THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 4844−4849

pubs.acs.org/JPCB

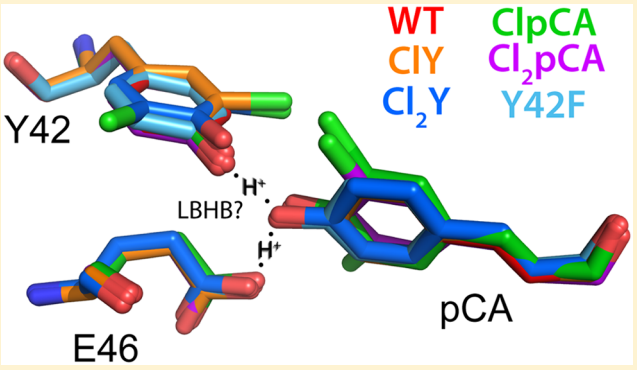
Perturbation of Short Hydrogen Bonds in Photoactive Yellow Protein via Noncanonical Amino Acid Incorporation

Benjamin Thomson,[†] Johan Both,[‡] Yufan Wu,[§] Robert M. Parrish, Todd J. Martínez, and Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Supporting Information

ABSTRACT: Photoactive yellow protein (PYP) is a small photoreceptor protein that has two unusually short hydrogen bonds between the deprotonated *p*-coumaric acid chromophore and two amino acids, a tyrosine and a glutamic acid. This has led to considerable debate as to whether the glutamic acid-chromophore hydrogen bond is a low barrier hydrogen bond, with conflicting results in the literature. We have modified the pK_a of the tyrosine by amber suppression and of the chromophore by chemical substitution. X-ray crystal structures of these modified proteins are nearly identical to the wild-type protein, so the heavy atom distance between proton donor and acceptor is maintained, even though these modifications change the relative proton affinity between donor and acceptor. Despite a considerable change in relative proton affinity, the NMR chemical shifts of the hydrogen-bonded protons are only moderately affected. QM/MM calculations were used to explore the protons' potential energy surface and connect the calculated proton position with empirically measured proton chemical shifts. The results are inconsistent with a low barrier hydrogen bond but in all cases are consistent with a localized proton, suggesting an ionic hydrogen bond rather than a low barrier hydrogen bond.



INTRODUCTION

Short hydrogen bonds, hydrogen bonds (H-bonds) with donor−acceptor (DA) distances less than 2.7 Å, are a common motif in enzyme active sites,[1] and as such, their role in enzymatic catalysis is the subject of intense scrutiny and debate.[2−4] This prevalence of short H-bonds has led to the suggestion that an active site H-bond whose DA distance is shorter than ∼2.5 Å could benefit energetically from being a so-called "low-barrier H-bond" (LBHB).[2] LBHBs are a particular class of short H-bonds in which the proton's zero point energy exceeds the potential energy barrier to DA proton transfer (Figure 1A). LBHBs require not only short DA distances under ∼2.5 Å but also very close matching of DA proton affinities.[2] Under these conditions, the proton could be delocalized between the donor and acceptor, and LBHBs are predicted to be roughly 10 to 20 kcal per mol stronger than normal H-bonds.[5,6] For example, formation of a LBHB with a substrate's transition state (but not the reactant or product ground states) could facilitate reaction catalysis.[7]

There is, however, considerable debate over whether LBHBs are necessary to describe the transition state stabilization afforded by short H-bond formation, as well as whether LBHBs ever exist in protein environments.[8−11] An alternative paradigm suggests that conventional electrostatic factors sufficiently explain the transition state stabilization provided by short H-bonds and that short H-bonds found in enzyme active sites are more likely to be short ionic H-bonds (SIHBs) due to polar solvation[8] (Figure 1B). SIHBs are defined as ionic H-bonds that satisfy the length requirements of a LBHB yet contain a proton that is localized on one of the D or A heteroatoms.[11−13] In contrast to LBHBs, SIHBs can occur between heteroatoms with mismatched proton affinities.

The proton in a LBHB is expected to exhibit a far downfield ^{1}H NMR chemical shift due to the deshielding caused by the proton not being closely associated with either H-bonding heteroatom.[2,14,15] Thus, putative LBHBs in proteins are typically identified indirectly via their particularly short DA distances[16] or by the unusually downfield ^{1}H NMR chemical shifts of their protons;[17,18] however, DA distance and/or chemical shift information are not sufficient to conclude that a particular H-bond is a LBHB. The most direct methods for assessing LBHB presence are those that reveal the location of a H-bond's proton, such as neutron diffraction or ultrahigh resolution X-ray crystallography,[13] noting that those methods report on the H-bond in a crystal environment rather than in solution. In at least one case, ultrahigh resolution X-ray crystallography has shown that an H-bond previously thought

Received: February 18, 2019
Revised: May 20, 2019
Published: May 22, 2019

DOI: 10.1021/acs.jpcb.9b01571
J. Phys. Chem. B 2019, 123, 4844−4849

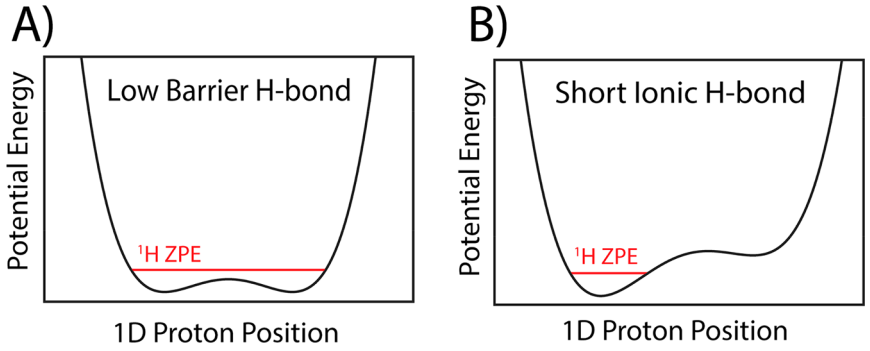


Figure 1. Characteristic 1D ^{1}H proton potential energy surfaces for (A) a short, pK_α matched LBHB and (B) a short, ionic hydrogen bond. The LBHB's barrier for proton transfer is lower than the proton zero-point energy (ZPE), leading to a delocalized proton with a positional expectation value exactly in between the two hydrogen bonding heteroatoms.

to be a LBHB due to its proton chemical shift[17] is likely instead a SIHB.[13]

Structural investigations of the *Halorhodospira halophila* photoactive yellow protein (PYP) have identified two short H-bonds between the protein's deprotonated *p*-coumaric acid chromophore (pCA) and nearby amino acids Y42 and E46, with Y42-pCA and E46-pCA H-bond lengths (heavy atom separations) of 2.48 and 2.58 Å, respectively[16,19] (Figure 2A).

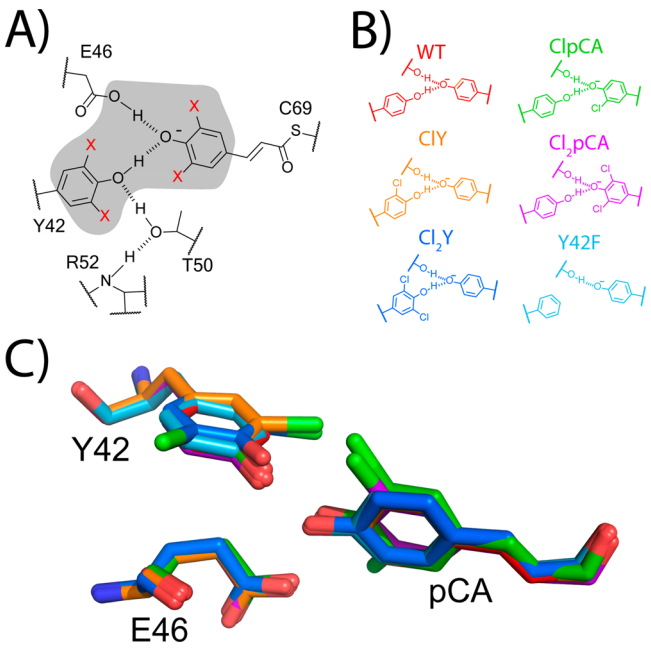


Figure 2. (A) PYP chromophore and surrounding amino acids showing hydrogen bonding network, including the E46-chromophore hydrogen bond, suggested to be a LBHB. The region where changes have been made is highlighted in gray. Red X's show the locations of chloro-substituents. (B) Variants and mutant studied here (color code used throughout). (C) Overlaid PYP variant active site X-ray structures (see Figure S2 for individual structures).

In 2009, Yamaguchi et al. published a 1.5 Å neutron diffraction crystal structure of PYP which appeared to show a deuteron centered in between the two oxygens in the E46-pCA H-bond.[20] This data was interpreted by Yamaguchi and co-workers as the first direct evidence of a LBHB in a protein. Subsequently, Yamaguchi and others have invoked the proposed E46-pCA LBHB to explain the behavior of PYP in solution, which is only reasonable if a LBHB exists in PYP in both crystals and in solution.[21−23]

The interpretation of Yamaguchi's neutron structure has been challenged by several computational studies.[24−28] QM/MM calculations suggest that a LBHB between E46 and pCA could only possibly exist in particular circumstances: in the gas phase,[27] if the Y42-pCA H-bond were not present,[26] or if the side chain of R52 (Figure 2A) were deprotonated,[24,28] an energetically improbable occurrence for a solvent-exposed arginine at neutral pH. Recent NMR experiments have shown that R52 is indeed protonated in solution,[29] and while these NMR experiments do not indicate the protonation state of R52 in the crystal state, Groenhof et al. noted solvent pores in the PYP crystal lattice next to R52, implying that R52 is solvated by water and should be protonated in the pH range under investigation.[27] Furthermore, Groenhof reanalyzed Yamaguchi's raw diffraction data and found that structures containing both a protonated R52 and a localized E46 proton are within the data's margins of error.[27]

It is also unclear whether E46 and deprotonated pCA are sufficiently close in proton affinity to support a LBHB. The solution pK_a's of pCA bound to the denatured PYP backbone (9.1) and free glutamic acid (4.25) are significantly different. It is well-known that the proton affinity of a residue can change significantly upon burial in a protein, and NMR titration experiments have shown little or no change in the protonation state of pCA in the folded protein between pH 3.4 and 11.4,[30] implying that the proton affinities of the chromophore and E46 in the folded protein environment are altered from those in solution. QM/MM calculated pK_a's of pCA and E46 in the protein environment, 5.4 and 8.6, respectively,[24] agree with the localized proton placement produced by other QM/MM calculations.

On the basis of the available data, it is more reasonable to conclude that the E46-pCA interaction is a SIHB rather than a LBHB, but the LBHB hypothesis continues to be invoked to explain observations of the PYP active site.[21−23,31] In the following we investigate the proposed LBHB by perturbing it using amber suppression and artificial chromophore incorporation, coupled to X-ray crystallography, ^{1}H NMR spectroscopy, and QM/MM computations. To examine the H-bond between E46 and the chromophore, we modified the hydrogen bond by producing five altered PYP variants with differentially perturbed E46-chromophore differential acidities (ΔpK_α's) (Figure 2B). We use the convention ΔpK_α to describe the closeness of acidity matching required to form an internal LBHB to avoid confusion with pK_a as defined in aqueous

DOI: 10.1021/acs.jpcb.9b01571
J. Phys. Chem. B 2019, 123, 4844−4849

solution when accounting for the proton affinity difference between solvated and buried acidic residues.[32,33] Since a LBHB requires a ΔpK_α near zero,[2,32] it is unlikely that more than one of these six PYPs contains a LBHB between E46 and the chromophore, due to varying ΔpK_α between variants. Thus, if WT PYP contained a LBHB, such pK_α perturbation should result in destruction of the LBHB and marked changes in observable indicators of a LBHB presence such as the ^{1}H NMR chemical shift of the E46-pCA proton. Because LBHBs have strict geometric constraints,[2] we ensured that the chromophore pocket geometry was conserved across all of our variants by X-ray crystallography. We also predicted the E46 proton one-dimensional DA potential energy surface and proton positional expectation value in each variant, using crystallographic coordinates to initialize our QM/MM computations. Finally, we used our computations in conjunction with an empirical relationship between proton position and ^{1}H NMR shift to generate predictions of E46 proton ^{1}H NMR shifts for each variant. We then compared experimental ^{1}H NMR shifts to our predicted shifts and to those predicted by a LBHB model.

METHODS

A full description of experimental methods and QM/MM methodology can be found in the supplementary text.

EXPERIMENTAL RESULTS

To significantly perturb the H-bond between the chromophore and E46, we increased ΔpK_α by modifying the proton affinity of the chromophore both directly and indirectly. Direct perturbation was accomplished by incorporation of chlorinated chromophore analogues (Figure 2B), which changes the phenolate oxygen's proton affinity as seen by a change in its pK_a under denaturing conditions (Figure S4). The solution pK_a's of the pCA phenolate oxygens were found to be 9.07, 7.56, and 5.74 for WT, ClpCA, and Cl_2pCA, respectively. Indirect perturbation was accomplished by exploiting the complementary nature of the chromophore's hydrogen bonding network[26,34] and modifying the acidity of Y42 by substituting chlorinated tyrosine residues at position 42 (Figure 2B) using amber suppression, as well as the conventional mutation Y42F. The solution pK_a's of the tyrosine phenolate oxygens have been shown to be 9.1, 8.3, and 6.4 for WT, ClY, and Cl_2Y, respectively.[35] The absorption spectrum of the chromophore in these variants shifts systematically (Figure S3).

We determined the crystal structures of the variants ClY, Cl_2Y, ClpCA, and Cl_2pCA; structures for WT (PDB 1NWZ) and Y42F (PDB 1F9I) PYP have been published previously.[19,36] The structures overlay very closely with WT as shown in Figure 2C; details of the structure determination and important DA distances are given in Tables S1 and S2, along with electron density maps in Figure S2. The presence of either a SIHB or LBHB is contingent on the general preservation of the chromophore H-bonding geometry, in particular the DA distance of the E46-chromophore H-bond. As shown in Table S3, we found that in each variant except Cl_2pCA, the E46-pCA DA distance required for a LBHB was conserved.

