Supporting Information for

"Mechanism of Color and Photoacidity Tuning for the Protonated Green Fluorescent Protein Chromophore"

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

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

*Correspondence to: chiyunl@stanford.edu; 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]. Point mutations were made using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's protocol. The residue numbering scheme follows GFPs without circular permutation. Supercharged GFP -30 gene in pET-29 was generously provided by the David Liu Lab at Harvard University [3] and used without further modification.

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 of interest, while the internal helix is denoted ih. Loop refers to a sacrificial loop with proteolytic cleavage sites. GFP refers to the remainder of the protein. A strike through an element indicates that the element has been removed. Synthetic elements are underlined. A dot is used to indicate a noncovalent interaction. For example, s10(203F) · ~~s10:loop~~:GFP denotes a synthetic β-s10 carrying the mutation T203F noncovalently bound to circularly permuted GFP with its original N-terminal s10 and loop removed.

Table S1. GFP constructs in this study, forming a subset of those characterized in our previous work [2]. The following entries were colored based on their parent circular permutants. The parent proteins for the colors orange, pink, and green are s10:loop:GFP, ih:GFP, and ih:loop:GFP, respectively. Red letters denote non-wild-type amino acids, and superscript "mat" indicates an internal helix with a matured chromophore. 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		s4	s7	s10	s11
	65	66	96	148	203	222
s10:loop:GFP	S	Y	R	H	T	E
s10:loop:GFP T203V	S	Y	R	H	V	E
s10:loop:GFP T203Y	S	Y	R	H	Y	E

ih:GFP T203(3-OMeY)	S	Y	R	H	3-OCH_3Y	E
s10:loop:GFP S65T	T	Y	R	H	T	E
ih:GFP S65T	T	Y	R	H	T	E
supercharged -30	T	Y	R	H	T	E
ih^{mat}(65T) · ~~ih:loop~~:GFP R96M	T	Y	M	H	T	E
s10:loop:GFP S65T T203V	T	Y	R	H	V	E
s10:loop:GFP S65T T203H	T	Y	R	H	H	E
s10(203F) · ~~s10:loop~~:GFP S65T	T	Y	R	H	F	E
s10(203(4-F_1F)) · ~~s10:loop~~:GFP S65T	T	Y	R	H	4-F_1F	E
s10(203(F_5F)) · ~~s10:loop~~:GFP S65T	T	Y	R	H	F_5F	E
s10(203(4-NH_2F)) · ~~s10:loop~~:GFP S65T	T	Y	R	H	4-NH_2F	E
ih:GFP S65T T203(3-OMeY)	T	Y	R	H	3-OCH_3Y	E
s10:loop:GFP S65T T203Y	T	Y	R	H	Y	E

DNA and Amino Acid Sequences

The sequences have been described in our previous work [2] in detail.

Synthetic Peptide Design

Peptides were designed to match native s10 of s10:loop:GFP and were synthesized by Elim Biopharmaceuticals.

s10(203F): LPDNHYLS**F**QTVLSKDPNE

s10(203(4-F_1F)): LPDNHYLS **(4-F_1F)** QTVLSKDPNE

s10(203(F_5F)): LPDNHYLS **(F_5F)** QTVLSKDPNE

s10(203(4-NH_2F)): LPDNHYLS **(4-NH_2F)** QTVLSKDPNE

Semi-synthetic Method for Split GFPs

The protocol has been described in our previous work [2] in detail, including the subsequent purification and characterization.

Sample Preparation for 77 K Absorption and Electronic Stark Spectroscopy

Glass forming solvents, such as ethanol or a 1:1 mixture of glycerol and aqueous buffer, are required for low-temperature electronic Stark spectroscopy experiments. The concentrated samples were mixed with an equal volume of glycerol (Fisher, CAS 56-81-

5) right before Stark measurements. 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 [2][5]. 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 [6] 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 of a lock-in amplifier (SR830; Stanford Research) with a frequency of 203 Hz, 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 – 3.0 kV, which amounts to a peak external field strength F_{ext} of 0.3 – 1.6

MV/cm given the sample thickness. The *X* and *Y* components of Stark signal Δ*I* were detected at the second harmonic of the applied field. The direct output voltage *I* was also simultaneously recorded. The Stark spectra were then obtained from the ratio: $\Delta A = \frac{2\sqrt{2}}{\ln 10}\frac{\Delta I}{I}$ as a function of the scanning wavelengths [5]. 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 (or ethanol). 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.

Stark Spectroscopy Data Analysis

All Stark spectra Δ*A* are shown with their corresponding absorbance *A* normalized. Δ*A* are also scaled to 1 MV/cm with χ = 90° according to their proportionality to $(F_{ext})^2$, where χ is the angle between probe light polarization and field direction. The Stark spectra (as functions of wavenumbers $\bar{\nu}$) 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 [5]:

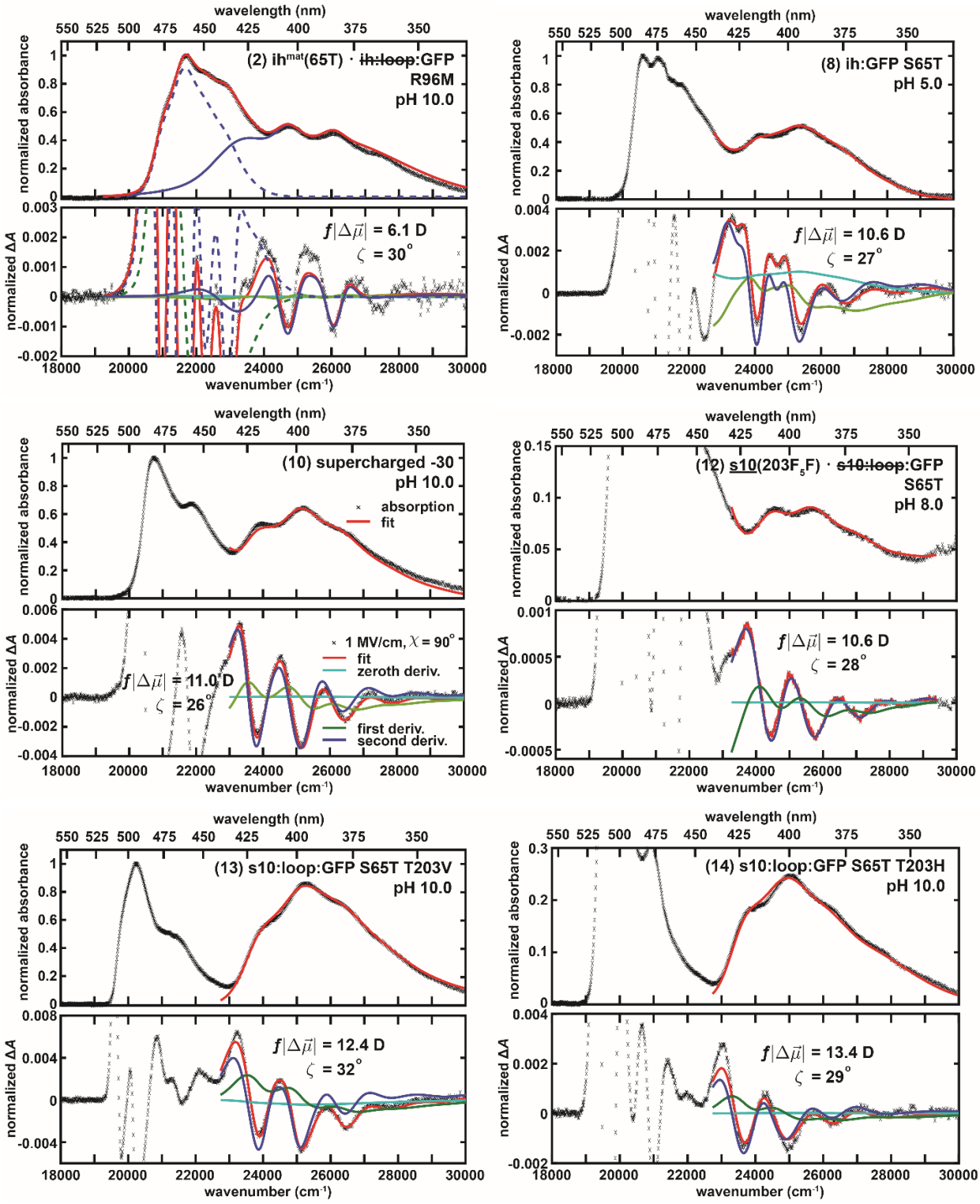
$$\begin{aligned}\Delta A(\bar{\nu}, F_{ext}) &= A(\bar{\nu}, F_{ext}) - A(\bar{\nu}, F_{ext} = 0) \\ &= (F_{ext})^2 \left[A_\chi A(\bar{\nu}) + \frac{B_\chi}{15hc}\bar{\nu}\frac{d}{d\bar{\nu}}\left(\frac{A(\bar{\nu})}{\bar{\nu}}\right) + \frac{C_\chi}{30h^2c^2}\bar{\nu}\frac{d^2}{d\bar{\nu}^2}\left(\frac{A(\bar{\nu})}{\bar{\nu}}\right)\right] \end{aligned} \quad \text{(S1)}$$

