J|A|C|S ARTICLES

Published on Web 05/25/2010

Visualizing Proton Antenna in a High-Resolution Green Fluorescent Protein Structure

Ai Shinobu,[†] Gottfried J. Palm,[*,‡] Abraham J. Schierbeek,[¶,§] and Noam Agmon[*,†]

The Fritz Haber Research Center, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, Institute for Chemistry and Biochemistry, Ernst-Moritz-Arndt-University, 17489 Greifswald, Germany, Bruker AXS B.V., Oostsingel 209, Delft NL-2612 HL, The Netherlands, and Rigaku Europe, Unit B6, Chaucer Business Park, Watery Lane, Kemsing, Sevenoaks, Kent TN15 6QY, England

Received February 14, 2010; E-mail: palm@uni-greifswald.de; agmon@fh.huji.ac.il

Abstract: "Proton-collecting antenna" are conjectured to consist of several carboxylates within hydrogen-bond (HB) networks on the surface of proteins, which funnel protons to the orifice of an internal proton wire leading to the protein's active site. Yet such constructions were never directly visualized. Here we report an X-ray structure of green fluorescent protein (GFP) of the highest resolution to date (0.9 Å). It allows the identification of some pivotal hydrogen atoms pertinent to uncertainties concerning the protonation state of the chromophore. Applying a computer algorithm for mapping proton wires in proteins reveals the previously observed "active site wire" connecting Glu222 with the surface carboxylate Glu5. In addition, it is now possible to identify what appears to be a proton-collecting apparatus of GFP. It consists of a negative surface patch containing carboxylates, threonines, and water molecules, connected by a HB network to Glu5. Furthermore, we detect exit points via Asn146 and His148 to a hydrophobic surface region. The more extensive HB network of the present structure, as compared with earlier GFP structures, is not accidental. A systematic investigation of over 100 mutants shows a clear correlation between the observed water content of GFP X-ray structures and their resolution. With increasing water content, the proton wires become progressively larger. These findings corroborate the scenario in which the photodissociated proton from wild-type GFP can leak outside, whereafter another proton is recruited via the proton-collecting apparatus reported herein.

Introduction

Most enzymes consume or produce protons during their catalytic activity. In particular, proton pumps are membrane proteins that transport protons across a membrane against its pH gradient.[1] Unlike electrons,[2] protons cannot migrate within a protein through either space or its carbon−carbon bonds. It is believed that their transport requires "proton wires" of closely spaced oxygen (nitrogen, sulfur) atoms.[3] For the wire to conduct, protons must arrive at an orifice on the surface of the protein where the wire begins. The probability that randomly moving particles in solution will hit a point on a surface is rather low. Thus, internal proton wires are conceivably complemented by a proton-collecting apparatus on the protein's surface that collects and funnels protons into the orifice via surface diffusion (Figure 1).

The idea that dimensionality reduction can enhance ligand binding to surface receptors has been discussed by several authors.[4−8] In biological proton transport, the concept of a "proton-collecting antenna" emerged in bioenergetics, in conjunction with membranal proteins such as bacterio-

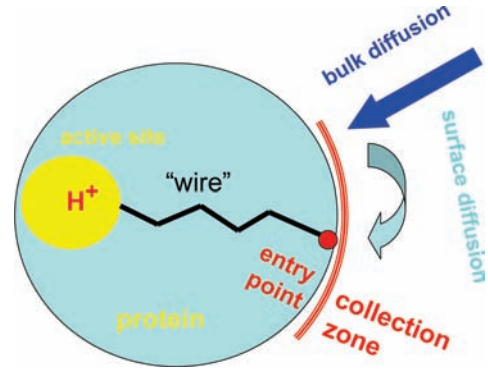

Figure 1. Schematic representation of the "central dogma of proton wires": A proton-collecting region on the surface of the protein (the proton antenna) attracts protons from the bulk and funnels them by surface diffusion to an orifice, where the internal wire begins. The latter connects with the active site of the protein, where protons are consumed.

[†] The Hebrew University of Jerusalem.
[‡] Ernst-Moritz-Arndt-University.
[¶] Bruker AXS.
[§] Rigaku Europe.

(1) Wraight, C. A. *Biochim. Biophys. Acta* **2006**, *1757*, 886–912.
(2) Gray, H. B.; Winkler, J. R. *Q. Rev. Biophys.* **2003**, *36*, 341–372.
(3) Nagle, J. F.; Morowitz, H. J. *Proc. Nat. Acad. Sci. U.S.A.* **1978**, *75*, 298–302.
(4) Bücher, T. *Adv. Enzymol.* **1953**, *14*, 1–48.
(5) Adam, G.; Delbrück, M. Reduction of Dimensionality in Biological Diffusion Processes. In *Structural Chemistry and Molecular Biology*; Rich, A., Davidson, N., Eds.; Freeman: San Francisco, 1968; pp 198−215.
(6) Berg, O. G. *Biophys. J.* **1985**, *47*, 1–14.
(7) Georgievskii, Y.; Medvedev, E. S.; Stuchebrukhov, A. A. *J. Chem. Phys.* **2002**, *116*, 1692–1699.

rhodopsin (BR),[9−12] cytochrome *c* oxidase (C*c*O),[13,14] and the bacterial reaction center.[14] In BR, two aspartates are identified near the orifice, whereas two histidines are there in C*c*O. These seem to connect with about three neighboring carboxylates, which were assigned the role of a proton-collecting antenna.

Taking BR as an example, its surface carboxylates, which reside on flexible loops connecting two α-helixes, cannot be identified in the X-ray structures. Their proximity was deduced from a multiparameter fit to kinetic data.[9−11] Surface water molecules are even more mobile, and therefore they can only be identified in high-resolution X-ray structures. In BR, only a portion of the internal proton wire could be traced. To date, there is no direct visualization of a complete hydrogen-bond (HB) network that may serve as the proton-collecting apparatus envisioned in Figure 1.

Soluble enzymes have a much larger surface area which is covered by hydrophilic carboxylate groups, potential components of "proton antennae".[15] Various mechanisms were suggested for conducting protons into enzymes: "proton shuttles" include a histidine ring flip in carbonic anhydrase (CA)[16] or the rotation of the carboxylate side chain of Asp15 in ferredoxin I, which transfers a proton between solution and a sulfur atom in the enzyme's active site.[17] The construction of the proposed proton wires is also quite diverse: from a water channel in succinate dehydrogenase[18] to a collection of two arginines and a glutamate in fumarate reductase.[19] Yet, in these cases also, an intact proton-collecting apparatus connected to these internal proton pathways was not visualized.

Some of us have recently developed a simple and fast computer code for the comprehensive mapping of proton wires within proteins and applied it to green fluorescent protein (GFP) and to CA.[20] The HB networks found in these proteins are much more extensive than those previously reported in the literature. In CA, there are surface wires extending from the active site to Glu69 and Asp72 on the rim of the "active site caldera", which may be part of a proton-collecting apparatus. The present work finds a considerably larger HB network on the surface of GFP, involving water molecules, carboxylates, and threonines, which may indeed function as a "proton antenna".

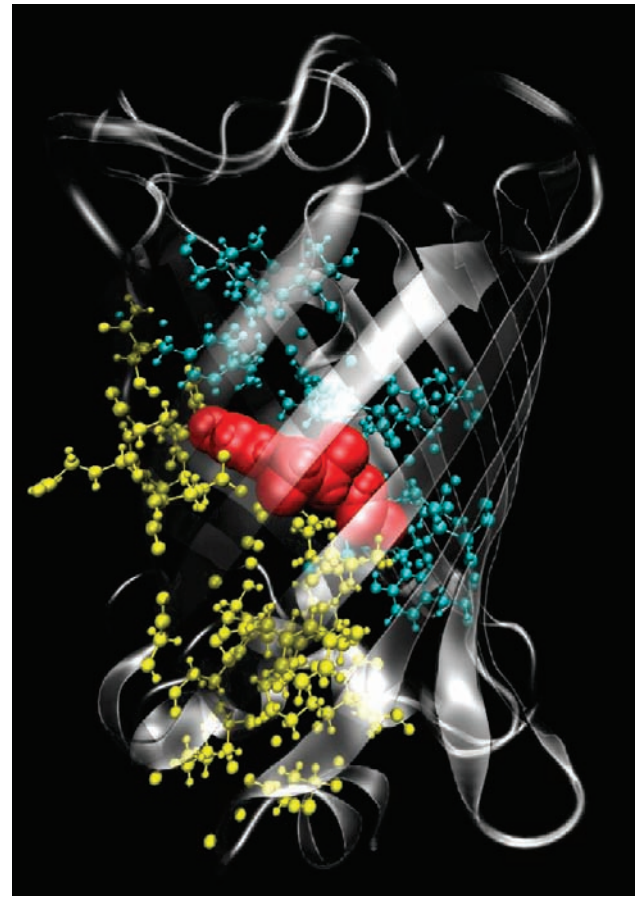

Figure 2. Overview of the two internal proton wires in wt-GFP (based on PDB code 2WUR). Transparent silver ribbons depict the backbone of the protein (β sheets and loops). The chromophore is shown in space-filling representation (red). Amino acid residues and water oxygens participating in the ASW are in yellow, whereas those participating in the biosynthesis wire are in cyan. Rendered using the VMD software.[37]

