Unusual Spectroscopic and Electric Field Sensitivity of a Chromophore with Short Hydrogen Bond: GFP and PYP as Model Systems

Chi-Yun Lin[1] and Steven G. Boxer[1]*

[1]Department of Chemistry, Stanford University, Stanford, CA 94305, USA.

*Correspondence to: sboxer@stanford.edu

ABSTRACT

Short hydrogen bonds, with heavy-atom distances less than 2.7 Å, are believed to exhibit proton delocalization and their possible role in catalysis has been widely debated. While spectroscopic and/or structural methods are usually employed to study the degree of proton delocalization, ambiguities still arise and no direct information on the corresponding potential energy surface is obtained. Here we apply an external electric field to perturb the short hydrogen bond(s) within a collection of green fluorescent protein S65T/H148D variants and photoactive yellow protein mutants, where the chromophore participates in the short hydrogen bond(s) and serves as an optical probe of the proton position. As the proton is charged, its position may shift in response to the external electric field, and the chromophore's electronic absorption can thus reflect the ease of proton transfer. The results suggest that low-barrier hydrogen bonds are not present within these proteins even when proton affinities between donor and acceptor are closely matched. Exploiting the chromophores as pre-calibrated electrostatic probes, the covalency of short hydrogen bonds as a non-electrostatic component was also revealed. No clear evidence was found for a possible contribution of unusually large polarizabilities of short hydrogen bonds due to proton delocalization; a theoretical framework for this interesting phenomenon is developed.

1.   INTRODUCTION

Hydrogen bonds are arguably one of the most important chemical bonds and are ubiquitous in biomolecules and materials [1][2]. Their intermediate strengths bridge between typical covalent and other noncovalent interactions [3], and they also play essential roles in mediating proton transfers [4]. Research on hydrogen bonds has flourished over nearly a century [5][6][7] since Linus Pauling first elucidated the nature of

hydrogen bonds in the late 1930s [8], nonetheless the exact correlation between geometries and energetics of hydrogen bonds remains controversial [9][10]. It is generally accepted that the topologies of potential energy surfaces (PESs) of protons in hydrogen bonds are strongly dependent on the heavy-atom distances R and the relative proton affinities ΔpK_α between donors and acceptors [11][12][13]. Note that pK_α (also proposed under the name pK_N [14]) is used here to describe the proton affinity of buried residues instead of the more commonly used pK_a, which is complicated by water solvation [15]. Because the proton is relatively light, nuclear quantum effects such as tunneling and delocalization can also be important especially for short hydrogen bonds [13][16], so this is a rich area of investigation.

The PESs for hydrogen bonds with heavy-atom separations R less than 2.7 Å are especially difficult to generalize and can only be examined on a case-by-case basis; the placement of the zero-point energy (ZPE) with respect to the barrier between wells is hotly debated [11][17][18][19][20]. If the proton affinity on each side is mismatched, one would expect the proton to be localized on the donor and the hydrogen bond is classified as a strong ionic hydrogen bond (SIHB) [21]. When the proton affinities of donors and acceptors are closely matched [12], the PESs become shallow and strongly anharmonic due to the strong coupling between the proton binding sites [13]. This could bring the ZPE close to or above the barrier and result in low-barrier hydrogen bonds (LBHBs), in which the proton is delocalized between donors and acceptors, and the corresponding PES is virtually indistinguishable from a single-well potential. Otherwise, the proton is equally probable to localize on the donor or acceptor, and we have a double-well potential [17]. Owing to the abundance and hypothesized strengths of short hydrogen bonds in proteins, especially at the active sites of many enzymes [22][23], the functional contribution of LBHBs to the catalytic power of enzymes has been actively debated [24][25]. The search for functionally important LBHBs in proteins has been rather difficult however [26][27][28], since no single approach can provide unambiguous evidence for the degree of delocalization [19]. X-ray and neutron diffraction are utilized to characterize the nuclear coordinates, while energetic information is mostly extracted from spectroscopic studies frequently combined with isotope substitution [2][10][29][30][31][32][33].

In our previous work, a short hydrogen bond with the hydroxyl group of the protonated neutral or A state of the chromophore was discovered in a particular green fluorescent protein (GFP) mutant, S65T/H148D using X-ray crystallography (Figure 1A) [34]. This was accompanied by an unusually featureless visible absorption band at 77 K compared with that of the normal A state of the chromophore when histidine is at position 148 [35]. We recently created an ideal condition for a LBHB to exist by systematically tuning the chromophore's pK_α via halogenation (Figures 1B and 1C) to test whether nearly zero ΔpK_α and short R between the chromophore and D148 are sufficient [36]. To characterize the energetics of the short hydrogen bond across the halogenated series, the spectral isotope effect (SIE) together with isotope fractionation factor at room temperature were measured by exploiting the halogenated chromophore as both an active participant in the short hydrogen bond and a sensitive optical probe of the proton position. The short O–O distance was demonstrated to persist throughout the variant series, but the data were inconsistent with a LBHB despite the close donor–acceptor proton affinity. In the present work, we extend the variant series (Figure 1C) and use electronic Stark spectroscopy at 77K to provide new insights [37]. Because proton transfer involves the movement of charge, one expects that it could be sensitive to an electric field, whether from the protein itself or an applied field. Thus, Stark spectroscopy can provide a novel approach for analyzing the extent of proton transfer between two wells and the degree of proton delocalization can be inferred, quite analogous to our previous applications to electron delocalization in mixed-valence systems [38].

Electronic Stark spectroscopy can also serve as a useful tool for extracting the underlying populations from the broad visible absorption bands associated with the S65T/H148D variants based on differences in Stark tuning rates. We associate the deconvolved populations with the proton being in each well, corresponding to a state with the *protonated* chromophore possessing a slightly lengthened O–H, which we will call an "A-like A state", and another state with the *deprotonated* chromophore engaging in a short hydrogen bond with its protonated partner D148, denoted a "B-like A state" (the B state of the GFP chromophore is the deprotonated form) (Figure 1B). We find that the correlations between the Stark tuning rates and the absorption maxima deviate from the calibration curves obtained through mutants with a normal hydrogen bond to the

chromophore, i.e., not involving the H148D mutation [39][40], suggesting the effect of the short hydrogen bond on the chromophore cannot be solely explained by electrostatics as was the case for these species with normal hydrogen bonds. Rather, it is likely that the covalency of the short hydrogen bond, owing to extended electron delocalization, alters the electronic properties of the neutral and deprotonated chromophores. Replacement of exchangeable protons with deuterons is also conducted to fine-tune the O–H(D) distance and gauge the influence of the proton (deuteron). To complement the results from the short-hydrogen-bond GFPs, photoactive yellow protein (PYP) and some of its mutants, which were previously characterized to be a closely related system [39] and possess two short hydrogen bonds with its chromophore (Figure 2) [18][19][20][41][42], are also analyzed and help strengthen the findings.

This study points out the fundamental differences between short hydrogen bonds, which have nonnegligible covalency [2][30][31][43][44], and other noncovalent interactions (e.g., normal hydrogen bonds, π stacking, and hydrophobic interactions), which can be adequately described by classical electrostatics [45]. We also shed light on the coupling between a π-conjugated system and short hydrogen bonds, which is of particular interest for understanding and designing molecular assemblies in the fields of resonance-assisted hydrogen bonds (RAHBs) [12][14] and hydrogen-bond mediated mixed-valence systems [46][47].
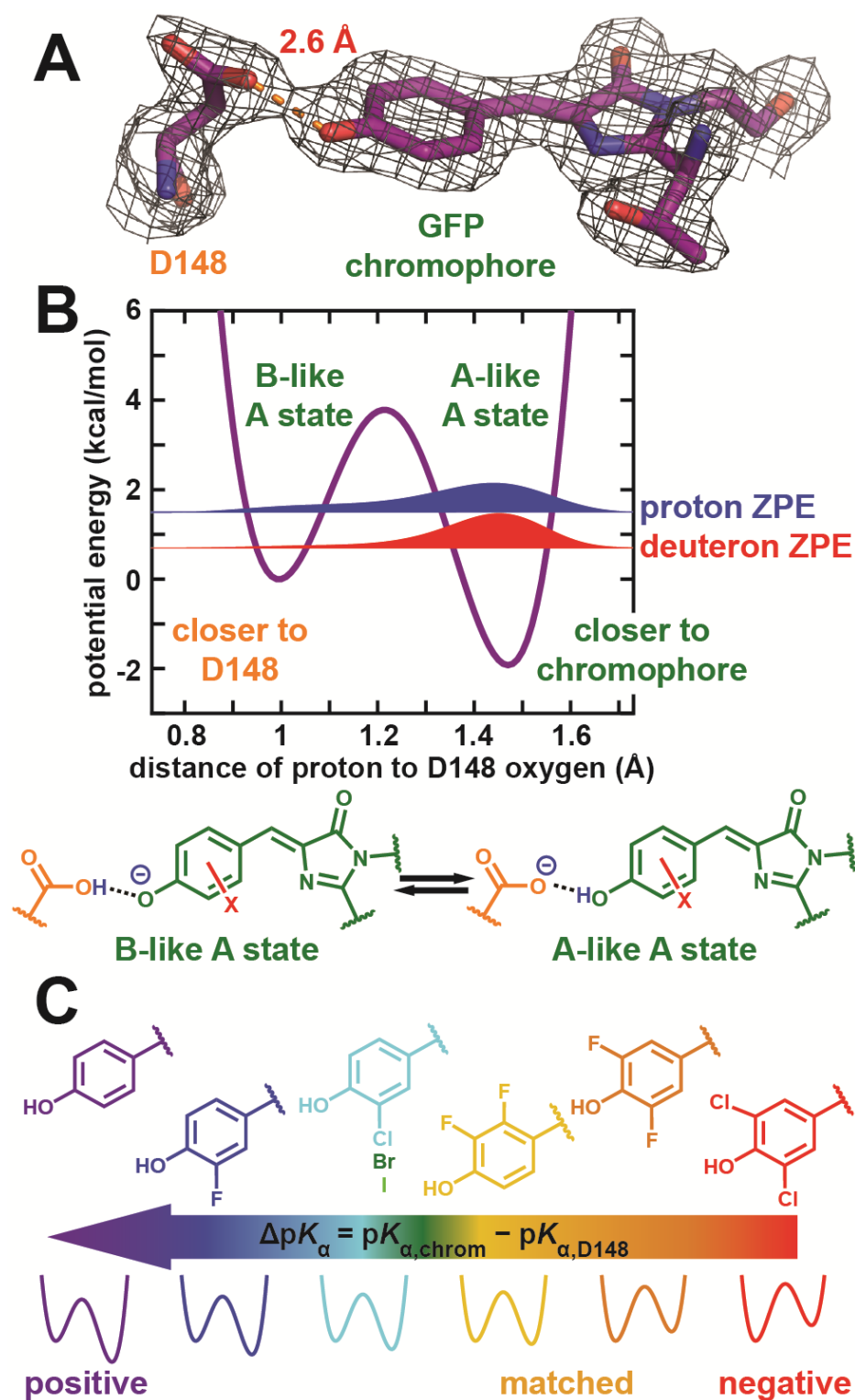
Figure 1. Structure and energetics of the short hydrogen bond in ih:GFP S65T H148D variants (see Section S1 for the nomenclature of circularly permuted GFPs). (A) Chromophore and D148 structure with the electron density ($2mF_o − DF_c$ contoured at 1σ) of ih:GFP S65T H148D (PDB: 4ZF3 [36]). Structures of other variants can be found in

Section S4 and figures therein; chromophore–D148 O–O distances for other variants are listed in Table 1. (B) A representative proton PES (in this case the Y66 variant) calculated from McKenzie's one-dimensional coupled Morse potential model [13] with parameters determined in the previous study [36]. The relative free energy between the two wells are governed by ΔpK_α (*vide infra*) through $\Delta G° = RT \ln 10 \Delta pK_\alpha$, where T is at 300 K. The two energy wells correspond to the proton residing more at D148 or at the chromophore, which we refer to as the A-like and B-like A states, respectively. The corresponding ground-state wavefunction and ZPE for proton or deuteron are shown in blue or red, respectively. Deuteration lowers the ZPE and reduces delocalization. (C) The relative proton affinity, ΔpK_α, of the short hydrogen bond is tuned via systematic introduction of halogen(s) to the chromophore in the H148D background, and the estimated PESs are shown below the arrow. ΔpK_α values for the variants are listed in Table 1.
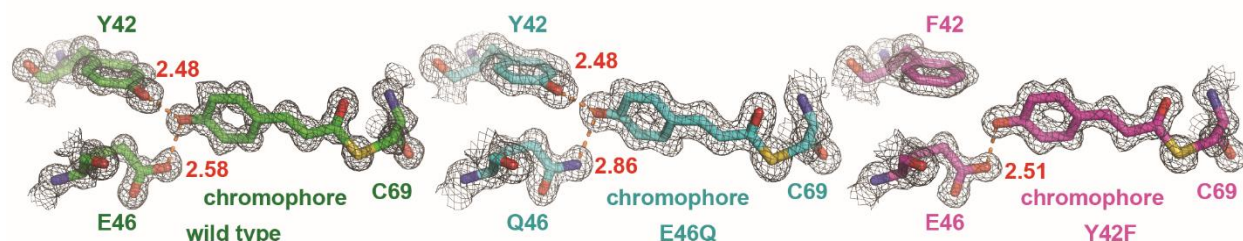

Figure 2. Local structure with the electron density ($2mF_o - DF_c$ contoured at 1σ) for each HhPYP mutant, including wild type (green, left, PDB: 1NWZ [48]), E46Q (cyan, middle, PDB: 1UGU [49]), and Y42F (magenta, right, PDB: 1F9I [50]). The hydrogen bond distances (in Å) are labeled in red and also listed in Table 1.

2.   EXPERIMENTAL METHODS

The experimental methods, including sample preparations, UV–vis absorption measurements, 77 K Stark spectroscopy, pre-resonance Raman spectroscopy, and X-ray crystallography, are detailed in the Supporting Information Sections S1–S3.

3.   RESULTS AND DISCUSSION

3.1.   Characterization of short hydrogen bonds and their proton PESs

3.1.1.   Mutant and variant design.

We introduced three more members, Y66(2,3-F_2Y), Y66(3-Br_1Y), and Y66(3-I_1Y), to the series of GFP chromophore variants (Figure 1C) in the background of S65T/H148D (short-hydrogen-bond GFP) [36] to better match the proton affinities between the donor

and acceptor (Table 1), as determined by titrating the chromophores under denaturing condition (Figure S24). X-ray crystallography revealed a consistent O–O distance of 2.4 – 2.6 Å for the short hydrogen bond across the variants within experimental error (Tables 1 and S8, see also Figure S6). It is not possible to assign the proton position even with the highest resolution X-ray structure at hand (Y66(2,3-F$_2$Y), 1.18 Å), but based on extensive empirical observations on short hydrogen bonds [2][30] and theoretical modeling [13][36], we can safely assume that the proton equilibrium position is slightly shifted away from either the proton donor or acceptor compared to the normal hydrogen bond. In particular, if the chromophore acts as a proton donor (i.e., having a larger pK_α than its partner D148), the O–H distance is estimated to lengthen from ~0.95 Å to ~1.05 Å due to the stronger coupling between the donor and acceptor O-H potentials [13][36].

PYP serves as another superb model system for elucidating the influence of short hydrogen bonds on chromophore properties. It is known from the crystal structures of wild-type *Halorhodospira halophila* PYP (HhPYP) that the chromophore, anionic *p*-coumaric acid, interacts with its neighboring residues Y42 and E46 through two short hydrogen bonds (Figure 2, Table 1). In addition to the thoroughly scrutinized HhPYP, we also choose to study PYP from *Salinibacter ruber* (SrPYP) [51], which has been found to exhibit larger SIEs at room temperature than HhPYP [52]. Even though no crystal structure has been solved for SrPYP to this date, the conserved Y42 and E46 from sequence alignment and the large SIE suggest the existence of at least one short hydrogen bond between the chromophore and these residues. In addition to the PYPs from two different species, we include a combination of HhPYP Y42F and E46Q mutations to break the hydrogen bond associated with the former and/or elongate that associated with the latter (Figure 2, Table 1) to probe the effect of short hydrogen bonds [41][49][50]. Unfortunately, SrPYP is effectively negatively supercharged (nearly 20% aspartate and glutamate), such that the Y42F mutation leads to chromophore protonation by raising the chromophore's proton affinity and breaking the short hydrogen bond(s) (Figure S13).

Note that we use "variants" and "mutants" to distinguish changes to the chromophore introduced by amber suppression and protein environment, respectively, to

facilitate the following discussion. As a useful shorthand, "anomalous A state" refers to a chromophore that engages in short hydrogen bond(s) ($R < 2.7$ Å), as in the S65T/H148D GFP variants and PYP mutants, while "normal A state" is reserved for neutral chromophores that only participate in normal hydrogen bond(s) ($R > 2.7$ Å) or other noncovalent interactions, such as those in the H148 chromophore variants. For clarity, since there is one proton present between D148 and the chromophore in these short-hydrogen-bond GFPs (Figure 1B), rather than referring to these two species as the conventional A and B states, we shall designate these species as A-like and B-like A states, the population ratio of which is internally governed by ΔpK_a instead of external pH [53]. We reserve the actual B state to the case in which both D148 and the chromophore are deprotonated at high pH, which is still redder than the red band (B-like A state) deconvolved from absorption of the anomalous A state (see below). The analogous A-like A state is absent in PYPs because of the large pK_a mismatch between the PYP chromophore and E46 or Y42F [18][20][42][52], resembling the case of the 3,5-Cl$_2$Y H148D GFP variant, so no further deconvolution is required for PYPs, as also suggested by Stark spectroscopy (Section S5, see also ref. [54]).

Table 1. The ΔpK_a's (defined in Figure 1C) and heavy-atom distances for short hydrogen bonds observed in crystal structures of short-hydrogen-bond GFP variants and PYP mutants in the *anomalous* A state. More discussion on structures and their correlations with ΔpK_a for GFPs can be found in Section S4 and figures therein.

short-hydrogen-bond GFP (ih:GFP S65T H148D)			
variants	ΔpK_a with D148	O–O distance with D148 (Å)[a,b]	resolution (Å)
Y66 (PDB: 4ZF3 [36])	+2.0	2.6	1.90
globally incorporated 3-F$_1$Y (PDB: 6OG8)	+1.3	2.6	1.60
Y66(3-Cl$_1$Y) (PDB: 4ZF4 [36])	+0.8	2.4	1.82
Y66(3-Br$_1$Y) (PDB: 6OGB)	+0.8	2.4	1.60
Y66(3-I$_1$Y) (PDB: 6OGD)	+0.9	2.5	1.65
Y66(2,3-F$_2$Y) (PDB: 6OGC)	+0.3	2.48	1.18
Y66(3,5-F$_2$Y)	-0.9	2.5	1.80

mutants	ΔpK$_a$ with E46 and/or Y42[c]	O–O/O–N distance with residue 46 (Å)[b]	O–O distance with residue 42 (Å)[b]	resolution (Å)
(PDB: 6OG9)				
Y66(3,5-Cl$_2$Y) (PDB: 4ZF5 [36])	-1.5	2.5		1.70
HhPYP mutants				
wild type (PDB: 1NWZ [48])		2.58	2.48	0.82
E46Q (PDB: 1UGU [49])	< -1.5	2.86	2.48	1.20
Y42F (PDB: 1F9I [50])		2.51	N/A	1.10

[a] The O–O distances for the GFP variants are averaged over two monomers within the asymmetric unit (in Table S8).

[b] The numbers of significant digits for the measured distances are dominantly determined by structure resolution: errors less than 0.1 Å typically require structural resolutions better than 1.3 Å [55].

[c] The estimation of PYP's ΔpK$_a$ being more negative than -1.5 is inferred because only the deprotonated chromophore is observed in spectroscopic and theoretical studies [19][20][42].

3.1.2. Proton PESs and deconvolution of short-hydrogen-bond protein electronic absorption spectra by Stark spectroscopy.

Unlike the 77 K electronic absorption spectra for most normal A states [40], the spectra for the A states from H148D GFP variants remain relatively featureless in a frozen glass at low temperature, but exhibit a consistent dip in the middle except for the Y66 counterpart (Figures 3A and S26). If the associated short hydrogen bond corresponds to a LBHB, the spectral dip could be the vibronic feature of absorption from a single species with a delocalized proton. However, deuteration not only widens the absorption band but also enhances the dip (Figures 3A and S26), especially for variants with nearly zero ΔpK$_a$ (Figure 3B).
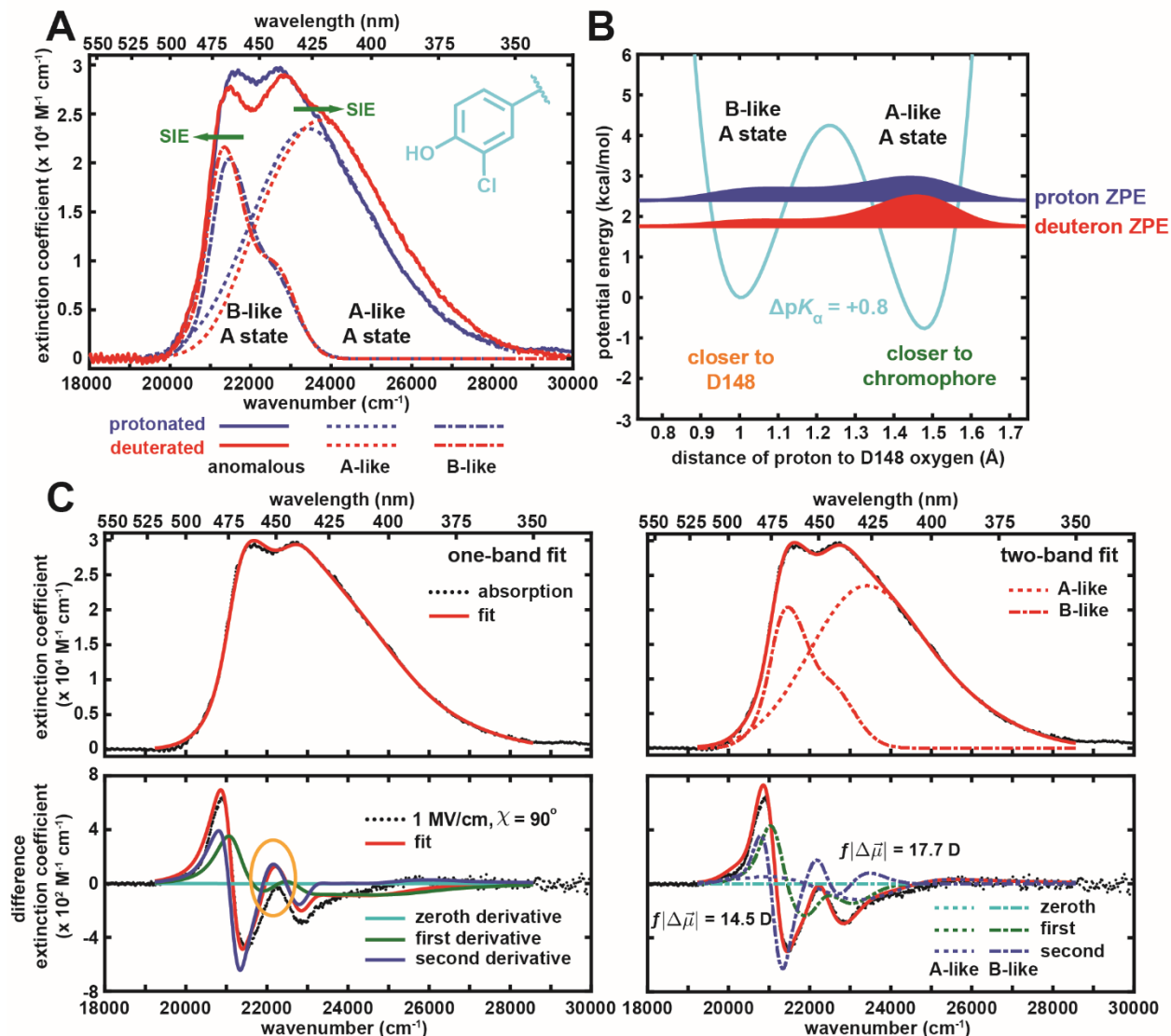
Figure 3. The absorption spectra, Stark spectra, and energetics of a representative S65T H148D GFP variant, ih:GFP S65T Y66(3-Cl$_1$Y) H148D. (A) The 77 K absorption spectra of the protonated (blue) and deuterated (red) species at pH 5 and pD 5, respectively. The corresponding A-like (dashed) and B-like (dash-dotted) A state bands are deconvolved from simultaneous fitting of the absorption and Stark spectra (Figure 3C). The direction of SIE upon deuteration for each underlying population is shown with a green arrow. Note that the maximum extinction coefficient of the normal protonated chromophore in GFP is about 60% of the deprotonated counterpart [53]. (B) The corresponding PES and ZPEs for the short hydrogen bond with $\Delta pK_a = +0.8$. The color coding follows Figure 1B. Deuteration further localizes the hydron wavefunction toward the donor and acceptor, causing the SIE seen in Figure 3A. (C) The 77 K absorption (upper panels) and 2ω Stark spectra (lower panels, scaled to 1 MV/cm) for the protonated Y66(3-Cl$_1$Y)/H148D variant at pH 5. The sum-of-derivative analysis is performed with one-band (left panels) and two-band (right panels) fits. One can see that the one-band fit is not satisfactory for both spectra simultaneously, especially in the region around 22000

cm^{-1} (circled), so an additional set of Stark parameters is required to fully recapitulate both bandshapes. The Stark tuning rates in Table 2 can then be extracted from the magnitudes of the second-derivative components [37].

When the Stark tuning rates are different for underlying populations, the Stark spectra, whose lineshapes are typically dominated by the second derivative of the absorption, are very useful for deconvolving the bands (Section S5). Through simultaneous fitting of both low-temperature absorption and Stark spectra, one homogeneous population with only one set of electro-optic parameters is insufficient to account for features across the entire absorption and Stark spectra (Figures 3C and S9), justifying the assertion that there are at least two populations. Assuming two populations and constraining with minimal spectral overlap, two distinct bands are resolved for all variants (Figures 3C and S9), except for Y66 and Y66(3,5-Cl$_2$Y) variants due to the larger degrees of pK_α mismatch (Table 1). Interestingly, the redder band resembles the typical B-state bandshape found in many normal GFPs [39] and its vibronic feature contributes to the dip. The bluer band is broad and featureless, behaving more similarly to the absorption band of the H148D Y66 variant than the vibronic structure observed in most other normal A-state spectra at low temperature [40]. Based on the population ratio and ΔpK_α (Figure S9), the red and blue bands can be intuitively assigned to species with the proton localized in each well of the PES. This is consistent with and reinforces the conclusion from the previous room-temperature study [36]. This assignment also agrees with the trend of SIE for each species: the underlying red band red-shifts and the blue band blue-shifts upon deuteration (Figures 3A and S26, Table S10). This SIE can be explained by the anharmonicity of the double-well PES causing a larger tendency for the deuteron to localize towards the donor or acceptor compared to the proton (Figures 1B and 3B), and these subtle changes in proton or deuteron positions can be sensitively probed by the chromophore absorption spectra. Population transfer between two wells caused by the external electric field should manifest as zeroth derivatives (a "non-classical" Stark effect [56]) rather than the typical second-derivative lineshapes from charge displacement upon excitation (a linear Stark effect [37]; see Section S5 for more discussion). Such a zeroth derivative component is not observed for each band in the Stark spectra within our ability to deconvolve the data (Figures 3C and S9), suggesting that external electric field driven proton transfer is not significant and hence there is a high

barrier in the PESs. This can be further attested by the 4ω (where ω is the field modulation frequency, see ref. [37]) spectra resembling second derivatives of 2ω spectra (Figure S11), as expected for charge displacement upon excitation rather than proton transfer within the short hydrogen bond [37].

Beside the sum-of-derivative analysis of the Stark spectra that has been discussed so far, the field strength F_{ext} dependence of Stark spectra (both 2ω and 4ω) provides additional qualitative evidence for the proposed topology of the proton PES. Significant deviation from the typical external field dependence of Stark spectra (i.e., 2ω and 4ω spectra scaling with F_{ext}^2 and F_{ext}^4, respectively) observed for charge displacement upon excitation is expected for borderline single-well/double-well cases due to proton transfer through the low barrier between the wells (Section S5, ref. [56]), while a single-well potential still follows the classical field dependence as the proton is extensively delocalized. The absence of deviations for all variants (Figure S12) suggests that either population exchange between two wells through tunneling and thermal activation is minimal at 77 K or the currently achievable strength for the applied electric field is still too low for significant external field-induced proton transfer to occur. Applied fields of ~ f ·1.4 MV/cm can be achieved on frozen glasses before dielectric breakdown, where f is the local field factor, which is necessary due to the larger field experienced by the chromophore compared to the externally applied field based on the unavoidable polarization effect of the chromophore environment [37]. Either way this rules out the existence of a LBHB in the S65T/H148D variants. A single-well potential for the short hydrogen bond could also be inferred just based on the observed field dependence and the second-derivative lineshapes, but this scenario is ruled out in combination with precedent evidence from low-temperature Stark spectroscopy and room-temperature experiments [36].

3.1.3. Proton PESs and deconvolution of short-hydrogen-bond protein vibrational spectra by Raman spectroscopy.

To further establish the topology of the PES, we also acquired the corresponding room-temperature vibrational spectra of the chromophore's phenol stretching mode (around 1230 – 1270 cm^{-1} [57][58], see also Section S6) for the S65T/H148D variants using pre-resonance Raman spectroscopy (Figure 4A). This method combines the

advantages of selectively enhancing signals from vibrational modes that are strongly coupled to the chromophore excitation (in contrast to IR spectroscopy), showing narrow peak widths (as opposed to UV-Vis spectroscopy), and possessing a fast intrinsic timescale comparable to electronic transitions (unlike IR and NMR spectroscopy). The phenol stretching mode blue-shifts as the chromophore becomes deprotonated [57][58] and is therefore a sensitive proton probe. If the population exchange between A-like and B-like A states is much slower than the electronic timescale, a split in the phenol stretching peak would be expected, provided the peak width is narrow enough to be resolvable. This is reminiscent of the spectroscopic strategies exploited to determine the ΔpK_α of a catalytically relevant hydrogen bond in ketosteroid isomerase (KSI) [59] and the degree of electron delocalization in 2-norbornyl cation [60][61][62] and mixed-valence compounds [63][64][65][66]. While the Raman spectra of the protonated $3\text{-Cl}_1\text{Y}$, $3\text{-Br}_1\text{Y}$, and $3\text{-I}_1\text{Y}$ variants show a single peak, two peaks can be readily seen upon deuterium exchange (Figures 4A and S19), indicating a placement of the deuteron ZPE below the barrier between two wells of the hydrogen-bond PES and setting a nonzero lower bound on the barrier height (Figure 4B). These findings can be well interpreted based on the ground-state wavefunctions of proton and deuteron within the PES calculated from McKenzie's one-dimensional coupled Morse potential model [13][36], in which the proton wavefunction is still fairly delocalized even with its ZPE below the barrier and may be the reason for the merging of the two underlying peaks (Figure 4B). The peak(s) of interest for the even better pK_α-matched fluorinated variants are unfortunately masked by the intense features from C–F stretching (Section S6, Figure S21), and thus no useful information could be obtained from these constructs.
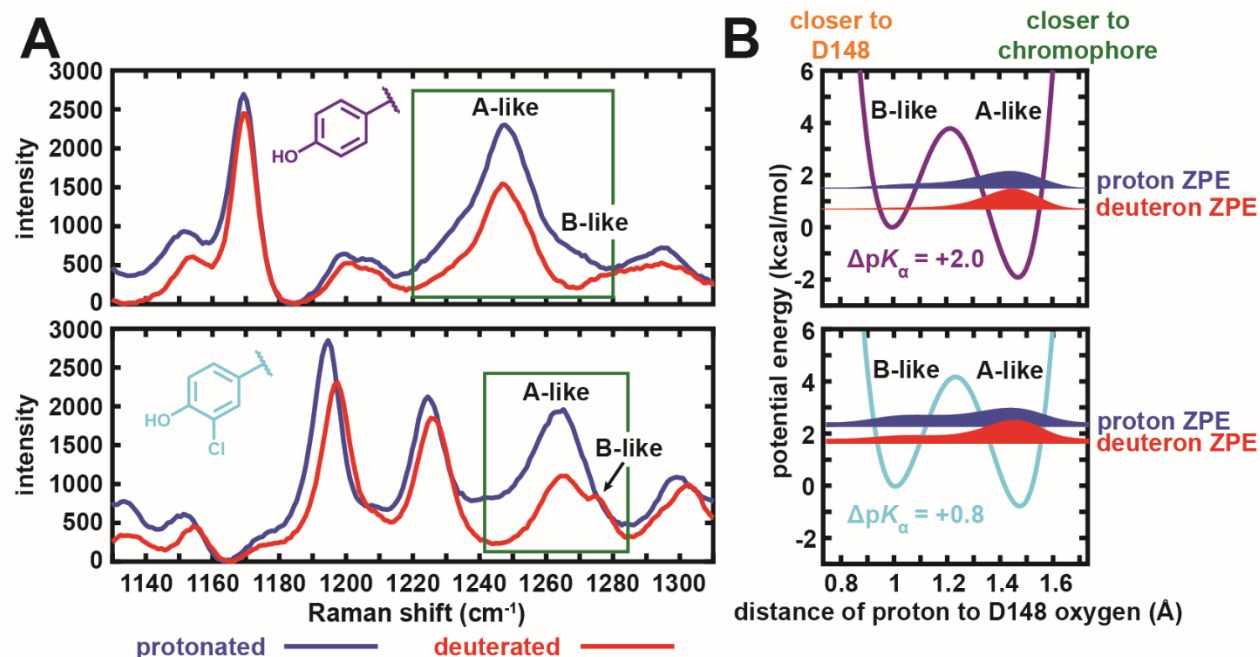
Figure 4. Pre-resonance Raman spectrum and derived PES's for pK_α matched GFPs. (A) Pre-resonance Raman spectra with 633 nm excitation of representative protonated (pH 5, blue traces) and deuterated (pD 5, red traces) S65T H148D GFP variants, which are Y66 (top) and Y66(3-Cl$_1$Y) (bottom) in this case. The peaks of interest, corresponding to a proton-sensitive phenol stretching mode, are highlighted within green boxes. Raman features outside the boxes are associated with other phenol modes that are not protonation sensitive [58]. Note that the observed Raman intensities are not concentration normalized. The appearances of the Raman bands can be understood through (B) their corresponding PESs and ZPEs, which are reproduced from Figures 1B and 3B. More discussion can be found in Section S6.

3.2   Perturbation of the short hydrogen bonds.

The strategy of modulating ΔpK_α via chromophore halogenation and using the chromophore absorption as a proton reporter is a double-edged sword: this approach is minimally perturbative in terms of structures (however, see Section S4 for more discussion on structures), yet the color of the chromophore can be affected by *both* the proton position and effects due to substituent-specific modulations in electronic distribution. Both are simultaneously changed by the substituents and therefore hard to be teased apart. Closer examination of the absorption maxima from the deconvolved bands reveals that while the trend in colors for both A-like and B-like A states across variants is primarily governed by the electronic distribution of the GFP chromophore, ΔpK_α only affects the A-like A states significantly [12][31]. Specifically, if we plot the B state

14

absorption maxima for the corresponding H148 variants (Table S11) against the A-like and B-like A states' absorption maxima, clear correlations for both can be seen, especially for the latter (Figure 5A, dashed lines). H148 variants are chosen because D148 is found to be twisted to the exterior of the protein and no longer engages in hydrogen bonding with the chromophore in the B state (see Section S6 in ref. [36]), and B states are chosen because there is no normal A state to compare against for these halogenated chromophores due to their low pK_a's. However, if we plot the absorption maxima against ΔpK_α, a more evident trend can only be observed for the A-like A state (Figure 5B), which likely suggests that the modification in O–H distances through ΔpK_α tuning is not sufficient to dominate the electronic effect for the B-like A state. This is not surprising, since the proton is closer to the chromophore in the A-like A state (Figure 1B). We will see that the same phenomenon is at work when we later scrutinize the Stark tuning rates for both anomalous A states. In hindsight, it was fortunate in our previous study [36] that the absorption maxima of the unresolved room-temperature anomalous A state bands from short-hydrogen-bond GFPs reflect the underlying population ratio from each well (Figure S25), and the resulting effect overwhelms the complicating electronic perturbation and allowed us to extract the direct consequences from ΔpK_α tuning.
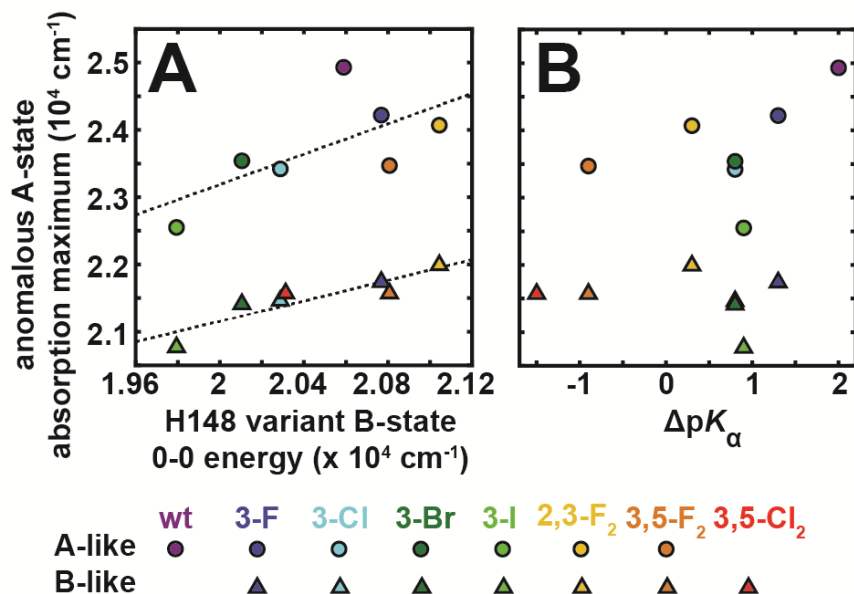

Figure 5. Correlation plots of the absorption maxima from the anomalous A states (Table S10) with (A) B-state 0–0 energy from their corresponding ih:GFP S65T H148 variants (Table S11) and (B) ΔpK_α (Table 1). The color coding is consistent with Figure 1C. The

plots are meant to gauge the contribution of electronic and proton position effects from halogenation to the absorption maxima of the anomalous A state. The electronic effect shows substantial influences on both the A-like and B-like A states, while the former is much more sensitive to the proton effect.

To study the influence of the short hydrogen bond on the chromophore's electronic structure, we extracted the Stark tuning rates for the population in each well of the PES from the protonated and deuterated S65T/H148D GFP variants (as in Figure 3C) and list them in Table 2. At first glance, other than the almost consistent and interesting decrease in Stark tuning rates upon deuteration, discussed in detail below, it is rather difficult to pinpoint an obvious trend as a function of ΔpK_α due to the aforementioned convolution of proton position and electronic effects from the halogen substituents. This convolution can be resolved, however, by comparing to the deprotonated and protonated chromophores under electrostatic influences [39][40] to isolate the additional perturbation of the short hydrogen bond from the electrostatic effect. This approach also relieves the severe problem of comparing H148D species with their H148 counterparts (Figure 5A), as both changes in hydrogen bond distances and hydrogen bond partners are accounted for by the electrostatic responses of the calibrated chromophore.

Table 2. A summary of measured Stark tuning rates from the short-hydrogen-bond GFP variants using electronic Stark spectroscopy. A more detailed table can be found in Table S10, and the corresponding spectral analysis is detailed in Section S5. For variants with larger pK_α mismatch (i.e., Y66 and Y66(3,5-Cl$_2$Y)), only one of the species is observed, hence the data are not available (N/A) for the other.