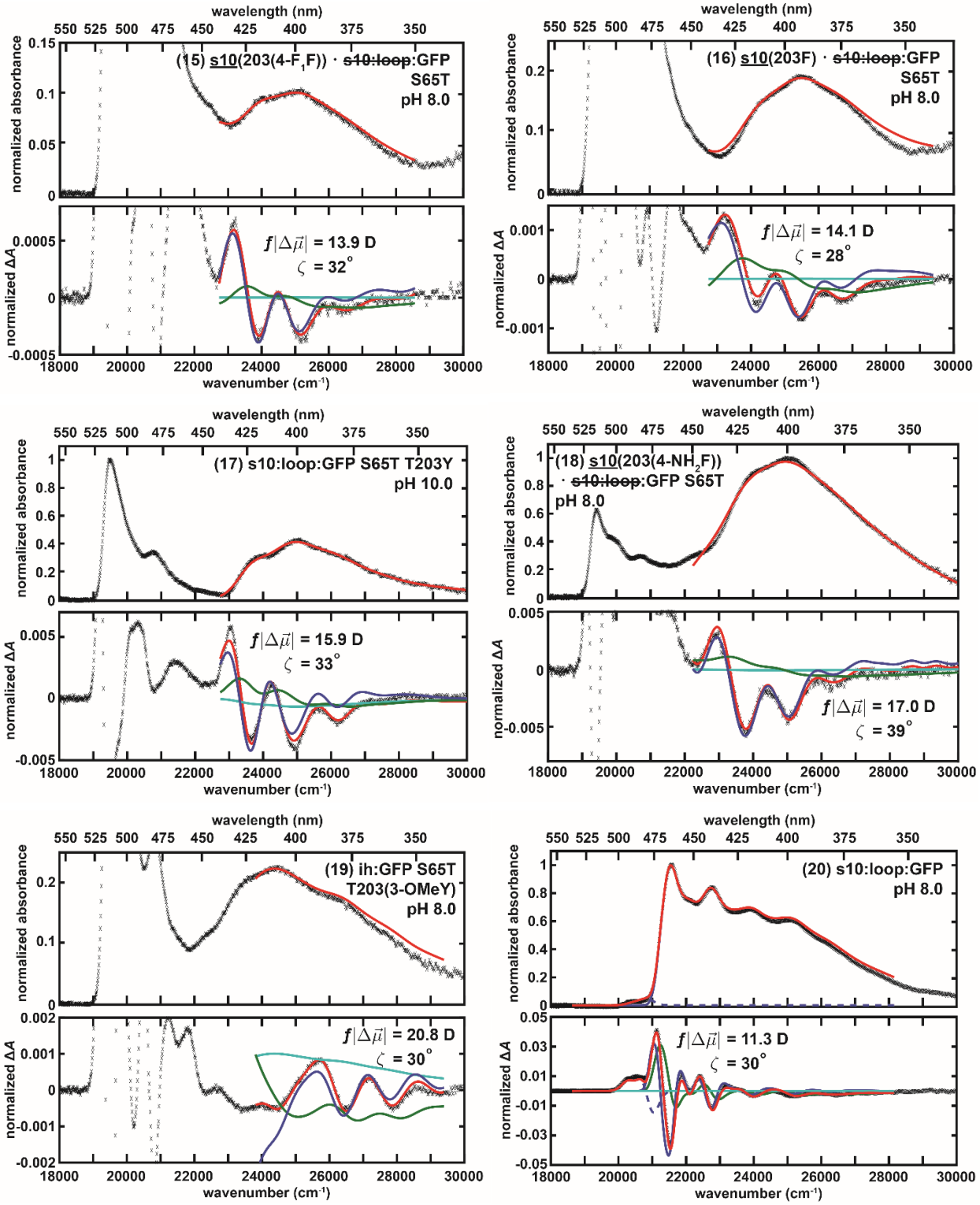
and

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

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 was processed by the MATLAB code kindly provided by Professor Robert Stanley at Temple University [7]. Simultaneous fitting of Δ*A* and *A* at χ = 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 the dominant contribution from $\Delta\mu_{app}$, no attempt was done 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 [5] (see also Section S6 in [2]).

