Supporting Information for:

"Ground-state proton transfer kinetics in Green Fluorescent Protein (GFP)"

Luke M. Oltrogge, Quan Wang, 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 the sequence of Superfolder GFP (sfGFP) which was used for all experiments due to its robust stability, high expression yield, and tolerance to circular permutation. These proteins differ from wtGFP by about a dozen mutations selected largely for improved folding properties, and they do not appreciably affect the fluorescence properties which remain nearly identical to wild-type.[1] The circular permutants have all had the native N- and C-termini fused with the linker sequence, GGTGGS. These constructs were designed by us and synthetically prepared by Genscript.

In all cases below the residue triad forming the chromophore (on *ih*) is highlighted in green, His148 (on *s7*) is highlighted in yellow, the original terminal linker is highlighted in cyan, and mutation locations are indicated with gray highlight and superscripts.

ih:GFP

MGHHHHHHSSGGKLPVPWPTLVTTLS^{a}YGVQCFSRYGTRGSGSIEGRHSGSGSPDHMKR
HDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGH
KLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQTV^{b}LSKDPNEKRDHMVLLE^{c}FVTAAGITHGMDELYGGTGGSASQGEELFTGV
VPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTT

a. S65T, **b.** V206K, **c.** E222Q

s7:GFP

MGHHHHHHSSGNSHNVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQTV^{a}LSKDPNEKRDHMVLLE^{b}FVTAAGITHGMDELYGGTGGSASQGE
ELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTTGKLPVPWPTLVTTLS^c
YGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRI
ELKGTDFKEDGNILGHKLEYNF

a. V206K, **b.** E222Q, **c.** S65T

s7+:GFP

MGHHHHHHRSSGKLEYNFNS**H**NVYITADKQKNGIKANFTVRHNVEDGSVQLADHYQQ
NTPIGDGPVLLPDNHYLSTQTKLSKDPNEKRDHMVLLE[a]FVTAAGITHGMDELY**GGTGG**
SASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTTGKLPVPWPT
LVTTL**TYG**VQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVKFEGD
TLVNRIELKGTDFKEDGNILGH

a. E222Q

s10:GFP

MGSSHHHHHHSSGLVPGGSHMLPDNHYLSTQTVLSKDPNEKRDHMVLHEYVNAAGIT
HGMDELY**GGTGGS**ASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKF
ISTTGKLPVPWPTLVTTL**SYG**VQAFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGK
YKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNS**H**NVYITADKQKNGIKAN
FTVRHNVEDGSVQLADHYQQNTPIGDGPVL

s1:GFP *(un-permuted)*

MGSSHHHHHHSSGLVPRGSHMGGTSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGE
GDATIGKLTLKFICTTGKLPVPWPTLVTTL**SYG**VQCFSRYPDHMKRHDFFKSAMPEGYV
QERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNS**H**NVYI
TADKQKNGIKANFTVRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKD
PNEKGTRDHMVLHEYVNAAGIT

S.2   Determination of Protonation Equilibrium by UV/Vis Absorbance

In order to estimate the fractional ionization of the chromophore at arbitrary pH it was necessary to establish the intrinsic extinction coefficients for the protonated and deprotonated basis states. Throughout this text we make the assumption that both neutral and both ionized forms (A, A′ and B, B′ from Fig. 6) have identical absorbance spectra, and, moreover, that there is no significant pH-dependence to either basis state. This approximation of only two optical states is well supported by the generally clean isosbestic points (*e.g.* Fig. 2B). The deviations that do exist are primarily due to partial denaturation at the extrema of pH.

In general the spectrum may be expressed as a linear combination of basis spectra using the Beer-Lambert law as,

$$(Eq. S1) \qquad\qquad S(\lambda) = bc\left[(1 - p_B)S_A(\lambda) + p_B S_B(\lambda)\right]$$

where b is the pathlength in cm, c is the molar concentration, p_B is the fraction ionized, and $S_A(\lambda)$ and $S_B(\lambda)$ are the basis spectra for the A and B states in units of $M^{-1}cm^{-1}$.

The basis spectra can be obtained in a straightforward way for the model titratable protein *ih*:GFP S65T. This protein features a pH titration well approximated by a single-site model giving an apparent pK_a of ~5.9. Consequently, at pH 8.0 and higher the protein has >99% B-state and thus directly provides $S_B(\lambda)$. Additionally, the A-state has negligible absorbance at the B-state maximum (~485 nm) so p_B may be accurately estimated within the pH titration from this relative absorbance. The integrated fluorescence emission from 485 nm excitation follows exactly the same pH trend as the absorbance and further confirms the absence of A-state absorbance at this wavelength. Together this information enables the calculation of the A-state basis spectrum, $S_A(\lambda)$. The scaling of $S_A(\lambda)$ and $S_B(\lambda)$ to extinction coefficients was performed

using the method of Ward *et al.* with 0.1 M NaOH denaturation and the known extinction coefficient of the denatured deprotonated chromophore of 44,100 $M^{-1}cm^{-1}$ at 447 nm.[2]

In practice, the concentration (*c*) and fraction deprotonated (p_B) were often determined using a two point calculation from the absorbances at the A-state and B-state maxima (~395 nm and ~485 nm, respectively).
(Table S1)

	ε_{395} ($M^{-1}cm^{-1}$)	ε_{485} ($M^{-1}cm^{-1}$)
A-state	30,700	0
B-state	4,600	57,200

After obtaining an experimental spectrum, $S(\lambda)$, *c* and p_B were calculated by simultaneous solution of Eq. S1 at 395 nm and 485 nm. The analytical equations for this procedure are,

(Eq. S2)
$$c = \frac{S(485)[S_A(395) - S_B(395)] + S(395)[S_B(485) - S_A(395)]}{S_A(395)S_B(485) - S_A(485)S_B(395)}$$

(Eq. S3)
$$p_B = \frac{S(485)S_A(395) - S(395)S_A(485)}{S(485)[S_A(395) - S_B(395)] + S(395)[S_B(485) - S_A(395)]}$$

S.3   FCS Experiments and Fitting

The autocorrelation functions (ACFs) were fit to a model containing three decay components: 1) the diffusional lifetime, 2) the light-dependent dark state formation—likely a triplet state, and 3) the pH-dependent protonation/deprotonation chemical exchange, which is the quantity we seek.

It was experimentally determined through measurements of a sub-diffraction sized fluorescent bead translated through the focal volume on a piezo-electric stage that the detection volume was well approximated by a Gaussian ellipsoid point spread function with $r_0 \sim 0.4\mu m$ and $z_0 \sim 1.6\mu m$. With this standard assumption for the focal volume the complete expression for the autocorrelation function (ACF) is,

$$(Eq. S4) \qquad ACF(t) \propto \left(1 - F_{dark} + F_{dark}e^{\frac{-t}{\tau_{dark}}} \right) * \left(1 - F_{prot} + F_{prot}e^{\frac{-t}{\tau_C}} \right) * \left(\frac{1}{1 + \frac{4Dt}{r_0^2}} \right) * \left(\frac{1}{1 + \frac{4Dt}{z_0^2}} \right)^{\frac{1}{2}}$$

where F_{dark} is the fraction of the population driven into a dark state with lifetime τ_{dark} at a given light intensity, F_{prot} is the equilibrium fraction of the population in the dark protonated state with apparent chemical exchange lifetime τ_C, D is the diffusion coefficient, and r_0 and z_0 describe the focal volume dimensions. For a single-site titration model the expression for τ_C can be written as a function of pH and the rate constants,

$$(Eq. S5) \qquad\qquad\qquad \tau_C(pH) = \frac{1}{k_{deprot} + k_{prot}10^{-pH}}$$

The protein samples were measured for a range of different pH values. Most trajectories were recorded with an irradiance of 10kW/cm^2, however, some were also measured with 2.5kW/cm^2 in order to assess the effect of the light intensity. It was observed that the pH-dependent flickering was unaffected by this change but that the short timescale decay was lessened in magnitude but unchanged in rate. This evidence is consistent with its tentative assignment to a dark triplet state.

In practice, the fitting was performed by first measuring the sample under high pH conditions in which the chromophore is always deprotonated and using only the diffusional and light-dependent triplet parts of the calculated ACF in the fit. The data were normalized at the ACF plateau occurring at one microsecond. We then assumed that the protein retains a similar diffusion coefficient and photophysical properties and fixed these parameters while varying only the chemical exchange parameters for the samples at all other pH values. Using the expression for τ_C from Eq. S5 we used k_{deprot} and k_{prot} as global fit parameters for a non-linear least squares optimization in Matlab.