Within GFP, we have found two internal proton wires, located on both sides of the chromophore (Figure 2). One (in cyan) was suggested to play a role in the GFP chromophore biosynthesis.[20] The other, the "active-site wire" (ASW, yellow), extends from the chromophore (the hydroxyl of Tyr66) to Glu5 on the bottom of the GFP β-barrel structure.[21] When GFP is illuminated by UV light, the pK_a of the chromophore drops dramatically, leading to excited-state proton transfer (ESPT)[22] (see ref 23 for a review of ESPT). The conventional scenario is that the photodissociated proton migrates via Ser205 to Glu222 and resides there for the duration of the excited-state lifetime.[24,25] Then, after radiative decay of the chromophore, the proton returns to the photoanion. An alternate scenario[21,26] is that the proton can diffuse on the ASW past Glu222, occasionally leaking out to the external solution via the Thr203-His148 gateway, and then (likely already in the ground state) a new proton is recruited from solution via Glu5.

(8) Agmon, N. *Chem. Phys.* **2010**, *370*, 232−237.
(9) Nachliel, E.; Yaniv-Checover, S.; Gutman, M. *Solid State Ionics* **1995**, *97*, 75–82.
(10) Checover, S.; Nachliel, E.; Dencher, N. A.; Gutman, M. *Biochemistry* **1997**, *36*, 13919–13928.
(11) Checover, S.; Marantz, Y.; Nachliel, E.; Gutman, M.; Pfeiffer, M.; Tittor, J.; Oesterhelt, D.; Dencher, N. A. *Biochemistry* **2001**, *40*, 4281–4292.
(12) Riesle, J.; Oesterhelt, D.; Dencher, N. A.; Heberle, J. *Biochemistry* **1996**, *35*, 6635–6643.
(13) Marantz, Y.; Nachliel, E.; Aagaard, A.; Brzezinski, P.; Gutman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8590–8595.
(14) Ädelroth, P.; Brzezinski, P. *Biochim. Biophys. Acta* **2004**, *1655*, 102–115.
(15) Gutman, M.; Nachliel, E. *Biochim. Biophys. Acta* **1990**, *1015*, 391–414.
(16) Silverman, D. N.; McKenna, R. *Acc. Chem. Res.* **2007**, *40*, 669–675.
(17) Chen, K.; Hirst, J.; Camba, R.; Bonagura, C. A.; Stout, C. D.; Burgess, B. K.; Armstrong, F. A. *Nature* **2000**, *405*, 814–817.
(18) Cheng, V. W. T.; Johnson, A.; Rothery, R. A.; Weiner, J. H. *Biochemistry* **2008**, *47*, 9107–9116.
(19) Pankhurst, K. L.; Mowat, C. G.; Rothery, E. L.; Hudson, J. M.; Jones, A. K.; Miles, C. S.; Walkinshaw, M. D.; Armstrong, F. A.; Reid, G. A.; Chapman, S. K. *J. Biol. Chem.* **2006**, *281*, 20589–20597.
(20) Shinobu, A.; Agmon, N. *J. Phys. Chem. A* **2009**, *113*, 7253–7266.
(21) Agmon, N. *Biophys. J.* **2005**, *88*, 2452–2461.
(22) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8362–8367.
(23) Agmon, N. *J. Phys. Chem. A* **2005**, *109*, 13–35.
(24) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2306–2311.
(25) Palm, G. J.; Zdanov, A.; Gaitanaris, G. A.; Stauber, R.; Pavlakis, G. N.; Wlodawer, A. *Nat. Struct. Biol.* **1997**, *4*, 361–365.
(26) Agmon, N. *J. Phys. Chem. B* **2007**, *111*, 7870–7878.

Possible experimental support for this mechanism was found in transient fluorescence from the acidic form of the chromophore.[26−28] If the proton were only transferred from Tyr66 to Glu222, the decay of the reactant (acidic) form of the chromophore would be exponential. Instead, the lifetime-corrected data exhibit a power-law tail in the nanosecond time scale. The $t^{-1/2}$ (lifetime corrected) fluorescence decay at low temperatures was taken as evidence for one-dimensional diffusion of the proton along an internalized proton wire, whereas its switchover to $t^{-3/2}$ at higher temperatures was ascribed to an activated process of proton leakage out of the protein, triggered by the side-chain rotation of Thr203. It was assumed that a new proton replenishes this loss via the ASW, yet no connection of Glu5 to a surface "antenna" has been observable in previously available X-ray structures of wild-type (wt) GFP.[24,29,30]

The present work is devoted to a detailed study of the ASW in GFP. We have examined over 100 GFP mutants with our program, finding that with increasing X-ray resolution the number of water molecules identified in the structure increases, along with the size of this cluster. HB clusters that are disjoint at lower resolution can merge into a single cluster at higher resolution.

Consequently, we have solved and examined the highest resolution X-ray structure of any GFP mutant available thus far, obtained herein for the wt-like mutant sg11 (0.9 Å resolution, PDB code 2WUR). In this structure one can see some hydrogen atoms which suggest, for example, an unprotonated nitrogen in the imidazolinone ring of the chromophore. Furthermore, we observe that the ASW exhibits an additional possible proton exit point to the proposed[21] rotation of Thr203 at the backbone carbonyl of Asn146. Perhaps more significant is the observation that the ASW continues beyond Glu5 on the GFP surface. We will show below that this continuation has the expected characteristics of a proton-collecting apparatus ("proton antenna") and thus may constitute the first example in which the construction in Figure 1 is visualized directly from X-ray data.

Methods

Purification and Crystallization of GFP. GFP variant sg11, which has the mutations F64L, I167T, and K238N (in addition to the Q80R mutation that appears in all GFP structures), was purified as published.[25] Briefly, *Escherichia coli* BL21(DE3) cells harboring the plasmid pTFred11, a pET21a vector with an insert for GFP sg11, were grown in LB medium supplemented with ampicillin (100 mg/L). After induction with isopropyl-β-D-thiogalactopyranoside (IPTG), the temperature was lowered from 37 to 22 °C, and cells were grown overnight. Cell extract prepared by French press and centrifugation was purified by anion-exchange chromatography on Q-sepharose (GE Healthcare, Freiburg, Germany) and gel filtration on Superdex 200 (GE Healthcare, Freiburg, Germany). After partial digestion with subtilisin (1:100) in 200 mM ammonium sulfate and 50 mM imidazole pH 7.5, the C-terminally truncated protein (after His231) was purified by anion chromatography with PorosHQ (Applied Biosystems, Darmstadt, Germany). For crystallization GFP was dialyzed against 20 mM Tris pH 8.0 and concentrated to 10 mg/mL. Using the hanging drop vapor diffusion method (drops with 2 μL of protein solution + 2 μL of well solution, 0.5 mL of well solution), crystals were obtained under numerous conditions with organic solvents (2-methyl-2,4-pentanediol, isopropanol, ethanol, dioxane, polyethylene glycol). The best crystal (900 × 500 × 200 μm) grew from 40% ethanol, 10% dioxane. The crystal was directly cryocooled in a 100 K nitrogen stream and subsequently annealed by blocking the nitrogen stream for 5 s.

