

## Energetic basis and design of enzyme function demonstrated using GFP, an excited-state enzyme

Chi-Yun Lin<sup>†,§,||</sup>, Matthew G. Romei<sup>†,§,▽</sup>, Irimpan I. Mathews<sup>‡</sup>, Steven G. Boxer<sup>\*,†</sup>

<sup>†</sup>Department of Chemistry, Stanford University, Stanford, CA 94305, USA.

<sup>‡</sup>Stanford Synchrotron Radiation Lightsource, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

<sup>\*</sup>To whom correspondence may be addressed. Email: sboxer@stanford.edu.

<sup>§</sup>C.-Y.L. and M.G.R. contributed equally to this work.

<sup>||</sup>Present address: Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA.

<sup>▽</sup>Present address: Antibody Engineering, Genentech Inc., South San Francisco, CA 94080, USA.

### ABSTRACT

The last decades have witnessed an explosion of *de novo* protein designs with a remarkable range of scaffolds. It remains challenging, however, to design catalytic functions that are competitive with naturally occurring counterparts as well as biomimetic or non-biological catalysts. Although directed evolution often offers efficient solutions, the fitness landscape remains opaque. Green fluorescent protein (GFP), which has revolutionized biological imaging and assays, is one of the most re-designed proteins. While not an enzyme in the conventional sense, GFPs feature competing excited-state decay pathways with the same steric and electrostatic origins as conventional ground-state catalysts, and they exert exquisite control over multiple reaction outcomes through the same principles. Thus, GFP is an “excited-state enzyme”. Herein we show that rationally designed mutants and hybrids that contain environmental mutations and substituted chromophores provide the basis for a quantitative model and prediction that describes the influence of sterics and electrostatics on excited-state catalysis of GFPs. As both perturbations can selectively bias photoisomerization pathways, GFPs with fluorescence quantum yields (FQYs) and photoswitching characteristics<sup>1–4</sup> tailored for specific applications could be predicted and then demonstrated. The underlying energetic

landscape, readily accessible via spectroscopy for GFPs, offers an important missing link in the design of protein function that is generalizable to catalyst design.

## 1. INTRODUCTION

Numerous methods have been employed in developing GFPs with desired behaviors<sup>5–17</sup>, including directed evolution and high-throughput screening of mutant libraries<sup>5–9</sup> that optimize brightness. Machine learning has afforded redder and brighter GFPs<sup>10,11</sup>, and *de novo* protein design has reduced the size of GFP<sup>12</sup>. Unfortunately, the former lacks physical insight, and the latter does not factor in structure–FQY relationships, leading to a FQY (~ 2%) substantially below those of GFPs derived from *Aequorea victoria* (avGFP; FQY ~ 80%). Only through further substantial screening and chromophore modification were brighter versions (FQY ~ 23%) obtained<sup>13</sup>. Photoswitching, the ability to toggle between strongly and weakly fluorescent states through irradiation<sup>18,19</sup>, is another useful function that facilitates super-resolution imaging and optogenetic applications<sup>20,21</sup>. One of the most common photoswitching mechanisms is photoisomerization (Figure 1A), an excited-state bond-rotation pathway that competes with fluorescence emission. Due to this competition, selecting for an efficient photoswitchable protein is difficult via high-throughput screens; past efforts have relied on naturally occurring photoisomerizable GFPs as starting points<sup>14</sup> and/or painstaking combinations of rational design and screening<sup>15–17</sup>. A physical framework capturing the protein environmental factors that control the FQY and photoisomerization in GFPs is necessary to guide more efficient designs, and this is intimately related to the challenge of catalyst design.

