Supporting Information for:

"Short Hydrogen Bonds and Proton Delocalization in Green Fluorescent Protein (GFP)"

Luke M. Oltrogge and Steven G. Boxer
Department of Chemistry, Stanford University, Stanford, CA 94305-5012, United States

Contents

S.1 Protein Sequences

All proteins used in this study are based on a circular permutant of Superfolder GFP.[1] The native N- and C-termini are fused with the linker sequence, GGTGGS (highlighted in blue). These constructs were designed by us and synthetically prepared by Genscript. This protein is the same as the one called *ih*:GFP used in Refs. [2] and [3]. "*ih*" indicates that the interior helix bearing the chromophore is the N-terminal structural element.

The residue triad forming the chromophore on the central helix is highlighted in green, Asp148 (on β-strand 7) is highlighted in yellow, the original terminal linker is highlighted in cyan, and the hexa-histidine purification tag is highlighted in gray.

ih:GFP S65T, H148D

MGHHHHHHSSGGKLPVPWPTLVTTLTY[a]GVQCFSRYGTRGSGSIEGRHSGSGSPDHMKR
HDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGH
KLEYNFNSD[b]NVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQTVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYGGTGGSASQGEELFTGVV
PILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTT

a. or halogen substituted Tyr66

b. or His148

S.2 Experimental Methods

Synthesis of Halide-Substituted Tyrosines

Of the four halide-substituted tyrosines used in this study only the 3-chloro-L-tyrosine was commercially available (Sigma Aldrich, St. Louis). The 3,5-dichloro-L-tyrosine, 3-fluoro-L-tyrosine, and 3,5-difluoro-L-tyrosine were all prepared synthetically. 3,5-dichloro-L-tyrosine was synthesized by exhaustive chlorination of L-tyrosine while the fluoro-substituted tyrosines were prepared by enzymatic synthesis from fluoro-substituted phenols.

3,5-dichloro-L-tyrosine synthesis

A 50 mL round-bottom flask was charged with a stirbar and 1.0 g (5.5 mmol) of L-tyrosine (Sigma) suspended in 10 mL of glacial acetic acid. Under nitrogen, 2.0 mL (25 mmol) of sulfuryl chloride was added dropwise and allowed to stir at room temperature for 20 hours.[4] The suspended reaction product was vacuum filtered, washed with cold diethyl ether, and dried. The product was confirmed pure by ^{1}H-NMR and used without further purification.

3-fluoro- and 3,5-difluoro-L-tyrosine synthesis

The plasmid for *Citrobacter freundii* tyrosine phenol lyase (TPL) was kindly provided by Prof. Robert Phillips. The TPL was recombinantly expressed and crudely purified by an ammonium sulfate salt cut using previously described methods.[5] To a beaker containing one liter of 100 mM pyruvic acid, 200 mM ammonium acetate, pH 8.25 buffer were added 13.7 mg pyridoxal-5'-phosphate (the TPL cofactor), 0.35 mL of β-mercaptoethanol, ~20 units of purified TPL, and 10 mmol of either 2-fluoro-phenol or 2,6-difluoro-phenol (Sigma). The reaction was allowed to sit covered overnight at room temperature. At 24 hour intervals, 5 mmol of the appropriate fluoro-phenol was added up to a final total of 30 mmol. The reaction progress was monitored daily by analytical HPLC prior to additions. The reaction mixture was purified on a column of approximately 200 mL of Dowex-50 cation exchange resin (Sigma) equilibrated with 2 M HCl. After washing with deionized water, the product was eluted with 2 M NH_4OH. The eluate was lyophilized and later redissolved in ~10 mL of boiling water before recrystallizing at

4°C. The product was vacuum filtered, washed with ice cold water and ethanol, and dried under vacuum prior to analysis by ^{1}H-NMR.

*Global incorporation of 3-fluoro-L-tyrosine (**F_1Y**)*

No nonsense suppression system currently exists for 3-fluoro-L-tyrosine due to its chemical similarity to L-tyrosine causing cross reactivity with the endogenous tyrosyl-tRNA synthetase. Therefore the incorporation of 3-fluoro-L-tyrosine was performed globally[6, 7]; this introduces fluorine into the chromophore and the other 8 tyrosines, but these have little or no effect on the properties of GFP. BL21(DE3) *E. coli* containing the vector pET-15b with *ih*:GFP C48S, S65T, H148D were grown in one liter of M9 minimal media with 100 mg/L ampicillin at 37°C to a cell density of approximately 0.5 OD_{600}. To the media was added 1 mmol each of L-phenylalanine and L-tryptophan along with 1 mmol of 3-fluoro-L-tyrosine.[a] After 5 min., protein expression was induced with 10 mM isopropyl thiogalactopyranoside (IPTG). The cells were harvested after 4 hours, resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, pH 8.0), and homogenized. The cellular debris was removed by centrifugation and the supernatant purified by hexahistidine-tag nickel affinity chromatography and further by anion exchange chromatography.

*Nonsense suppression non-canonical amino acid incorporation (**Cl_1Y**, **Cl_2Y**, and **F_2Y**)*

In the *ih*:GFP S65T, H148D pET-15b vector Tyr66 was mutated to the amber stop codon (TAG) using a Stratagene Quikchange Lightning mutagenesis kit according to manufacturer protocols. BL21(DE3) *E. coli* were co-transfected with GFP pET-15b vector with ampicillin resistance and each of the required tRNA/aminoacyl tRNA synthetase pEVOL vectors bearing chloramphenicol resistance. The two mono- and dichloro tyrosine plasmids were provided by Prof. Jiangyun Wang (ClYRS and Cl2YRS from Liu *et al.*)[8] while the difluoro-tyrosine plasmid was provided by Prof. Joanne Stubbe (E3 from Minnihan *et al.*).[9]

[a] Phenylalanine and tryptophan were added in order to generally suppress aromatic amino acid metabolism to keep levels of unlabeled tyrosine down.

To one liter of rich LB growth media was added 0.5 mmol of the corresponding substituted tyrosine prior to autoclaving. 5-mL starter cultures grown overnight were transferred to 100 mL of media and allowed to grow for 2 hours at 37°C. This 100-mL culture was transferred into a flask containing the remaining media in addition to IPTG at 0.25 g/L (for GFP induction) and D-arabinose at 0.2 g/L (for tRNA/aminoacyl tRNA synthetase induction) at 25°C. The cells were grown overnight at 25°C and harvested. The protein purification protocol was the same as described above. The purity and identity of all proteins were confirmed with electrospray ionization mass spectrometry.

pH Titrations

All pH titrations were performed using master buffers of 30 mM citrate, 30 mM phosphate, and 200 mM NaCl with small amounts of NaOH added to adjust the pH. Titrations of denatured protein included 6 M guanidinium hydrochloride. In these solutions a correction of +0.72 was applied to the electrode pH reading according to Garcia-Mira and Sanchez-Ruiz.[10] Concentrated solutions of protein were diluted at least 30-fold into the corresponding buffer and allowed to equilibrate for several minutes prior to absorbance measurements.

Equilibrium isotope fractionation

Two buffers were prepared with identical hydron activity with a composition of 30 mM citrate, 30 mM phosphate, 200 mM NaCl, and pH or pD 5.0. The pD was determined using the standard pH electrode correction (*i.e.* $pD = pH_{app} + 0.42$).[11] These two buffers were mixed in the proper ratios to give solutions of 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0 mole fraction deuterium (χ_D). Concentrated protein solutions were diluted more than 30-fold into each of these buffers to give a maximum optical density of about 0.2 at 1 cm pathlength. The absorbance spectra were measured using a Cary 6000i UV/Vis spectrometer.

Protein crystallization

The proteins **Y**, **Cl$_1$Y**, and **Cl$_2$Y** were all crystallized by hanging drop vapor diffusion. **Y** and **Cl$_2$Y** were crystallized under the same conditions with 50 mM sodium acetate, 100 mM NaCl, 5% (wt/vol) PEG 3550, pH 5.0. The conditions for **Cl$_1$Y** were 50 mM sodium acetate, 100 mM NaCl, 15% (w/vol) PEG 3550, pH 5.0. In all cases protein of concentration 10 mg/mL was mixed 1:1 with the mother liquor in the initial drop. All proteins gave initial crystals of rather poor quality which were improved considerably by microseeding. This was performed by crushing the low quality crystal by vortexing with a glass bead, serially diluting the resulting nanocrystals, and using a cat whisker to transfer the microseeds to another set of equilibrated hanging drops prepared with lower precipitant concentrations. The cryoprotectant was the mother liquor with sucrose added to 2 M.

