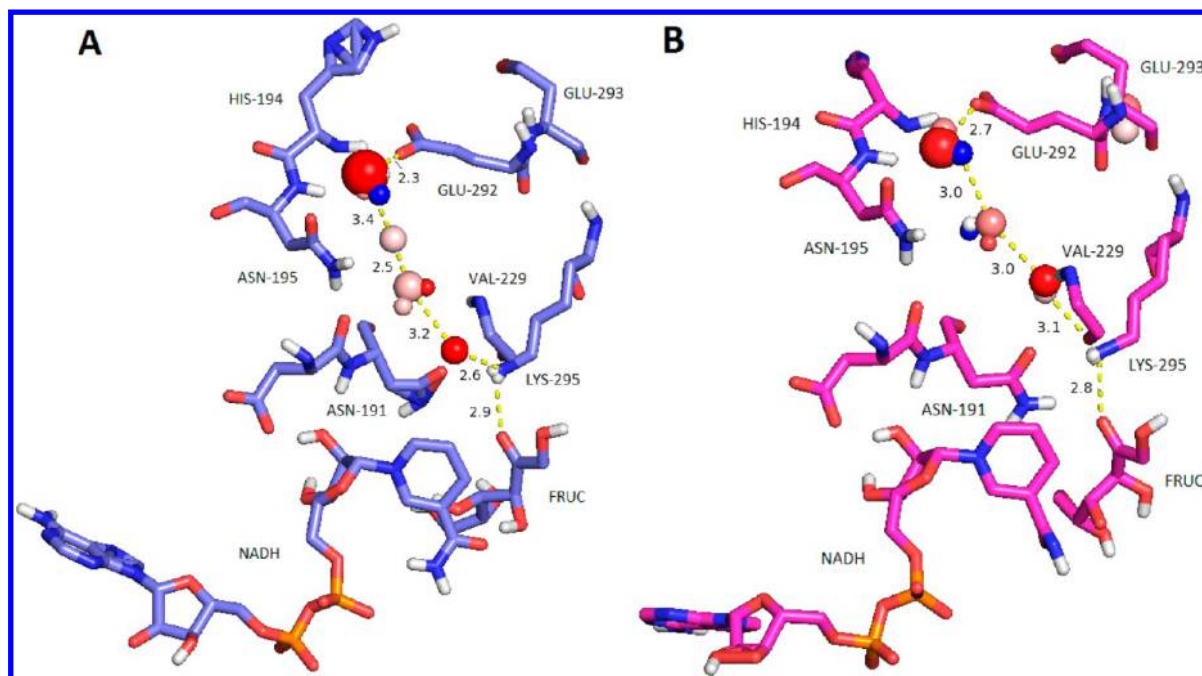
Figure 10. Representative time-averaged structures of the PfM2DH−NADH−fructose complex for the E292⁻/K295⁺ protonation state, illustrating the H-bonding network that would allow for proton escape from the catalytic site to bulk solvent after mannitol oxidation: (A) state 1 and (B) state 2.

agreement between the solute structures for all runs, but samples two distinct solvent states along the channel, characterized by either 3 or 4 solvation sites (see Figures 10 and S7A). E292⁻/K295 and E292/K295⁺ display higher structural diversity. Their various states are mainly defined by the prevailing rotamer adopted either by E292⁻ in the former case (see Figure S7B) or by K295⁺ in the latter case (see Figure S7C), changes that are closely coupled with shifts in the solvation sites. In contrast, a dominant state is easily identified for E292/K295 (see Figure S7D), whose representative time-averaged structure nearly matches that of the enzyme-NADH-mannitol complex with the same protonation state (see Figure 7D). Note that for the sake of brevity, we exclude from this work further details concerning the analysis of the ternary fructose-bound structures.