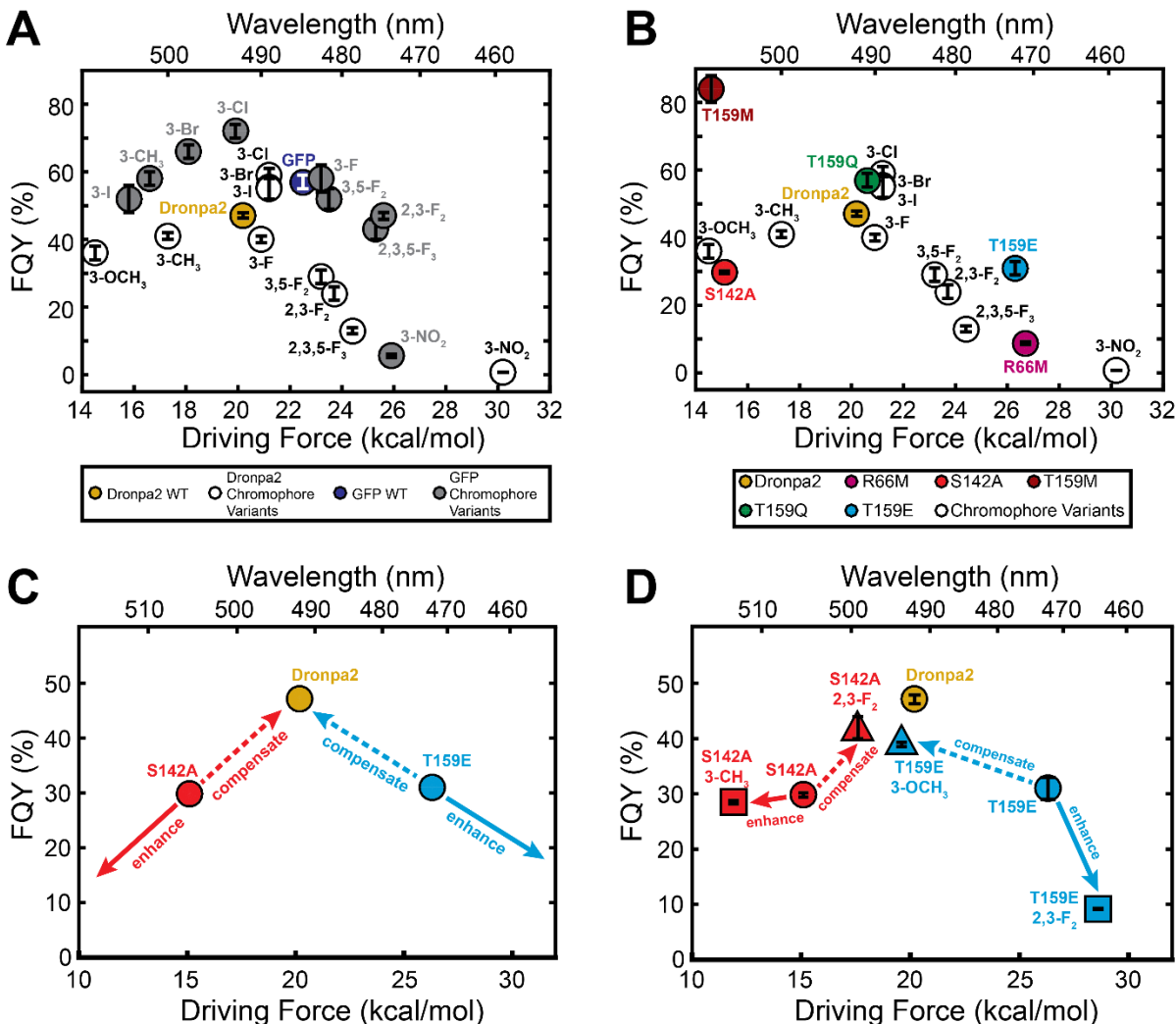


In earlier work, we discovered that the FQY of the *anionic* GFP chromophore embedded in the fixed native protein environments of Dronpa2 or superfolder GFP can be modulated through the introduction of electron-donating and -withdrawing substituents<sup>22</sup>. The FQY exhibits a peaked trend when correlated with transition energy (Figure 2A; now converted into driving force, *vide infra*); the shift in transition energy reflects the extent of electronic perturbation conferred by the substituents. This observation reveals two competing nonradiative photoisomerization pathways (Figure 1A), with the probability of each influenced by the *electrostatic* interaction between the protein environment and the electron flow within the chromophore during photoisomerization<sup>24,25</sup>. Because the twisting about the two exocyclic bonds (the P and I bonds) in the excited state is associated with opposite electron flow directions (Figure 1A), *electrostatics* can cause bond-selective photoisomerization of the chromophore, complementing the more commonly argued role of steric hinderance in suppressing chromophore (photo)isomerization<sup>3,26,27</sup>. The relative barrier heights  $E_P$  and  $E_I$  determine the outcome, and control of these barrier heights is analogous to conventional concepts in catalysis.

To quantify this electrostatic perturbation, we use the driving force  $\Delta\bar{\nu}$  (Figure 1B)<sup>23,28</sup>, which is the relative energy between the P and I resonance forms of the chromophore.  $\Delta\bar{\nu}$  is obtained from the observed transition energy (absorption peak maximum)  $\bar{\nu}_{abs}$  through the Marcus–Hush treatment<sup>23,28</sup>:

$$\bar{\nu}_{abs} = \sqrt{(\Delta\bar{\nu})^2 + (2V_0)^2} \quad (1)$$

where  $V_0$  ( $= 9530 \text{ cm}^{-1}$ <sup>23</sup>) is the electronic coupling between the two resonance forms. With respect to the wild-type environment or chromophore, any decrease or increase in  $\Delta\bar{\nu}$  caused by modifications results in a red or blue shift, respectively (Figure 1C). The driving force can be perturbed through either direct modification of the chromophore or through changes in the protein environment, so it can serve as an ideal quantity to reflect the electron distribution of the chromophore<sup>23</sup>, unify both sources of perturbations<sup>29</sup>, and connect to the underlying theme of electrostatic catalysis.