X-ray data collection and structure refinement

The x-ray diffraction data for **Cl$_1$Y** was collected on BL 7-1 at the Stanford Synchrotron Radiation Lightsource (Menlo Park, CA) while the data for **Y** and **Cl$_2$Y** was acquired at BL 5.0.2 at the Advanced Light Source (Berkeley, CA).[b] All data was obtained at 100 K. The data were indexed and reduced with iMosflm[12] and scaled with CCP4.[13] Molecular replacement was carried out in PHENIX[14] using the original S65T, H148D structure (PDB entry: 2DUF) as the search model. Mutations and cycles of model building and refinement were performed with Coot[15] and PHENIX. The substituted chromophore structures and parameters were modified with the Phenix REEL program. To reduce model bias in the location of D148 the sidechain was removed and then modeled back in to the electron density following simulated annealing. The overall fold of the protein, despite being a circular permutant, was virtually identical to those of both Superfolder GFP (PDB ID: 2B3P) and the original S65T, H148D GFP (PDB ID: 2DUF).

[b] Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

^{1}H-NMR spectra of protein

1D ^{1}H-NMR spectra were collected for **Y**, **Cl_1Y**, and **Cl_2Y** using a Varian 800 MHz spectrometer with a cryoprobe. All samples were approximately 300 μM protein with 5% D_2O for locking. Water suppression was performed using the 1-3-3-1 binomial sequence[16] in order to minimize spin saturation for protons in chemical exchange.

S.3 Crystallographic Data

	Y	**Cl_1Y**	**Cl_2Y**
Data Collection			
Space group	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1
Cell dimensions			
a, b, c (Å)	46.37, 68.26, 57.87	51.84, 68.59, 60.88	46.14, 68.13, 58.00
α, β, γ (°)	90.0, 102.2, 90.0	90.0, 100.5, 90.0	90.0, 102.33, 90.0
Total Reflections	136,600	125,462	269,998
Unique Reflections	27,729	37,043	38,429
Resolution (Å)	37.8-1.9	36.3-1.8	56.7-1.7
R_{merge}	0.069	0.107	0.079
$I/\sigma I$	11.0	6.8	9.8
Completeness (%)	99.3	98.6	99.3
Redundancy	4.9	3.4	7.0
Refinement			
Resolution (Å)	1.9	1.8	1.7
R_{work} / R_{free}	0.22/0.26	0.20/0.25	0.20/0.23
No. molecules	2	2	2
No. non-H atoms			
protein	3344	3535	3458
ligand	44	46	48
solvent	26	99	58
B-factors			
protein ($Å^3$)	38.0	22.8	28.5
ligand ($Å^3$)	28.2	18.9	21.2
solvent ($Å^3$)	28.7	21.0	23.5
rmsd bond lengths (Å)	0.008	0.007	0.009
rmsd bond angles (°)	1.15	1.06	1.25

Table S.1

X-ray diffraction data collection and refinement statistics.

protein	*chain*	$r_{O\text{-}O}$ (Å)	ϕ_P (°)	ϕ_I (°)
Y	A	2.41±0.22	12.8	16.8
	B	2.74±0.26	8.5	15.6
Cl_1Y	A	2.41±0.20	24.9	10.0
	B	2.52±0.20	18.6	7.9
Cl_2Y	A[c]	---	9.6	7.3
	B[d]	2.51±0.19	18.6	7.9
S65T, H148D pH 5.6 (PDB ID: 2DUF)[17]	A	2.32	7.2	16.8
S65T, H148D pH 10.0 (PDB ID: 2DUE)[17]	A	---	4.2	5.7
sfGFP[e] (PDB ID: 2B3P)[1]	A	---	0.4	1.4

Table S.2

Chromophore geometric parameters. Distance uncertainties were estimated using the method of Gurusaran *et al.*[18]

[c] This A-chain is deprotonated, not to be confused with the protonated form called the A-state. Asp148 is pointed outward to the solvent thus an $r_{O\text{-}O}$ distance is not given. See section S.6 and Fig. S.5 for details.
[d] This B-chain has a protonated chromophore.
[e] This protein is deprotonated in this structure. Also no $r_{O\text{-}O}$ distance is provided since this is His148.

S.4 Spectroscopic Data

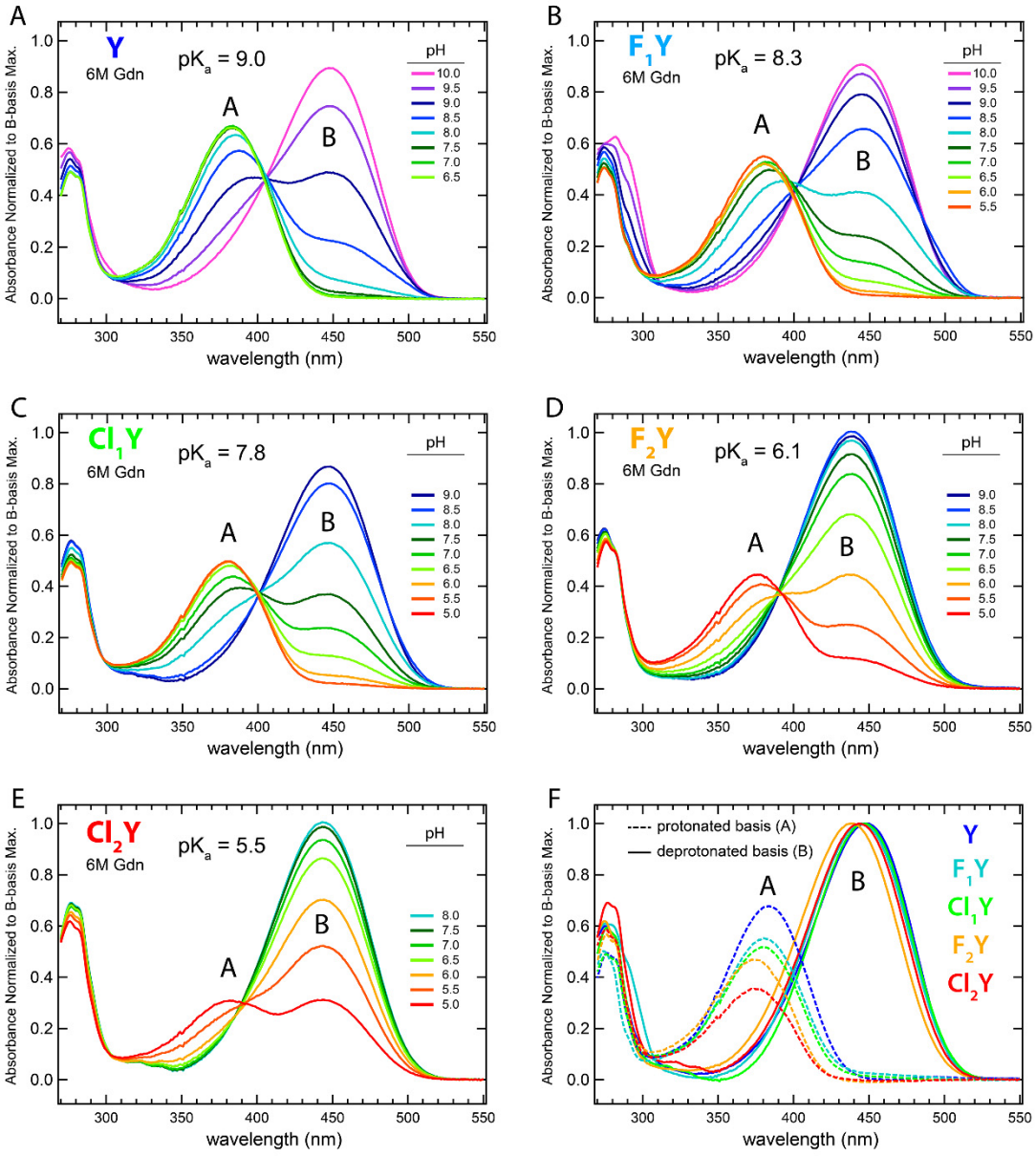

Figure S.1

A-E) UV/Vis pH titrations of substituted GFPs denatured in 6M guanidinium HCl. F) Comparison of denatured protonated and deprotonated basis spectra normalized to the deprotonated maximum absorbance.