Proton NMR spectra were obtained for each variant in order to probe the chemical shift of the two protons between E46 and Y42 [Figure 3]. Both the Y42 and E46 H-bonding protons were detected in each variant, except Y42F in which the proton

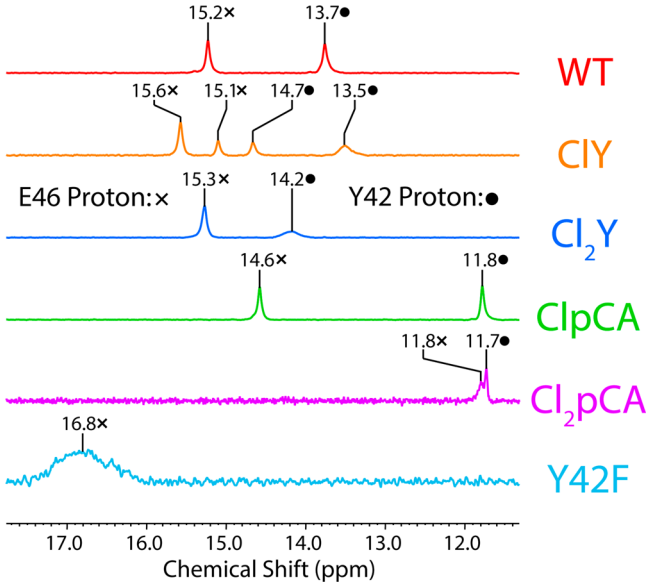


Figure 3. Downfield 1D ^{1}H NMR spectra of each PYP variant. Peaks assigned to the E46 H-bonded proton are labeled with ×, while peaks assigned to the Y42 protons are labeled with ● (see Figures S5–S9 for details of assignment). The four peaks in the ClY spectrum are likely due to two ring flipped conformations of the 3-chlorotyrosine residue observed in solution but not in the X-ray structures.

is absent. Our data for WT PYP matched what has been previously published.[37] ClY displayed four downfield peaks, rather than the expected two. This was interpreted as the presence of two coplanar solution conformations of the 3-chlorotyrosine residue with the ring flipped. Only one such conformation is observed in the crystallographic electron density map for ClY (Figure S2); however, that does not preclude two conformations from existing in solution-phase ClY. Multiple 3-chlorotyrosine conformations have been observed in the crystal structure of apo-ClY57 ketosteroid isomerase, a system with a similar short H-bonding network to ClY PYP.[38,39]

There are only modest variations in the chemical shifts of the E46 proton among the conformationally similar variants (WT, ClY, Cl2Y, ClpCA, and Y42F). Discussion of peak assignments via NOESY data can be found in section S6.

COMPUTATIONAL RESULTS AND DISCUSSION

We used a combination of crystallographic coordinates, quantum mechanical geometry optimizations and scans, and empirical fits to predict ^{1}H NMR shifts from both the LBHB model and the SIHB model and compared those predicted shifts to measured values.

QM/MM geometry optimizations for each variant predicted a proton distance of approximately 1 Å from the E46 oxygen, clearly localized on E46 (Figure 4). In conjunction with our experimental determination of the E46-chromophore DA distances for WT, ClY, Cl_2Y, ClpCA, and Y42F, these predicted proton positions satisfy our definition of a SIHB for all of these variants. In the QM/MM study, constrained optimizations (with the proton constrained to be transferred to the pCA oxygen atom) followed by relaxed optimizations were performed to attempt to find local minima with the proton transferred to the pCA oxygen atom. In all cases, when the constraint was removed, the proton relaxed back to the Y42 or E46 oxygen position. This indicates that in the absence of

DOI: 10.1021/acs.jpcb.9b01571
J. Phys. Chem. B 2019, 123, 4844–4849

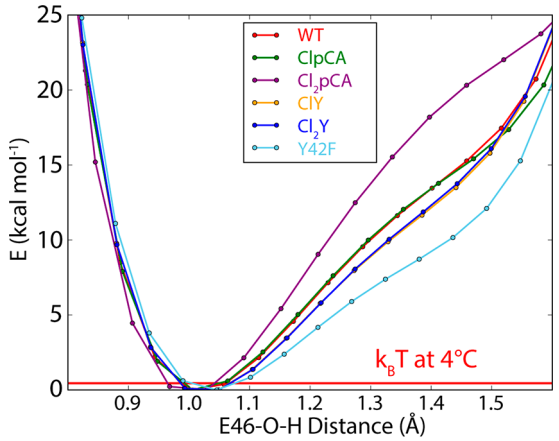


Figure 4. QM/MM 1D potential energy surfaces (rigid proton scans) of proton transfer between E46 and the pCA chromophore, with scan performed from the QM/MM optimal geometry. Computations were carried out at ωPBE(ω = 0.3)-D3/6-31G**//AMBER-ff14SB. Note the lack of a second potential energy minimum in any of the six variants and the predicted proton position at 4 °C approximately 1 Å from the E46 oxygen in all variants.

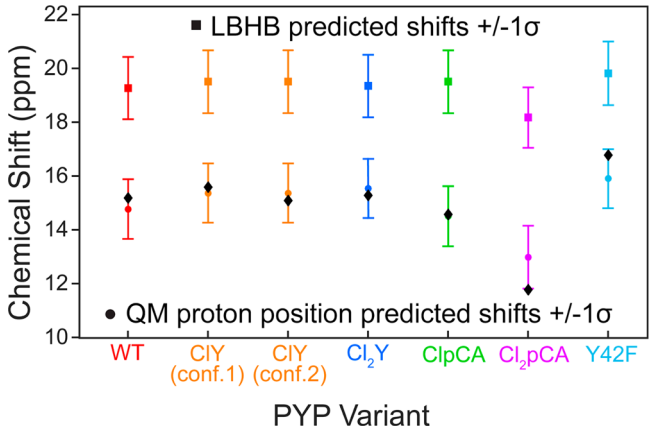


Figure 5. Experimental E46 proton ^{1}H NMR chemical shifts for each variant (♦), compared with corresponding empirical predictions of NMR chemical shift assuming the central proton position produced by a LBHB (square upper symbols), as well as assuming a SIHB (circle lower symbols) using the proton positions given by QM/MM computed predictions. Error bars represent a 1σ deviation given by the linear fit statistics. The experimental data fit with the QM/MM predicted shift rather than with the LBHB predicted shift in each variant, as well as with the relatively small variance in shift between multiple variants despite varying ΔpK_α. ClY is listed twice, as we believe the 3-chlorotyrosine residue exists in two possible conformations in solution, leading to two different E46 ^{1}H NMR chemical shifts (see Figure 3 and text).

more-significant large-scale conformational changes, all hydrogen bond potentials exhibit a single well, with the proton bound to the Y42 or E46 oxygen. We also investigated the effect of full QM treatment of the protein and its crystal water molecules. As shown in Figure S12 of the SI, the proton potential energy curve softens somewhat when the entire protein is treated quantum mechanically and more so when the fully QM protein is aqueously solvated with a polarizable continuum model. However, this softening is far from what would be required for a LBHB. The results from our QM/MM study were used in conjunction with an empirical model developed from previously published data relating the ^{1}H NMR shift of a H-bonded proton to the proton's distance to the acceptor heteroatom.

Jeffrey and Yeon[14] compiled a set of small molecule ionic crystal DA distances determined via neutron diffraction and high-resolution X-ray crystallography, along with the associated H-bonding proton's solid state ^{1}H NMR chemical shifts. Jeffrey and Yeon noted an empirical linear relationship between these two observables across a range of DA distances. We identified their data set for ionic H-bonds as most relevant to PYP due to the deprotonated state of the chromophore and obtained a least-squares fit to model this data as shown in Figure S13. The linear fit was then used to make two sets of predictions of the PYP variants' E46 proton chemical shifts: a set of shifts assuming the existence of a SIHB, and a set of shifts assuming the existence of a LBHB. The SIHB set was produced by using the linear fit in conjunction with proton positions predicted by our QM/MM study, while the LBHB set was produced by using the linear fit in conjunction with proton-acceptor distances derived by dividing our measured E46 DA distance in two, under the assumption that a LBHB would contain a perfectly centered proton. These predictions were plotted against the experimental E46 proton chemical shifts for each variant (Figure 5). It should be noted that, while we have included Cl_2pCA in this analysis, we do not claim that it has a short enough DA distance to be a SIHB or LBHB. We instead classify it as a traditional ionic H-bond due to its longer DA distance. The range of DA distances covered by the linear fit produced from Jeffrey and Yeon's data includes that of Cl_2pCA, as well as the shorter DA distances of WT, ClY, Cl_2Y, ClpCA, and Y42F.

The experimental shifts (Figure 3) for WT, ClY, Cl_2Y, ClpCA, and Y42F were all within a 1σ confidence interval of the QM/MM SIHB predicted shifts, while the experimental shift for Cl_2pCA was slightly outside the 1σ confidence interval of its QM/MM predicted shift. However, the experimental shifts for all variants were far outside the 1σ confidence interval of the LBHB predicted shifts. Furthermore, there is little absolute difference in the E46 proton chemical shift between WT and the variants with conserved E46-chromophore DA bond geometry but assuredly altered ΔpK_α. All shifts in the group of variants WT, ClY, Cl_2Y, and ClpCA fall within the range of 14.6 to 15.6 ppm. Both this lack of significant change in the E46 proton chemical shift among variants of differing ΔpK_α and the agreement of empirically predicted chemical shifts for H-bonds with localized protons fail to support the hypothesis that a LBHB exists between E46 and the chromophore. A normal ionic hydrogen bond model is able to accurately explain the measured ^{1}H NMR shift for Cl_2pCA.

CONCLUSIONS

We investigated the existence of a proposed LBHB in *H. halophila* PYP by perturbing that H-bond directly. Because LBHBs require close pK_α matching between the DA partners,[2,39] varying ΔpK_α between the two DA partners in the E46-chromophore H-bond in a series of variants was expected to significantly alter the behavior of a LBHB, if present. At most one, if any, of these variants could reasonably be expected to contain a LBHB, which would be identifiable from QM/MM computations and NMR experiments.[2,14,40] If one were not present, the behavior of this series of variants should be adequately explained by an alternative SIHB model. We created such a series of six PYP variants with varying ΔpK_α by utilizing artificial chromophore incorporation, amber

DOI: 10.1021/acs.jpcb.9b01571
J. Phys. Chem. B 2019, 123, 4844−4849

nonsense suppression, and conventional mutagenesis, but found no evidence to support the notion that a LBHB is present.

X-ray crystal structures were used to initiate QM/MM geometry optimizations of the E46 protons' optimized one-dimensional potential energy surfaces. These computations predicted that none of our variants contained a LBHB. Rather, they were all predicted to be SIHBs or conventional ionic H-bonds. We then tested these predictions experimentally through ^{1}H NMR using an established empirical linear relationship[14] between the DA distance and ^{1}H NMR chemical shift of the proton in short H-bonds. Application of this linear fit showed, for every variant, good agreement between the variant's experimental E46 proton chemical shifts with the DA distances predicted by the QM/MM calculations. None of the variants showed agreement between experimental chemical shifts and chemical shifts expected from LBHB protons. Furthermore, the E46 proton chemical shifts of the WT, ClY, Cl_2Y, and ClpCA variants were all relatively similar to one another (±0.6 ppm), indicating similar degrees of shielding for the protons in all of these variants and contradicting the notion that one of these variants has a fundamentally different proton environment. These experimental observations agree with our QM/MM predictions and do not support the presence of a LBHB in any of the PYP variants.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.9b01571.

Detailed experimental methods (S.1, S.2), X-ray crystallography data collection statistics (S.3), UV−vis spectroscopic data (S.4, S.5), protein NMR spectra (S.6), detailed computational methods (S.7), and H-bonding proton chemical shift predictions (S.8) (PDF)

Optimized QM/MM structures (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Sboxer@stanford.edu.

ORCID

Todd J. Martínez: 0000-0002-4798-8947

Steven G. Boxer: 0000-0001-9167-4286

Present Addresses

†B.T.: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138.

‡J.B.: Boston Consulting Group, San Francisco, CA 94111.

§Y.W.: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Professor Jiangyun Wang for the gift of 3-chlorotyrosine and 3,5-dichlorotyrosine nonsense suppression systems and Dr. Marc Deller and the Stanford Macromolecular Structure Knowledge Center, as well as staff at SSRL and ALS, for expertise in protein crystallization and X-ray data collection. We also acknowledge Dr. Jeffrey Pelton and Dr. Stephen Lynch of the Berkeley QB3 and Stanford Chemistry NMR facilities, respectively, for their assistance and expertise on collecting NMR data. Part of this work was supported by NIH R35 GM118044 (to S.G.B.); support from the National Science Foundation's CCI Phase I: NSF Center for First-Principles Design of Quantum Processes (CHE-1740645) is gratefully acknowledged.

■ REFERENCES

(1) Rajagopal, S.; Vishveshwara, S. Short Hydrogen Bonds in Proteins. *FEBS J.* **2005**, *272*, 1819−1832.

(2) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. The Low Barrier Hydrogen Bond in Enzymatic Catalysis. *J. Biol. Chem.* **1998**, *273*, 25529−25532.

(3) Katz, B. A.; Spencer, J. R.; Elrod, K.; Luong, C.; Mackman, R. L.; Rice, M.; Sprengeler, P. A.; Allen, D.; Janc, J. Contribution of Multicentered Short Hydrogen Bond Arrays to Potency of Active Site-Directed Serine Protease Inhibitors. *J. Am. Chem. Soc.* **2002**, *124*, 11657−11668.

(4) Jang, D. S.; Choi, G.; Cha, H. J.; Shin, S.; Hong, B. H.; Lee, H. J.; Lee, H. C.; Choi, K. Y. Contribution of a Low-Barrier Hydrogen Bond to Catalysis is Not Significant in Ketosteroid Isomerase. *Mol. Cells* **2015**, *38*, 409−415.