**Figure 2.** Correlation plots of FQY and driving force. (A) Relationship between FQY and driving force (Figure 1C; converted from eq 1) for unsubstituted and substituted chromophores within Dronpa2 and GFP. In both Dronpa2 and GFP, varying the electronic properties of the chromophore using substituents leads to a nonmonotonic peaked trend. (B) The dependence of FQY on the chromophore's driving force for environmental mutants (colored circles) and chromophore variants (white) of Dronpa2. (C) and (D) The dependence of FQY on the chromophore's driving force in the Dronpa2 compensating and enhancing hybrids schematically (panel C) and experimentally (panel D).

## 2. RESULTS AND DISCUSSION

**2.1. Tuning Electrostatics with Mutants and Hybrids.** Figure 1D shows the chromophore environment of Dronpa2, which exhibits a balance between emission and photoisomerization. To isolate the electrostatic effects, residues immediately surrounding

the chromophore were replaced with amino acids that minimized differences in size. The S142A mutation causes a red shift by destabilizing the P form through removal of a hydrogen bond to the phenolate oxygen (Figures 1B, 1C, S1A, and S2A). The blue-shifted R66M mutant results from I-form destabilization via the removal of the favorable electrostatic interaction between the arginine and the imidazolinone oxygen (Figures 1B, 1C, S1A, and S2B). Within an isosteric T159 mutant series (T159M, T159Q, T159E), T159M is the most red-shifted (by 15 nm compared to wild type), while increasing polarity and/or charge causes a blue shift in T159Q/E; the glutamine and glutamate in T159Q and T159E mutants, respectively, replace S142 as the primary hydrogen bonding partner to the phenolate oxygen and preferentially stabilize the P form (Figures S1A and S2C–S2F).

We next measured the FQYs (Table S1) and plotted them against the corresponding driving forces (eq 1) to determine the electrostatic effect on photoisomerization (Figure 2A). S142A and R66M have a decreased FQY along with strong red- and blue-shifted peak maxima, respectively, recapitulating the peaked trend for chromophore variants (Figure 2B). In contrast, the isosteric T159 mutant series displays a linear correlation with peak maximum, rendering Dronpa (T159M) an outlier of the trend. We attribute this to an increased steric effect for the isosteric series in conjunction with the electrostatic mechanism (*vide infra*). Nevertheless, we still find that the FQY can be tuned electrostatically through environmental mutations.