Data Collection and Structure Determination. Data were collected on a Microstar microfocus rotating anode generator (Montel multilayer graded optics with Cu Kα radiation) with a 135 mm SMART6000 CCD detector mounted on a MACH3 four-circle goniostat (Bruker AXS, Delft, The Netherlands). To complete a 0.9 Å data set, 8 low- and 21 high-resolution scans were run over ca. 100 h. Data were collected using the program Proteum2 from Bruker AXS[31] and processed with the program SAINT.[31] Scaling and absorption correction was done using the program SADABS.[31] Data collection statistics are given in Table S1 of the Supporting Information (SI). The structure was solved by sulfur-SAD phasing as well as by molecular replacement. Final refinement was done with the program SHELXL-97,[32] with anisotropic *B*-factors for all non-hydrogen atoms. Riding hydrogen atoms were included for all residues except for water molecules. The protonation state of deprotonable groups was set on the basis of their calculated pK_a value (PDB2PQR server http://www.poissonboltzmann.org/pdb2pqr/).[33,34] Because the pH was 8.0 in the crystal, occupancies of acidic protons were set to 0.0 for pK_a < 7.0, 0.2 for 7.0 < pK_a < 7.5, 0.5 for 7.5 < pK_a < 8.5, 0.8 for 8.5 < pK_a < 9.0, and 1.0 for 9.0 < pK_a. No significant changes of the protein chain are visible in the model compared to published structures of F64L/I167T mutants (PDB codes: 1EMC, 1EME, and 1EML). A total of 320 water molecules were modeled, compared to 105 at 2.30 Å resolution (1EML) and 74 at 2.50 Å resolution (1EME). This is consistent with the increase in water content with increasing resolution,[35] as reported below. The resulting structure was deposited in the Protein Data Bank (PDB) under the code 2WUR. Note, however, that numbering of the water molecules in the PDB entry is shifted by 2000 compared to this article.

A hydrogen omit map was calculated to visualize possible hydrogen positions. To this end, all hydrogens were omitted from the final model, and maps were calculated after 10 least-squares refinement cycles in SHELXL-97.[32]

Computer Algorithm for Proton Wire Search. We have developed a general algorithm for identifying and mapping proton wires within proteins.[20] Its goal is to divide all oxygen atoms in the structure (backbone, side chain, and water oxygens) into exclusive hydrogen-bonded clusters as shown in Figure 3. Only clusters with more than three atoms are considered. Within a cluster, there is a HB pathway connecting each atom to every other atom, but between atoms belonging to different clusters there is no such path. In a second step, we add to each cluster all the nitrogen atoms which are hydrogen-bonded to oxygen atoms already in the cluster (cutoff 3.0 Å). HBs between backbone nitrogens and oxygens are not considered. Presently, we do not look for longer pathways containing nitrogens, assuming that conduction through nitrogen atoms is less efficient than through oxygens.

In practice, the algorithm involves identifying hydrogen-bonded pairs (based on distance and angle criteria), followed by a tree search to construct the clusters. Subsequently, we determine the surface

(27) Leiderman, P.; Huppert, D.; Agmon, N. *Biophys. J.* **2006**, *90*, 1009–1018.
(28) Agmon, N. *Chem. Phys. Lett.* **2006**, *417*, 530–534.
(29) Yang, F.; Moss, L. G.; Phillips, G. N., Jr. *Nat. Biotechnol.* **1996**, *14*, 1246–1251.
(30) van Thor, J. J.; Georgiev, G. Y.; Towrie, M.; Sage, J. T. *J. Biol. Chem.* **2005**, *280*, 33652–33659.
(31) Bruker AXS, http://www.bruker-axs.com.
(32) Sheldrick, G. M. *Acta Crystallogr. A* **2008**, *64*, 112–122.
(33) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, W665–W667.
(34) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, W522–W525.
(35) Carugo, O.; Bordo, D. *Acta Crystallogr. D* **1999**, *55*, 479–483.
(36) Hubbard, S. J.; Thornton, J. M. *NACCESS* Computer Program; Department of Biochemistry and Molecular Biology, University College London, 1993.
(37) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

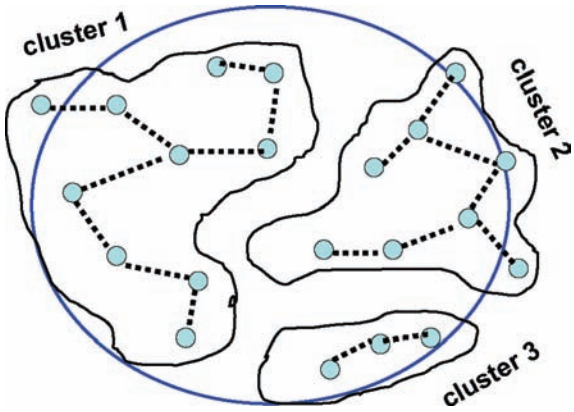

Figure 3. Schematic depiction of the partition of atoms into mutually exclusive hydrogen-bonded clusters. Small clusters with less than three HBs (such as cluster 3 here) are ignored.

accessibility of each atom on the cluster by running NAccess[36] and use VMD[37] to visualize it. Additional details are given in ref 20.

Results

Hydrogen Electron Densities in the Chromophore Region. We have determined the structure of GFP sg11 to 0.90 Å resolution. While this protein has the mutations F64L, I167T, and K238N (in addition to the usual Q80R mutation), its two absorption peaks at 395 and 468 nm (of the protonated A-state and deprotonated B-state, respectively), with estimated ratio of 2:1 (in 100 mM Tris, pH 8.0), and emission peak at 506 nm are within 5 nm from those of the wt.[25] Hence, one expects GFP sg11 to be similar in structure to wt-GFP. In addition, both mutations are on the opposite face of the chromophore with respect to the ASW and are thus not expected to interfere with its study.

The high resolution is manifested in the hydrogen omit electron density maps, Figure 4 and Figure 5, in which hydrogen atoms begin to be visible (although the hydrogens could not be independently modeled). Denoting, as customary, the observed Fourier transform map by F_o and the calculated one by F_c, the $2F_o - F_c$ hydrogen omit map (blue) should cover the model, if correct. The superimposed $F_o - F_c$ hydrogen omit map (green) shows experimental electron density missing from the calculated model. Since hydrogens were omitted, the green density is either the missing hydrogens or noise. The figures show that most of the carbon-bound protons and some of the acidic protons become visible in this X-ray diffraction experiment.

Focusing first on the chromophore and its vicinity, Figure 4A shows that, at the expected position of an aromatic proton at the $N_{\varepsilon 2}$ position in the imidazolinone ring, no density is visible (even at 0σ). This supports the assertion that the 5-ring is deprotonated (hence neutral) in its ground electronic state, as proposed in ref 38. Previously, Voityuk et al.[39] have favored a protonated imidazolinone in their calculations, which becomes zwitterionic after the ESPT reaction. Indeed, the absorbance peak of GFP at 398 nm matches that of the isolated *cationic* chromophore in solution, whereas its neutral form in solution absorbs around 370 nm.[40] Because our X-ray data provide direct experimental indication for a neutral imidazolinone, the absorption peak of A-state GFP must be red-shifted by various interactions within the protein, most notably with Arg96 and Gln94, which were calculated to produce a 30−40 nm red shift.[41]

The hydrogen on Tyr66-C_β is visible, as well as the four aromatic hydrogens on the phenol ring, but the phenolic proton, Try66-O_ηH, is barely visible and thus does not point persistently toward the adjacent water molecule (W285), the expected HB acceptor (see Figure 4B). Another observation is the orientation of the hydroxyl of Thr203, which forms a HB with Try66-O_ηH. Usually, this is characteristic of the B-state GFP,[24] whereas in the A-state the side chain of Thr203 is rotated away from the chromophore. This correlates to the absorption spectrum of 0.1 mg/mL GFP sg11 in 40% ethanol (similar to the crystallization solution) showing only the B-peak (data not shown), in contrast to that in aqueous buffer. How far the equilibrium is shifted back toward the A-state due to the high protein concentration in the crystal is unclear. The observed orientation of Thr203 may open an exit pathway for the proton via the backbone carbonyl of His148.[21,26]

Water 285, in turn, exhibits a marked hydrogen density along the HB connecting it to the backbone carbonyl of Asn146 (Figure 4B). However, its second hydrogen appears to be disordered between two HBs: one to Tyr66-O_ηH and the other to Ser205-O_γH. Thus, this water might be swinging around its HB with Asn146 between two sites, explaining the rather low X-ray signal of its second hydrogen atom. In general, we note that the proton wire connecting Tyr66 and Glu222 does not display well-defined HB directions, either because of the mixture of the A- and B-states in our crystal or because the hydrogens within this wire segment are disordered due to rotations around all of the relevant RC−OH dihedral angles.