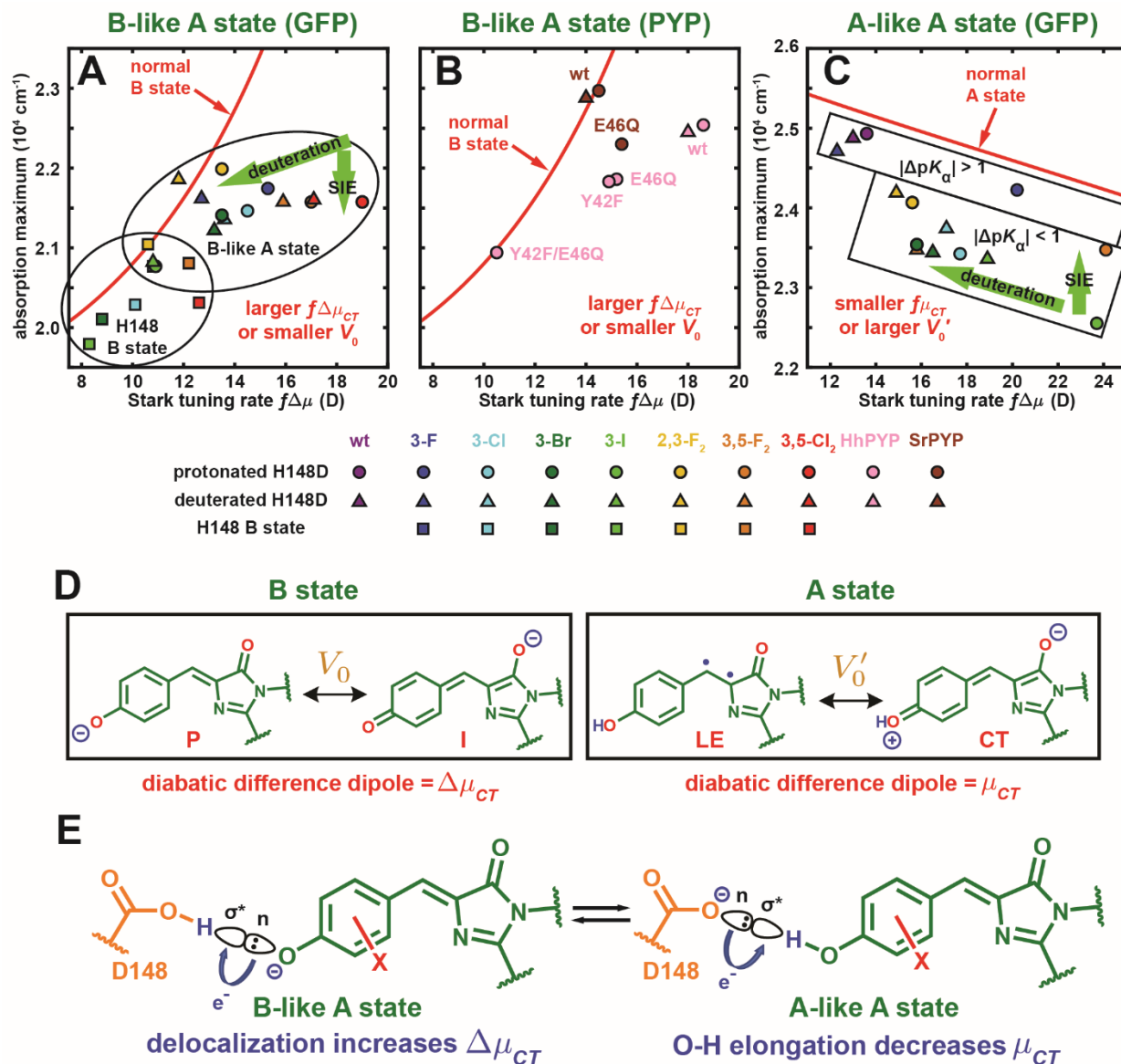| short-hydrogen-bond GFP variants	ΔpK_α with D148	A-like A state Stark tuning rate (Debye)		B-like A state Stark tuning rate (Debye)	
		protonated	deuterated	protonated	deuterated
Y66	+2.0	13.6	13.0	N/A	
globally incorporated 3-F$_1$Y	+1.3	20.2	12.3	15.3	12.7
Y66(3-Cl$_1$Y)	+0.8	17.7	17.1	14.5	13.6
Y66(3-Br$_1$Y)	+0.8	15.8	16.5	13.5	13.2
Y66(3-I$_1$Y)	+0.9	23.7	18.9	10.9	10.8
Y66(2,3-F$_2$Y)	+0.3	15.6	14.9	13.5	11.8
Y66(3,5-F$_2$Y)	-0.9	24.1	15.8	17.0	15.9
Y66(3,5-Cl$_2$Y)	-1.5	N/A		19.0	17.1

Figure 6. Correlation of absorption maxima and Stark tuning rates and contributing resonance forms for (A) B-like A state from short-hydrogen-bond GFP variants, (B) PYP mutants, and (C) A-like A state from short-hydrogen-bond GFP variants (Tables 2 and S10). The calibration curves for the corresponding normal states are reproduced from references [39] and [40] and shown in red. The trends of Stark tuning rate change and spectral shift (SIE) upon deuteration are represented with green thick arrows. (D) The diabatic difference dipole moments and couplings between the underlying resonance forms of the deprotonated (left) [39] and protonated chromophore (right) [40]. (E) The molecular orbital picture of the short hydrogen bond within each anomalous A state, involving donation of electrons (shown as a lone pair) from the proton acceptor's nonbonding orbital (n) to the donor–H's σ* orbital through overlap.

17

We examine the B-like A state first (Figure 6A). The red line in Figure 6A is the absorption maximum *vs*. Stark tuning rate correlation curve for the unsubstituted normal B state anionic chromophore to quantitatively capture its electrostatic response. This curve is a fit from the data of S65T environmental mutants using the Marcus–Hush model and diabatic states in Figure 6D (left) developed in our previous work [39]. We have previously demonstrated that it is meaningful to treat chromophore variants and environmental mutants the same way in terms of electron density modulation [39] (see also ref. [67]), so we can safely perform the same analysis here even with the change in chromophore identities. Data from H148 with halogenated chromophore variants are also shown in squares. As previously detailed in ref. [39], while the variants follow a general trend consistent with the fitted curve from environmental mutants, some nonnegligible deviations from the environmental mutants can be observed, reflecting characteristics from individual substituted chromophores and complicating the use of these chromophore variants as consistent electrostatic probes. We expect the same deviation also occurs in the B-like A state for H148D variants. Indeed, as we plot the corresponding data in Figure 6A, the same general pattern persists, albeit with bluer absorption maxima and larger Stark tuning rates. On the other hand, no obvious trend related to ΔpK_{α} can be identified (as opposed to the A-like A state, *vide infra*), and if a trend does exist, it is likely masked by the substitution-specific electronic effect. Data from H148D variants exhibit a larger spread compared to the H148 counterparts. Both data sets also noticeably biased towards the side of larger $f\Delta\mu_{CT}$ or smaller V_0 within the framework of the Marcus–Hush treatment (Figure 6D, left) [39]. At this point, it is not possible to distinguish which of the two Marcus–Hush parameters, $f\Delta\mu_{CT}$ and V_0, is affected more by the short hydrogen bond because of the intrinsic electronic variations due to substituents on the chromophores.

To help clarify and resolve this GFP chromophore variant complexity, HhPYP mutants can offer an incisive answer, since they exist in a B-like A state and possess the same chromophore but various hydrogen-bonding patterns through mutagenesis at Y42 and E46 (Table 1 and Figure 2). We have concluded in our previous work [39] that by comparing the correlation between the Stokes shift and absorption maximum from PYP and GFP mutants, both anionic chromophores follow the same *quantitative* behavior and share the same electronic coupling V_0 even in the presence of short hydrogen bonds. In

contrast, when analyzing with the correlation plot between the absorption maximum and Stark tuning rate (Figure 6B), as we alter the number of short hydrogen bonds in HhPYP from two (Y42/E46) to one (Y42/E46Q and Y42F/E46) and finally to zero (Y42F/E46Q), the data point moves closer and closer to the GFP-based B-state fit curve and finally sits exactly on it. This shows an unambiguous trend of decreasing $f\Delta\mu_{CT}$ or increasing V_0 as we remove more short hydrogen bonds. Because we know that V_0 is unchanged across this mutant series, $f\Delta\mu_{CT}$ becomes the only factor that is influenced by the short hydrogen bond. By extension, we can argue that the S65T/H148D GFP variants likely share the same behavior, but it is concealed by the additional substituent-specific electronic effects from halogenation. Conversely, the SrPYP mutants exhibit the opposite trend upon E46Q mutation, which is different from that expected from the room-temperature SIE study [52], suggesting a nontrivial interplay between the two putative short hydrogen bonds that could be studied through NMR [68].

For the A-like A state, the same exercise can be duplicated (Figure 6C). The A-state fit curve [40] is exploited to reflect the electrostatic response of the unsubstituted normal A state chromophore. Deviations of H148D variants from the A-state fit curve can again be seen, but this time they correspond to larger V_0' or smaller $f\mu_{CT}$ than the normal A state according to the three-form model (Figure 6D, right) [40]. Since there is no valid comparison for the A-like A state with the corresponding halogenated chromophores in the normal A state due to the low pK_a's of halogenated chromophores, we cannot rule out the possibility of substituent-specific electronic effects to explain the absence of any obvious pattern of these data points. However, the fact that values with larger deviations correspond to better pK_α-matched variants suggests that the electronic structure of the protonated chromophore is modulated through the proton when it serves as a proton donor in a short hydrogen bond. Relying on the robust insensitivity of V_0 for the deprotonated halogenated chromophore from the Stokes shift study [39] and the previous conclusion from the B-like A state, we speculate that $f\mu_{CT}$ could again be the only parameter that is subject to the perturbation of the short hydrogen bond. Just as how we scrutinized the variation in V_0 of the deprotonated chromophore through its Stokes shift, this claim could in principle be tested with the correlation of Stokes shift from A* emission and absorption maximum (as in Figure 8 in ref. [39]) by considering an additional vibronic

coupling, but the relevant measurements are so far inaccessible for most GFP mutants and variants given the efficiency of excited-state proton transfer and the chromophore's low pK_a.

In summary, we have identified that $f\Delta\mu_{CT}$ for the deprotonated chromophore and $f\mu_{CT}$ for the protonated chromophore increases and decreases, respectively, in the presence of short hydrogen bond(s). Both quantities are products of the (difference) dipole moment $(\Delta)\mu_{CT}$ and the local field factor f, which arise from the electron distribution of the π system *per se* and the polarizability of the π system's *environment*, respectively. In the following we will propose a mechanism for the short hydrogen bond to perturb each of the quantities and examine the corresponding plausibility.

3.2.1 Perturbation on dipole moments.

Based on the short hydrogen bond's ability to modulate the O–H distance, the (difference) dipole moments of the chromophores can be evidently tuned by the short hydrogen bond. Specifically, the chromophore's O–H distance is lengthened or shortened when the chromophore behaves as a proton donor or acceptor, respectively [30][31]. We have asserted in our previous work that the A state behavior is not at all akin to the B state counterpart owing to the covalency of the O–H bond [40], so we should expect a deviation in the character of the A-like A state from an electrostatically perturbed normal A state. By elongating the O–H bond, we expect the CT form (Figure 6E, right) to be less charge localized on both ends of the chromophores, thereby reducing μ_{CT} (Figure 6E, right).

To understand the B-like A state's response to the short hydrogen bond, we need to invoke the covalent nature of the short hydrogen bonds to explain the deviation from the electrostatically modulated B state, which includes examples perturbed by normal hydrogen bonds [39]. The covalency of short hydrogen bonds as opposed to normal ones has been documented with NMR [69]. It results from the delocalization of electrons between the proton donor and acceptor, where the exchange interaction is necessary for the short hydrogen bond to overcome the unfavorable van der Waals repulsion between the overlapping heavy atoms of the proton donor and acceptors, and so a model with only classical electrostatics is insufficient to predict the existence of short hydrogen bonds [43].

This also rationalizes the inability of using classical molecular dynamics to simulate and maintain the short hydrogen bond (O–O distance ~ 2.6 Å) between Y16 of KSI and its substrate analog 19-nortestoterone (Supplementary Text 4 from ref. [70]) in contrast to QM/MM simulations in ref. [71]. In the language of molecular orbitals, the lone pair electrons in the nonbonding orbital of the proton acceptor partially occupies the vacant σ^* orbital of the donor–proton covalent bond [31][43][44], resulting in a more delocalized electron distribution within the short hydrogen bond and a weakening and lengthening of the donor–proton bond (Figure 6E, left). The interaction strength strongly depends on the energy matching of the two involved orbitals, which is equivalent to pK_α matching. Therefore, when the chromophore behaves as a proton acceptor in the B-like A state, it donates its electrons and leads to a more spread-out electron distribution, hence a larger $\Delta\mu_{CT}$.

By virtue of forming a short hydrogen bond with its partner or the covalent modification by halogens, the number of electrons within the chromophore is no longer the same due to electron delocalization across the hydrogen bond and the slightly different identity of the chromophore. Since the proton is attached to the chromophore through a true covalent bond (a hydroxyl group) rather than a partially covalent hydrogen bond, the A-like A state is more sensitive to changes in ΔpK_α, while the B-like A state is understandably dominated by the electronic effect from substitutions, yet another covalent interaction, in the chromophore variants. This is reminiscent of the effect on the absorption maxima observed in Figure 5. The phenomenon of short hydrogen bonds modulating π systems has also been extensively investigated, notably in the context of RAHB [12][14] and hydrogen-bond mediated mixed-valence complexes [46][47], in which electrons within hydrogen bonds can be thought of as part of the delocalized π systems [14].

3.2.2. Perturbation on local field factors.

In addition to the change in electronic dipole moments discussed so far, another provocative proposal is that proton polarization within the short hydrogen bond(s) might manifest itself as a local field factor f, as this could also increase $f\Delta\mu_{CT}$ for the B-like A state. While we will argue below that this is likely not making a significant contribution, the

underlying concepts are interesting and may apply to other short hydrogen bond systems, so the basic idea is developed in the following and in further detail in Sections S7 and S8. The local field factor is typically considered as originating from field-induced molecular rotation of solvent (the reaction field), nuclear displacement, and electronic distortion in the environment. We are not aware of any analysis suggesting that f could also specifically be affected by electric-field-induced proton polarization, which is analogous to magnetic-field-induced ring currents that account for the (de)shielding of nuclei in NMR [72] (see footnote [100] in the Supporting Information). While we have shown in Section 3.1.2 that the proton in the short hydrogen bond of GFP variants cannot be transferred across wells via experimentally accessible external electric fields, it could be displaced slightly within each well.

Treating the problem classically, since a proton carries a positive elementary charge $+e$, an external field F_{ext} can shift its mean position, whose displacement Δr is determined by the *local curvature* of the hydrogen bond PES, $V_{HB}''(r_0)$, and reflects the degree of proton delocalization (Figure 7, more in Sections S7 and S8):

$$\Delta r = \frac{eF_{ext}}{V_{HB}''(r_0)} \tag{1}$$

where r_0 is the most probably position of the proton wavefunction in the absence of an applied field, and Δ refers to the difference between situations with and without the external field. The displaced proton subsequently generates an induced dipole field ΔF_{HB} and effectively *amplifies* the external field F_{ext}, leading to a larger internal field F_{int} sensed by the GFP or PYP chromophore (Figure 7):

$$F_{\text{int}} = F_{\text{ext}} + \Delta F_{\text{HB}} + \Delta F_{non-HB} = \left[1 + \frac{2e^2}{4\pi\epsilon_0 r^3}\frac{1}{V_{HB}''(r_0)} + (f_{non-HB} - 1)\right]F_{ext} \equiv fF_{ext} \tag{2}$$

where r is the distance between the proton and the phenol(ate) oxygen (Section S7), and ϵ_0 is the vacuum permittivity. The induced fields from other sources are collectively denoted as $\Delta F_{\text{non-HB}}$ (= $(f_{\text{non-HB}} - 1)F_{\text{ext}}$) and factored in according to the superposition principle, assuming they are not simultaneously affected by the polarized short hydrogen bond. Note that the overall local field factor f reduces to $f_{\text{non-HB}}$ in the absence of a short hydrogen bond. Since the principal axes for the polarizable moieties are in general not colinear with externally applied fields, any local field factor should be a tensor rather than a scalar as shown in Equation 2, and we reserve a more explicit tensor treatment for

Section S7. The concept of proton polarizability was strongly advocated by Georg Zundel, who invoked the existence of such a notion for the pK_α-matched short hydrogen bond in the Zundel cation ($H_2O\cdots H^+\cdots OH_2$) to explain the broad continuum IR band of diluted aqueous acid due to the susceptibility of the proton position to environmental electrostatic fluctuations [73]. Along the same vein, Perrin and Lau showed that protons that are found to be equally shared between donors and acceptors of short hydrogen bonds in crystals could localize on either side owing to their sensitivity to the environment that leads to symmetry breaking [17][74], reminiscent of electron localization in the Creutz–Taube ion at 77 K [38].
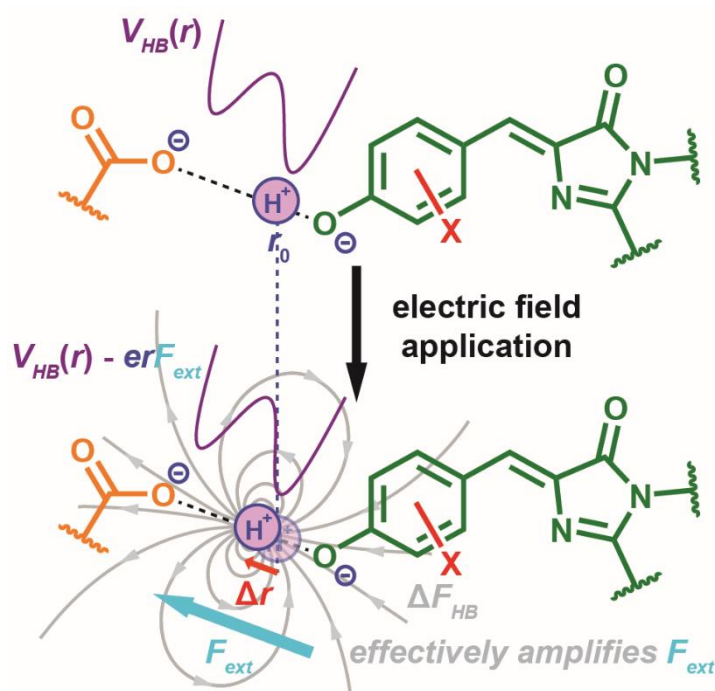
Figure 7. Proton polarization within the short hydrogen bond (Zundel polarizability) may lead to an increase in the local electric field sensed by the chromophore. Without the external field (top), the equilibrium position of the proton is r_0 according to the proton PES $V_{HB}(r)$ (purple curve). The externally applied field F_{ext} (cyan arrow) displaces the proton (bottom, displacement exaggerated) by perturbing the PES and thereby induces a dipole field ΔF_{HB} (gray field lines). Since the chromophore experiences both F_{ext} and ΔF_{HB} in the same direction, the proton polarization effectively amplifies the external field and consequently contributes to the local field factor f. Since shorter hydrogen bonds are expected to possess shallower PESs, the corresponding Zundel polarizability should be appreciable.

Based on Equation 2, the shallower the PES, which corresponds to shorter heavy-atom distances and/or better pK_α matching [13], the more delocalized the associated proton and the more polarizable the hydrogen bond, causing a larger local field factor f. Even though we argue that the PESs for the short hydrogen bond in the protonated S65T/H148D GFP variants are double-welled, the local curvatures at the energetic minima should be smaller than that of a normal hydrogen bond (Section S8). As the GFP chromophore can sense the delocalized proton within each well, smaller local curvatures could result in the relatively featureless electronic absorption bands observed from short-hydrogen-bond GFPs (Figure 3A). However, because the proton is closer to the chromophore for the A-like A state than the B-like A state, the former should have a larger f than the latter (Equation 2), which conflicts with the decrease in $f\mu_{CT}$ due to the short hydrogen bond for the A-like A state, indicating that the proton polarizability could not be the major contributor to the anomaly in the observed Stark tuning rates.

Since f is positively correlated with the degree of proton delocalization, a simple way to assess its existence is to measure the Stark tuning rates of proteins where protons are exchanged for deuterons. The consistent decrease in Stark tuning rates for both A-like and B-like A states of GFPs upon deuterium exchange in Table 2 looks very suggestive, as it can be readily understood as the deuteron being more localized than the proton and thus showing a smaller f (Section S8). Slight decreases in the Stark tuning rates of deuterated PYPs were also found as compared to the protonated counterparts (Figure 6B), lending support to this argument. However, by examining Figures 6A and 6C and viewing the GFP variant series as a whole rather than individually, the data points from deuterated samples do not follow a noticeably different trend when comparing with the protonated counterparts, which is in contradiction with a smaller f expected for deuterated species, so it appears that proton polarizability cannot explain the observed isotope effect. Rather, it could be plausibly rationalized as an electronic effect in response to the isotope substitution, the same origin as the electrostatic modulation of the chromophore's color [39][40]. Specifically, the SIE can explain this observation: because deuteration causes a blue and red shift of A-like and B-like A states, respectively [36], it reduces the Stark tuning rates in both cases based on the Stark tuning rate *vs.* absorption maximum correlation from the chromophores in their corresponding protonation states

(Figures 6A and 6C). The degree of Stark tuning rate decrease also correlates with the magnitude of the SIE (Table S10), further strengthening the plausibility of electrostatic modulation over proton polarizability. The anomalies in the observed Stark tuning rates from short-hydrogen-bond GFP variants can therefore be fully explicated with the covalency of the short hydrogen bond (Section 3.2.1), and the observed large $f\Delta\mu$ for PYPs can also be accounted for using a similar argument.

Even though it seems unlikely that the Zundel polarizability is a dominant mechanism for the perturbation from short hydrogen bonds in these proteins, it might be more prominent in systems with more delocalized protons, such as true LBHBs, where the anticipated polarization could be less masked by the substituents' electronic effect [75]. Past evidence for the Zundel polarizability has relied heavily on vibrational observables through IR or Raman spectroscopy to avoid substantial electronic polarization from the probes, where the electric fields were provided by condensed-phase environments (e.g., solvents, ions, and protein residues) [73] rather than externally applied. It would be very interesting to search for proteins with pre-calibrated vibrational probes [45] participating in short hydrogen bonds and to directly apply an external field in order to validate the existence of proton polarizability through enhanced Stark tuning rates within these proteins.

4.   CONCLUSIONS

Using the normal A and B states as well-calibrated electrostatic probes, we identify how the electron distribution of the chromophore is perturbed when engaging in a short hydrogen bond. This leads to a smaller Stark tuning rate for the protonated chromophore but enhances the Stark tuning rate for the deprotonated chromophores. The correlation between Stark tuning rates and absorption maxima obtained from these situations deviates from that expected from pure electrostatic color tuning. Based on the applicability of classical electrostatics, we are thus able to distinguish short hydrogen bonds from other noncovalent interactions, including normal hydrogen bonds, π–π stacking, and cation/anion–π interactions. Specifically, electrons are still confined within the π system of the chromophore when participating in noncovalent interactions, and this can be

treated well with classical electrostatics. Conversely, covalent modifications, such as halogenation and short hydrogen bonds, allow electrons to smear outside the original conjugation through exchange interactions due to the quantum nature of electrons and this is combined with the classical electrostatically induced electron polarization. While a large proton polarizability of the short hydrogen bond(s) could also be a plausible source of a larger-than-expected Stark tuning rate of the probing chromophore, we failed to collect strong evidence to support this claim from GFPs and PYPs. This is likely because the proton(s) involved is not delocalized enough such that the anticipated effect is masked by the electronic component. Furthermore, the small geometrical differences among chromophore variants (Section S4), even though designed to minimize this issue, limit our ability to selectively detect the proton polarizability effect since hydrogen bonds are especially sensitive to geometries.

With a proton-sensitive electronic probe, we also demonstrate how either electronic Stark or pre-resonance Raman spectroscopy can be a useful diagnostic tool for LBHBs. We use these methods to reinforce our previous argument that short-hydrogen-bond GFPs do not contain LBHBs regardless of how closely we match the donor–acceptor pK_α [36]. Given the ubiquity of short hydrogen bonds in proteins [22][23] in combination with the scarce examples characterized with LBHBs [22][27][28], there seems to be lack of an energetic advantage for LBHBs to exist, either due to the aforementioned symmetry breaking originating from solvation [74] or the difficulty of stabilizing a delocalized charge within polar environments [76][77]. Given the large proton polarizability within these short hydrogen bonds [73], it is evident how the corresponding structure–energetics relationship is strongly context dependent [10][18] [19][78][79][80][81], and the role of the environment is the key to resolving conflicting results for this long-standing problem. In this regard, more local vibrational probes could be more suitable for measuring the environmental subtleties [45] and identifying the appropriate parameter(s) for quantifying different environments of short hydrogen bonds.

ACKNOWLEDGMENTS

REFERENCES

[1]   Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: New York, 1991.

[2]   Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76.

[3]   Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books: Sausalito, California, 2006; pp 145–204.

[4]   Ishikita, H.; Saito, K. Proton transfer reactions and hydrogen-bond networks in protein environments. *J. R. Soc. Interface* **2014**, *11*, 20130518.

[5]   Pimentel, G. C.; McClellan, A. L. *The Hydrogen Bond*; W. H. Freeman: San Francisco, California, 1960.

[6]   Schuster, P.; Zundel, G.; Sandorfy, C. *The Hydrogen Bond: Recent Developments in Theory and Experiments*; North-Holland: Amsterdam, NL, 1976.

[7]   Gilli, G.; Gilli, P. *The Nature of the Hydrogen Bond: Outline of a Comprehensive Hydrogen Bond Theory*; Oxford University Press: Oxford, UK, 2009.

[8]   Smith, D. A. A Brief History of the Hydrogen Bond. In *Modeling the Hydrogen Bond*; Smith, D. A., Ed.; ACS Symposium Series 569; American Chemical Society: Washington, DC, 1994; pp 1–5.

[9]   Perrin, C. L. Are short, low-barrier hydrogen bonds unusually strong? *Acc. Chem. Res.* **2010**, *43*, 1550–1557.

[10]   Herschlag, D.; Pinney, M. M. Hydrogen bonds: Simple after all? *Biochemistry* **2018**, *57*, 3338-3352.

[11]   Perrin, C. L.; Nielson, J. B. "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.* **1997**, *48*, 511–544.

[12]   Gilli, P.; Pretto, L.; Bertolasi, V.; Gilli, G. Predicting hydrogen-bond strengths from acid-base molecular properties. The pK_a slide rule: Toward the solution of a long-lasting problem. *Acc. Chem. Res.* **2009**, *42*, 33–44.

[13]   McKenzie, R. H.; Bekker, C.; Athokpam, B.; Ramesh, S. G. Effect of quantum nuclear motion on hydrogen bonding. *J. Chem. Phys.* **2014**, *140*, 174508.

[14]   Sobczyk, L.; Grabowski, S. J.; Krygowski, T. M. Interrelation between H-bond and π-electron delocalization. *Chem. Rev.* **2005**, *105*, 3513–3560.

[15]   Fried, S. D.; Boxer, S. G. Thermodynamic framework for identifying free energy inventories of enzyme catalytic cycles. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12271–12276.

[16] Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chem. Rev.* **2016**, *116*, 7529–7550.

[17] Perrin, C. L. Symmetries of hydrogen bonds in solution. *Science* **1994**, *266*, 1665–1668.

[18] 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.

[19] 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.

[20] Thomson, B.; Both, J.; Wu, Y.; Parrish, R. M.; Martínez, T. J.; Boxer, S. G. Perturbation of short hydrogen bonds in photoactive yellow protein via noncanonical amino acid incorporation. *J. Phys. Chem. B* **2019**, *123*, 4844–4849.

[21] Meot-Ner, M. Update 1 of: Strong ionic hydrogen bonds. *Chem. Rev.* **2012**, *112*, PR22–PR103.

[22] Hosur, M. V.; Chitra, R.; Hegde, S.; Choudhury, R. R.; Das, A.; Hosur, R. V. Low-barrier hydrogen bonds in proteins. *Crystallogr. Rev.* **2013**, *19*, 3–50.

[23] Zhou, S.; Wang, L. Unraveling the structural and chemical features of biological short hydrogen bonds. *Chem. Sci.* **2019**, *10*, 7734–7745.

[24] Cleland, W. W.; Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **1994**, *264*, 1887–1890.

[25] Gerlt, J. A.; Kreevoy, M. M.; Cleland, W. W.; Frey, P. A. Understanding enzymic catalysis: The importance of short, strong hydrogen bonds. *Chem. Biol.* **1997**, *4*, 259–267.

[26] 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. USA* **2009**, 106, 440–444.

[27] Agback, P.; Agback, T. Direct evidence of a low barrier hydrogen bond in the catalytic triad of a serine protease. *Sci. Rep.* **2018**, *8*, 10078.

[28] Dai, S.; Funk, L.-M.; Rabe von Pappenheim, F.; Sautner, V.; Paulikat, M.; Schröder, B.; Uranga, J.; Mata, R. A.; Tittmann, K. Low-barrier hydrogen bonds in enzyme cooperativity. *Nature* **2019**, *573*, 609–613.

[29] Kreevoy, M. M.; Liang, T.; Chang, K.-C. Structures and isotopic fractionation factors of complexes AHA-1. *J. Am. Chem. Soc.* **1977**, *99*, 5207–5209.

[30] Gilli, P.; Bertolasi, V.; Ferretti, V.; Gilli, G. Covalent nature of the strong homonuclear hydrogen bond. Study of the O-H---O system by crystal structure correlation methods. *J. Am. Chem. Soc.* **1994**, *116*, 909–915.

[31] Gilli, G.; Gilli, P. Towards an unified hydrogen-bond theory. *J. Mol. Struct.* **2000**, *552*, 1–15.

[32] Hansen, P. E.; Spanget-Larsen, J. NMR and IR investigations of strong intramolecular hydrogen bonds. *Molecules* **2017**, *22*, 552.

[33] Feyer, V.; Prince, K. C.; Coreno, M.; Melandri, S.; Maris, A.; Evangelisti, L.; Caminati, W.; Giuliano, B. M.; Kjaergaard, H. G.; Carravetta, V. Quantum effects for a proton in a low-barrier, double-well potential: Core level photoemission spectroscopy of acetylacetone. *J. Phys. Chem. Lett.* **2018**, *9*, 521–526.

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

[35] Shi, X.; Abbyad, P.; Shu, X.; Kallio, K.; Kanchanawong, P.; Childs, W.; Remington, S. J.; Boxer, S. G. Ultrafast excited-state dynamics in the green fluorescent protein variant S65T/H148D. 2. Unusual photophysical properties. *Biochemistry* **2007**, *46*, 12014–12025.

[36] Oltrogge, L. M.; Boxer, S. G. Short hydrogen bonds and proton delocalization in green fluorescent protein (GFP). *ACS Cent. Sci.* **2015**, *1*, 148–156.

[37] Bublitz, G. U.; Boxer, S. G. Stark spectroscopy: Applications in chemistry, biology, and materials science. *Annu. Rev. Phys. Chem.* **1997**, *48*, 213–242.

[38] Silverman, L. N.; Kanchanawong, P.; Treynor, T. P.; Boxer, S. G. Stark spectroscopy of mixed-valence systems. *Philos. Trans. A Math. Phys. Eng. Sci.* **2008**, *366*, 33–45.

[39] Lin, C.-Y.; Romei, M. G.; Oltrogge, L. M.; Mathews, I. I.; Boxer, S. G. Unified model for photophysical and electro-optical properties of green fluorescent proteins. *J. Am. Chem. Soc.* **2019**, *141*, 15250–15265.

[40] Lin, C.-Y.; Boxer, S. G. Mechanism of color and photoacidity tuning for the protonated green fluorescent protein chromophore. *J. Am. Chem. Soc.* **2020**, *142*, 11032–11041.

[41] Anderson, S.; Crosson, S.; Moffat, K. Short hydrogen bonds in photoactive yellow protein. *Acta Cryst.* **2004**, *D60*, 1008–1016.

[42] Saito, K.; Ishikita, H. Energetics of short hydrogen bonds in photoactive yellow protein. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 167–172.

[43] Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*, 1st ed.; Cambridge University Press: Cambridge, UK, 2005; pp 593–661.

[44] Hunt, P. A.; Ashworth, C. R.; Matthews, R. P. Hydrogen bonding in ionic liquids. *Chem. Soc. Rev.* **2015**, *44*, 1257–1288.

[45] Fried, S. D.; Boxer, S. G. Measuring electric fields and noncovalent interactions using the vibrational Stark effect. *Acc. Chem. Res.* **2015**, *48*, 998–1006.

[46] Goeltz, J. C.; Kubiak, C. P. Mixed valency across hydrogen bonds. *J. Am. Chem. Soc.* **2010**, *132*, 17390–17392.

[47] Porter, T. M.; Heim, G. P.; Kubiak, C. P. Stable mixed-valent complexes formed by electron delocalization across hydrogen bonds of pyrimidinone-linked metal clusters. *J. Am. Chem. Soc.* **2018**, *140*, 12756–12759.

[48] Getzoff, E. D.; Gutwin, K. N.; Genick, U. K. Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation. *Nat. Struct. Biol.* **2003**, *10*, 663–668.

[49] Sugishima, M.; Tanimoto, N.; Soda, K.; Hamada, N.; Tokunaga, F.; Fukuyama, K. Structure of photoactive yellow protein (PYP) E46Q mutant at 1.2 Å resolution suggests how Glu46 controls the spectroscopic and kinetic characteristics of PYP. *Acta Cryst.* **2004**, *D60*, 2305–2309.

[50] 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.

[51] Memmi, S.; Kyndt, J.; Meyer, T.; Devreese, B.; Cusanovich, M.; Van Beeumen, J. Photoactive yellow protein from the halophilic bacterium *Salinibacter ruber*. *Biochemistry* **2008**, *47*, 2014–2024.

[52] 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.

[53] Oltrogge, L. M.; Wang, Q.; Boxer, S. G. Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry* **2014**, *53*, 5947–5957.

[54] Premvardhan, L. L.; van der Horst, M. A.; Hellingwerf, K. J.; van Grondelle, R. Stark spectroscopy on photoactive yellow protein, E46Q, and a nonisomerizing derivative, probes photo-induced charge motion. *Biophys. J.* **2003**, *84*, 3226–3239.

[55] Gurusaran, M.; Shankar, M.; Nagarajan, R.; Helliwell, J. R.; Sekar, K. Do we see what we should see? Describing non-covalent interactions in protein structures including precision. *IUCrJ* **2014**, *1*, 74–81.

[56] Treynor, T. P. Non-Classical Stark Effects of Photosynthetic Reaction Centers. Ph.D. Dissertation, Stanford University, Stanford, CA, 2003; pp 193–207.

[57] Bell, A. F.; He, X.; Wachter, R. M.; Tonge, P. J. Probing the ground state structure of the green fluorescent protein chromophore using Raman spectroscopy. *Biochemistry* **2000**, *39*, 4423–4431.

[58] He, X.; Bell, A. F.; Tonge, P. J. Isotopic labeling and normal-mode analysis of a model green fluorescent protein chromophore. *J. Phys. Chem. B* **2002**, *106*, 6056–6066.

[59] Fried, S. D.; Boxer, S. G. Evaluation of the evaluation of the energetics of the concerted acid–base mechanism in enzyme catalysis: The case of ketosteroid isomerase. *J. Phys. Chem. B* **2012**, *116*, 690–697.

[60] Olah, G. A. The σ-bridged 2-norbornyl cation and its significance to chemistry. *Acc. Chem. Res.* **1976**, *9*, 41–52.

[61] Olah, G. A.; Prakash, G. K. S.; Saunders, M. Conclusion of the classical-nonclassical ion controversy based on the structural study of the 2-norbornyl cation. *Acc. Chem. Res.* **1983**, *16*, 440–448.

[62] Schleyer, P. von R.; Mainz, V. V.; Strom, E. T. Norbornyl Cation Isomers Still Fascinate. In *The Foundations of Physical Organic Chemistry: Fifty Years of the James Flack Norris Award*; Strom, E. T., Mainz, V. V., Eds.; ACS Symposium Series 1209; American Chemical Society: Washington, DC, 2015; pp 139–168.

[63] Kanchanawong, P.; Dahlbom, M. G.; Treynor, T. P.; Reimers, J. R.; Hush, N. S.; Boxer, S. G. Charge delocalization in the special-pair radical cation of mutant reaction centers of *Rhodobacter sphaeroides* from Stark spectra and nonadiabatic spectral simulations. *J. Phys. Chem. B* **2006**, *110*, 18688–18702.

[64] Crutchley, R. J. Intervalence charge transfer and electron exchange studies of dinuclear ruthenium complexes. *Adv. Inorg. Chem.* **1994**, *41*, 273–325.

[65] Londergan, C. H.; Kubiak, C. P. Electron transfer and dynamic infrared-band coalescence: It looks like dynamic NMR spectroscopy, but a billion times faster. *Chem. Eur. J.* **2003**, *9*, 5962–5969.

[66] Day, P.; Hush, N. S.; Clark, R. J. H. Mixed valence: Origins and developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **2008**, *366*, 5–14.

[67] Marder, S. R.; Gorman, C. B.; Meyers, F.; Perry, J. W.; Bourhill, G.; Brédas, J.-L.; Pierce, B. M. A unified description of linear and nonlinear polarization in organic polymethine dyes. *Science* **1994**, *265*, 632–635.

[68] Pinney, M. M.; Natarajan, A.; Yabukarski, F.; Sanchez, D. M.; Liu, F.; Liang, R.; Doukov, T.; Schwans, J. P.; Martínez, T. J.; Herschlag, D. Structural coupling throughout the active site hydrogen bond networks of ketosteroid isomerase and photoactive yellow protein. *J. Am. Chem. Soc.* **2018**, *140*, 9827–9843.

[69] Arnold, W. D.; Oldfield, E. The chemical nature of hydrogen bonding in proteins via NMR: *J*-couplings, chemical shifts, and AIM theory. *J. Am. Chem. Soc.* **2000**, *122*, 12835–12841.

[70] Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **2014**, *346*, 1510–1514.

[71]   Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18454–18459.

[72]   McConnell, H. M. Theory of nuclear magnetic shielding in molecules: I. Long-range dipolar shielding of protons. *J. Chem. Phys.* **1957**, *27*, 226–229.

[73]   Zundel, G. Hydrogen Bonds with Large Polarizability and Proton Transfer Processes in Electrochemistry and Biology. In *Advances in Chemical Physics, Volume 111*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons, Inc.: New York, 2000; pp 1–217.

[74]   Perrin, C. L.; Lau, J. S. Hydrogen-bond symmetry in zwitterionic phthalate anions: Symmetry breaking by solvation. *J. Am. Chem. Soc.* **2006**, *128*, 11820–11824.

[75]   Janoschek, R.; Weidemann, E. G.; Pfeiffer, H.; Zundel, G. Extremely high polarizability of hydrogen bonds. *J. Am. Chem. Soc.* **1972**, *94*, 2387–2896.

[76]   Warshel, A.; Papazyan, A.; Kollman, P. A. On low-barrier hydrogen bonds and enzyme catalysis. *Science* **1995**, *269*, 102–106.

[77]   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. USA* **1996**, *93*, 13665–13670.

[78]   Scheiner, S.; Redfern, P.; Szczęśniak, M. M. Effects of external ions on the energetics of proton transfers across hydrogen bonds. *J. Phys. Chem.* **1985**, *89*, 262–266.

[79]   Trommsdorff, H. P. Localized and delocalized proton levels in hydrogen bonds: Fluorescence line narrowing and Stark effect measurements. *J. Lumin.* **1987**, *38*, 129–133.

[80]   Bała, P.; Lesyng, B.; Truong, T. N.; McCammon, J. A. Ab initio studies and quantum-classical molecular dynamics simulations for proton transfer processes in model systems and in enzymes. In *NATO Advanced Science Institutes Series C Volume 368: Molecular Aspects of Biotechnology: Computational Models and Theories*; Bertrán, J., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1992; pp 299–326.

[81]   Sigala, P. A.; Ruben, E. A.; Liu, C. W.; Piccoli, P. M. B.; Hohenstein, E. G.; Martínez, T. J.; Schultz, A. J.; Herschlag, D. Determination of hydrogen bond structure in water versus aprotic environments to test the relationship between length and stability. *J. Am. Chem. Soc.* **2015**, *137*, 5730–5740.
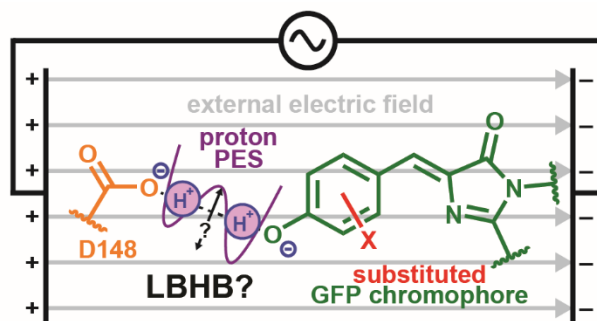
For Table of Contents Only

Supporting Information for

"Unusual Spectroscopic and Electric Field Sensitivity of a Chromophore with Short Hydrogen Bond: GFP and PYP as Model Systems"

Chi-Yun Lin[1] and Steven G. Boxer[1]*

[1]Department of Chemistry, Stanford University, Stanford, CA 94305, USA.

*Correspondence to: sboxer@stanford.edu

Table of Contents

S1   Sample Preparation

Plasmid Construction

The logic of GFP plasmid design followed our previous works on Superfolder GFPs [1][2][3]. Point mutations were made using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's protocol. The specific circular permutant ih:loop:GFP was chosen for short hydrogen bond studies because the noncanonical amino acids were originally introduced to residue 66 via semi-synthetic strategies with split GFPs [4]. However, amber suppression [5] was later found to be more efficient and afforded superior protein yields. To further increase the yields, we removed the proteolytic cleavage loop and restored cysteine at residue 70 since denaturation was unnecessary. We note that this study could have been carried out with any circular permutant at hand or even with *Aequorea victoria* GFPs (avGFPs) [6]. However, while retaining the short hydrogen bond of interest, some structural differences can be observed between the avGFP and ih:GFP S65T/H148D mutants (Section S4, Figure S4). Notice that the protein sequences given in the previous publication [1] were not entirely correct, so please refer to the next section for the exact sequences. The residue numbering scheme follows GFPs without circular permutation.