S.4 $^{13}C_\zeta$-Tyrosine Spectrum and Normalization Procedure

All of the protein variants contained nine tyrosine residues. The $^{13}C_\zeta$ label was incorporated into all of these as confirmed by a shift of +9 Da observed in intact protein electrospray ionization mass spectrometry (data not shown). In small unstructured peptides the $^{13}C_\zeta$ resonances of Tyr have been measured to occur at 155.5 ppm in the neutral form and 166.3 ppm in the ionized form.[3] The cluster of peaks between 154 and 158 ppm shown in Fig. S1 arises from the eight protonated tyrosines in the protein. The ninth tyrosine is Tyr66 which is incorporated into the chromophore has its chemical shift increased to 162.0 ppm in the protonated form and 177.1 ppm in the deprotonated form by virtue of the extended electronic conjugation. The tyrosine cluster showed only a small dependence on pH through the titration series due the relatively high pK_a of all tyrosines (~10).[4]

All NMR spectra were normalized by the same method using the protonated tyrosine peaks. The integrated area underneath these peaks was normalized to eight. Our choice to use the tyrosine $^{13}C_\zeta$ peaks for normalization—as opposed to some other feature such as the aliphatic region—was based on the large intensity due to these peaks (see Fig. S2) and the fact that their abundance should be directly related to the chromophore peak intensity arising from the incorporation of the same label and thus insensitive to variations in labeling efficiency. That the backbone carbonyl resonance bands overlap nicely in Fig. 3 A&C and Fig. 4 B&C reflects the fact that the isotope labeling was reproducible and essentially complete—in accordance with the mass spectrometry results.
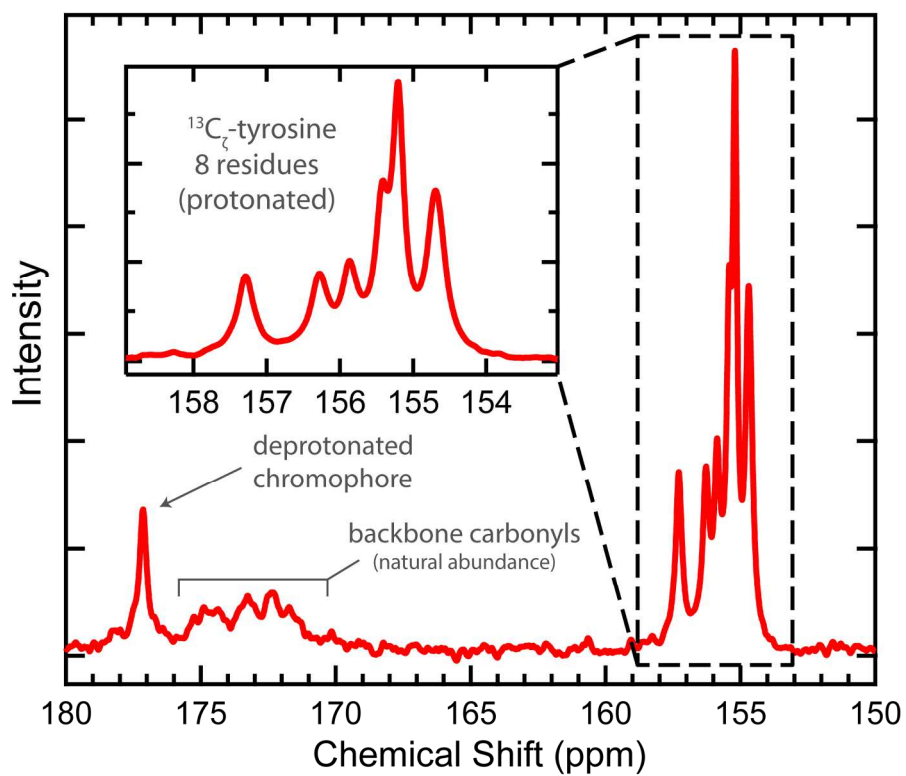
Fig. S1

1D ^{13}C-NMR spectrum of *ih*:GFP S65T at pH 8.0. The spectrum is composed of approximately 8,000 transients collected on a 300MHz Varian spectrometer with a 10 Hz exponential line-broadening function.
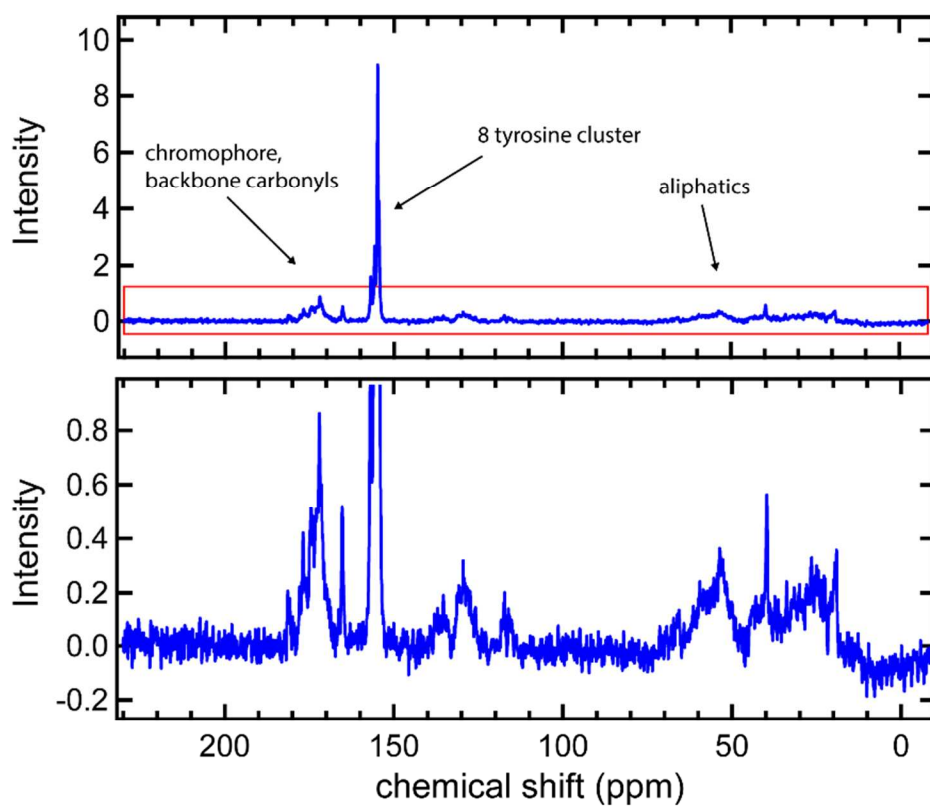
Fig. S2

Complete 1D ^{13}C-NMR spectrum of a representative protein sample (*s7+*:GFP S65T at pH 6.6) which highlights the relative intensity of the labeled peaks to the peaks arising from ^{13}C at natural abundance from the rest of the protein.
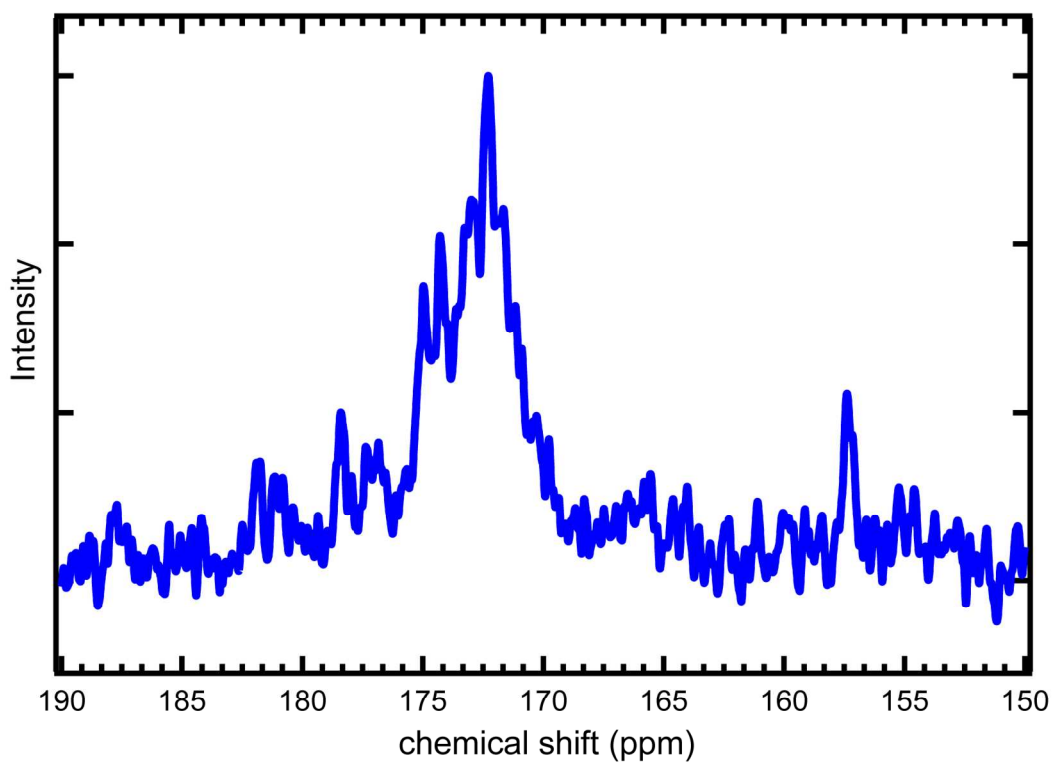
Fig. S3