S3 Stark Spectra and Fitting of GFP Mutants and HBDI




wavelength (nm)
550 525 500 475 450 425 400 375 350
normalized absorbance
(15) s10(203(4-F1F)) · s10:loop:GFP S65T pH 8.0
(16) s10(203F) · s10:loop:GFP S65T pH 8.0
(17) s10:loop:GFP S65T T203Y pH 10.0
(18) s10(203(4-NH2F)) · s10:loop:GFP S65T pH 8.0
(19) ih:GFP S65T T203(3-OMeY) pH 8.0
(20) s10:loop:GFP pH 8.0
normalized ΔA
f|Δμ| = 13.9 D
ζ = 32°
f|Δμ| = 14.1 D
ζ = 28°
f|Δμ| = 15.9 D
ζ = 33°
f|Δμ| = 17.0 D
ζ = 39°
f|Δμ| = 20.8 D
ζ = 30°
f|Δμ| = 11.3 D
ζ = 30°
wavenumber (cm-1)
18000 20000 22000 24000 26000 28000 30000

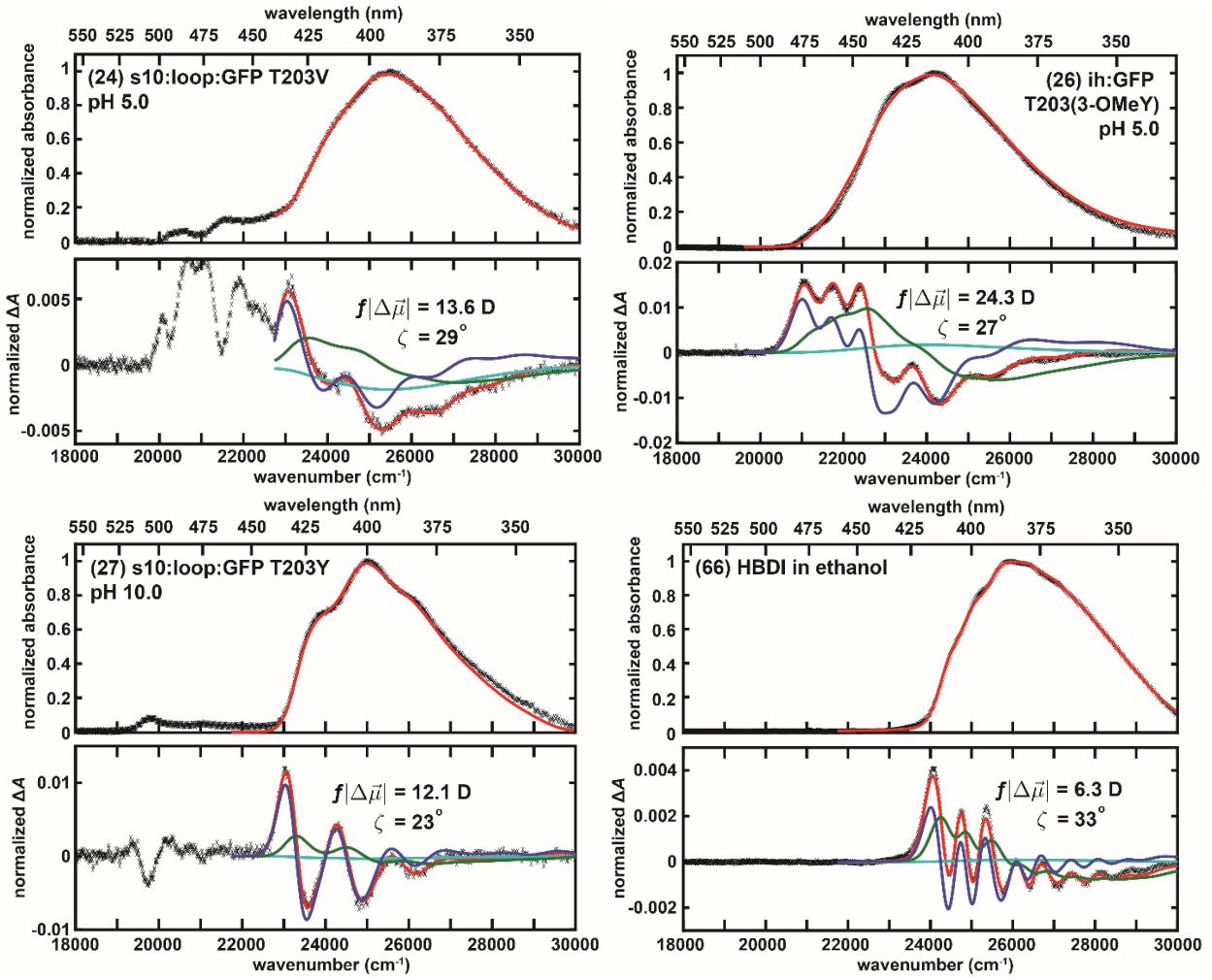


Figure S1. The classical sum-of-derivative analysis for 77 K UV–Vis absorption and Stark spectra, which are dominated by second-derivative lineshapes (Figure S2). The corresponding B-state analysis can be found in Figure S10 of [2]. The panels are numbered and listed according to Table S4. The absorption spectra are normalized to 1 at the maximum absorbance in 18000 – 30000 cm^{-1} and are magnified if the corresponding normalized absorbance for A state is less than 0.3. 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 panel 10, is consistent throughout the figure. Solid lines represent the band of interest (the neutral state), for which the measured Stark tuning rate (±5%) and the ζ angle (±5°) is noted, while dashed lines in panel 2 represent other absorption bands that require simultaneous fitting to extract out electro-optic parameters from the higher energy band. In most cases, in which either only one dominant band or sufficient spectral separation between bands is observed, assigning one set of electro-optic parameters with occasionally a limited spectral range is preferred, even when two bands with distinct origins (such as A and B states) share similar Stark parameters (panel 20). An extra set of electro-optic parameters is only considered when the fit demands such a scenario due to significant overlap of two bands with different Stark tuning rates (panel 2). Due to less sharp vibronic features of A band compared to B band, the A-state

Stark signals are much less significant than the B-state counterparts and tend to be buried when both bands overlap.

S4 The Three-Form Coupling Model for Protonated GFP Chromophore

In this section, we present a detailed derivation for the relation between Stark tuning rate and absorption maximum for the neutral GFP chromophore from the three-form model (Figure 4C), which is based on Olsen's multi-configurational calculation [8]. Even though the model contains three forms, it is effectively a two-form model since the GS form is decoupled from the other two forms by assumption due to the relatively large energy gap. By setting the energy difference between the LE and GS forms to be $\Delta\bar{\nu}_{LE}$ ($\equiv \bar{\nu}_{LE} - \bar{\nu}_{GS}$) and that between the CT and LE forms to be $\Delta\bar{\nu}$ ($\equiv \bar{\nu}_{CT} - \bar{\nu}_{LE}$) (see Figure 4C in main text), we can write down the potential energy matrix to describe the coupling between the LE and CT forms:

$$V = \begin{pmatrix} \Delta\bar{\nu}_{LE} & {V_0}' \\ {V_0}' & \Delta\bar{\nu}_{LE} + \Delta\bar{\nu} \end{pmatrix} \quad \text{(S3)}$$

where V_0' is the associated electronic coupling and the diagonal element are the energies for the LE and CT forms from the GS form. We do not include vibrational degrees of freedom in contrast to our previous B-state color tuning model [2]; this is justified below. The relative energies for the resulting two excited adiabatic states with respect to the GS form can be solved by diagonalizing Equation S3 and are:

$$\Delta\bar{\nu}_{1,2} = \Delta\bar{\nu}_{LE} + \frac{\Delta\bar{\nu}}{2} \pm \frac{\sqrt{\Delta\bar{\nu}^2 + 4{V_0}'^2}}{2} \quad \text{(S4)}$$

the lower of which is the absorption energy from the S_0 to S_1 state:

$$\bar{\nu}_{abs} = \Delta\bar{\nu}_{LE} + \frac{\Delta\bar{\nu}}{2} - \frac{\sqrt{\Delta\bar{\nu}^2 + 4{V_0}'^2}}{2} \quad \text{(S5)}$$