The original pQE plasmid with photoactive yellow protein (HhPYP) gene from *Halorhodospira halophila* was generously provided by Professor Marius Schmidt at University of Wisconsin Milwaukee and the SrPYP gene from *Salinibacter ruber* was obtained from the NCBI Gene database [7][8][9] and synthesized by GenScript. The genes were later cloned into pET-15b between the same restriction sites as previously reported GFPs to keep the endogenous His-tags and for better protein expression [10][11]. The HhPYP gene was redesigned according to the Xie lab at the Oklahoma State University [12]. The residue numbering scheme follows HhPYP [9].

GFP Constructs in This Study

We adopted the nomenclature devised for split GFP circular permutants in our previous works [4]. Labels describe elements (separated by colons) of GFP progressing

from the N terminus to the C terminus when read from left to right. Specific β-strands in the GFP β-barrel are denoted sX, where X is the number of the strand, while the internal helix is denoted ih. GFP refers to the remainder of the protein.

Table S1. GFP constructs in this study. Red letters denote non-wild-type amino acids. To facilitate readability, the mutation carried by the synthetic strand is enclosed by parentheses rather than superscripted as in our previous publications.

GFP Constructs	ih		s7	s10	s11
	65	66	148	203	222
ih:GFP S65T	T	Y	H	T	E
ih:GFP S65T Y66(2,3-F_2Y)	T	2,3-F_2Y	H	T	E
ih:GFP S65T Y66(3,5-F_2Y)	T	3,5-F_2Y	H	T	E
ih:GFP S65T with globally incorporated 3-F_1Y	T	3-F_1Y	H	T	E
ih:GFP S65T Y66(3-Cl_1Y)	T	3-Cl_1Y	H	T	E
ih:GFP S65T Y66(3,5-Cl_2Y)	T	3,5-Cl_2Y	H	T	E
ih:GFP S65T Y66(3-Br_1Y)	T	3-Br_1Y	H	T	E
ih:GFP S65T Y66(3-I_1Y)	T	3-I_1Y	H	T	E
ih:GFP S65T Y66(2,3,5-F_3Y)	T	2,3,5-F_3Y	H	T	E
ih:GFP S65T Y66(3-NO_2Y)	T	3-NO_2Y	H	T	E
ih:GFP S65T Y66(3-OMeY)	T	3-OMeY	H	T	E
ih:GFP S65T H148D	T	Y	D	T	E
ih:GFP S65T Y66(2,3-F_2Y) H148D	T	2,3-F_2Y	D	T	E
ih:GFP S65T Y66(3,5-F_2Y) H148D	T	3,5-F_2Y	D	T	E
ih:GFP S65T H148D with globally incorporated 3-F_1Y	T	3-F_1Y	D	T	E
ih:GFP S65T Y66(3-Cl_1Y) H148D	T	3-Cl_1Y	D	T	E
ih:GFP S65T Y66(3,5-Cl_2Y) H148D	T	3,5-Cl_2Y	D	T	E
ih:GFP S65T Y66(3-Br_1Y) H148D	T	3-Br_1Y	D	T	E
ih:GFP S65T Y66(3-I_1Y) H148D	T	3-I_1Y	D	T	E

DNA Sequences

ih:GFP C48S S65T H148D (with T65, Y66, D148, and T203 codons in bold)

```
ATGGGGCATCATCATCATCATCATAGCAGCGGCGGCAAACTACCGGTGCCGTGGCCGACCCTGG
TGACCACCTTAACCTATGGCGTGCAGTGCTTTAGCCGCTATCCGGATCATATGAAACGCCATGA
TTTTTTTAAAAGCGCGATGCCGGAAGGCTATGTGCAGGAACGCACCATTAGCTTTAAAGATGAT
GGCAAATATAAAACCCGCGCGGTGGTGAAATTTGAAGGCGATACCCTGGTGAACCGCATTGAAC
TGAAAGGCACCGATTTTAAAGAAGATGGCAACATTCTGGGGCATAAACTGGAATATAACTTTAA
CAGCGATAACGTGTATATTACCGCGGATAAACAGAAAAACGGCATTAAAGCGAACTTTACCGTG
CGCCATAACGTGGAAGATGGCAGCGTGCAGCTGGCGGATCATTATCAGCAGAACACCCCGATTG
GCGATGGCCCGGTGCTGCTGCCGGATAACCATTATCTGAGCACCCAGACCGTGCTGAGCAAAGA
TCCGAACGAAAAACGCGATCACATGGTGCTGCTGGAATTTGTGACCGCAGCGGGCATTACACAC
```

```
GGCATGGATGAACTGTATGGCGGCACCGGCGGCAGCGCGAGCCAGGGCGAAGAACTGTTTACCG
GCGTGGTGCCGATTCTGGTGGAACTGGATGGCGATGTGAACGGCCATAAATTTAGCGTGCGCGG
CGAAGGCGAAGGCGATGCGACCATTGGCAAACTGACCCTGAAATTTATTTCCACCACCTAA
```

HhPYP (with Y42 and E46 codons in bold)

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGG
AACACGTAGCCTTCGGTAGCGAGGACATCGAGAACACCCTCGCCAAGATGGACGACGGCCAGCT
CGACGGCCTGGCCTTCGGCGCCATCCAGCTCGACGGCGACGGCAACATCCTTCAGTACAACGCC
GCGGAGGGCGACATCACCGGCCGCGACCCGAAGCAGGTCATCGGCAAGAACTTCTTCAAGGACG
TGGCCCCGTGCACTGACAGCCCGGAGTTCTACGGCAAGTTCAAGGAAGGGGTGGCCTCGGGCAA
CCTGAACACGATGTTCGAGTACACCTTCGATTACCAAATGACGCCCACGAAGGTGAAGGTGCAC
ATGAAGAAGGCCCTCTCCGGCGACAGCTACTGGGTCTTCGTCAAGCGCGTCTAA
```

SrPYP (with Y42 and E46 codons in bold)

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGG
CTGACTCTCAGAATCCGTACTCGTACCTGCGTGAAGACGACCCGGACTCCGCCCCCGGCGACTC
GGGCGACGCCGACGAGCCTGAGCCGCCGGCAACGGACCTCGCGTTCGACGACGAGGGGGTGGGG
GAGGAGCTGCGTCACGTCGACGAGGACGAGCTAAATGCGGCCCCGTTCGGCATCATTCAGATTG
ACGACGCGGGGGTCGTGCAGTTCTACAACCGCTACGAGTCGAACCTAAGCGGCATCGATCCGGC
AGACGCCGTCGGGGCCAACTTCTTCACGGAGCTGGCGCCGTGCAGCAACAACCCGCTGTTCTTC
GGCCGCTTCAAGGACGGGGTTCGGGAGGGGGGGCTCGACGAGTACTTCACGTACACGTTCACCT
ACCAGATGCGCCCCACCCTCGTCGACGTGCGGCTCTACCGCGACGAGGCGGAGAACAACTGGAT
CCTGATTCAGAAGCGGTGA
```

Protein Sequences

ih:GFP C48S S65T H148D (with critical mutation sites in bold and tyrosines labeled in blue)

```
MGHHHHHHSSGGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDD
GKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSDNVYITADKQKNGIKANFTV
RHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMVLLEFVTAAGITH
GMDELYGGTGGSASQGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFISTT
```

Note that the incorrect sequence listed in reference [1] had the proteolytic cleavage loop inserted.

HhPYP (with thrombin cleavage site in red and Y42, E46, and C69 in bold)

```
MGSSHHHHHHSSGLVPR▼GSHMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYN
AAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKV
HMKKALSGDSYWVFVKRV
```

SrPYP (with thrombin cleavage site in red and Y42, E46, and C69 in bold)

```
MGSSHHHHHHSSGLVPR▼GSHMADSQNPYSYLREDDPDSAPGDSGDADEPEPPATDLAFDDEGV
GEELRHVDEDELNAAPFGIIQIDDAGVVQFYNRYESNLSGIDPADAVGANFFTELAPCSNNPLF
FGRFKDGVREGGLDEYFTYTFTYQMRPTLVDVRLYRDEAENNWILIQKR
```

Noncanonical Amino Acid Incorporation into Green Fluorescent Proteins

The protocol for synthesizing halogenated tyrosines is detailed in refs. [1][2][11]. All chemicals were purchased from Sigma Aldrich unless otherwise stated. The procedures for global incorporation of $3\text{-}F_1Y$ and amber suppression incorporation of $3\text{-}Cl_1Y$ (using ClYRS from Professor Jiangyun Wang at Chinese Academy of Sciences [13]), $3,5\text{-}Cl_2Y$ (using Cl2YRS from Professor Jiangyun Wang [13]) and $3,5\text{-}F_2Y$ (using $F_nY\text{-}RS$ E3 from Professor JoAnne Stubbe at MIT [14]) were previously described [1], but the inducer for the aminoacyl tRNA synthetases (aaRSs) should be L-arabinose (CAS 10323-20-3) instead. The expression protocol after induction was modified to be at 25°C with overnight shaking. Incorporation of $3\text{-}Br_1Y$ and $3\text{-}I_1Y$ were achieved using ClY-RS with slightly less efficiencies and $2,3\text{-}F_2Y$ was incorporated using E3 with a better yield than that of $3,5\text{-}F_2Y$. The proteins were all purified with Ni-NTA columns and anion exchange for spectroscopy and crystallography [1][10]. The final yields were usually around 10 – 50 mg per liter LB, with $3,5\text{-}F_2Y$ at the lower end and $3\text{-}Cl_1Y$ at the higher end with amber suppression in LB (cf. wild-type yield is around 100 mg per liter LB). The yield for global incorporation of $3\text{-}F_1Y$ in glucose minimal media was around 30 mg per liter media. The purity and identity of all proteins were confirmed with electrospray ionization mass spectrometry (ESI-MS) measured with LC-MS (Waters 2795 HPLC with ZQ single quadrupole MS in Stanford University Mass Spectrometry (SUMS) facility):

Table S2. Expected and observed mass for each GFP construct. Those not listed here were previously characterized [3].

GFP Mutants and Variants	Expected Mass[a] (Da)	Observed Mass[b] (Da)	
	Intact	Intact	N-terminus Modified[c]

ih:GFP	28041	28045	28224 (gluc)
ih:GFP Y66(3-Cl$_1$Y)	28076	28082	28260 (gluc)
ih:GFP Y66(3,5-Cl$_2$Y)	28110	28119	28296 (gluc)
ih:GFP Y66(3-Br$_1$Y)	28120	28124	28304 (gluc) 28380 (phosphogluc)
ih:GFP Y66(3-I$_1$Y)	28167	28174	28352 (gluc) 28430 (phosphogluc)
ih:GFP Y66(2,3-F$_2$Y)	28077	28083	28262 (gluc)
ih:GFP Y66(3,5-F$_2$Y)	28077	28081	28263 (gluc)
ih:GFP with global 3-F$_1$Y	28203	28207	-
ih:GFP H148D	28019	28023	28202 (gluc)
ih:GFP Y66(3-Cl$_1$Y) H148D	28054	28060	28238 (gluc)
ih:GFP Y66(3,5-Cl$_2$Y) H148D	28088	28098	28274 (gluc)
ih:GFP Y66(3-Br$_1$Y) H148D	28098	28102	28282 (gluc) 28360 (phosphogluc)
ih:GFP Y66(3-I$_1$Y) H148D	28145	28150	28330 (gluc) 28406 (phosphogluc)
ih:GFP Y66(2,3-F$_2$Y) H148D	28055	28061	28237 (gluc)
ih:GFP Y66(3,5-F$_2$Y) H148D	28055	28059	28240 (gluc)
ih:GFP H148D with global 3-F$_1$Y	28181	28185	28363 (gluc)

[a] Predicted from the primary sequence with N-terminal methionine removed [15].
[b] Proteins with ~ 30 kDa have ±10 Da deviations, depending on the protonation states.
[c] Proteins with His-tags tend to be modified at the N-terminus, such as gluconoylation (+178 Da, gluc) and phosophogluconoylation (+258 Da, phosphogluc) [16].

All GFPs are stored at 4°C after concentrating from anion exchange for future usage. The protocol for preparing the ih:GFP S65T Y66(2,3,5-F$_3$Y), Y66(3-NO$_2$Y), and Y66(3-OMeY) was described in our previous work [2].

$^{13}C_\zeta$-L-Tyrosine Global Incorporation into Green Fluorescent Proteins

The protocol has been described in our previous work [17] in detail, including the subsequent purification. The purity and identity of all proteins were characterized by ESI-MS as described above:

Table S3. Expected and observed mass for each ^{13}C-labeled GFP constructs.

GFP Mutants and Variants	Expected Mass[a] (Da)	Observed Mass[b] (Da)	
	Intact	Intact	N-terminus Modified[c]
ih:GFP with globally incorporated $^{13}C_\zeta$-L-tyrosine	28050	28055	28234 (gluc)
ih:GFP H148D with globally incorporated $^{13}C_\zeta$-L-tyrosine	28028	28034	28213 (gluc)

[a] Predicted from the primary sequence with N-terminal methionine removed [15].
[b] Proteins with ~ 30 kDa have ±10 Da deviations, depending on the protonation states.
[c] Proteins with His-tags tend to be modified at the N-terminus, such as gluconoylation (+178 Da, gluc) and phosophogluconoylation (+258 Da, phosphogluc) [16].

Photoactive Yellow Protein (PYP) Expression and Purification

The procedure was adapted from the Schmidt lab at University of Wisconsin, Milwaukee [18] and the Getzoff lab at the Scripps Research Institute [19]. The pET-15b plasmids containing the PYP genes were transformed into BL21(DE3) *E. coli* cells, which were then grown up in 2 L of Terrific Broth (TB) with 100 mg/L ampicillin at 37°C and shaken at 180 rpm. TB was chosen to maximize the protein yields per liter media, especially for double mutants that were harder to be expressed. At OD_{600} 0.6 – 0.7, induction was performed with 0.25 g/L isopropyl β-D-1-thiogalactopyranoside (IPTG; Santa Cruz Biotechnology, CAS 367-93-1), and the media was cooled down to 20°C for overnight expression. The cells were pelleted the next day by centrifugation at 6500 rcf for 30 min, and resuspended in lysis buffer, an aqueous buffer at pH 7.5 containing 50 mM potassium phosphate monobasic, 300 mM sodium chloride (Fisher, CAS 7647-14-5) and 5 mM imidazole (Aldrich, CAS 288-32-4). The cells were subsequently lysed with a high-pressure homogenizer, and the resulting lysate was centrifuged twice at 25000 rcf

for 1 hr each. In the meanwhile, 220 mg of *p*-coumaric acid (*p*CA; CAS 501-98-4) was activated with 240 mg of 1,1'-carbonyldiimidazole (CDI; CAS 530-62-1) in 30 mL anhydrous tetrahydrofuran (THF; CAS 109-99-9) under argon at room temperature for at least 1 h. THF was then removed under reduced pressure using a rotary evaporator (Laborota 4000; Heidolph). The remaining yellowish oil or solid of activated *p*CA was incubated with the apoprotein supernatant from the spun down lysate at 20°C and shaken at 200 rpm for at least 90 min. The mixture turned turbid and bright yellow (or light yellow depending on the yield of PYP), which was again centrifuged at 25000 rcf for 30 min to pellet down the insoluble *p*CA. The Ni-NTA column and anion exchange purification followed our normal GFP protocol [10]. Notably, while GFPs elute around 15% – 20% B (150 – 200 mM NaCl) at pH 8.0 for anion exchange, HhPYPs elute at 5% B (50 mM NaCl), and SrPYPs elute at 40% B (400 mM NaCl) due to its high aspartate and glutamate content. His-tags were then removed by adding 1 – 2 units of bovine thrombin (Calbiochem, plasminogen-free) per mg of PYP in buffer A for anion exchange (low salt and imidazole free) and allowed to incubate for 4 h at room temperature. The mixture was subjected to another anion exchange purification to remove thrombin, and the resulting proteins were used for subsequent spectroscopic study. Typical yields range from 50 – 200 mg per liter TB, with Y42F E46Q mutants at the lower end. The purity and identity of all proteins were characterized by ESI-MS as described above.

Table S4. Expected and observed mass for each PYP constructs.

PYP Mutants	Expected Mass[a] (Da)	Observed Mass[b] (Da)
HhPYP	14301	14304
HhPYP E46Q	14300	14303
HhPYP Y42F	14285	14288
HhPYP Y42F E46Q[c]	14284	14140 (apo), 14287
SrPYP	17960	17965
SrPYP E46Q	17959	17963
SrPYP Y42F	17944	17948
SrPYP Y42F E46Q[c]	17943	17801 (apo), 17947

Predicted from the primary sequence with N-terminal His-tag removed. The chromophore increases the mass by 146 Da.
[b] Proteins with ~18 kDa have ±5 Da deviations, depending on the protonation states.
[c] Partial incorporation of pCA were observed, which could be due to insufficient pCA activation.

All PYPs are stored at 4°C after concentrating from anion exchange for future usage.

Synthesis of the PYP Model Chromophore: Methyl Thioester of pCA

The procedure of synthesizing the PYP model chromophore has been described elsewhere [20]. Instead of THF, equal moles of pCA and CDI were mixed together in anhydrous dimethylformamide (CAS 68-12-2) under argon at room temperature for at least 1 h. One equivalent of sodium thiomethoxide (Acros, CAS 5188-07-88) was added subsequently and the solution immediately turned bright orange. The mixture was allowed to stir overnight under argon at room temperature and applied to a silica gel column pre-equilibrated with petroleum ether (Fisher, CAS 101316-46-5). The column was then washed with 2 column volumes of petroleum ether and eluted with 2 column volumes of 1:1 mixture of petroleum ether and ethyl acetate (Fisher, CAS 141-78-6). The fractions were pooled according to visual inspection of the yellow color and concentrated under vacuum with a rotary evaporator followed by a Schlenk line. The resulting product was a bright yellow claylike solid, and was characterized via UV–vis (for both neutral and anionic states) [21] and ESI-MS. The observed mass was 193 Da in the negative ion mode, which was 1 Da less than the expected mass due to deprotonation.

Buffers for Spectroscopic Studies

The pH 5, pD 5, pH 10 buffers for proteins contain 30 mM phosphate, 30 mM citrate and 200 mM NaCl to achieve a broad buffering range. To ensure consistency and minimal proton contamination for deuterated buffer, the buffers were made from anhydrous sodium phosphate tribasic (Aldrich, CAS 7601-54-9) and sodium citrate dihydrate (CAS 6132-04-3), and then titrated to the correct pH (pD) with hydrochloric acid (or 35% deuterium chloride in D_2O (CAS 7698-05-7) with the consideration of $pD = pH_{app} + 0.42$

[22]. The pH 8 buffer is 50 mM Tris hydrochloride (Fisher, CAS 1185-53-1) and 200 mM NaCl titrated with sodium hydroxide. The pH 10 buffer for the PYP model chromophore is 50 mM sodium carbonate (CAS 497-19-8) titrated with hydrochloric acid. The guanidinium buffers for chromophore titration were prepared as previously described [1].

Sample Preparation for Raman Spectroscopy

Even with the confocal microscope setup and pre-resonance condition with 633 nm (or 638 nm) excitation, protein samples for Raman spectroscopy need to be more than 3 mM to achieve a good signal-to-noise ratio. Therefore, the purified GFPs were concentrated with spin filters (Amicon Ultra 0.5 mL, 10 kDa cutoff) and thoroughly exchanged into the appropriate buffers more than 5 times. The deuterated proteins were characterized within one or two days after deuterium exchange, or slow back exchange from ambient moisture can be observed.

Sample Preparation for 77 K Absorption and Electronic Stark Spectroscopy

Glass forming solvents, such as 1:1 mixture of glycerol and aqueous buffer, are required for low-temperature electronic Stark spectroscopy experiments. Protein samples were concentrated and exchanged into appropriate buffers as described for Raman spectroscopy. The concentrated samples were then mixed with an equal volume of glycerol (Fisher, CAS 56-81-5) or glycer(ol-d_3) (CAS 7325-16-8) in the case of deuterated samples right before Stark measurements. For SrPYPs in acidic buffers, however, glycerol was added immediately after concentration to stabilize the proteins due to significant precipitation observed over time. For the PYP model chromophore, due to its relatively low solubility in the carbonate buffer, the glycerol buffer mixture was made first to dissolve the chromophore. The sample was also freshly prepared right before measurement to avoid base-catalyzed hydrolysis of the thioester. The final sample concentrations for Stark spectroscopy were checked with a NanoDrop spectrometer (ND-1000 Spectrometer; NanoDrop) to ensure a maximum absorbance of 0.2 – 0.9 for a 25 μm path length, the optimum OD for good signal-to-noise ratio in low temperature absorption.

S2   Spectroscopic Methods

UV–vis Absorption Measurements

UV–vis absorption spectra at room temperature were all measured with a PerkinElmer Lambda 25 UV–vis spectrometer and a 1 mL quartz cuvette. Data acquisition was performed every 1.0 nm at a maximum scan rate of 480 nm/min.

Low-Temperature (77 K) Absorption Measurements and Electronic Stark Spectroscopy

The detailed method has been reviewed in our previous works [3][23]. The cell consisted of a pair of 12.7 mm diameter by 1 mm thick fused silica windows (FOCtek Photonics, Inc.) deposited with 45 Å of nickel on the surfaces facing the sample. The windows were separated by a pair of 27-micron thick Teflon spacers and held in place with a metal clamp and four adjustable screws. The interference fringes were optimized under a fluorescent lamp, and the path length was determined by the undulations in UV–vis absorption from 500 – 1100 nm. The path length was then used to calculate the electric field strength applied during the measurement knowing the applied voltage. The Stark cell was mounted onto a home-built rod with electrical wires and alligator clips attached to the nickel electrodes. The whole apparatus was insulated with electrical tape, and a sample (at most 10 µL) was loaded into the cell by capillary uptake. The whole rod was then rapidly plunged into an immersion cryostat [24] pre-filled with liquid nitrogen to allow the sample to form a transparent glass upon flash freezing. Protein samples with glycerol were centrifuged at 17000 rcf for at least 40 min prior to sample loading.

The custom-built spectrometer could be switched between Stark spectroscopy and absorption modes with the latter dual-beamed. For Stark spectroscopy, the sinusoidal high voltage signal was generated from the sample channel lock-in amplifier (SR830; Stanford Research) with a frequency of 203 Hz and amplified 1000-fold via a high-voltage power supply (TREK 10/10; TREK), and the voltage was applied through the rod onto the sample. The root-mean-square voltage (V_{rms}) applied before dielectric breakdown can range from 0.6 – 2.7 kV, which amounts to a peak external field strength F_{ext} of 0.3 – 1.4

MV/cm given the sample thickness. The X and Y components of Stark signal $\Delta I\,(2\omega)$ were detected at the second harmonic (2ω) of the applied field. The direct output voltage I was also simultaneously recorded. The Stark spectra were then obtained from the ratio: $\Delta A(2\omega) = \frac{2\sqrt{2}}{\ln 10}\frac{\Delta I(2\omega)}{I}$ as a function of the scanning wavelengths [23]. Higher-order Stark spectra [23], specifically 4ω, can be obtained by $\Delta A(4\omega) = \frac{8\sqrt{2}}{\ln 10}\frac{\Delta I(4\omega)}{I}$. A wavelength scan rate of 0.3 nm/s and a time constant of 300 ms were chosen. The polarized probe light was set to be horizontal, and depolarization along the beam path was carefully checked. χ angles between the applied electric field and the polarization of the probe light at 90°, 70°, and 50° were sampled at each applied field strength with an increment of 0.3 kV in V_{rms} to ensure a complete data set.

For absorption spectroscopy, the sample channel was reconfigured by replacing the polarizer with another beam chopper, and the reference channel was employed. The magnitudes of output signals were detected at the first harmonic of the chopper modulation frequency (3029 Hz). The scanning rate and time constant were set to match those of the Stark measurements. The blank sample was prepared by carefully blowing the Stark sample out of the cell with air and then loading the cell with a 1:1 buffer and glycerol mixture. The absorbance A was determined at normal incidence with an absolute uncertainty around ±0.01. The final absorbance was obtained by averaging over three to four scans for each sample. LabView programs were used to facilitate data collection in both modes. Undulation can be occasionally seen in the baseline at the red-edge of the absorption spectra due to light interference between two windows of the sample cell.

Note that since wild-type HhPYP and its E46Q mutant still undergo significant photoisomerization at 77 K [25], the samples were scanned several times from 300 to 600 nm to reach the photostationary state before data acquisition (Figure S13). This was unnecessary for SrPYPs since a significant increase in fluorescence and no appreciable spectral change were observed upon freezing and subsequent irradiation.

Stark Spectroscopy Data Analysis

All Stark spectra ΔA are shown with their corresponding absorbance A normalized. Occasionally, the spectra are concentration normalized to facilitate comparison between protonated and deuterated samples. ΔA are also scaled to 1 MV/cm with $\chi = 90°$ (χ is the angle between probe light polarization and field direction), according to its proportionality to $(F_{ext})^2$ and $(F_{ext})^4$ for the 2ω and 4ω spectra, respectively (see Figures S12 and S15). The Stark spectra (as functions of wavenumbers $\bar{v}$) were analyzed as linear combinations of wavenumber-weighted zeroth, first, and second derivatives of the absorbance spectra with coefficients A_χ, B_χ, and C_χ as functions of χ, respectively, to extract the apparent Stark tuning rates $\Delta\mu_{app}$ (= $|\Delta\vec{\mu}_{app}|$) and the measured angles ζ between difference dipoles and transition dipoles [23]:

$$\Delta A(\bar{v}, F_{ext}) = A(\bar{v}, F_{ext}) - A(\bar{v}, F_{ext} = 0)$$

$$= (F_{ext})^2 \left[A_\chi A(\bar{v}) + \frac{B_\chi}{15hc} \bar{v} \frac{d}{d\bar{v}} \left(\frac{A(\bar{v})}{\bar{v}} \right) + \frac{C_\chi}{30h^2c^2} \bar{v} \frac{d^2}{d\bar{v}^2} \left(\frac{A(\bar{v})}{\bar{v}} \right) \right] \qquad (S1)$$

and

$$C_\chi = \left(\Delta\mu_{app} \right)^2 [5 + (3\cos^2\chi - 1)(3\cos^2\zeta - 1)] \qquad (S2)$$

where F_{ext} is the strength of the externally applied field through the parallel-plate capacitor. The magnitude of a vector quantity is denoted by dropping the vector notation. The data were processed by the MATLAB code kindly provided by Professor Robert Stanley at Temple University [26]. Simultaneous fitting of ΔA and A at $\chi = 90°$, 70°, and 50° were performed with a minimal number of Gaussian components and their analytical derivatives to model the vibronic progression and effectively smooth the absorbance spectra. No real physical meaning is associated with the individual peak positions of these fit Gaussians, and any attempt to do so should be treated with great caution. One set of electro-optical parameters ($\Delta\mu_{app}$, ζ, A_χ, and B_χ) was first assigned to recapitulate the transition with the dominant Stark effect. More bands were employed only if the result from the one-band fit was unsatisfactory (Section S5). Due to nonnegligible contribution from $\Delta\mu_{app}$, no attempt was made to isolate the difference polarizabilities $\Delta\alpha$ from B_χ. The uncertainties in $\Delta\mu_{app}$ from both fitting and duplicates amounted to ±5%, while those in ζ

were ±5°, unless the bands were too small ($A < 0.1$) to be properly analyzed. Throughout this study, $\Delta\mu_{app}$ was treated as the product of the true difference dipole moment of the chromophore $\Delta\mu$ and the local field factor f, with the latter assumed to be a constant scalar across different environments. The necessity of including f reflects our lack of certainty over the magnitude of the local field sensed by the chromophore [23] (see also Section S6 in [3]).

Extinction Coefficient Determination of GFP Variants

Precise determination of protein concentrations in this study was only necessary to facilitate the comparison between protonated and deuterated Stark samples for the short-hydrogen-bond GFP variants, and thus for PYP mutants only normalized absorbances were shown due to almost identical absorption features for protonated and deuterated species. The apparent Stark tuning rates can still be obtained with ΔA and A without knowing the sample concentrations (Equation S1).

Concentrations of GFPs with the wild-type chromophore (Y66) were determined by measuring the UV–vis absorbance at 447 nm in 0.2 M NaOH (Fisher BioReagents, CAS 497-19-8) and scaling by the known extinction coefficient of the deprotonated chromophore in the denatured protein (44100 $M^{-1}cm^{-1}$) [27]. For variants with halogenated chromophores, we have determined the relevant values in reference [2] assuming that the corresponding samples are 100% pure, such that the extinction coefficients can be normalized against the 290 nm absorption from the protein (mostly the deprotonated tyrosine [28]) in 0.2 M NaOH (27000 $M^{-1}cm^{-1}$) for wild-type ih:GFP S65T. The extinction coefficients for the halogenated variants in 0.2 M NaOH are shown in Table S5.

Table S5. Extinction coefficients for denatured GFP variants in 0.2 M NaOH.

Variants	Extinction coefficient at the visible absorption maximum ($M^{-1}cm^{-1}$)	Absorption maximum (nm)
WT	44100	447
Y66(3-F_1Y)	35700	443

Y66(3-Cl$_1$Y)	36000	445
Y66(3-Br$_1$Y)	35300	447
Y66(3-I$_1$Y)	33900	451
Y66(2,3-F$_2$Y)	36000	438
Y66(3,5-F$_2$Y)	36100	438
Y66(3,5-Cl$_2$Y)	29500	444

Pre-resonance Raman Spectroscopy

The sample cell for Raman spectroscopy was constructed by attaching a single-well gasket (cut from Secure-Seal spacer; Molecular Probes) to a 24 mm × 50 mm glass cover slide (VWR). Vacuum grease (Dow Corning) was applied around the gasket and 2 – 3 μL highly concentrated protein samples were added to the center of the gasket. A second cover slip was then dropped onto and lightly pressed against the assembly to form a tightly sealed sample cell allowing minimal evaporation.

The Raman spectra were acquired with a Horiba LabRAM HR Evolution (excitation at 633 nm) or a Horiba XploRA+ (excitation at 638 nm) Raman confocal microscope to achieve a pre-resonance condition for preferentially enhancing the chromophore signals over the protein background. The instruments were pre-calibrated with silicon, and the samples were focused on with a 100x objective and irradiated with 20 – 30 mW diode lasers for 20 min. The spectra were obtained from averaging over 20 scans of 1 min accumulations. No significant changes in Raman spectra were observed after strong laser illumination. The data were then baselined to remove the constant background. Baseline undulation can be occasionally seen from the interference between the two slides, but it does not affect our data interpretation.

S3   X-ray Crystallography

Protein Crystallization

All short-hydrogen-bond GFPs (ih:GFP S65T/H148D) whose structures were not previously reported [1] (globally incorporated 3-F_1Y, Y66(2,3-F_2Y), Y66(3,5-F_2Y), Y66(3-Br_1Y), and Y66(3-I_1Y)) were crystallized in similar conditions as described in [1] (i.e., 50 mM sodium acetate, 100 mM sodium chloride, ~ 5% (wt/vol) PEG 3350, *pH 5.0*), with slight differences in PEG 3350 concentrations (*vide infra*). Three additional GFP variants (ih:GFP S65T Y66(2,3,5-F_3Y), Y66(3-NO_2Y), and Y66(3-OMeY)) were also crystallized in similar conditions (i.e., 150 mM ammonium acetate, ~ 30% (wt/vol) PEG 3350) as reported before [3]. All solutions were made with highest possible concentration stocks and sterile filtered with 0.22 μm syringe filters (EMD Millipore Millex). Proteins were all exchanged into the anion exchange buffer A (50 mM Tris-HCl, 10 mM NaCl, pH 8.0) and concentrated to 10 mg/mL based on the Nanodrop absorbance measurement. The concentrated protein stocks were centrifuged at 17000 rcf for 10 min prior to use to remove potential dusts and particulates. Hanging drop trials were all manually performed with 24-well VDX plates with sealant (Hampton Research) and 22 mm siliconized glass circle cover slides (Hampton Research). Specifically, 1 μL protein stock was mixed with 1 μL mother liquor on the glass slide, which was then flipped and sealed above the well with 500 μL mother liquor for each condition.

For short-hydrogen-bond GFPs (S65T/H148D), an initial attempt with a sparse screen of various concentrations of PEG 3350 (Hampton Research, CAS 25322-68-3), 50 mM sodium acetate (J. T. Baker, CAS 127-09-3), 100 mM NaCl, and pH 5.0 gave almost no hit except for Y66(3-Br_1Y) with multicrystalline broom-like appearances at conditions with 15% PEG or above after a week. The Y66(3-Br_1Y) and Y66(3-Cl_1Y) (previously obtained [1]) crystals were pipetted into 1.5 mL eppendorf tubes, diluted with the corresponding mother liquor and crushed with seed beads (Hampton Research) with a vortexer. The seed stocks were subsequently serially diluted further with mother liquor by $10^2 - 10^7$ fold and streak seeded into trays from the initial attempt with cat whiskers. All seeded drops (except for those with PEG 3350 concentrations lower than 5%) turned completely colorless overnight and green crystals with diverse sizes depending on the

seed concentrations were observed under a microscope. This suggests the nucleation processes were highly stochastic. Microseeding gave rise to higher quality crystals, but the appearances were still far from perfect. Lower PEG concentrations tended to form single crystals without appreciable defects, but too low of a concentration only resulted in clear drops. Further iterations of streak seeding were performed with pre-equilibrated drops with finely adjusted PEG concentrations and new seed stocks of the corresponding mutants to find the optimal point. Y66(3-F_1Y) and Y66(3,5-F_2Y) were crystallized best in 25% PEG, while Y66(3-Br_1Y), Y66(3-I_1Y), and Y66(2,3-F_2Y) required 7 – 8% PEG. Crystals were plate like with sizes around 0.2 mm for the longest dimension.

For normal GFPs (S65T), initial attempts with a sparse screen of 26% – 34% PEG and 150 mM ammonium acetate (CAS 631-61-8) led to precipitation of Y66(2,3,5-F_3Y), Y66(3-NO_2Y), and Y66(3-OMeY) and no crystal formation after overnight equilibration, as opposed to the Y66 counterpart [3]. Previously obtained ih:GFP S65T seed stocks were prepared as described above, further diluted with mother liquor by 10^1 – 10^2 fold, and streak seeded into trays from the initial attempt with cat whiskers. Crystals formed in all drops for Y66(2,3,5-F_3Y) and only formed in conditions with 28 – 34% PEG for Y66(3-NO_2Y) and Y66(3-OMeY) in two days. Note that lower PEG concentration (< 30%) tended to yield broom-like crystals, while higher concentration gave rise to much better results but still required effort to isolate single crystals. Crystals were plate like with sizes around 0.1 mm for the longest dimension (the best ones were usually found hidden under the protein precipitates).

The best-looking crystals were looped with 0.1 – 0.2 mm Mounted CryoLoops (Hampton Research), dipped into their corresponding cryoprotectants and flash frozen into a Stanford Synchrotron Radiation Lightsource (SSRL) style cassette [29] within a week after crystal formation. For ih:GFP S65T/H148D Y66(3-F_1Y) and Y66(3,5-F_2Y), the original cryoprotectant was used (2 M sucrose, 50 mM sodium acetate, 100 mM NaCl, 5% PEG 3350 and, pH 5.0), but it became white upon freezing, indicating undesired ice formation. An improved version of cryoprotectant was used for the 3-Br_1Y and 3-I_1Y counterparts, namely 30% glycerol, 50 mM sodium acetate, 100 mM NaCl, 9% PEG and pH 5.0, which led to better diffraction. The crystals of the 2,3-F_2Y counterpart were

observed to crack due to osmotic shock upon transferring to this cryoprotectant, so perfluoropolyether cryo oil (Hampton Research, CAS 69991-67-9) was used instead and resulted in superior diffraction (~1.2 Å). The cause of these cracks were later found to correlate with electron density for glycerol in the ih:GFP S65T/H148D Y66(3-Br$_1$Y) and Y66(3-I$_1$Y) structures (Section S4, Figure S8). From these experiences, only the cryo oil was used for ih:GFP S65T Y66(2,3,5-F$_3$Y), Y66(3-NO$_2$Y), and Y66(3-OMeY).

X-ray Data Collection and Structure Determination

X-ray data for all crystals were acquired at SSRL (Menlo Park, CA) [30] at 100 K. Specifically, diffraction of ih:GFP S65T/H148D Y66(3-F$_1$Y) and Y66(3,5-F$_2$Y) were collected at BL 7-1, Y66(2,3-F$_2$Y) at BL 9-2, and the rest at BL 12-2 (microfocus). Data processing was performed with XDS [31][32] using the autoxds script [33]. Chromophore restraint files were built using REEL and eLBOW in PHENIX. Molecular replacement was performed in PHENIX [34] using the ih:GFP S65T H148D structure (PDB: 4ZF3) as the search model. Numerous rounds of model building and refinement were carried out with Coot [35] and PHENIX. The resulting data collection and refinement statistics are summarized in Table S6. More discussion on protein structures can be found in Section S4.