1D ^{13}C-NMR spectrum of unlabeled protein in the region of interest. The broad cluster of peaks centered around 173 ppm is due to backbone and side-chain carbonyl carbons at natural abundance. The feature at 158 ppm is likely due to the arginine $^{13}C_\zeta$ with its sharpness due to the relatively free rotation of the solvent exposed side-chain.[5]

S.5   Controls for NMR and FCS Artifacts

Beyond the quantitative disagreement with the kinetic selective inversion experimental results, the qualitative appearance alone of the low pH ^{13}C-NMR spectra—in which the B-state peak decreases in intensity without changing position—precludes rates faster than ~1ms from lineshape predictions (Fig. 3C). Seeing the striking difference between NMR and FCS we sought to investigate the possible artifacts arising from either technique to explain the discrepancy. The two most significant differences between the experimental conditions are the concentration (~500pM for FCS and 1-2mM for NMR) and the illumination (~10kW/cm^2 for FCS and 0W/cm^2 for NMR). Many FPs are known to oligomerize at high concentration with wtGFP having a dimerization constant reported as ~100μM.[6] It is possible that sfGFP is forming dimers at the high NMR concentrations and changing the structural dynamics of the protein or even occluding the proton entry point. On the other hand, in FCS the protein is subject to very intense illumination not present in the NMR. Perhaps this light is driving some saturating behavior in which the proton transfer dynamics are altered.

S.5.a   Concentration Effects

We tested the effect of concentration by creating limiting situations: first, by guaranteeing monomers in the NMR measurement and second, by engineering dimers for an FCS measurement.

NMR of Monomers

It was noted very early by Ward and coworkers that the absorbance spectrum of wtGFP depended on the concentration in a way that caused a shift towards greater A-state absorbance.[2]

We measured the absorbance spectra of *ih*:GFP S65T up to concentrations as high as 1mM, taking care with the pathlength to avoid saturation and maintain accurate readings. Indeed, we also see a concentration-dependent increase in the A-state absorbance that suggests a dimerization constant around 100μM (see Fig. S4A, green). The mutation V206K, first reported by Tsien and coworkers,[7] introduces a positive charge in the middle of the hydrophobic dimer interface on strand 10 pointing outward into the solution; it was measured to change the dimerization constant from ~100μM to ~3mM by ultracentrifugation. This mutation was introduced to make *ih*:GFP S65T, V206K and the absorbance experiment was repeated. The resulting absorption spectra were unchanged up to at least 1mM thereby suggesting that V206K is in fact preventing dimerization (see Fig. S4A, blue).

The 1D ^{13}C-NMR spectrum of the monomeric *ih*:GFP S65T, V206K labeled with ^{13}C$_\zeta$-tyrosine was measured at pH 6.0 and 200μM on a 500MHz Bruker spectrometer with an indirect detection cryoprobe. The B-state peak stayed at exactly the same chemical shift indicating that indeed it remained in a slow exchange regime on the NMR timescale (Fig. S4B). At concentrations appreciably higher than 100μM *ih*:GFP S65T is effectively a dimer and, as a result, has a chromophore pK_a which is perturbed upward. This fact allowed us to observe higher A-state fractions for any given pH while avoiding the problems of protein precipitation encountered at pH less than 6.0. For this reason a complete NMR titration and the time-resolved NMR experiments were only practical using the protein not bearing the V206K mutation. Consequently the NMR results from Fig. 3C&D are more accurately described as representing the proton dynamics of the dimer. Nonetheless, the NMR spectra in Fig. S4B establish even in the strictly monomeric protein that there is a long-lived deprotonated species.

FCS of Dimers

Nearly all x-ray structures of GFPs show a hydrophobic protein interface centered at residue 206 on strand 10.[6] A covalent dimer version of the protein was created with the use of a disulfide bridge across the native dimer interface. We prepared the mutant *ih*:GFP S65T, V206C and allowed air oxidation of the sulfhydryl before size exclusion purification of the dimer which was confirmed pure by mass spectrometry and SDS-PAGE. FCS measurements were made for the complete pH titration as for the other samples. The diffusion coefficient resulting from the fit in the high pH limit was approximately 80% of the value for the monomer. This is in close agreement with the expectation from the Stokes-Einstein equation if we assume a spherical shape and a doubling of the hydrodynamic volume.

The pH-dependent flickering rates were somewhat slower than that observed in the monomer but not nearly enough to bring it into agreement with the NMR results in Fig. 3C&D. It is interesting to note that the dimer ACFs in Fig. 5B fit very poorly to a single-site model and qualitatively imply negative cooperativity (fits not shown). This result is further evidence for the multi-site character of the titration.

In summary the experiments above led us to conclude that the *ih*:GFP S65T monomer was still undergoing a slow exchange process that led to a particularly long-lived deprotonated species. The FCS measurements on an engineered covalent dimer of *ih*:GFP S65T revealed dynamics that were slightly slower but still much faster than the apparent rate from NMR. This small reduction in rate may be due to partial occlusion of the putative proton entry site through strand 7 which is immediately adjacent to strand 10 which lies at the center of the homodimer interface.
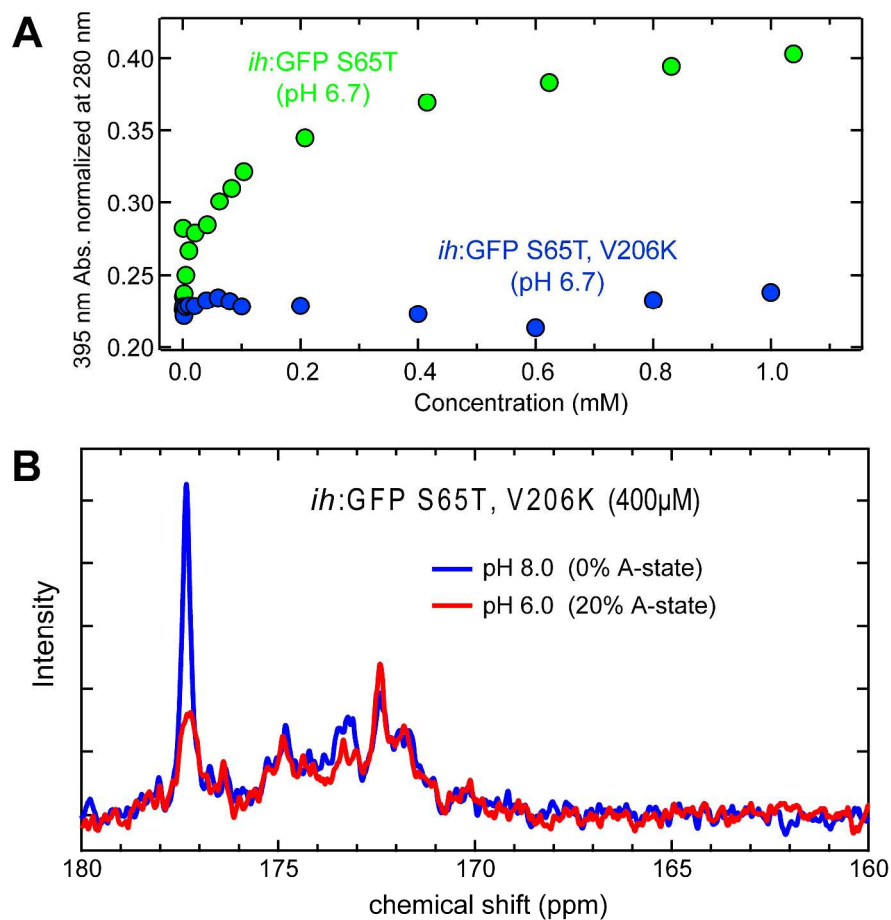
Fig. S4