Therefore, when $\Delta\bar{\nu}$ is much larger than $2V_0$', no appreciable mixing between the two states is expected, and $\bar{\nu}_{abs}$ approaches $\Delta\bar{\nu}_{LE}$ as the bluest possible absorption for the neutral GFP chromophore. As $\Delta\bar{\nu}$ becomes smaller or more negative, the coupling lowers the LE form's energy and red shifts the absorption (Figure 5). It does not matter whether we use the 0–0 electronic excitation energy or the absorption maximum (which is technically 0–1 excitation energy) as $\bar{\nu}_{abs}$, since it should only differ by a constant frequency (Figure 3D), which is absorbed by $\Delta\bar{\nu}_{LE}$ and justifies why we neglect the energy offset. After mixing between the CT and LE forms, the S_1 state inherits some dipolar character from the CT form, which carries a dipole moment of μ_{CT}. Since the GS form has

a much smaller dipole moment than the CT form, the Stark tuning rate $\Delta\mu$ associated with the absorption becomes:

$$\Delta\mu = \left(\frac{1}{2} - \frac{\Delta\bar{\nu}}{2\sqrt{\Delta\bar{\nu}^2 + 4{V_0}'^2}}\right)\mu_{CT} \tag{S6}$$

(cf. Equation 2 in the main text). When $\Delta\bar{\nu}$ is much larger than $2V_0'$, the Stark tuning rate becomes zero because the S_1 state exists purely as the neutral LE form (Figure 5). The fact that zero Stark tuning rate corresponds to the bluest possible absorption for the neutral chromophore is not a mere coincidence, as discussed below. Since $\Delta\bar{\nu}$ is not an observable and assumed to be the only quantity that can be modulated by the chromophore's environment through electrostatic interactions, as developed in detail for the driving force in the Marcus–Hush model for the B state [2], we can combine Equations S5 and S6 by eliminating $\Delta\bar{\nu}$ to obtain the correlation between $\bar{\nu}_{abs}$ and $\Delta\mu$:

$$\boxed{\bar{\nu}_{abs} = \Delta\bar{\nu}_{LE} - {V_0}' \frac{2\frac{\Delta\mu}{\mu_{CT}}}{\sqrt{1-\left(2\frac{\Delta\mu}{\mu_{CT}}-1\right)^2}}} \tag{S7}$$

Note that we include the local field factor f in the corresponding equation in the main text (Equation 1) to emphasize the fact that all experimentally determined dipole moments are associated with f due to environmental polarization in response to the externally applied field.

In addition to allowing us to extract parameters for a specific model from the correlation plot of absorption maxima and Stark tuning rates for various mutants, we advocate that the correlation plot is a useful strategy for understanding a chromophore's electrostatic color tuning behavior even if the underlying model is unknown. The Stark tuning rate is the linear response of the chromophore's color to the effective electric field $\vec{F}$ experienced by the chromophore:

$$\Delta\mu = \left|\frac{\partial\bar{\nu}_{abs}}{\partial\vec{F}}\right| \tag{S8}$$

which is a *derivative* of $\bar{\nu}_{abs}$ with respect to $\vec{F}$. By plotting $\bar{\nu}_{abs}$ against $\Delta\mu$ for one-dimensional systems, we can in principle capture the behavior of the function $\bar{\nu}_{abs}(\vec{F})$ without knowing the function itself explicitly. This is in fact a widely used strategy in different contexts. For instance, in classical Hamiltonian mechanics, it is useful to plot the

velocity (which is the time *derivative* of the position) of a particle against its position to illustrate the phase space trajectory and analyze the dynamical information without directly solving the equation of motion [9]. Similarly, when encountering a nonlinear ordinary differential equation (ODE) of a function *y*(*t*) that is not analytically solvable, a plot with $\frac{dy}{dt}$ against *y* itself, known as the phase portrait, is an invaluable tool to understand the dynamical behavior of the ODE, especially in terms of fixed points, flows, and limit cycles [10]. In the case of electrostatic color tuning, *y* is $\bar{\nu}_{abs}$, while the external field plays the role of time. Therefore, the correlation curve between $\bar{\nu}_{abs}$ and $\Delta\mu$ can serve as a calibration curve for electrostatic color tuning for a specific chromophore.

Furthermore, fixed points, defined as *y* at which $\frac{dy}{dt} = 0$, correspond to the extremal values of *y*(*t*). Analogously, even if the function $\bar{\nu}_{abs}(\vec{F})$ that encodes all information of electrostatic color tuning of a given chromophore is unknown, we can still figure out where the reddest or bluest possible absorption maximum is by knowing where $\Delta\mu = 0$ occurs, which is in fact the fixed point(s). Whether it is the reddest or bluest can be readily determined from the correlation plot itself: if $\Delta\mu$ is a decreasing or increasing function of $\bar{\nu}_{abs}$, the fixed point corresponds to the bluest or reddest possible absorption maximum as for the neutral and anionic GFP chromophore [2], respectively.

The aforementioned analysis is only strictly applicable to one-dimensional systems (and two-form systems from the perspective of the minimal Hilbert space for diabatic states) since the direction of $\Delta\vec{\mu}$ stays the same (or at most 180° flipped) across mutants. The direct consequence of being one-dimensional is that the correlation plots are monotonic and environmental effects from any combinations of mutations can be completely captured by the relative energy between the two underlying forms (i.e. driving forces $\Delta\bar{\nu}$). Therefore, the *magnitude* of $\Delta\vec{\mu}$ from each mutant is sufficient to encompass all electro-optic properties and can be readily measured through electronic Stark spectroscopy. The correlation plot for red fluorescent proteins, on the other hand, fails to show the simple monotonic trend [11], suggesting the necessity of incorporating at least one more coupled form likely due to the additional acylimine tail from the GFP chromophore [12][13]. Since at least three energetically close forms with noncolinear charge distributions are involved, the *direction* of $\Delta\vec{\mu}$ likely changes across mutants and

is sensitive to the direction of the electric field exerted by the protein environment. In other words, the effects of mutations can no longer be encoded by only one driving force, so presumably one has to determine both *x* and *y* components of $\Delta\vec{\mu}$ for each mutant (assuming the *z* axis is normal to the chromophore plane) in order to conduct the same analysis. That is to say, a correlation plot with data points of ($\Delta\mu_x$, $\Delta\mu_y$, $\bar{\nu}_{abs}$) is required to evaluate multiple driving forces and electronic couplings between the forms [14] and understand the electrostatic color tuning behavior for non-one-dimensional systems, such as porphyrins, chlorins, and bacteriochlorins [15][16].