(5) Garcia-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M. Theoretical Study of the Low-Barrier Hydrogen Bond in the Hydrogen Maleate Anion in the Gas Phase. Comparison with Normal Hydrogen Bonds. *J. Am. Chem. Soc.* **1997**, *119*, 1081−1086.

(6) McAllister, M. A. Characterization of Low-Barrier Hydrogen Bonds. 3. Hydrogen Maleate. An ab initio and DFT Investigation. *Can. J. Chem.* **1997**, *75*, 1195−1202.

(7) Cleland, W.; Kreevoy, M. Low-Barrier Hydrogen Bonds and Enzymic Catalysis. *Science* **1994**, *264*, 1887−1890.

(8) Warshel, A.; Papazyan, A.; Kollman, P. On Low Barrier Hydrogen Bonds and Enzyme Catalysis. *Science* **1995**, *269*, 102−106.

(9) Schutz, C. N.; Warshel, A. The Low-Barrier Hydrogen Bond (LBHB) Proposal Revisited: The Case of the Asp—His Pair in Serine Protease. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 711−723.

(10) Warshel, A.; Papazyan, A. Energy Considerations Show that Low-Barrier Hydrogen Bonds Do Not Offer a Catalytic Advantage over Ordinary Hydrogen Bonds. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13665−13670.

(11) Kaledhonkar, S.; Hara, M.; Stalcup, T. P.; Xie, A.; Hoff, W. D. Strong Ionic Hydrogen Bonding Causes a Spectral Isotope Effect in Photoactive Yellow Protein. *Biophys. J.* **2013**, *105*, 2577−2585.

(12) Meot-Ner, M. The Ionic Hydrogen Bond. *Chem. Rev.* **2005**, *105*, 213−284.

(13) Fuhrmann, C.; Daugherty, M.; Agard, D. Subangstom Crystallography Reveals that Short Ionic Hydrogen Bonds, and Not a His-Asp Low-Barrier Hydrogen Bond, Stabilize the Transition State in Serine Protease Catalysis. *J. Am. Chem. Soc.* **2006**, *128*, 9086−9102.

(14) Jeffrey, G. A.; Yeon, Y. The Correlation Between Hydrogen-Bond Lengths and Chemical Shifts in Crystals. *Acta Crystallogr., Sect. B: Struct. Sci.* **1986**, *42*, 410−413.

(15) Pinney, M. M.; Natarajan, A.; Yabukarski, F.; Sanchez, D. M.; Doukov, T.; Schwans, J. P.; Martinez, T. J.; Herschlag, D.; Liu, F.; Liang, R. Structural Coupling Throughout the Active Site Hydrogen Bond Networks of Ketosteroid Isomerase and Photoactive Yellow Protein. *J. Am. Chem. Soc.* **2018**, *140*, 9827−9843.

(16) Anderson, S.; Crosson, S.; Moffat, K. Short Hydrogen Bonds in Photoactive Yellow Protein. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1008−1016.

(17) Frey, P. A.; Whitt, S. A.; Tobin, J. B. A Low-Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases. *Science* **1994**, *264*, 1927−1930.

(18) Zhao, Q.; Abeygunawardana, C.; Talalay, P.; Mildvan, A. NMR Evidence for the Participation of a Low-Barrier Hydrogen Bond in the Mechanism of Δ^5-3-Ketosteroid Isomerase. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8220−8224.

DOI: 10.1021/acs.jpcb.9b01571
J. Phys. Chem. B 2019, 123, 4844−4849

(19) Getzoff, E.; Gutwin, K.; Genick, U. Anticipatory Active-Site Motions and Chromophore Distortion Prime Photoreceptor PYP for Light Activation. *Nat. Struct. Mol. Biol.* **2003**, *10*, 663−668.
(20) Yamaguchi, S.; Kamikubo, H.; Kurihara, K.; Kuroki, R.; Niimura, N.; Shimizu, N.; Yamazaki, Y.; Kataoka, M. Low-Barrier Hydrogen Bond in Photoactive Yellow Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 440−444.
(21) Mizuno, M.; Kamikubo, H.; Kataoka, M.; Mizutani, Y. Changes in the Hydrogen-Bond Network around the Chromophore of Photoactive Yellow Protein in the Ground and Excited States. *J. Phys. Chem. B* **2011**, *115*, 9306−9310.
(22) Kita, Y.; Kamikubo, H.; Kataoka, M.; Tachikawa, M. Theoretical Analysis of the Geometrical Isotope Effect on the Hydrogen Bonds in Photoactive Yellow Protein With Multi-Component Density Functional Theory. *Chem. Phys.* **2013**, *419*, 50−53.
(23) Kanematsu, Y.; Kamikubo, H.; Kataoka, M.; Tachikawa, M. Vibrational Analysis on the Revised Potential Energy Curve of the Low-Barrier Hydrogen Bond in Photoactive Yellow Protein. *Comput. Struct. Biotechnol. J.* **2016**, *14*, 16−19.
(24) Saito, K.; Ishikita, H. Energetics of Short Hydrogen Bonds in Photoactive Yellow Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 167−172.
(25) Saito, K.; Ishikita, H. H. Atom Positions and Nuclear Magnetic Resonance Chemical Shifts of Short H Bonds in Photoactive Yellow Protein. *Biochemistry* **2012**, *51*, 1171−1177.
(26) Saito, K.; Ishikita, H. Formation of an Unusually Short Hydrogen Bond in Photoactive Yellow Protein. *Biochim. Biophys. Acta, Bioenerg.* **2013**, *1827*, 387−394.
(27) Graen, T.; Inhester, L.; Clemens, M.; Grubmüller, H.; Groenhof, G. The Low Barrier Hydrogen Bond in the Photoactive Yellow Protein: A Vacuum Artifact Absent in the Crystal and Solution. *J. Am. Chem. Soc.* **2016**, *138*, 16620−16631.
(28) Nadal-Ferret, M.; Gelabert, R.; Moreno, M.; Lluch, J. M. Are There Really Low-Barrier Hydrogen Bonds in Proteins? The Case of Photoactive Yellow Protein. *J. Am. Chem. Soc.* **2014**, *136*, 3542−3552.
(29) Yoshimura, Y.; Oktaviani, N. A.; Yonezawa, K.; Kamikubo, H.; Mulder, F. A. Unambiguous Determination of Protein Arginine Ionization States in Solution by NMR Spectroscopy. *Angew. Chem., Int. Ed.* **2017**, *56*, 239−242.
(30) Oktaviani, N. A.; Pool, T. J.; Yoshimura, Y.; Kamikubo, H.; Scheek, R. M.; Kataoka, M.; Mulder, F. A. A. Active-Site pKa Determination for Photoactive Yellow Protein Rationalizes Slow Ground-State Recovery. *Biophys. J.* **2017**, *112*, 2109−2116.
(31) Kuramochi, H.; Takeuchi, S.; Yonezawa, K.; Kamikubo, H.; Kataoka, M.; Tahara, T. Probing the Early Stages of Photoreception in Photoactive Yellow Protein with Ultrafast Time-Domain Raman Spectroscopy. *Nat. Chem.* **2017**, *9*, 660−666.
(32) Oltrogge, L. M.; Boxer, S. G. Short Hydrogen Bonds and Proton Delocalization in Green Fluorescent Protein. *ACS Cent. Sci.* **2015**, *1*, 148−156.
(33) Fried, S. D.; Boxer, S. G. Thermodynamic Framework for Identifying Free Energy Inventories of Enzyme Catalytic Cycles. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 12271−12276.
(34) Joshi, C. P.; Otto, H.; Hoersch, D.; Meyer, T. E.; Cusanovich, M. A.; Heyn, M. P. Strong Hydrogen Bond Between Glutamic Acid 46 and Chromophore Leads to the Intermediate Spectral Form and Excited State Proton Transfer in the Y42F Mutant of the Photoreceptor Photoactive Yellow Protein. *Biochemistry* **2009**, *48*, 9980−9993.
(35) Liu, X.; Jiang, L.; Li, J.; Wang, L.; Yu, Y.; Zhou, Q.; Lv, X.; Gong, W.; Lu, Y.; Wang, J. Significant Expansion of Fluorescent Protein Sensing Ability through the Genetic Incorporation of Superior Photo-Induced Electron-Transfer Quenchers. *J. Am. Chem. Soc.* **2014**, *136*, 13094−13097.
(36) Brudler, R.; Meyer, T. E.; Genick, U. K.; Devanathan, S.; Woo, T. T.; Millar, D. P.; Gerwert, K.; Cusanovich, M. A.; Tollin, G.; Getzoff, E. D. Coupling of Hydrogen Bonding to Chromophore Conformation and Function in Photoactive Yellow Protein. *Biochemistry* **2000**, *39*, 13478−13486.
(37) Sigala, P. A.; Tsuchida, M. A.; Herschlag, D. Hydrogen Bond Dynamics in the Active Site of Photoactive Yellow Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9232−9237.
(38) Wu, Y.; Fried, S. D.; Boxer, S. G. A Critical Test of the Electrostatic Contribution to Catalysis with Noncanonical Amino Acids in Ketosteroid Isomerase. *Biochemistry* **2015**, *54*, 7110−7119.
(39) Wu, Y.; Fried, S. D.; Boxer, S. G. Dissecting Proton Delocalization in an Enzyme's Hydrogen Bond Network with Unnatural Amino Acids. *Biochemistry* **2015**, *54*, 7110−7119.
(40) Hibbert, F.; Emsley, J. Hydrogen Bonding and Chemical Reactivity. In *Advances in Physical Organic Chemistry*; Elsevier, 1990; Vol. *26*, pp 255−379.

DOI: 10.1021/acs.jpcb.9b01571
J. Phys. Chem. B 2019, 123, 4844−4849

Supplementary Information for Perturbation of Short Hydrogen Bonds in Photoactive Yellow Protein via Noncanonical Amino Acid Incorporation

Benjamin Thomson[†], Johan Both[§], Yufan Wu[‡], Robert M. Parrish,
Todd J. Martínez, and Steven G. Boxer*
Department of Chemistry, Stanford University
Stanford, CA 94305

Address correspondence to: Sboxer @stanford.edu

[†] Current address: Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138
[§] Current address: Boston Consulting Group, San Francisco, CA 94111
[‡] Current address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Contents

S.1 Synthesis of chromophores and amino acids

Synthesis of Chloro-substituted Tyrosines and para-coumaric Acid Chromophores

3-chloro-L-tyrosine was obtained from Sigma Aldrich, while 3,5-dichloro-L-tyrosine, 3-chloro-*para*-coumaric acid, and 3,5-dichloro-*para*-coumaric acid were

prepared synthetically. 3,5-dichloro-L-tyrosine was prepared by exhaustive chlorination of L-tyrosine, while the chloro-substituted chromophores were prepared via Knoevenagel condensation of the corresponding benzylic aldehydes[1].

Synthesis of 3,5-dichlorotyrosine:

1.0 g of L-tyrosine (5.51 mmol, Sigma) was suspended in 10 mL glacial acetic acid with a stir bar in a 50 mL round bottom flask. Under nitrogen, 2.0 mL of sulfuryl chloride (25 mmol, Sigma) was added dropwise[2]. The reaction was stirred at room temperature for 20 hours. The suspended precipitate was vacuum filtered, washed with cold diethyl ether, and dried. The product was confirmed to match previous characterizations by ^{1}H NMR[3] and used without further purification.

Synthesis of 3-chloro-para-coumaric Acid:

156 mg 3-chloro,4-hydroxybenzaldehyde (1.0 mmol, Sigma) and 229 mg malonic acid (2.2 mmol, Sigma) were dissolved in 450 µL pyridine (Sigma), and 18 µL piperidine (Sigma) in a 10 mL round-bottom flask. The mixture was stirred under reflux at 80°C for 5 hours. The mixture was then cooled to room temperature and 9 mL of 2M HCl was added slowly. The resulting off-white precipitate was filtered, washed with additional acid, and dried. The product was confirmed to match previous characterizations by ^{1}H NMR[1] and used without further purification.

Synthesis of 3,5-dichloro-para-coumaric Acid

191 mg 3,5-dichloro,4-hydroxybenzaldehyde (1.0 mmol, Matrix Scientific) and 229 mg malonic acid (2.2 mmol, Sigma) were dissolved in 450 µL pyridine (Sigma), and 18 µL piperidine (Sigma) in a 10 mL round-bottom flask. The mixture was stirred under reflux at 80°C for 1 hour. The mixture was then cooled to room temperature and 9 mL of 2M HCl was added slowly. The resulting off-white precipitate was filtered, washed with additional acid, and dried. The product was confirmed by ^{1}H NMR, and used without further purification.