A) Concentration dependence of A-state absorbance normalized to the absorbance at 280nm. The apparent pK_a of the dimer is higher than the monomer and so this relative absorbance is a proxy for the extent of dimerization. *ih*:GFP S65T (green) shows a clear dimerization isotherm with an apparent K_D of ~100μM. Upon addition of an aggregation-suppressing mutation *ih*:GFP S65T, V206K (blue) shows no sign of dimerization up to 1mM. **B)** 1D ^{13}C-NMR spectra of *ih*:GFP S65T, V206K at 400μM and pH 8.0 and 6.0. Under these conditions all the protein is monomeric. The fact that the peak at 177.2 ppm does not shift with pH implies that rate of proton transfer is still in the slow exchange regime.

S.5.b *Light Effects*

Fluorescent proteins are known to undergo light-dependent processes on many timescales from ESPT (ps) to triplet state formation (μs) to blinking and photobleaching (s). As mentioned above in *FCS Experiments and Fitting* (§S.3), we tested the sensitivity of the ACFs to light intensity by measuring fluorescent trajectories at 10 and 2.5 kW/cm^2 and the only effect was on the magnitude of the presumed triplet state formation. It is still possible that, even at this lower intensity, we are still under sufficiently high irradiance to be at saturation for some other light-dependent process with consequences for PT kinetics. To control for this possibility we conducted a 1D ^{13}C-NMR experiment under laser illumination of the B-state absorbance.

In Situ Illumination 1D ^{13}C-NMR

The *in situ* illumination inside the magnet was done in a similar fashion to Mizuno *et al.*[8] Briefly, a 473nm diode-pumped solid state laser was coupled into a multimode optical fiber which was run into a 10-mm Shigemi tube (Fig. S5A). The plunger portion of the tube was partially filled with water for better refractive index matching with the glass. To mitigate the inner-filter effect as much as possible the protein concentration was lowered to 400μM and the plunger was set to create a disk of sample 4mm thick. The final output power from the optical fiber into the sample was 105mW. In order to minimize photobleaching over the 12 hour experiment, a shutter was triggered by the auxiliary NMR pulse output to open 100 ms before and then close after the acquisition time (0.5 s) for the duration of the interpulse delay interval (Fig. S5C).

1D ^{13}C-NMR measurements of *ih*:GFP S65T at pH 6.6 were made with this setup first for 12 hrs. in the dark and then 12 hrs. with irradiation. It was found that the sample suffered less

than 2% photobleaching over the acquisition as determined by UV/Vis absorbance (Fig. S5B). A comparison of the dark and illuminated NMR spectra revealed that they were identical within the noise (Fig. S5D). More specifically, the peak at the B-state basis (177ppm) did not shift which indicates that it remains in a slow exchange regime even under illumination. In the FCS experiment lowering the power had no effect on the ACF aside from changing the dark triplet state component. This suggests that the light intensity is far within a saturating regime of some unknown light-driven process if it is playing a role at all. The fact that the significantly dimmer NMR experiment also experiences no change is consistent with the hypothesis that the proton transfer kinetics are unaffected by illumination.
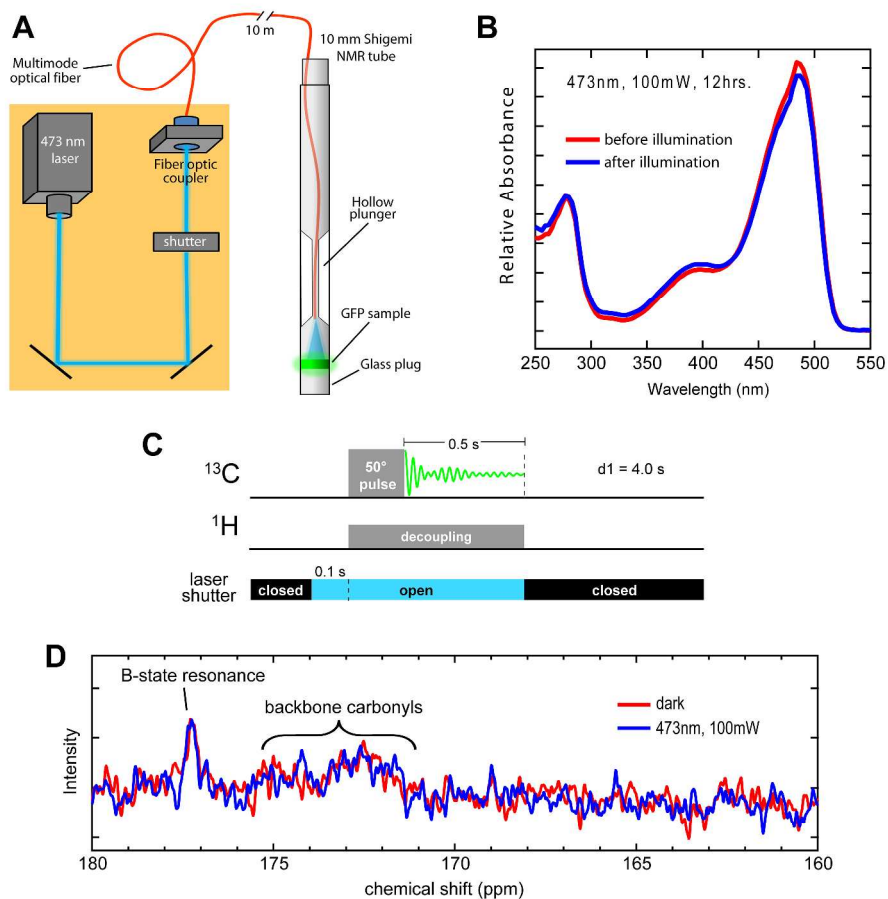
Fig. S5

A) Schematic drawing of the laser-illumination setup (drawing not to scale). **B)** Absorbance spectra of *ih*:GFP S65T before and after NMR measurement. The protein suffers about 2% photobleaching over the course of the experiment. **C)** Diagram of pulse sequence used in illumination experiments. **D)** 1D ^{13}C-NMR spectra of *ih*:GFP S65T for identical experiments in the dark (red) and with 100mW, 473 nm laser illumination (blue). The spectra are indistinguishable from each other within the noise and imply that the light has no effect of proton transfer kinetics.

S.6   NMR Saturation Transfer

To establish the population connectivity among putative chromophore ionization states we employed a series of NMR saturation transfer experiments. In Fig. S6A a low power 400-ms pulse was applied to the A-state resonance at 162 ppm (red) and in another experiment at 192 ppm as an equidistant control (blue). Only the on-resonance pulse caused a loss of all B-state magnetization as evidenced by the disappearance of the peak at 177 ppm. This shows that the species at 177 ppm is interconverting with that in the broad peak at 162 ppm many times during the saturation pulse thus losing magnetic polarization. The rest of the spectrum was unaffected.