The key variations identified through the study of the binary and ternary complexes facilitate the isolation of distinctive milestones that may conform to biologically relevant steps along the reactive process. Overall, our results support the postulate that proton escape due to mannitol binding and oxidation proceeds along a water wire between the substrate and bulk solvent, with E292 acting both as a gate-keeper that modulates solvation of the channel and a bridge for proton translocation to bulk solvent, once this residue is facing out toward bulk solvent. Based on our findings, release of a proton to bulk solvent via direct shuttling of E292 is an unlikely scenario, because this residue appears to be protonated in the binary complex. Concerning a potential mechanism for this process, our simulations suggest that upon mannitol binding, the ε-amino group of K295⁺ is displaced by ∼0.7−1.2 Å (relative to that in the binary crystal) toward solvation site W1563 due to electrostatic hindrance with the reactive hydroxyl of mannitol. This movement compresses solvation site W1563, in turn, forcing E292 out toward bulk solvent. Upon exposure to solvent, E292 deprotonates and allows the entrance of water into the channel, whose occupancy is

concentrated at W6601, W6799, and W6872. In particular, the local network stabilized by these solvation sites, and by the δ-oxygens of N299 and N191, is optimized such that proton transfer between K295⁺ and bulk solvent readily occurs. Consequently, the deprotonated K295 side chain is capable of shifting back toward the reactive hydroxyl of mannitol thereby "activating" the substrate for hydride transfer and oxidation. The proton released from this reaction can traverse to bulk along a wire comprised by a now readjusted network of three or four solvation sites and the side chain of E292⁻ (see Figure 10). Return to the binary complex state proceeds via a series of steps that are presumably triggered by the dissociation of fructose. Elucidation of a tentative mechanism for this process is beyond the scope of this work.

As a final and rather supplementary observation, we note the high flexibility of the side chains of E293 and H194 (except when acting as a proton trap, see Figure 9), as evidenced from the time-averaged structures in Figures 6 and 7 and to some extent from the crystal structures (moderately high β-factors are reported for these side chains in 1m2w but not in 1lj8). The palpable mobility of these two residues arises from the inability of these side chains to establish a more favorable free energetic interaction among themselves relative to the ones they can independently form with solvating water. Strictly speaking, the entropic cost of having these side chains fixed at the surface of the protein outweighs the enthalpic gain attained when they are permanently H-bonded to each other. The latter suggests that, unlike what has been previously proposed,[64] the H194−E293 H-bond is not long-lived and thus plays no major role in the mechanism of proton translocation to bulk solvent. However, this does not imply that these residues cannot act as a proton buffer at the entrance of the channel, given their ability to modulate the local electrostatic environment. Also, intimately related to the latter arguments is the finding that W6875 is not properly reproduced in any of the cases illustrated in Figure 7. Instead, we observed several alternate high-occupancy sites that

can potentially accept a proton from E292 (e.g., two in Figure 7A and B and one in C). This behavior is consistent with the propensity of water to maximize its entropy while still roughly satisfying its H-bond requirements and those of the surface residues by localizing multiple lower-occupancy sites around the enzyme, rather than retaining a single higher-density water at a fixed position (as observed in the crystal structure). In general, it is clear that care should be taken when carrying out design efforts or inferring mechanisms based on crystal conformations of solutes and solvents at protein surfaces.

■ CONCLUSIONS AND FINAL REMARKS

The study of H-bond networks within and between the solvation layer and protein surface provides valuable insights about the fundamental structure−function relationships of aqueous systems, as supported by this work and by ample preceding research. Of particular interest in this regard is the cost of water and proton rearrangements, as well as the coupling between solute rearrangements and the cost of water transfer into and out of the solvation layer. Experimental elucidation of such networks is limited to high-density water positions (either H-bond-enriched or trapped water molecules) that have been accurately determined via high-resolution crystal structures. Elucidation of low-density, H-bond-depleted water positions exhibiting high β-factors falls largely within the realm of computation, and in particular, the analysis of MD simulation trajectories of both solute atoms and solvation. Thus, analysis of H-bond networks using high resolution crystallographic data will prove more fruitful when coupled with dynamic studies. Although this is a conventional practice in some settings, its increased use within a pharmaceutical research environment can help improve the current success of the drug discovery process.