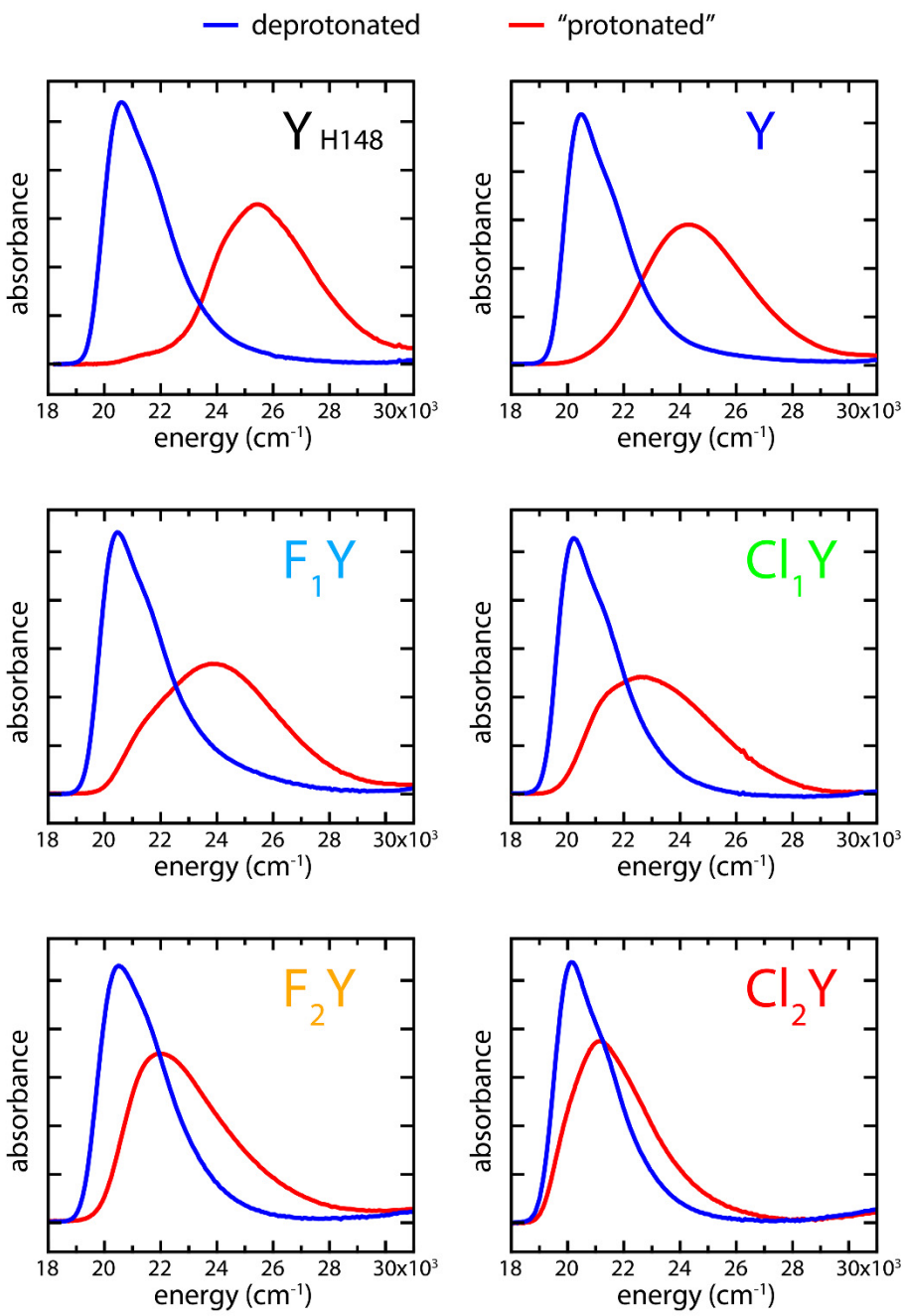

Figure S.2

Absorbance basis spectra for natured protein in low (red, "protonated") and high (blue, deprotonated) pH limits. "Protonated" is in quotes to indicate that rather than simply being protonated the chromophore is instead sharing the proton with Asp148. The extent to which the chromophore owns the proton depends on the relative pK_a.

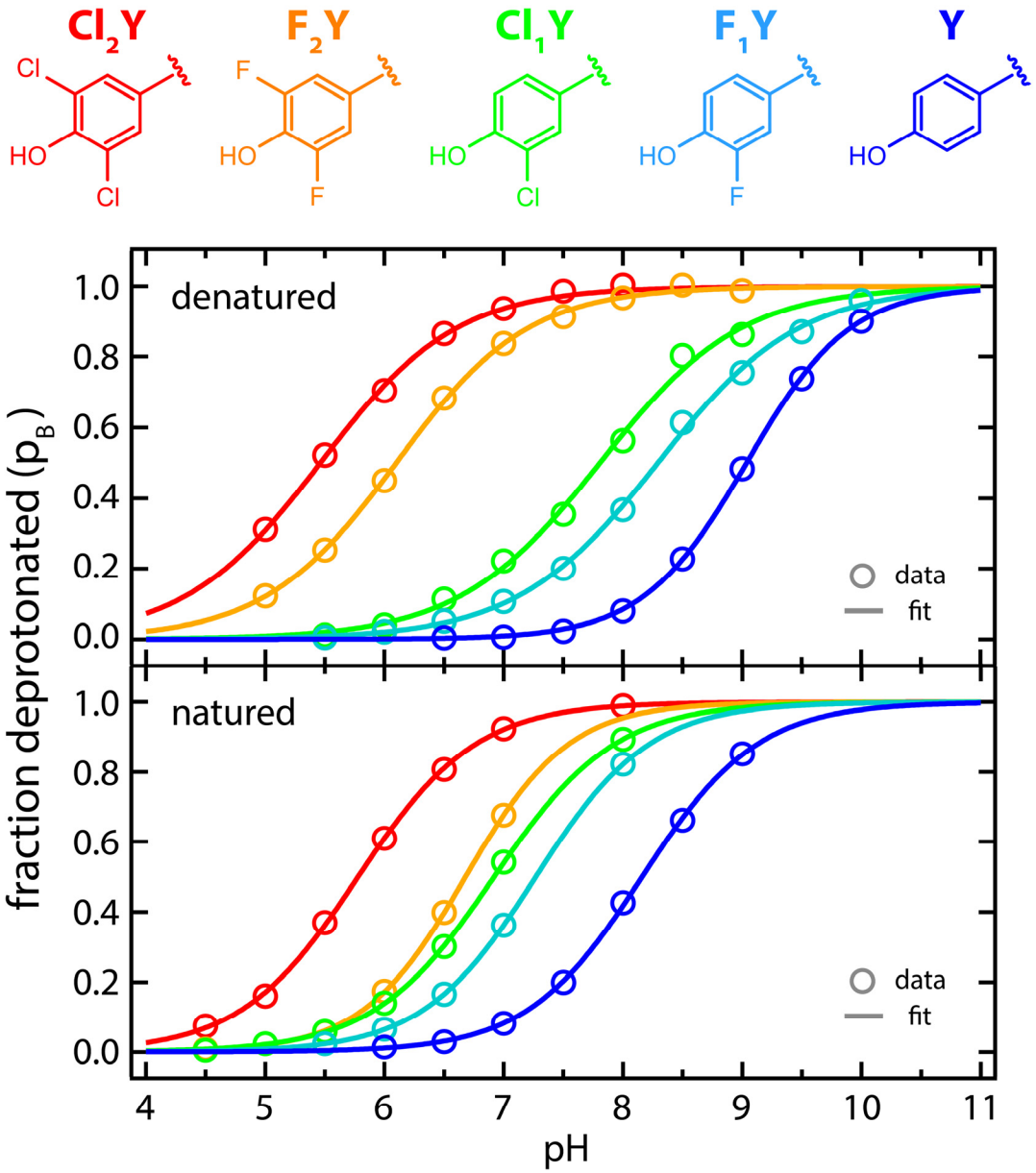

Figure S.3

pH titrations of substituted chromophores under denaturing and naturing conditions. It should be noted that the denatured titrations report directly on the acid dissociation of the substituted chromophore itself, and thus we refer to the resulting pK_a's as the *intrinsic* pK_a's (see Table S.3). In contrast, the natured titrations report on the acid dissociation of the proton shared by the chromophore and Asp148 and thus represents the collective behavior of these strongly coupled sites. This does not tell us anything about the relative affinity of the Asp148 and chromophore proton binding sites with respect to one another.