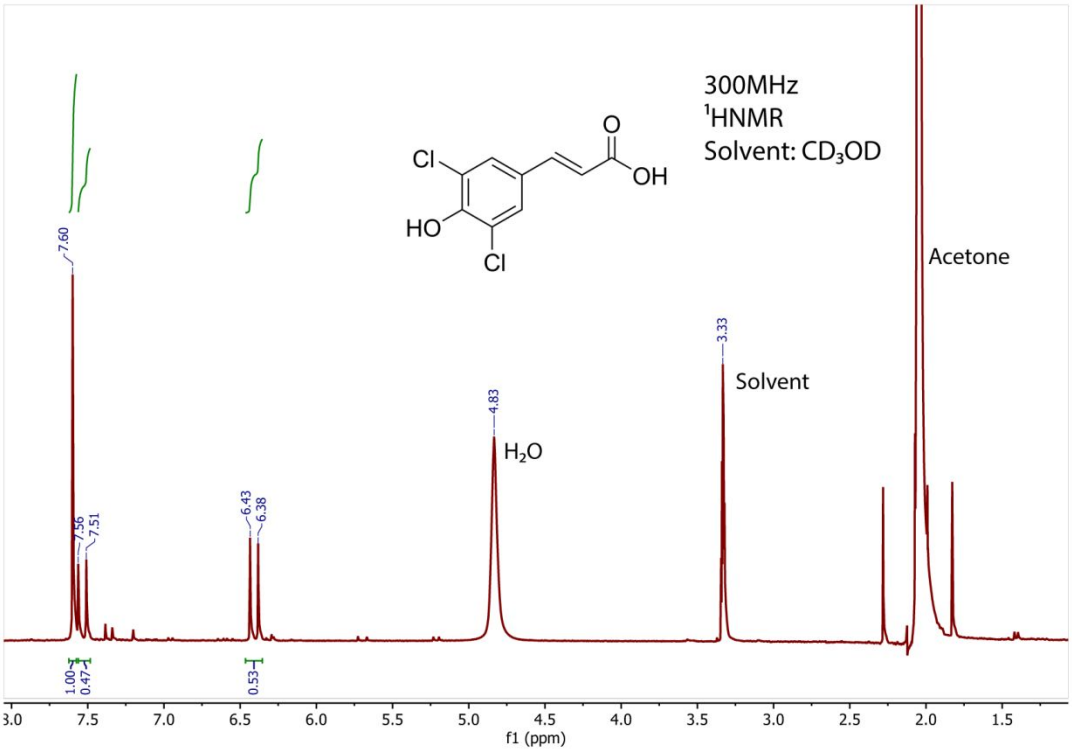


Figure S1: [1]HNMR spectrum of 3,5-dichloro-para-coumaric Acid

Functionalization of PYP Chromophores

PYP chromophores were functionalized for incorporation into PYP by their conversion into acid anhydrides. To ensure complete incorporation of the chromophore into PYP, several hundred equivalents of chromophore were functionalized for addition into the protein lysate solution. 0.8 mmol of the corresponding coumaric acid and 162 mg carbonyldiisoimidazole (1.0 mmol, ThermoFisher) were added to a 50 mL round-bottom flask containing 20 mL THF. The mixture was stirred under nitrogen for 1 hour. The solvent was then evaporated via rotary evaporation, yielding the functionalized chromophore product as a yellow oil[4].

S.2 Protein preparation, purification and mutagenesis

Mutagenesis of PYP

The pET-15b WT PYP plasmid was used as a template for mutagenesis of other PYP variants. The Agilent Genomics QuickChange Lightning DNA mutagenesis kit was used to introduce the following mutations into the pET-15b WT PYP plasmid:

Y42 → Amb (TAG) for production of ClY and Cl_2Y PYP via amber nonsense suppression (codon bold and underlined)
Y42 → F (TTT) for production of Y42F PYP

The PYP DNA sequence was:

```
ATG GAA CAC GTA GCC TTC GGT AGC GAG GAC ATC GAG AAC ACC CTC
GCC AAG ATG GAC GAC GGC CAG CTC GAC GGC CTG GCC TTC GGC GCC
ATC CAG CTC GAC GGC GAC GGC AAC ATC CTT CAG TAC AAC GCC GCG
GAG GGC GAC ATC ACC GGC CGC GAC CCG AAG CAG GTC ATC GGC AAG
AAC TTC TTC AAG GAC GTG GCC CCG TGC ACT GAC AGC CCG GAG TTC
TAC GGC AAG TTC AAG GAA GGG GTG GCC TCG GGC AAC CTG AAC ACG
ATG TTC GAG TAC ACC TTC GAT TAC CAA ATG ACG CCC ACG AAG GTG
AAG GTG CAC ATG AAG AAG GCC CTC TCC GGC GAC AGC TAC TGG GTC
TTC GTC AAG CGC GTC TAA
```

Preparation of WT PYP, Y42F PYP, ClpCA and Cl_2pCA

Cell Growth:

WT PYP in a pET-15b vector was obtained from Genscript. The plasmid was transformed into competent Bl21 (DE3) *E. coli*, and grown on agar plates containing 100 µg/mL ampicillin overnight at 37°C. Colonies were picked and suspended in 5 mL starter cultures of LB containing 100 µg/mL ampicillin. Starter cultures were grown overnight at 37°C with shaking at 180 rpm. Starter cultures were then added to 2L pre-warmed baffle flasks containing 1L Terrific Broth (ThermoFisher), and 100 µg/mL ampicillin. 1L cultures were shaken at 180 rpm and 37°C for 2-4 hours, until OD_{600} reached 0.8. The cultures were then induced with 143 mg IPTG (.6mmol, ThermoFisher), and shaken overnight at 100 rpm and 20°C.

Cell Harvest:

Cells were harvested via centrifugation at 6000 rpm for 1 hour. Cell pellets were resuspended in 30 mL PYP Lysis Buffer (300 mM NaCl, 5 mM imidazole, 50 mM potassium phosphate pH 7.5), lysed via homogenization and the broken cells were spun

down via centrifugation at 15000 rpm for 1 hour, and the resulting supernatant was collected.

Chromophore Incorporation:

10 mL of water were then added to the yellow oil product of the chromophore functionalization reaction, and the resulting suspension was added slowly to the PYP lysate solution. The solution was then softly shaken at 100 rpm, turning bright yellow in minutes. After 1 hour, the incorporated PYP solution was centrifuged at 15000 rpm for 1 hour to spin down excess suspended chromophore.

Purification and 6X His-Tag Cleavage:

The resulting supernatant was filtered and loaded onto a Ni-ion Histidine affinity column equilibrated with PYP Lysis Buffer. The column was then washed with Wash Buffer (300 mM NaCl, 10 mM imidazole, 50 mM potassium phosphate pH 7.5), and PYP was eluted with Elution Buffer (300 mM NaCl, 150 mM imidazole, 50 mM potassium phosphate pH 7.5). The resulting protein solution was transferred into Thrombin Cleavage Buffer (150 mM NaCl, 50 mM potassium phosphate pH 7.5) and incorporated with Thrombin (1 U per mg PYP, Sigma) for 1 hour to cleave the 6XHis tag. The resulting solution of cleaved protein was transferred into Anion Exchange Buffer A (10mM NaCl, 50mM TRIS, pH 8.0) and purified through anion exchange chromatography. Purity and his-tag cleavage were confirmed for all variants via protein ESI LCMS. The resulting pure, cleaved PYP was transferred to Storage Buffer (20mM potassium phosphate, pH 7.0) for storage at 4°C.

Preparation of ClY and Cl_2Y PYP via Amber Suppression

PET-15b PYP plasmids containing the Y42Amb mutation were co-transformed into competent BL21 (DE3) *E. coli* with the appropriate nonsense suppression tRNA synthetase plasmid obtained from Professor Jiangyun Wang[5]. Cells were grown overnight at 37°C on agar plates containing 100 µg/mL ampicillin and 35 µg/mL chloramphenicol. Resulting colonies were picked and suspended in 5 mL starter cultures of LB containing 100 µg/mL ampicillin and 35 µg/mL chloramphenicol and

grown overnight at 37°C with shaking at 180 rpm. Starter cultures were then added to 2L pre-warmed baffle flasks containing 1L Terrific Broth (ThermoFisher), 100 µg/mL ampicillin, 35 µg/mL chloramphenicol, and 0.6 mM of the appropriate chloro-substituted L-tyrosine. For ease of dissolution, the tyrosine derivative was added to the 2L baffle flask before autoclaving. 1L cultures were shaken at 180 rpm and 37°C for 2-4 hours, until OD_{600} reached 0.8. The cultures were then induced with 143 mg IPTG (.6mmol, ThermoFisher), and shaken overnight at 100 rpm and 20°C. Procedures for cell harvest, chromophore incorporation, purification, and 6X his-tag cleavage of ClY and Cl_2Y PYP are identical to those for WT PYP and all proteins were confirmed by mass spectrometry.

S.3 X-ray crystallography

Preparation of Protein Crystals for X-Ray Crystallography

Protein structures for ClY, Cl_2Y, ClpCA, Cl_2pCA, and E46V PYP were determined via X-ray crystallography. Protein crystals for ClY PYP were grown using the hanging-drop vapor diffusion method. Optimal pH, NaCl concentration, and $(NH_4)_2SO_4$ well solution concentrations were screened. Crystal nucleation was found to occur at pH 6.0, 1M NaCl, and 2.7M $(NH_4)_2SO_4$. Initial crystals were stellate, and unsuitable for crystallography. Thus, they were made into a crystal microseed stock which was used to nucleate crystals in wells with pH 6.0, 1M NaCl, and 2.4M $(NH_4)_2SO_4$ solutions. In these conditions, crystals grew as even hexagonal prisms. The microseed stock generated from ClY crystals was used to nucleate drops for all other crystal variants[6].

X-Ray Crystallography

Crystal X-Ray diffraction images were recorded at Stanford Synchrotron Radiation Lightsource at SLAC National Accelerator Laboratory, beamlines 14-1 and 12-2, and at the Advanced Light Source at Lawrence Berkeley National Laboratory, beamline 501. Diffraction images were indexed, integrated and scaled using HKL2000. Molecular replacement and refinement were performed using Refmac and Phenix Refine. Molprobity was used for structure verification, and manual refinement was

performed in Coot. Data collection and analysis statistics are shown in Table S1, key distances in Table S2 and overlays of the structures are in Figure 2C; individual structures are shown in greater detail in Figure S2.

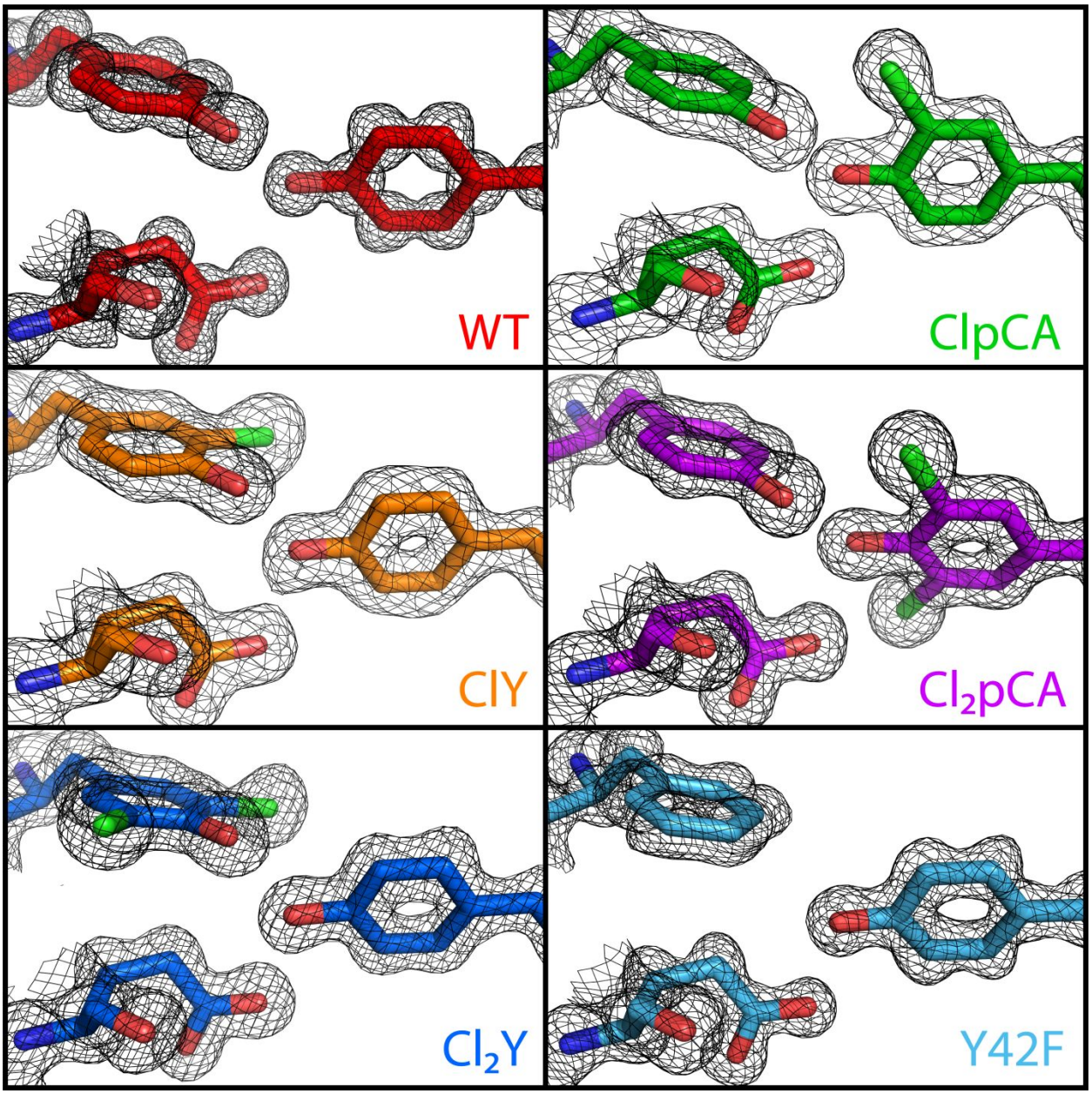


Figure S2: Detail of active site crystal structures, with electron density maps shown at 1σ. WT and Y42F structures and electron density maps from PDB entries 1NWZ[7] and 1F9I[8] respectively.

Table S1 – data collection statistics

Crystal	ClY	Cl_2Y	ClpCA	Cl_2pCA
Space group	P3	P63	P63	P63
Unit-cell parameters				
a (Å)	65.91	65.50	66.25	65.97
b (Å)	65.91	65.50	66.25	65.97
c (Å)	40.5	40.38	40.66	40.42
α, β, γ (°)	90,90,120	90,90,120	90,90,120	90,90,120
Wavelength (Å)	0.98	1.18	1.00	0.89
Temperature (K)	80	80	80	80
Unique reflections	56619	13140	17696	24905
Resolution range (Å)	33.03-1.23	32.90-1.60	33.17-1.67	33.00-1.30
Redundancy	8.6	9.9	21.7	5.9
Completeness (%)	98.7	99.9	99.9	100
<I/σ(I)>	1.84	4.83	5.11	1.77
R_{merge}	0.07	0.07	0.124	0.05
No. of PYP molecules per asymmetric unit	2	1	1	1
R_{work}/R_{free} (%)	15.3 / 18.1	16.7 / 18.8	17.0 / 21.3	14.3 / 18.6
R.m.s. Deviations				
Bond Lengths (Å)	0.02	0.01	0.01	0.02
Bond Angles (°)	1.69	0.90	0.87	1.72
Ramachandran plot (%)				
Preferred	97.50	95.00	95.93	95.93
Allowed	2.50	5.00	4.07	4.07
Outliers	0	0	0	0
PDB code	6MMD	6MKT	6MHN	6MHI

Table S2 - key active site distances from the x-ray structures

Variant	E46-Chromophore O Distance (Å)	Y42-Chromophore O Distance (Å)	Crystal Structure Resolution (Å)
WT[a]	2.58	2.48	0.89
ClY	2.55	2.54	1.23
Cl2Y	2.57	2.50	1.60
ClpCA	2.57	2.55	1.67
Cl2pCA	2.72	2.60	1.30
Y42F[b]	2.51	N/A	1.20

[a] From [7].