S5 Analogy with Special Pair in Bacterial Photosynthetic Reaction Center

As a side note of interest in the spectroscopy of photobiological systems, it is intriguing to draw an analogy between the GFP chromophore and special pair P (i.e. bacteriochlorophyll dimer $BChl_2$) in the *Rhodobacter sphaeroides* photosynthetic reaction center with regards to color tuning and charge transfer. In particular, the B-state GFP chromophore resembles the oxidized P (P^+) and both chromophores can both be treated as mixed-valence systems since the electron and hole are strongly delocalized within the chromophores, respectively [17]. On the other hand, owing to the large energetic asymmetry between the two interacting moieties within the chromophore, the A state is more similar to a heterodimer D, which is formed by selectively replacing one of the BChls with a bacteriopheophytin (BPhe): both have a CT form ($phenol^+imidazolinone^-$ or $BChl^+BPhe^-$) close in energy to an LE-type form (or an exciton state) that can be excited from the ground state [18]. An obvious difference, however, between the special pair (P^+ and D) and the GFP chromophore (A and B states) is that the electronic coupling between the two rings of the latter is an order of magnitude larger than that for the former due to the direct π conjugation. For the charged states P^+ and B, V_0 is approximately 1000 cm^{-1} [17] and 9500 cm^{-1} [2], respectively; for the neutral states D and A, V_0' is approximately 500 cm^{-1} [18] and 5000 cm^{-1}, respectively. Following this analogy, it might be possible to understand why the redder A bands tend to be broader through the Fano theory [19], in which a broadened absorption band is caused by a decrease in excited lifetime of the LE form by virtue of coupling to the CT form (homogeneous broadening) [18]. It could also be explained by more significant inhomogeneous broadening from the environmental electric field distribution experienced by the chromophore with a larger Stark tuning rate, as explicated by Drobizhev *et al.* [20] and our previous work [2].

S6 Supplementary Figures

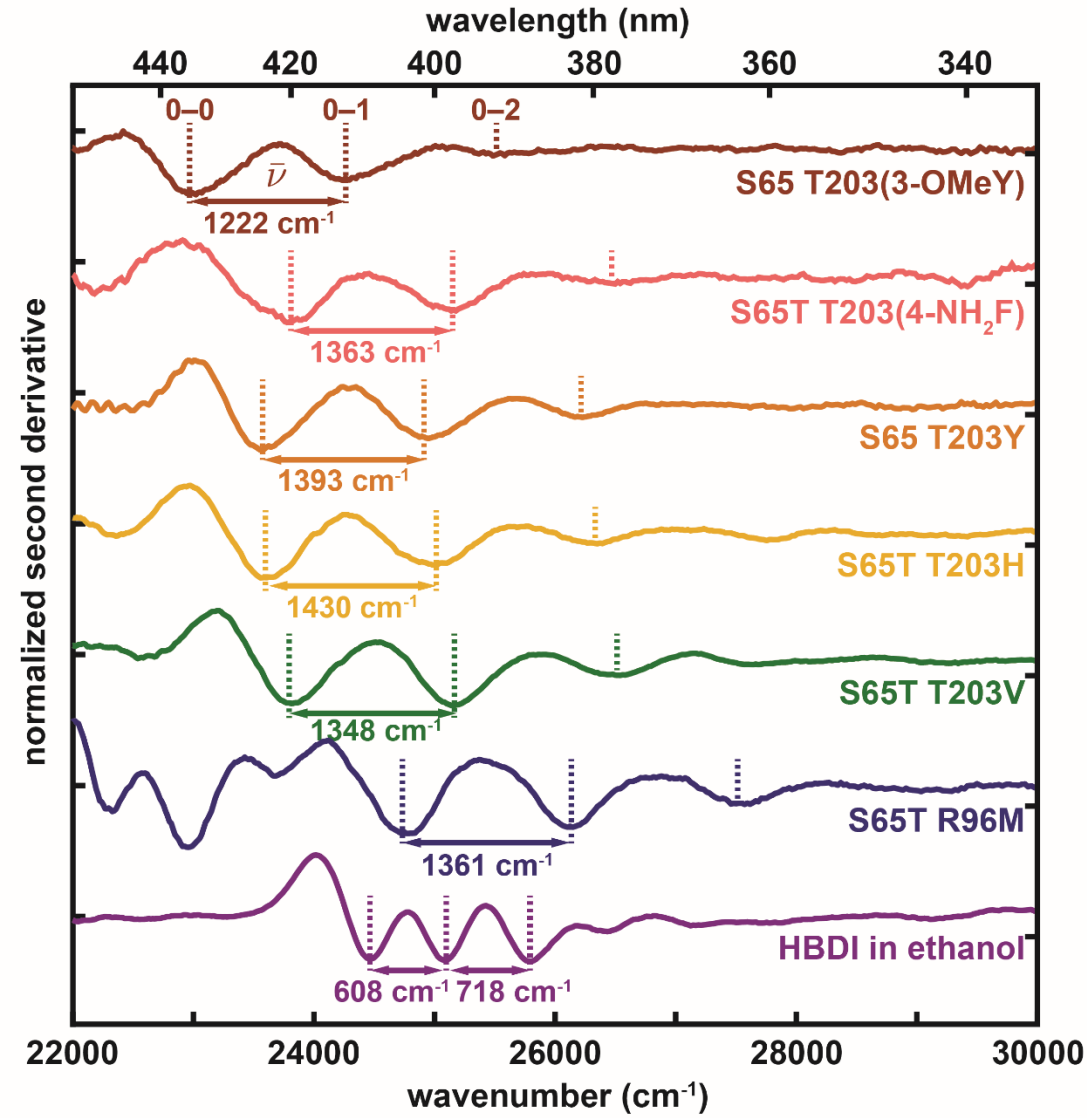


Figure S2. Method for estimating the difference between 0–0 and 0–1 transitions using the second derivative of 77 K absorption spectra (Figures 3A and 3B) of protonated GFP mutants and HBDI. 0–0 and 0–1 energies are assigned from the negative peak positions; this difference corresponds to the BLA and other possible normal modes with similar frequencies for GFP mutants (Figure 3D). The color coding follows that of Figures 3A and 3B. For HBDI in ethanol, the frequency spacings are about half of those from GFP mutants, suggesting that the 0–2 feature assigned in the figure could also correspond to a 0–1 transition as in protein mutants, while the 0–1 feature results from another vibrational mode that is more vibronically coupled in ethanol than in the GFP environment. Interestingly, this additional feature is not seen for the anionic HBDI in ethanol (Figure S24 in [2]). The redder features from S65T R96M GFP correspond to the B and I bands [2], which are also observed in the absorption spectrum (Figure 3B).