protein	*denatured* (energy in cm^{-1})				*natured* (energy in cm^{-1})			
	pK_a	A_{max}	B_{max}	ΔE	pK_a	A_{max}	B_{max}	ΔE
Y (H148)	9.0	26,040	22,320	3,720	5.9	25,450	20,620	4,830
Y	9.0	26,040	22,320	3,720	8.2	24,300	20,490	3,810
F_1Y	8.3	26,320	22,470	3,850	7.3	23,810	20,450	3,360
Cl_1Y	7.8	26,320	22,320	4,000	6.9	22,570	20,200	2,370
F_2Y	6.1	26,740	22,780	3,960	6.7	21,980	20,490	1,490
Cl_2Y	5.5	26,740	22,570	4,170	5.8	21,140	20,160	980

Table S.3

pK_a's , peak maxima, and energy differences between A- and B-states (ΔE) for denatured and natured substituted protein species.

S.5 Coupled Morse Potential Model

The basic framework for modeling the PES for the short hydrogen bonds and the effects of chromophore pK_a changes was adapted from that of Ross McKenzie.[19, 20] In this model the PES is generated by coupling two Morse potentials of the form,

Equation S.1
$$V(r) = D_e\left[\exp\left(-2a(r-r_0)\right) - 2\exp\left(-a(r-r_0)\right)\right]$$

where D_e is the bond dissociation energy, r_0 is the equilibrium bond length, and a is a parameter controlling the well width. We use parameters characteristic of a typical O-H bond with D_e = 110 kcal/mol, r_0 = 0.96 Å, and a = 2.12 Å^{-1}. This yields a bond with a 3500 cm^{-1} stretch frequency. The coupling energy we use is that provided by McKenzie for a linear hydrogen bond (*i.e.* with the proton falling on the line between the donor and acceptor),

Equation S.2
$$\Delta_{DA}(R) = \Delta_1 \exp\left(-b\left(R - 2r_0 - \frac{1}{a}\right)\right)$$

where R is the distance between donor and acceptor atoms, r_0 and a are constants as given above for the O-H bond, and Δ_1 and b are fit parameters. In the present work we only consider R = 2.45 Å as determined from crystal structures of the S65T H148D variants so Δ_{DA} is itself just a constant throughout. McKenzie determined that $\Delta_1 = 0.4\ D_e$ and $b = a$ provided reasonable global parameters for hydrogen bonds of type -O··H-O-.[19] Thus the Hamiltonian may be written as,

Equation S.3
$$\hat{H} = \begin{pmatrix} V_{Asp148}(r) & \Delta_{DA} \\ \Delta_{DA} & V_{Cro}(R_{O-O} - r) \end{pmatrix}$$

By diagonalizing Equation S.3 the adiabatic potential energy surfaces are obtained from the eigenvalues with the one corresponding to the ground-state given by,

Equation S.4
$$E(r,R) = \tfrac{1}{2}\left(V_{Asp}(r) + V_{Cro}(R-r)\right) - \tfrac{1}{2}\left(\left(V_{Asp}(r) - V_{Cro}(r-R)\right)^2 + 4\Delta_{DA}(R)\right)$$

where $V_{Asp}(r)$ and $V_{Cro}(R\text{-}r)$ are the Morse potentials corresponding to the proton bound to Asp148 and the chromophore, respectively. In order to account for the pK_a differences in the

various substituted chromophores the dissociation energy (D_e) was adjusted by amounts equal to the differences acid dissociation free energies,

Equation S.5

$$\Delta\Delta G^\circ = \ln 10 * RT\ \Delta pK_a$$

This makes the assumption that ΔΔG° is dominated by the enthalpy difference (ΔΔH°) rather than any contributions from the entropy (ΔΔS°) which is likely near zero for all substituted chromophores due to their structural similarity. The eigenvector associated with the ground-state eigenvalue describes the contribution of each of the diabats to the final PES. We make the approximation that the excitation energy is given by the linear combination of A- and B-state bases weighted according to the relative contribution due to the Asp148 and chromophore diabats. In this formulation a proton bound 100% to the chromophore would result in a pure A-state absorbance while a proton bound 100% to Asp148 would result in a pure B-state absorbance. Note that we specifically are not considering transitions between the adiabatic states obtained from diagonalization of Equation S.3 which would represent local excitation of the bond itself. Rather, the collective behavior of the entire chromophore must be taken into account. The simple model we use cannot encompass this complexity and so we take a semi-classical approach which recognizes the primacy of the protonated state in determining the spectrum and abstracts all other degrees of freedom into this coordinate. We use a normalized energy scale such that the excitation energy of the B-state is arbitrarily set to zero while the A-state is set to one. In this scale the excitation energy is given by,

Equation S.6

$$E_{ex}(r) \cong \frac{v_{Cro}^{\ 2}}{v_{Asp}^{\ 2} + v_{Cro}^{\ 2}} = \frac{1}{1 + \left(\frac{V_{Asp}(r) - V_{Cro}(R-r) - \sqrt{\left(V_{Asp}(r) - V_{Cro}(R-r)\right)^2 + 4\Delta_{DA}(R)}}{2\Delta_{DA}(R)} \right)^2}$$

where v_{Cro} and v_{Asp} are the eigenvector elements associated with the chromophore and Asp148 diabats respectively. This equation provides a mapping function with which to calculate the excitation energy as a function of the proton transfer coordinate (r).

Now with the functional form of the PES and the excitation energy the spectral prediction can be made with a knowledge of the proton (or deuteron) probability distribution. This was

calculated numerically using a finite difference method. Starting with the 1D time-independent Schrödinger equation,

Equation S.7 $$\hat{H}\psi(x) = \frac{-\hbar}{2m}\frac{\psi^2(x)}{dx^2} + V(x)\psi(x) = E\psi(x)$$

with *V(x)* as the coupled Morse potential described above and *m* as the reduced mass for the asymmetric stretch mode of O-H-O or O-D-O, the expression was discretized with $x = n\Delta x$ which makes,

Equation S.8 $$\frac{\psi^2(x)}{dx^2} = \frac{[\psi((n+1)\Delta x) - \psi(n\Delta x)] - [\psi(n\Delta x) - \psi((n-1)\Delta x)]}{\Delta x^2}.$$

With *n* from 1 to *N* the nuclear Hamiltonian can be written as an *N*x*N* matrix,

Equation S.9

$$\hat{H} = \begin{pmatrix} \frac{\hbar}{m\Delta x^2} + V(\Delta x) & \frac{-\hbar}{2m\Delta x^2} & \cdots & 0 \\ \frac{-\hbar}{2m\Delta x^2} & \frac{\hbar}{m\Delta x^2} + V(2\Delta x) & \vdots & \vdots \\ \vdots & \ddots & \frac{\hbar}{m\Delta x^2} + V((N-1)\Delta x) & \frac{-\hbar}{2m\Delta x^2} \\ 0 & \cdots & \frac{-\hbar}{2m\Delta x^2} & \frac{\hbar}{m\Delta x^2} + V(N\Delta x) \end{pmatrix}.$$

The nuclear Hamiltonian matrix was numerically diagonalized in Matlab. The eigenvalues represent the vibrational energy levels and the associated eigenvectors are the corresponding wavefunctions. In order to obtain equilibrium populations, Maxwell-Boltzmann statistics were used to thermally average the probability density.