The same experiment was repeated on the other variant with slow-exchange characteristics, *s7+*:GFP S65T (Fig. S6B). The black trace is the 1D ^{13}C-NMR spectrum, and red and blue are the spectra following 400-ms saturation at the resonances at 166.4 ppm and 177.2 ppm, respectively. The fact that peaks being irradiated directly disappear entirely indicates that the pulse is sufficiently long for saturation. More importantly, the saturation also leads to depletion of the distant peak to confirm chemical exchange between them. The diminution of the downfield shoulder of the backbone carbonyl band is likely due to overlap with the saturation pulse. This interpretation is further supported by the fact that the saturation at 166.4 ppm (red trace) does not lead to a decrease in the carbonyl region. If the broad decrease between ~177-173 ppm were due instead to a broad chemical exchange feature of the chromophore one would expect to observe a decrease in magnetization in this region upon 166.4 ppm saturation, which is not seen.
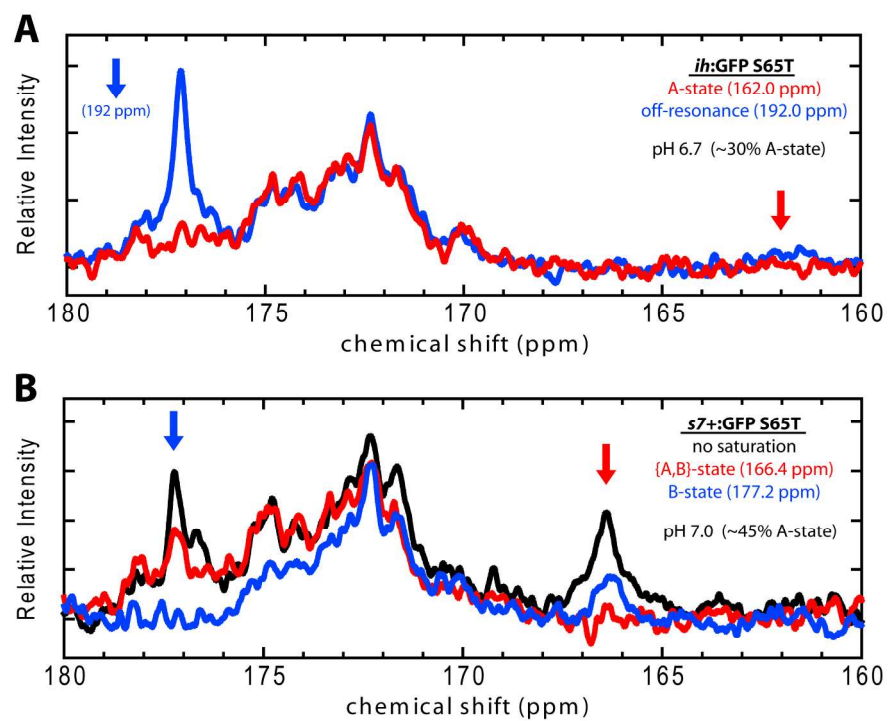
Fig. S6

NMR saturation transfer spectra for *ih*:GFP S65T at pH 6.7 (**A**) and *s7+*:GFP S65T at pH 7.0 (**B**).

S.7   NMR Selective Inversion Transfer

The time-resolved selective inversion experiment observes the change in intensity of the B-state magnetization as a function of the chemical mixing time since the A-state band has very poor signal-to-noise as a result of exchange broadening. If a single-site titration framework is assumed and the T_1 contribution neglected then the equation for the apparent chemical exchange is given by,

$$(Eq.\ S6) \qquad\qquad M_{BZ}(t) = M_{B0}[(2p_B - 2p_B^2)\,e^{-t/\tau_C} + (2p_B^2 - p_B)]$$

where M_{B0} is the equilibrium magnetization magnitude due to the B-state population, p_B is the fractional population in the B-state, and τ_C is the chemical exchange lifetime at a given pH. Recognizing the fact that $K_a = k_{deprot}/k_{prot}$, the chemical exchange lifetime (τ_C) may be expressed in terms of p_B and k_{deprot} by manipulation of Eq. S5,

$$(Eq.\ S7) \qquad\qquad \tau_C(pH) = \frac{10^{pH-pK_a}}{k_{deprot}(1+10^{pH-pK_a})} = \frac{p_B(pH)}{k_{deprot}}$$

The fractional B-state population which was calculated from the UV/Vis spectrum for the pH 6.70 sample to be 79% (see SI §S.2). In practice the triplicate data was fit to Eq. S6 to obtain τ_C which was then used to calculate k_{deprot} with Eq. S7 to get 74 s^{-1}.

In addition to the data shown in Fig. 3D for pH 6.7, kinetic selective inversion experiments were also conducted for *ih*:GFP S65T at pH 8.0 and 6.3. The latter two conditions were measured once for each mixing time while the former was done in triplicate. Fig. S7 shows the model from Eq. S6 applied globally. Aside from the deprotonation constant calculated from the single exponential fit to the pH 6.7 data in blue there are no additional free parameters. The

values for p_B were calculated from the visible absorbance spectra using the procedure in SI §S.2 and the apparent chemical exchange lifetimes were calculated as $\tau_C = p_B/k_{deprot}$ with $k_{deprot} = 74\ s^{-1}$. The relative magnitude of the B-state magnetization was determined by taking the integrated area of the peak at 177 ppm, subtracting the un-labeled blank to remove the backbone carbonyl overlap, and then normalizing to the area of the pH 8.0 peak with 0 ms mixing time.

It is notable how closely the model matches the data given the parameter constraints. The fact that the pH 8.0 sample experiences no loss of magnetization over the longest mixing time is entirely consistent with a non-exchanging deprotonated state in the limiting case of high pH. Comparing the pH 6.7 and 6.3 samples, the difference in net magnetization is slight when the mixing time is very short, reflective of the relatively small difference in the fraction deprotonated (p_B). However, the long time limit sees a much greater divergence in intensity preceded by differing rates of decay. Both of these features are quantitatively predicted from the global model and highlight the data's internal consistency.
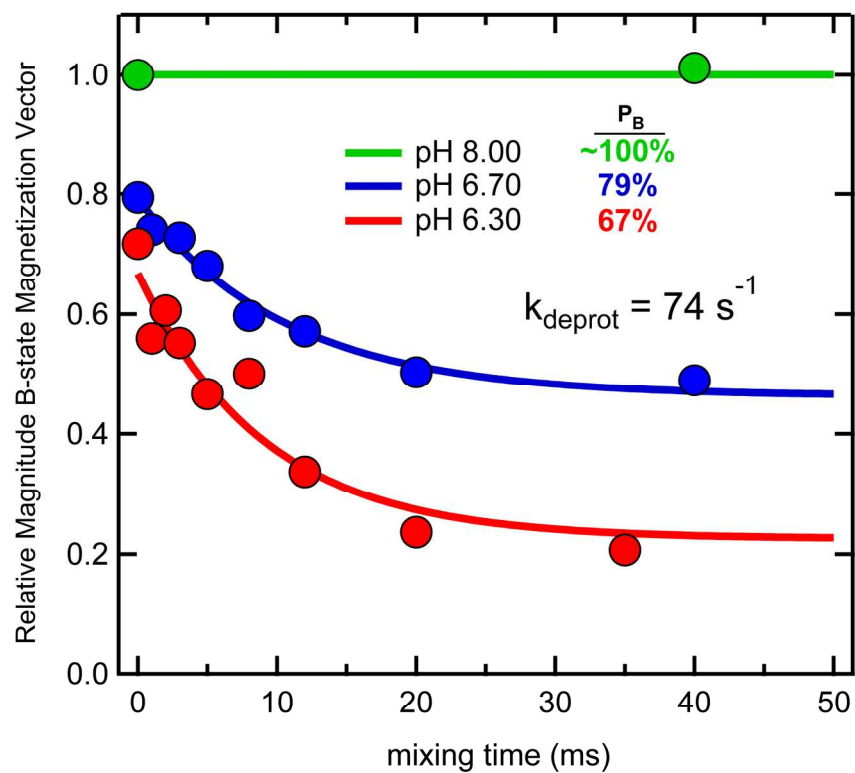
Fig. S7

Global model of NMR kinetic selective inversion data of *ih*:GFP S65T at pH 8.0, 6.7, and 6.3 according to Eq. S6. The sole fit parameter (k_{deprot}) was derived from the single exponential decay constant for the pH 6.7 data (blue).

S.8 Two-Site Kinetic Model

As introduced in the text, a two-site model of proton transfer can be expressed through a kinetic master equation. It is possible to construct a rate matrix, $\underline{\underline{K}}$, where K_{ij} is the rate constant from the state i to state j with the diagonal given by $K_{ii} = -\sum_{j \neq i} K_{ij}$ to ensure detailed balance. The connectivity and rates of conversion between the four states arising from this treatment are represented by a 4x4 kinetic rate matrix ($\underline{\underline{K}}$) written explicitly below.