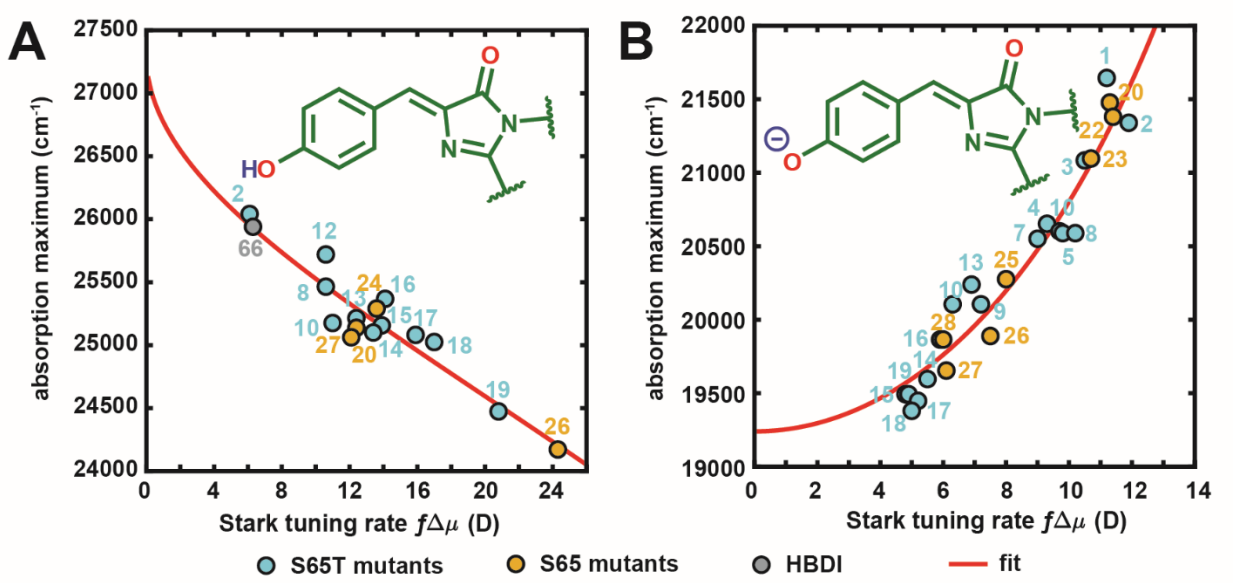


Figure S3. Correlation between the absorption maximum and Stark tuning rate for GFP mutants and HBDI in the (A) A and (B) B states at 77 K. This figure is reproduced from Figure 4 to include numerical labels defined in Table S4.

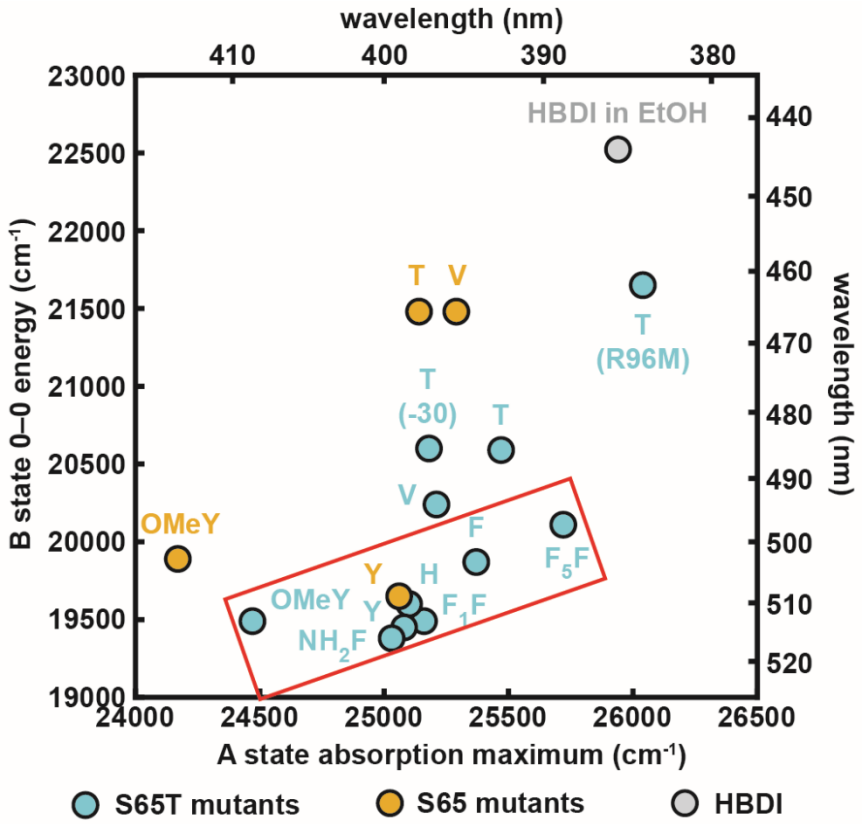


Figure S4. Correlation plot between the B state 0–0 energy (which is approximately the absorption maximum [2]) and A state absorption maximum for GFP mutants and HBDI in ethanol (Table S3). The data points are labeled with the identities of residue 203, and additional mutations are noted in the parentheses. The overall trend shows a positive correlation of these two quantities, but is not even roughly linear for the following reasons. First, from the color tuning mechanisms discussed in Figure 4, it is easier to tune the A state for redder mutants while the B state is more tunable when blue. Second, there are significant hydrogen bonding network rearrangement in the vicinity of the chromophore between A and B states, especially for mutants with S65, T203, or V203 [21] or HBDI in ethanol. This causes an extra stabilization in the P form of the B state, leading to a bluer B state absorption than expected and deviating from the trend of S65T π–π stacking series (red box), for which no significant structural rearrangement should occur. Given the nonlinearity of the absorption maximum to electrostatics (e.g. Equation S5), it is better to study the correlation in terms of driving forces (Figure S5).