[b] From [8].

S.4 UV-Vis Spectroscopic Data

UV-Vis spectroscopy was performed using a Varian Cary 6000i spectrophotometer with a scan rate of 240 nm/minute, resolution of .5 nm, and integration time of 0.133 s. Protein samples were made at approximately 7.5 μM in 10 mM pH 7.5 phosphate buffer to completely deprotonate each protein's chromophore. Denatured UV-Vis titration spectra were measured in 10mM phosphate buffer with 6M GdHCl.. A pH correction of +0.72 was made to pH measurements in GdHCl according to Garcia-Mira and Sanchez-Ruiz[9]. Titration spectra were obtained via an in-cuvette titration of a protein sample by addition of small volumes of HCl. Sample pH was monitored using a Ross PerpHect Micro pH Combination Electrode.

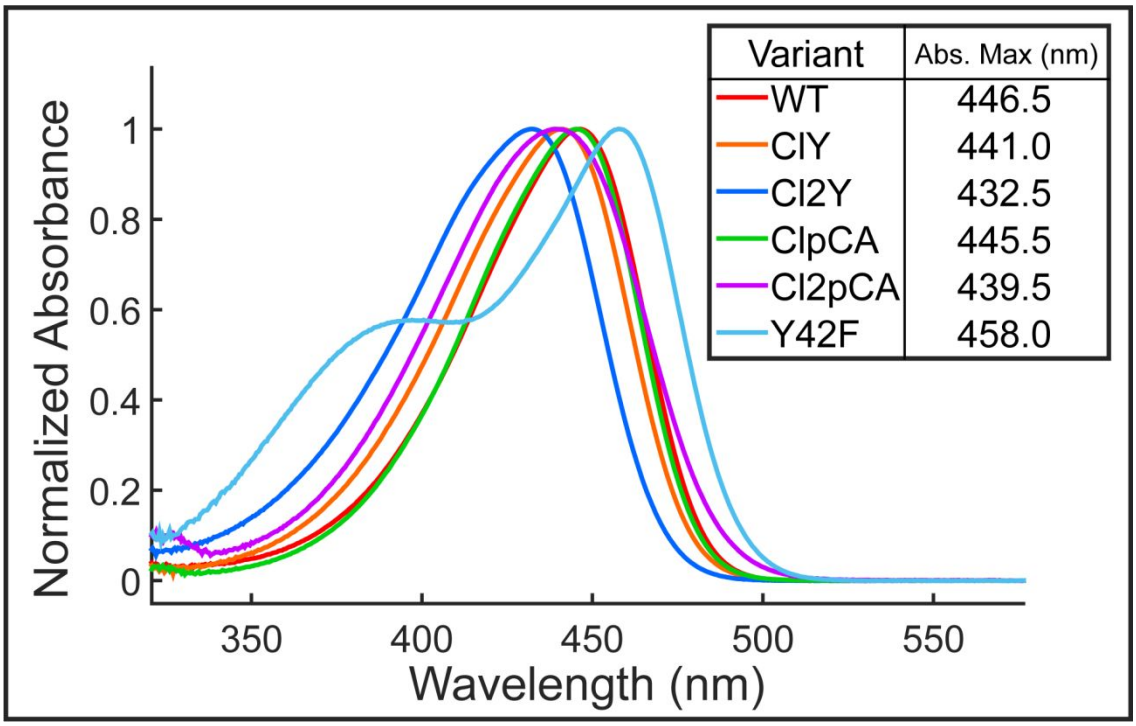


Figure S3: Absorption spectra of PYP variants at pH 7.5, where the chromophore is deprotonated. The presence of two peaks in the Y42F spectrum is consistent with previous observations[10] and is likely not due to different protonation states of the chromophore as titration to high pH did not eliminate the smaller blue-shifted peak. It is possible that the active site of Y42 exists in two different conformations in solution.

S.5 Denatured Chlorinated Chromophore pK_a Determination by UV-Vis Titration

Chlorination of the chromophore has a significant effect on the ΔpK_α between the chromophore and E46 oxygens. UV-Vis titration experiments were performed in triplicate for WT, ClpCA and Cl_2pCA PYP denatured in 6M GdHCl and the maximum absorption was fitted to the sigmoid derived from a modified Henderson-Hasselbalch Equation,

Eq. S2

$$A = \frac{1}{1 + 10^{n(pKa - pH)}}$$

where A represents the normalized peak absorption of a titration PYP mutant sample at a given pH (Fig. S4). It has been previously shown that chlorination of tyrosine at the 3 and 5 positions[5] and incorporation into the GFP chromophore[2] results in similar significant solution pK_a decreases.

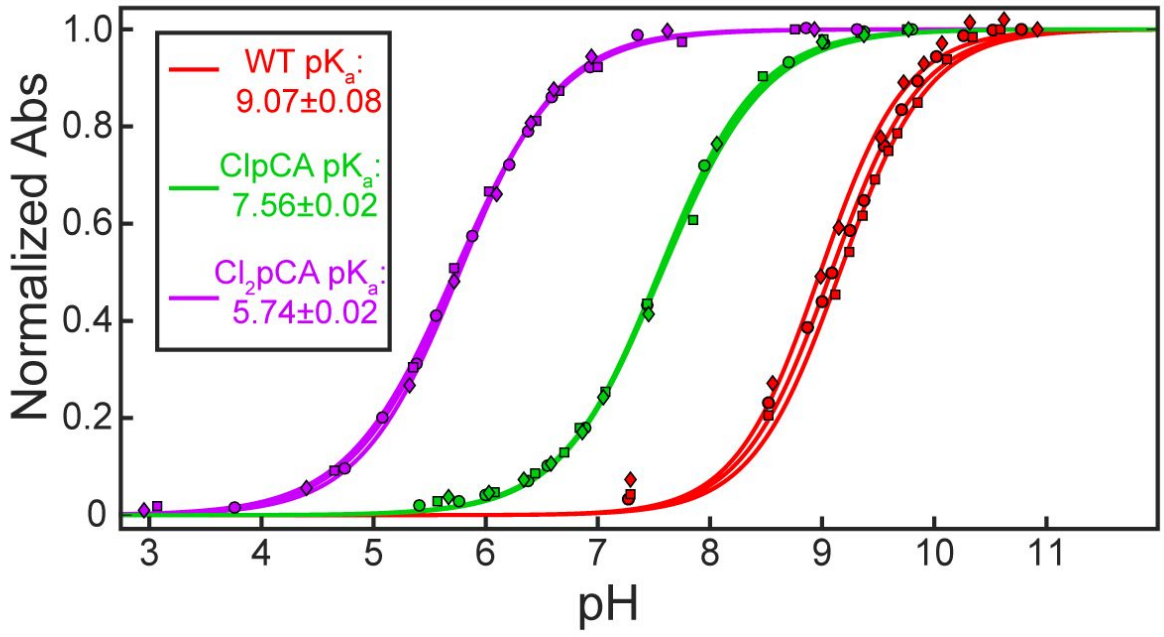


Figure S4: Denatured titration curves of chromophores in WT, ClpCA and Cl_2pCA PYP with fitted pK_a's. Titrations were performed in triplicate by monitoring the absorbance at the denatured, deprotonated absorption peak of each variant (398.5 nm, 397.5 nm, and 396.5 nm for WT, ClpCA and Cl_2pCA respectively). Absorbances were normalized to the absorbance at the highest pH measurement.

S.6 Assignments of H-bond Proton Peaks in NMR Spectra

One-Dimensional NMR Spectroscopy

One-Dimensional NMR Spectra were acquired on a Varian Inova 600 MHz NMR Spectrometer at Stanford. NMR samples consisted of 0.5 mM PYP variant, 50mM potassium phosphate buffer, 5% D_2O, 1mM EDTA, and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as reference, pH 7.5. Data were collected using a 1:1 water suppression pulse sequence with a spectral width of 30 ppm, and a recycle delay of 1.9 s. 13000 points were collected over 256 scans per sample. Data were processed in MNova by applying a multi-point baseline over the downfield peaks of interest. Chemical shifts were referenced internally to DSS.

Two-Dimensional NMR Spectroscopy

Two-Dimensional ^{1}H-^{1}H NOESY NMR spectra for WT, ClpCA, Cl_2pCA, ClY and Cl_2Y were recorded using the QB3 900 MHz NMR Spectrometer at UC Berkeley. Spectra for WT, ClpCA, and Cl2pCA were recorded using the 3-9-19 water suppression sequence with flip-back. A spectral width of 24.8 ppm was acquired with 2048 total points, a recycle delay of 1.55 s, and an NOE mixing time of 50 ms.

Assignments of the downfield proton peaks in the WT PYP ^{1}HNMR spectrum have been previously determined using a ^{1}H-^{1}H NOESY experiment[11]. The ^{1}H-^{1}H NOESY revealed NOE peaks in the aromatic region coupled with each of the downfield H-bonded protons: one aromatic peak at 6.2 ppm appeared coupled with the 15.2 ppm proton, while two aromatic peaks at 6.2 ppm and 6.6 ppm appeared coupled with the 13.7 ppm proton. An earlier assignment of the ^{1}H nmr spectrum of PYP[12] revealed that the 6.2 ppm NOEs corresponded to the protons on the ortho-position of the chromophore, while the 6.6 ppm NOE corresponded to the protons at the ortho position of Y42. Because NOEs are highly dependent on the spatial distance between coupled protons, it is reasonable to infer that the 13.7 ppm proton is in closer spatial proximity to the Y42 ortho protons than the 15.2 ppm proton. Thus, the 13.7 ppm peak is assigned to the Y42 H-bonding proton, and the 15.2 ppm peak is assigned to the E46 H-bonding proton. These data are shown in Fig. S5 and confirm the earlier work.

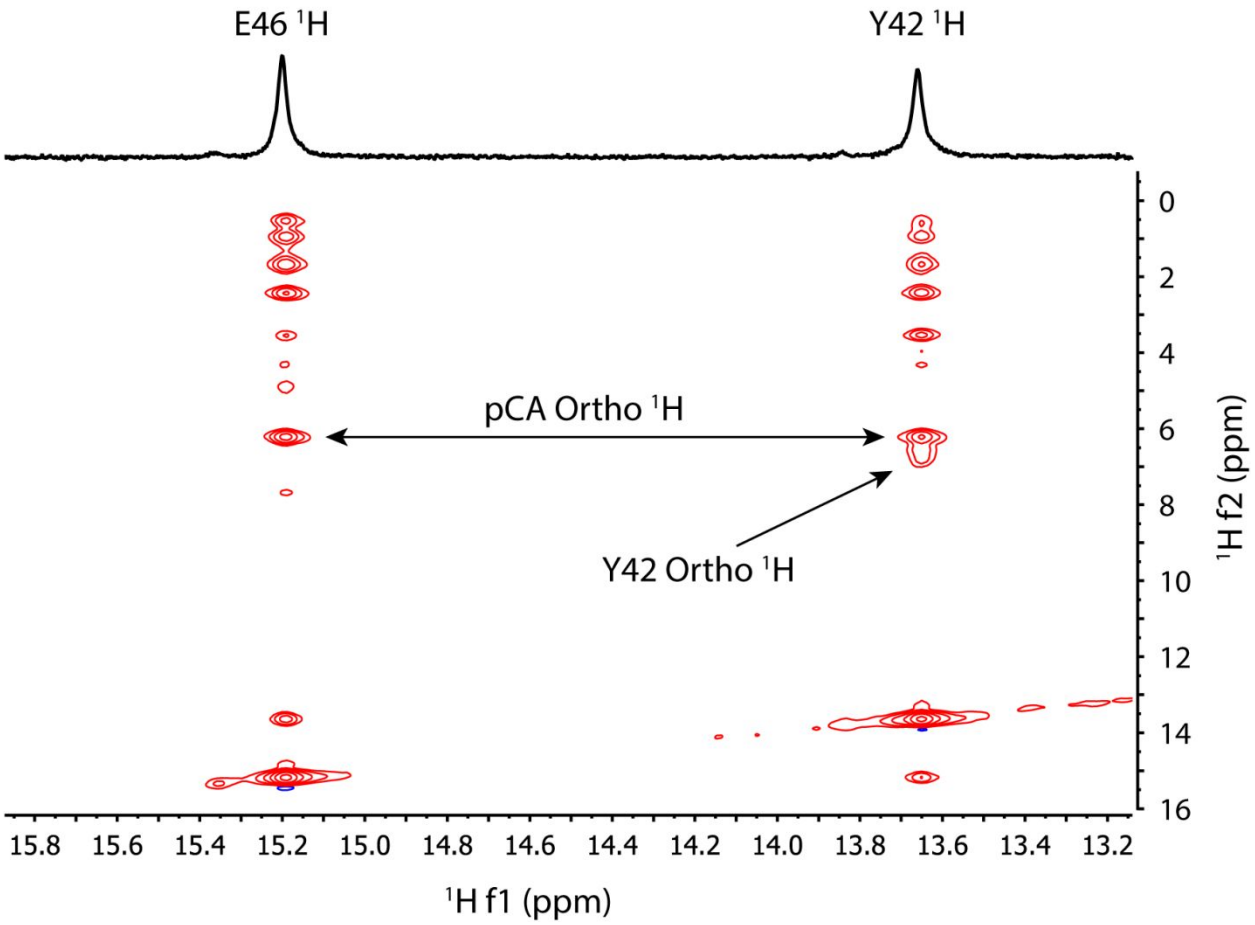


Figure S5: WT ^{1}H-^{1}H NOESY

ClpCA NOESY

As in the WT 2D spectrum, the proton peaks were assigned based on the presence of multiple aromatic NOEs associated with the farther upfield of the two peaks. The 11.7 ppm peak was therefore assigned to the Y42 proton, and the 14.6 ppm peak to the E46 proton.