Table S6. X-ray data collection and refinement statistics for ih:GFP S65T/H148D and ih:GFP S65T variants.

	ih:GFP S65T/H148D global 3-F$_1$Y	ih:GFP S65T/H148D Y66(3,5-F$_2$Y)	ih:GFP S65T/H148D Y66(3-Br$_1$Y)	ih:GFP S65T/H148D Y66(3-I$_1$Y)	ih:GFP S65T/H148D Y66(2,3-F$_2$Y)
PDB entry	**6OG8**	**6OG9**	**6OGA**	**6OGB**	**6OGC**
Data collection statistics					
beamline	BL 7-1		BL12-2		BL 9-2
wavelength (Å)	1.127		0.9795		0.9795
detector distance (mm)	200	200	300	275	198
resolution range (Å)	27.28 – 1.599 (1.657 – 1.599)	36.59 – 1.798 (1.864 –1.798)	34.3 – 1.6 (1.657 – 1.6)	36.18 – 1.65 (1.709 – 1.65)	36.5 – 1.18 (1.222 – 1.18)
space group	P 1 2$_1$ 1 (No.4)				
unit cell dimensions a, b, c (Å) α, β, γ (°)	50.18, 70.32, 61.26 90.0, 95.9, 90.0	50.12, 70.13, 61.10 90.0, 96.3, 90.0	51.60, 68.60, 61.45 90.0, 99.8, 90.0	51.80, 68.69, 61.64 90.0, 99.8, 90.0	52.28, 68.87, 60.97 90.0, 100.3, 90.0
total observations	189867 (10094)	142519 (13435)	312673 (31570)	289832 (26882)	1761544 (123970)
unique reflections	54415 (4414)	38845 (3852)	54833 (5471)	50444 (4904)	134643 (11607)
multiplicity	3.5 (2.3)	3.7 (3.5)	5.7 (5.8)	5.7 (5.5)	13.1 (10.6)
completeness (%)	97.2 (79.7)	98.8 (97.6)	98.3 (98.8)	98.3 (95.6)	95.3 (83.7)
mean I/σI	14.2 (1.4)	12.6 (1.6)	13.4 (1.4)	13.1 (1.6)	29.2 (3.8)
Wilson B-factor (Å^2)	12.2	14.3	21.2	21.2	9.9
R$_{merge}$	0.068 (0.60)	0.105 (0.741)	0.074 (1.09)	0.080 (0.96)	0.048 (0.55)
R$_{meas}$	0.081 (0.79)	0.123 (0.87)	0.081 (1.20)	0.087 (1.06)	0.050 (0.57)
R$_{pim}$	0.043 (0.51)	0.064 (0.45)	0.033 (0.49)	0.036 (0.44)	0.014 (0.17)
CC$_{1/2}$	0.998 (0.546)	0.996 (0.644)	0.999 (0.592)	0.999 (0.656)	1 (0.917)

CC*	0.999 (0.840)	0.999 (0.885)	1 (0.863)	1 (0.890)	1 (0.978)
Refinement statistics					
reflections used	54370 (4412)	38639 (3852)	54821 (5471)	50437 (4904)	132849 (11588)
reflections used for R_{free}	2719 (221)	1934 (193)	2742 (274)	2523 (246)	6634 (578)
R_{work}	0.1593	0.1690	0.1630	0.1559	0.1449
R_{free}	0.1897	0.2006	0.1886	0.1778	0.1621
chromophore three-letter code	**MFC**	**MHY**	**B2C**	**I1C**	**MJ1**
number of non-H atoms	4709	4476	4094	4137	4604
protein	3890	3731	3672	3702	3877
ligand	92	48	76	96	72
solvent	727	697	346	339	655
protein residues	455	452	455	460	462
RMSD bond lengths (Å)	0.012	0.006	0.008	0.007	0.015
RMSD bond angles (°)	1.47	1.07	0.93	0.87	1.68
Ramachandran favored (%)	98.50	98.63	98.64	99.10	99.11
Ramachandran allowed (%)	1.50	1.37	1.36	0.90	0.89
Ramachandran outliers (%)	0.00	0.00	0.00	0.00	0.00
rotamer outliers (%)	0.94	1.43	0.98	1.95	1.16
clashscore	5.70	4.13	6.62	6.80	5.95
Average B-factor (Å^2)	16.25	18.87	28.32	28.05	15.63
protein	14.12	17.01	27.45	27.08	13.82
ligand	8.17	11.67	33.60	34.56	8.89
solvent	28.72	29.31	36.37	36.82	27.09

	ih:GFP S65T Y66(2,3,5-F$_3$Y)	ih:GFP S65T Y66(3-NO$_2$Y)	ih:GFP S65T Y66(3-OMeY)
PDB entry	**6UN5**	**6UN6**	**6UN7**
Data collection statistics			
beamline	BL12-2		
wavelength (Å)	0.9795		
detector distance (mm)	188	188	200
resolution range (Å)	38.93 – 1.36 (1.409 – 1.36)	38.53 – 1.5 (1.554 –1.5)	34.76 – 1.5 (1.554 – 1.5)
space group	P 1 2$_1$ 1 (No.4)		
unit cell dimensions a, b, c (Å) α, β, γ (°)	48.39, 68.46, 60.60 90.0, 102.1, 90.0	47.95, 67.72, 58.36 90.0, 102.3, 90.0	50.88, 68.60, 60.19 90.0, 101.6, 90.0
total observations	1118532 (108386)	762543 (73068)	846173 (81526)
unique reflections	81555 (8115)	57445 (5413)	63506 (6186)
multiplicity	13.7 (13.3)	13.3 (13.0)	13.3 (13.1)
completeness (%)	98.2 (98.1)	98.0 (93.4)	97.8 (96.1)
mean I/σI	22.9 (0.7)	17.0 (1.1)	20.3 (1.1)
Wilson B-factor (Å^2)	19.3	21.8	22.0
R$_{merge}$	0.113 (1.80)	0.074 (1.98)	0.056 (2.53)
R$_{meas}$	0.118 (1.87)	0.077 (2.06)	0.058 (2.63)
R$_{pim}$	0.031 (0.50)	0.021 (0.56)	0.016 (0.72)
CC$_{1/2}$	0.998 (0.605)	0.999 (0.588)	1 (0.620)
CC*	1 (0.868)	1 (0.861)	1 (0.875)

Refinement statistics			
reflections used	81513 (8113)	57231 (5412)	63409 (6184)
reflections used for R_{free}	4075 (404)	2862 (271)	3171 (310)
R_{work}	0.1764	0.1603	0.1699
R_{free}	0.2079	0.1807	0.1909
chromophore three-letter code	**QC4**	**QCA**	**QCD**
number of non-H atoms	4106	4023	4061
protein	3727	3710	3733
ligand	50	50	48
solvent	329	263	280
protein residues	454	456	456
RMSD bond lengths (Å)	0.005	0.007	0.007
RMSD bond angles (°)	1.18	1.27	1.28
Ramachandran favored (%)	98.86	98.64	98.19
Ramachandran allowed (%)	1.14	1.13	1.81
Ramachandran outliers (%)	0.00	0.23	0.00
rotamer outliers (%)	2.33	0.47	0.46
clashscore	5.70	3.49	5.57
Average B-factor (Å^2)	32.20	31.12	37.57
protein	31.85	30.78	37.51
ligand	21.90	21.62	25.24
solvent	37.65	37.68	40.36

S4   Discussion on GFP Structures

Introduction of halogen(s) to the GFP chromophore results in asymmetries in the substitution pattern, except for the 3,5-substituted ones. Therefore, it is important to determine their structures to learn whether dual conformers are present and to facilitate the spectroscopic analyses. In combination with our previous study [1], here we are able to obtain structures with medium-to-high (1.2 – 1.9 Å) resolution for all short-hydrogen-bond GFPs (Figure S1). Intriguingly, the heavy halogens in the monobrominated (PDB: 6OGA) and the monoiodinated chromophore (PDB: 6OGB) assume *syn* orientations with respect to the double-bonded nitrogen of the imidazolinone ring, consistent with the observed monochlorinated counterpart (PDB: 4ZF4, [1]), also adopted by normal GFPs with monosubstituted chromophores (PDB: 6OFL (Cl), 6OFM (CH_3), 6UN6 (NO_2), and 6UN7 (OMe), [3] and this work). In contrast, the *anti* conformers are identified in the corresponding Dronpa2 variants (PDB: 6NQL (Cl), 6NQN (Br), 6NQO (I), 6NQR (NO_2), 6NQS (OMe), and 6NQV (CH_3), [2]). It is not clear what exact factors in the chromophores' immediate environment, either based on these structures or some aspect of chromophore maturation, lead to different preferences in the substituent orientation for GFP and Dronpa2, also noted in our prior experience working with other monohalogenated proteins, e.g., ketosteroid isomerases (KSIs) [36][37] and HhPYP [11]. For example, it is tempting to conclude that the substituents prefer to stay in a pocket with more polar residues, such as T203 and E222 in GFP, but the consistent trend observed from CH_3 variants compared with others is inconsistent with this explanation. In addition, in the monochlorinated rsEGFP2 (PDB: 6PFR, [38]), another protein derived from GFP, we observe an *anti* conformer instead, disfavoring the intuitive arguments of steric or electrostatic environmental factors.

As opposed to bulkier substituents, the fluorines in the Y66(2,3-F_2Y) variant (PDB: 6OGC) show *anti* orientations, which are also adopted by the fluorines in the 2,3-F_2 variant of Dronpa2 (PDB: 6NQP, [2]). Similarly, the *anti* orientation is also adopted by the 2-position fluorines of the 2,3,5-F_3 variants for both ih:GFP (PDB: 6UN5) and Dronpa2 (PDB: 6NQQ, [2]). All these observations confirm the hypothesized steric clash between

the 2-position substituent and the imidazolinone nitrogen [40], causing the peculiar preferences in the *anti*-orientation for the asymmetric multifluorinated chromophores.
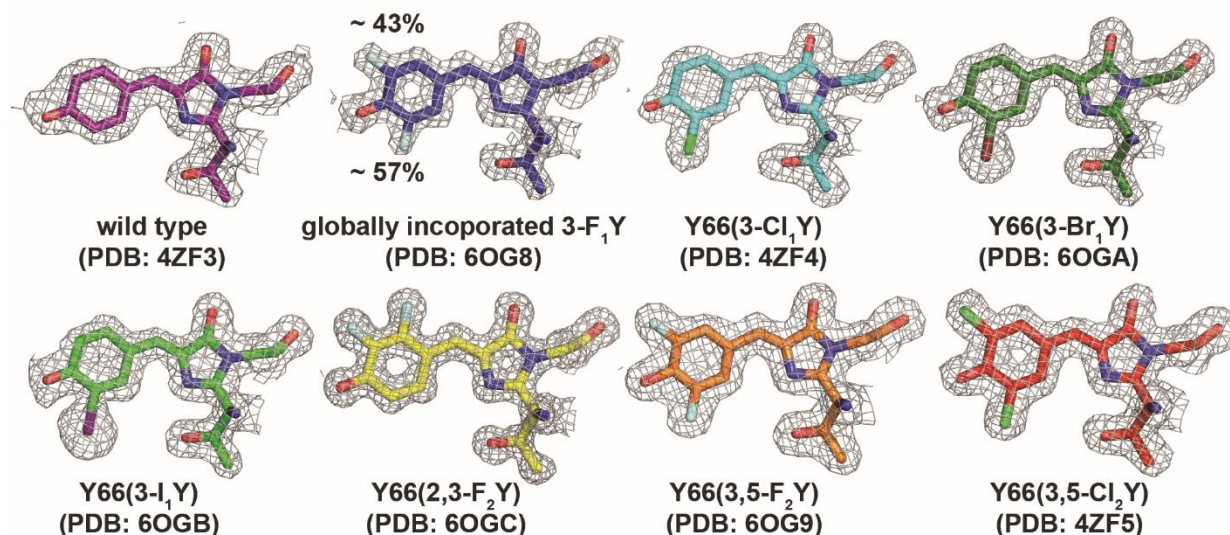

Figure S1. Orientations and populations of the halogen atom(s) for halogenated chromophores in short-hydrogen-bond GFPs, shown with their corresponding $2mF_o–DF_c$ maps contoured at 1σ. The color coding is consistent with Figure 1C. The variations in mesh densities among different chromophores are due to their different resolutions.

Among all substituted chromophores, only the monofluorinated chromophore assumes both *anti* and *syn* conformations and with nearly equal populations. This has also been observed in *Aequorea victoria* GFP (avGFP) with globally incorporated $3-F_1Y$ [39] and Dronpa2 with site-specifically incorporated $3-F_1Y$ at the chromophore (PDB: 6NQK, [2]), the latter confirmed by an ^{19}F NMR measurement [2]. As noted in the Stark analysis of the B-state chromophore (Section S11 in [3]) in Dronpa2 $Y63(3-F_1Y)$, the fitting does not demand a second set of electro-optic parameters to account for the two conformers, suggesting that these two conformers are spectroscopically indistinguishable in terms of electronic absorption. Based on this, we only use a single set of electro-optic parameters when fitting the Stark spectrum for the short-hydrogen-bond GFP with globally incorporated $3-F_1Y$ (Section S5, Figure S9).

Since we have the structure for ih:GFP S65T H148D with globally incorporated $3-F_1Y$ at hand (Figure S2), it is interesting to compare the orientation and population for each $3-F_1Y$ with those determined by both X-ray crystallography [39][40] and ^{19}F NMR [41] from avGFP with globally incorporated $3-F_1Y$. Even within the same protein, $3-F_1Y$

can be either locked at one conformation (3-F$_1$Y at residues 74, 106, and 200) or exhibit two conformers with almost equal populations (3-F$_1$Y at residues 66, 92, 143, and 151). Residue 182 shows the opposite behavior for the two chains in the unit cell, suggesting a high variability even at a certain position. When one compares these observations with those drawn from the avGFP counterparts using 100 K X-ray crystallography (Table S7), they are remarkably consistent in conformer orientations even though the two proteins are far from identical (pH 5.0 *vs* pH 8.5; H148D *vs* H148; in addition to circular permutation and tens of mutations), implying the existence of some cryptic principles that govern these behaviors. Based on the ^{19}F NMR study, residues 151 and 182, which exhibit dual conformers, show population exchange on a faster timescale (< ms), while residues 92 and 143 belong to the slow-exchange regime. Pal *et al.* [39] hypothesized that it is the solvent accessibility that dictates 3-F$_1$Ys' rotameric behavior, where exposed 3-F$_1$Ys tend to show dual conformers with fast exchange rates while buried 3-F$_1$Ys assume a single rotamer locked by local interactions. This is evidently not the full story since it cannot explain some observations (e.g., dual conformers for the fluorinated chromophore). It could be informative if more proteins with globally incorporated 3-F$_1$Y can be studied in this comprehensive manner [42], though results from X-ray crystallography could be biased by the crystal packing. For example, Pal *et al.* also noticed that the surface-exposed residue 200 does not exhibit two orientations and suggested that this is prevented by crystal contacts. Temperature-varying solution-phase ^{19}F NMR could also be complementary to the crystallographic approach, but more dual conformers imply an increasing difficulty for peak assignments, which could be relieved by more sophisticated 2D ^{19}F–^{19}F EXSY experiments [43].
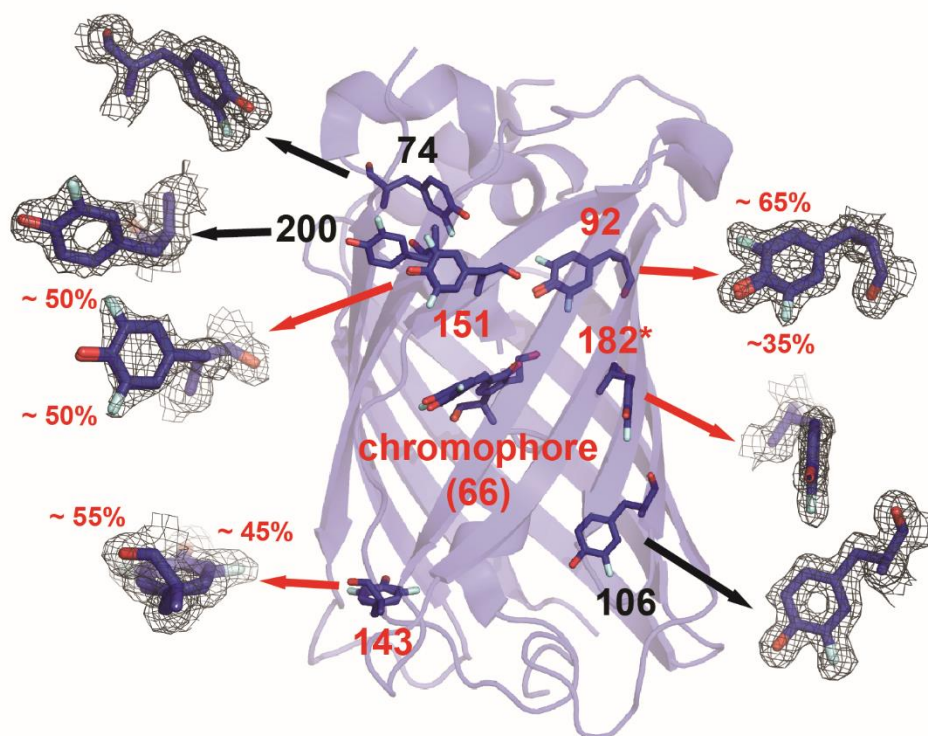
Figure S2. Orientations and populations of the fluorine atom for each 3-F_1Y in ih:GFP S65T H148D with globally incorporated 3-F_1Y (PDB: 6OG8, chain B). The numbering follows that of the avGFP. Residue numbers associated with singly and doubly oriented 3-F_1Y are colored black and red, respectively. The populations are averaged from two chains within the asymmetric unit, except for 182 (labeled with asterisk), at which 3-F_1Y assumes one conformer for chain B (shown) and 50:50 population ratio for chain A (not shown). The corresponding 2mF_o–DF_c maps contoured at 1σ are also shown, except for the chromophore, which is shown in Figure S1. Even though there are nine 3-F_1Ys in total, one is not resolved at the C–N linker for circular permutation, so the corresponding 3-F_1Y is non-existent in the fluorinated avGFP (Table S7). A similar figure for the avGFP counterpart can be found in [40], though not all residues are shown. The corresponding structure was deposited in PDB as 1RRX, but unfortunately the structure factors were not documented and residues with dual conformers were only modeled as single [40].

Table S7. Comparison of the conformational behaviors of 3-F_1Y at different positions in ih:GFP S65T/H148D (this work) and avGFP [39][41], characterized by X-ray crystallography and ^{19}F NMR. The words "dual" are colored red to be consistent with the convention in Figure S2. Solvent accessibility is qualitatively inferred from the X-ray structures and the presence of photo-chemically induced dynamic nuclear polarization (CIDNP) in ^{19}F NMR. Even though the ^{19}F chemical shift is known for its sensitivity to the environment [44], it is still technically challenging to assign the extensively overlapping NMR peaks from 10 fluorine atoms, and Khan *et al.* had to rely on exhaustive mutations

from tyrosine to phenylalanine for non-perturbative cases and photo-CIDNP to differentiate 3-F_1Y at different positions.

residue	strand	solvent accessibility	ih:GFP S65T/H148D	avGFP	
			X-ray crystallography (100 K, pH 5.0)	X-ray crystallography (100 K, pH 8.5) [40]	^{19}F NMR (300 K, pH 7.2) [42]
39	loop 2-3	exposed	N/A (Y39I)	N/A (not mentioned)	single (CIDNP)
66 (chromophore)	ih	buried	dual		not resolved
74	loop ih-4		single		
92	4		dual		
106	5		single		
143	loop 6-7	partially exposed	dual		
145	7	buried	N/A (Y145F)	single	
151		exposed	dual		single (CIDNP)
182	9		dual (chain A) single (chain B)	N/A (not resolved)	
200	10		single		

3-F_1Ys are not the only residues that show two conformers; in the atomic-resolution (1.18 Å) structure of ih:GFP S65T/H148D Y66(2,3-F_2Y) (PDB: 6OGC), E222 also occupies two rotameric states with a nearly 50:50 population ratio (Figure S3). The phenol oxygen of the chromophore engages with two branches of hydrogen bonding networks, one of which is the short hydrogen bond with D148 that we are interested in throughout this study and the other is the long network encompassing the phenol oxygen, T203, water, T205 (S205 in avGFP), E222, and finally the T65 oxygen of the chromophore. The latter is the excited-state proton transfer (ESPT) chain, which is defunct in the case of GFPs harboring S65T chromophores but can be efficiently rescued by the H148D mutation [45]. Even though the traditional ESPT chain is disabled, the hydrogen bonding network is still clearly present and complete in the crystal structure shown here. To complete the hydrogen bond network, E222 exhibits two conformers: one forms a hydrogen bond (2.7 Å) with T205 (rotamer A) and the other (rotamer B) interacts with T65

and a water via two hydrogen bonds (2.7 Å and 2.6 Å, respectively). With only rotamer A, the O–O distance between E222 and T65 becomes 3.4 Å, which is slightly too long for an effective hydrogen bond; with only rotamer B, the hydrogen bond between T205 and E222 is broken. This tridentate electron density for E222 has been explicitly documented three times in the literature for EGFP without the H148D mutation [46][47][48] and has been conjectured to cause two distinct fluorescence lifetimes for EGFP [49]. A recently reported nearly ultrahigh-resolution (0.85 Å) structure of GFP S65T (PDB: 6JGI, [50]) also shows E222 modeled with two rotamers, albeit a population ratio of 4:1. Given the importance of E222's alternate conformations, it is curious how both rotamers are not routinely captured by X-ray crystallography and observed in the PDB, especially for the 1.15 Å ih:GFP S65T structure we previously obtained (PDB: 6OFK, [3]).
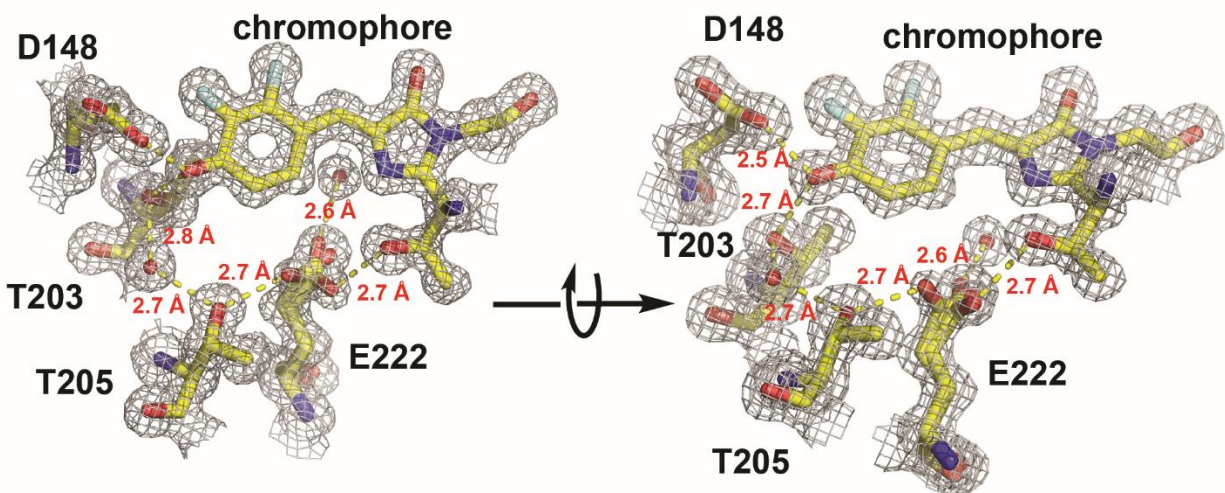
Figure S3. The local interactions between the doubly fluorinated chromophore and its neighboring residues (including water) in ih:GFP S65T/H148D Y66(2,3-F$_2$Y) (PDB: 6OGC). The electron density map 2mF$_o$–DF$_c$ contoured at 1σ is also shown. The hydrogen bonding network is denoted with yellow dashed lines and the corresponding hydrogen bond lengths are labeled in red. E222 assumes two rotamers to complete the hydrogen bonding network from the phenol oxygen to the T65 oxygen of the chromophore.

Close examination of the structures also allows us to better understand how the chromophore substitution perturbs the geometry of the chromophore and its immediate environment. First, we consider the effects of circular permutation and the short hydrogen bond from the H148D mutation. We noted in our previous work [3] that except for certain loop regions where the peptide chain is disconnected, circular permutation has a negligible impact on the global and local structures of GFPs with anionic chromophores.

By contrast, while the T203's oxygen diverts away from the protonated chromophore as seen in both avGFP S65T H148 and H148D at acidic pHs (Figure S4, left panel), a structural change well-recognized in the early literature of GFP [51], T203 retains its hydrogen bonding with the chromophore in the ih:GFP counterparts (Figure S4, middle and right panels). The hydrogen bond between T203 and the chromophore is also consistently observed across all halogenated H148D variants (Figures S3 and S5, except for some partial occupancy of the other T203 rotamer in ih:GFP S65T H148D with globally incorporated 3-F_1Y) and seems to be part of the defunct ESPT network. In addition, even though the short hydrogen bond between D148 and the chromophore survives when switching the background from avGFP to ih:GFP, the angle of D148 has twisted such that the aspartate overlays better with the original histidine.
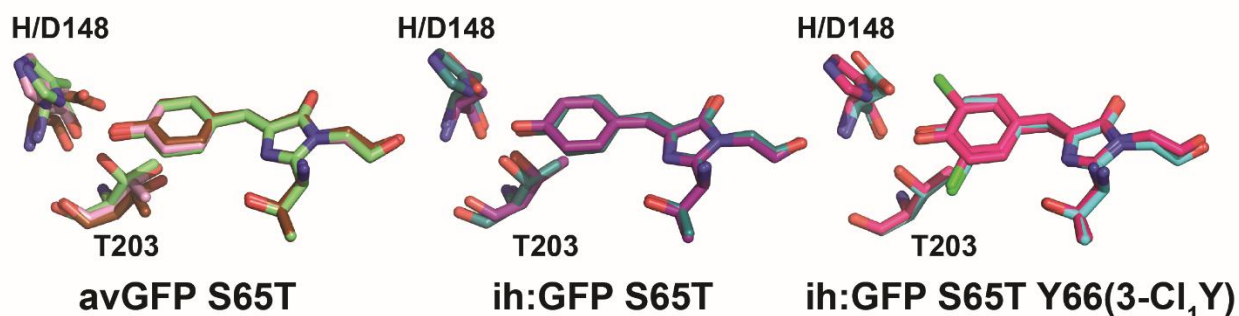
Figure S4. Overlay of X-ray structures for avGFP S65T (not circularly permuted, left), ih:GFP S65T (middle), and ih:GFP S65T Y66(3-Cl_1Y) (right) mutants shown with the same perspective. The left panel includes avGFP S65T at pH 8.5 (PDB: 1Q4A, lime, [52]), avGFP S65T at pH 5.5 (PDB: 1Q4B, pink, [52]), and avGFP S65T H148D at pH 5.6 (PDB: 2DUF, brown, [6]); the middle panel includes ih:GFP S65T at pH 7.0 (PDB: 6OFK, aqua, [3]) and ih:GFP S65T H148D at pH 5.0 (PDB: 4ZF3, purple, [1]); the right panel includes ih:GFP S65T Y66(3-Cl_1Y) at pH 7.0 (PDB: 6OFL, magenta, [3]) and ih:GFP S65T Y66(3-Cl_1Y) H148D at pH 5.0 (PDB: 4ZF4, cyan, [1]). Residues 148 (H or D) and 203 are labeled. Note that there are two chlorine orientations in ih:GFP S65T Y66(3-Cl_1Y) [3].

While chromophore substitution leads to almost no change in the *anionic* chromophore's immediate environment, as noted in [3] and except for some notable exceptions for E222 and T205 in some cases (Figure S5B), significant displacements up to 1 Å in residues participating in the hydrogen bonding network (i.e., D148, T203, T205, and E222) can be identified when comparing structures from ih:GFP S65T/H148D chromophore variants in the anomalous A state (Figure S5A). The latter structural rearrangement seems to be a direct consequence of the need to accommodate the

greater span of the substituted chromophores' dihedral angles φ along the P bond (see [53] for the definition), which can be evidently inferred from the relatively conserved O–O distances within the ESPT chains and the short hydrogen bonds across the variants (Figure S6) except for those involving the aforementioned E222 rotamer. In contrast, the imidazolinone side, including Q94 and R96, is much less affected upon chromophore substitution. It might be tempting to conclude that the protonated chromophore has more single-bond character for the P bond than the deprotonated chromophore and thus is more susceptible to any structural perturbation. However, as we will see next, the structural overlay can be misleading and more quantitative analysis informs us that the deprotonated chromophore does not in fact show a smaller range of φ upon substitution.

In Table S8 we list distances and angles extracted from our X-ray structures along with other related ones found in the PDB. The geometry of the interaction between residue H148 and the chromophore is of central importance in this study, and especially because the qualitative behavior of the hydrogen bond is strongly sensitive to its length [54], we also estimate the uncertainties in the O–O distance from the X-ray crystallography and the subsequent structural modeling according to Gurusaran *et al.* [55] (see footnote b, Table S8). We can only infer static geometries from the crystal structures and thus no dynamic inference can be drawn from the following analysis. For the deprotonated chromophore, its hydrogen bond length with H148 is consistently around 2.8 – 2.9 Å and the relative angle at which they interact is in a narrow range between 136° – 140° irrespective of differences in substitutions. By contrast, a slightly larger variations in these quantities (2.3 – 2.6 Å; 119° – 130°) can be found when examining X-ray structures of short-hydrogen-bond GFP variants, which could be due to the slightly worse overall resolutions in these structures but is more likely caused by the stronger perturbations from chromophore substitution.
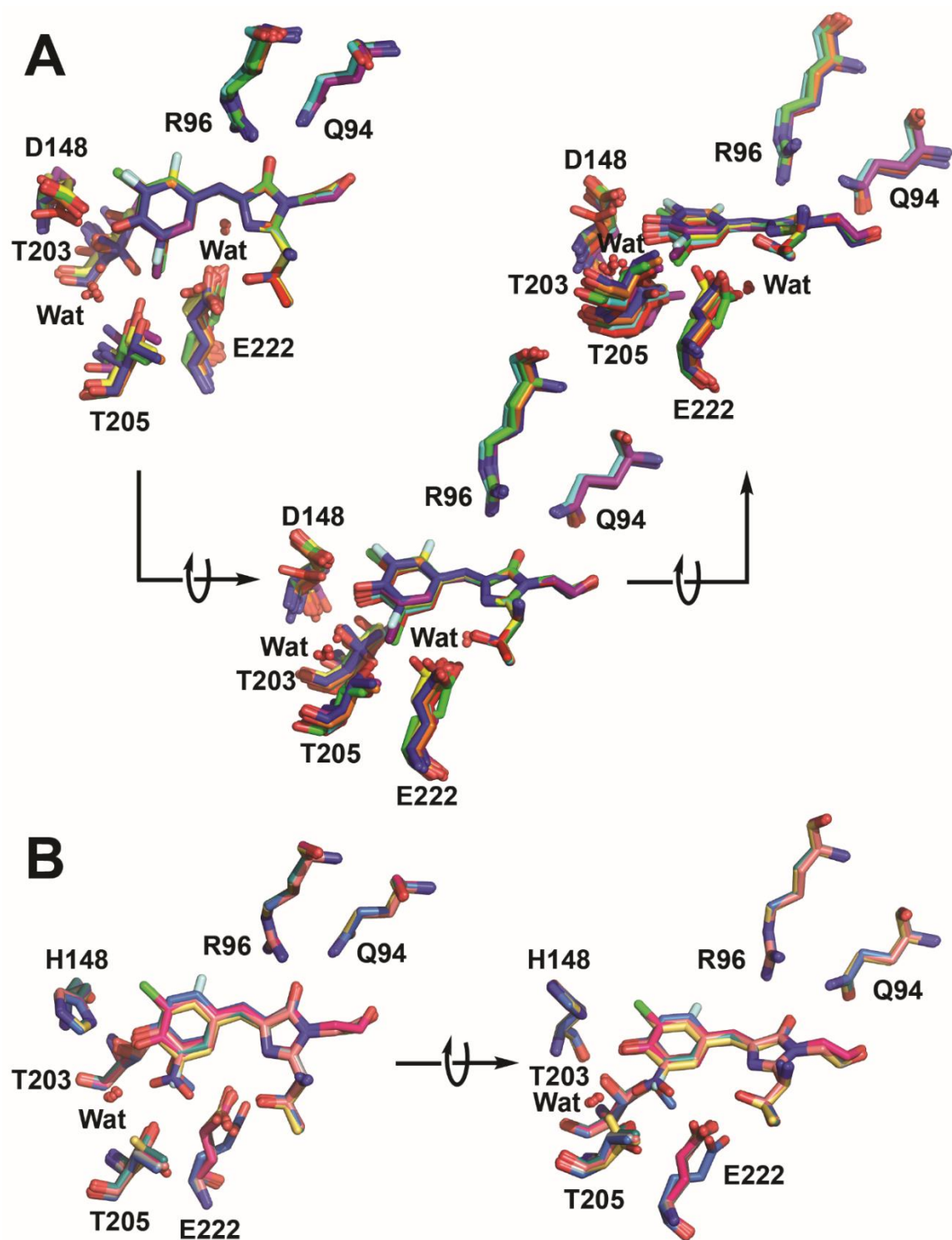
Figure S5. Overlay of X-ray structures of (A) ih:GFP S65T H148D chromophore variants at pH 5.0 and (B) ih:GFP S65T H148 chromophore variants and pH 7.0, with critical residues and waters labeled and various perspectives shown. Panel A contains all structures listed in Figure S1, namely Y66 (PDB: 4ZF3, purple, [1]), globally incorporated 3-F₁Y (PDB: 6OG8, dark blue, this work), Y66(3-Cl₁Y) (PDB: 4ZF4, cyan, [1]), Y66(3-Br₁Y)

(PDB: 6OGA, dark green, this work), Y66(3-I$_1$Y) (PDB: 6OGB, green, this work), Y66(2,3-F$_2$Y) (PDB: 6OGC, yellow, this work), Y66(3,5-F$_2$Y) (PDB: 6OG9, orange, this work), and Y66(3,5-Cl$_2$Y) (PDB: 4ZF5, red, [1]). Panel B contains Y66 (PDB: 6OFK, aqua, [3]), Y66(3-Cl$_1$Y) (PDB: 6OFL, magenta, [3]), Y66(2,3,5-F$_3$Y) (PDB: 6UN5, pale yellow, this work), Y66(3-NO$_2$Y) (PDB: 6UN6, salmon, this work), Y66(3-CH$_3$Y) (PDB: 6OFM, silver, [3]), and Y66(3-OMeY) (PDB: 6UN7, blue, this work). In stark contrast to ih:GFPs carrying anionic chromophores (panel B), which are mostly reasonably overlaid, short-hydrogen-bond ih:GFPs show a large variation in the chromophore's environment across the halogenated species.
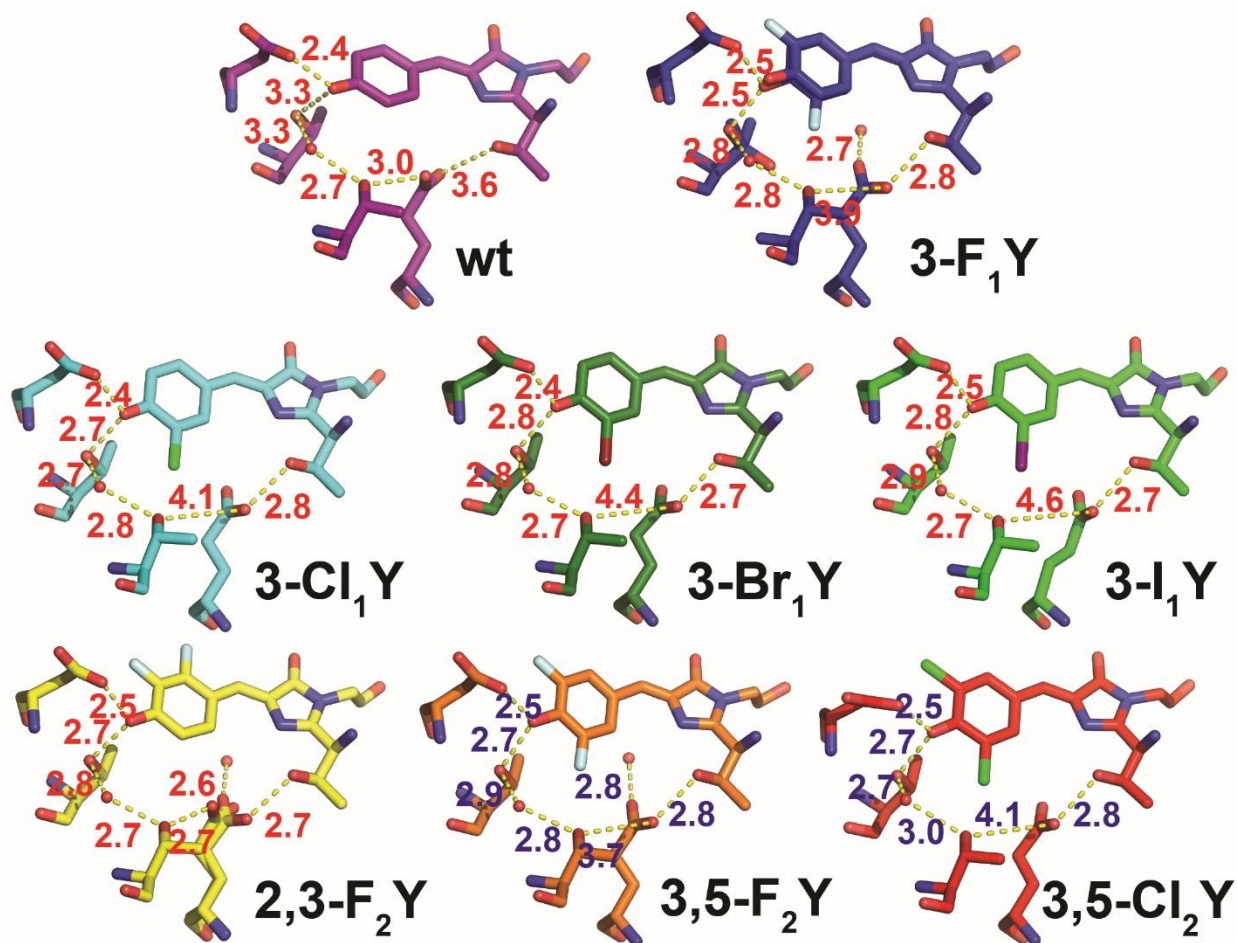
Figure S6. Hydrogen-bonding network for ih:GFP S65T H148D halogenated chromophore variants, denoted by the corresponding identity of residue 66 and shown with the same perspective. The color scheme is consistent with Figures S1 and S5. The hydrogen-bonding network are explicitly shown in yellow dashed lines and the O–O distances are labeled with numbers in Å.

Table S8. Geometry of the chromophore and its relative position to its hydrogen bonding partner H/D148 in GFPs mutants and variants. For crystal structures with the asymmetric units containing two monomers, the average values from the two chains are also provided and shown in bold; otherwise the values from the single monomer are shown in bold to facilitate comparison.

species and structure	resolution (Å)	chain and rotamer[a]	residue 148-chromophore O–O/N–O distance (Å)[b]	residue 148 O/N–phenol O–C$_\zeta$ angle	φ along the P bond	τ along the I bond
short-hydrogen-bond protonated state						
avGFP S65T H148D (PDB: 2DUF)	1.50	A	**2.32 (± 0.08)**	**124.4°**	**-7.2°**	**16.8°**
ih:GFP S65T H148D (PDB: 4ZF3)	1.90	A	2.41	132.6°	-12.8°	16.8°
		B	2.74	117.6°	-8.5°	15.6°
		average	**2.58 (± 0.23)**	**125.1°**	**-10.7°**	**16.2°**
ih:GFP S65T global 3-F$_1$Y H148D (PDB: 6OG8)	1.60	A (*syn*, 54%)	2.61	117.9°	-9.1°	2.2°
		A (*anti*, 46%)	2.35	128.3°	-21.6°	13.0°
		B (*syn*, 59%)	2.49	119.7°	-16.1°	3.9°
		B (*anti*, 41%)	2.21	131.2°	-16.4°	15.3°
		average (*syn*)	**2.55 (± 0.13)**	**118.8°**	**-12.6°**	**3.1°**
		average (*anti*)	**2.28 (± 0.13)**	**129.8°**	**-19.0°**	**14.2°**
ih:GFP S65T Y66(3-Cl$_1$Y) H148D (PDB: 4ZF4)	1.82	A	2.41	119.5°	-24.8°	10.0°
		B	2.52	120.8°	-18.6°	7.9°
		average	**2.42 (± 0.17)**	**120.2°**	**-21.7°**	**9.0°**
ih:GFP S65T Y66(3-Br$_1$Y) H148D (PDB: 6OGA)	1.60	A	2.37	118.3°	-27.3°	14.3°
		B	2.48	123.9°	-24.8°	10.7°
		average	**2.43 (± 0.13)**	**121.1°**	**-26.1°**	**12.5°**
ih:GFP S65T Y66(3-I$_1$Y) H148D (PDB: 6OGB)	1.65	A	2.47	120.4°	-27.1°	14.4°
		B	2.57	123.6°	-23.0°	10.3°
		average	**2.52**	**122.0°**	**-25.1°**	**12.4°**

			(± 0.15)			
ih:GFP S65T Y66(2,3-F$_2$Y) H148D (PDB: 6OGC)	1.18	A (*syn*, 13%)	2.46	120.6°	-6.1°	-3.9°
		A (*anti*, 87%)	2.49	124.5°	-17.4°	7.6°
		B	2.47	121.4°	-16.3°	6.8°
		average	**2.48 (± 0.04)**	**123.0°**	**-16.9°**	**7.2°**
ih:GFP S65T Y66(3,5-F$_2$Y) H148D (PDB: 6OG9)	1.80	A	2.51	129.6°	-27.9°	10.6°
		B	2.54	127.6°	-29.4°	9.2°
		average	**2.53 (± 0.52)**	**128.6°**	**-28.7°**	**9.9°**
ih:GFP S65T Y66(3,5-Cl$_2$Y) H148D (PDB: 4ZF5)	1.70	A	N/A (not in the protonated state [1])			
		B	**2.51 (± 0.15)**	**129.7°**	**-19.3°**	**2.2°**
normal protonated state						
avGFP S65T (PDB: 1Q4B)	1.48	A	**3.18 (± 0.11)**	**152.5°**	**-0.9°**	**5.2°**
deprotonated state						
avGFP S65T (PDB: 1Q4A)	1.45	A	**2.81 (± 0.08)**	**140.5°**	**-2.6°**	**1.7°**
ih:GFP S65T (PDB: 6OFK)	1.15	A	2.82	137.1°	-8.2°	2.8°
		B	2.81	135.2°	-9.3°	2.0°
		average	**2.82 (± 0.04)**	**136.2°**	**-8.8°**	**2.4°**
ih:GFP S65T Y66(3-Cl$_1$Y) (PDB: 6OFL)	1.25	A (*syn*, 82%)	2.83	137.0°	-10.7°	-0.2°
		A (*anti*, 18%)	2.91	134.2°	-29.0°	22.9°
		B (*syn*, 79%)	2.84	135.9°	-11.4°	-0.5°
		B (*anti*, 21%)	2.97	134.9°	-34.5°	23.7°
		average (*syn*)	**2.84 (± 0.04)**	**136.5°**	**-11.1°**	**-0.4°**
ih:GFP S65T Y66(2,3,5-F$_3$Y) (PDB: 6UN5)	1.36	A	2.80	134.6°	-19.1°	8.6°
		B	2.81	137.3°	-28.7°	11.4°
		average	**2.81 (± 0.08)**	**136.0°**	**-23.9°**	**10.0°**
ih:GFP S65T Y66(3-NO$_2$Y) (PDB: 6UN6)	1.50	A	2.89	138.8°	-16.6°	7.3°
		B	2.88	135.7°	-15.2°	6.1°
		average	**2.89 (± 0.12)**	**137.3°**	**-15.9°**	**6.7°**
ih:GFP S65T	1.48	A	2.90	137.4°	-13.4°	3.5°

		B	2.94	141.1°	-14.1°	3.7°
Y66(3-CH₃Y) (PDB: 6OFM)		average	**2.92 (± 0.10)**	**139.3°**	**-13.8°**	**3.6°**
ih:GFP S65T Y66(3-OCH₃Y) (PDB: 6UN7)	1.50	A	2.90	138.7°	-29.9°	8.4°
		B	2.77	137.6°	-23.8°	6.1°
		average	**2.84 (± 0.11)**	**138.2°**	**-26.9°**	**7.3°**

ª Multiple conformations may decrease the reliability of the geometries from modeling, especially for minor conformers. *Syn* conformer denotes the substituent is on the same side with the double-bonded imidazolinone nitrogen (opposite to the carbonyl) and *anti* is the converse.