Equation S.10 $$p(x) = \frac{\sum_{i=1}^{N} |\psi_i(x)|^2 e^{-E_i/k_BT}}{\sum_{i=1}^{N} e^{-E_i/k_BT}}$$

E_i is the energy and $|\psi_i(x)|^2$ is the probability density of the *i*th vibrational eigenstate.

The probability density from Equation S.10 was then used in combination with Equation S.6 to produce a distribution of excitation energies ($p(E_{ex})$). In order to generate a more realistic

spectrum with which to compare directly to experimental data, the calculated excitation energy distribution was convolved with a normalized Gaussian for the final output spectrum, $S(E_{ex})$.

Equation S.11

$$S(E_{ex}) = p(E_{ex}) \otimes \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-E_{ex}^{\,2}}{2\sigma^2}\right)$$

where σ specifies the peak width and was set to 0.289 in order approximate the experimental A- and B-state absorbance band widths.

In final form the model takes as global fit parameters the coupling energy (Δ_{DA}) and the proton binding energy offset whereby the set of substituted chromophore diabats could be collectively adjusted relative to the Asp148 diabat. These parameters together with the constants characteristic of the O-H bond (r_0, a, and D_e) fully specify the model potential energy surfaces from which the spectra are calculated. The workflow for the process is shown schematically in Figure S.4. It should be emphasized that for a given set of global parameters the diabats from the substituted chromophores are identical in every respect except for the small deviations in well depth due to the pK_a offset. The best fit global parameters were calculated using nonlinear least-squares optimization in Matlab. The fit parameters obtained thusly were then used to recalculate the vibrational eigenvalues, probability densities, and spectra with the deuteron reduced mass. The spectral isotope effects were obtained directly from these spectra. The isotope fractionation factor is defined as the equilibrium constant for the H/D exchange reaction with water.[21]

Equation S.12 [f]

$$H\text{-}X + D(L_2O) \rightleftharpoons D\text{-}X + H(L_2O)$$

Equation S.13 [f]

$$\phi_{HD} = \frac{[D-X]}{[H-X]}\left(\frac{H}{D}\right)_{L_2O}$$

We assume that the chemical potential difference for this reaction is given by the differences in zero-point energies for the O-H (O-D) stretch in the complex relative to solvent. From this the isotope fractionation factor may be calculated directly.

[f] L_2O is being used as a shorthand for the particular solvent composition which may contain a mixture of H and D. For example the appearance of $D(L_2O)$ in Equation S.12 indicates that the deuteron in the reaction is obtained from the solvent.

Equation S.14
$$\phi_{HD} = \exp\left(\frac{-1}{k_B}\left(ZPE_{DX} - ZPE_{HX} + ZPE_{H(L_2O)} - ZPE_{D(L_2O)}\right)\right)$$

This treatment is most certainly an overestimation of the magnitude of the effect (*i.e.* the calculated ϕ_{HD} is too small) because it neglects all other degrees of freedom which experience considerably less perturbation. Kreevoy and Liang estimate that in LBHBs the true value for ϕ_{HD} should be ~1.7 times higher.[21] This correction factor puts our observed ϕ_{HD} in near quantitative agreement with the model. Regardless, the primary motivation was to compare trends rather than absolute magnitudes.

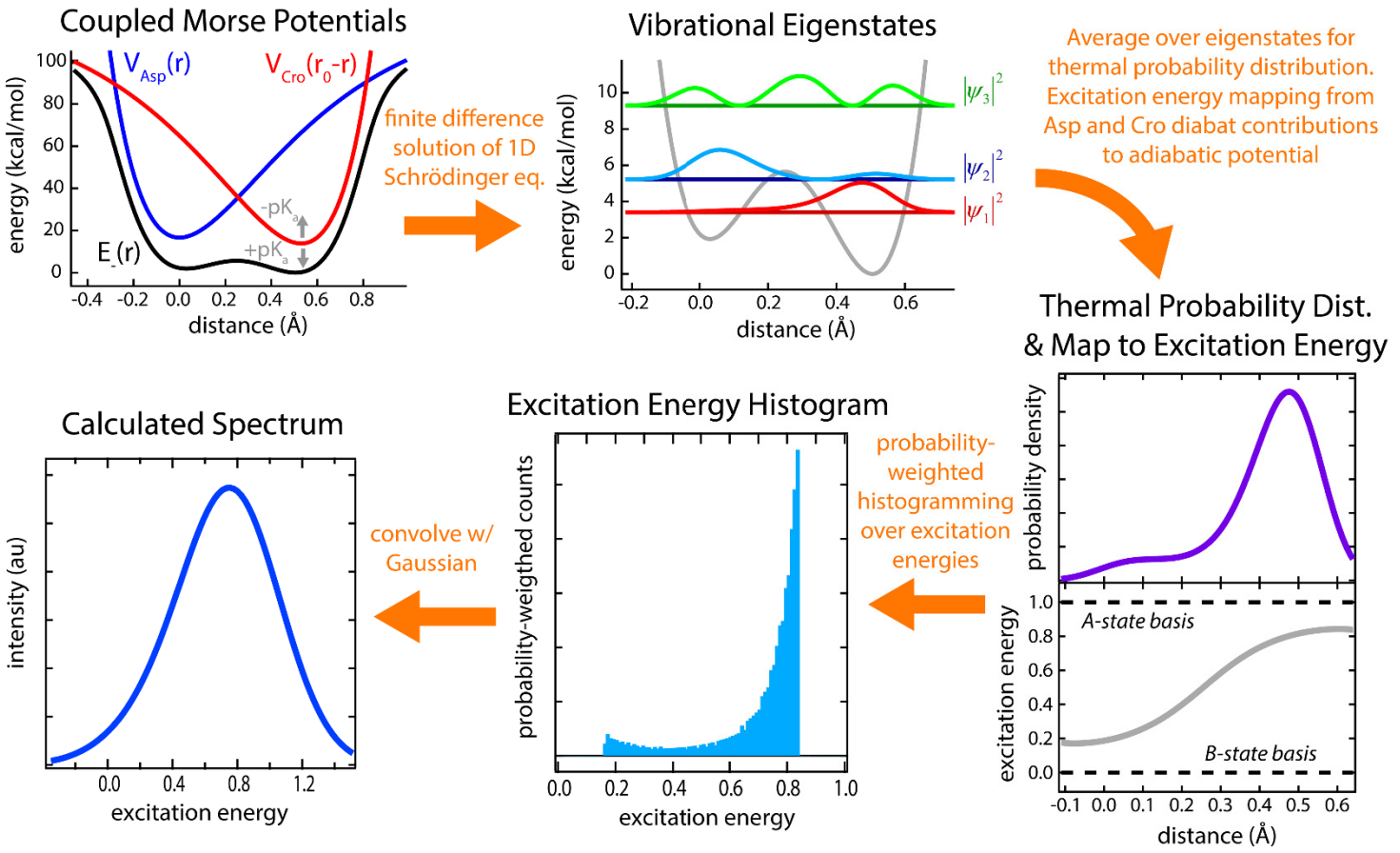

Figure S.4

Workflow for calculating trial absorbance spectra from input coupled Morse potentials.

S.6 Structural Changes Dependent on Ionization State

While this study is concerned primarily with the specific interaction between the chromophore and Asp148 it has also provided new structural information that bears on the issues of ground-state proton transfer and attendant conformational change which will be of interest to those more concerned with properties of GFP itself. In the course of structural characterization of these GFPs, two forms of **Cl_2Y** differing by chromophore ionization state were found in the same crystal. Analysis of the differences between these structures brings some insight into the structural changes accompanying proton transfer and also a plausible mechanistic explanation for the very slow (> 10 ms) chemical exchange rates between these states.

UV/Vis absorbance spectra from pH titrations of natured protein of all S65T, H148D mutants (including **Y**, **F_1Y**, **Cl_1Y**, **F_2Y**, **Cl_2Y**, and S65T, H148D from Shu *et al.*[17]) have found clean isosbestic points between a deprotonated state at high pH with absorbance peaked around 490 nm and another peak at low pH with a variable absorbance maximum dependent on the chromophore pK_a (see Figure S.2 and Table S.3). As discussed above and in the text this state is hypothesized to be a result of proton delocalization across a short hydrogen bond. The behavior of this state is *not* pH-dependent, however, the transitions between this state and the deprotonated state are pH-dependent.