In contrast, a well-defined HB donated from Thr62-$O_{\gamma 2}$ to His181-$N_{\varepsilon 2}$ (which is clearly deprotonated) is seen in Figure 5. This HB is part of the "biosynthesis wire" (cyan in Figure 2) in the vicinity of the chromophore. This figure also shows several carbon-bound protons, such as the three protons at Thr63-$C_{\gamma 2}$ and one (out of three) at Thr62-$C_{\gamma 2}$. Thus, the methyl group of Thr63 is probably not freely rotating within the GFP.

Cluster Statistics in GFP. We ran the proton-wire program on GFP sg11, noting that the hydrogen-bonded clusters appear to be significantly larger than those identified[21] in previous wt structures[24] of a lower X-ray resolution. In order to check whether this is an expected effect of the increased resolution obtained for GFP sg11 in the present work, we have conducted an extensive statistical analysis on 103 additional GFP mutant structures from the PDB, some of which consist of several subunits. These structures serve as a "GFP library" against which to compare the new results. The PDB codes of all structures in the library and their various statistical attributes are listed in Table S2 of the SI.

wt-GFP consists of 238 amino acids. Of these, a trailing sequence of about 8 amino acids is typically deleted or not visible in the density, so that the entries contain ca. 230 amino acids consisting on average of 1796 atoms (excluding hydrogens). From these, we find that *on average* 193 atoms participate in 22 clusters (full statistics available in Table S3 of the SI). These numbers tend to increase at higher resolution. The effect is not due to unresolved amino acids, because GFP is a rigid

(38) Weber, W.; Helms, V.; McCammon, J. A.; Langhoff, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6177–6182.
(39) Voityuk, A. A.; Michel-Beyerle, M.-E.; Rösch, N. *Chem. Phys.* **1998**, *231*, 13–25.
(40) Dong, J.; Solntsev, K. M.; Tolbert, L. M. *J. Am. Chem. Soc.* **2006**, *128*, 12038–12039.
(41) Laino, T.; Nifosì, R.; Tozzini, V. *Chem. Phys.* **2004**, *298*, 17–28.

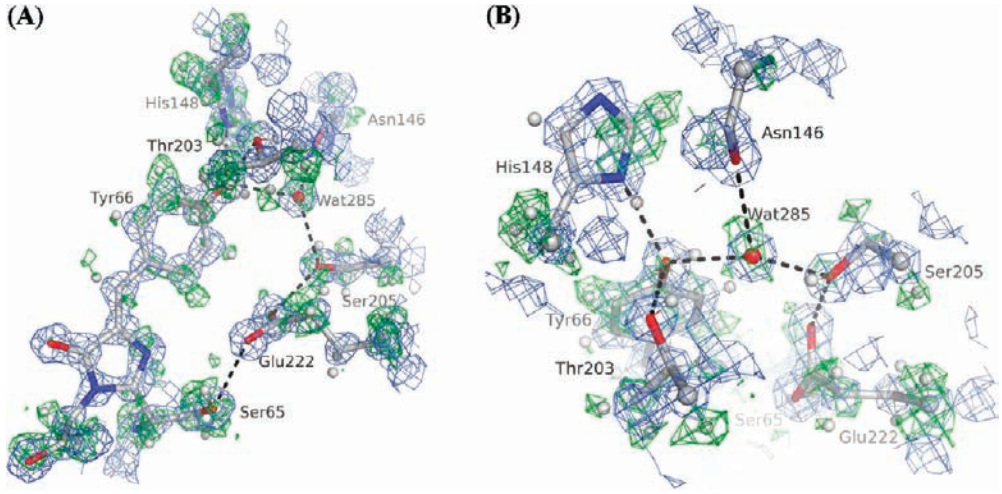

Figure 4. Hydrogen omit electron density maps of the chromophore region (residues 65, 66, 146, 148, 203, 205, and 222, and water molecule 285). (A) Side view with all of the chromophore atoms and the HBs from W285 to Ser65 clearly visible. (B) Top view of the chromophore. Three HBs of the phenolic oxygen of the chromophore and of W285 are depicted. HBs are drawn as dashed lines between the non-hydrogen atoms. Contour levels are 2.0σ for both the $F_o - F_c$ map (in green) and the $2F_o - F_c$ map (in blue). Hydrogen atoms were included in the final model for all C−H groups and for less acidic O−H and N−H groups ($pK_a > 7.0$). They are indicated by thin sticks and gray balls. C_α atoms are indicated by larger spheres. Figure prepared with PyMOL.[51]

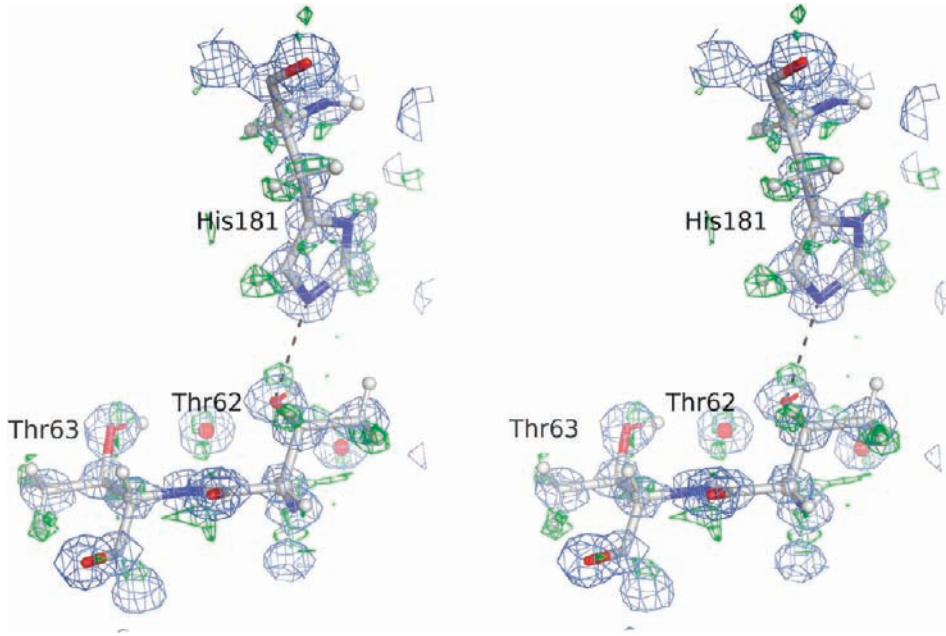

Figure 5. Stereoscopic view of the hydrogen omit electron density map of a well-ordered region within GFP (part of the biosynthesis wire with Thr62, a proposed catalytically active residue). The map was calculated after refinement without hydrogens in the model. The difference electron density ($F_o - F_c$ map shown in green at 2.2σ) indicates mostly the positions of hydrogens (besides noise). The $2F_o - F_c$ map (blue) is contoured at 2.0σ. Figure prepared with PyMOL.[51]

protein and all its atoms are accurately determined (except for hydrogens and a few N- and C-terminal residues). Rather it stems from an increasing number of water molecules observed at higher resolution. These interconnect more amino acid residues, leading to larger clusters.

Figure 6A depicts the fraction of water molecules in GFP X-ray structures, $f_w \equiv N_w/(N_a + N_w)$, where N_w is the number of water oxygens and N_a the number of observable protein atoms (including carbon atoms, but excluding hydrogens) in the X-ray structure, as a function of the X-ray resolution, d. The positive correlation observed in this figure is similar to the correlation noted in previous work.[35] The higher the resolution (the smaller d), the more water molecules are observed. This is due to the lability of water molecules, which leads to smeared electron density that requires higher resolution to detect. The numbers of oxygen and nitrogen protein atoms and water oxygen atoms participating in clusters (of at least three atoms), n_a and n_w, respectively, increase with the water content, see Table 1. It can be seen there that the correlation between cluster size and water content is better than that between cluster size and resolution, supporting our assessment that the cluster size depends directly on water content.