ᵇ The uncertainties in atom distances are estimated via the formula provided in [55]:

$$\text{coordinate error of atom} = \text{DPI} \sqrt{\frac{B_{atom}}{B_{average}}}$$

in which DPI, the diffraction precision index, is calculated using an online server [56] and largely depends on the resolution and refinement statistics [57]. The ratio of the chromophore's B factor and the average B factor amounts to roughly 60 – 70% for typical GFP structures (Table S7 in [3] and Section S3, Table S6), so the value 65% is used throughout the table. Distance error can be estimated from $\sqrt{2}$ times the coordinate error of atoms, assuming the coordinate errors are the same for both the chromophore and its hydrogen bonding partner. With all these considerations, the uncertainties in atom distances are estimated by 1.14 times the chromophore's DPI. For structures with moderate resolutions (> 1.5 Å), it is thus inappropriate to show the distances up to two decimal places in Å. In theory, the same exercise can be performed on the angle measurements, but it is unnecessary in our cases given the insensitivity of electronic coupling to small dihedral angles (absolute values < 30°) [3].

It has been proposed by Sigala *et al.* that the heavy-atom distance of a short hydrogen bond is largely dependent on the mismatch in the donor-acceptor proton affinities (ΔpK_α) for small molecules [58]. In that study, O–O distances from 2.4 to 2.8 Å were observed upon increasing the ΔpK_α from 0 to 20 for small-molecule complexes within environments ranging from crystals, non-protic solvents (chloroform and acetone), to water, and a line with a slope of 0.02 Å/pK_α unit gave a good fit. Here we can construct an analogous correlation plot (Figures S7A and S7B) using the ΔpK_α inferred from the denatured chromophore titration (Figure S3 in [1] and Figure S24) and modeling. In Figures S7A and S7B, it is inconclusive whether we can reproduce the trend noted by Sigala *et al.* because of the large uncertainties posed by the insufficient resolution or data quality. If an obvious trend were to be obvious based on the ratio of 0.02 Å/pK_α unit, a difference of 0.04 Å would have to be resolved between the most pK_α matched and

mismatched cases, which is clearly beyond the ability of macromolecular X-ray crystallography even if all variants reached ~ 1 Å resolution. In other words, the O–O distance across these halogenated variants are indistinguishable within experimental error, and it is therefore not possible to conclude whether the protein scaffold plays an active role of constraining short hydrogen bond distances. Also, by virtue of plotting against ΔpK_α, we are disregarding other aspects of the protein environment which is non-existent in the small-molecule studies, such as electrostatics from the environment with limited reorganization ability or sterics imposed by the substituents. While similar works have been realized in KSIs and HhPYP studied with ^{1}H-NMR [59][60], these systems contain more than one neighboring short hydrogen bonds, which could affect each other and lead to additional structural perturbations [61]. Other proteins seem to show a great dispersion in O–O distances despite similar ΔpK_α (see Figure 5A in [62]). Further investigation with a larger pK_α span for the GFP chromophore and a consistently better protocol for high-resolution structure determination could help extend conclusions from Sigala *et al.* to proteins or identify other environmental factors that could perturb the hydrogen bond lengths. Unfortunately the downfield shifted ^{1}H-NMR signal associated with the short hydrogen bond has never been observed in these GFP variants (see Section S7 in [1]) potentially due to fast solvent exchange. This influence of solvent exchange on the ^{1}H-NMR signal has been more extensively tested in a bacterial serine protease also bearing a short hydrogen bond [63]. The angle between the short hydrogen bond and the phenol C–O bond is also plotted against ΔpK_α (Figure S7C) and $|pK_\alpha|$ (Figure S7D) based on the limited data points we have obtained. With uncertainties in the atomic positions under consideration, these angles are consistent within experimental errors.

Dihedral angles φ and τ are also subject to variations upon chromophore substitution for both A and B states. Across all the chromophores compared in Table S8, the dihedral angles for the B-state chromophore show ranges of -27° – -3° and 0° – 10° for φ and τ, respectively, while for the A-state chromophore the ranges become -29° – -7° and 3° – 17° for φ and τ, respectively. Therefore, the variations in dihedral angles exhibited by the anionic chromophore are not significantly less than those of the neutral chromophore (*vide supra*). However, as argued in Section S12 of [3], these variations are

not likely to impact the assumption of constant electronic coupling across all these chromophores. Dihedral angles φ and τ are plotted against ΔpK_α in Figures S7E and S7F, respectively, but no clear decisive factor for the dihedral angle change can be identified, further limited by the fact that we only have two sets of crystal structures with the same substitution (or the lack thereof) (Figure S4). It would be useful if more X-ray structures of the ih:GFP S65T H148 harboring dihalogenated or other monohalogenated chromophores can be determined to discern whether the substitution pattern is the sole factor for the variations in dihedral angles or whether the short hydrogen bond also participates in the perturbation.
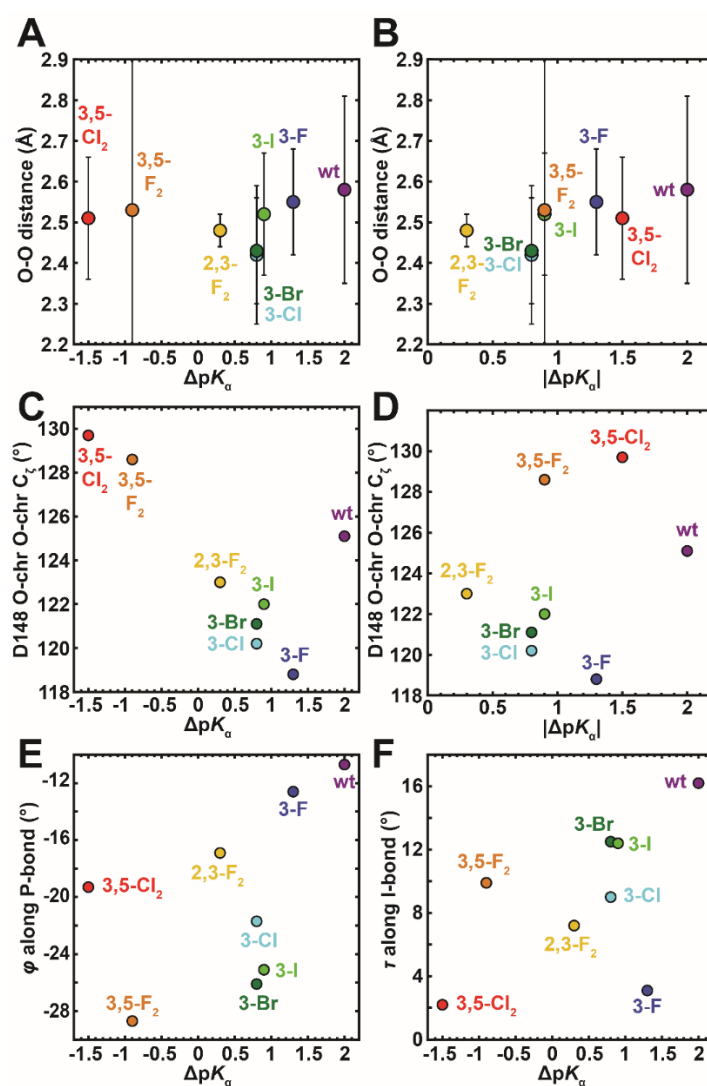
Figure S7. Correlation plots between (A & B) distances, (C & D) angles, and (E & F) dihedral angles listed in Table S8 and ΔpK_α or $|\Delta pK_\alpha|$ for the ih:GFP S65T/H148D chromophore variants. The difference proton affinity ΔpK_α is defined as the proton affinity

of the chromophore minus that of D148, assumed to be constant (Figure 2C) [1]. The average values (in bold in Table S8) are used for these plots. Only the geometry from the *syn* conformer of the 3-F variant is plotted here to match the other monohalogenated variants. In panels A and B, parts of the error bars for the 3,5-F_2 variant are not displayed due to the large uncertainties.

Finally, we have noticed that when cryoprotecting the ih:GFP S65T H148D Y66(2,3-F_2Y) crystals with glycerol, significant cracks developed during the soaking stage and rendered the cryoprotectant used for the 3-Br_1Y and 3-I_1Y variants inappropriate (Section S3). Further inspection of the crystal structures from the last two reveals electron density due to a few glycerol molecules (Figure S8), suggesting the active role of glycerol diffusing into the water channel of the crystals for cryoprotection. This may be why such a sudden influx of glycerol might destroy the crystals in a few seconds while soaking. Usually this phenomenon is attributed to osmotic shock [64], resulting from the osmotic pressure difference between the mother liquor (contains 7 – 8% PEG) and the cryoprotectant (contains 9% PEG and 30% glycerol). However, it is unclear to us why this cryoprotectant is unsuitable for the 2,3-F_2 variant given that its crystals were harvested from the same exact conditions as the 3-Br and 3-I variants. Nevertheless, such a "shock" turned out to be a blessing in disguise and forced us to find perfluoropolyether as a superior cryoprotectant which affords high-resolution structures in our previous and subsequent studies.

Y66(3-Br_1Y)
(PDB: 6OGA)
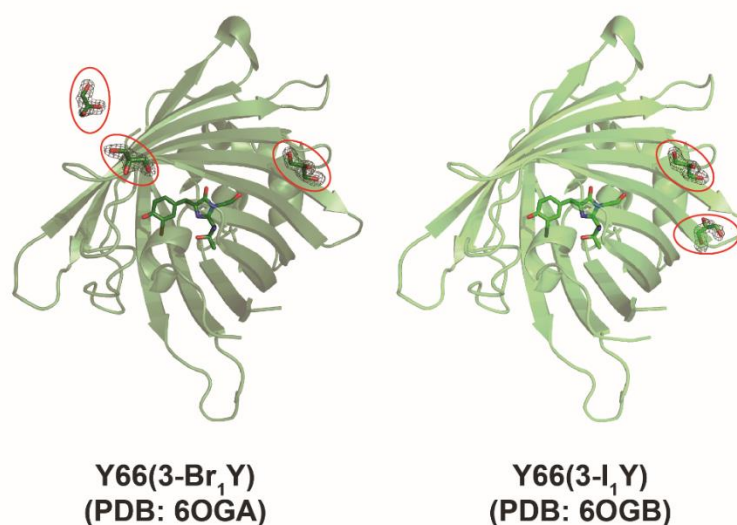
Y66(3-I_1Y)
(PDB: 6OGB)

Figure S8. Glycerol molecules and their electron densities ($2mF_o$–DF_c maps contoured at 1σ, highlighted with red ovals) observed in crystal structures of ih:GFP S65T/H148D

Y66(3-Br$_1$Y) and Y66(3-I$_1$Y), suggesting glycerol molecules actually diffuse into the water channels within the protein crystals in an ordered fashion as a cryoprotectant during the brief soak prior to freezing.

S5   Stark Spectra and Fitting for GFP Variants and PYP Mutants

Electric field application can perturb energetics associated with any charge transfer processes, such as electron redistribution within the chromophore upon excitation and proton transfer across the short hydrogen bond in our cases, rendering Stark spectroscopy a potentially powerful tool for determining the barrier height of the short hydrogen bond involved. Within the realm of Stark spectroscopy, we classify the effects arising from electron redistribution as classical Stark effects, while those for proton transfer are deemed non-classical [65][66]. This is not only because electronic Stark spectroscopy was originally conceived for extracting Stark tuning rates (the change in dipole moment for a transition) [67], but also the effects caused by these two mechanisms can be distinguished based on their distinct characteristics. In the following we will discuss the relevant phenomena using short-hydrogen-bond GFPs and PYP mutants as examples, yet the argument is generally applicable to any similar systems.

For the classical 2ω Stark spectrum, $\Delta A(2\omega)$, of an isotropic and immobilized sample, where ω is the external field modulation frequency, we expect a dominant second derivative lineshape for each electronic transition (i.e., A-like and B-like A states) when the Stark tuning rates are significant. If one considers the possibility of field-induced proton transfer that transfers populations between A-like and B-like A states, a net increase/decrease in the absorbance for each species should be observed upon field application, leading to a zeroth derivative contribution for each electronic transition with opposite signs.

Another powerful aspect of this technique is the use of a sinusoidal field profile and lock-in detection, allowing for isolation of the various external field magnitude (F for short) dependences of the signal. For isotropic and immobilized samples, the Stark signals only depend on even powers in field magnitude (i.e., F^2, F^4, …) as centrosymmetry would be violated otherwise. For a sinusoidal field with frequency ω, the corresponding Stark signals with F^2 dependence modulate at 2ω, F^4 dependent components modulate at both 2ω and 4ω, and so on. Therefore, 2ω Stark spectrum contain F^2 and also higher order terms and 4ω Stark spectrum similarly contain F^4 and higher order terms. For classical Stark spectra and experimentally achievable applied fields ($\sim$ 1 MV/cm), since the F^4

dependence is linked to a *fourth* derivative of absorption lineshape, which is much smaller in magnitude than a second derivative, higher order contaminations are negligible compared to the lower order features. As such, classical $\Delta A(2\omega)$ and $\Delta A(4\omega)$ spectra are proportional to F^2 and F^4, respectively, and 4ω spectral lineshapes are the second derivative of the 2ω spectra with a scale factor that depends on the Stark tuning rate [23]. By contrast, for Stark spectra associated with population transfer due to the applied field, manifesting itself as a *zeroth* derivative lineshape in the F^4 dependent component, deviations from being proportional to F^2 and F^4 (classical Stark effect) can be more easily seen for the $\Delta A(2\omega)$ and $\Delta A(4\omega)$ spectra, respectively. When the ratio, $\frac{\mu_{PT}fF}{k_BT}$ (where μ_{PT} is the electric dipole moment associated with the proton moving across the two wells, f is the local field factor, k_B is the Boltzmann constant, and T is the temperature) exceeds 1, the deviations from the classical Stark effect become large. Since all the following Stark spectra were acquired at 77 K, this condition corresponds to $\mu_{PT}fF$ being greater than 0.67 e · Å · MV/cm, which is in fact not hard to achieve with fields slightly stronger than 1 MV/cm, since the traveling distance of a proton is ∼ 0.5 Å within the short hydrogen bond (Figure 1B) and f is larger than 1. A more thorough analysis has shown that depending on the intrinsic bias (i.e., ΔpK_a) between the two wells, $\Delta A(2\omega)$ can either have subquadratic or superquadratic field dependence [66], thus providing a further diagnostic that such non-classical effects are present.

Based on this, we should be able to detect the presence of proton transfer upon field application via performing the classical sum-of-derivative (wavenumber dependence, Equation S1) and field dependence analysis on $\Delta A(2\omega)$ and $\Delta A(4\omega)$ spectra. We summarize the characteristics of Stark effects caused by the classical mechanism and proton transfer in Table S9. In the subsequent subsections we will examine the Stark data obtained from short-hydrogen-bond GFP variants and PYP mutants.
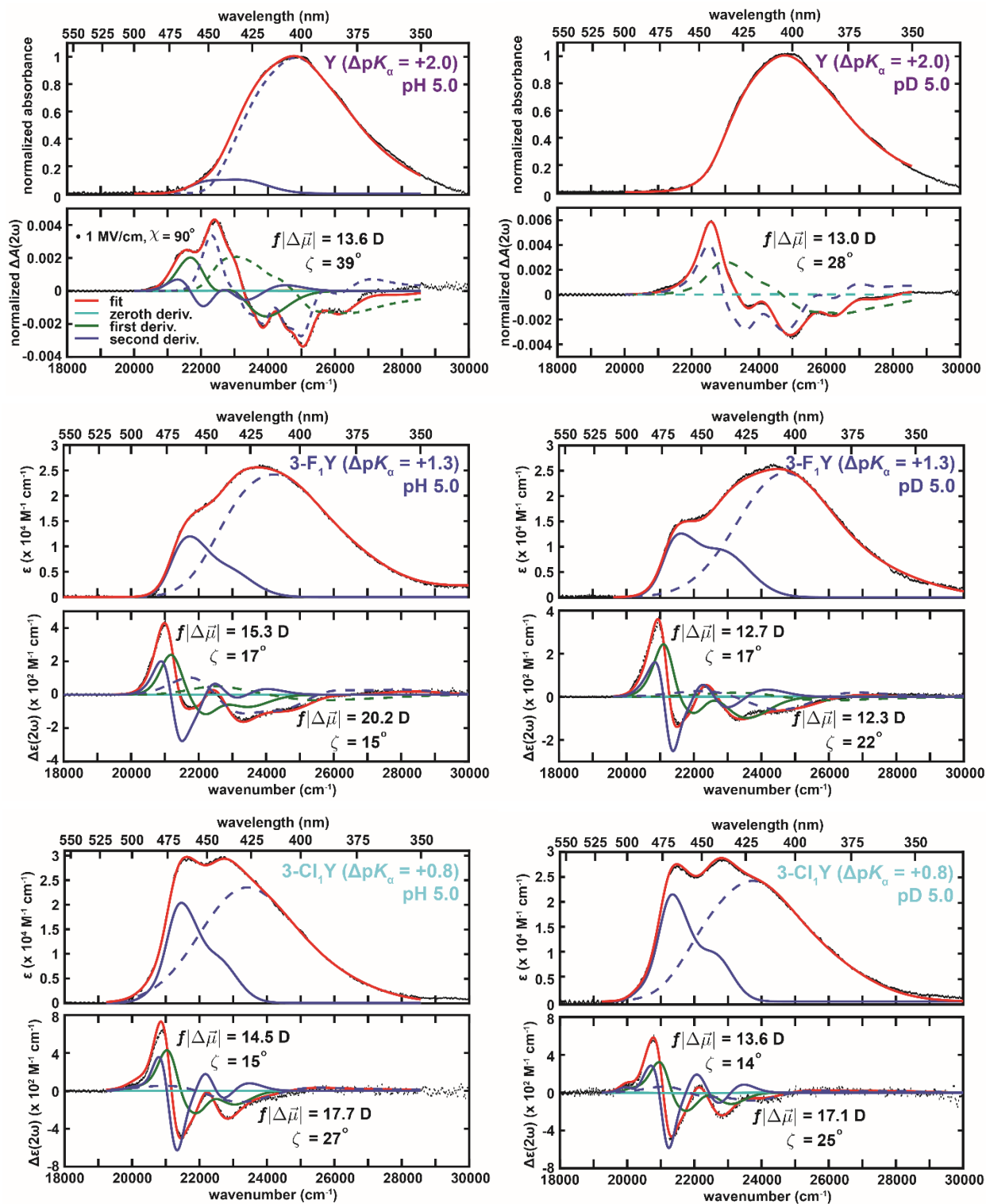
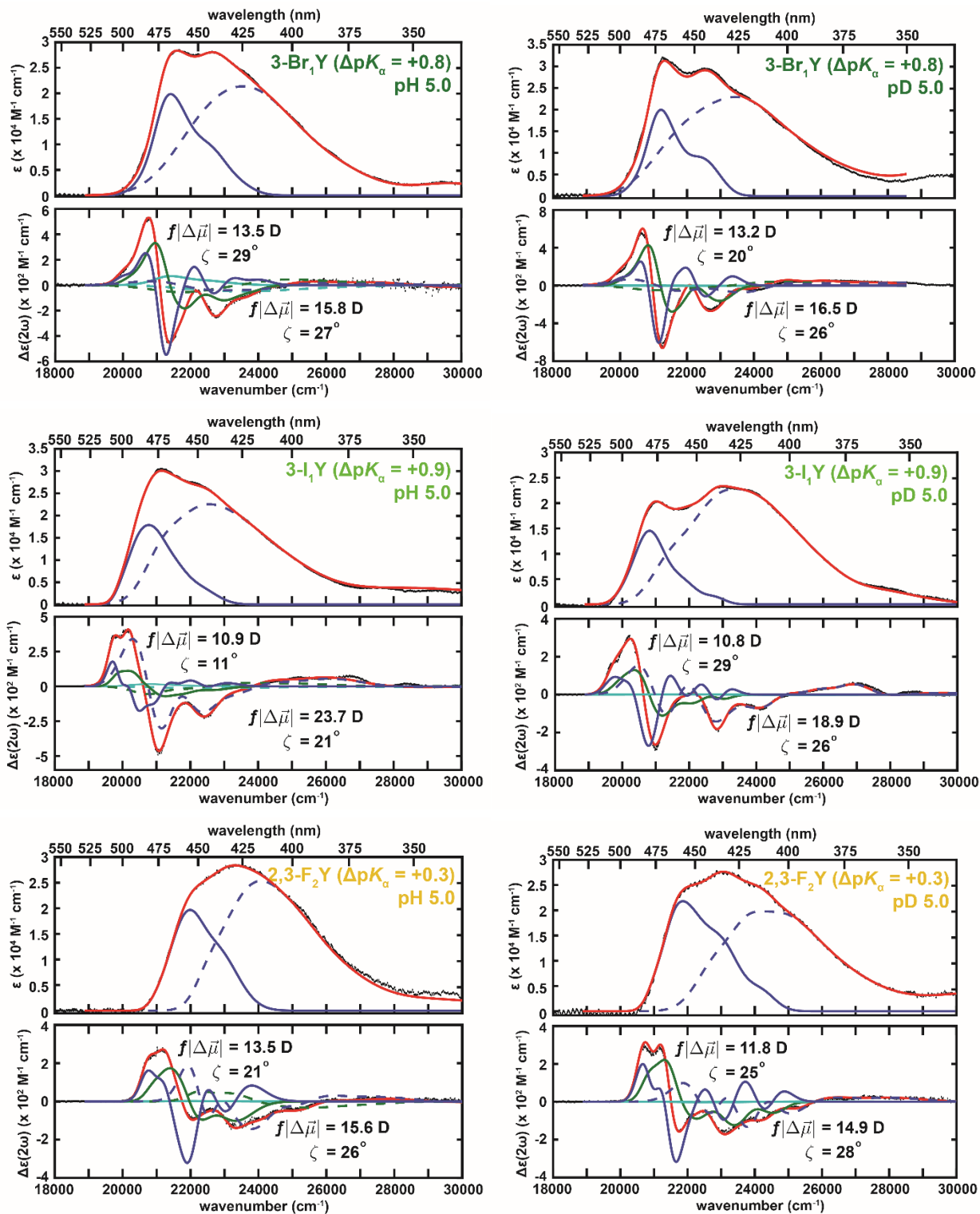**Table S9.** Lineshapes and field dependence from classical and non-classical Stark effects.

	$\Delta A(2\omega)$ spectrum	$\Delta A(4\omega)$ spectrum
classical mechanism (electron redistribution)	second derivative F^2 dependence	fourth derivative F^4 dependence

non-classical mechanism	zeroth derivative
(proton transfer)	contaminated by higher order terms at high F

Short-Hydrogen-Bond GFP Variants

We start with sum-of-derivative analysis on $\Delta A(2\omega)$ spectra, as shown in Figure S9. As noted in Figure 3C in the main text, all cases except for Y66 at pD 5 require two sets of electro-optic parameters with maximal spectral separation to simultaneously give a good fit to both the absorption and Stark spectra. In addition, the $\Delta A(2\omega)$ spectra for both features are dominated by second-derivative lineshapes and no appreciable zeroth derivative component is needed for the fit, suggesting the absence of substantial proton transfer between two wells even when high field strength (~ $f \cdot 1.4$ MV/cm) is applied. By following the trend from negative to positive ΔpK_a cases (i.e., more halogenated to less halogenated chromophores), as the A-like A state grows at the expense of the B-like A state, one can see that the A-like A state has less prominent vibronic feature than the normal A state [68], such that its corresponding Stark signal is greatly reduced and easily overwhelmed by that of the B-like A state, best illustrated by the Stark spectra from monochlorinated and monobrominated variants (the second derivative of a broad *vs* sharp feature). In contrast, B-like A states still preserve the BLA vibronic structure akin to that of the B state chromophore [3], and this accounts for the dip in the overall absorption spectra at low temperature (Figure S9 except for the Y66 variant). As seen before in both GFP and Dronpa2 with monofluorinated chromophores (Section S11 in [3]), even though dual chromophore conformers are observed in the X-ray structure for the 3-F_1Y variants (Figure S1), no additional set of electro-optic parameters is required.

Y (ΔpK$_\alpha$ = +2.0)
pH 5.0

normalized absorbance

wavelength (nm)
550 525 500 475 450 425 400 375 350

• 1 MV/cm, χ = 90°

$f|\Delta\vec{\mu}|$ = 13.6 D
ζ = 39°

fit
zeroth deriv.
first deriv.
second deriv.

normalized $\Delta A(2\omega)$

wavenumber (cm^{-1})
18000 20000 22000 24000 26000 28000 30000

Y (ΔpK$_\alpha$ = +2.0)
pD 5.0

$f|\Delta\vec{\mu}|$ = 13.0 D
ζ = 28°

3-F$_1$Y (ΔpK$_\alpha$ = +1.3)
pH 5.0

ϵ (× 10^4 M^{-1} cm^{-1})

$f|\Delta\vec{\mu}|$ = 15.3 D
ζ = 17°

$f|\Delta\vec{\mu}|$ = 20.2 D
ζ = 15°

$\Delta\epsilon(2\omega)$ (× 10^2 M^{-1} cm^{-1})

3-F$_1$Y (ΔpK$_\alpha$ = +1.3)
pD 5.0

$f|\Delta\vec{\mu}|$ = 12.7 D
ζ = 17°

$f|\Delta\vec{\mu}|$ = 12.3 D
ζ = 22°

3-Cl$_1$Y (ΔpK$_\alpha$ = +0.8)
pH 5.0

$f|\Delta\vec{\mu}|$ = 14.5 D
ζ = 15°

$f|\Delta\vec{\mu}|$ = 17.7 D
ζ = 27°

3-Cl$_1$Y (ΔpK$_\alpha$ = +0.8)
pD 5.0

$f|\Delta\vec{\mu}|$ = 13.6 D
ζ = 14°

$f|\Delta\vec{\mu}|$ = 17.1 D
ζ = 25°

Figure S9. The sum-of-derivative analysis for 77 K UV–vis absorption and Stark spectra of short-hydrogen-bond GFP variants (H148D) in protonated (left panels) and deuterated buffers (right panels) are mostly dominated by second-derivative lineshapes and lack of zero-derivative features. The variants are labeled by the identities of residue 66 followed by their corresponding ΔpK_α in parentheses. The absorption spectra are normalized to 1 at the maximum absorbance for the Y66 variant, while the rest are concentration normalized and shown as extinction coefficients ε due to significant spectral isotope effects (SIEs). The Stark spectra are measured at $\chi = 90°$ and scaled to 1 MV/cm to facilitate comparison. The color scheme of fit lines and data, as shown in the first panel, is consistent throughout the figure. Dashed and solid lines for the underlying bands represents the A-like and B-like A state, respectively. Dash-dotted lines for the Y66(3,5-Cl$_2$Y) variant represents the normal B state, which is shown to contaminate the B-like A state at pH (pD) 5.0 [1].

To get around the problem of overlapping bands which can undermine the uniqueness of the sum-of-derivative analysis due to cancelling positive and negative features [23][26] (see also Section S11 in [3]), we also take advantage of the difference absorption and Stark spectra between protonated and deuterated samples such that the

underlying species can be unveiled in a model-free way (Figure S10). In addition to its diagnostic power of disentangling various charge transfer mechanisms, Stark spectroscopy is also good at enhancing subtle differences in absorption that are hardly noticeable, as exemplified by the red-edge absorption of the protonated and deuterated Y66 variant. The difference spectra reveal vibronic and Stark features of the B-like A state (see also the first panel in Figure S9), as suggested by the PES in Figure 1B. Unfortunately, this deconvolution method only strictly applies to the Y66 variant, as significant SIEs change the absorption peak positions and bandshapes going from protonated to deuterated species. Nevertheless, some A-like A state's Stark feature can still be observed through difference spectra in the Y66(3-F_1Y) case. The method ceases to be useful as the Stark contribution from the B-like A state largely overwhelms the A-like A state counterpart for variants with lower ΔpK_α due to the former possessing more prominent vibronic features.
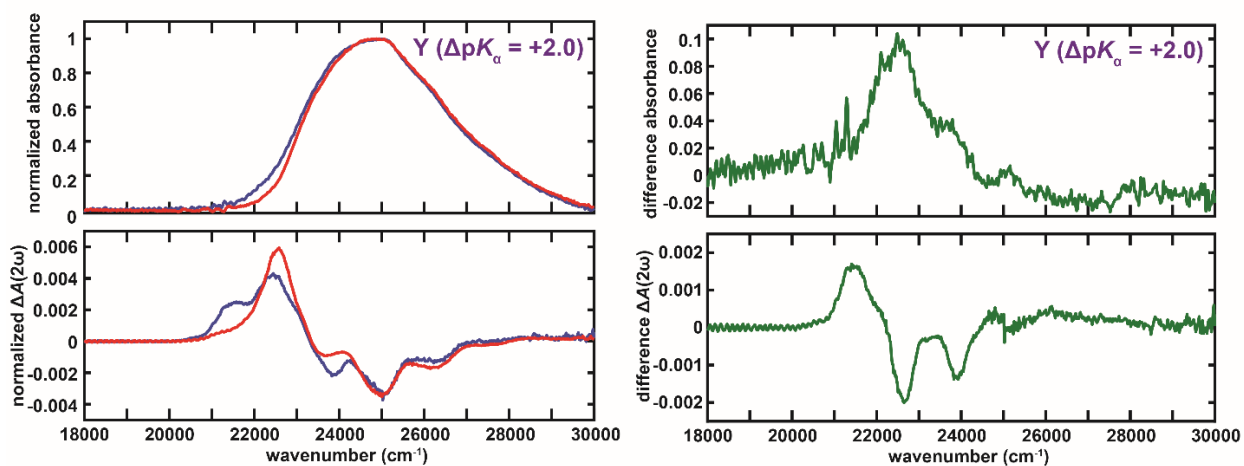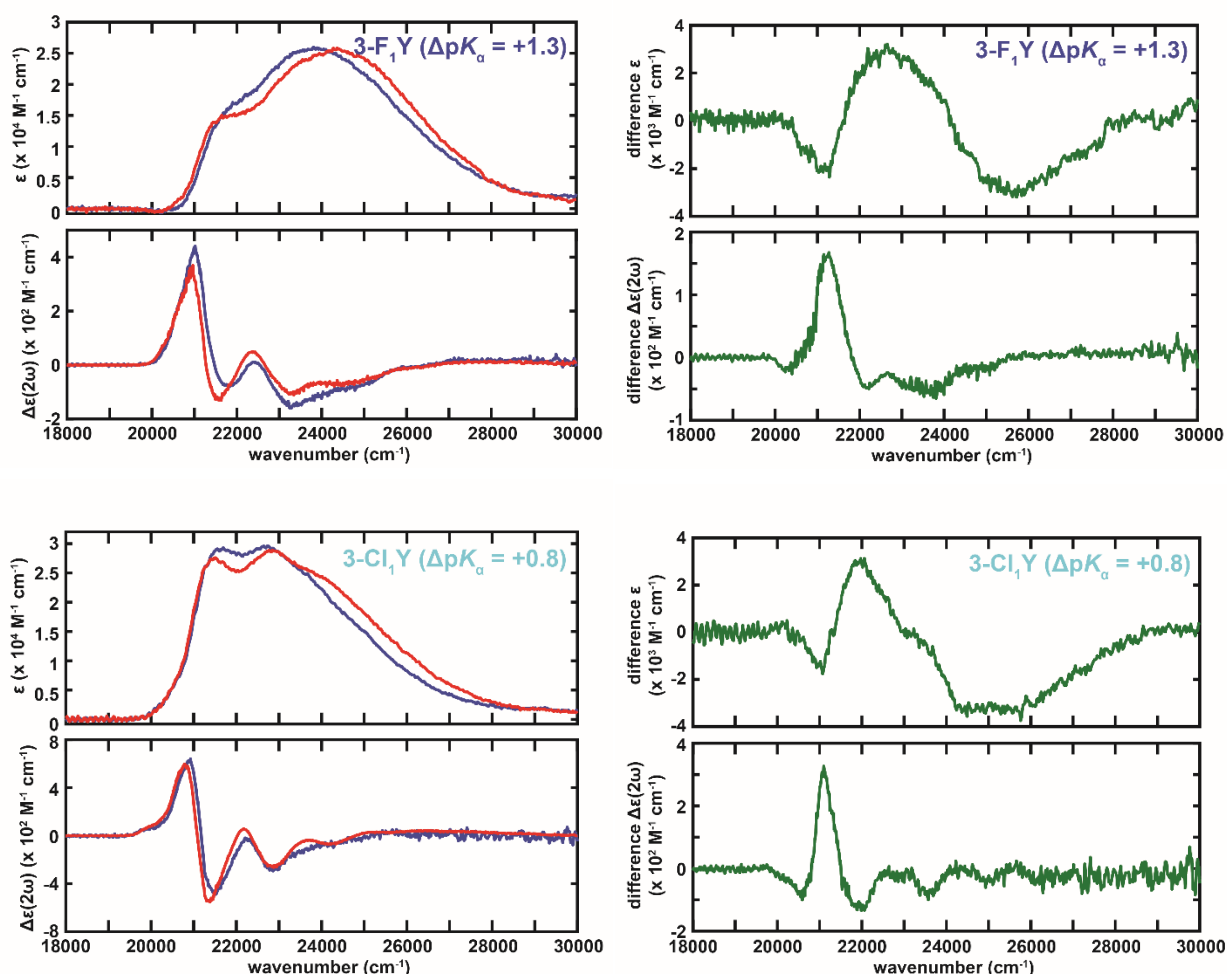
Figure S10. Comparison of 77 K absorption and Stark spectra for protonated and deuterated short-hydrogen-bond GFP variants (left panels), and their corresponding difference spectra (right panels). Blue, red, and green traces are for protonated samples, deuterated samples, and their spectral differences (protonated – deuterated). For the Y66 variant, a clear B-like A state feature can be seen from both difference spectra (cf. Figure 6 in [69]). For the Y66(3-F$_1$Y) variant, while a narrowing lineshape (inverted second derivative) is observed for the difference extinction coefficient due to the SIE, an A-like A state Stark feature can be still seen from the difference Stark spectrum. The same feature is however lost for the Y66(3-Cl$_1$Y) counterpart owing to the growth of the B-like A state band, so the same devolution strategy is not useful for variants with lower ΔpK_α. See Figure S26 for more direct comparisons of 77 K absorption spectra between protonated and deuterated GFP variants.

Another model-free method to assess the dominant charge transfer mechanism is through the 4ω Stark spectra (Figure S11), as they depend on the fourth power of the difference dipole moment and thus are much more sensitive to mechanisms with larger difference dipole moments. Even though Stark contributions from proton transfer could

be at a disadvantage due to its shorter charge transfer distance, the fourth derivative lineshapes associated with electron redistribution upon excitation are strongly suppressed compared to the zeroth derivative lineshape. To be more precise, since the proton transfer distance across the short hydrogen bond and electron transfer distance within the chromophore upon excitation typically amounts to 0.5 and 4 Å, respectively, the proton transfer contribution for the 4ω spectra is suppressed by 8^4, which is nearly 4 orders of magnitude. However, as each differentiation reduces the order of magnitude by 2, the fourth-derivative feature $\bar{v}\frac{d^4}{d\bar{v}^4}\left(\frac{A(\bar{v})}{\bar{v}}\right)$ would be about 8 orders of magnitude smaller than the corresponding zeroth-derivative feature $A(\bar{v})$. In conclusion, the zeroth-derivative contribution that would be expected from proton transfer is still enhanced by 4 orders of magnitude or so in the 4ω spectra, and so should be detectable unless the proton transfer itself is strongly disfavored even upon field application. From Figure S11, all 4ω spectra mirror the second derivative of the corresponding 2ω spectra, with no evidence for a zeroth derivative component as anticipated from proton transfer, thus ruling out the existence of a *low-barrier* hydrogen bond (LBHB) in these variants. It is possible that proton transfer could still be observed if the proton probe's Stark tuning rate is greatly reduced, such as using vibrational probes instead of electronic ones, yet the effect must be small for it not to manifest under the conditions of our experiments based on the preceding argument. Obviously, such *external* field induced proton transfer would be even less likely for a normal hydrogen bond. We note, however, that very large electric fields can arise *locally* within organized systems such as proteins, and these fields could change substantially upon conformational transitions, ligand binding, and the like. Thus, it is possible that proton positions within short or even normal hydrogen bonds could shift if very large and directional internal field changes occur.
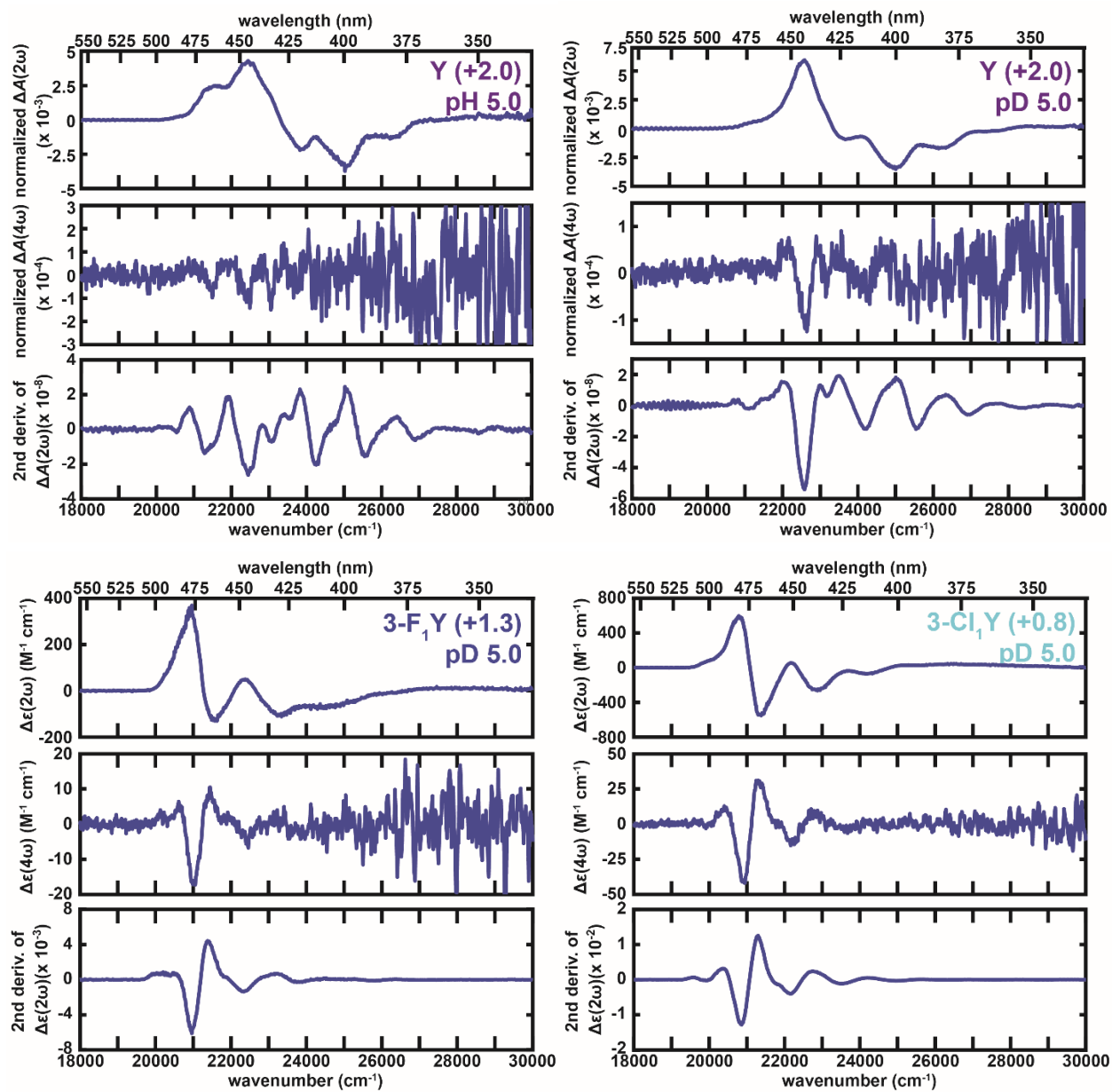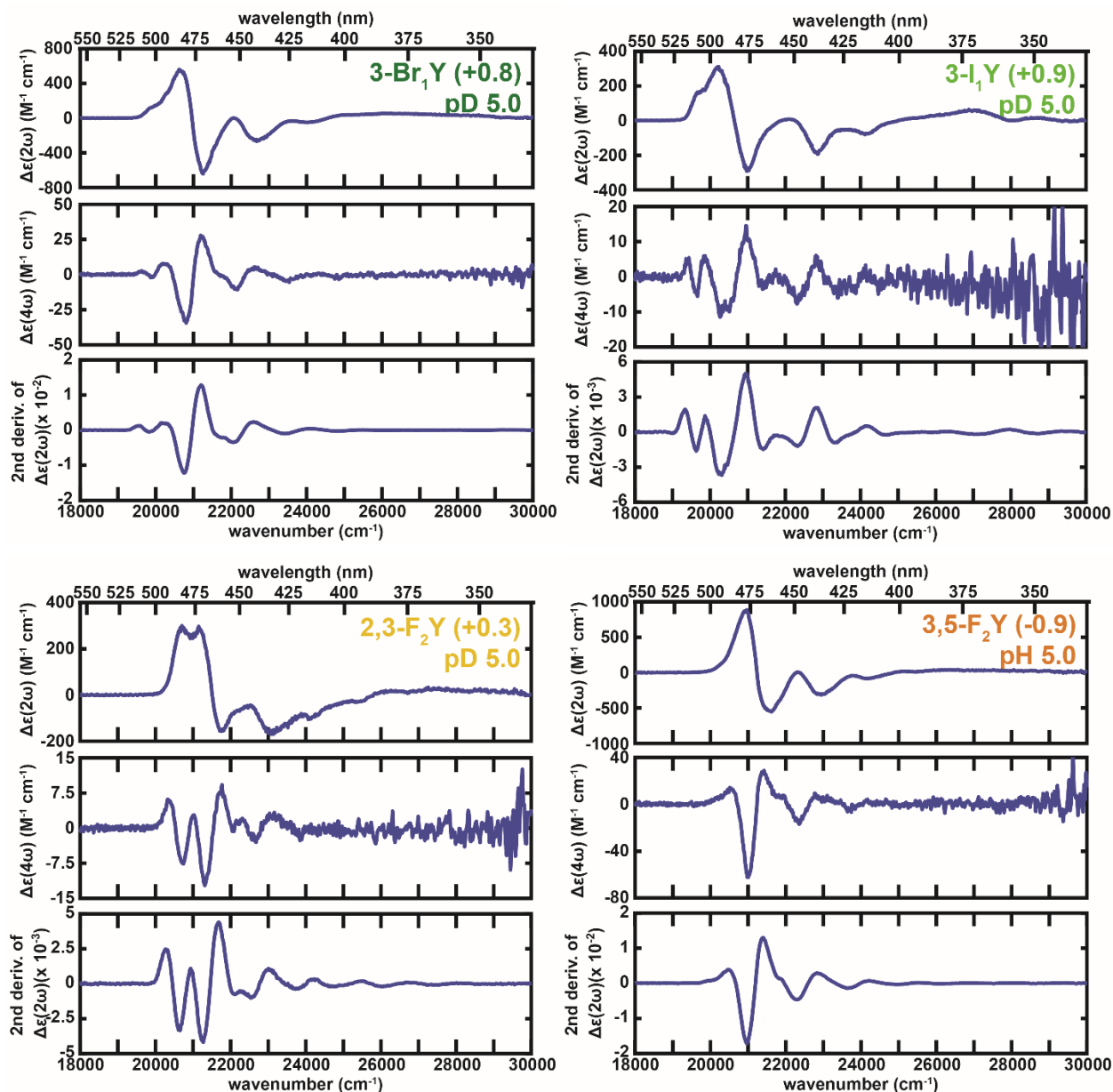
Figure S11. Absorbance or concentration normalized 2ω Stark spectra, 4ω Stark spectra, and wavenumber-weighted second derivative of the 2ω Stark spectra for short-hydrogen-bond GFP variants. Because 4ω spectra are harder to acquire than 2ω spectra, as they require high field strengths and enough photons to increase the signal-to-noise ratio, the signal-to-noise is poorer at the high wavenumber side. As dielectric breakdown frequently occurs with field strengths at which 4ω spectra have sufficient signal-to-noise ratio, we are unable to obtain 4ω spectra for all conditions. Note the strong resemblance between the 4ω Stark spectra and the second derivative of the 2ω Stark spectra, as opposed to the predicted zeroth-derivative lineshape if proton transfer were operative.