The MD-based methodology introduced in this work shows promise as a tool to aid such advancements, given the accuracy and simplicity with which the resulting "single-snapshot" solute−solvent structures recapitulate the ensemble-averaged dynamic behavior of a system. This not only allows for a direct comparison of computational and crystallographic data, but also provides a detailed picture of occupancy by all water oxygens and water hydrogens at fine-grained positions within the simulation box. Such information cannot be obtained from the X-ray structure, yet it is critical for elucidation of the variation in directionality and stability of water in different microscopic regions of the H-bond network. Herein, we have made use of this approach to study proton transfer networks of the green fluorescent protein (GFP) and the *Pseudomonas fluorescens* mannitol 2-dehydrogenase (PfM2DH).

Regarding GFP, our results are consistent with previous studies[78] suggesting the presence of an extended H-bond network through which the proton can either exit to bulk or diffuse internally, possibly undergoing reversible geminate recombination with the anionic phenol of the chromophore. Our time-averaged structures also reproduce the relative distances between those residues/waters with low β-factor, and successfully capture minor structural changes due to point mutations S205V and S65T. More so, our study affords information that may be valuable to further understand GFP's proton transfer mechanism in light of available fluorescence data, including a previously unidentified solvation site and water/side chain phase space sampling proclivity.

Similarly for PfM2DH, our simulations are in general agreement with prior studies suggesting that the formation of a proton wire connecting bulk solvent with the catalytic site upon mannitol binding is promoted by the shift of a channel-capping glutamic acid side chain toward bulk solvent, with the concomitant rearrangement and deprotonation of a lysine side chain at the catalytic site. Nonetheless, our analysis suggests several differences in the reaction mechanism, relative to those proposed previously. Moreover, we find that both the E292 and K295 side chains appear to be protonated in the binary crystal complex (and deprotonated in the ternary crystal complex). While challenging previous kinetic studies, our hypothesis leads to a more consistent H-bonding network along the substrate's proton escape pathway. Our results further highlight the ability of our methodology to properly reorganize the solute−solvent structure based on changes in a single protonation state of a ligand group or a side chain involved in the proton transfer process.

Overall, detailed correspondence between MD simulations and high-resolution crystal structures representing the dominant state at physiological conditions is a central requirement for quantitative validation of modeling results. Furthermore, omission of this step may jeopardize the robustness of the outcome, possibly leading to erroneous conclusions. Surprisingly, the use of MD ensemble-averaged structures as a validation tool for MD simulations has, to our knowledge, been greatly overlooked, and MD runs are generally interpreted via order parameter analyses and/or the study of single snapshots considered descriptive of the conformation of interest. Our methodology offers a detailed, yet efficient framework for model scrutiny and subsequent prediction, which is apt for dynamic analysis of solvated systems. Additional remarks on the use of time-averaged structures are given in Appendix 2 of the Supporting Information.

Looking forward, we believe that fully dynamic solvation-based free energy analysis will prove to be an important advancement in the elucidation of a wide range of structure-free energy relationships relative to the "static" analysis of crystal structures and conventional analysis of explicit solvent MD simulations performed using global protein restraints and other simplifying assumptions.[4−9] In particular, we are currently evaluating a novel theoretical formulation to compute solvation thermodynamic quantities from all-occupancy solute−solvent statistics of MD simulations that employ only anchoring restraints for the protein.

■ ASSOCIATED CONTENT

⑤ Supporting Information

Appendix 1 includes additional computational details of the WATMD software, and Appendix 2 provides a supplementary discussion pertaining the use and analysis of time-averaged solute and solvent structures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: camilo.velez-vega@novartis.com
*E-mail: daniel.mckay@novartis.com.

Notes
The authors declare no competing financial interest.

current modeling techniques, and Dr. Bernd Nidetzky for his feedback on the interpretation of the kinetics experiments reported for PfM2DH.[64]

■ REFERENCES

(1) Pearlstein, R. A.; Hu, Q.-Y.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V. K.; Daniels, D.; Hanrahan, S.; Sherman, W.; Abel, R. New Hypotheses about the Structure-Function of Proprotein Convertase Subtilisin/kexin Type 9: Analysis of the Epidermal Growth Factor-like Repeat A Docking Site Using WaterMap. *Proteins* **2010**, *78*, 2571−2586.