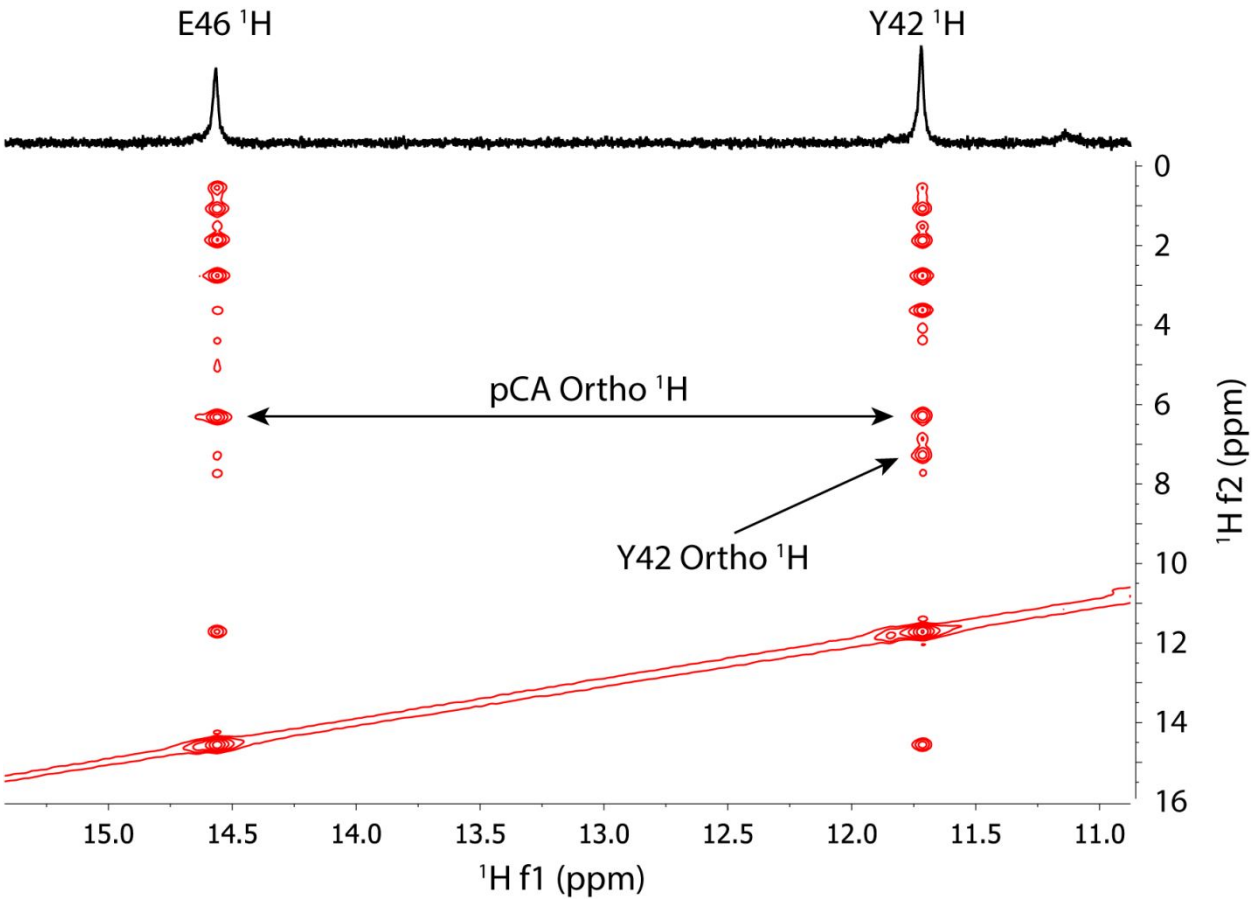


Figure S6: ^{1}H-^{1}H ClpCA NOESY Spectrum

Cl_2pCA NOESY

The two peaks in the Cl_2pCA 1D spectrum are quite close together; however, differing NOE peaks are discernable. The two ortho protons on the chromophore are replaced with chlorines in this variant, so that a single aromatic NOE peak would indicate spatial proximity to the Y42 ortho protons; the peak at 11.7 ppm, the farther upfield of the two peaks, shows such an NOE peak. The 11.7 ppm peak was thus assigned to the Y42 proton, and the 11.8 ppm peak to the E46 proton by default.

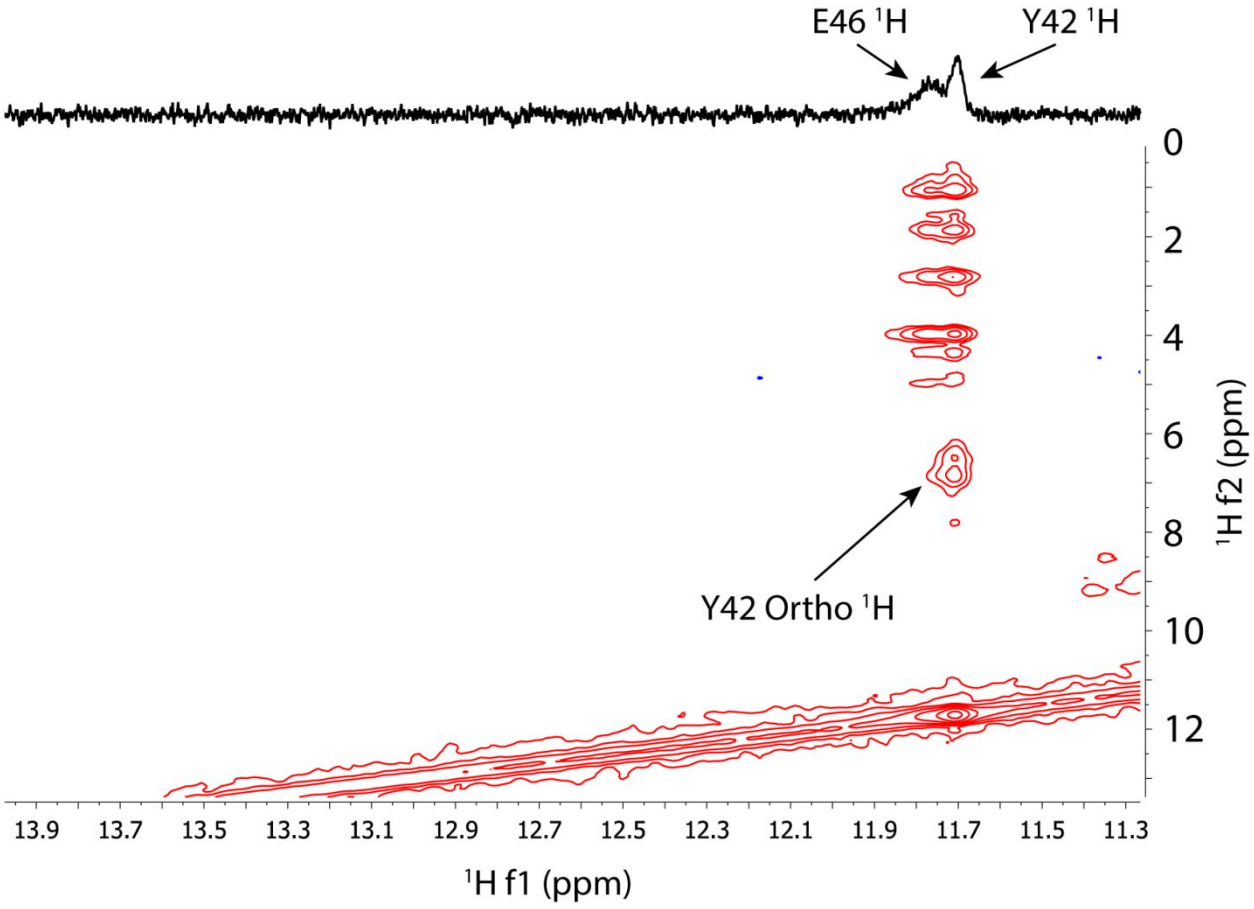


Figure S7: ^{1}H-^{1}H Cl_2pCA NOESY Spectrum

ClY NOESY

The 1D ClY NMR spectrum shows four peaks. A likely source is two possible conformations of the active site tyrosine that share the same ring planarity, but differ in orientation of the chloro substituent. Only one conformation of this tyrosine was evident in the solved crystal structure (SI Fig. S2), however, this may be an artifact of crystallization and does not necessarily imply that only one conformation exists in solution.

Because these two distinct conformations exist, the peaks must be assigned pairwise. Two of the peaks represent E46 protons, while the other two represent Y42 protons. The corresponding pair of peaks at 15.1 ppm and 14.7 ppm were assigned in a manner identical to the WT peaks. The NOESY spectrum of ClY reveals multiple aromatic NOEs for the 14.7 ppm peak, but only one aromatic NOE for the 15.1 ppm peak. Thus, the 14.7 ppm peak was assigned to the Y42 proton, and the 15.1 ppm peak to the E46 proton. Unfortunately, the broad peak at 13.5 ppm did not produce any NOEs (except for a solvent exchange peak), and is thus impossible to directly assign. However, the presence of only one aromatic NOE coupled to the 15.6 ppm peak implies that it arises from the E46 proton. This leaves the Y42 proton to be assigned to the 13.5 ppm peak, assuming that the 15.6 ppm peak and the 13.5 ppm peak arise from the same conformation.

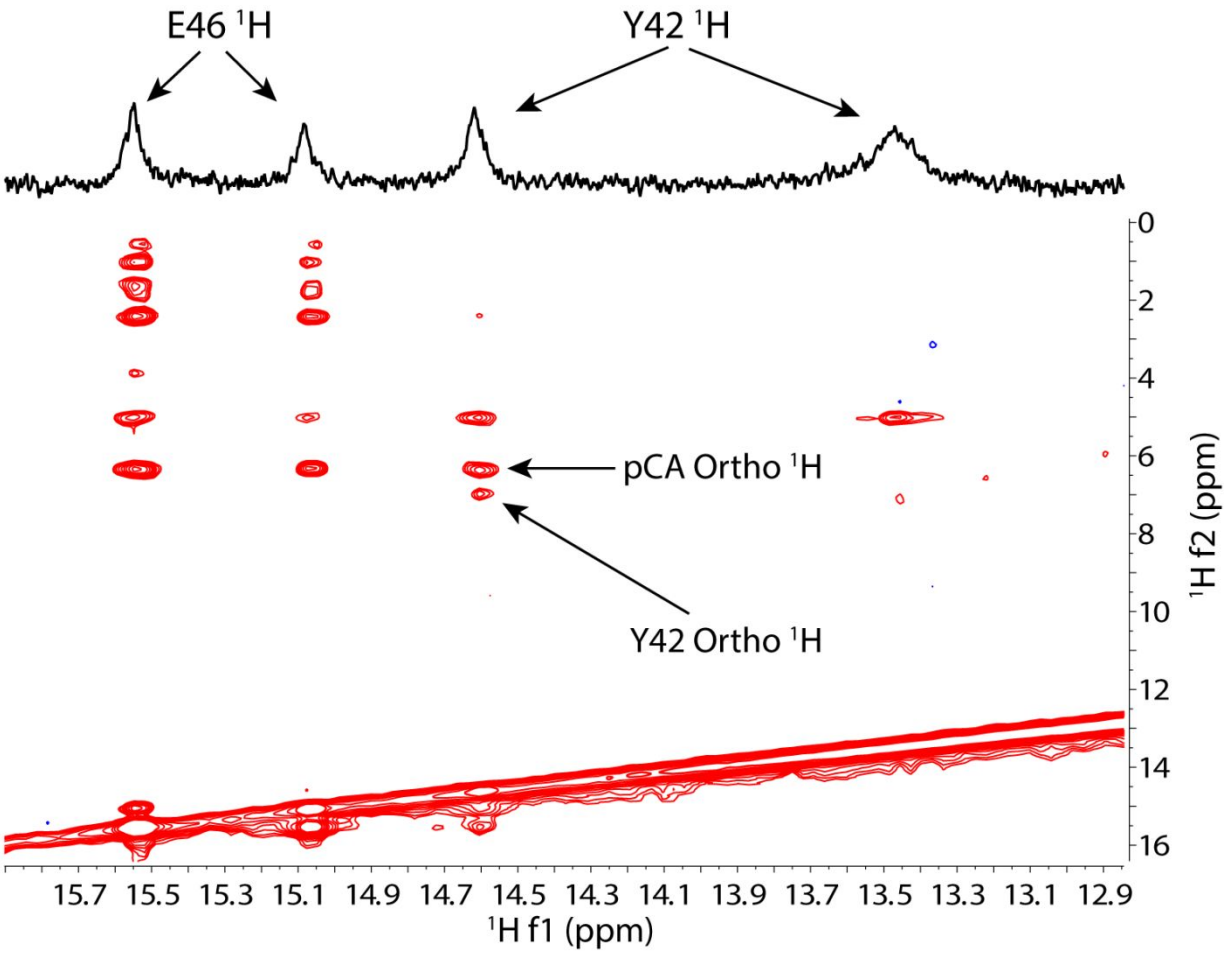


Figure S8: ^{1}H-^{1}H CIY NOESY Spectrum

Cl_2Y NOESY

The peaks in the Cl_2Y spectrum were the most difficult to assign with certainty, as the lack of ortho protons on Y42 in this variant prevents us from assigning peaks using their aromatic NOEs. Furthermore, the 1D Cl_2Y spectrum contains a broad peak at 14.3 ppm that did not produce any aromatic NOEs. For this reason, we postulated that the farther downfield 15.3 ppm peak arose from the E46 proton, as in every other variant, that proton appeared farthest downfield.