To circumvent the confounding steric effect, we created hybrids by introducing substituted chromophores into environmental mutants. We first chose one red-shifted (S142A) and one blue-shifted (T159E) mutant with the wild-type Dronpa2 chromophore. We then introduced electron-donating or -withdrawing chromophore substituents to the P ring, which would be predicted to either respectively enhance or compensate for the electronic effect of the mutant with respect to wild-type properties. For example, as the S142A mutation destabilizes the P form, an “enhancing” chromophore modification would be electron-donating and push the electronic properties of the chromophore (driving force and FQY) even further from wild type. A “compensating” modification with an electron-withdrawing group would stabilize the P form, countering the mutational effect and creating a more wild-type-like chromophore (Figure 2C). Note that the same substituent

can act as enhancing or compensating in different environmental contexts according to electrostatic FQY tuning.

For the hybrids, we can quantitatively predict the optimal substituent, within the range available<sup>22</sup>, to pair with a given mutant based on driving force additivity (Table 1). Each point mutant has a driving force, to which a fixed value is added or subtracted based on the chromophore substituent, obtained from the difference between the driving force of Dronpa2 with a natural and substituted chromophore<sup>23</sup>. For the compensating hybrids, the optimal substituents to bring the driving force of S142A and T159E close to wild type are 2,3-F<sub>2</sub> and 3-OCH<sub>3</sub>, respectively. For the enhancing hybrids, we chose substituents with low steric bulk but that still provide a large perturbation to the driving force: S142A/3-CH<sub>3</sub> and T159E/2,3-F<sub>2</sub>. The observed absorption peak maximum for each hybrid agrees well with the predictions (Table 2; Figures S1B and S1C): incorporation of electron-donating and -withdrawing substituents leads to the predicted red and blue shift, respectively. Figure 2D shows the correlation between FQY and driving force for the Dronpa2 hybrids. Both enhancing hybrids (S142A/3-CH<sub>3</sub> and T159E/2,3-F<sub>2</sub>) have a decreased FQY, pushing the values further from wild type as anticipated from electrostatic FQY tuning. Remarkably, both compensating hybrids (S142A/2,3-F<sub>2</sub> and T159E/3-OCH<sub>3</sub>) have an increased FQY compared to the respective mutant with the unsubstituted chromophore, bringing the values closer to the wild-type value. This observation implies that the electronic effect of the chromophore substituent successfully compensates for the electrostatic perturbation caused by the environmental mutation. Either the chromophore substituents (2,3-F<sub>2</sub> or 3-OCH<sub>3</sub>) or the environmental mutations (S142A or T159E) alone each cause a decrease in FQY compared to the wild-type Dronpa2, so the observation of an increased FQY in these compensating hybrids suggests cooperativity (“reciprocal sign epistasis”)<sup>8,30</sup> between deleterious perturbations that cannot otherwise be explained without electrostatic FQY tuning.

**Table 1. Driving force  $\Delta\bar{\nu}$  predictions for each Dronpa2 hybrid.** The left side shows either the additive or subtractive effect of a particular chromophore substituent on the driving force (chromophore substituents: 2,3-F<sub>2</sub>, blue; 3-CH<sub>3</sub>, orange; 3-OCH<sub>3</sub>, pink). The right side shows the predicted driving force for each hybrid combining the effect of the point mutant and the chromophore substituent. Driving force values are extracted from ref. 23 and calculated from eq 1 with an electronic coupling  $V_0$  of 9530 cm<sup>-1</sup>. The

chromophore modified with OCH<sub>3</sub> possesses a somewhat smaller  $V_0$  than the unsubstituted counterpart<sup>23</sup> but for the current purpose the same  $V_0$  is used for driving force evaluation.