(2) Tran, Q.-T.; Williams, S.; Farid, R.; Erdemli, G.; Pearlstein, R. The Translocation Kinetics of Antibiotics through Porin OmpC: Insights from Structure-Based Solvation Mapping Using WaterMap. *Proteins* **2013**, *81*, 291−299.

(3) Pearlstein, R. A.; Sherman, W.; Abel, R. Contributions of Water Transfer Energy to Protein-Ligand Association and Dissociation Barriers: Watermap Analysis of a Series of p38α MAP Kinase Inhibitors. *Proteins* **2013**, *81*, 1509−1526.

(4) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 044101.

(5) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein−ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808−813.

(6) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. *Computational Drug Discovery and Design*; Methods in Molecular Biology; Springer: Clifton, NJ, 2012; Vol. *819*, pp 393−404.

(7) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.* **2009**, *131*, 12430−12440.

(8) Gerogiokas, G.; Calabro, G.; Henchman, R. H.; Southey, M. W. Y.; Law, R. J.; Michel, J. Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *J. Chem. Theory Comput.* **2013**, *10*, 35−48.

(9) Haider, K.; Huggins, D. J. Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 2571−2586.

(10) Altarsha, M.; Benighaus, T.; Kumar, D.; Thiel, W. How Is the Reactivity of Cytochrome P450cam Affected by Thr252X Mutation? A QM/MM Study for X = Serine, Valine, Alanine, Glycine. *J. Am. Chem. Soc.* **2009**, *131*, 4755−4763.

(11) Altarsha, M.; Wang, D.; Benighaus, T.; Kumar, D.; Thiel, W. QM/MM Study of the Second Proton Transfer in the Catalytic Cycle of the D251N Mutant of Cytochrome P450cam. *J. Phys. Chem. B* **2009**, *113*, 9577−9588.

(12) Mikulski, R.; West, D.; Sippel, K. H.; Avvaru, B. S.; Aggarwal, M.; Tu, C.; McKenna, R.; Silverman, D. N. Water Networks in Fast Proton Transfer during Catalysis by Human Carbonic Anhydrase II. *Biochemistry (Mosc.)* **2013**, *52*, 125−131.

(13) Fishelovitch, D.; Shaik, S.; Wolfson, H. J.; Nussinov, R. How Does the Reductase Help to Regulate the Catalytic Cycle of Cytochrome P450 3A4 Using the Conserved Water Channel? *J. Phys. Chem. B* **2010**, *114*, 5964−5970.

(14) Tatoli, S.; Zazza, C.; Sanna, N.; Palma, A.; Aschi, M. The Role of Arginine 38 in Horseradish Peroxidase Enzyme Revisited: A Computational Investigation. *Biophys. Chem.* **2009**, *141*, 87−93.

(15) Maupin, C. M.; Saunders, M. G.; Thorpe, I. F.; McKenna, R.; Silverman, D. N.; Voth, G. A. Origins of Enhanced Proton Transport in the Y7F Mutant of Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2008**, *130*, 11399−11408.

(16) Taraphder, S.; Hummer, G. Protein Side-Chain Motion and Hydration in Proton-Transfer Pathways. Results for Cytochrome p450cam. *J. Am. Chem. Soc.* **2003**, *125*, 3931−3940.

(17) Shinobu, A.; Agmon, N. Mapping Proton Wires in Proteins: Carbonic Anhydrase and GFP Chromophore Biosynthesis†. *J. Phys. Chem. A* **2009**, *113*, 7253−7266.

(18) Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. A Short, Strong Hydrogen Bond in the Active Site of Human Carbonic Anhydrase II. *Biochemistry (Mosc.)* **2010**, *49*, 249−251.

(19) Leiderman, P.; Huppert, D.; Agmon, N. Transition in the Temperature-Dependence of GFP Fluorescence: From Proton Wires to Proton Exit. *Biophys. J.* **2006**, *90*, 1009−1018.

(20) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. A Potential Energy Function for Heterogeneous Proton-Wires. Ground and Photoactive States of the Proton-Wire in the Green Fluorescent Protein. *J. Chem. Theory Comput.* **2008**, *4*, 1138−1150.