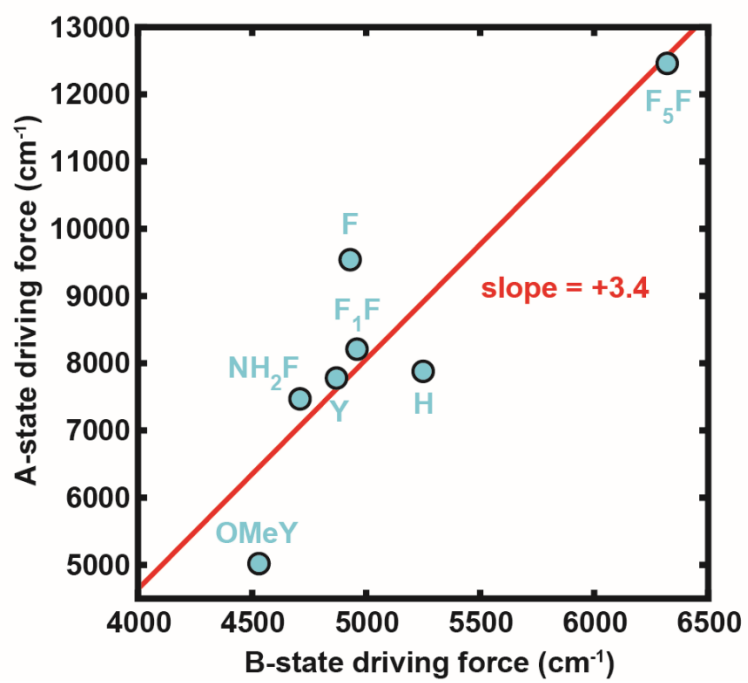


Figure S5. Correlation plot between the A-state and B-state driving forces from π–π stacking S65T mutants (aromatic residue at position 203), whose values are highlighted in blue in Table S5. The data points are labeled with the identities of residue 203 for the mutants. This series is chosen since π–π stacking allows for minimal structural rearrangement between the environments of the protonated and deprotonated chromophore, rendering the comparison between A and B states valid within these mutants. A linear fit (red) has a slope of 3.4, which agrees well to the ratio of the diabatic dipole moments between the two protonation states ($\frac{\mu_{CT}}{\Delta\mu_{CT}} \approx 3.1$ from Figure 4). From this plot, we can estimate the absorption maximum for the protonated GFP chromophore in vacuum, given the experimentally determined gas phase value for the deprotonated chromophore (479 nm from Nielsen and colleagues [22]). The latter corresponds to a B-state driving force of 8100 cm^{-1}. By extrapolation, the A-state chromophore in vacuum should possess a driving force of 18700 cm^{-1}, which translates to an absorption maximum at 382 nm according to Equation S5 using the parameters determined from Figure 4. The exact experimental value is still under debate. The Nielsen group determined it to be 370 ± 5 nm [23], which is close to our estimation; however, 340 ± 5 nm was obtained using multiphoton ionization from Greenwood *et al.* [24]. In our opinion, the latter seems unlikely as it is bluer than the bluest possible absorption $\Delta\bar{\nu}_{LE}$ (= 366 nm) obtained from Figure 4A, unless the chromophore geometry in the gas phase deviates substantially from that in the GFP environment.

S7 Supplementary Tables

Table S2. Summary of the second-derivative analysis of GFP mutants and the model chromophore HBDI in ethanol at 77 K in the protonated A state (Figures 3D and S2), presented in order of the 0–0 position). The numbering scheme follows Table S12 in [2]. Even though the difference between 0–1 and 0–0 energies for HBDI in ethanol is an apparent outlier, it is curiously half as much as a typical BLA vibrational frequency (~ 1320 cm^{-1}), as also evidenced by its 0–2 energy at 25800 cm^{-1} (Figure S2).

number	species	0–0 position (nm)	0–0 energy (cm^{-1})	0–1 energy (cm^{-1})	difference between 0–1 and 0–0 energies (cm^{-1})
S65T GFP mutants					
2	ih^{mat}(65T) · ~~ih:loop~~:GFP R96M	403.5	24783	26144	1361
12	s10(203F_5F) · ~~s10:loop~~:GFP S65T	410.4	24366	25700	1334
16	s10(203F) · ~~s10:loop~~:GFP S65T	413.1	24207	25465	1258
8	ih:GFP S65T	415.2	24085	25368	1283
15	s10(203(4-F_1F)) · ~~s10:loop~~:GFP S65T	417.0	23981	25233	1253
10	supercharged -30	420.0	23810	25214	1405
13	s10:loop:GFP S65T T203V	420.0	23810	25157	1348
18	s10(203(4-NH_2F)) · ~~s10:loop~~:GFP S65T	420.6	23776	25138	1363
17	s10:loop:GFP S65T T203Y	423.3	23624	24931	1308
14	s10:loop:GFP S65T T203H	424.5	23557	24988	1430
19	ih:GFP S65T T203(3-OMeY)	428.1	23359	24618	1259
S65 GFP mutants					
24	s10:loop:GFP T203V	413.1	24207	25484	1277
20	s10:loop:GFP	417.9	23929	25195	1266
27	s10:loop:GFP T203Y	424.5	23557	24950	1393
26	ih:GFP T203(3-OMeY)	434.4	23020	24242	1222
GFP model chromophore					

66	HBDI in ethanol	408.6	24474	25082	608

Table S3. A-state and B-state absorption maxima (the latter cited from Table S14 in [2]) for GFP mutants at 77 K, listed in the order of decreasing A-state absorption energy. The numbering scheme follows Table S12 in [2]. Mutants colored with shades of orange carry threonine at the 203 position, which assume different rotamers in two different protonation states (Figure 1 in the main text [21]).

number	species	A-state absorption maximum (nm)	A-state absorption maximum (cm^{-1})	B-state absorption maximum (nm)	B-state absorption maximum (cm^{-1})
S65T GFP mutants					
2	ih^{mat}(65T) · ~~ih:loop~~:GFP R96M	384.0	26040	462.0	21650
12	s10(203F_5F) · ~~s10:loop~~:GFP S65T	388.8	25720	497.4	20110
8	ih:GFP S65T	392.7	25470	485.7	20590
16	s10(203F) · ~~s10:loop~~:GFP S65T	394.2	25370	503.4	19870
13	s10:loop:GFP S65T T203V	396.6	25210	494.1	20240
10	supercharged -30	397.2	25180	485.4	20600
15	s10(203(4-F_1F)) · ~~s10:loop~~:GFP S65T	397.5	25160	513.0	19490
14	s10:loop:GFP S65T T203H	398.4	25100	510.3	19600
17	s10:loop:GFP S65T T203Y	398.7	25080	514.2	19450
18	s10(203(4-NH_2F)) · ~~s10:loop~~:GFP S65T	399.6	25030	516.0	19380
19	ih:GFP S65T T203(3-OMeY)	408.6	24470	513.0	19490
S65 GFP mutants					
24	s10:loop:GFP T203V	392.4	25290	465.6	21480
20	s10:loop:GFP	397.8	25140	465.6	21480
27	s10:loop:GFP T203Y	399.0	25060	508.8	19650
26	ih:GFP T203(3-OMeY)	413.7	24170	502.8	19890

Table S4. Absorption maxima and Stark tuning rates for A states of GFP mutants at 77 K listed in the order of decreasing absorption maximum. The numbering scheme is the same as Table S12 in [2]. For mutants where only B-state properties are measured (Figure S3), the corresponding reasons are given below. The energy gap between the CT and LE forms (denoted as $\Delta\bar{\nu}$) and degree of CT character in S_1 state is estimated using Equation S5 and S6, respectively. Note that for all mutants we have accessed, the CT form is consistently higher in energy than the LE form and none of the estimated CT degree exceeds 50%.

number	species	A-state absorption maximum (cm^{-1})	A-state Stark tuning rate (D)	energy gap between CT and LE $\Delta\bar{\nu}$ (cm^{-1})	degree of CT character in S_1 state
S65T GFP mutants					
2	ih^{mat}(65T) · ~~ih:loop~~:GFP R96M	26040	6.1	16360	6.7%
12	s10(203F_5F) · ~~s10:loop~~:GFP S65T	25720	10.6	12460	10.1%
8	ih:GFP S65T	25470	10.6	10250	13.2%
16	s10(203F) · ~~s10:loop~~:GFP S65T	25370	14.1	9540	14.4%
13	s10:loop:GFP S65T T203V	25210	12.4	8550	16.4%
10	supercharged -30	25180	11.0	8320	16.9%
15	s10(203(4-F_1F)) · ~~s10:loop~~:GFP S65T	25160	13.9	8210	17.2%
14	s10:loop:GFP S65T T203H	25100	13.4	7880	17.9%
17	s10:loop:GFP S65T T203Y	25080	15.9	7780	18.2%
18	s10(203(4-NH_2F)) · ~~s10:loop~~:GFP S65T	25030	17.0	7470	18.9%
19	ih:GFP S65T T203(3-OMeY)	24470	20.8	5020	26.5%
1	ih^{mat}(65T)·~~ih:loop~~:GFP R96E E222K	no A state observed [2]			
3	ih:GFP S65T E222Q				
4	s10:loop:GFP S65T				
5	supercharged +36 s10-				
7	supercharged +36				