Finally, we analyzed the field strength dependence of both the 2ω and 4ω Stark spectra and found no appreciable deviation can be observed up to $f \cdot 1.4$ MV/cm (Figure

S12) across the GFP variant series. This observation is again consistent with the previous analyses and together suggest a dominant classical mechanism rather than field-induced proton transfer.
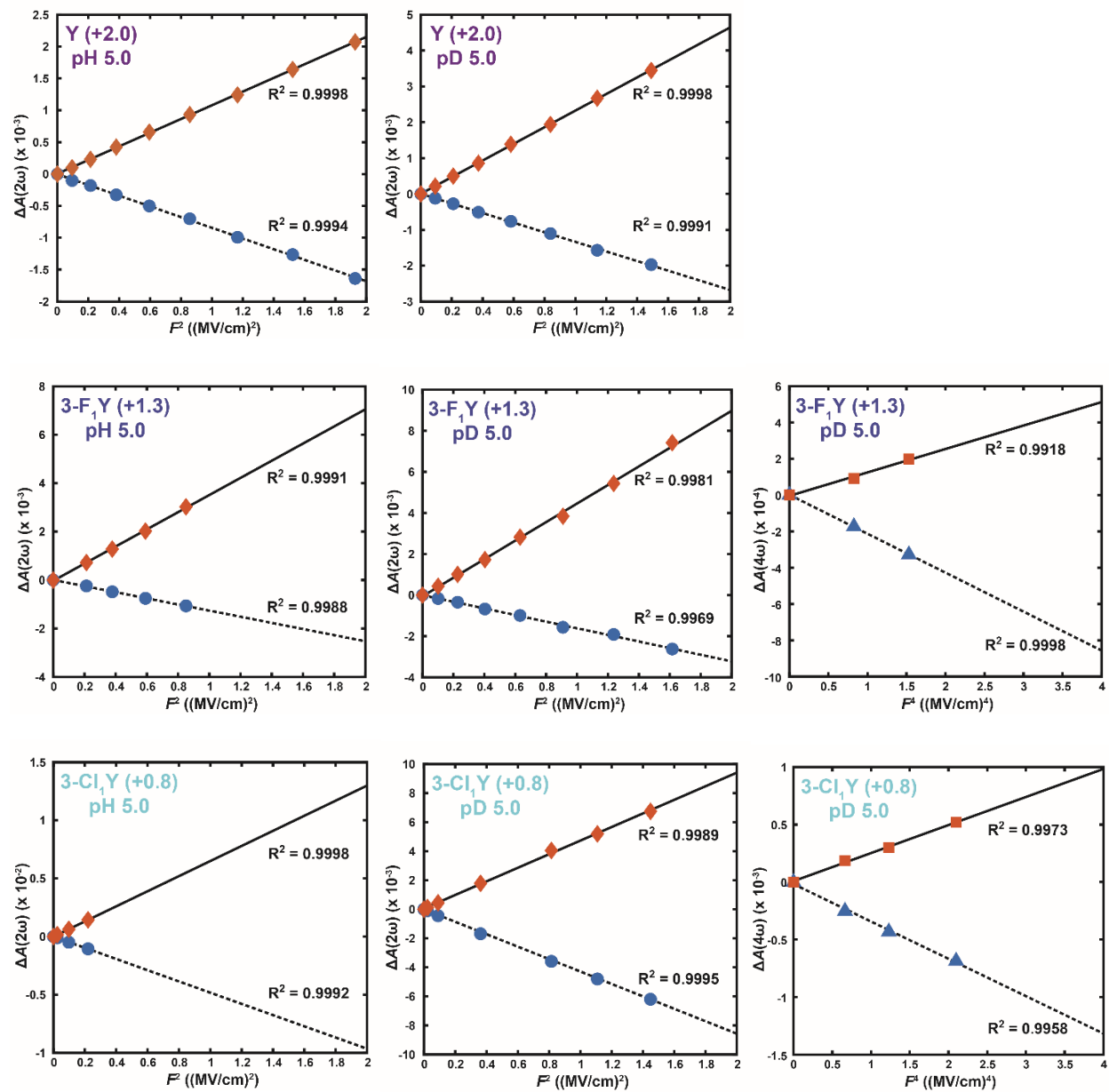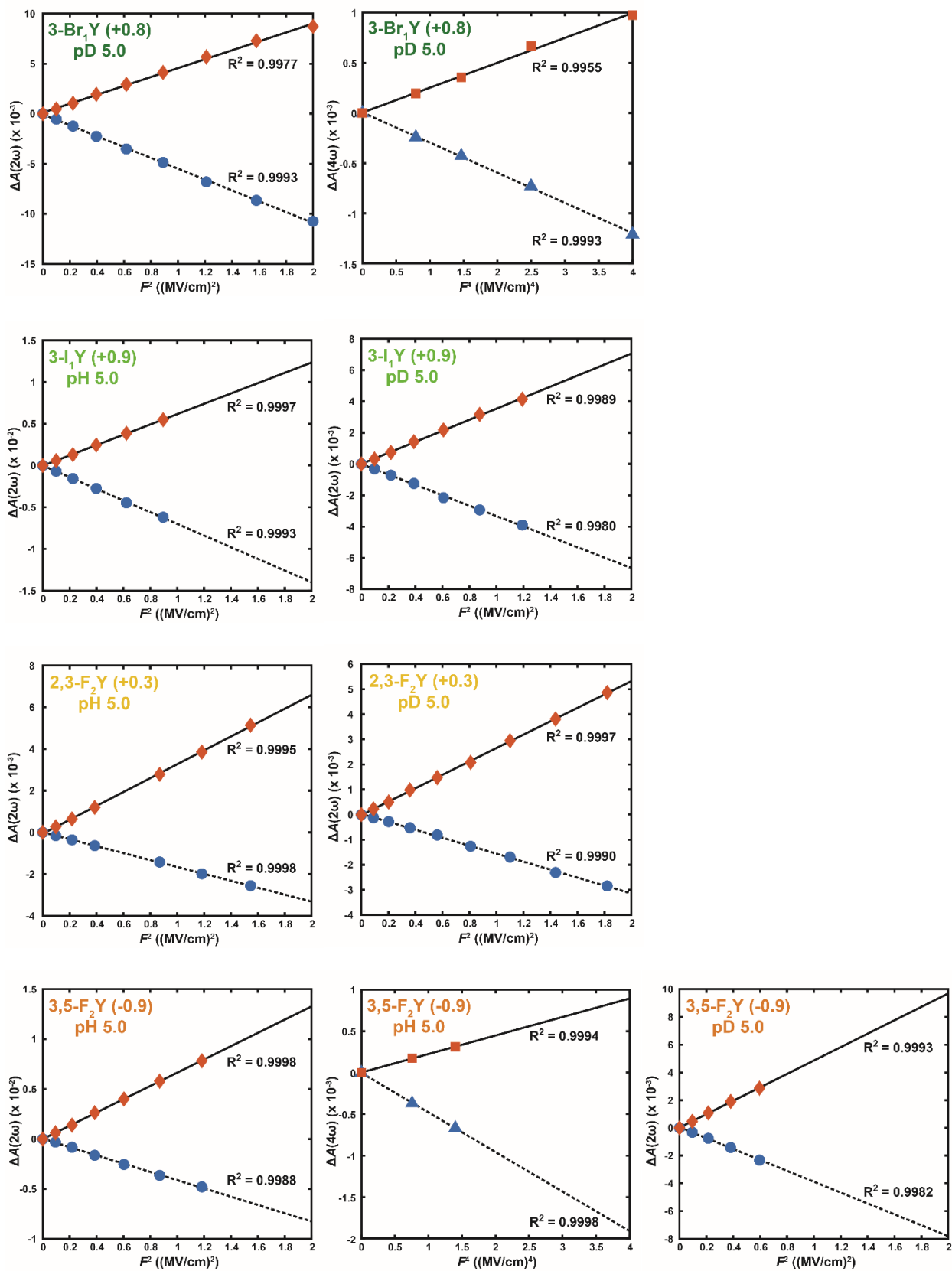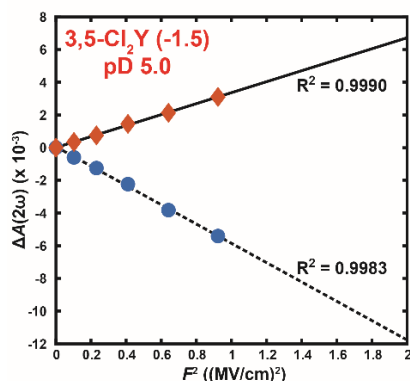
Figure S12. Field dependence analysis for 2ω and 4ω Stark spectra of GFP variants. Values from raw (i.e., not normalized and not scaled to 1 MV/cm) 2ω and 4ω Stark data measured at $\chi = 90°$ are plotted against F^2 and F^4, respectively. Orange and blue data points are global maximum and minimum from each spectrum, and solid and dashed lines are the corresponding linear fits. Data for some conditions are not shown due to either insufficient signal-to-noise ratio at low fields, or only one spectrum or none could be taken prior to dielectric breakdown. Note that the actual field strength is amplified by a local field factor f, whose value is greater than 1 [23].

PYP Mutants and Model Chromophore

The same set of analyses can be conducted with PYP mutants, though the corresponding pK_α's are much more mismatched and proton transfer is even less likely to occur, as substantiated by the following data (Figures S13 – S16). In addition, we note that our measured Stark tuning rates of HhPYP mutants and the model chromophore are consistently smaller than those reported from van Grondelle *et al.* by a factor of 1.4 [20][25], suggesting an overestimation of charge transfer extent in the past literature.
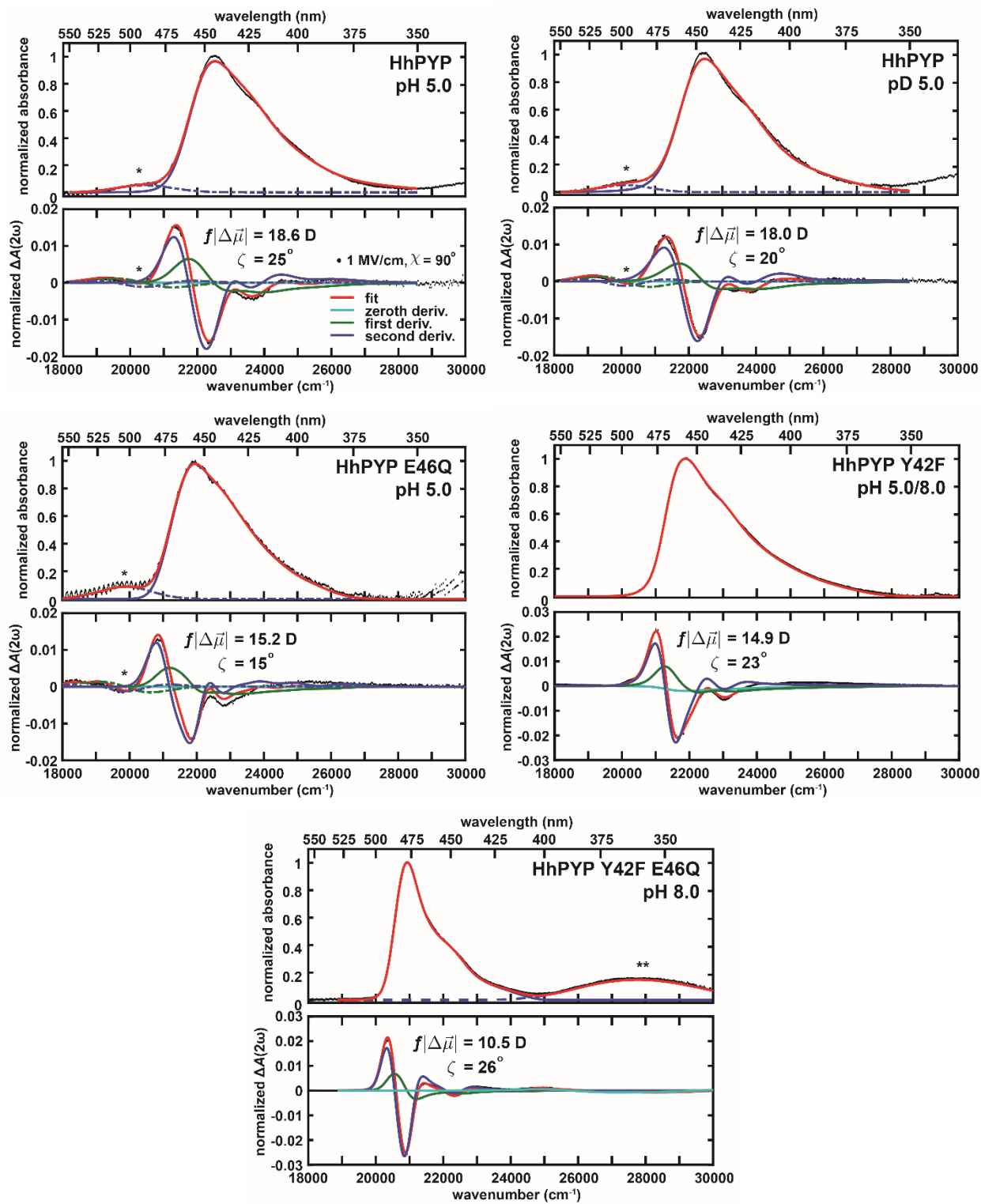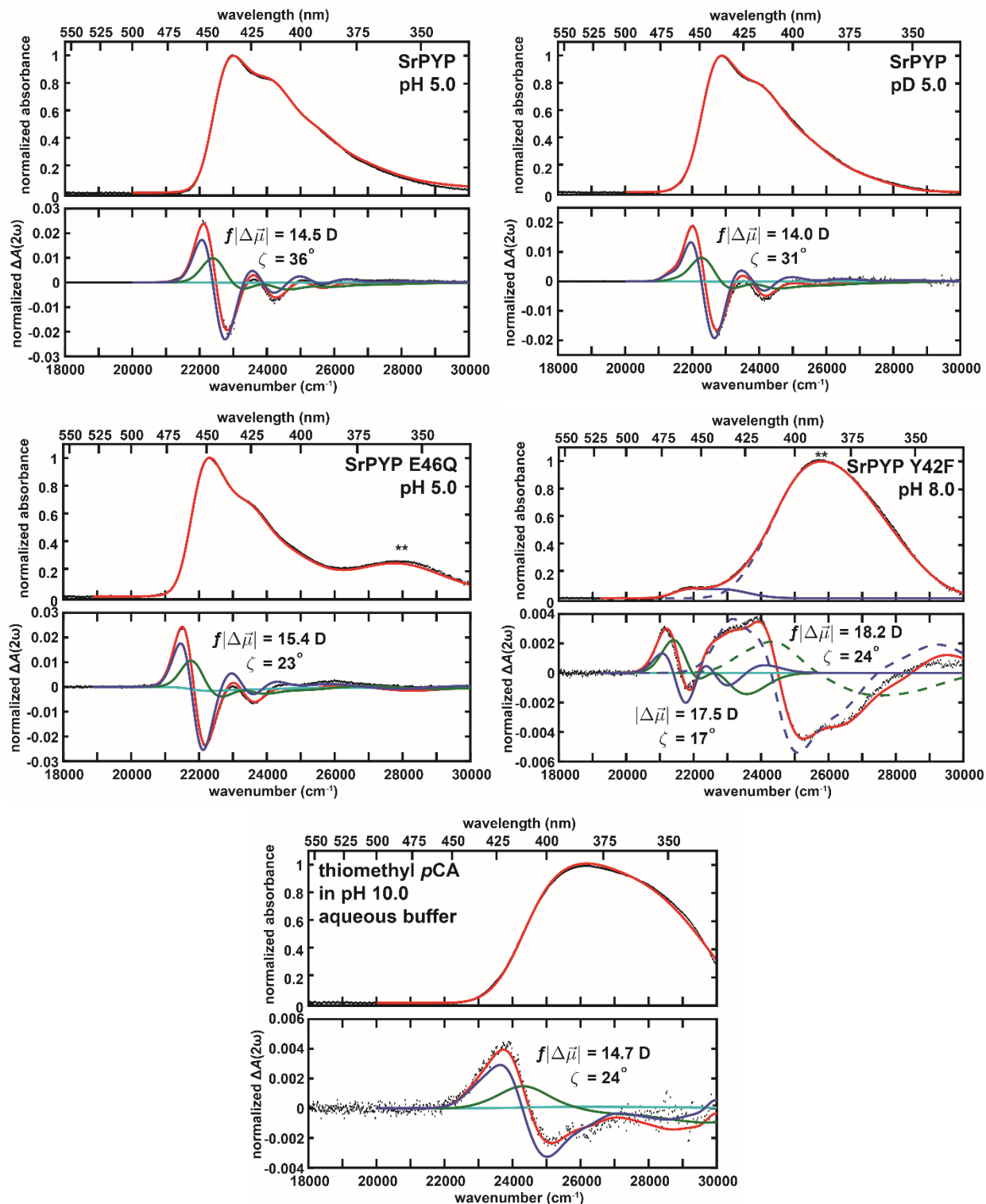
Figure S13. The sum-of-derivative analysis for 77 K UV–vis absorption and Stark spectra of PYP mutants are mostly dominated by second-derivative lineshapes and lack zero-derivative features. The absorption spectra are normalized to 1 at the maximum absorbance. The Stark spectra are measured at χ = 90° and scaled to 1 MV/cm to facilitate comparison. The color scheme of fit lines and data, as shown in the first panel,

is consistent throughout the figure. For wild-type and E46Q HhPYPs, a redder band (indicated by single asterisk) appears as the sample is irradiated during wavelength scans, and the Stark spectra shown here are at the photostationary state. This band has been previously observed before and is assigned to pR (I_0), the first intermediate within the photocycle [25][70], which can be trapped at 77 K due to a much smaller photoisomerization barrier preceding its formation than subsequent ones. This photoisomerization barrier is significantly enhanced by the Y42F mutation, as evidenced by the high fluorescence quantum yields [71][72] and the lack of pR accumulation for the corresponding mutants. As more hydrogen bonds are removed, the chromophore tends to prefer the protonated state (normal A state, labeled by double asterisks) even at pH 8.0. Since SrPYP is effectively negatively supercharged (~ 20% aspartate + glutamate), the preference is even more dominant, leading to unreliable Stark measurement for the B-like A state of the Y42F mutant. The Stark spectrum for SrPYP Y42F E46Q is not shown for the same reason.
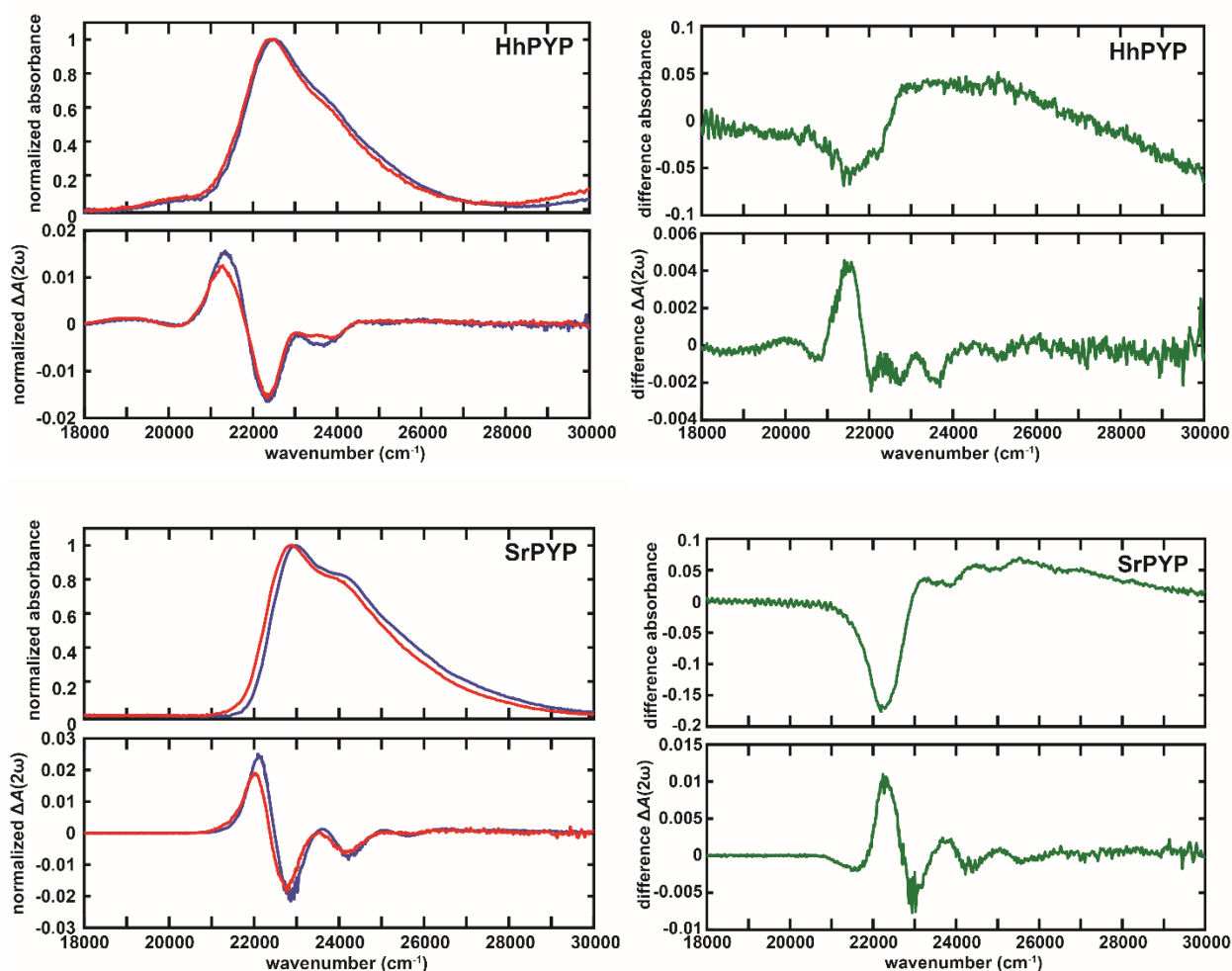
Figure S14. Comparison of 77 K absorption and Stark spectra for protonated and deuterated PYPs (left panels), and their corresponding difference spectra (right panels). Blue, red, and green traces are for protonated samples, deuterated samples, and their spectral differences (protonated – deuterated). Both peaks red shift upon deuteration,

following the expected SIE trend for B-like A states. Even though the differences in peak maxima are comparable for both PYPs (~ 90 cm⁻¹, Table S10), SrPYP indeed demonstrates a larger spectral change upon deuteration at 77 K compared to HhPYP, consistent with their room-temperature behaviors [73].
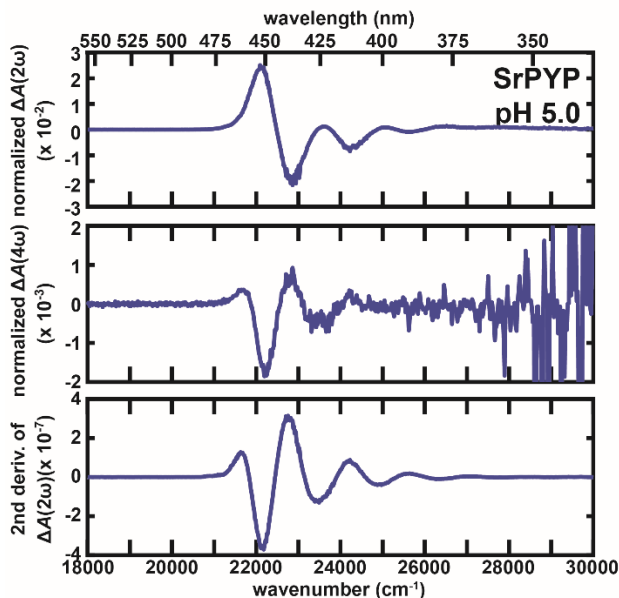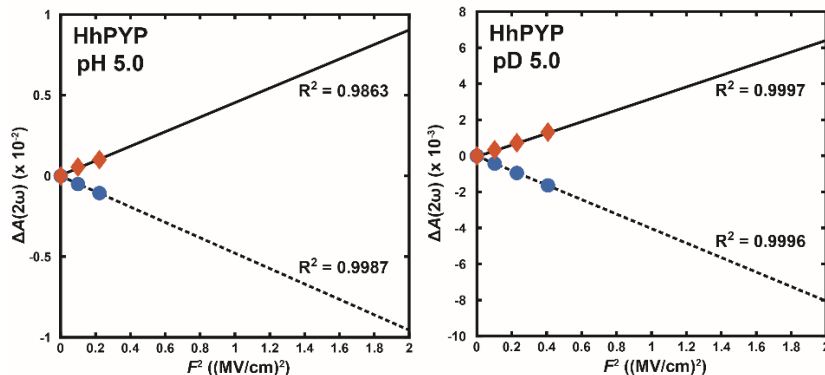

Figure S15. Normalized 2ω Stark spectra, normalized 4ω Stark spectra, and wavenumber-weighted second derivative of the 2ω Stark spectra for wild-type SrPYP at pH 5.0, which has the largest SIE among all PYP constructs. Note the strong resemblance between the 4ω Stark spectra and the second derivative of the 2ω Stark spectra, as opposed to the zeroth-derivative lineshape if proton transfer were operative.
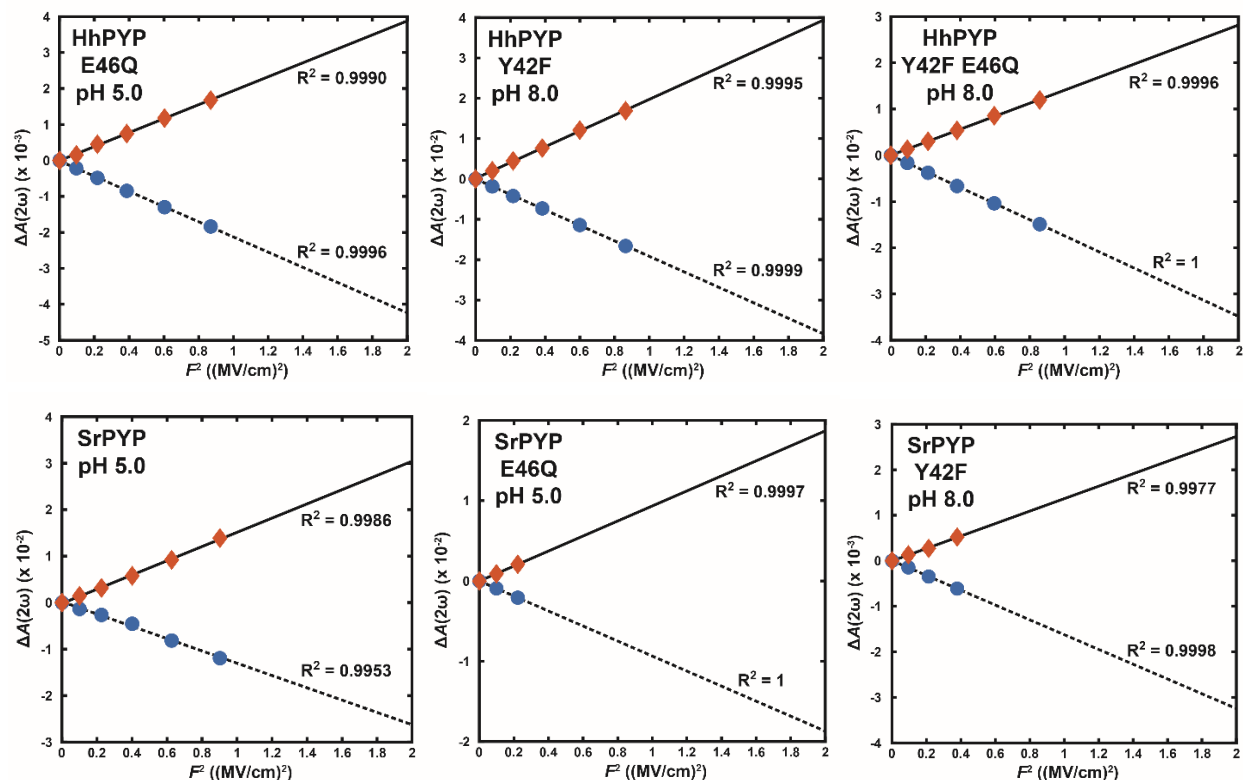
Figure S16. Field dependence analysis for 2ω Stark spectra of PYP mutants. Values from raw (i.e., not normalized and not scaled to 1 MV/cm) 2ω Stark data measured at χ = 90° are plotted against F^2, respectively. Orange and blue data points are global maximum and minimum from each spectrum, and solid and dashed lines are the corresponding linear fits. Data for some conditions are not shown due to either insufficient signal-to-noise ratio at low fields, or only one spectrum or none could be taken prior to dielectric breakdown. Note that the actual field strength should be amplified by a local field factor f, of which the value is greater than 1 [23].

S6   Pre-resonance Raman Spectroscopy on Short-Hydrogen-Bond GFPs

In this section we will elaborate how we determine the qualitative shape of the hydrogen bond PES (i.e., single- or double-welled) through monitoring the splitting of a Raman band upon deuteration combined with ΔpK_a tuning across the short-hydrogen-bond GFP variants. As with the use of NMR for evaluating the extent of chemical exchange [74], we first need to find a vibrational mode whose frequency is sensitive to the proton position and that allows us to assess whether the barrier of proton transfer is high or the proton is delocalized between the donor and acceptor. From detailed assignment of the chromophore's IR and Raman spectra through (de)protonation, isotope labeling, and density functional theory [75][76][77][78], there are several candidates that are present in both the neutral and anionic chromophores and exhibit sufficiently intense Raman scattering signals. Specifically, Raman peaks around 1620, 1540, and 1260 cm^{-1} for the anionic chromophore become 1640, 1560, and 1240 cm^{-1} upon protonation in GFP mutants (Figure S17). The first two peaks are mainly assigned to the bridge C=C and imidazolinone C=N stretching, respectively, and their blue-shifting behaviors after protonation can be understood through the resonance picture of the GFP chromophore in the ground state [79]. The mode we are interested in, however, has the lowest vibrational frequency among the aforementioned candidates and is frequently attributed to the phenol C–O stretching; its red-shifting behavior can also be understood via resonance. From the recently reported high resolution structures [50], the C–O bond is observed to be significantly shorter in the anionic B state than the protonated A state. This mode is chosen because the phenol C–O bond is in direct interaction with the proton we are interested in. The sensitivity of the vibrational frequency to the proton position has been utilized for probing the ultrafast process of ESPT through femtosecond Raman spectroscopy [80]. Note also that the NMR chemical shift of the corresponding ^{13}C is sensitive to the chromophore's protonation states [17].

It is useful to confirm the assignment of the phenol C–O stretching using GFPs with all tyrosines ^{13}C-labeled at the ζ carbon (including the chromophore's) created in the course of our earlier NMR study [17]. Tonge's isotopic labeling strategy was applied to the bridge and the imidazolinone moiety of the model chromophore [78], while the

assignment of the C–O stretching is only inferred from other aromatic compounds and its proton sensitivity [75]. If C–O stretching were a reliable assignment, we would expect a large red-shift (~ 20 cm^{-1} based on reduced masses) with ^{13}C substitution. Curiously, after comparing the pre-resonance Raman spectra of both ih:GFP S65T and ih:GFP S65T/H148D with ^{12}C and ^{13}C at the ζ carbons (Figure S17), the peak of interest red-shifts at most 2 cm^{-1}. This is in stark contrast with tyrosine itself: while the corresponding peak in the same region still changes from 1250 to 1270 cm^{-1} upon protonation, ^{13}C labeling at the ζ carbon red-shifts the former and the latter to 1228 and 1246 cm^{-1}, respectively [81]. This suggests that by conjugating to the imidazolinone moiety, the C–O stretching mode in the GFP chromophore is no longer as local as in tyrosine, presumably due to coupling to other stretching modes that are close in frequency, and thereby diluting the contribution of the C–O stretching local mode. Our observation is also supported by another GFP study, in which the ring carbons of the phenol moiety were all ^{13}C-labeled: no shift was observed for the 1240 cm^{-1} peak when comparing the UV pump-infrared probe spectra from the A-state proteins with and without ^{13}C labeling [82]. Nevertheless, the "C–O stretching" band is proton sensitive and its non-locality does not defeat our original purpose. Even the higher frequency C=C and C=N stretching modes are not completely local due to their couplings to nearby C=C, C=N and C=O modes [76][77][79]. In the following, we will refer to these observed Raman peaks as the simplest associated modes and drop the quotation marks for convenience, while keeping in mind that the meaning of these assignments as local modes is not completely correct. As a side note, even though DFT calculations have been extensively employed to help assign normal modes and the predicted frequencies are mostly reliable, we performed a simple calculation with Gaussian [83] using the B3LYP functional and a basis set of 6-311++g(d,p) with geometry optimization and found that it misleadingly predicts a shift from 1264 to 1241 cm^{-1} in the 633 nm excited pre-resonance Raman spectra for the deprotonated HBDI *in vacuum* upon ^{13}C labeling at the ζ carbon. This suggests any *in vacuo* DFT calculations should be treated with caution when comparing against experimental data performed in condensed phases.
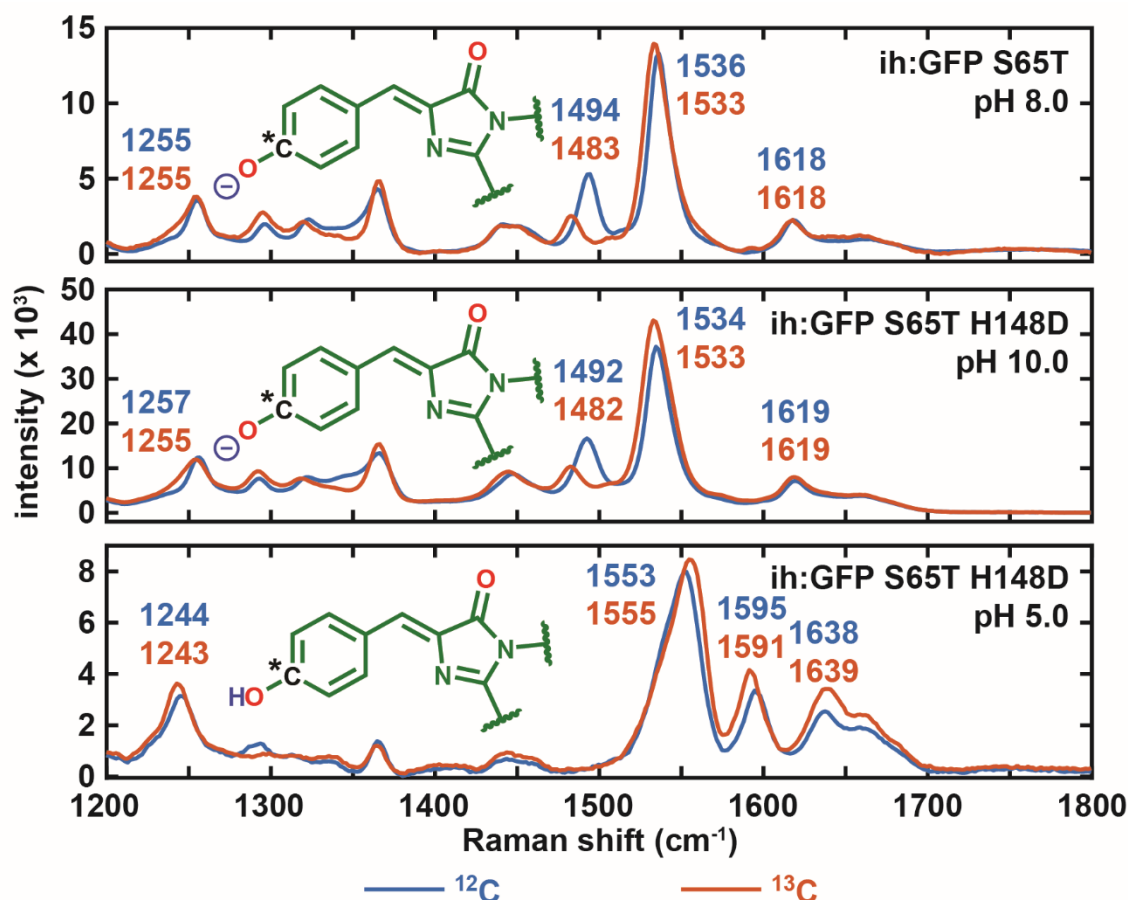
Figure S17. Pre-resonance Raman spectra excited at 633 nm for GFP mutants in both protonation states with ^{12}C (blue trace) and ^{13}C (orange trace) at the ζ carbons (labeled with asterisk in the structure). The corresponding peak positions are also labeled and color coded. The "C–O stretching" band, which is at 1244 cm^{-1} for the protonated and 1255 cm^{-1} for the deprotonated chromophore, barely shifts upon ^{13}C substitution, as opposed to the tyrosine counterpart [81]. However, the 10 cm^{-1} red shift for the 1494 cm^{-1} band upon ^{13}C substitution, corresponding to a phenolate stretching mode [78], is retained within the deprotonated GFP chromophore. A similar yet less dramatic red shift also occurs to the phenol stretching mode (1595 cm^{-1}) [77] for the protonated chromophore.

We begin the analysis of the short-hydrogen-bond GFPs with an unsubstituted chromophore (Y66), which is the most well-studied case. As Tonge and coworkers have demonstrated [75], some Raman peak positions strongly correlate with the chromophore's electronic transition energy and therefore reflect the fractions of double- and single-bond characters for the GFP chromophore, or the relative contributions of the P and I forms (or GS and CT forms for the protonated chromophore [68]) to the ground-state structure from the perspective of Marcus-Hush theory [3]. As a reference, the C=C

and C=N stretching bands appear at 1618 (1641) and 1536 (1555) cm^{-1} for the B (A) state, respectively, for ih:GFP S65T. For ih:GFP S65T/H148D, with a short hydrogen bond, while the B state peak positions (1619 and 1534 cm^{-1}) match the H148 counterparts nicely, the A state peaks (1638 and 1553 cm^{-1}) are slightly red-shifted from the H148 counterparts, suggesting that the short hydrogen bond imparts some anionic character to the protonated chromophore through elongating the O–H distance, hence the A-like A state. Similarly, while the C–O stretching peaks for most GFP mutants appear at 1240 cm^{-1} in the A state, the corresponding peak shows up at 1244 cm^{-1} and exhibits an even bluer shoulder for the protonated short-hydrogen-bond GFP (Figures 4A and S17). Upon deuteration, the blue shoulder disappears, and the main band remains at 1244 cm^{-1} (Figure 4A), which likely suggests a more localized deuteron wavefunction than the proton counterpart probed by the C–O stretching mode (Figure 4B). The blue shoulder can be explained by the tiny B-like A state fraction from the anomalous A state, which is also revealed by deconvolution using electronic Stark spectroscopy (Section S5, Figure S10). Notably, as opposed to the 16 cm^{-1} red shift of the C–O stretching peak after deuteration of avGFP reported previously in the ultrafast ESPT study (see the Supplementary Information of [80]), we clearly see no measurable change in the main band position in our GFP mutant even though the short hydrogen bond should exhibit a larger isotope effect.