(Eq. S8)

$$\underline{\underline{K}} = \begin{bmatrix} -k''_{B'B}*10^{-pH} - k''_{B'A}*10^{-pH} & k''_{B'B}*10^{-pH} & k''_{B'A}*10^{-pH} & 0 \\ k'_{BB'} & -k'_{BB'} - k'_{BA} - k''_{BA'}*10^{-pH} & k'_{BA} & k''_{BA'}*10^{-pH} \\ k'_{AB'} & k'_{AB} & -k'_{AB'} - k'_{AB} - k''_{AA'}*10^{-pH} & k''_{AA'}*10^{-pH} \\ 0 & k'_{A'B} & k'_{A'A} & -k'_{A'B} - k'_{A'A} \end{bmatrix}$$

The above rate matrix $\left(\underline{\underline{K}} \right)$ is written in terms of the microscopic unimolecular rate constants for deprotonation (′) and bimolecular rate constants for protonation (″). The parameters in Table 1, however, are given in terms of intrinsic pK_a's, the energetic coupling W, and effective rate constants k_{XY} that describe the equilibrium relaxation rate at the pH equal to the pK_a for each particular microscopic step. Specifically, the microscopic pK_a for each process is shown in Fig. 6 in terms of $pK_{A'A}$, $pK_{A'B}$, and W. If one considers a single acid dissociation reaction, the forward rate is given by the unimolecular deprotonation rate constant (k'_{deprot}) while the reverse reaction rate is the product of the bimolecular protonation rate constant and the proton concentration ($k''_{prot}[H^+] = k''_{prot}*10^{-pH}$). K_a is the ratio of the forward and reverse rate

constants ($K_a = k'_{deprot} / k''_{prot}$). Using the definition of the pK_a each of the following three

equalities can be written,

(Eq. S9a) $\quad 10^{-pK_a} = \dfrac{k'_{deprot}}{k''_{prot}}$ $\qquad$ (Eq. S9b) $\quad k'_{deprot} = k''_{prot} * 10^{-pK_a}$ $\qquad$ (Eq. S9c) $\quad k''_{prot} = k'_{deprot} * 10^{pK_a}$

It can be seen that the protonation reaction rate ($k''_{prot} * 10^{-pH}$) is equal to Eq. S9b when the pH is

equal to the pK_a.* Furthermore, the effective relaxation rate toward equilibrium is the sum of the

forward and reverse rates. As shown above these pseudo-unimolecular rates are the same when

the pH is equal to the pK_a. Thus, the effective rate is simply twice k'_{deprot}, and k''_{prot} is

determined by substitution into Eq. S9c. Collectively this treatment allows us to create the rate

matrix ($\underline{\underline{K}}$) using the microscopic pK_a's and effective rate constants from Table 1 while

maintaining detailed balance.

From this rate matrix, $\underline{\underline{K}}$, one may calculate the transition probability matrix as

$\underline{\underline{T}}(\tau) = \exp(\underline{\underline{K}} * \tau)$. Each element of the transition matrix, T_{ij}, describes the probability that the

system is found in state j after a time interval τ when starting in state i. The left eigenvectors

($\vec{q}_i$) of the transition matrix and their associated eigenvalues (λ_i) provide important equilibrium

and dynamical information about the chemical network. There is always an eigenvector having

an eigenvalue of exactly 1 which describes the relative equilibrium concentrations of all

chemical species. The other eigenvectors report on transition modes within the network and—to

ensure mass balance—have elements that sum to zero. Roughly speaking, the transition modes

---

* Note that the internal transfer between A and B is pH independent. In this case the equilibrium constant only describes a unimolecular intramolecular proton transfer.

can describe the kinetic clustering of states of like-sign and the equilibrium chemical flux through the transition network between clusters of opposite-sign. The associated eigenvalues have values between 0 and 1 and specify the implied timescale associated with these modes given by the equation,

(Eq. S10)
$$\tau_i^* = \frac{-\tau}{\ln \lambda_i}$$

in which τ is the time interval for the transition matrix, $\underline{\underline{T}}(\tau)$, and λ_i is the eigenvalue of the ith eigenvector. Specifically, modes having eigenvalues near 0 are very fast while those with eigenvalues near 1 are slow.[9]

All thermodynamic and kinetic information about the system is contained within $\underline{\underline{K}}$—or equivalently $\underline{\underline{T}}(\tau)$—and can be projected onto the observables measured in FCS and NMR. Below we describe the numerical methods used for calculating the theoretical FCS autocorrelation functions and NMR spectra.

S.8.a *FCS Autocorrelation Functions*

The equations for the autocorrelation function decay were adapted from Pramanik and Widengren.[10] Under the assumption that the diffusion coefficients of all protonation states are equal, the ACF can be expressed as the product of the diffusional and chemical exchange components, $G(\tau) = G_{dif}(\tau) * X(\tau)$. The separable chemical exchange part is given by,

$$(\text{Eq. S11}) \qquad X(\tau) = \frac{\displaystyle\sum_{i=1}^{M}\sum_{j=1}^{M} Q_i Q_j q_i^{eq} T_{ij}(\tau)}{\displaystyle\sum_{i=1}^{M} Q_i q_i^{eq}}$$

where M is the number of chemical species, Q_i is the relative brightness of the ith chemical species, $T_{ij}(\tau)$ are elements of the transition probability matrix, and q_i^{eq} is the equilibrium concentration of the ith chemical species. In the absence of any evidence to the contrary we have just assumed that both deprotonated states of the chromophore, B and B′, have the same brightness which we arbitrarily set to one and that both protonated states, A and A′, have no fluorescence. In practice, the final predicted ACF was the product of the empirically determined diffusional and dark triplet parts with the chemical exchange as in Eq. S4.

S.8.b *NMR Spectra Predictions*

The 1D-^{13}CNMR spectra were numerically calculated in the time domain from the coupled Bloch-McConnell equations.[11] J-coupling does not play a significant role in the system in this study so the density matrix treatment is not necessary and this simpler approach is entirely satisfactory. In matrix form the Bloch-McConnell equation is given by,

(Eq. S12)
$$\frac{d\vec{M}(t)}{dt} = -\left(i\underline{\underline{L}} + \underline{\underline{R}} - \underline{\underline{K}}\right) * \vec{M}(t)$$

where $\vec{M}(t)$ is the vector of magnetizations due to each chemical species, i is $\sqrt{-1}$, $\underline{\underline{L}}$ is a diagonal matrix with the frequency of each chemical species along the diagonal, $\underline{\underline{R}}$ is the relaxation matrix with the R_2 relaxation parameters for each species along the diagonal, and $\underline{\underline{K}}$ is the rate matrix as defined above. The solution to Eq. S12 can be written as,

(Eq. S13)
$$\vec{M}(t) = \exp\left(-\left(i\underline{\underline{L}} + \underline{\underline{R}} - \underline{\underline{K}}\right)t\right)\vec{M}(0)$$

We use the relative equilibrium concentration of all species as the initial magnetization, $\vec{M}(0) = \vec{q}^{\,eq}$. The net magnetization vector was calculated as the sum of all elements of the vector $\vec{M}(t)$. The frequency-domain signal was calculated with a fast Fourier transform of the discrete time-domain signal.

The B-state basis chemical shift used for all calculations was 177.15 ppm as determined from the peak position of the *ih*:GFP S65T sample at pH 8.0. The A-state basis was estimated from the linear extrapolation of the fit to the fraction ionized versus the peak chemical shift of the *int*PT series to be 162.00 ppm (see Fig. 3B). No distinction in chemical shift was made between A′ and A or B′ and B since the frequency is assumed to be dominated by the protonation state of the adjacent hydroxyl group. On account of the similar linewidths for the *ih*:GFP S65T pH 8.0 peak (~100% B-state) and the *s7*:GFP peak (~5% B-state) we assumed that the intrinsic linewidths of the A and B forms are nearly equal and used a full-width at half maximum of 0.22 ppm. Most of the NMR measurements were performed with a Varian Inova 300 MHz