The proximity of Asp148 to the chromophore presents a structural difficulty in the deprotonated state since it would entail the energetically prohibitive formation of two anions within 2.5 Å of one another. It is reasonable therefore to suspect that the chromophore deprotonation is achieved either by some pH-dependent modification of the environment which increases the proton affinity of Asp148 and thus allows complete proton transfer, or that there is a significant structural change that increases the distance to accommodate the two anions. Our data clearly support this latter hypothesis. The evidence for this claim is two-fold with NMR spectra providing evidence for slow inferred dynamics and x-ray crystallography providing an atomistic perspective on the mechanism.

In previous work we utilized an NMR method to follow proton transfer in GFP by introducing a ^{13}C label with high sensitivity to the chromophore ionization state.[3] The resulting

1D ^{13}C-NMR spectra contain information on both the thermodynamics and kinetics of this process. We found that β-strand 7 (residue 148 is in the middle of strand 7) plays an important role in setting the timescale of proton transfer to and from the solution and that this process can be quite slow (>10 ms) thus suggesting large-scale structural motion. We employed the same methodology to **Y** and found very slow chemical exchange between protonation states (Figure S.5). However, in addition, we have been able to complement this data with structural information on the closely related **Cl$_2$Y** protein.

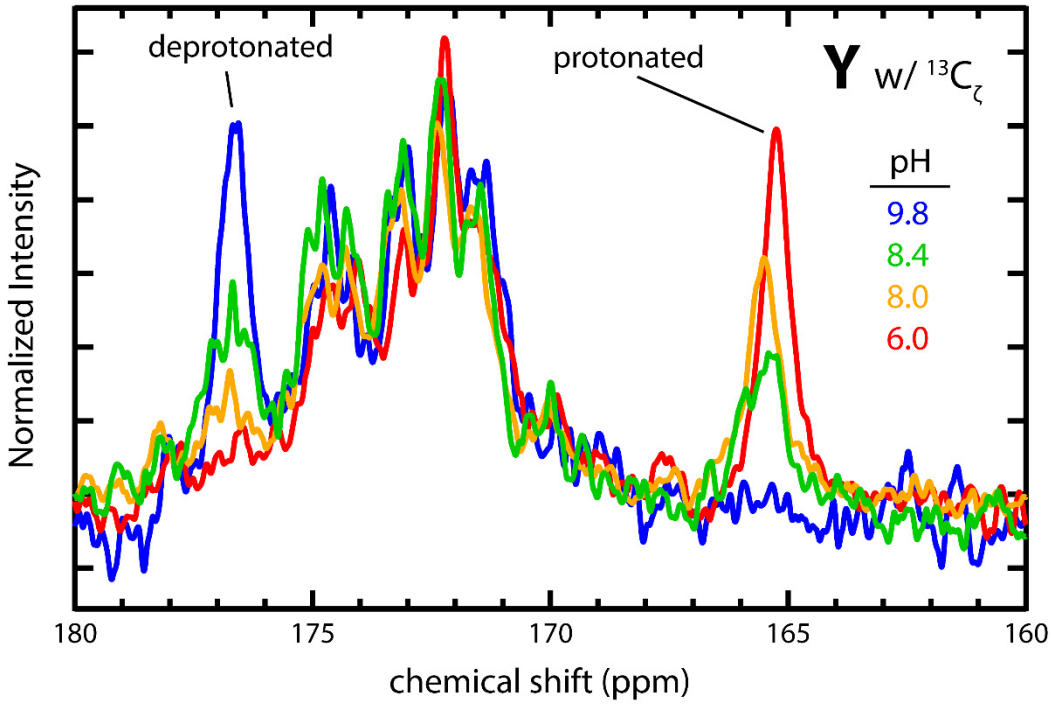

Figure S.5

1D ^{13}C-NMR of **Y** for a pH titration series. The presence of two very clear peaks implies that the chemical exchange rate is far less than the inverse frequency difference between the two basis states. $\tau >> \frac{1}{|\nu_B - \nu_A|}$. These spectra were collected on a 300 MHz NMR (75 MHz for ^{13}C) such that the relevant reference timescale is 11.6 ms. The broad cluster of peaks between 170 and 175 ppm is due to backbone carbonyls at natural abundance.

Shu *et al.* solved crystal structures for GFP S65T, H148D in conditions of pH 5.6 (PDB ID: 2DUF) and pH 10.0 (PDB ID: 2DUE) for pure populations of the two observed optical

states.[17] The pH 5.6 structure was the one first demonstrating the close interaction between the chromophore and Asp148. The pH 10.0 structure, however, had poorly ordered electron density in the region of strand 7 containing Asp148 and could not be modeled. This was the first evidence that significant structural changes might be accompanying changes of ionization state.

In the present work all proteins were crystallized at pH 5.0 in the spacegroup P 1 2_1 1 and had a unit cell comprised of two GFP protein chains. For **Y** and **Cl_1Y** chains A and B were essentially identical. However, the crystal of **Cl_2Y** was fortuitously such that chain A had a deprotonated chromophore and chain B had a protonated chromophore (or more accurately a proton shared between the chromophore and Asp148). The measured solution pK_a for natured **Cl_2Y** was 5.8. At pH 5.0 one would expect there to be a relatively significant fraction of the deprotonated form of the chromophore, however due to increased protein destabilization and denaturation at low pH we were unable to push to lower pH for a more homogenous population. Ordinarily this scenario would be expected to lead to static disorder in the crystal. In this case however, presumably aided by the very slow chemical exchange process deduced from NMR, the crystal formed in such a way that alternating protein chains were in the two different conformations enabling simultaneous solution of both structures.

The resolution of the **Cl_2Y** structure at 1.7 Å is not sufficient to directly observe hydrogen atoms, therefore the positions of protons must be indirectly inferred. A number of factors informed the assignment. First, as mentioned above, the proximity of anions that a deprotonated assignment to chain B would necessitate creates an energetically unrealizable situation such that we may be confident this chain has a proton shared between these groups. Conversely, the conformational change in chain A which relocates Asp148 to the exterior of the protein would alleviate this destabilization and allow both positions to carry a negative charge. Secondly, the putative deprotonated chromophore (chain A) has three potential hydrogen bonding partners within 2.9 Å (Ser147, Thr203, and an ordered water) while the putative protonated chromophore (chain B) has just two (Asp148 and Thr203). This situation is consistent with that from other studies on the structural changes upon chromophore titration.[22] Lastly, chain A has a chromophore structure which is much more planar than the chain B chromophore. This is also an observation in accord with previous experimental measurements[10, 14] and theoretical calculations concerning the chromophore deformability.[17, 22, 23]