For the monohalogenated variants (excluding the monofluorinated one, *vide infra*), the assignments become tricky. Experimentally, isotope labeling fails to identify the C–O stretching band and the chromophores' pK_a are too low to allow for complete titrations. Even worse, the asymmetric substitution pattern breaks the approximate C_{2v} symmetry of the phenol(ate) moiety (Figure S1) and completely scrambles the normal modes of the chromophore, as evinced by comparing IR and Raman spectra for free 3-F$_1$Y and Y combined with related DFT studies [84][85]. In other words, given the non-locality of the normal modes at low frequencies and the change in identities of normal modes across the unsubstituted and substituted model chromophores, it is not possible to unambiguously single out the Raman band that inherits the most C–O character or resembles the wild-type chromophore's C–O stretching band the most through DFT calculations (Figure S18). Nevertheless, we believe that at least one of the bands in the

same region (1200 – 1300 cm^{-1}) should still behave similarly to the one examined for the unsubstituted chromophore, and we present the measured Raman spectra for the protonated and deuterated monohalogenated variants in Figures 4A and S19. Across the 3-Cl$_1$Y, 3-Br$_1$Y, and 3-I$_1$Y variants, the band at 1266 cm^{-1} consistently splits into two and shows another clearly resolved peak at 1275 cm^{-1} upon deuteration, while none of the other bands in this region of the spectra show as much change. The appearance of the blue peak upon deuteration in consistent with the B-like A state and can be understood using the ground-state proton/deuteron wavefunction calculated from our previous study [1] (Figure 4B), so the two peaks could suggest a placement of the deuteron zero-point energy (ZPE) below the barrier within the short hydrogen bond, providing further evidence against its designation as a LBHB. The proton wavefunction might be too delocalized such that even if its corresponding ZPE is below the barrier, the underlying two populations can still be unresolved (Figure 4B). The two peaks merge again a few days after the initial deuteron exchange due to ambient moisture, indicating that the spectral change is reversible.
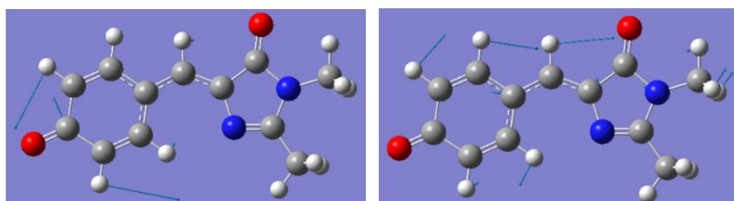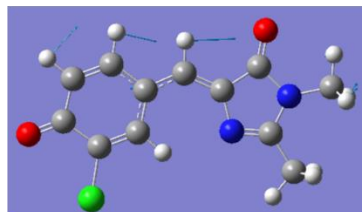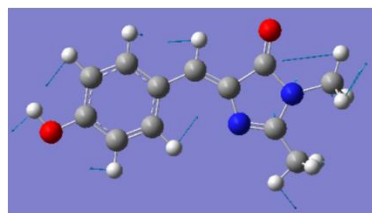
Mode #: 45
Frequency: 1264 cm^{-1}
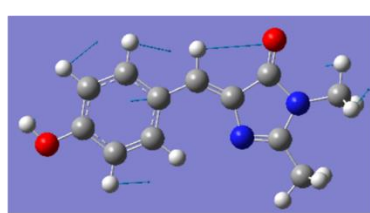
Mode #: 46
Frequency: 1278 cm^{-1}
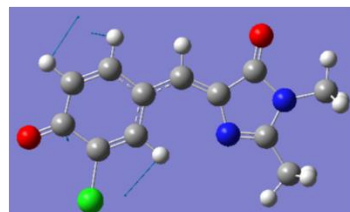
Mode #: 46
Frequency: 1221 cm^{-1}

Mode #: 47
Frequency: 1275 cm^{-1}

Mode #: 46
Frequency: 1203 cm^{-1}

Mode #: 47
Frequency: 1252 cm^{-1}
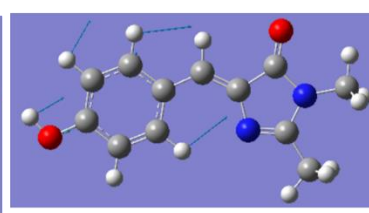
Mode #: 48
Frequency: 1289 cm^{-1}
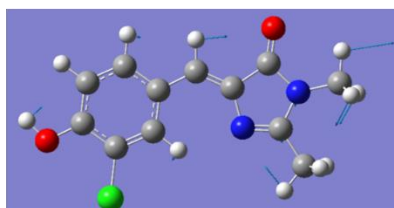
Mode #: 47
Frequency: 1202 cm^{-1}
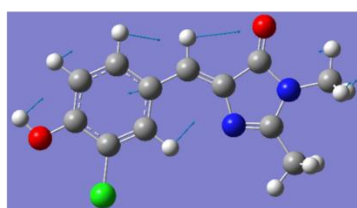
Mode #: 48
Frequency: 1244 cm^{-1}
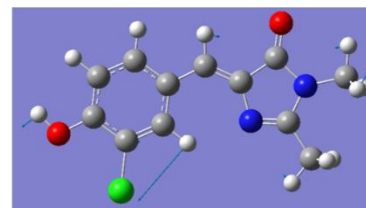
Mode #: 49
Frequency: 1292 cm^{-1}

Figure S18. DFT normal mode analysis for (A) deprotonated and (B) protonated model chromophores, including the unsubstituted and the monochlorinated chromophore, in vacuum. Only modes with calculated frequencies of 1200 – 1300 cm^{-1} are shown. The results are obtained using the B3LYP functional and a basis set of 6-311++g(d,p) with geometry optimization. The mode number (Mode #) are ordered from the lowest to the highest frequency modes according to Gaussian. Due to the change in coupling patterns between the local vibrational modes, normal modes with the same mode number are not similar.
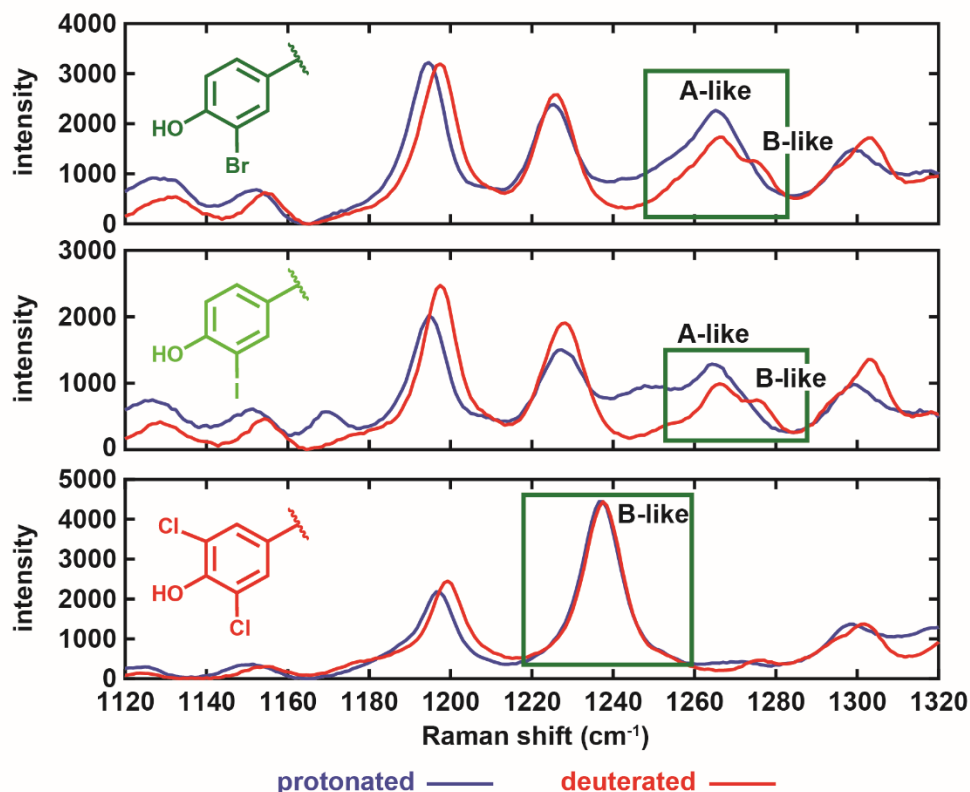
Figure S19. Pre-resonance Raman spectra with 633 nm excitation of protonated (pH 5, blue traces) and deuterated (pD 5, red traces) S65T H148D GFP variants not including those with fluorines. The corresponding spectra for the Y66 and Y66(3-Cl$_1$Y) constructs can be found in Figure 4A. The peaks of interest, corresponding to a proton-sensitive phenol stretching mode, are highlighted within green boxes.

The only problem left with treating 1266 cm^{-1} as the C–O stretching band for the monohalogenated chromophores is its higher frequency than the unsubstituted counterpart (1248 cm^{-1}), which seems counterintuitive since introduction of a heavy atom should decrease the vibrational mode frequencies. However, there are two additional governing factors. First, since halogens are electron-withdrawing groups, modifying the phenol(ate) with them leads to further electron delocalization across the chromophore [3] and blue-shifts the C–O stretching mode. The additional C–halogen bond could also interact with the environment and perturb the chromophore's vibrational frequencies. Second, due to the previously mentioned symmetry breaking, the coupling pattern between the local modes is no longer conserved after halogenation. Since vibrational frequencies for the C–halogen stretching modes (< 800 cm^{-1}) are much lower than those for C–O stretching, the substituted side of the phenol(ate) moiety becomes heavier than

the unsubstituted side and some vibrational modes (especially those involving the phenol(ate)) are thus blue-shifted upon halogen substitution [83][84]. We have also identified this trend for normal modes that can be readily compared between the halogenated and substituted chromophores through DFT calculations (Figure S20).
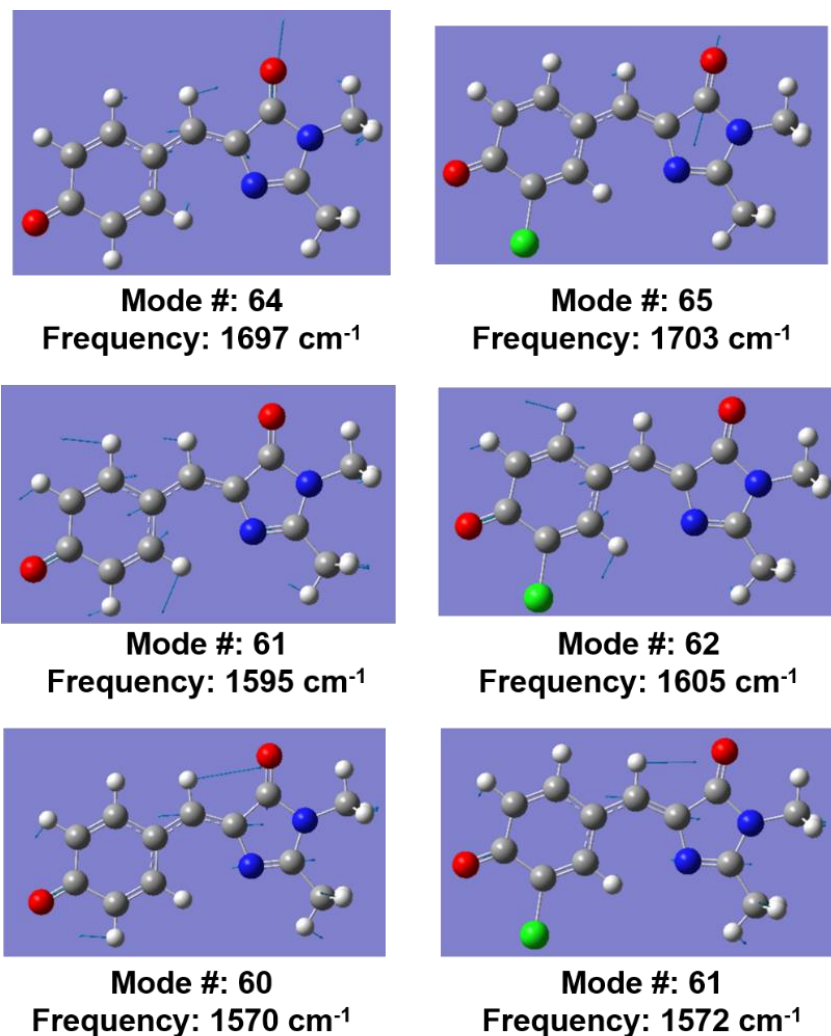

Mode #: 64
Frequency: 1697 cm^{-1}

Mode #: 65
Frequency: 1703 cm^{-1}

Mode #: 61
Frequency: 1595 cm^{-1}

Mode #: 62
Frequency: 1605 cm^{-1}

Mode #: 60
Frequency: 1570 cm^{-1}

Mode #: 61
Frequency: 1572 cm^{-1}

Figure S20. Comparison of similar normal modes between those of the unsubstituted (left) and the monochlorinated model chromophores (right) in vacuum. The results are obtained using the B3LYP functional and a basis set of 6-311++g(d,p) with geometry optimization. Monochlorination can lead to blue shifts in vibrational frequencies.

For the 3,5-Cl$_2$ variant, only one dominant peak (1238 cm^{-1}) can be observed in 1200 – 1300 cm^{-1} region, and we cannot spot any significant difference between the Raman spectra for the protonated and deuterated proteins (Figure S19), as expected due to the large pK_α mismatch (ΔpK_α = -1.5). The low frequency of this band follows our

intuition with chorine substitution and is consistent with the restoration of the phenol(ate)'s approximate C$_{2v}$ symmetry. Unfortunately, for the fluorinated variants, including the most pK$_\alpha$ matched cases, the band of interest is masked by the intense features from C–F stretching (Figure S21) due to the close vibrational frequencies of the C–F and C–O stretching modes [84][85]. The situation is especially dire for the 3,5-F$_2$ variant, since the Raman scattering signal from the symmetrically substituted fluorines is much more enhanced.
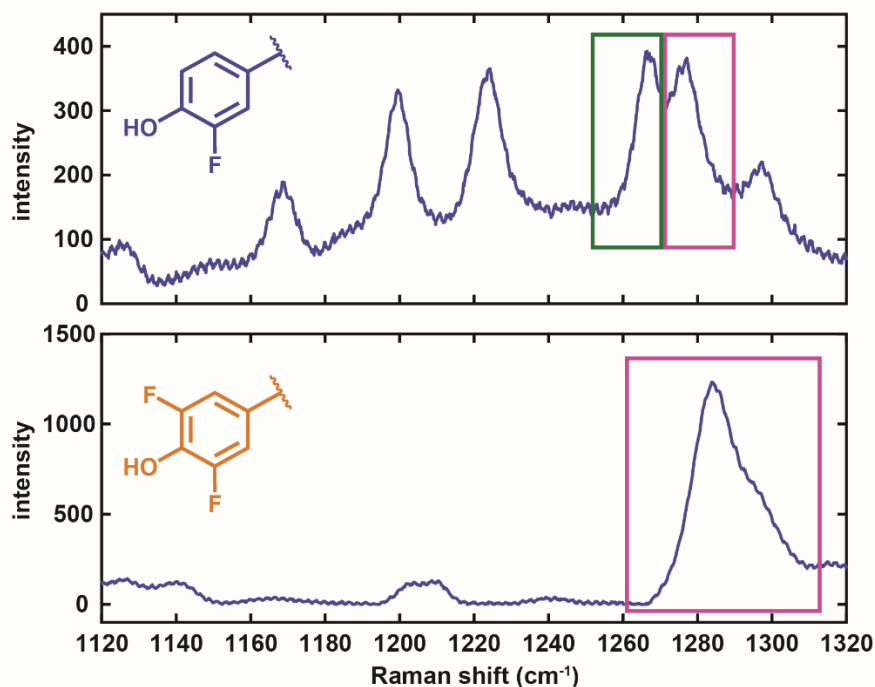

Figure S21. Pre-resonance Raman spectra with 633 nm excitation of protonated (pH 5) S65T H148D GFP variants with fluorines. The peak of interest, corresponding to a proton-sensitive phenol stretching mode, is highlighted within a green box. The C–F stretching peak [83][84], highlighted by purple boxes, partially and completely mask the peak of interest for the 3-F$_1$Y and 3,5-F$_2$Y variants, respectively.

In short, while we are unable to obtain any useful information from monitoring the change in Raman spectra for the protonated and deuterated fluorinated variants due to the overwhelming C–F stretching signals, we can infer the qualitative shape of the short hydrogen bond PES from the rest via the presumed C–O stretching mode. The Raman bandshapes are consistent with the proton/deuteron wavefunctions calculated from the PES model using the pre-determined ΔpK$_\alpha$ and reinforce our conclusions from electronic Stark spectroscopy (Section S5). These room- and low-temperature methods altogether

suggest a delocalized proton within the short hydrogen bond of the S65T/H148D GFP variants, but not delocalized enough for us to conclude the presence of an LBHB [1][86][87].

S7   Local Field Factor f_{HB} due to Proton Polarizability of Short Hydrogen Bonds

Consider an isotropically distributed sample of the short-hydrogen-bond GFP molecules under an applied constant electric field $\vec{F}_{ext}$. To understand the electrostatic properties of the critical structural components, namely the chromophore and the short hydrogen bond, we invoke some approximations to provide a reductionist picture of the system. We will treat the short hydrogen bond in terms of its dipole induced by the external field:

$$\vec{\mu}_{PT} = \alpha_{PT}\vec{F}_{ext} \tag{S3}$$

where the subscripts PT mean "proton transfer" and α_{PT} is the proton's electric polarizability in the short hydrogen bond; proton transfer in this context means any shift in the proton position due the external field. In this expression we already introduced several assumptions. First, we assume linear response of the proton to the external field. This is justified by the fact that under experimentally accessible external field strengths (up to ~ 1 MV/cm), higher order terms (i.e., hyperpolarizabilities) can be ignored altogether according to Zundel's calculations [88][89]. Second, instead of treating α_{PT} as a tensor, here we treat it as a scalar, i.e., as the zz component of the tensor, placing the hydrogen bond along the z direction. With the much shallower proton PES along the linear path between the heavy atoms than along the orthogonal directions due to the stronger coupling between two O–H potentials [54], it should be easier to perturb the proton position along that linear path. This assertion is further supported by analyzing the linearity of short hydrogen bonds with structures from crystallography [90][91][92] or computation [93][94], where protons are mostly located colinearly between the heavy atoms of short hydrogen bonds within various electrostatic environments. Proton conductivities along proton chains were also found to be greater than in the perpendicular direction [95]. Finally, since the field-induced proton displacement is tiny (on the order of 10^{-3} to 10^{-4} Å for short hydrogen bonds for applied fields under 1 MV/cm, Section S8, see also [96]) compared to the proton-chromophore distance (on the order of Å), proton polarization can indeed be well approximated as a point dipole.

For the chromophore, we can use the difference dipole moment $\Delta\vec{\mu}$ between its electronic excited and ground states to understand the electrostatic response of its

transition energy to the electric field it experiences [23]. While the magnitude of $\Delta\vec{\mu}$ is obtained via electronic Stark spectroscopy [23], where a *homogeneous* field is applied, the induced field from the proton displacement is *inhomogeneous*, and thus it is not at all obvious how one can justify the use of $\Delta\vec{\mu}$ in the current scenario. From another perspective, it would be nice to approximate the proton – chromophore interaction as an induced dipole–difference dipole interaction but given the close proximity of the proton to the chromophore, this simple picture might fail to hold. If we were to adopt this simple approximation, it would be necessary to discuss where we should place this difference dipole (see the discussion in Section S8 of [3]), as any choice of the dipole position would affect the proton-chromophore distance r, which is a critical parameter discussed below. Since we have previously argued that the effective electric field experienced by the chromophore is the electric potential difference between the oxygens divided by their distance based on the diabatic-state model (Section S8 of [3]) and the proton field is far more significant at the phenol(ate) oxygen than the one on the imidazolinone ring (which is ~ 9 Å away), it is reasonable to approximate r as the distance between the proton and the phenol(ate) oxygen. On top of the insidious problem of locating the difference dipole moment, the electron cloud is highly polarizable, and the proton displacement could thereby induce a change in the chromophore's difference dipole moment. However, as we will later estimate in Section S8, the strength of the induced proton field is comparable to the experimentally achievable electric field strength, which amounts to 1 MV/cm and is much smaller than the effective field from the GFP environment itself (~ 20 MV/cm) [97]. One can therefore neglect the difference electronic polarizability term for the chromophore, supporting the use of $\Delta\vec{\mu}$ with a constant magnitude.

Given these assumptions, we can now start analyzing this simplified difference dipole-induced dipole model with relative geometries of the chromophore and the short hydrogen bond considered. The goal is to show this model indeed generates Stark spectra that are proportional to F_{ext}^2 (as in classical Stark spectroscopy) but with an enhanced (or diminished) apparent Stark tuning rate $\Delta\mu_{app}$ due to the influence of proton polarization.
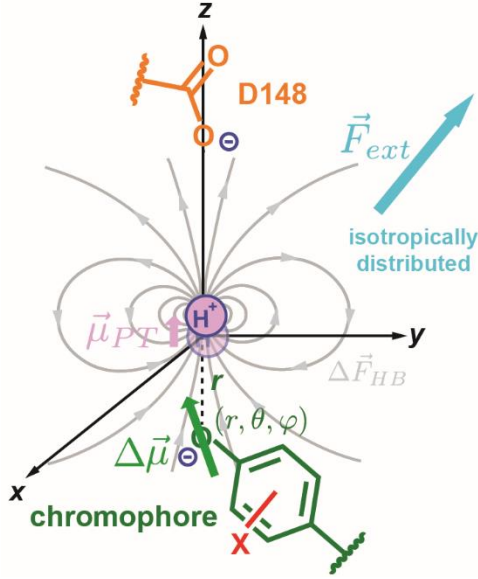
Figure S22. The difference dipole-induced dipole model in the molecular frame. The difference dipole $\Delta\vec{\mu}$ of the chromophore (green arrow) is assumed to be located at the phenolate oxygen, whose position is (r, θ, φ). The intrinsic orientation of $\Delta\vec{\mu}$ is characterized by a pair of angles (Θ, Φ) (not labeled for clarity). The induced dipole $\vec{\mu}_{PT}$ of the proton (pink arrow) is placed at the origin and points toward the positive z direction. In the molecular frame, the external electric field $\vec{F}_{ext}$ is isotropically distributed (aqua arrow). Due to the linearity of the hydrogen bond, only the z component of $\vec{F}_{ext}$ can effectively induce an appreciable displacement of the proton and lead to the corresponding induced dipole field $\Delta\vec{F}_{HB}$ (gray field lines).

There are two vectors ($\vec{\mu}_{PT}$ and $\Delta\vec{\mu}$) associated with the isotropically distributed GFP molecules while there is only one vector $\vec{F}_{ext}$ with a constant orientation, suggesting that it is easier to deal with the problem in the molecular frame, where $\vec{\mu}_{PT}$ sits along the z axis at the origin (Figure S22). $\Delta\vec{\mu}$ locates at the spherical coordinates (r, θ, φ) (where θ is the polar angle and φ is the azimuthal angle). The intrinsic direction of $\Delta\vec{\mu}$ is defined by another set of angles (Θ, Φ), observing an isotropic distribution of $\vec{F}_{ext}$. The magnitude of the induced dipole field from proton polarization $\Delta\vec{F}_{HB}$ is proportional to the induced dipole μ_{PT} [98]:

$$\Delta\vec{F}_{HB}(r,\theta,\varphi) = \frac{\mu_{PT}}{4\pi\epsilon_0 r^3}(2\cos\theta\,\hat{r} + \sin\theta\,\hat{\theta})$$

(S4)

in which ϵ_0 is the vacuum permittivity, and $\hat{r}$ and $\hat{\theta}$ are the unit vectors in the polar coordinate system and change directions depending on the position (r, θ, φ). We sacrifice the elegance of Equation S4 and express the dipole field in terms of Cartesian unit vectors:

$$\Delta \vec{F}_{HB}(r, \theta, \varphi) = \frac{\alpha_{PT} F_{ext,z}}{4\pi\epsilon_0 r^3} [3 \sin \theta \cos \theta \cos \varphi \, \hat{x} + 3 \sin \theta \cos \theta \sin \varphi \, \hat{y} + (3 \cos^2 \theta - 1)\hat{z}]$$

(S5)

where we also plug in Equation S3. Here the induced dipole field is linearly proportional to the z component of the external field due to proton polarization. The total internal field $\vec{F}_{int}$ experienced by the chromophore, with the "zero-field" reference state as the GFP itself without the external field, can then be expressed as (cf. Equation 2 in the main text)

$$\vec{F}_{int}(r, \theta, \varphi) = \vec{F}_{ext} + \Delta \vec{F}_{HB}(r, \theta, \varphi) = \underline{f}_{HB}(r, \theta, \varphi) \cdot \vec{F}_{ext}$$

(S6)

with the local field factor from proton polarization $\underline{f}_{HB}$ as a tensor (hence the underscore):

$$\underline{f}_{HB}(r, \theta, \varphi) = \begin{pmatrix} 1 & 0 & \frac{3\alpha_{PT}}{4\pi\epsilon_0 r^3} \sin \theta \cos \theta \cos \varphi \\ 0 & 1 & \frac{3\alpha_{PT}}{4\pi\epsilon_0 r^3} \sin \theta \cos \theta \sin \varphi \\ 0 & 0 & 1 + \frac{\alpha_{PT}}{4\pi\epsilon_0 r^3} (3 \cos^2 \theta - 1) \end{pmatrix}$$

(S7)

which is anisotropic due to the directionality of the polarizable hydrogen bond, but whose components are nonetheless independent of the external field strength. With this local field factor $\underline{f}_{HB}$, the probe $\Delta \vec{\mu}$ experiences a different overall field $\vec{F}_{int}$ compared to $\vec{F}_{ext}$. However, the field distribution is no longer isotropic because of the preferential direction (i.e., anisotropy) of the hydrogen bond polarization. This seems to contradict to the assumption of isotropy on which the observed Stark spectra being proportional to F_{ext}^2 relies (Figure S12). This conundrum is saved by the fact that the Stark shift $\Delta \bar{v}$ of the chromophore only relies on the field projection on $\Delta \vec{\mu}$:

$$\Delta \bar{v} = -\Delta \vec{\mu}^T \vec{F}_{int} = -\Delta \vec{\mu}^T (\underline{f}_{HB} \vec{F}_{ext}) = -(\underline{f}_{HB}^T \Delta \vec{\mu})^T \vec{F}_{ext} = -\Delta \vec{\mu}_{app}^T \vec{F}_{ext} \qquad \text{(S8)}$$

where T means transpose, and the dot products are written in terms of matrix notations. Equation S8 tells us that even though $\Delta\vec{\mu}$ by itself does not experience an isotropic distribution of fields $\vec{F}_{int}$, it is still possible to find a rotated and scaled vector $\Delta\vec{\mu}_{app} \equiv \underline{f_{HB}}^T \Delta\vec{\mu}$ which experiences an isotropically distributed $\vec{F}_{ext}$, such that the observed Stark shifts stay the same. In other words, the observed Stark spectra would still be proportional to F_{ext}^2, but the extracted apparent Stark tuning rate will be $\Delta\mu_{app} \equiv \left| \underline{f_{HB}}^T \Delta\vec{\mu} \right|$ instead of $|\Delta\vec{\mu}|$. More generally, as long as the local field factor is a constant tensor, the classical Stark spectroscopy analysis (Equations S1 and S2) is still obeyed even if the factor is anisotropic, which explains why even though most molecules are non-spherical and cause an anisotropic local field factor from the reaction field mechanism [99], their Stark spectra are still proportional to F_{ext}^2 (unless basic assumptions, intact bandshape and population upon field applications of the *classical* Stark spectroscopy analysis are violated [67]).

From Equation S8, we can now calculate the apparent Stark tuning rate $\Delta\mu_{app}$ obtained in this short hydrogen bond-chromophore system:

$$\Delta\mu_{app}^2 = \Delta\vec{\mu}^T \underline{f_{HB}} \underline{f_{HB}}^T \Delta\vec{\mu} \equiv (f_{HB}\Delta\mu)^2 \tag{S9}$$

where the proportionality constant f_{HB}, which is now a scalar, depends on the angles in the model:

$$f_{HB}(r, \theta, \varphi, \Theta, \Phi)$$

$$= \sqrt{\sin^2\Theta + \left[\cos\Theta + \frac{\alpha_{PT}}{4\pi\epsilon_0 r^3}(3\sin\theta\cos\theta\sin\Theta\cos\Delta\phi + (3\cos^2\theta - 1)\cos\Theta) \right]^2}$$

$$\tag{S10}$$

where the two azimuthal angles have been reduced to their difference $\Delta\phi \equiv \varphi - \Phi$, which makes sense because it only takes three independent angles to characterize the relative geometry between two non-coplanar vectors. As a check of limiting cases, one finds f_{PT} to be 1 when there is no proton polarization and $1 + \frac{2\alpha_{PT}}{4\pi\epsilon_0 r^3}$ when the hydrogen bond is colinear with the difference dipole moment ($\theta = 0$ or π and $\Theta = 0$ or π) as in the main text (Equation 2 with α_{PT} evaluated classically using Equation S15). If we assume $\frac{\alpha_{PT}}{4\pi\epsilon_0 r^3} = 1$ and $\Delta\phi = 0$ (the two dipole moments are coplanar), we obtain the dependence of f_{HB} on

the probe position's polar angle θ and the probe direction's polar angle Θ (Figure S23). Interestingly, depending on the relative orientation, f_{HB} can be either greater or less than 1 because the short hydrogen bond is an anisotropic source of the induced field, which enhances the external field at some positions but partially cancels out the external field at others [100]. Another interesting consequence is that the measured angle ζ_{app} between the apparent difference dipole $\Delta\vec{\mu}_{app}$ and the transition dipole $\vec{m}$ of the chromophore can also change by virtue of rotation/scaling from the anisotropic local field factor:

$$\cos \zeta_{app} = \frac{(\underline{f_{HB}}^T \Delta\vec{\mu})^T \vec{m}}{\left|\underline{f_{HB}}^T \cdot \Delta\vec{\mu}\right| |\vec{m}|}$$

(S11)

in which the transition dipole presumably does not carry $\underline{f_{HB}}$ because the UV–visible probe light frequency (~ 10^4 cm^{-1}) is too fast to keep the hydrogen bond polarized (with an intrinsic frequency of $10^2 - 10^3$ cm^{-1}). However, this effect might be masked by the intrinsic angle differences among various halogenated chromophores (Table S8) [3].

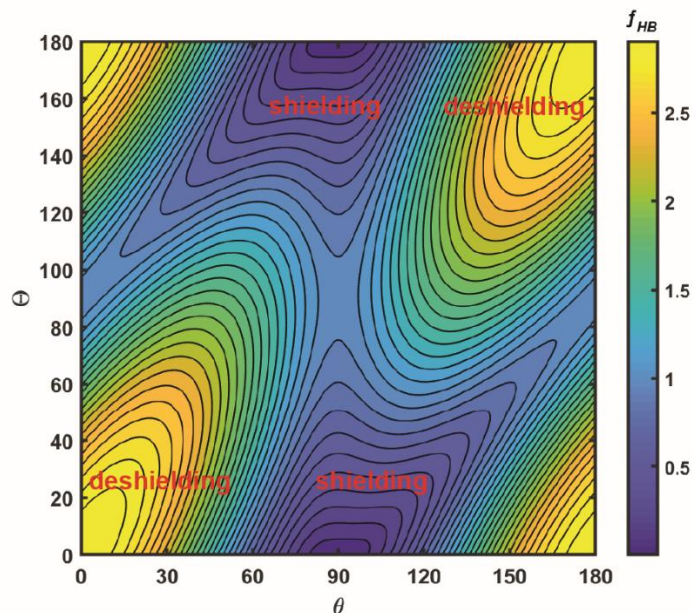**Figure S23.** Local field factor anisotropy from proton polarization as a function of two polar angles θ *and* Θ, where $\frac{\alpha_{PT}}{4\pi\epsilon_0 r^3} = 1$ and $\Delta\phi = 0$ are assumed. In the region where f_{HB} is greater than 1 (green to yellow zones), the internal field sensed by the chromophore is enhanced, hence "deshielding". With certain relative geometries between the proton and

the chromophore, f_{HB} can also be less than 1 (deep blue to purple zones, "shielding"). Equation 2 in the main text corresponds to the head-on probing case, which locates at the four corners of this plot.

In conclusion, we expect a change in the apparent Stark tuning rate of a chromophore and the classical Stark analysis is still valid if there is a linearly polarizable moiety in proximity. This polarizable moiety could be a short hydrogen bond, and the associated local field factor depends on the hydrogen bonding geometry.

S8   Evaluating Proton Polarizabilities from Classical and Quantum Models

The relation between the PES local curvatures, degrees of proton delocalization and proton polarizabilities can be intuitively understood as follows: the shallower the potential, the more delocalized the proton and thereby easier to be polarized. In this section we aim to provide a quantitative argument on these connections based on simple classical and quantum models. The latter are anticipated to offer more accurate descriptions for short hydrogen bonds; nonetheless the former can still be useful starting points. We will also present some back-of-the-envelope estimations on proton polarizabilities.

Let us first consider a one-dimensional double-well PES of a hydrogen bond $V_{HB}(r)$ under an electric field F_{ext}, which introduces a linear potential $-eF_{ext}r$ that biases the potential's shape (Figure 7). If a proton stays in one of the wells, by applying an electric field it is not possible to move it over the barrier in the classical model due to the lack of a tunneling mechanism. Instead, the position of the energy minimum r_0 would be slightly shifted by Δr:

$$\left.\frac{\partial V_{tot}}{\partial r}\right|_{r=r_0+\Delta r} = V_{HB}'(r_0 + \Delta r) - eF_{ext} = 0$$

(S12)

where the total potential energy is $V_{tot}(r) = V_{HB}(r) - eF_{ext}r$. Since the applied field is a small perturbation, we can expand Equation S12 to first order in Δr:

$$V_{HB}''(r_0)\Delta r - eF_{ext} = 0 \tag{S13}$$

in which we recognize that $V_{HB}'(r_0) = 0$ at the energy minimum. We can then obtain the induced dipole as

$$\mu_{PT} = e\Delta r = \frac{e^2 F_{ext}}{V_{HB}''(r_0)}$$

(S14)

which is linear in F_{ext}. The corresponding classical expression for proton polarizability is

$$\alpha_{PT} = \frac{e^2}{V_{HB}''(r_0)}$$

(S15)

The polarizability is inversely proportional to the local curvature of the hydrogen bond PES as expected. Due to the decrease in local curvature, we anticipate the proton polarizability to rise as the heavy-atom distance shortens [54]. Using the symmetrically coupled Morse potential model parameterized by McKenzie [54] to get a rough idea of the magnitude of the effect, one finds the local curvatures to be 1140 and 600 kcal/(mol Å^2) for O–O distances at 3 and 2.45 Å, respectively. The corresponding proton polarizabilities are then 0.29 and 0.56 Å^3 (see Section S9 for unit conversion), which might seem modest but actually is not so if one considers the relevant parameter $\frac{\alpha_{PT}}{r^3}$ in Equation S10 (or Equation 2 in the main text) with $r \sim 1$ Å. Remarkably, the proton displacement Δr amounts to 4×10^{-4} Å for O–O distance at 2.45 Å with an 1 MV/cm external field, while the corresponding induced dipole field from proton polarizability is comparable to 1 MV/cm (Equation 2), owing to the short distance r.

The classical model is simple and physically transparent. However, it predicts α_{PT} to be independent of the mass of hydrons (i.e., proton and deuteron), because the mass effect is only manifested in terms of quantum nuclear effects (e.g., zero-point energy), while the classical model is the infinite mass limit. A slightly improved way of incorporating the mass effect would be to think in terms of wavefunctions. Because steeper potentials and larger masses yield less delocalized ground-state wavefunctions, we can qualitatively connect between proton polarizability and proton delocalization and infer larger polarizabilities for protons than deuterons in the same potential well [102].

A full quantum mechanical treatment can be initiated using time-independent perturbation theory, justified because the external applied field modulation frequency ($\omega \sim 200$ Hz) is orders of magnitude less than the hydrogen bond vibrational frequencies ($10^{12} - 10^{13}$ Hz) and therefore quasi-static. The perturbed ground state due to an electric field is

$$|0'\rangle = |0\rangle + eF_{ext} \sum_{n \neq 0} \frac{\langle n|x|0 \rangle}{E_n - E_0} |n\rangle$$

(S16)

which is a superposition of the original ground state and various excited states $|n\rangle$ weighted by their accessibilities (i.e., transition dipole moments and energy gaps) from the ground state. We can calculate the induced dipole moment as

$$\mu_{PT} = e\langle 0'|x|0'\rangle - e\langle 0|x|0\rangle = \alpha_{PT}F_{ext} + O(F_{ext}{}^2) \tag{S17}$$

where the corresponding linear polarizability is

$$\alpha_{PT} = 2e^2 \sum_{n\neq 0} \frac{|\langle n|x|0\rangle|^2}{E_n - E_0}$$

$$\tag{S18}$$

This expression is notoriously hard to evaluate because of its sum-over-state nature, which is also encountered in nonlinear optics (e.g., Raman spectroscopy and two-photon absorption [103]). For harmonic oscillators, it is in fact simple because only the first excited state has a nontrivial transition dipole moment with the ground state:

$$\alpha_{PT} = 2e^2 \frac{|\langle 1|x|0\rangle|^2}{E_1 - E_0} = \frac{2e^2}{\hbar\omega} \cdot \frac{\hbar}{2m\omega} = \frac{e^2}{m\omega^2} = \frac{e^2}{k}$$

$$\tag{S19}$$

in which m is the proton or deuteron mass, ω is the angular frequency and k is the corresponding spring constant. The result is curiously the same as its classical counterpart (Equation S15) and mass independent. We can therefore conclude that the mass effect arises from the anharmonicity of the potential.

With anharmonic potentials, we cannot really proceed without any approximation. For a double-well potential, one can treat the energy levels as a two-level system (tunneling approximation) by recognizing the ground state $|0^+\rangle$ (even parity) being close in energy to the first excited state $|0^-\rangle$ (odd parity) due to tunneling, while higher excited states are rather energetically inaccessible. We can then ignore the contributions from higher energy levels:

$$\alpha_{PT} \approx 2e^2 \frac{|\langle 0^-|x|0^+\rangle|^2}{\hbar\omega}$$

$$\tag{S20}$$

where $\hbar\omega$ is the energy gap between $|0^+\rangle$ and $|0^-\rangle$, namely the tunneling splitting. Equation S20 is also applicable to any two-level system. To be slightly more realistic, one

can also include the polarizability from $|0^-\rangle$ (note the negative sign in the difference energy) and Boltzmann weigh the contributions:

$$\alpha_{PT} = \alpha_+ + \alpha_- = 2e^2 \frac{|\langle 0^-|x|0^+\rangle|^2}{\hbar\omega} \tanh\frac{\hbar\omega}{2k_B T}$$

(S21)

which is exactly the expression proposed by Zundel [88]. For a mildly coupled *symmetric* double-well potential, we can approximate the two states by the superposition of ground states in the left and right wells:

$$|0^+\rangle = \frac{1}{\sqrt{2}}(|0_L\rangle + |0_R\rangle), \qquad |0^-\rangle = \frac{1}{\sqrt{2}}(|0_L\rangle - |0_R\rangle)$$

(S22)

and the transition dipole moment between the two states is

$$e\langle 0^-|x|0^+\rangle = e\langle 0_L|x|0_L\rangle = e\int \psi_L^*(x)(x-r_0)\psi_L(x)dx + e\int \psi_L^*(x)r_0\psi_L(x)dx \approx er_0$$

(S23)

The first equality is from $\langle 0_L|x|0_R\rangle = \langle 0_R|x|0_L\rangle$ and $\langle 0_L|x|0_L\rangle = -\langle 0_R|x|0_R\rangle$; the first integral term is approximately zero with r_0 as the distance between the local minima and the local maximum (center) of the proton PES which is around 0.5 and 0.25 Å when O–O distances are 3 and 2.45 Å, respectively. Zundel argued based on Equation S21 that polarizabilities of protons in hydrogen bonds can be comparable to those of electronic systems [88]. For example, the tunneling splitting for short hydrogen bonds can reach 500 cm^{-1} according to McKenzie's model [54], while the S_0 to S_1 energy gap of the anionic GFP chromophore is around 21000 cm^{-1} [3]. As evaluated above, the transition displacement of a proton in the short hydrogen bond is about 0.25 Å, while that of the electron in the anionic GFP chromophore is about 1.2 Å [3]. Ground-state polarizabilities are therefore estimated to be 29 and 16 Å^3 for protons and electrons, respectively, in the short-hydrogen-bond GFPs. The value given by Zundel (462 Å^3) was an order of magnitude larger than ours due to his overestimation of transition dipoles and underestimation of tunneling frequencies [88], but his calculation nevertheless captured the idea of the decrease in proton polarizabilities when the potential is biased either by external fields or proton affinity differences [104]. From his *ab initio* simulation on the Zundel cation $H_5O_2^+$, the proton polarizability is around 50 Å^3 instead and closer to our

estimation here [102]. One might question why the quantum estimation of the proton polarizability is two orders of magnitude more than its classical counterpart. The quantum model not only considers the proton to be delocalized but tunneling through the barrier also allows the "population" in one well to trickle into the other as the field biases the potential and the proton effectively travels a longer distance rather than staying within each well.