Chromophore Variant	Driving Force $\Delta\bar{\nu}$ (cm <sup>-1</sup> )	Difference from Dronpa2 $\Delta\Delta\bar{\nu}$ (cm <sup>-1</sup> )	Hybrid Protein	Point Mutant Driving Force $\Delta\bar{\nu}$ (cm <sup>-1</sup> )	Substituent Driving Force $\Delta\Delta\bar{\nu}$ (cm <sup>-1</sup> )	Predicted Combined Driving Force $\Delta\bar{\nu}$ (cm <sup>-1</sup> )
Dronpa2 ("wild type")	7010	0	S142A 2,3-F <sub>2</sub>	5300	+1290 (compensating)	6590
2,3-F <sub>2</sub>	8300	+1290	S142A 3-CH <sub>3</sub>	5300	-820 (enhancing)	4480
3-CH <sub>3</sub>	6190	-820	T159E 2,3-F <sub>2</sub>	9200	+1290 (enhancing)	10490
3-OCH <sub>3</sub>	5070	-1940	T159E 3-OCH <sub>3</sub>	9200	-1940 (compensating)	7260

**Table 2. Predicted and observed driving forces, absorption peak maxima, and FQYs for each Dronpa2 hybrid.** The observed driving force is calculated from eq 1 with an electronic coupling  $V_0$  of 9530 cm<sup>-1</sup> <sup>23</sup>. SD: standard deviation. The predicted absorption peak maxima are at most 5 nm from the observed ones, with better accuracy for redder species. The entries are colored according to the involved chromophore substituents: 2,3-F<sub>2</sub>, blue; 3-CH<sub>3</sub>, orange; 3-OCH<sub>3</sub>, pink.

Hybrid Protein	Predicted Combined Driving Force $\Delta\bar{\nu}$ (cm <sup>-1</sup> )	Predicted Absorption Peak Maximum (nm)	Observed Driving Force $\Delta\bar{\nu}$		Observed Absorption Peak Maximum (Transition Energy $\bar{\nu}_{abs}$ )			FQY (%)	FQY SD (%)
			(cm <sup>-1</sup> )	(kcal/mol)	(cm <sup>-1</sup> )	(nm)	(kcal/mol)		
T159E 2,3-F <sub>2</sub> (enhancing)	10490	459.6	9990	28.6	21520	464.7	61.5	9.2	0.1
T159E	N/A		9190	26.3	21160	472.5	60.5	31	2
T159E 3-OCH <sub>3</sub> (compensating)	7260	490.3	6870	19.6	20260	493.6	57.9	38.9	0.4
Dronpa2 ("wild type")	N/A		7070	20.2	20310	492.4	58.1	46	2



S142A 2,3-F <sub>2</sub> (compensating)	6590	495.9	6160	17.6	20030	499.3	57.3	42	2
S142A	N/A		5290	15.1	19780	505.5	56.6	29.8	0.4
S142A 3-CH <sub>3</sub> (enhancing)	4480	510.7	4170	11.9	19510	512.6	55.8	28.5	0.3

## 2.2. Predictive Model for Steric and Electrostatic Effects on Excited-State Catalysis.

The FQY  $\varphi_{\text{fl}}$  is the ratio between the intrinsic spontaneous emission rate  $k_{\text{fl}}$  and the total excited-state decay rate constants<sup>31</sup> (Figure 1A):

$$\varphi_{\text{fl}} = k_{\text{fl}}\tau = \frac{k_{\text{fl}}}{k_{\text{fl}} + k_{\text{iso}} + k_{\text{other}}} \quad (2)$$

where  $k_{\text{iso}}$  and  $k_{\text{other}}$  denote the total rate constant for excited-state isomerization and other nonradiative pathways, respectively;  $\tau$  is the fluorescence lifetime. We can then dissect the temperature, electrostatic, and steric dependence of each term to understand how the chromophore's FQY is influenced by its environment.  $k_{\text{fl}}$  is minimally tunable through electrostatics as evidenced by the nearly constant transition dipole moment across different GFP mutants<sup>23,32</sup>; steric effects are irrelevant since emission is a Franck–Condon process. The only way the protein environment can tune the FQY is through modulating the competing nonradiative decay pathways.  $k_{\text{fl}}$  is estimated to be (3.5 ns)<sup>-1</sup><sup>33</sup>, so any nonradiative process much slower than this value cannot tune FQYs.