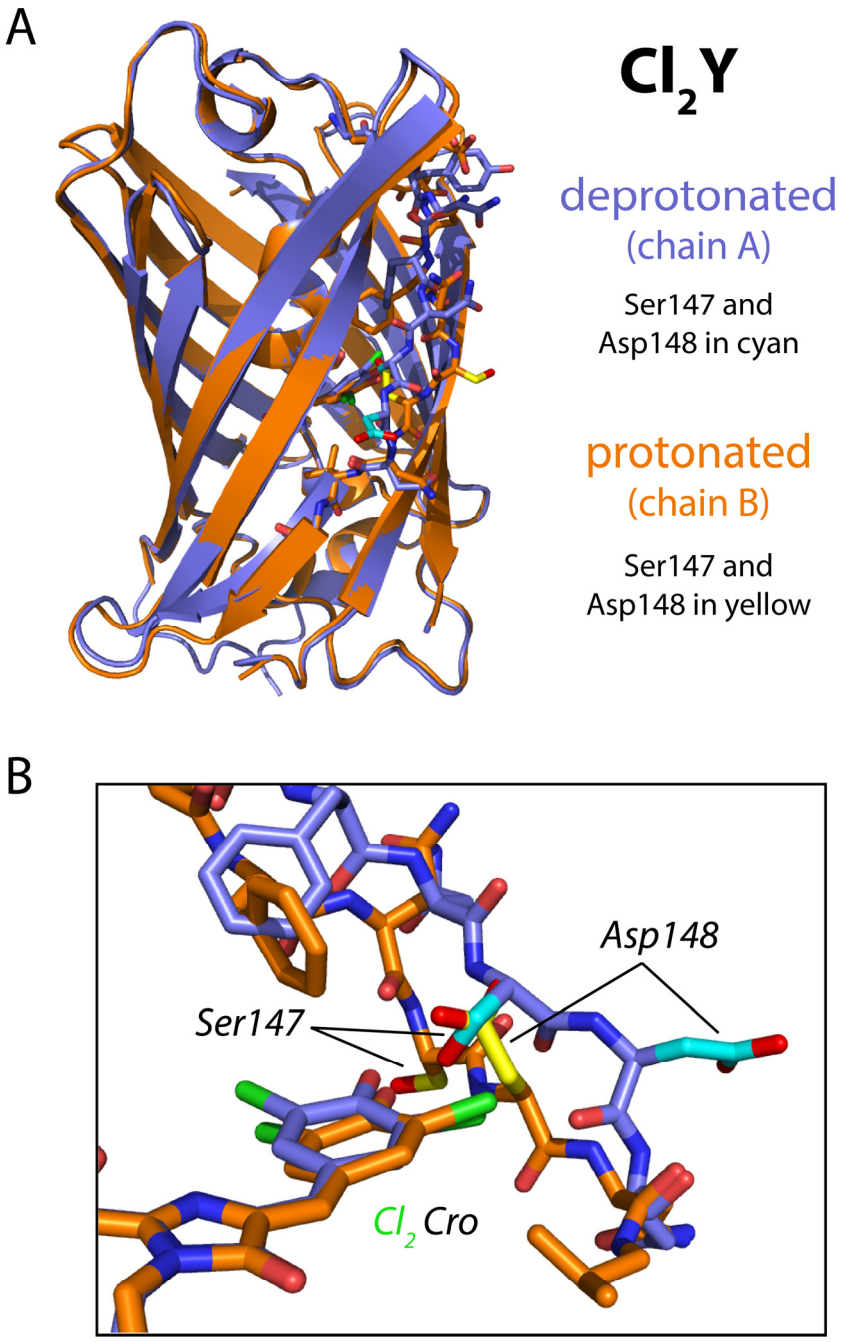

Figure S.6

A) Cartoon structure of aligned structures of chain A (deprotonated) and chain B (protonated) of **Cl_2Y** with the residues of greatest deviation shown in stick form. B) A close-up of the dichloro-chromophore and the positional changes of the key residues Ser147 and Asp148 on strand 7.

Alignment of the two **Cl_2Y** chains in PyMOL (Schrödinger, LLC) reveals that the structural transition linking the two conformations is largely localized to β-strand 7 (see Figure S.6). The two chains were essentially superimposed elsewhere having a backbone RMSD of 0.238 Å excluding strand 7. The significant deviations begin at position 143 and continue on to 149 with C_α distances reaching a maximum difference of 4.4 Å at Ser147. The most important functional difference between the two conformations is the swap of residues interacting with chromophore. In the protonated structure (chain B) Asp148 is participating in the short hydrogen bond with the chromophore while Ser147 is directed to the protein exterior. Upon deprotonation (chain A) the Asp148 sidechain has a net motion of 8.6 Å for the interacting oxygen which positions it on the exterior while the Ser147 hydroxyl group has a net motion of 6.0 Å as it establishes a hydrogen bond to the chromophore phenolate (see Figure S.6B).

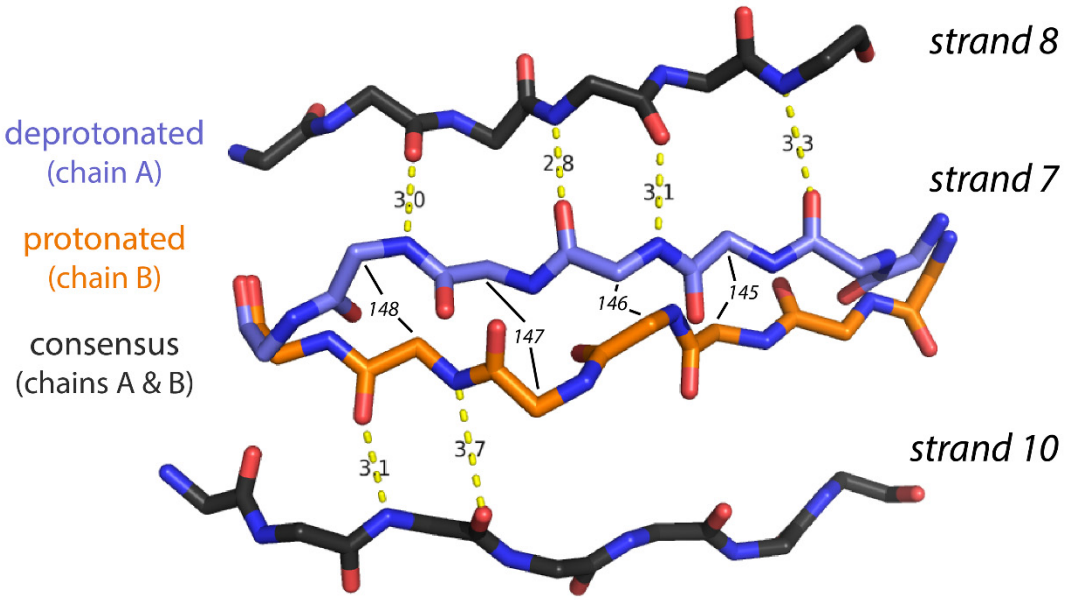

Figure S.7

Backbone representation of the structural changes occurring between protonated (chain B) and deprotonated (chain A) structures. The regions of consensus structure are superimposable between the two chains. Backbone hydrogen bonds are indicated in yellow with the N-O distances given in Ångstroms. Residue C_α positions on strand 7 are indicated.

Uniquely among the 11 β-strands comprising the GFP structure, strand 7 possesses a conventional anti-parallel β-sheet for only about half the length of barrel. Beyond this point the backbone is twisted and, instead of interacting with the adjacent strands, satisfies backbone amide and carbonyl hydrogen bonds with sidechains, ordered water, or bulk solvent. The bordering β-strands, 8 and 10, are actually separated by a distance which makes it impossible for strand 7 to simultaneously engage in backbone hydrogen bonding with both. These factors have been implicated as giving rise to dynamic structural heterogeneity of strand 7 in solution computationally[24, 25] and also through NMR experiments[26] including H/D exchange.[27] It is interesting therefore to examine the structural change accompanying deprotonation of **Cl_2Y** in this context. The protonated form (chain B) has very close structural homology to Superfolder GFP (PDB ID: 2B3P) and S65T, H148D GFP (PDB ID: 2DUF) and the backbone conformation is essentially superimposable on both of these structures. The deprotonated **Cl_2Y** structure (chain A), however, is to our knowledge unique. The exclusion of Asp148 from the protein interior requires considerable changes in the backbone geometry. Unexpectedly this accommodation resulted in the establishment of four additional backbone hydrogen bonds to strand 8 relative to the common strand 7 conformation which makes none of these contacts but rather shares two backbone hydrogen bonds to strand 10 (Figure S.7).

The preceding discussion has been focused primarily on H148D. We hypothesize, however, that structural transitions involving this toggling between strands 8 and 10 may be a general phenomenon in fluorescent proteins. Beyond the role of gating proton transfer between the solution and chromophore this structural mechanism could be important to the large class of GFP-based sensors which incorporate sensing domains inserted within strand 7.