9	ih:GFP S65T H148D	A state with a short hydrogen bond with D148 [25]			
S65 GFP mutants					
24	s10:loop:GFP T203V	25290	13.6	9030	15.4%
20	s10:loop:GFP	25140	11.3	8100	17.4%
27	s10:loop:GFP T203Y	25060	12.1	7670	18.4%
26	ih:GFP T203(3-OMeY)	24170	24.3	3960	30.6%
22	s10:loop:GFP E222Q	no A state observed [2]			
23	avGFP	significant overlap between A and B states [2]			
25	s10:loop:GFP T203V E222Q	no A state observed [2]			
28	s10:loop:GFP T203Y E222Q				
GFP model chromophore					
66	HBDI in ethanol	25940	6.3	14950	7.7%

Table S5. The energy gaps (or driving forces) between diabatic forms for the neutral and anionic GFP chromophores evaluated in the GFP environments. The latter is reproduced here from Table S13 in [2]. The numbering scheme follows Table S12 in [2]. Mutants with a S65T chromophore and an aromatic residue at the 203 position are highlighted in blue. Since within this set of mutants, the π–π interaction stays relatively rigid and S65T chromophore disallows hydrogen bonding network rearrangement upon chromophore protonation, the comparison between A-state and B-state values is meaningful unlike others (Figure S5). In contrast, there must be a significant change in solvation structures of ethanol when interacting with the neutral and anionic chromophores, so the comparison between the two corresponding driving forces is unrealistic.

number	species	energy gap between CT and LE for A state (cm^{-1})	energy gap between I and P for B state (cm^{-1})
S65T GFP mutants			
2	ih^{mat}(65T) · ~~ih:loop~~:GFP R96M	16360	10270
12	s10(203F_5F) · ~~s10:loop~~:GFP S65T	12460	6320
8	ih:GFP S65T	10250	7880
16	s10(203F) · ~~s10:loop~~:GFP S65T	9540	4930
13	s10:loop:GFP S65T T203V	8550	6170
10	supercharged -30	8320	7540
15	s10(203(4-F_1F)) · ~~s10:loop~~:GFP S65T	8210	4960
14	s10:loop:GFP S65T T203H	7880	5250
17	s10:loop:GFP S65T T203Y	7780	4870
18	s10(203(4-NH_2F)) · ~~s10:loop~~:GFP S65T	7470	4710
19	ih:GFP S65T T203(3-OMeY)	5020	4530
S65 GFP mutants			
24	s10:loop:GFP T203V	9030	7600
20	s10:loop:GFP	8100	9770
27	s10:loop:GFP T203Y	7670	5680
26	ih:GFP T203(3-OMeY)	3960	6340

GFP model chromophore			
66	HBDI in ethanol	14950	12200

S8 References

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

[2] 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.

[3] Zuris, J. A.; Thompson, D. B.; Shu, Y.; Guilinger, J. P.; Bessen, J. L.; Hu, J. H.; Maeder, M. L.; Joung, J. K.; Chen, Z.-Y.; Liu, D. R. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat. Biotechnol.* **2015**, *33*, 73-80.

[4] 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.

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

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

[7] 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.

[8] Olsen, S. Locally-excited (LE) versus charge-transfer (CT) excited state competition in a series of para-substituted neutral green fluorescent protein (GFP) chromophore models. *J. Phys. Chem. B* **2015**, *119*, 2566-2575.

[9] Goldstein, H.; Poole, C.; Safko, J. *Classical Mechanics*, 3rd ed.; Addison-Wesley: Boston, Massachusetts, 2002; pp 483-509.

[10] Strogatz, S. H. *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering*, 1st ed.; Perseus: Cambridge, Massachusetts, 1994.

[11] Drobizhev, M.; Tillo, S.; Makarov, N. S.; Hughes, T. E.; Rebane, A. Color hues in red fluorescent proteins are due to internal quadratic Stark effect. *J. Phys. Chem. B* **2009**, *113*, 12860-12864.

[12] Hasegawa, J.; Fujimoto, K. J.; Hakatsuji, H. Color tuning in photofunctional proteins. *ChemPhysChem* **2011**, *12*, 3106-3115.

[13] Drobizhev, M.; Hughes, T. E.; Stepanenko, Y.; Wnuk, P.; O'Donnell, K.; Scott, J. N.; Callis, P. R.; Mikhaylov, A.; Dokken, L.; Rebane, A. Primary role of the chromophore bond length alternation in reversible photoconversion of red fluorescence proteins. *Sci. Rep.* **2012**, *2*, 688.

[14] Heckmann, A.; Lambert, C. Organic mixed-valence compounds: a playground for electrons and holes. *Angew. Chem. Int. Ed.* **2012**, *51*, 326-392.

[15] Gouterman, M. Study of the effects of substitution on the absorption spectra of porphin. *J. Chem. Phys.* **1959**, *30*, 1139-1161.

[16] Gouterman, M. Spectra of porphyrins. *J. Mol. Spectrosc.* **1961**, *6*, 138-163.

[17] 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.

[18] Zhou, H.; Boxer, S. G. Charge resonance effects on electronic absorption line shapes: application to the heterodimer absorption of bacterial photosynthetic reaction centers. *J. Phys. Chem. B* **1997**, *101*, 5759-5766.

[19] Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **1961**, *124*, 1866-1878.

[20] 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.

[21] Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **2009**, *38*, 2865-2875.

[22] Nielsen, S. B.; Lapierre, A.; Andersen, J. U.; Pedersen, U. V.; Tomita, S.; Andersen, L. H. Absorption spectrum of the green fluorescent protein chromophore anion *in vacuo*. *Phys. Rev. Lett.* **2001**, *87*, 228102.

[23] Rajput, J.; Rahbek, D. B.; Andersen, L. H.; Rocha-Rinza, T.; Christiansen, O.; Bravaya, K. B.; Erokhin, A. V.; Bochenkova, A. V.; Solntsev, K. M.; Dong, J.; Kowalik, J.; Tolbert, L. M.; Petersen, M. Å.; Nielsen, M. B. Photoabsorption studies of neutral green fluorescent protein model chromophores *in vacuo*. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9996-10002.

[24] Greenwood, J. B.; Miles, J.; De Camillis, S.; Mulholland, P.; Zhang, L.; Parkes, M. A.; Hailes, H. C.; Fielding, H. H. Resonantly enhanced multiphoton ionization spectrum of the neutral green fluorescent protein chromophore. *J. Phys. Chem. Lett.* **2014**, *5*, 3588-3592.

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