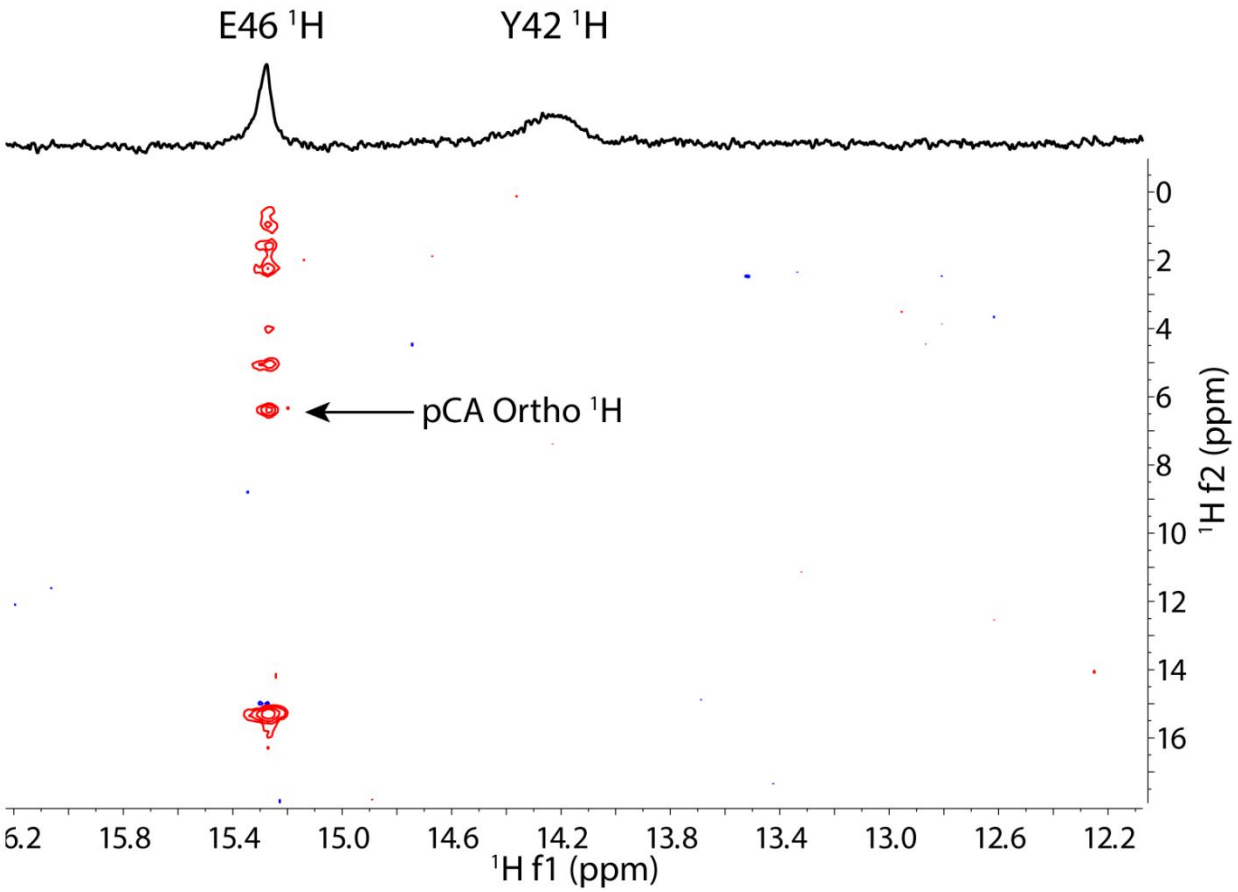


Figure S9: ^{1}H-^{1}H Cl_2Y NOESY Spectrum

S.7 Computational methods

QM/MM geometry optimizations and potential energy scans were carried out using the TeraChem GPU-accelerated electronic structure package for WT, ClY, Cl_2Y, ClpCA, Cl_2pCA, and Y42F PYP[13-17]. Our QM/MM implementation follows previous work,[26,27] including the usual link atom treatment where H atoms are placed along cut covalent bonds. The QM and MM regions are illustrated in Fig. S10. AMBER14 was used for force field model setup and initial optimizations[18]. GAFF models for the pCA, ClpCA, and Cl_2pCA chromophore residues and the Y42, ClY42, Cl_2Y42, and F42 residues were constructed. The parameters for Y42 and F42 are already present in AMBER but are parametrized by GAFF within this study to use similar parameter selection across various standard and non-standard amino acids and chromophores[19]. RESP charges at HF/6-31G* are used to complete the GAFF models[20]. The protonation states for the titratable protein residues were determined by PROPKA 3.1 at pH 7.0. Standard amino acid residues were parametrized with ff14SB, and waters were parametrized with SPCFW. Hydrogens were added in Leap, and initial optimizations of water and hydrogen atom coordinates were carried out in SANDER. Manual rotation of a few hydrogen positions was required to provide the correct hydrogen bond topology from that provided by Leap, prior to starting SANDER optimizations – this was particularly relevant for the E46 – pCA ionic hydrogen bond.

The selected QM region included the pCA chromophore from CB onward, Y42 from CB onward, E46 from CG onward, and T50 from CB onward (Fig. S9). Thus, Y42F had 50 QM atoms, all other variants had 51 QM atoms, and all variants had a total charge of -1 in the QM region and 4 covalent links to the MM region. During QM/MM optimizations, the positions of all backbone CA atoms were constrained, with the exception of the CA atoms on pCA, Y42, E46, and T50 (not frozen to allow extra flexibility in the region of the pCA chromophore hydrogen-bonding pocket).

QM/MM geometry optimizations were carried out in Cartesian coordinates with restrained CA atoms in TeraChem using the DL-FIND optimization package. The results for the key hydrogen bond distances are shown in Figure S11 and in SI Table S3 below. The starting geometries were taken from two sources to probe the sensitivity of the

optimized structures to the initial conditions: (1) with heavy atom positions taken from WT geometry modulo variations from WT which were performed in Avogadro, and (2) with heavy atom positions taken from available crystal structures. Note that the selection of the heavy atom positions is performed at the very beginning of the analysis, prior to the construction of AMBER/GAFF modeling and initial MM optimizations in SANDER. Additionally, two different levels of theory [B3LYP-D3/6-31G** and ωPBE(ω=0.3)-D3/6-31G**] were used to probe the sensitivity of the optimized geometries to level of theory[21-24]. As seen in Figure S11 the optimized geometries from different starting points and levels of theory are all in rather close agreement of 0.01 – 0.02 Å, with the exception of the Y42—pCA hydrogen bond distance in the Cl2Y variant, where the choice of WT vs. crystal structure starting geometry leads to a discrepancy of ~ 0.06 Å in the bond distance. Overall, this agreement indicates that the optimized geometries are not particularly sensitive to the choice of level of theory or starting geometry. All levels of theory generally reproduce the bond distances found in crystal structures (despite, for example, possibly very different starting bond distances before optimization), with the exception of the Y42 – pCA H-bond distance in ClY, which differs from the crystal structure value by ~0.1 Å. From here onward, we will use the ωPBE(ω=0.3)-D3/6-31G** results with starting geometries taken from mutant crystal structures for analysis.

To probe the possibility for an LBHB, the proton for each H-bond was moved from its optimized position near Y42 or E46 and placed 0.96 Å from the pCA oxygen atom. The proton and pCA oxygen are fixed, and another geometry optimization is performed – this is intended to relax the surrounding coordinates of the H-bonding pocket with the new H-bond topology. Finally, the constraint on the proton and pCA oxygen atom was removed and another optimization was performed. In all such analyses (11 total), the proton reverts to affiliating with the Y42 or E46 oxygen, rather than the pCA oxygen. This indicates that in the absence of more-significant large-scale conformational changes, all H-bond potentials exhibit a single well, with the proton bound to the Y42 or E46 oxygen.

We have performed additional rigid scans of the proton transfer coordinate (without relaxation) as shown in Figure 4 of the primary manuscript. While these are

admittedly upper bounds to the potential energy along the true proton-transfer reaction coordinate (as these scans lack relaxation of the surrounding coordinates), the large increases in potential energy along this scan coordinate indicate that it is unlikely that significant differences in H-bond potential energy surfaces exist among these variants.

It is of interest to probe the potential effects of treating the entire protein quantum mechanically, with or without polarizable continuum (PCM) treatment of the surroundings. We carried out such computations for the proton transfer scan shown in Figure 4 of the main text. The calculations in this case use the 6-31G basis set and include 2367 atoms with 8936 electrons and 12689 basis functions. The ωPBE(ω=0.3) functional is used and the PCM treatment follows previous work. The total charge of the protein and included crystal waters is -6. Results are shown in Figure S12. Full quantum mechanical treatment of the entire protein softens the proton potential negligibly. The potential is further softened by inclusion of a surrounding polarizable continuum with dielectric constant corresponding to water (ε=80). Nevertheless, the softening is far from sufficient to provide any indication of a possible LBHB.

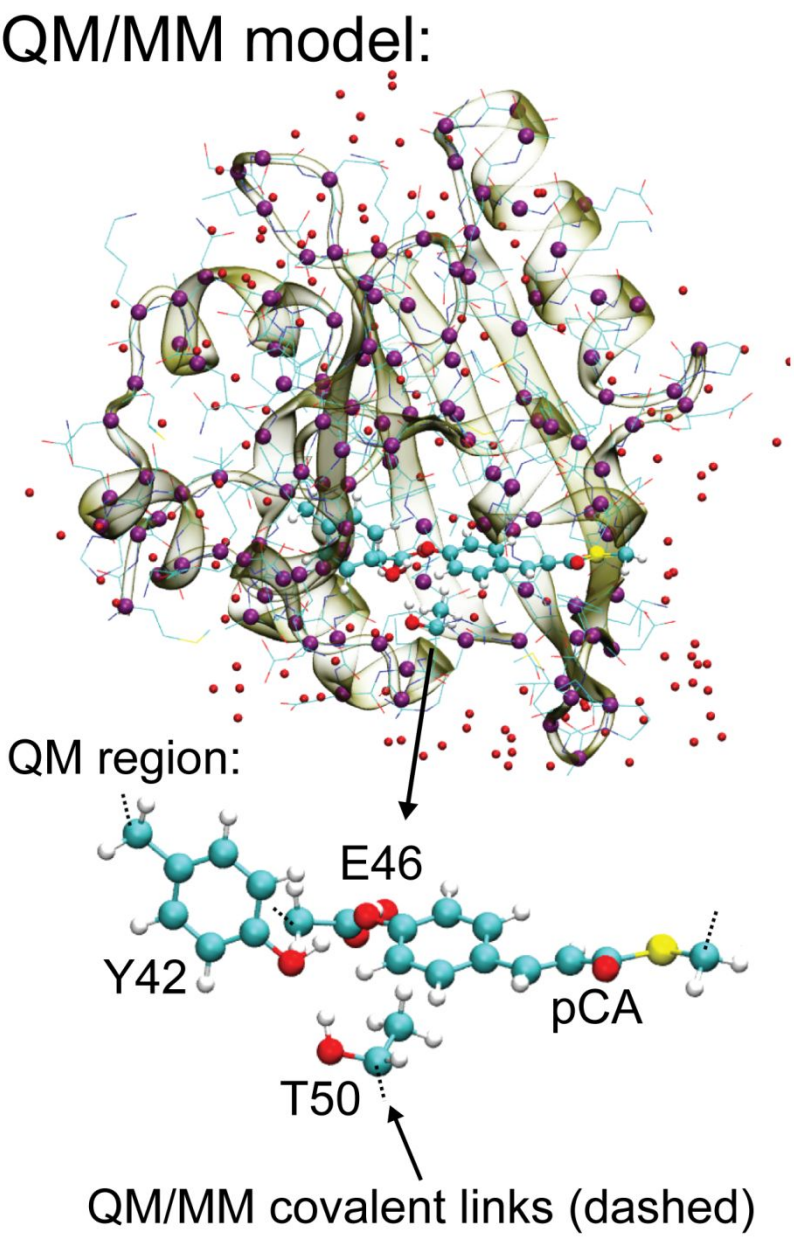


Figure S10: Schematic of QM/MM model geometry for WT PYP, indicating QM region, MM region, QM/MM links, and restrained CA backbone atoms. Hydrogens are not shown in MM region for clarity.

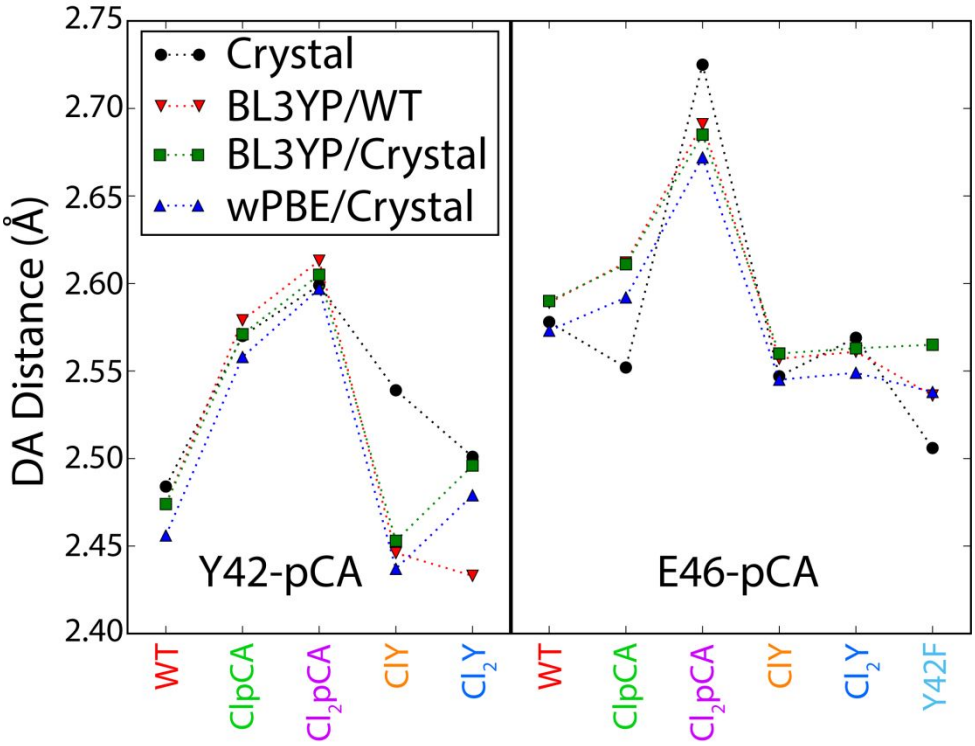


Figure S11: Comparison of crystal structure vs. theory predicted O—O bond distances in critical hydrogen bonds between Y42—pCA (left) and E46 – pCA (right). B3LYP refers to B3LYP-D3/6-31G**, wPBE refers to ωefe(ωe0.3)-D3/6-31G**. WT refers to heavy atom geometry for optimization guess taken from WT PYP. Crystal refers to heavy atom geometry for optimization guess taken from mutant crystal structures. Dashed lines are a guide for the eye – no data are present along these lines.