The number of clusters also increases with increasing f_w. Most of the observed clusters are small and on the surface of the protein. Two large internal clusters can be identified on both

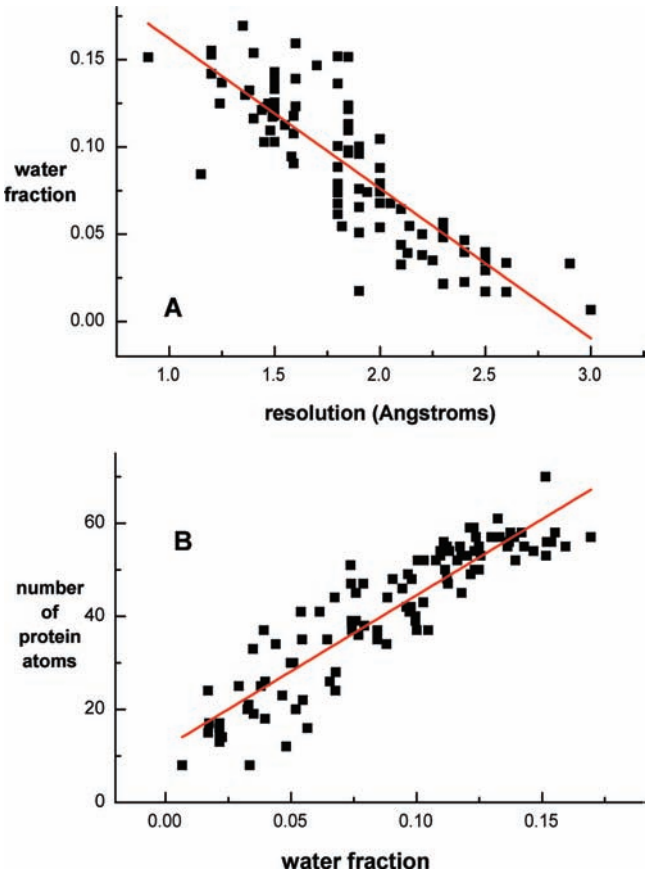

Figure 6. Correlations within the GFP library in Table S2 of the SI. (A) Dependence of the fraction of water molecules in GFP structures, f_w, on their X-ray resolutions, d. Red line is best linear fit, obeying $f_w = 0.248 - 0.0859d$, with a correlation coefficient of -0.82. Alternately, we find $N_w/N_a = 0.290 - 0.102d$, as compared with $0.301 - 0.095d$ in ref 35. These are identical within statistical error limits. (B) Number (n_a) of protein atoms from the GFP sg11-based reference list (Table S4 of the SI) appearing in clusters in various GFP mutants (Table S2), as a function of the corresponding water fraction in these X-ray structures, f_w. The red line is the best linear correlation, obeying $n_a = 11.9 + 327f_w$, with a correlation coefficient of 0.91.

Table 1. Statistical Properties of HB Clusters in GFP Mutants[a]

	atoms included				
	$\langle n_\alpha \rangle$	A	B	CC	|CC′|
n_a summed over all clusters	193	9.82	2054	0.96	0.83
n_w summed over all clusters	135	−42.7	1988	0.98	0.80
$n_{a+w} \equiv n_a + n_w$ in largest cluster	63	−16.3	892	0.76	0.66

[a] For the 104 GFP structures we have plotted the number of atoms of a given type, n_α (α = a, w, or a+w), that participate in whatever cluster as a function of the water fraction, f_w, for the corresponding protein. The fitting results to the linear correlation $n_\alpha = A + Bf_w$ are reported herein. $\langle n_\alpha \rangle$ is the average number of atoms in the cluster (averaged over the 104 structures in our GFP library). CC is the correlation coefficient, whereas |CC′| is the absolute value of the correlation coefficient when correlated with resolution instead of the water content.

sides of the chromophore (Figure 2): the ASW and the "biosynthesis wire".

Active-Site Wire. The ASW, which connects to the hydroxyl moiety of Tyr66 (part of the GFP chromophore), is the main subject of the present study. In GFP sg11, it contains 147 atoms, of which 75 are (non-hydrogen) protein atoms and 72 are water oxygens. Four of the ASW protein atoms which are less common in the other X-ray structures were omitted, and the remaining 71 protein atoms and 72 water oxygens are depicted in Figure 7. Figure 7A shows a three-dimensional projection, whereas Figure 7B provides a schematic depiction of it, in which it is easier to identify all the participating atoms. It can be seen that the ASW is composed of three parts: (i) a central internal part (cyan HBs), which is buried within the protein (from Asn146, via Tyr66, Glu222, Ser72 to Glu5, on the bottom of the β-barrel) [the residues participating in this segment are denoted in yellow in Figure 2]; (ii) two adjacent surface clusters beyond the carbonyl moieties of Asn146 and His148 (orange HBs), which we term "exit clusters"; and (iii) a surface cluster beyond Glu5 (green HBs), which we term the "entrance cluster".

The ASW of GFP sg11 is considerably larger than what was previously observed in the lower resolution structures of wt-GFP (1GFL, 1.9 Å resolution;[29] 1EMB, 2.1 Å resolution[24]). In order to check whether this is indeed due to the increase in water content for higher resolutions, we have conducted a comparative analysis of the ASW size. The 71 (non-hydrogen) protein atoms in the ASW of GFP sg11 were taken as a reference list (cf. Table S4 of the SI). For each structure in the GFP library, we counted the number of protein atoms from the reference list that participate in HB clusters of that structure. Figure 6B shows that this number is positively correlated with the water content. A similar analysis for three individual ASW segments is shown in Figure S2 of the SI, with linear regression parameters in Table S5 there. The GFP sg11 structure falls on this trend, which argues that its more extensive ASW is indeed an expected result of its higher resolution.

Figure 8 depicts the two exit segments of the GFP sg11 ASW. Indeed, this face of the protein is partly hydrophobic and nearly devoid of carboxylates, and thus not likely to be functioning in proton capture. Once the proton crosses the backbone carbonyl of Asn146 to W216, it is essentially located in bulk water. We find that about 60% of the structures in our GFP library exhibit this connectivity. Consequently, it appears that the photodissociated proton may leak out via the Asn146 pathway, but not enter. A similar conclusion was previously reached for the Thr203-His148 exit.[21] The figure shows that if the proton can cross the backbone carbonyl of His148 to W219, it is again in bulk solution.

Figure 9 shows the internal portion of the wire, which is quite similar to the depiction in ref 21 using lower resolution wt-GFP structures.[24,29] Within the internal segment, there are two water molecules that separate Ser72-O_γ from Glu222-$O_{\varepsilon 1}$, W108/109 and W307/308 (each of these two water molecules have two nearly equally probable positions in the 2WUR structure). In the schematic depiction of Figure 10, they are denoted by W(α) and W(β), respectively. An additional water molecule on a bypass loop to Ser65-O_γ, W104 of 2WUR in Figure 9, is denoted W(γ) in Figure 10. In some of the structures it is found hydrogen-bonded to W(β) rather than to W(α). Two additional water molecules, W(δ) and W(ε), help solvate W(α) and W(β), respectively. Therefore, W(α) or W(β) may be able to solvate a hydronium,[26] as found computationally in CcO.[42] Consequently, more than a single proton may participate in the excited-state dynamics of GFP.[26]

The water molecules near Ser72 are not always arranged to yield a continuous hydrogen-bonded pathway between Glu5 and Glu222. In 37 of the structures in the GFP library, all three waters exist and there is a continuous HB pathway between Glu222 (or, at least, Ser65) and Ser72 (entries with green

(42) Xu, J.; Sharpe, M. A.; Qin, L.; Ferguson-Miller, S.; Voth, G. A. *J. Am. Chem. Soc.* **2007**, *129*, 2910–2913.