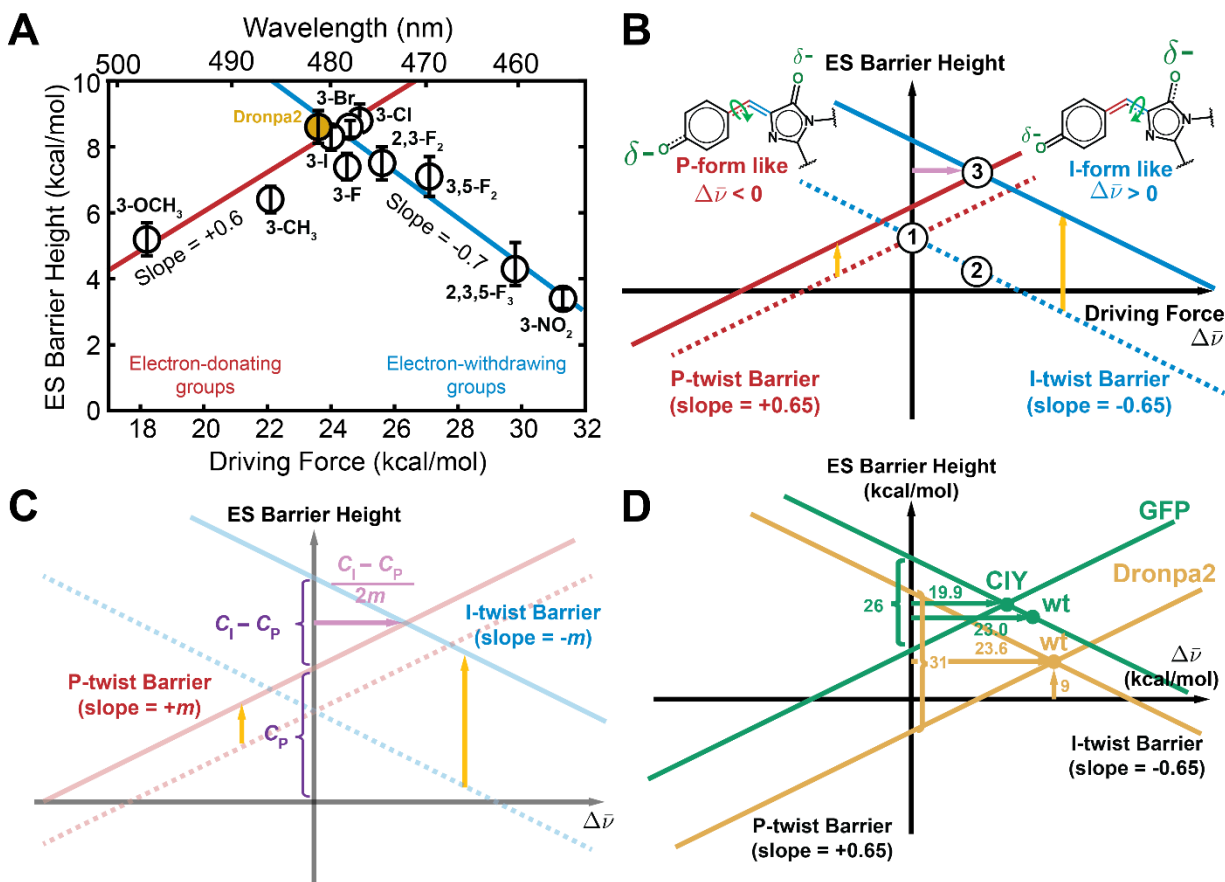
$k_{\text{other}}$  arises from both direct internal conversion and intersystem crossing, but the latter is much less competitive than other excited-state processes<sup>34</sup>. Accordingly, we can approximate  $k_{\text{other}}$  with a single rate constant from direct internal conversion  $k_{\text{IC}}$  due to vibrational wavefunction overlap between the ground and excited electronic states, which is relatively temperature insensitive (see Section S6 of ref. 22 and Section S11 of ref. 30). To obtain  $k_{\text{IC}}$ , we examine a GFP mutant series in which the threonine at position 203 is replaced with aromatic side chains that  $\pi$ – $\pi$  stack with the chromophore P ring and can be varied in electron richness. The corresponding FQYs are nearly constant around 77% despite the modified electrostatic interaction (Figure S3 and Table S2). Steric hinderance