(21) Braun-Sand, S.; Strajbl, M.; Warshel, A. Studies of Proton Translocations in Biological Systems: Simulating Proton Transport in Carbonic Anhydrase by EVB-Based Models. *Biophys. J.* **2004**, *87*, 2221−2239.

(22) Toba, S.; Colombo, G.; Merz, K. M. Solvent Dynamics and Mechanism of Proton Transfer in Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **1999**, *121*, 2290−2302.

(23) Gu, W.; Helms, V. Tightly Connected Water Wires Facilitate Fast Proton Uptake at The Proton Entrance of Proton Pumping Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 2080−2081.

(24) Pomès, R.; Roux, B. Structure and Dynamics of a Proton Wire: A Theoretical Study of H+ Translocation along the Single-File Water Chain in the Gramicidin A Channel. *Biophys. J.* **1996**, *71*, 19−39.

(25) Wood, M. L.; Schow, E. V.; Freites, J. A.; White, S. H.; Tombola, F.; Tobias, D. J. Water Wires in Atomistic Models of the Hv1 Proton Channel. *Biochim. Biophys. Acta* **2012**, *1818*, 286−293.

(26) Cukierman, S. The Transfer of Protons in Water Wires inside Proteins. *Front. Biosci. J. Virtual Libr.* **2003**, *8*, s1118−s1139.

(27) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The Formation and Dynamics of Proton Wires in Channel Environments. *Biophys. J.* **2001**, *80*, 1691−1702.

(28) Wikström, M.; Verkhovsky, M. I.; Hummer, G. Water-Gated Mechanism of Proton Translocation by Cytochrome c Oxidase. *Biochim. Biophys. Acta* **2003**, *1604*, 61−65.

(29) Hasan, S. S.; Yamashita, E.; Baniulis, D.; Cramer, W. A. Quinone-Dependent Proton Transfer Pathways in the Photosynthetic Cytochrome b6f Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 4297−4302.

(30) Freier, E.; Wolf, S.; Gerwert, K. Proton Transfer via a Transient Linear Water-Molecule Chain in a Membrane Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11435−11439.

(31) Akeson, M.; Deamer, D. W. Proton Conductance by the Gramicidin Water Wire. Model for Proton Conductance in the F1F0 ATPases? *Biophys. J.* **1991**, *60*, 101−109.

(32) Cao, Y.; Váró, G.; Chang, M.; Ni, B. F.; Needleman, R.; Lanyi, J. K. Water Is Required for Proton Transfer from Aspartate-96 to the Bacteriorhodopsin Schiff Base. *Biochemistry (Mosc.)* **1991**, *30*, 10972−10979.

(33) Karahka, M. L.; Kreuzer, H. J. Charge Transport along Proton Wires. *Biointerphases* **2013**, *8*, 13.

(34) Cross, J. B.; Duca, J. S.; Kaminski, J. J.; Madison, V. S. The Active Site of a Zinc-Dependent Metalloproteinase Influences the Computed pKa of Ligands Coordinated to the Catalytic Zinc Ion. *J. Am. Chem. Soc.* **2002**, *124*, 11004−11007.

(35) Viswanathan, U.; Basak, D.; Venkataraman, D.; Fermann, J. T.; Auerbach, S. M. Modeling Energy Landscapes of Proton Motion in Nonaqueous, Tethered Proton Wires. *J. Phys. Chem. A* **2011**, *115*, 5423−5434.

(36) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601−604.

(37) Maupin, C. M.; Voth, G. A. Proton Transport in Carbonic Anhydrase: Insights from Molecular Simulation. *Biochim. Biophys. Acta* **2010**, *1804*, 332−341.

(38) Lian, P.; Li, J.; Wang, D.-Q.; Wei, D.-Q. Car−Parrinello Molecular Dynamics/Molecular Mechanics (CPMD/MM) Simulation Study of Coupling and Uncoupling Mechanisms of Cytochrome P450cam. *J. Phys. Chem. B* **2013**, *117*, 7849−7856.

(39) Remington, S. J. Green Fluorescent Protein: A Perspective. *Protein Sci. Publ. Protein Soc.* **2011**, *20*, 1509−1519.