S.7 ^{1}H-NMR in search of LBHBs

One of the most common experimental observables used to identify LBHB's in proteins is the presence of ^{1}H-NMR peaks occurring abnormally far downfield (~12 – 20 ppm).[28-30] This method was used on **Y** and **Cl$_1$Y** and no diagnostic downfield proton resonances were observed.[g] Using the empirical correlation between O-O distance and proton chemical shift from Berglund and Vaughan we would have expected peaks around 18 ppm for the measured distance of ~2.45 Å.[28] A possible explanation for the missing peaks is that there is rapid proton exchange with water on the NMR timescale. In this event the peak of interest would disappear into the solvent peak provided that chemical exchange were taking place in approximately a millisecond or less. This result may be related to the observation of Stoner-Ma *et al.*[31] that no vibrational signature of neutral Asp148 could be detected following ultrafast ESPT in visible pump/ IR probe experiments on GFP S65T, H148D whereas such a signal was clearly seen for Glu222 in wtGFP.[32] Perhaps the proton is efficiently shuttled through an alternative relay which may also be operative for proton exchange in the ground state. Alternatively, it is possible that, on account of a smaller than expected coupling energy, the O-H distance is closer to its normal length and thus results in a peak that is further upfield and obscured by the protein background. **Cl$_1$Y** is most closely affinity matched and therefore should be expected to have the furthest downfield proton resonance. However, our model for **Cl$_1$Y** finds an O-H bond lengthening of about 0.07 Å relative to r_0 [h] which is consistent with empirical findings of Berglund and Vaughan and should still give rise to a downfield peak[28] so we favor the hypothesis of rapid exchange to solvent.

[g] ^{1}H-NMR was also performed on GFP S65T, H148D and likewise no unusual downfield proton resonance was observed.[17]

[h] These length change was calculated by using a value for the bond length determined by the maximum in the proton probability density.

S.8 References

(1) Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat. Biotechnol. 24*, 79-88.

(2) Kent, K. P., Oltrogge, L. M., and Boxer, S. G. (2009) Synthetic Control of Green Fluorescent Protein, *J. Am. Chem. Soc. 131*, 15988-15989.

(3) Oltrogge, L. M., Wang, Q., and Boxer, S. G. (2014) Ground-State Proton Transfer Kinetics in Green Fluorescent Protein, *Biochemistry 53*, 5947-5957.

(4) McCubbin, J. A., Maddess, M. L., and Lautens, M. (2008) Enzymatic Resolution of Chlorohydrins for the Synthesis of Enantiomerically Enriched 2-Vinyloxiranes, *Synlett 2008*, 289-293.

(5) Chen, H., Gollnick, P., and Phillips, R. S. (1995) Site-Directed Mutagenesis of His343→Ala in Citrobacter freundii Tyrosine Phenol-Lyase, *Eur. J. Biochem. 229*, 540-549.

(6) Brooks, B., and Benisek, W. F. (1994) Mechanism of the Reaction Catalyzed by .DELTA.5-3-Ketosteroid Isomerase of Comamonas (Pseudomonas) testosteroni: Kinetic Properties of a Modified Enzyme in Which Tyrosine 14 Is Replaced by 3-Fluorotyrosine, *Biochemistry 33*, 2682-2687.

(7) Brooks, B., Phillips, R. S., and Benisek, W. F. (1998) High-Efficiency Incorporation in Vivo of Tyrosine Analogues with Altered Hydroxyl Acidity in Place of the Catalytic Tyrosine-14 of Δ5-3-Ketosteroid Isomerase of Comamonas (Pseudomonas) testosteroni: Effects of the Modifications on Isomerase Kinetics, *Biochemistry 37*, 9738-9742.

(8) Liu, X., Jiang, L., Li, J., Wang, L., Yu, Y., Zhou, Q., Lv, X., Gong, W., Lu, Y., and Wang, J. (2014) Significant Expansion of Fluorescent Protein Sensing Ability through the Genetic Incorporation of Superior Photo-Induced Electron-Transfer Quenchers, *J. Am. Chem. Soc. 136*, 13094-13097.

(9) Minnihan, E. C., Young, D. D., Schultz, P. G., and Stubbe, J. (2011) Incorporation of Fluorotyrosines into Ribonucleotide Reductase Using an Evolved, Polyspecific Aminoacyl-tRNA Synthetase, *J. Am. Chem. Soc. 133*, 15942-15945.

(10) Garcia-Mira, M. M., and Sanchez-Ruiz, J. M. (2001) pH corrections and protein ionization in water/guanidinium chloride, *Biophys. J. 81*, 3489-3502.

(11) Glasoe, P. K., and Long, F. A. (1960) Use of glass electrodes to measure acidities in deuterium oxide, *The Journal of Physical Chemistry 64*, 188-190.

(12) Leslie, A. G. W., and Powell, H. R. (2007) Processing Diffraction Data with Mosflm, In *Evolving Methods for Macromolecular Crystallography*, pp 41-51.

(13) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallographica Section D 67*, 235-242.

(14) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica Section D 66*, 213-221.

(15) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica Section D 66*, 486-501.

(16) Hors, P. J. (1983) A new method for water suppression in the proton NMR spectra of aqueous solutions, *Journal of Magnetic Resonance (1969) 54*, 539-542.

(17) Shu, X., Kallio, K., Shi, X., Abbyad, P., Kanchanawong, P., Childs, W., Boxer, S. G., and Remington, S. J. (2007) Ultrafast excited-state dynamics in the green fluorescent protein variant S65T/H148D. 1. Mutagenesis and structural studies, *Biochemistry 46*, 12005-12013.

(18) Gurusaran, M., Shankar, M., Nagarajan, R., Helliwell, J. R., and Sekar, K. (2014) Do we see what we should see? Describing non-covalent interactions in protein structures including precision, *IUCrJ 1*, 74-81.

(19) McKenzie, R. H. (2012) A diabatic state model for donor-hydrogen vibrational frequency shifts in hydrogen bonded complexes, *Chem. Phys. Lett. 535*, 196-200.

(20) McKenzie, R. H., Bekker, C., Athokpam, B., and Ramesh, S. G. (2014) Effect of quantum nuclear motion on hydrogen bonding, *J. Chem. Phys. 140*, -.

(21) Kreevoy, M. M., and Liang, T. M. (1980) Structures and isotopic fractionation factors of complexes, A1HA2, *J. Am. Chem. Soc. 102*, 3315-3322.

(22) Elsliger, M. A., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (1999) Structural and spectral response of green fluorescent protein variants to changes in pH, *Biochemistry 38*, 5296-5301.

(23) Olsen, S., Lamothe, K., and Martínez, T. J. (2010) Protonic Gating of Excited-State Twisting and Charge Localization in GFP Chromophores: A Mechanistic Hypothesis for Reversible Photoswitching, *J. Am. Chem. Soc. 132*, 1192-1193.

(24) Shinobu, A., and Agmon, N. (2015) The Hole in the Barrel: Water Exchange at the GFP Chromophore, *J. Phys. Chem. B 119*, 3464-3478.

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

(26) Seifert, M. H. J., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B., and Holak, T. A. (2003) Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution, *Biochemistry 42*, 2500-2512.

(27) Huang, J.-r., Craggs, T. D., Christodoulou, J., and Jackson, S. E. (2007) Stable Intermediate States and High Energy Barriers in the Unfolding of GFP, *J. Mol. Biol. 370*, 356-371.

(28) Berglund, B., and Vaughan, R. W. (1980) Correlations between proton chemical shift tensors, deuterium quadrupole couplings, and bond distances for hydrogen bonds in solids, *J. Chem. Phys. 73*, 2037-2043.

(29) Jeffrey, G. A., and Yeon, Y. (1986) The correlation between hydrogen-bond lengths and proton chemical shifts in crystals, *Acta Crystallographica Section B 42*, 410-413.

(30) Frey, P., Whitt, S., and Tobin, J. (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases, *Science 264*, 1927-1930.

(31) Stoner-Ma, D., Jaye, A. a., Ronayne, K. L., Nappa, J., Meech, S. R., and Tonge, P. J. (2008) An alternate proton acceptor for excited-state proton transfer in green fluorescent protein: rewiring GFP, *J. Am. Chem. Soc. 130*, 1227-1235.

(32) Stoner-Ma, D., Jaye, A. A., Matousek, P., Towrie, M., Meech, S. R., and Tonge, P. J. (2005) Observation of excited-state proton transfer in green fluorescent protein using ultrafast vibrational spectroscopy, *J Am Chem Soc 127*, 2864-2865.