By definition, one issue from the crude two-level approximation is that contributions from higher excited states are omitted, especially if the corresponding transition dipole moments are non-negligible. Instead of evaluating the summation term by term in Equation S18, we can approximate the denominator by an average energy gap ΔE (the Unsöld approximation [105]):

$$\alpha_{PT} \approx \frac{2e^2}{\Delta E} \sum_{n \neq 0} \langle 0|x|n \rangle \langle n|x|0 \rangle = \frac{2e^2}{\Delta E} [\langle 0|x^2|0 \rangle - \langle 0|x|0 \rangle^2] = \frac{2e^2}{\Delta E} \langle (\Delta x)^2 \rangle_0$$

(S24)

where the second to last equality is obtained by recognizing the completeness relation $\sum_n |n \rangle \langle n| = 1$. Notice that this relates the spread of ground-state wavefunction $\langle (\Delta x)^2 \rangle_0$, which effectively measures the delocalization, to the proton polarizability. If ΔE is substituted with the tunneling splitting (i.e., the ground- and excited-state difference energy), the result imposes an upper bound for α_{PT}. Approximating ΔE itself, however, is rarely mentioned. A compromising method is utilizing the Thomas–Reiche–Kuhn sum rule, derived from potentials that are independent of momenta (e.g., without magnetic fields) [106][107]:

$$\sum_n (E_n - E_0)|\langle n|x|0 \rangle|^2 = \sum_{n \neq 0} (E_n - E_0)|\langle n|x|0 \rangle|^2 = \frac{\hbar^2}{2m}$$

(S25)

If we invoke the Unsöld approximation again, we obtain:

$$\frac{1}{\Delta E} \approx \frac{2m}{\hbar^2} \sum_{n \neq 0} |\langle n|x|0 \rangle|^2 = \frac{2m}{\hbar^2} \langle (\Delta x)^2 \rangle_0$$

(S26)

By combining Equations S24 and S26, we finally arrive at

$$\alpha_{PT} \approx \frac{4me^2}{\hbar^2} \langle (\Delta x)^2 \rangle_0^2$$

<div align="right">(S27)</div>

which can be shown to be exact for harmonic oscillators and accurate within an order of magnitude for other systems. Here one can clearly observe a direct connection between delocalization and polarizability: the more delocalized a proton, the higher its polarizability (a familiar concept for electron polarization). It might be tempting to conclude from Equation S27 that the polarizability is directly proportional to the particle mass without considering the mass effect on delocalization. To explicitly see the mass effect, we apply perturbation theory on a harmonic oscillator with an additional cubic term $V(x) = \frac{1}{2}kx^2 + \gamma x^3$ to model the anharmonicity and calculate its corresponding polarizability to the lowest order term in γ:

$$\alpha_{PT} \approx \frac{e^2}{k} \left(1 + N\gamma^2 \frac{\hbar}{m^{\frac{1}{2}}k^{\frac{5}{2}}} \right)$$

<div align="right">(S28)</div>

where N is a numerical factor that is irrelevant to our analysis. One can see that the correction from anharmonicity is quantum in nature (i.e., linear in $\hbar$) and inversely proportional to the square root of mass, which is consistent with the polarizability being lower when the hydrogen bond is deuterated.

S9   Unit Conversion

Energy:

1 D · MV/cm = 16.7 cm^{-1} = 0.0477 kcal/mol = k_B · 24.0 K = 0.312 pK_a unit at 77 K

Dipole moment:

1 D = 0.208 e · Å = 16.7 cm^{-1}/(MV/cm)

Polarizability:

1 Å^3 = 1.112 × 10^{-40} C · m^2/V = 1.99 × 10^{-4} D^2/cm^{-1} = 3.31 × 10^{-3} D/(MV/cm)

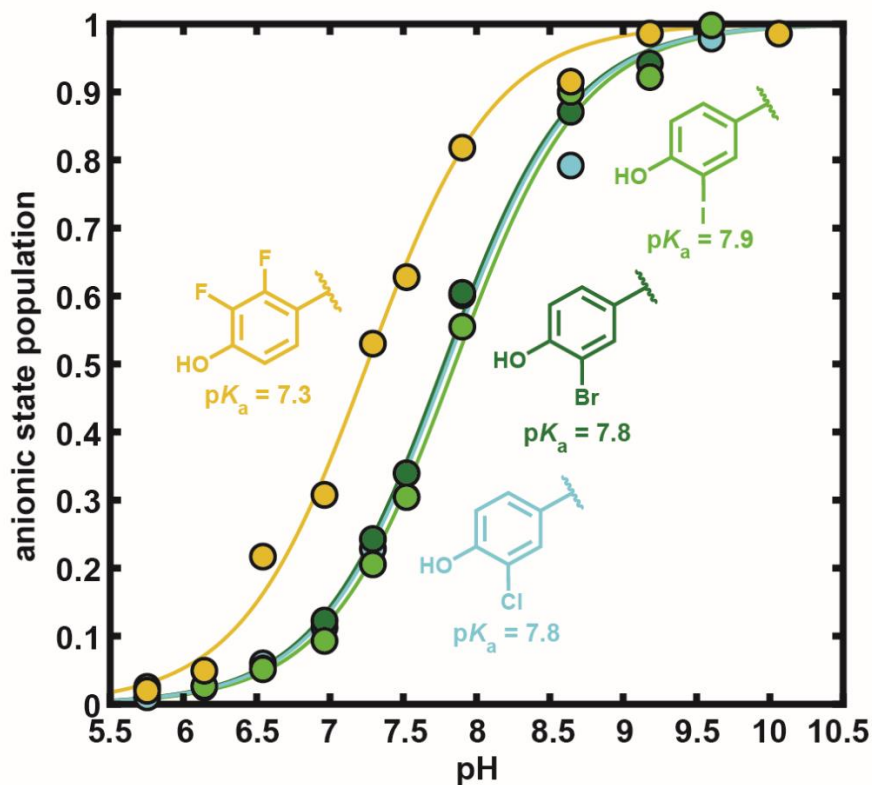= 5.54 × 10^{-2} cm^{-1}/(MV/cm)^2

Figure S24. pH titration of substituted chromophores in ih:GFP S65T/H148D under the denaturing condition (6M guanidium chloride), through which the intrinsic pK_a's are obtained and approximated as pK_α's (Tables 1 and 2). Only the data from the newly introduced variants in this work are included in this figure, except for those of the monochlorinated variant, which are shown as a control. For the titration curves determined from the rest of the variant series, please consult Figure S3 in [1]. The procedure is detailed in the Supporting Information of [1]. In short, the maximum absorbance A of the anionic state (Table S5) is monitored as a function of the solution's pH, and is subsequently fit to and normalized with the sigmoidal function $A(pH) = \frac{a+b\cdot10^{-(pH-pK_a)}}{1+10^{-(pH-pK_a)}}$ to extract the corresponding pK_a and the anionic state population (vertical axis). Note that the trend in the pK_a values mirrors those from the tyrosine analogues [28].
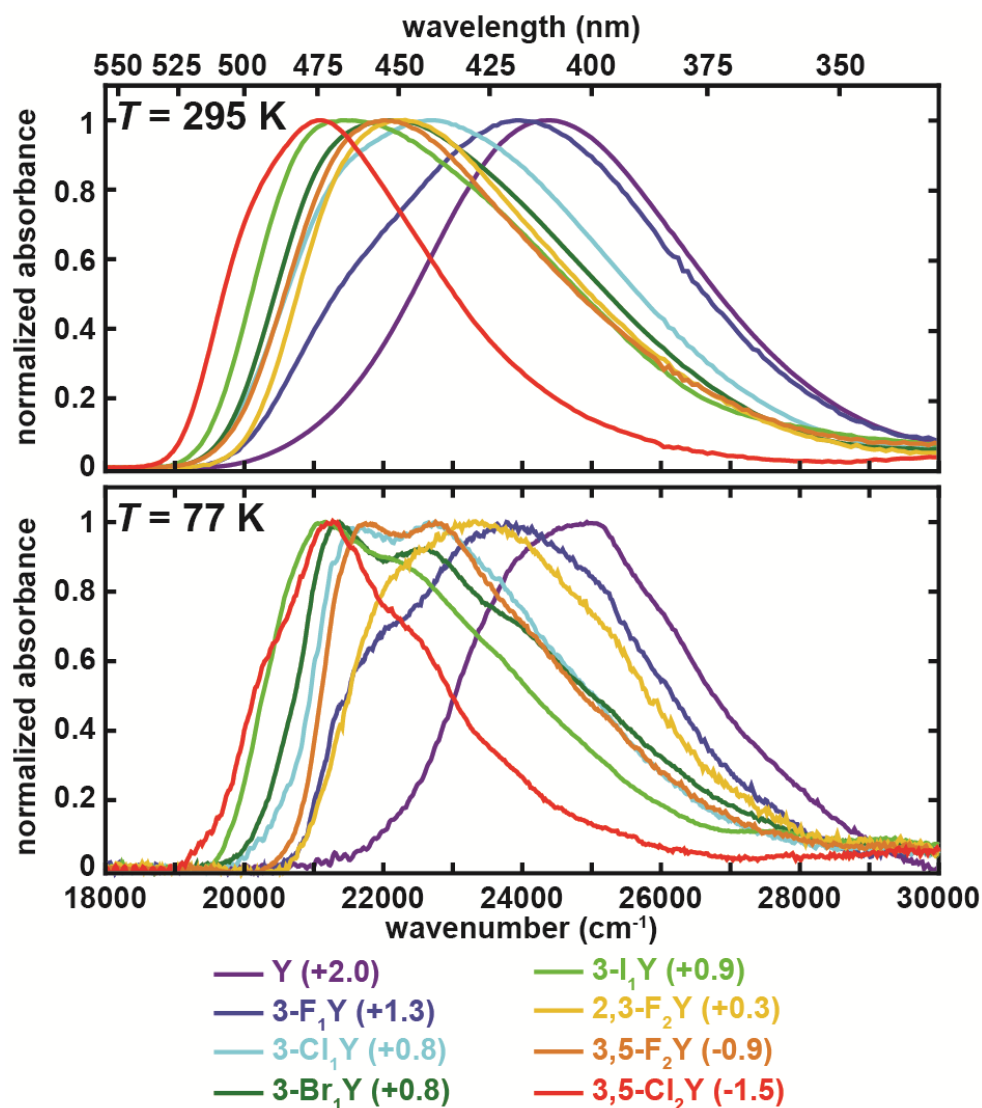
Figure S25. Room-temperature (top) and 77 K (bottom) UV–vis absorption spectra of short-hydrogen-bond GFP variants. The legend is denoted based on the identities of residue 66. The rainbow color scheme is the same as Figure 1 and assigned according to the order of the short hydrogen bond's ΔpK_α (in parentheses). While the room-temperature absorption maxima roughly follow the trend of ΔpK_α as observed in Figure 3 of [1], except for chromophores with heavy halogens, the correlation is more scrambled at 77 K, suggesting a nontrivial contribution from the electronic effect of the substituents. This effect can be better seen after spectral deconvolution (Figure S9).
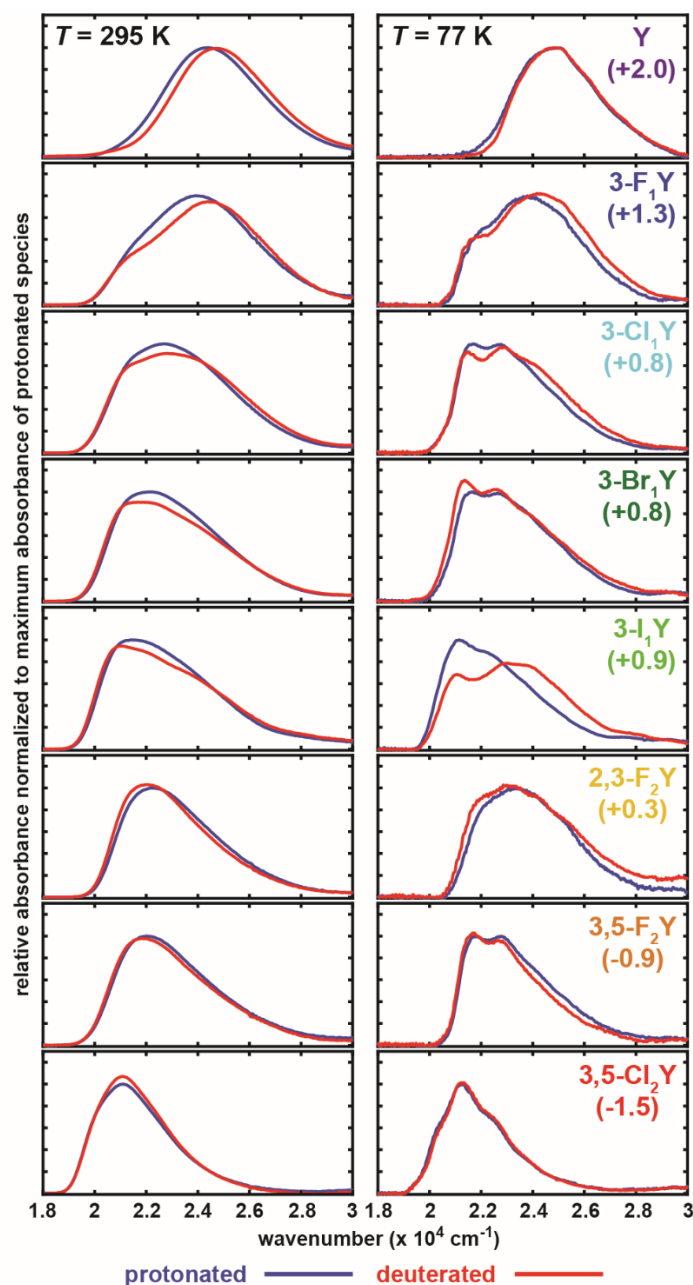
Figure S26. UV–vis absorption spectra of protonated (blue traces) and deuterated (red traces) short-hydrogen-bond GFP variants at room temperature (left panels) and 77 K (right panels). Every spectrum is normalized to the maximum absorption of the protonated counterpart. Each row is labeled with the identity of residue 66. Numbers in the parentheses are the corresponding ΔpK_a's of the short hydrogen bond in the variants. Most 77 K spectra mirror the expected SIE behavior observed in the room-temperature spectra upon deuteration, except for the mono-iodinated species, suggesting that the combination of the bulky iodine, deuteration, and low temperature could destabilize the short hydrogen bond.

S11   Supplementary Tables

Table S10. Absorption maxima and Stark tuning rates for ih:GFP S65T H148D variants and PYP mutants at 77 K. An abridged version can be found in Table 2. Values from protonated and deuterated proteins are separated by a slash within each entry. For most of the PYP mutants the deuterated data are not determined (N.D.) except for wild-type constructs. For constructs whose $|\Delta pK_a| \geq 1.5$, only one population exists within the short hydrogen bond(s), so the absorption maxima can be readily obtained and the data for the other population are unavailable (N/A). For those in which both populations exist, the absorption maxima are determined by deconvolution through Stark analysis (Section S5, Figure S9). The spectral isotope effect (SIE) is defined by the deuterated absorption maximum minus the protonated counterpart and shown within a parenthesis with its sign emphasized by red or blue. The error for the absorption maximum is ~ 100 cm^{-1}, so any SIE with a similar magnitude should be treated as negligible within error. Note that deuteration is carried out through buffer exchange, and thus all exchangeable protons within the experimental timescale (within a week), including most of the amide protons, are replaced by deuterons [108]. In other words, the chromophore is not the only deuterated moiety.

species	A-like A state protonated/deuterated		B-like A state protonated/deuterated	
	absorption maximum (and SIE) in cm^{-1}	Stark tuning rate (D)	absorption maximum (and SIE) in cm^{-1}	Stark tuning rate (D)
ih:GFP S65T H148D variants				
Y66	24930/24880 (-50)	13.6/13.0	N/A	
globally incorporated 3-F$_1$Y	24220/24710 (+490)	20.2/12.3	21740/21620 (-120)	15.3/12.7
Y66(3-Cl$_1$Y)	23420/23740 (+320)	17.7/17.1	21460/21350 (-110)	14.5/13.6
Y66(3-Br$_1$Y)	23540/23440 (-100)	15.8/16.5	21410/21220 (-190)	13.5/13.2
Y66(3-I$_1$Y)	22550/23360 (+810)	23.7/18.9	20770/20820 (+150)	10.9/10.8
Y66(2,3-F$_2$Y)	24070/24190 (+120)	15.6/14.9	21990/21860 (-130)	13.5/11.8
Y66(3,5-F$_2$Y)	23470/23470 (0)	24.1/15.8	21570/21570 (0)	17.0/15.9
Y66(3,5-Cl$_2$Y)	N/A		21570/21600 (+30)	19.0/17.1
PYP mutants				
HhPYP	N/A		22540/22450 (-90)	18.6/18.0

HhPYP E46Q		21860/N.D.	15.2/N.D.
HhPYP Y42F		21830/N.D.	14.9/N.D.
HhPYP Y42F/E46Q		20940/N.D.	10.5/N.D.
SrPYP		22970/22880 (-90)	14.5/14.0
SrPYP E46Q		22300/N.D.	15.4

Table S11. B-state 0-0 positions and Stark tuning rates of the ih:GFP S65T H148 variants for comparison to isolate the electronic effect from individual substituents. The values are extracted from Table S14 in [3].

variants	0-0 position (cm^{-1})	Stark tuning rate (D)
ih:GFP S65T	20589	10.2
ih:GFP S65T Y66(3-F_1Y)	20768	10.8
ih:GFP S65T Y66(3-Cl_1Y)	20288	10.1
ih:GFP S65T Y66(3-Br_1Y)	20105	8.8
ih:GFP S65T Y66(3-I_1Y)	19794	8.3
ih:GFP S65T Y66(2,3-F_2Y)	21044	10.6
ih:GFP S65T Y66(3,5-F_2Y)	20807	12.2
ih:GFP S65T Y66(3,5-Cl_2Y)	20313	12.6

S12   References

[1]   Oltrogge, L. M.; Boxer, S. G. Short hydrogen bonds and proton delocalization in green fluorescent protein (GFP). *ACS Cent. Sci.* **2015**, *1*, 148–156.

[2]   Romei, M. G.; Lin, C.-Y.; Mathews, I. I.; Boxer, S. G. Electrostatic control of photoisomerization pathways in proteins. *Science* **2020**, *367*, 76–79.

[3]   Lin, C.-Y.; Romei, M. G.; Oltrogge, L. M.; Mathews, I. I.; Boxer, S. G. Unified model for photophysical and electro-optical properties of green fluorescent proteins. *J. Am. Chem. Soc.* **2019**, *141*, 15250–15265.

[4]   Kent, K. P.; Oltrogge, L. M.; Boxer, S. G. Synthetic control of green fluorescent protein. *J. Am. Chem. Soc.* **2009**, *131*, 15988–15989.

[5]   Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

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

[7]   SRU_2224 photoactive yellow protein [*Salinibacter ruber* DSM 13855]: https://www.ncbi.nlm.nih.gov/gene/3851003 (accessed Aug 23, 2020).

[8]   Mongodin, E. F.; Nelson, K. E.; Daugherty, S.; DeBoy, R. T.; Wister, J.; Khouri, H.; Weidman, J.; Walsh, D. A.; Papke, R. T.; Perez, G. S.; Sharma, A. K.; Nesbø, C. L.; MacLeod, D.; Bapteste, E.; Doolittle, W. F.; Charlebois, R. L.; Legault, B.; Rodriguez-Valera, F. The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18147–18152.

[9]   Memmi, S.; Kyndt, J.; Meyer, T.; Devreese, B.; Cusanovich, M.; Van Beeumen, J. Photoactive yellow protein from the halophilic bacterium *Salinibacter ruber*. *Biochemistry* **2008**, *47*, 2014–2024.

[10]   Lin, C.-Y.; Both, J.; Do, K.; Boxer, S. G. Mechanism and bottlenecks in strand photodissociation of split green fluorescent proteins (GFPs). *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2146–E2155.

[11]   Thomson, B.; Both, J.; Wu, Y.; Parrish, R. M.; Martínez, T. J.; Boxer, S. G. Perturbation of short hydrogen bonds in photoactive yellow protein via noncanonical amino acid incorporation. *J. Phys. Chem. B* **2019**, *123*, 4844–4849.

[12]   Kaledhonkar, S. Structural Dynamics of Photoactive Yellow Protein. Ph.D. Dissertation, Oklahoma State University, Stillwater, OK, 2013.

[13]   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.

[14]   Minnihan, E. C.; Young, D. D.; Schultz, P. G.; Stubbe, J. Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.* **2011**, *133*, 15942–15945.

[15]   Hirel, P.-H.; Schmitter, J.-M.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8247–8251.

[16]   Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroh, J. G. Spontaneous α-N-6-phosphogluconoylation of a "His Tag" in *Escherichia coli*: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **1999**, *267*, 169–184.

[17]   Oltrogge, L. M.; Wang, Q.; Boxer, S. G. Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry* **2014**, *53*, 5947–5957.

[18]   Large-scale overexpression and purification of photoactive yellow protein (PYP) from *E. halophila* in *E. coli*: https://people.uwm.edu/smarius/files/2017/01/SAHIL_Protocol_PYP_0410-1hljf2r.pdf (accessed Aug 23, 2020).

[19]   Genick, U. K.; Devanathan, S.; Meyer, T. E.; Canestrelli, I. L.; Williams, E.; Cusanovich, M. A.; Tollin, G.; Getzoff, E. D. Active site mutants implicate key residues for control of color and light cycle kinetics of photoactive yellow protein. *Biochemistry*, **1997**, *36*, 8–14.

[20] Premvardhan, L. L.; Buda, F.; van der Horst, M. A.; Lührs, D. C.; Hellingwerf, K. J.; van Grondelle, R. Impact of photon absorption on the electronic properties of *p*-coumaric acid derivatives of the photoactive yellow protein chromophore. *J. Phys. Chem. B* **2004**, *108*, 5138–5148.

[21] van der Horst, M. A.; Arents, J. C.; Kort, R.; Hellingwerf, K. J. Binding, tuning and mechanical function of the 4-hydroxy-cinnamic acid chromophore in photoactive yellow protein. *Photochem. Photobiol. Sci.* **2007**, *6*, 571–579.

[22] Glasoe, P. K.; Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **1960**, *64*, 188–190.

[23] Bublitz, G. U.; Boxer, S. G. Stark spectroscopy: Applications in chemistry, biology, and materials science. *Annu. Rev. Phys. Chem.* **1997**, *48*, 213–242.

[24] Andrews, S. S.; Boxer, S. G. A liquid nitrogen immersion cryostat for optical measurements. *Rev. Sci. Inst.* **2000**, *71*, 3567–3569.

[25] Premvardhan, L. L.; van der Horst, M. A.; Hellingwerf, K. J.; van Grondelle, R. Stark spectroscopy on photoactive yellow protein, E46Q, and a nonisomerizing derivative, probes photo-induced charge motion. *Biophys. J.* **2003**, *84*, 3226-3239.

[26] Pauszek, R. F.; Stanley, R. J. A "How-To" Guide to the Stark Spectroscopy of Flavins and Flavoproteins. In *Methods in Molecular Biology (Methods and Protocols) Volume 1146: Flavins and Flavoproteins*; Weber, S., Schleicher, E., Eds.; Humana Press: New York, 2014; pp 443–466.

[27] Ward, W. W. Properties of the Coelenterate Green-Fluorescent Proteins. In *Bioluminescence and Chemiluminescence*; DeLuca, M. A.; McElroy, W. D., Eds.; Academic Press, 1981; pp 235–242.

[28] Seyedsayamdost, M. R.; Yee, C. S.; Stubbe, J. Site-specific incorporation of fluorotyrosines into the R2 subunit of *E. coli* ribonucleotide reductase by expressed protein ligation. *Nat. Protoc.* **2007**, *2*, 1225–1235.

[29] Using the SSRL automated mounting (SAM) system: http://smb.slac.stanford.edu/users_guide/manual/Using_SSRL_Automated_Mounti.html (accessed Aug 23, 2020).

[30] Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of

[31]   Kabsch, W. XDS. *Acta Cryst.* **2010**, *D66*, 125–132.

[32]   Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* **2010**, *D66*, 133–144.

[33]   Using the *autoxds* script: http://smb.slac.stanford.edu/facilities/software/xds/#autoxds_script (accessed Aug 23, 2020).

[34]   Adams, P. D.; Afonine, P. V.; Bunkóczi, 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.; Zwart, P. H. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **2010**, *D66*, 213–221.

[35]   Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Cryst.* **2004**, *D60*, 2126–2132.

[36]   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.

[37]   Wu, Y.; Boxer, S. G. A critical test of the electrostatic contribution to catalysis with noncanonical amino acids in ketosteroid isomerase. *J. Am. Chem. Soc.* **2016**, *138*, 11890–11895.

[38]   Chang, J.; Romei, M. G.; Boxer, S. G. Structural evidence of photoisomerization pathways in fluorescent proteins. *J. Am. Chem. Soc.* **2019**, *141*, 15504–15508.

[39]   Pal, P. P.; Bae, J. H.; Azim, M. K.; Hess, P.; Friedrich, R.; Huber, R.; Moroder, L.; Budisa, N. Structural and spectral response of *Aequorea victoria* green fluorescent protein to chromophore fluorination. *Biochemistry* **2005**, *44*, 3663–3672.

[40]   Hoesl, M. G.; Merkel, L.; Budisa, N. Synthetic Biology of Autofluorescent Proteins. In *Fluorescent Proteins I: From Understanding to Design*; Jung, G., Ed.; Springer Series on Fluorescence 11; Springer-Verlag: Heidelberg, Germany, 2012; pp 99–130.

[41]   Khan, F.; Kuprov, I.; Craggs, T. D.; Hore, P. J.; Jackson, S. E. ^{19}F NMR studies of the native and denatured states of green fluorescent protein. *J. Am. Chem. Soc.* **2006**, *128*, 10729–10737.

[42]   Xiao, G.; Parsons, J. F.; Tesh, K.; Armstrong, R. N.; Gilliland, G. L. Conformational changes in the crystal structure of rat glutathione transferase M1-1 with global substitution of 3-fluorotyrosine for tyrosine. *J. Mol. Biol.* **1998**, *281*, 323–339.

[43]   Sušac, L.; Eddy, M. T.; Didenko, T.; Stevens, R. C.; Wüthrich, K. A$_{2A}$ adenosine receptor functional states characterized by ^{19}F-NMR. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12733–12738.

[44]   Danielson, M. A.; Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 163–195.

[45]   Stoner-Ma, D.; Jaye, A. A.; Ronayne, K. L.; Nappa, J.; Meech, S. R.; Tonge, P. J. An alternate proton acceptor for excited-state proton transfer in green fluorescent protein: Rewiring GFP. *J. Am. Chem. Soc.* **2008**, *130*, 1227–1235.

[46]   Royant, A.; Noirclerc-Savoye, M. Stabilizing role of glutamic acid 222 in the structure of enhanced green fluorescent protein. *J. Struct. Biol.* **2011**, *174*, 385–390.

[47]   Arpino, J. A. J.; Rizkallah, P. J.; Jones, D. D. Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PLoS One* **2012**, *7*, e47132.

[48]   Shibazaki, C.; Shimizu, R.; Kagotani, Y.; Ostermann, A.; Schrader, T. E.; Adachi, M. Direct observation of the protonation states in the mutant green fluorescent protein. *J. Phys. Chem. Lett.* **2020**, *11*, 492–496.

[49]   Ghosh, A.; Isbaner, S.; Veiga-Gutiérrez, M.; Gregor, I.; Enderlein, J.; Karedla, N. Quantifying microsecond transition times using fluorescence lifetime correlation spectroscopy. *J. Phys. Chem. Lett.* **2017**, *8*, 6022–6028.

[50]   Takaba, K.; Tai, Y.; Eki, H.; Dao, H.-A.; Hanazono, Y.; Hasegawa, K.; Miki, K.; Takeda, K. Subatomic resolution X-ray structures of green fluorescent protein. *IUCrJ* **2019**, *6*, 387–400.

[51]   Craggs, T. D. Green fluorescent protein: Structure, folding and chromophore maturation. *Chem. Soc. Rev.* **2009**, *38*, 2865–2875.

[52]   Jain, R. K.; Ranganathan, R. Local complexity of amino acid interactions in a protein core. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 111–116.

[53]   Maddalo, S. L.; Zimmer, M. The role of the protein matrix in green fluorescent protein fluorescence. *Photochem. Photobiol.* **2006**, *82*, 367–372.

[54]   McKenzie, R. H.; Bekker, C.; Athokpam, B.; Ramesh, S. G. Effect of quantum nuclear motion on hydrogen bonding. *J. Chem. Phys.* **2014**, *140*, 174508.

[55]   Gurusaran, M.; Shankar, M.; Nagarajan, R.; Helliwell, J. R.; Sekar, K. Do we see what we should see? Describing non-covalent interactions in protein structures including precision. *IUCrJ* **2014**, *1*, 74–81.

[56]   A web server to calculate the Diffraction Precision Index (DPI): http://cluster.physics.iisc.ernet.in/dpi/ (accessed Aug 23, 2020).

[57]   Kumar, K. S. D.; Gurusaran, M.; Satheesh, S. N.; Radha, P.; Pavithra, S.; Tharshan, K. P. S. T.; Helliwell, J.R.; Sekar, K. *Online_DPI*: A web server to calculate the diffraction precision index for a protein structure. *J. Appl. Cryst.* **2015**, *48*, 939–942.

[58]   Sigala, P. A.; Ruben, E. A.; Liu, C. W.; Piccoli, P. M. B.; Hohenstein, E. G.; Martínez, T. J.; Schultz, A. J.; Herschlag, D. Determination of hydrogen bond structure in water versus aprotic environments to test the relationship between length and stability. *J. Am. Chem. Soc.* **2015**, *137*, 5730–5740.

[59]   Sigala, P. A.; Tsuchida, M. A.; Herschlag, D. Hydrogen bond dynamics in the active site of photoactive yellow protein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9232–9237.

[60]   Hanoian, P.; Sigala, P. A.; Herschlag, D.; Hammes-Schiffer, S. Hydrogen bonding in the active site of ketosteroid isomerase: Electronic inductive effects and hydrogen bond coupling. *Biochemistry* **2010**, *49*, 10339–10348.

[61] Pinney, M. M.; Natarajan, A.; Yabukarski, F.; Sanchez, D. M.; Liu, F.; Liang, R.; Doukov, T.; Schwans, J. P.; Martínez, T. J.; Herschlag, D. Structural coupling throughout the active site hydrogen bond networks of ketosteroid isomerase and photoactive yellow protein. *J. Am. Chem. Soc.* **2018**, *140*, 9827–9843.

[62] Herschlag, D.; Pinney, M. M. Hydrogen bonds: Simple after all? *Biochemistry* **2018**, *57*, 3338–3352.

[63] Haddad, K. C.; Sudmeier, J. L.; Bachovchin, D. A.; Bachovchin, W. W. α-Lytic protease can exist in two separately stable conformations with different His[57] mobilities and catalytic activities. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1006–1011.

[64] Pflugrath, J. W. Practical macromolecular cryocrystallography. *Acta Cryst.* **2015**, *F71*, 622–642.

[65] Boxer, S. G. Stark realities. *J. Phys. Chem. B* **2009**, *113*, 2972–2983.

[66] Treynor, T. P. Non-Classical Stark Effects of Photosynthetic Reaction Centers. Ph.D. Dissertation; Stanford University, Stanford, CA, 2003; pp. 193–207.

[67] Liptay, W. Optical absorption of molecules in liquid solutions in an applied external electric field (electrochromism). *Ber. Bunsenges. Phys. Chem.* **1976**, *80*, 207–217.

[68] Lin, C.-Y.; Boxer, S. G. Mechanism of color and photoacidity tuning for the protonated green fluorescent protein chromophore. *J. Am. Chem. Soc.* **2020**, *142*, 11032–11041.

[69] Shi, X.; Abbyad, P.; Shu, X.; Kallio, K.; Kanchanawong, P.; Childs, W.; Remington, S. J.; Boxer, S. G. Ultrafast excited-state dynamics in the green fluorescent protein variant S65T/H148D. 2. Unusual photophysical properties. *Biochemistry* **2007**, *46*, 12014–12025.

[70] Hutchison, C. D. M.; van Thor, J. J. Populations and coherence in femtosecond time resolved X-ray crystallography of the photoactive yellow protein. *Int. Rev. Phys. Chem.* **2017**, *36*, 117–143.

[71] 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.

[72] Getzoff, E. D.; Gutwin, K. N.; Genick, U. K. Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation. *Nat. Struct. Biol.* **2003**, *10*, 663–668.

[73] 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.

[74] McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **1958**, *28*, 430–431.

[75] Bell, A. F.; He, X.; Wachter, R. M.; Tonge, P. J. Probing the ground state structure of the green fluorescent protein chromophore using Raman spectroscopy. *Biochemistry* **2000**, *39*, 4423–4431.

[76] Schellenberg, P.; Johnson, E.; Esposito, A. P.; Reid, P. J.; Parson, W. W. Resonance Raman scattering by the green fluorescent protein and an analogue of its chromophore. *J. Phys. Chem. B* **2001**, *105*, 5316–5322.

[77] Esposito, A. P.; Schellenberg, P.; Parson, W. W.; Reid, P. J. Vibrational spectroscopy and mode assignments for an analog of the green fluorescent protein chromophore. *J. Mol. Struct.* **2001**, *569*, 25–41.

[78] He, X.; Bell, A. F.; Tonge, P. J. Isotopic labeling and normal-mode analysis of a model green fluorescent protein chromophore. *J. Phys. Chem. B* **2002**, *106*, 6056–6066.

[79] Tozzini, V.; Nifosì, R. Ab initio molecular dynamics of the green fluorescent protein (GFP) chromophore: An insight into the photoinduced dynamics of green fluorescent protein. *J. Phys. Chem. B* **2001**, *105*, 5797–5803.

[80] Fang, C.; Frontiera, R. R.; Tran, R.; Mathies, R. A. Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy. *Nature* **2009**, *462*, 200–204.

[81] Noguchi, T.; Inoue, Y.; Tang, X.-S. Structural coupling between the oxygen-evolving Mn cluster and a tyrosine residue in photosystem II as revealed by Fourier transform infrared spectroscopy. *Biochemistry* **1997**, *36*, 14705–14711.

[82] Peuker, S.; Andersson, H.; Gustavsson, E.; Maiti, K. S.; Kania, R.; Karim, A.; Niebling, S.; Pedersen, A.; Erdelyi, M.; Westenhoff, S. Efficient isotope editing of

proteins for site-directed vibrational spectroscopy. *J. Am. Chem. Soc.* **2016**, *138*, 2312–2318.

[83] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09* (Gaussian, Inc., Wallingford CT, 2009).

[84] Ayala, I.; Perry, J. J. P.; Szczepanski, J.; Tainer, J. A.; Vala, M. T.; Nick, H. S.; Silverman, D. N. Hydrogen bonding in human manganese superoxide dismutase containing 3-fluorotyrosine. *Biophys. J.* **2005**, *89*, 4171–4179.

[85] Reid, P. J.; Loftus, C.; Beeson, C. C. Evaluating the potential of fluorinated tyrosines as spectroscopic probes of local protein environments: A UV resonance Raman study. *Biochemistry* **2003**, *42*, 2441–2448.

[86] Perrin, C. L.; Nielson, J. B. "Strong" hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.* **1997**, *48*, 511–544.

[87] Perrin, C. L. Are short, low-barrier hydrogen bonds unusually strong? *Acc. Chem. Res.* **2010**, *43*, 1550–1557.

[88] Weidemann, E. G.; Zundel, G. Field-dependent mechanism of anomalous proton conductivity and the polarizability of hydrogen bonds with tunneling protons. *Z. Naturforsch.* **1970**, *25a*, 627–634.

[89] Janoschek, R.; Weidemann, E. G.; Pfeiffer, H.; Zundel, G. Extremely high polarizability of hydrogen bonds. *J. Am. Chem. Soc.* **1972**, *94*, 2387–2896.

[90]   Hasegawa, M.; Noda, H. Distribution of hydrogen bond angles in molecular crystals. *Nature* **1975**, 254, 212.

[91]   Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76.

[92]   Grabowski, S. J. What is the covalency of hydrogen bonding? *Chem. Rev.* **2011**, *111*, 2597–2625.

[93]   Scheiner, S. Extraction of the Principles of Proton Transfers by Ab Initio Methods. In *NATO Advanced Science Institutes Series B Volume 291: Proton Transfer in Hydrogen-Bonded Systems*; Bountis, T., Ed.; Springer: New York, 1992; pp 29–47.

[94]   Meot-Ner, M. Update 1 of: Strong ionic hydrogen bonds. *Chem. Rev.* **2012**, *112*, PR22–PR103.

[95]   Pang, X.-F.; Feng, Y.-P. *Quantum Mechanics in Nonlinear Systems*, 1st ed.; World Scientific Publishing: Hackensack, New Jersey, 2005.

[96]   Park, Y.; Kang, H.; Kang, H. Electric field effect on condensed-phase molecular systems. IX. Control of proton displacement in matrix-isolated hydrogen chloride-water complexes. *J. Phys. Chem. C* **2020**, *124*, 1129–1134.

[97]   Drobizhev, M.; Callis, P. R.; Nifosì, R.; Wicks, G.; Stoltzfus, C.; Barnett, L.; Hughes, T. E.; Sullivan, P.; Rebane, A. Long- and short-range electrostatic fields in GFP mutants: Implications for spectral tuning. *Sci. Rep.* **2015**, *5*, 13223.

[98]   Griffiths, D. J. *Introduction to Electrodynamics*, 3rd ed.; Prentice-Hall: Upper Saddle River, New Jersey, 1999.

[99]   Böttcher, C. J. F.; van Belle, O. C.; Bordewijk, P.; Rip, A. *Theory of Electric Polarization Vol 1*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 1973.

[100] This is also seen in the context of magnetic anisotropy of NMR, where induced magnetic field by ring currents can either shield or deshield the nuclear spins depending on their positions with respect to the current source, and thus reflected in their chemical shifts [101]. One difference between the NMR ring current and the short hydrogen bond system is that the probe spins align along the externally applied magnetic field in competition with thermal fluctuations and thus are more or less isotropic with respect to the current source in an isotropic sample, while the

chromophore's difference dipole moment is always pointed at a specific direction with respect to the hydrogen bond, rendering the conventional NMR shielding cone concept inapplicable.

[101] McConnell, H. M. Theory of nuclear magnetic shielding in molecules: I. Long-range dipolar shielding of protons. *J. Chem. Phys.* **1957**, *27*, 226–229.

[102] Zundel, G. Hydrogen Bonds with Large Polarizability and Proton Transfer Processes in Electrochemistry and Biology. In *Advances in Chemical Physics, Volume 111*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons, Inc.: New York, 2000; pp 1–217.

[103] Craig. D. P.; Thirunamachandran, T. *Molecular Quantum Electrodynamics: An Introduction to Radiation-Molecule Interactions*, 1st ed.; Dover: Mineola, New York, 1998.

[104] Eckert, M.; Zundel, G. Proton polarizability, dipole moment, and proton transitions of an AH···B $\rightleftharpoons$ A$^-$···H$^+$B proton-transfer hydrogen bond as a function of an external electrical field: An ab initio SCF treatment. *J. Phys. Chem.* **1987**, *91*, 5170–5177.

[105] Norman, P.; Ruud, K.; Saue, T. *Principles and Practices of Molecular Properties: Theory, Modeling and Simulations*, 1st ed.; John Wiley & Sons: West Sussex, UK, 2018.

[106] Sakurai, J. J.; Tuan, S. F. *Modern Quantum Mechanics*, revised ed.; Addison-Wesley Publishing Company: Reading, Massachusetts, 1994.

[107] Butcher, P. N.; Cotter, D. *The Elements of Nonlinear Optics*, 1st ed.; Cambridge University Press: Cambridge, UK, 1990.

[108] Hsu, S.-T. D.; Blaser, G.; Jackson, S. E. The folding, stability and conformational dynamics of β-barrel fluorescent proteins. *Chem. Soc. Rev.* **2009**, *38*, 2951–2965.