(40) Kavanagh, K. L.; Klimacek, M.; Nidetzky, B.; Wilson, D. K. Crystal Structure of Pseudomonas Fluorescens Mannitol 2-Dehydrogenase: Evidence for a Very Divergent Long-Chain Dehydrogenase Family. *Chem. Biol. Interact.* **2003**, *143−144*, 551−558.

(41) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. Structural Basis for Dual Excitation and Photoisomerization of the Aequorea Victoria Green Fluorescent Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2306−2311.

(42) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. Ultra-Fast Excited State Dynamics in Green Fluorescent Protein: Multiple States and Proton Transfer. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8362−8367.

(43) Di Donato, M.; van Wilderen, L. J. G. W.; Van Stokkum, I. H. M.; Stuart, T. C.; Kennis, J. T. M.; Hellingwerf, K. J.; van Grondelle, R.; Groot, M. L. Proton Transfer Events in GFP. *Phys. Chem. Chem. Phys. PCCP* **2011**, *13*, 16295−16305.

(44) Van Thor, J. J.; Ronayne, K. L.; Towrie, M.; Sage, J. T. Balance between Ultrafast Parallel Reactions in the Green Fluorescent Protein Has a Structural Origin. *Biophys. J.* **2008**, *95*, 1902−1912.

(45) Agmon, N. Proton Pathways in Green Fluorescence Protein. *Biophys. J.* **2005**, *88*, 2452−2461.

(46) Van Oort, B.; ter Veer, M. J. T.; Groot, M. L.; van Stokkum, I. H. M. Excited State Proton Transfer in Strongly Enhanced GFP (sGFP2). *Phys. Chem. Chem. Phys. PCCP* **2012**, *14*, 8852−8858.

(47) Stoner-Ma, D.; Jaye, A. A.; Matousek, P.; Towrie, M.; Meech, S. R.; Tonge, P. J. Observation of Excited-State Proton Transfer in Green Fluorescent Protein Using Ultrafast Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2005**, *127*, 2864−2865.

(48) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.

(49) Abbruzzetti, S.; Bizzarri, R.; Luin, S.; Nifosì, R.; Storti, B.; Viappiani, C.; Beltram, F. Photoswitching of E222Q GFP Mutants: "concerted" Mechanism of Chromophore Isomerization and Protonation. *Photochem. Photobiol. Sci.* **2010**, *9*, 1307−1319.

(50) Bizzarri, R.; Serresi, M.; Cardarelli, F.; Abbruzzetti, S.; Campanini, B.; Viappiani, C.; Beltram, F. Single Amino Acid Replacement Makes Aequorea Victoria Fluorescent Proteins Reversibly Photoswitchable. *J. Am. Chem. Soc.* **2010**, *132*, 85−95.

(51) Stoner-Ma, D.; Jaye, A. A.; Ronayne, K. L.; Nappa, J.; Meech, S. R.; Tonge, P. J. An Alternate Proton Acceptor for Excited-State Proton Transfer in Green Fluorescent Protein: Rewiring GFP. *J. Am. Chem. Soc.* **2008**, *130*, 1227−1235.

(52) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal Structure of the Aequorea Victoria Green Fluorescent Protein. *Science* **1996**, *273*, 1392−1395.

(53) Shu, X.; Leiderman, P.; Gepshtein, R.; Smith, N. R.; Kallio, K.; Huppert, D.; Remington, S. J. An Alternative Excited-State Proton Transfer Pathway in Green Fluorescent Protein Variant S205V. *Protein Sci. Publ. Protein Soc.* **2007**, *16*, 2703−2710.

(54) Nemukhin, A. V.; Grigorenko, B. L.; Savitsky, A. P. Computer Modeling of the Structure and Spectra of Fluorescent Proteins. *Acta Naturae* **2009**, *1*, 33−43.

(55) Lill, M. A.; Helms, V. Proton Shuttle in Green Fluorescent Protein Studied by Dynamic Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2778−2781.

(56) Patnaik, S. S.; Trohalaki, S.; Pachter, R. Molecular Modeling of Green Fluorescent Protein: Structural Effects of Chromophore Deprotonation. *Biopolymers* **2004**, *75*, 441−452.