by the aromatic ring overwhelms electrostatics and renders  $k_{\text{iso}}$  uncompetitive; the remaining 23% of excited-state decay can be ascribed to internal conversion;  $k_{\text{IC}}$  is  $(12 \text{ ns})^{-1}$  and imposes an upper limit for GFP's FQY of approximately 80%<sup>35</sup>, close to that of avGFP. Extensive mutational studies also demonstrate that avGFP is indeed located at the local maximum of the fitness landscape for brightness<sup>8</sup>. Any approach that slows excited-state isomerization down to tens of nanoseconds is sufficient to maximize FQY.

In contrast with other processes, excited-state isomerization requires crossing over an energy barrier along with significant electronic and nuclear motion (Figure 1A), so the isomerization rate  $k_{\text{iso}}$  is almost solely responsible for the temperature, electrostatic, and steric dependence of FQY<sup>22</sup>. The associated barriers are typically  $> 3 \text{ kcal/mol}$  for GFPs<sup>22</sup>, and the corresponding rate constants are comparable with  $k_{\text{fl}}$  (ns timescale). The rapid intramolecular vibrational energy redistribution (ps timescale)<sup>31,39</sup> right after excitation renders the system thermally equilibrated before emission and isomerization, so the assumption for Arrhenius behavior, also common for ground-state catalysis, is met for isomerization. A pre-exponential factor  $A$  and an energy barrier  $E$  can thus be assigned for each isomerization pathway:

$$k_{\text{iso}} = A_{\text{P}} \exp\left(-\frac{E_{\text{P}}}{RT}\right) + A_{\text{I}} \exp\left(-\frac{E_{\text{I}}}{RT}\right) \approx A \exp\left(-\frac{E}{RT}\right) \quad (3)$$

where  $k_{\text{iso}}$  is then approximated with a single Arrhenius expression when we measure the excited-state energy barrier  $E$  of Dronpa2 variants using the temperature dependence of their fluorescence lifetimes<sup>22</sup>. If  $A_{\text{P}}$  and  $A_{\text{I}}$  are close in value,  $A$  should be close to both  $A_{\text{P}}$  and  $A_{\text{I}}$ , and the measured excited-state barrier height  $E$  can well approximate the lesser of the two barriers,  $E_{\text{P}}$  or  $E_{\text{I}}$  (Figure 1A).  $A$  is  $10^3 - 10^5 \text{ ns}^{-1}$ <sup>22</sup>, agreeing well with the value estimated from transition state theory ( $\frac{kT}{h} \sim 10^{13} \text{ s}^{-1}$ ). This suggests that when the excited-state barrier exceeds 9 kcal/mol (i.e.,  $k_{\text{iso}}$  being 1% of  $k_{\text{fl}}$  at 300 K), as for the  $\pi$ - $\pi$  stacking GFP mutants (Figure S3), no further increase in FQY can be seen as it reaches the upper limit.