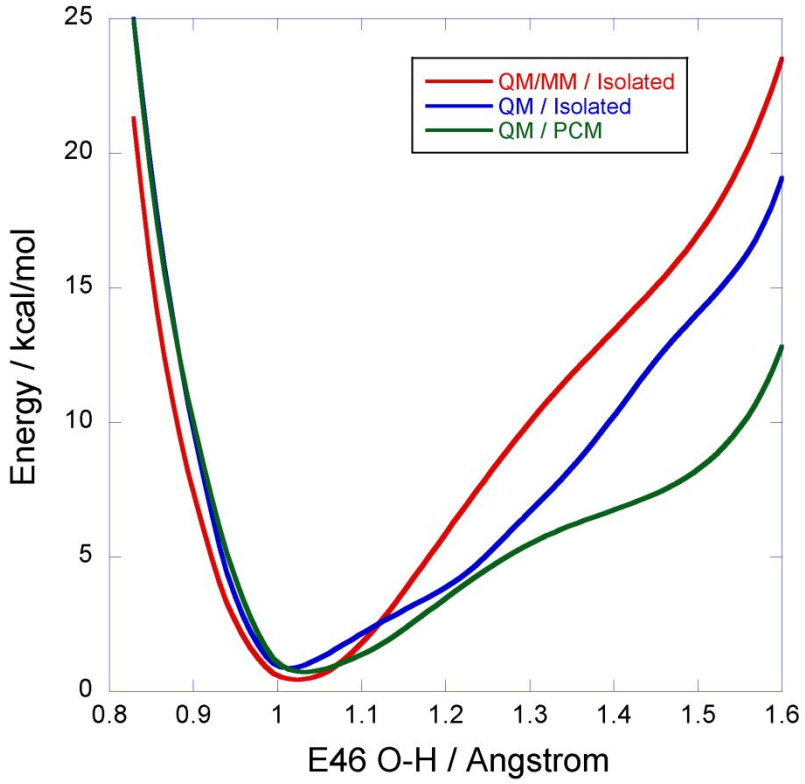


Figure S12: Proton transfer scan (analogous to Figure 4 of the main text) for the wild type PYP protein. The QM/MM results from the main text are shown (red line), as well as the effect of modeling the entire protein quantum mechanically (blue line) and further including aqueous solvation with PCM (green line). Although both QM treatment of the entire protein and further aqueous solvation lead to softening of the proton transfer potential curve, it is far from sufficient to lead to an LBHB.

S.8 ^{1}HNMR Chemical Shift Predictions

Predictions of ^{1}HNMR chemical shift using the SIHB model and the LBHB model were performed using the following least-squares linear empirical fit correlating proton-acceptor heteroatom distance with ^{1}HNMR chemical shift using the data from Jeffrey and Yeon[25]:

$$\delta = -15.618 * r + 39.41 \qquad \text{Eqn. S1}$$

where δ = proton chemical shift in ppm, and r = the proton-acceptor O distance in Å (Figure S13). Only anionic H-bonds were used in the fitting. SIHB model H-acceptor

distances were derived by subtracting the QM/MM predicted E46-proton distance from the crystallographic DA distance in each variant (Table S3). LBHB model H-acceptor distances were taken to be half of the H-bond DA distance, implying a completely centralized proton.

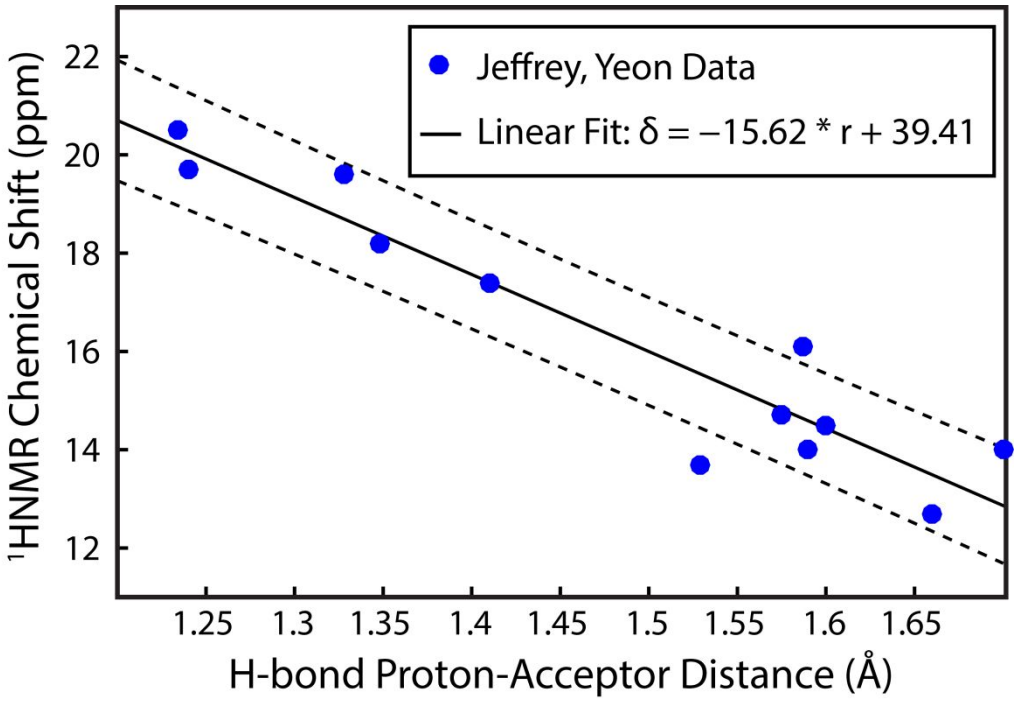


Figure S13. Empirical linear relationship between an H-bonded proton's ^{1}H-NMR chemical shift and the proton-acceptor distance in Å. Each blue data point represents an anionically H-bonded compound described by Jeffrey and Yeon[25] where the proton-acceptor distance was determined by neutron diffraction or high resolution X-ray crystallography and the 1H-NMR chemical shift by solid state NMR. A least-squares fit trendline (solid line) with 1σ confidence bounds (dashed lines) is shown.

Table S3: QM/MM Predicted Expectation Values between E46 and Y42 Protons and the Chromophore Oxygen with Predicted and Experimental Proton NMR Shifts. The two values given for the ClY experimental shift result from the two distinct solution conformations of chlorotyrosine as detected via NMR spectroscopy.

Mutant	SIHB Predicted Proton to Chromophore O Distance (Å)	LBHB Predicted E46 Proton to Chromophore O Distance (Å)	SIHB Predicted E46 Proton NMR Chemical Shift (ppm)	LBHB Predicted Proton NMR Chemical Shift (ppm)	Experimental Proton NMR Shift (ppm)
WT	1.577	1.290	14.8	19.3	15.2
ClY	1.539	1.275	15.4	19.5	15.6/15.1
Cl2Y	1.528	1.285	15.6	19.4	15.3
ClpCA	1.594	1.285	14.5	19.4	14.6
Cl2pCA	1.691	1.360	13.0	18.2	11.8
Y42F	1.505	1.255	15.9	19.8	16.8

Supplementary Information References:

1. Wuensch, C.; Pavkov-Keller, T.; Steinkellner, G.; Gross, J.; Fuchs, M.; Hromic, A.; Lyskowski, A.; Fauland, K.; Gruber, K.; Glueck, S.M.; Faber, K. Regioselective Enzymatic β-Carboxylation of *para*- Hydroxy-Styrene Derivatives Catalyzed by Phenolic Acid Decarboxylases. *Adv. Synth. Catal.* **2015**, 357, 1909-1918.
2. Oltrogge, L.; Boxer, S.G.; Short Hydrogen Bonds and Proton Delocalization in Green Fluorescent Protein (GFP). *ACS Cent. Sci.* **2015**, 1, 148-156.
3. Klausmeyer, P.; Howard, O.M.Z.; Shipley, S.M.; McCloud, T.G.; An Inhibitor of CCL2-Induced Chemotaxis from the Fungus *Leptoxyphium* sp. *J. Nat. Prod.* **2009**, 72, 1369-1372.
4. Imamoto, Y.; Ito, T.; Kataoka, M.; Tokunaga, F. Reconstitution of Photoactive Yellow Protein from Apoprotein and *p*-Coumaric Acid Derivatives. *FEBS Lett.* **1995**, 374, 157-160.
5. Liu, X.; Jiang, L.; Li, J.; Wang, L.; Yu, Y.; Zhou, Q.; Lv, X.; Gong, W.; Lu, Y.; Wang, J. Significant Expansion of Fluorescent Protein Sensing Ability through

the Genetic Incorporation of Superior Photo-Induced Electron-Transfer Quenchers. *J. Am. Chem. Soc.* **2014**, 136, 13094-13097.

6. Yamaguchi, S.; Kamikubo, H.; Shimizu, N.; Yamazaki, Y.; Imamoto, Y.; Kataoka, M. Preparation of Large Crystals of Photoactive Yellow Protein for Neutron Diffraction and High Resolution Crystal Structure Analysis. *Photochem. Photobiol.* **2007**, 83, 336-338.
7. Getzoff, E.; Gutwin, K.; Genick, U. Anticipatory Active-Site Motions and Chromophore Distortion Prime Photoreceptor PYP for Light Activation. *Nature Structural Biology* **2003**, 10(8):663-668.
8. Brudler, R.; Meyer, T.E.; Genick, U.K.; Devanathan, S.; Woo, T.T.; Millar, D.P.; Gerwert, K.; Cusanovich, M.A.; Tollin, G.; Getzoff, E.D.; Coupling of Hydrogen Bonding to Chromophore Conformation and Function in Photoactive Yellow Protein. *Biochemistry* **2000**, 39, 13478-13486.
9. Garcia-Mira, M.M.; Sanchez-Ruiz, J.M.; pH Corrections and Protein Ionization in Water/Guanidinium Chloride. *Biophys. J.* **2001**, 81, 3489-3502.
10. Philip, A.F.; Nome, R.A.; Papadantonakis, G.A.; Scherer, N.F.; Hoff, W.D. Spectral Tuning in Photoactive Yellow Protein by Modulation of the Shape of the Excited State Energy Surface. *Proc. Nat. Acad. Sci.* **2010**, 107, 5821-5826.
11. Sigala, P.A.; Tsuchida, M.A.; Herschlag, D.; Hydrogen Bond Dynamics in the Active Site of Photoactive Yellow Protein. *Proc. Nat. Acad. Sci.* **2009**, 106, 9232-9237.
12. Düx, P.; Rubinstenn, G.; Vuister, G.W.; Boelens, R.; Mulder, F.A.A.; Hård, K.; Hoff, W.D.; Kroon, A.R.; Crielaard, W.; Hellingwerf, K.J.; Kaptein, R. Solution Structure and Backbone Dynamics for the Photoactive Yellow Protein. *Biochemistry* **1998**, 37, 12689-12699.
13. Ufimtsev, I.S.; Martinez T.J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, 4, 222-231.
14. Ufimtsev, I.S.; Martinez T.J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, 5, 1004-1015.

15. Ufimtsev, I.S.; Martinez T.J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. **2009**, *J. Chem. Theory Comput.* 5, 2619-2628.
16. Luehr, N.; Ufimtsev, I.S.; Martinez, T.J. Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs). *J. Chem. Theory Comput.* **2011**, 7, 949-954.
17. Isborn, C.M.; Luehr, N.; Ufimtsev, I.S.; Martinez, T.J. Excited-State Electronic Structure with Configuration Interaction Singles and Tamm–Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units. *J. Chem. Theory Comput.* **2011**, 7, 1814–1823.
18. Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.A.; Cheatham, T.E., III; Darden, T.A.; Duke, R.E.; Gohlke, H.; Goetz, A.W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Paesani, F.; Roe, D.R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.L.; Smith, W.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R.M.; Wu, X.; Kollman, P.A. AMBER 14, **2014**, University of California, San Francisco.
19. 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.
20. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Kollman, P.A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, 115, 9620-9631.
21. Becke, A.D. Density‐Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, 98, 5648-5652.
22. Tawada, Y.; Tsuneda, T.; Yunagisawa, S.; Yinai, T.; Hinao, K. A Long-Range-Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2004**, 120, 8425.
23. Vydrov, O.A.; Scuseria, G.E. Assessment of a Long-Range Corrected Hybrid Functional. *J. Chem. Phys.* **2006**, 125, 234109.

24. Runge, E.; Gross, E.K.U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1986**, 52, 997.
25. Jeffrey, G.A.; Yeon, Y. The Correlation Between Hydrogen-Bond Lengths and Chemical Shifts in Crystals. *Acta Crystallogr. B*. **1986**, 42, 410–413.
26. Walker, R. C.; Crowley, M. F.; Case, D. A. The Implementation of a Fast and Accurate QM/MM Potential Method in Amber. *J. Comp. Chem.* **2008**, 29, 1019-1031.
27. Gotz, A. W.; Clark, M. A.; Walker, R. C. An Extensible Interface for QM/MM Molecular Dynamics Simulations with AMBER. *J. Comp. Chem.* **2014**, 35, 95-108.