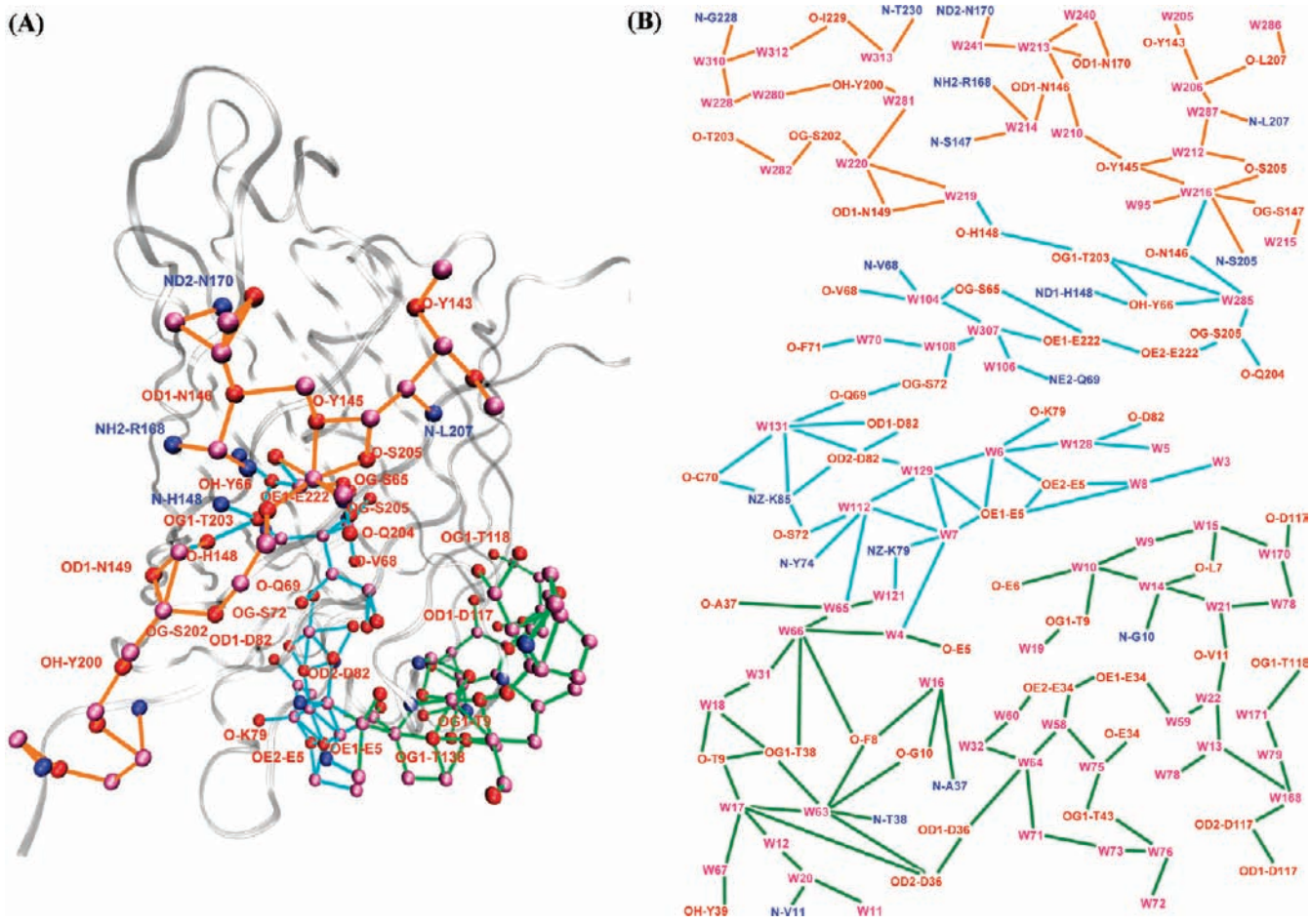

Figure 7. Active-site wire of the GFP sg11 mutant. HBs are depicted by colored lines corresponding to their subcluster: entrance (green), internal (cyan), and exit (orange). Atoms participating in the wire are rendered as colored spheres: red, protein oxygen; pink, water oxygen; blue, protein nitrogen. (A) Output from our program as displayed by a three-dimensional projection from VMD,[37] with GFP backbone in silver. (B) Schematic depiction of the connectivity in this cluster.

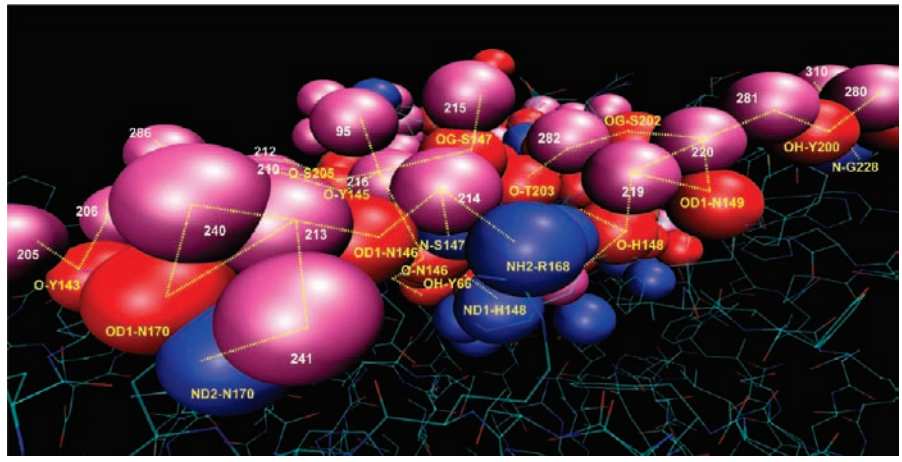

Figure 8. Segments of the ASW in GFP sg11 that are external to the backbone carbonyls of Asn146 and His148 (space-filling representation, attached to wire-frame representation of the remainder of the protein). Color scheme: red, protein oxygens; pink, water oxygens (with white numbers); blue, nitrogens. HBs are depicted in dotted yellow lines. Rendered using the VMD software.[37]

background in Table S2). In 67 structures, however, one or more of these water molecules are missing, so there is a discontinuity near Ser72. We group the 104 structures into four types with regard to these water molecules:

• Type A: W(α), W(β), and W(γ) all exist.

• Type B: Only W(α) is present. This case is rare, occurring only in five structures (2DUE, 1Z1P, 2DUH, 2G3D, and 2H6V).

• Type C: W(α) is missing. Either W(β) or W(γ), or both are present.

• Type D: None of the waters exist.

Table 2 characterizes these four types. Type A structures are on average of higher resolution than types C and D. Thus, high resolution may be a necessary condition for observing a continuous pathway past Ser72, but not a sufficient one,

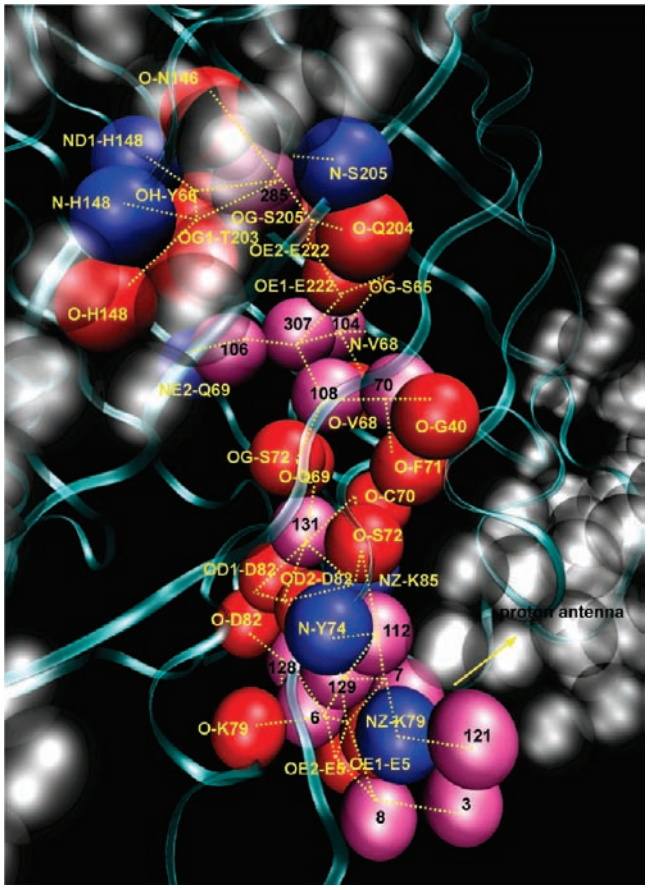

Figure 9. Internal part of the ASW in GFP sg11. Same color scheme as in Figure 8. Water numbers in black. Rendered using the VMD software.[37]

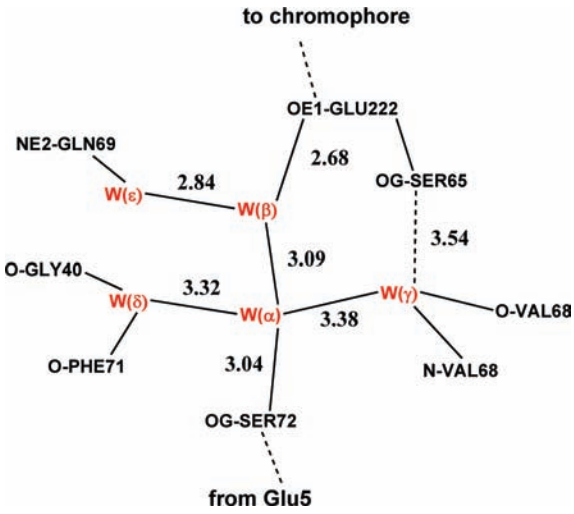

Figure 10. Schematic depiction of the five-water-molecule cluster (in red) separating Glu222 and Ser72. Numbers indicate average HB length (in Å) over all the structures which have a continuous pathway at Ser72.

because several high-resolution structures belong to the other three types.