(57) Helms, V.; Straatsma, T. P.; McCammon, J. A. Internal Dynamics of Green Fluorescent Protein. *J. Phys. Chem. B* **1999**, *103*, 3263−3269.

(58) Simkovitch, R.; Huppert, A.; Huppert, D.; Remington, S. J.; Miller, Y. Proton Transfer in Wild-Type GFP and S205V Mutant Is

Reduced by Conformational Changes of Residues in the Proton Wire. *J. Phys. Chem. B* **2013**, *117*, 11921−11931.

(59) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. Operation of the Proton Wire in Green Fluorescent Protein. A Quantum Dynamics Simulation. *J. Phys. Chem. B* **2008**, *112*, 5500−5511.

(60) Grigorenko, B. L.; Nemukhin, A. V.; Polyakov, I. V.; Morozov, D. I.; Krylov, A. I. First-Principles Characterization of the Energy Landscape and Optical Spectra of Green Fluorescent Protein along the A→I→B Proton Transfer Route. *J. Am. Chem. Soc.* **2013**, *135*, 11541−11549.

(61) Plapp, B. V. In *Isotope Effects In Chemistry and Biology*; Kohen, A., Limbach, H.-H.; Eds.; CRC Press: Boca Raton, FL, 2005; pp 811−835.

(62) Kovaleva, E. G.; Plapp, B. V. Deprotonation of the Horse Liver Alcohol Dehydrogenase-NAD+ Complex Controls Formation of the Ternary Complexes. *Biochemistry (Mosc.)* **2005**, *44*, 12797−12808.

(63) Klimacek, M.; Nidetzky, B. The Oxyanion Hole of *Pseudomonas Fluorescens* Mannitol 2-Dehydrogenase: A Novel Structural Motif for Electrostatic Stabilization in Alcohol Dehydrogenase Active Sites. *Biochem. J.* **2010**, *425*, 455−463.

(64) Klimacek, M.; Brunsteiner, M.; Nidetzky, B. Dynamic Mechanism of Proton Transfer in Mannitol 2-Dehydrogenase from Pseudomonas Fluorescens Mobile GLU292 Controls Proton Relay through a Water Channel That Connects the Active Site with Bulk Solvent. *J. Biol. Chem.* **2012**, *287*, 6655−6667.

(65) Zhang, W.; Hou, T.; Schafmeister, C.; Ross, W. S.; Case, D. A. *Leap and Sleap*, 2011.

(66) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(67) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(68) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

(69) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(70) *QUACPAC 1.6.3.1*; OpenEye Scientific Software, Santa Fe, NM, 2013.

(71) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(72) Bayly, C. I.; McKay, D.; Truchon, J.-F. An Informal AMBER Small Molecule Force Field: parm@Frosst. http://www.ccl.net/cca/data/parm_at_Frosst/parm_Frosst_note.pdf (accessed Nov 21, 2014).

(73) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(74) Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; Merz, K.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.; Seetin, M.; Sagui, C.; Babin, V.; Kollman, P. *Amber 11*; University of California, San Francisco, 2010.

(75) Cheatham, T.; Crowley, M.; Tsui, V.; Pitera, J.; Case, D.; Gohlke, H.; Tanner, S.; Absgarten, E.; Roe, D.; Frybarger, P.; Walker, R. *Ptraj and Cpptraj*; University of California, San Francisco, 2010.

(76) *The PyMOL Molecular Graphics System*, Version 1.3r1; Schrödinger; 2010.

(77) Holliday, G. L.; Mitchell, J. B. O.; Thornton, J. M. Understanding the Functional Roles of Amino Acid Residues in Enzyme Catalysis. *J. Mol. Biol.* **2009**, *390*, 560−577.

(78) Shinobu, A.; Palm, G. J.; Schierbeek, A. J.; Agmon, N. Visualizing Proton Antenna in a High-Resolution Green Fluorescent Protein Structure. *J. Am. Chem. Soc.* **2010**, *132*, 11093−11102.

(79) Leiderman, P.; Gepshtein, R.; Tsimberov, I.; Huppert, D. Effect of Temperature on Excited-State Proton Tunneling in Wt-Green Fluorescent Protein. *J. Phys. Chem. B* **2008**, *112*, 1232−1239.