spectrometer. Therefore all of the simulations presented in Fig. 7 were performed for 7.0 T at which the carbon-13 Larmor frequency is approximately 75 MHz.
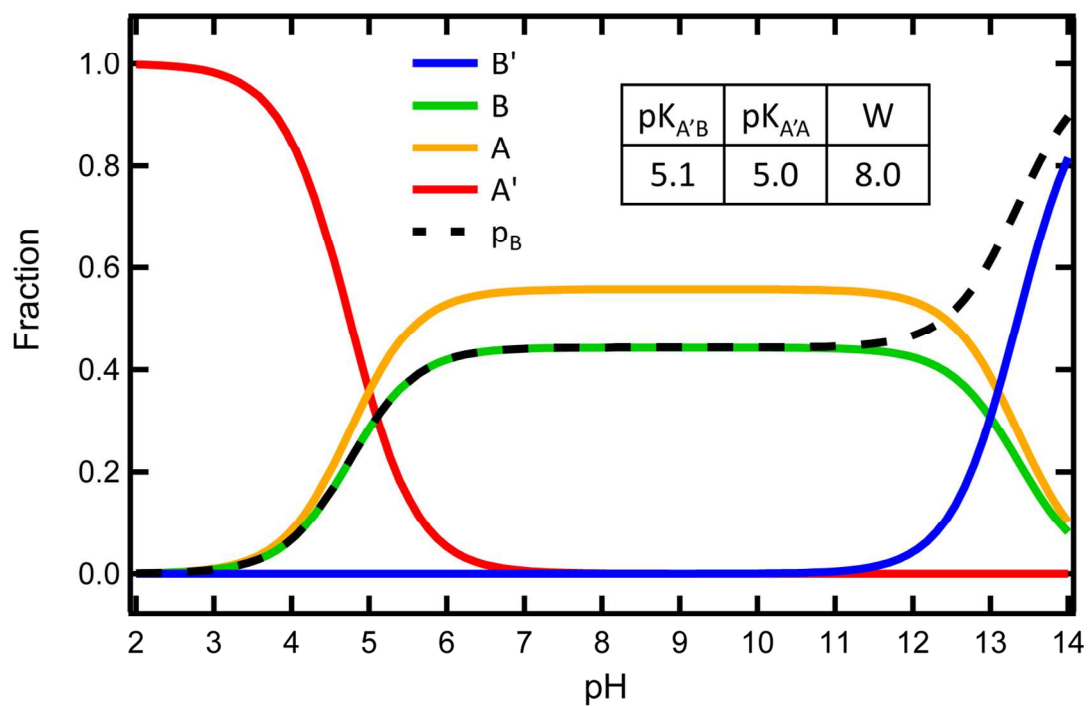
Fig. S8

Example titration curve for a prototypical *int*PT protein (compare to *ih*:GFP S65S in Fig. 2C). Over the range from ~6 to 12 the states B and A are dominant with their ratio determined by the difference of the intrinsic pK_a's. Specifically, $\dfrac{[B]}{[A]} = 10^{-(pK_{A'A} - pK_{A'B})}$

S.9   Titration of E222Q

Due to its established role as the terminal proton acceptor for ESPT in wtGFP[12] as well as being one of the most proximate ionizable groups to the GFP chromophore, E222 has attracted attention as a likely candidate for site "X" in Fig. 6.[13, 14] We made the isosteric mutation E222Q to assess the effect of removing the contribution of this ionizable site. Previously it was established that E222Q in wtGFP causes a breakdown in ESPT and results in a titratable *ext*PT protein superficially similar to S65T.[15, 16]

A master buffer solution was prepared with 20mM citric acid, 20mM sodium phosphate, 10mM glycine, and 150mM NaCl all adjusted to the appropriate pH by additions of 1M NaOH. Concentrated protein samples were diluted by approximately 100-fold into each buffer and the absorbance spectra were measured. The deprotonated fraction was calculated according to Eq. S3 in SI §S.2. The resulting data was fit to the single-site titration equation given in Fig. S9.

Fig. S9 shows absorbance pH titrations of *ih*:GFP S65T, *s7*:GFP S65T, and *s7+*:GFP S65T with the addition of E222Q. All of these proteins are fit reasonably well to a single-site titrations model and do not show the telltale signs of coupled sites such as negative cooperativity or non-zero A-state fractions in high pH limits as can be seen in Fig. 2C. Interestingly, while *ih*:GFP (green) and *s7+*:GFP (blue) have very similar pK_a's near 5.1, *s7*:GFP (red) has a much higher pK_a of 7.0. This value is getting closer to the pK_a of the model chromophore HBDI of 7.8[17] and further highlights the sensitivity of the chromophore to nearby structural disturbances.

Legend (inset):
- s7:GFP S65T, E222Q (orange)
- s7+:GFP S65T, E222Q (blue)
- ih:GFP S65T, E222Q (green)

$$p_B(pH) = \frac{p_B \max}{1 + 10^{pK_a - pH}}$$

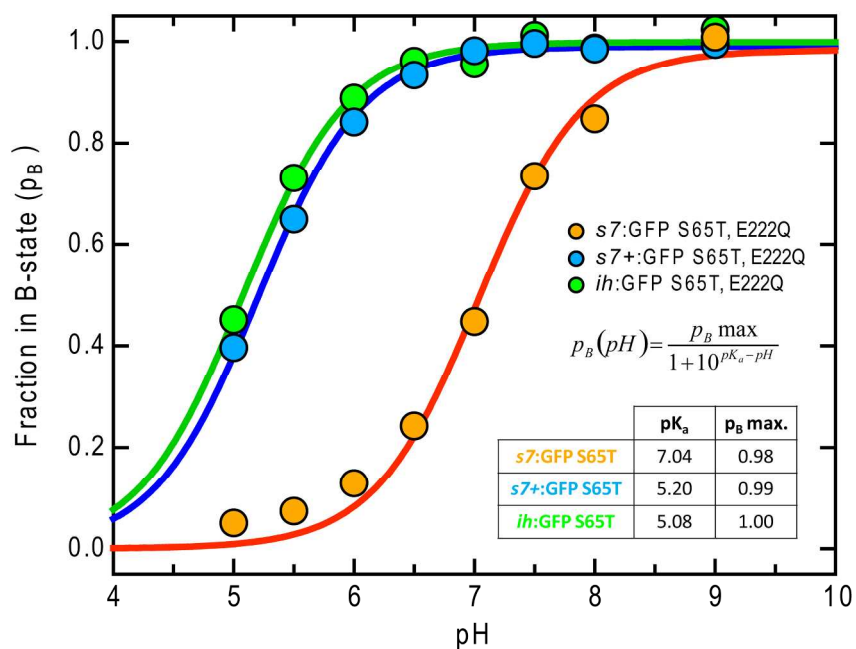|	pK_a	p_B max.
s7:GFP S65T	7.04	0.98
s7+:GFP S65T	5.20	0.99
ih:GFP S65T	5.08	1.00

Fig. S9

B-state fraction calculated from the UV/vis spectra as a function of pH for selected E222Q mutants. The solid lines are fits to the single-site titration equation inset. The values for the pK_a's and limiting B-state fractions (p_B) are given in the inset table.

S.10   Local His-tag Effects

Hexa-histidine tags are routinely used for efficient affinity-based purification of recombinant proteins. The effect of the His-tag is often assumed to be negligible with respect to the protein structure and function[18] and many of the counterexamples are a result of binding site occlusion.

We examined the effect of relocation and proteolytic removal of the His-tag on the two *s7* CPs (*s7*:GFP and *s7+*:GFP) due to its relative proximity to the chromophore. We found that the presence of the His-tag on the N-terminus gave rise to a significant shift in the apparent chromophore pK_a towards higher values. In the case of *s7*:GFP S65T and *s7+*:GFP S65T the magnitude of these shifts were 0.5 and 0.8 pH units, respectively (see Fig. S10A).

In our view, there are two general modes in which His-tags might affect the pK_a of a buried ionizable group. Firstly, the positive charge borne by the His-tag could electrostatically favor chromophore deprotonation. Secondly, the His-tag may act as a reservoir of protons gathered from solution thus causing an apparent increase in their local concentration. Our observation that presence of the His-tag favors the protonated chromophore form inclines us towards this latter explanation. The electrostatic effects are likely of less consequence due to the solvent exposure of the His-tag and the accompanying charge shielding by water and counterions. If we posit a simple local proton enrichment factor α, such that $[H^+]_{local} = \alpha[H^+]_{sol}$, we can write an expression for the modified $pK_{a\ obs}$,

(Eq. S14)
$$pK_{a\ obs} = pK_{a\ in} + \log\alpha$$

$pK_{a\ obs}$ is the observed pK_a in the presence of the His-tag, $pK_{a\ in}$ is the intrinsic pK_a in the absence of the His-tag, and α is the local proton enrichment factor which will, in general, be a function of pH.