To investigate this issue further, we present the ASW fragments from two structures of decarboxylated wt-GFP[30] in Figure S3 of the SI. The crystal of wt-GFP was decarboxylated by short-wavelength UV radiation at 100 K (PDB structure 1W7T) and then annealed at 170 K (PDB structure 1W7U). Figure S3 clearly shows that, while in structure 1W7T this water

Table 2. Properties of the Four Types of Structures with Regard to the Three-Water Connection between Glu222 and Ser72[a]

type	number	⟨resolution⟩ (Å)	$\langle f_w \rangle$
A	37	1.67	0.11
B	5	1.45	0.13
C	37	1.96	0.075
D	21	2.05	0.068

[a] Excluding the 1W7T and 1W7U tetramers, which do not have Glu222.

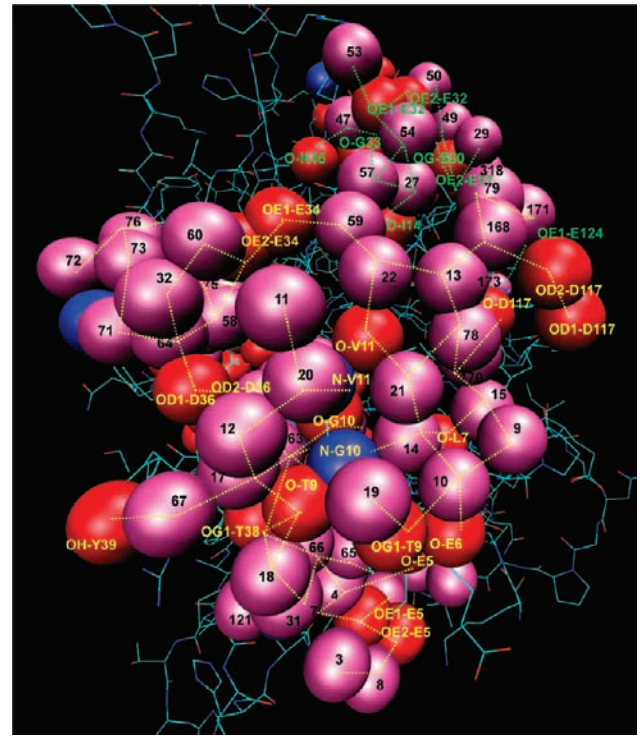

Figure 11. Presumed proton-collecting antenna of GFP, connected to Glu5, as visualized in the high-resolution structure of the GFP sg11 mutant. Same color scheme as in Figure 9. Protein atoms and HBs of the antenna are labeled in yellow, and protein atoms and HBs of the antenna extension are labeled in green. Rendered using the VMD software.[37]

cluster is absent, in the annealed structure (1W7U) three water molecules connect Ser72-O_γ with Ser65-O_γ (whose connection via Glu222 to the active site is now missing due to the decarboxylation of Glu222). This suggests that in some of the crystals these water molecules were trapped in local minima, so that annealing may provide a preferable strategy for investigating proton wires in their global minimum conformation. As noted in the Methods section, our sg11 crystal has been annealed by blocking the nitrogen stream for 5 s.

Proton-Collecting Antenna. Figure 11 shows the ASW segment beyond Glu5, on the surface of the protein. This proton-wire segment was not previously mentioned in the literature. Of the 104 structures in our GFP library, there are only 18 with a complete internal wire (from Tyr66 to Glu5), which is connected to (some fragment of) an entrance cluster at Glu5. As suggested earlier for proton-collecting antennae,[15] this subcluster is composed of neighboring carboxylates (Glu5, Asp36, Glu34, Asp117, and perhaps Glu6). In GFP, these carboxylates are intercalated by threonines (numbers 9, 38, 43, and 118). At the current resolution, a sufficient number of surface-bound water molecules is observed to allow one to visualize the connectivity between these residues. We also show in the figure a close-by surface cluster, which may connect to

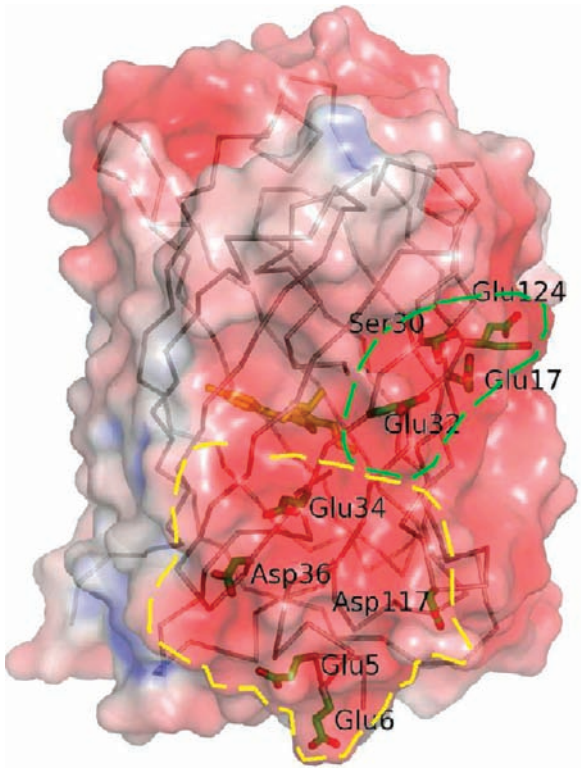

Figure 12. Electrostatic potential at the solvent-accessible surface of the sg11 GFP mutant obtained from the Poisson equation solver APBS[52] and displayed in PyMOL.[51] The range is + 1 k_BT/e (blue, positively charged) to −5 k_BT/e (red, negative), thereby emphasizing positively charged areas. Some of the carboxylic residues of the conjectured proton antenna are marked in black. The dashed lines encircle the surface areas that contain the HBs of the entrance cluster (yellow) and its extension (green), cf. Figure 11. It is seen that the protein surface is the most negative in the region of these clusters.

the entrance cluster via a few unresolved water molecules (e.g., by forming a water bridge between Glu34 and Glu32). It contains three more carboxylates in addition to the four that are already in the entrance cluster.

Figure 12 shows the electrostatic surface of the protein. The face on which the entrance segment resides is negative (colored red), due to the predominance of carboxylates. This further corroborates the identification of this region as a proton collection zone. In contrast, the opposite face of the protein (not shown here) is either positive or neutral, and thus would not attract positive ions from solution.

Experimental Implications. It is interesting to consider some possible experimental implications for the pathway model suggested by our X-ray data. In this model, the ASW is an entrance pathway because it emanates from the tentative proton antenna structure. Thus, while Glu222 acts as a proton acceptor in fast ESPT,[24,25] at longer times (e.g., in the ground state) it is actually a proton donor.[21] In addition, there are two short exit pathways in this model, one via Thr203-His148 and the other via Asn146. A test of this idea could involve iso-voluminous point mutations at critical sites along these pathways.

There is abundant mutation work on GFP, most of it geared toward obtaining new colors, improved fluorescence quantum yields, pH sensitivity, and the like. For example, mutations at the key positions 65, 203, and 222 were considered by Jung et al.[43] The absorption spectra of these mutants show a behavior which is not easily explainable.[44] While wt-GFP has mostly A-state absorption (ROH, 397 nm) and only a small fraction of B-state absorption (RO^-, 476 nm), the mutations E222Q and T65G exhibit exclusively B-state absorption, whereas T203V becomes pure A-state. The double mutants T65G/T203V and T203V/E222Q show B-state absorption.

As pointed out earlier,[21] this can be naively explained by our pathway model. Disconnecting the ASW at E222 or Thr65 does not enable the replenishment of protons following proton escape via one of the two exit pathways, making the chromophore anionic. Conversely, elimination of the Thr203-His148 exit pathway by the T203V mutation hampers proton exit, rendering the chromophore neutral. In the double mutants, the proton may escape via the second, Asn146 pathway, but another proton cannot be recruited if the ASW is disconnected.