This simple analysis can be further extended to estimate the consequences for chemical exchange kinetics if the same microscopic rate constants are retained and the factor α is treated as a constant over this approximately one unit pH window. It is straightforward to show that the apparent rate of chemical exchange at a pH equal to the pK_a for an ionization step to be,

(Eq. S15) $$k_{app} = 2\,k_{deprot}$$

If the factor α is included for a His-tag perturbation, algebraic manipulation leads to an expression for the apparent chemical exchange at a pH equal to the new perturbed pK_a (see Eq. S14) in the presence of the His-tag,

(Eq. S16) $$k_{app} = k_{deprot}\left(1+\frac{1}{\alpha}\right)$$

In the limit of no local pH perturbation (i.e. $\alpha = 1$) the result is identical to Eq. S15. However, with $\alpha > 1$ the prediction is that chemical exchange rate in the presence of the perturbation is slower than in its absence.

The predicted slowing effect is evident in our data and can be seen most clearly by comparison of the B-state (~177ppm) behavior upon pH titration (see Fig. S10 B&C). Specifically, in $s7+$:GFP S65T with the close His-tag, the peak retains its width and position and merely decreases in intensity with decreasing pH. In contrast the peak in $s7+$:GFP S65T lacking

the His-tag decreases in intensity but simultaneously increases dramatically in width and begins
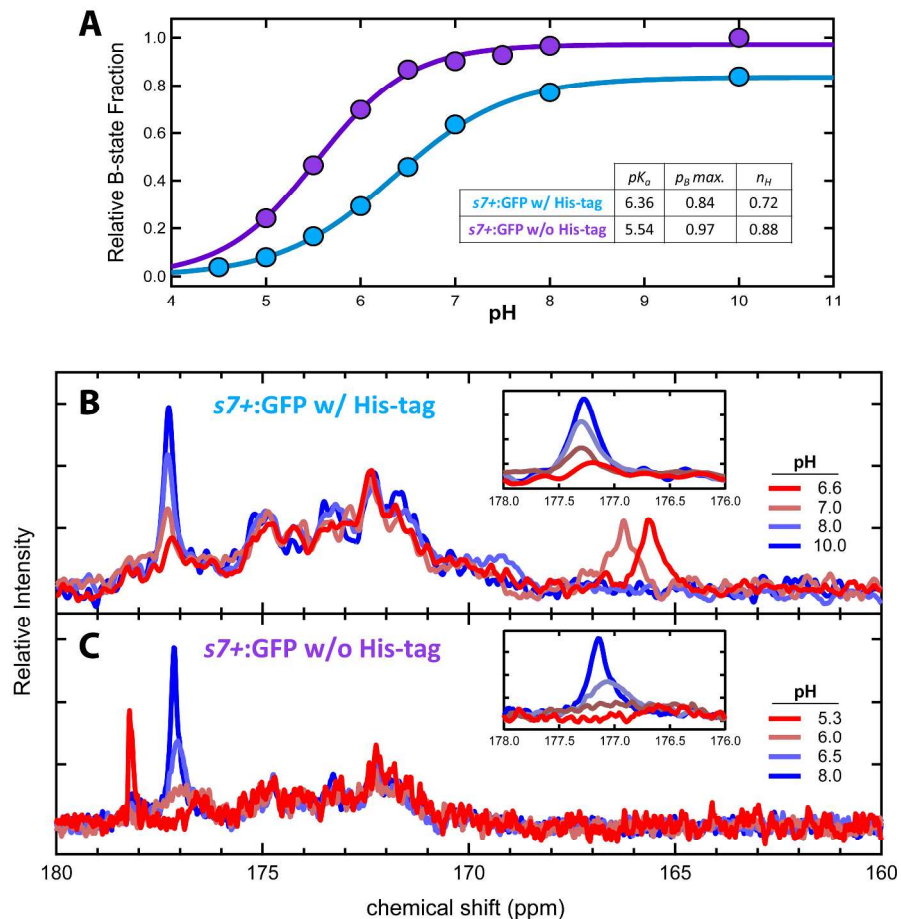
to shift upfield.

Fig. S10

A) pH titrations of *s7+*:GFP S65T with and without an N-terminal hexa-histidine affinity tag. The solid lines are fits to a Hill titration equation of the form, $p_B(pH) = \dfrac{p_B \max}{1 + 10^{n_H(pK_a - pH)}}$ where p_B *max* is the fraction in the B-state in the high pH limit, the pK_a is the pH value at which the p_B is half the value of p_B *max*, and n_H is the Hill coefficient describing the extent of cooperativity (values less than one are negatively cooperative). The fit parameters are shown in the inset table. **B)** 1D ^{13}C-NMR spectra of *s7+*:GFP S65T with the His-tag collected at 2.0 mM on a 300 MHz NMR. **C)** 1D ^{13}C-NMR spectra of *s7+*:GFP S65T without the His-tag collected at 0.3 mM on a 500 MHz NMR (the peak near 178 ppm at pH 5.3 is due to the citrate buffer).

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, *Nature
Biotechnology 24*, 79-88.

(2) Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., and Reeves, S. C. (1982) Spectral
Perturbations of the Aequorea Green-Fluorescent Protein, *Photochemistry and
Photobiology 35*, 803-808.

(3) Richarz, R., and Wuthrich, K. (1978) C-13 Nmr Chemical-Shifts of Common Amino-Acid
Residues Measured in Aqueous-Solutions of Linear Tetrapeptides H-Gly-Gly-X-L-Ala-
Oh, *Biopolymers 17*, 2133-2141.

(4) Grimsley, G. R., Scholtz, J. M., and Pace, C. N. (2009) A summary of the measured pK
values of the ionizable groups in folded proteins, *Protein Science 18*, 247-251.

(5) Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M.,
Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent
Wenger, R., Yao, H., and Markley, J. L. (2008) BioMagResBank, *Nucleic Acids
Research 36*, D402-D408.

(6) Phillips, G. N. (1997) Structure and dynamics of green fluorescent protein, *Current Opinion
in Structural Biology 7*, 821-827.

(7) Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002) Partitioning of lipid-
modified monomeric GFPs into membrane microdomains of live cells, *Science 296*, 913-
916.

(8) Mizuno, H., Mal, T. K., Walchli, M., Fukano, T., Ikura, M., and Miyawaki, A. (2010) Molecular basis of photochromism of a fluorescent protein revealed by direct (13)C detection under laser illumination, *Journal of Biomolecular Nmr 48*, 237-246.

(9) Noe, F., and Fischer, S. (2008) Transition networks for modeling the kinetics of conformational change in macromolecules, *Current Opinion in Structural Biology 18*, 154-162.

(10) Pramanik, A., and Widengren, J. (2006) Fluorescence Correlation Spectroscopy (FCS), In *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, Wiley-VCH Verlag GmbH & Co. KGaA.

(11) McConnell, H. M. (1958) Reaction rates by nuclear magnetic resonance, *Journal of Chemical Physics 28*, 430-431.

(12) 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.

(13) Bizzarri, R., Nifosi, R., Abbruzzetti, S., Rocchia, W., Guidi, S., Arosio, D., Garau, G., Campanini, B., Grandi, E., Ricci, F., Viappiani, C., and Beltram, F. (2007) Green fluorescent protein ground states: the influence of a second protonation site near the chromophore, *Biochemistry 46*, 5494-5504.

(14) Scharnagl, C., Raupp-Kossmann, R., and Fischer, S. F. (1999) Molecular basis for pH sensitivity and proton transfer in green fluorescent protein: protonation and conformational substates from electrostatic calculations, *Biophys J 77*, 1839-1857.

(15) 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.

(16) Kent, K. P., Childs, W., and Boxer, S. G. (2008) Deconstructing green fluorescent protein, *Journal of the American Chemical Society 130*, 9664-+.

(17) Conyard, J., Kondo, M., Heisler, I. A., Jones, G., Baldridge, A., Tolbert, L. M., Solntsev, K. M., and Meech, S. R. (2011) Chemically Modulating the Photophysics of the GFP Chromophore, *Journal of Physical Chemistry B 115*, 1571-1576.

(18) Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and DeLucas, L. J. (2007) His-tag impact on structure, *Acta Crystallographica Section D-Biological Crystallography 63*, 295-301.