There are several problems with this interpretation. First, the X-ray structures of these mutants are unknown (except, sometimes, in combination with other mutations), so it is not clear that their sole effect is in disconnecting one pathway or the other. Often, HBs make alternate pathways in response to a given mutation. For example, the substitution S205V produces a GFP variant that still exhibits the B-state green fluorescence, apparently due to an alternate pathway formed between Tyr66 and Glu222 via Thr203.[45]

Nevertheless, in some B-state GFPs the structure is known and supports the notion of a bisected ASW. For example, S65T GFP is the most widely used anionic GFP, and the X-ray structure shows that the side chain of Glu222 is rotated to form a HB with Thr65, thereby eliminating the connection with Ser205.[24] In decarboxylated Glu222 wt-GFP,[30] the connection between Ser205 and Ser65 is severed, as can be seen in Figure S3 of the SI, and it also exhibits B-state absorption (Figure 3B in ref 30).

Yet these observations are not unambiguous, because the considered mutations are very close to the chromophore, so that the altered interactions with it may be responsible for the observed change. A more interesting test of the pathway model might therefore involve a mutation in a more distant site. For example, if Ser72 is indeed a bottleneck along the ASW, it may be interesting to check the effect of the S72A mutation. It has thus far been implemented only in conjunction with other mutations in the vicinity of the chromophore (e.g., ref 46). One would then need to determine the X-ray structure of this mutant, preferably following annealing, to verify that the nearby water molecules do not form a new HB pathway bypassing Ala72. Subsequently, steady-state and time-resolved spectra (absorption, fluorescence, and IR) would need to be acquired, investigating the effect on nanosecond ESPT kinetics (e.g., ref 27) and microsecond ground-state pH jump kinetics (e.g., ref 44).

Conclusion

This work reports a GFP structure of the highest available resolution (0.90 Å), a computerized routine for mapping proton wires, and statistical analysis of the influence of the X-ray

(43) Jung, G.; Wiehlery, J.; Zumbusch, A. *Biophys. J.* **2005**, *88*, 1932–1947.
(44) Bizzarri, R.; Nifosì, R.; Abbruzzetti, S.; Rocchia, W.; Guidi, S.; Arosio, D.; Garau, G.; Campanini, B.; Grandi, E.; Ricci, F.; Viappiani, C.; Beltram, F. *Biochemistry* **2007**, *46*, 5494–5504.
(45) Shu, X.; Leiderman, P.; Gepshtein, R.; Smith, N. R.; Kallio, K.; Huppert, D.; Remington, S. J. *Protein Sci.* **2007**, *16*, 2703–2710.
(46) Cormack, B. P.; Valdivia, R. H.; Falkow, S. *Gene* **1996**, *173*, 33–38.

resolution on the observed hydrogen-bonded clusters within this protein. Starting with the chromophore, our data provide direct evidence that it is anionic (and not zwitterionic) in the B-state, possibly settling a long dispute over this question.[38−40]

Most of our work focused on the ASW that connects to the chromophore at Tyr66-O_η. The most frequently discussed segment of the ASW extends up to Glu222,[24,25] which is the initial proton acceptor following ultrafast ESPT. The HB network along this segment seems to be disordered. The hydroxyl hydrogen of Tyr66 is barely visible, perhaps because of its rotation around the C−O bond. Similarly, W285 appears to be anchored by one HB to the carbonyl group of Asn146, whereas the other hydrogen may be rotating. Also contributing to the difficulty of observing the HBs along this segment is the fact that in our crystal there might be only a small A-state fraction.

Previously, the ASW was found to extend up to Glu5, on the bottom of the β-barrel.[21] The present work shows that, with increasing resolution, the size of this cluster increases, predominantly because additional water molecules can now be identified. We verified this by analyzing over 100 GFP mutants, for which clear correlations were established between resolution and water content on the one hand, and between water content and cluster size on the other hand. The ultimate resolution of GFP sg11 allows us to trace its HB wires not only internally but also on the surface of the protein.

The extended wire has two clear characteristics that were not discussed before. First, there is a possible exit via the backbone carbonyl of Asn146 that leads to a hydrophobic patch on the protein surface (and therefore it is likely an exit rather than an entry point). Exit along this pathway is favored by the permanent HB between W285 and Asn146. This adds to the previously suggested[21] gateway that may form by the rotation of Thr203. Thus, the proton does not always shuttle reversibly between Tyr66 and Glu222, but it may also leak outside.

Second, and possibly more significant, is the observation of what appears to be a "proton-collecting antenna" which leads to the conjectured entry point at Glu5. It is located on a negatively charged face of the protein and, as suggested for proton antennae in proteins,[15] consists of several carboxylates in close proximity. In addition, we find several threonines within this cluster. The side-chain oxygen atoms of these amino acids are connected, on the GFP surface, by 1−4 intervening water molecules. To our knowledge, this is the first time that a nearly complete HB network possibly functioning as a proton-collecting apparatus of a protein is visible from X-ray data.

With the structural data at hand, we may conjecture regarding the mode of operation of this apparatus. Likely, the negative charge of the surface carboxylates attracts protons from solution. If protons are carried by negatively charged buffer molecules like phosphates, they would first dissociate from the solvated buffer molecule at some distance from the protein and then arrive at the surface as hydronium ions. Recent work suggested that proton transfer between acid−base pairs in solution may occur by a concerted proton transfer across several water layers.[47] Protons discharged near the protein surface may traverse an even larger distance between the buffer and the "antenna", propelled by the enhanced Coulombic field of the clustered carboxylates.

Surface-bound protons were proposed to encounter a barrier for leaving the interface,[48] and thus they may migrate preferentially along the surface, rather than leaving to solution. Experimental observations indicate preferential (and fast) lateral migration of protons along negatively charged membranes,[49] so that conceivably a similar effect may operate for negatively charged protein surfaces. As a proton moves between carboxylates, intervening threonines may rotate to prevent the reverse transfer (e.g., Thr38) and thus help funnel the proton in the direction of the orifice. Indeed, the role of threonines as microswitches along proton wires has emerged in several previous studies.[20,21] Once at the orifice, rotation of the orificial carboxylate side chain may help to internalize the proton.[17]

The observation of such an extensive proton-collecting apparatus in GFP lends support to the hypothesis[21] that, following proton exit from GFP, the ASW recruits a new proton from solution. Clearly, direct experimental evidence for this is desired, such as by utilizing judiciously placed mutations on the antenna or the internal pathway connecting it to the chromophore. As discussed in the Experimental Implications section above, mutation work performed to date has focused on the vicinity of the chromophore, and thus cannot provide decisive evidence for or against a functional role for this conjectured proton-collecting apparatus.

Why is such an apparatus needed, if the sole role of GFP were to fluoresce green? One explanation lies in its known role as a fluorescent enhancer of the light emitted by the photoprotein aequorin. An A-state GFP dimer forms a complex with aequorin.[50] After the luminescence reaction of the aequorin, the heterooligomer would disassemble because of transition to the B-state and concurrent weakening of the GFP dimer. This could allow the active site of aequorin to be solvent accessible and be recharged with coelenterazine. The B-state GFP monomer captures a proton, dimerizes, and re-forms the complex. Another reason could be an evolutionary relic of a GFP ancestor functioning as a primitive proton pump.[21]

Acknowledgment. We thank Marco Kloos for help in structure refinement. This research was supported by the Israel Science Foundation (grant no. 122/08). The Fritz Haber Center is supported by the Minerva Gesellschaft für die Forschung, München, FRG.

Supporting Information Available: Supporting tables (S1, X-ray data collection and refinement statistics; S2, list of GFP structures and their statistics; S3, cluster statistics; S4, ASW atom reference list; S5, statistics for protein atoms in ASW subclusters; S6, statistics for water oxygens in ASW subclusters) and supporting figures (S1, schematics of ASW connectivity; S2, cluster size correlations; S3, ASW of decarboxylated GFP). This material is available free of charge via the Internet at http://pubs.acs.org.

JA1010652

(47) Cox, M. J.; Timmer, R. L. A.; Bakker, H. J.; Park, S.; Agmon, N. *J. Phys. Chem. A* **2009**, *113*, 6599–6606.

(48) Mulkidjanian, A. Y.; Heberle, J.; Cherepanov, D. A. *Biochim. Biophys. Acta* **2006**, *1757*, 913–930.

(49) Alexiev, U.; Mollaaghababa, R.; Scherrer, P.; Khorana, H. G.; Heyn, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 372–376.

(50) Ward, W. W. *Photochem. Photobiol. Rev.* **1979**, *4*, 1–57.

(51) DeLano, W. L. *The PyMOL molecular graphics system*; DeLano Scientific: San Carlos, 2002.

(52) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.