**Figure 3.** Qualitative and quantitative analysis of sterics and electrostatics of the GFP chromophore within a protein environment. (A) Excited-state energy barriers for Dronpa2 chromophore variants plotted against driving force. The fit through the electron-withdrawing and -donating group points is shown as a blue and red line, respectively, with wild-type Dronpa2 shown in gold at the apex. (B) Schematic showing effects of sterics around the P and I rings of the chromophore on the magnitude and apex position of the excited-state energy barrier, shown as blue and red lines for I and P twist, respectively. Without steric effects (dashed lines), the apex lies at zero driving force (case 1). Shifting the driving force to positive energy (i.e., to the right) leads to a preference for I-bond rotation due to a lower barrier (case 2). Greater steric confinement (solid lines) around the I ring (i.e., longer yellow arrow for I-twist than P-twist) causes the apex to shift to the right (positive driving forces, case 3). (C) The algebraic relationship between the apex shift and differential sterics according to panel B. (D) Interplay between steric and electrostatic effects for the excited-state barrier height of GFP (gold, inferred from Figure 2A) and Dronpa2 (green, panel A). See also Figure 4A.

We now replot the excited-state barriers from Dronpa2 variants (Figure 3B in ref. 22) against the corresponding driving forces to better understand the electrostatic effect (Figure 3A). Linear fits to the electron-donating and -withdrawing substituent exhibit

slopes of +0.6 and -0.7, reflecting the electrostatic sensitivity of  $E_P$  and  $E_I$ , respectively. These slopes are about equal in magnitude ( $\sim 0.65$  within experimental errors) and opposite in sign; the signs agree well with a model treating the chromophore as an allylic anion<sup>22</sup>. Analogous to electrostatic enzyme catalysis<sup>40,41</sup>, this electrostatic sensitivity originates from chromophore charge redistribution during photoisomerization interacting with the protein environment (Figure 1A), effectively an excited-state enzyme that selectively catalyzes either P- or I-bond rotation. We expect these slopes in Figure 3A to be directly transferable to different environments around the chromophore, since the driving force is the only parameter responsible for the electrostatic sensitivity of the entire PES<sup>22</sup>:

$$E_P = 0.65\Delta\bar{v} + C_P \quad \text{and} \quad E_I = -0.65\Delta\bar{v} + C_I \quad (4)$$

where the steric effects, including the intrinsic barrier to bond isomerization in the absence of any external steric constraint, can be separated out in terms of empirical constants  $C_P$  and  $C_I$  ( $y$ -intercepts of red and blue lines in Figure 3B, respectively). We can then rewrite eqs 2 and 3 to explicitly show the electrostatic and steric dependence of the FQY:

$$\varphi_{\text{fl}}(T, \Delta\bar{v}, \text{sterics}) \approx \frac{k_{\text{fl}}}{k_{\text{fl}} + k_{\text{other}} + A \left[ \exp\left(-0.65 \frac{\Delta\bar{v}}{RT} - \frac{C_P(\text{sterics})}{RT}\right) + \exp\left(0.65 \frac{\Delta\bar{v}}{RT} - \frac{C_I(\text{sterics})}{RT}\right) \right]} \quad (5)$$

Two factors mediate excited-state pathway selection: sterics, which acts upon large scale nuclear motion of two rings during isomerization, and electrostatics, which interacts with electronic redistribution during isomerization (or driving force). The electrostatic influence of the red fluorescent protein environment on the corresponding chromophore's FQY is also extensively discussed by a recent paper<sup>42</sup>, while our physical model treats electrostatics differently and explicitly incorporates the steric component (see Section S2 in Supporting Information). According to eq 5, FQY is a nonlinear function of  $\Delta\bar{v}$ , and thus the linear additivity of driving force does not translate to an additivity of FQY, as observed from the compensating hybrids (Figure 2D and Table 2). Cooperativity between mutations, a phenomenon that renders protein design and even directed evolution challenging<sup>29,43</sup>, could similarly be partly explained by a nonlinear function (i.e., FQY) encoding two (or more) pathways dependent on an additive underlying parameter (i.e., driving force)<sup>23</sup>. Steric effects  $C_P$  and  $C_I$  serve as an alternative tuning mechanism

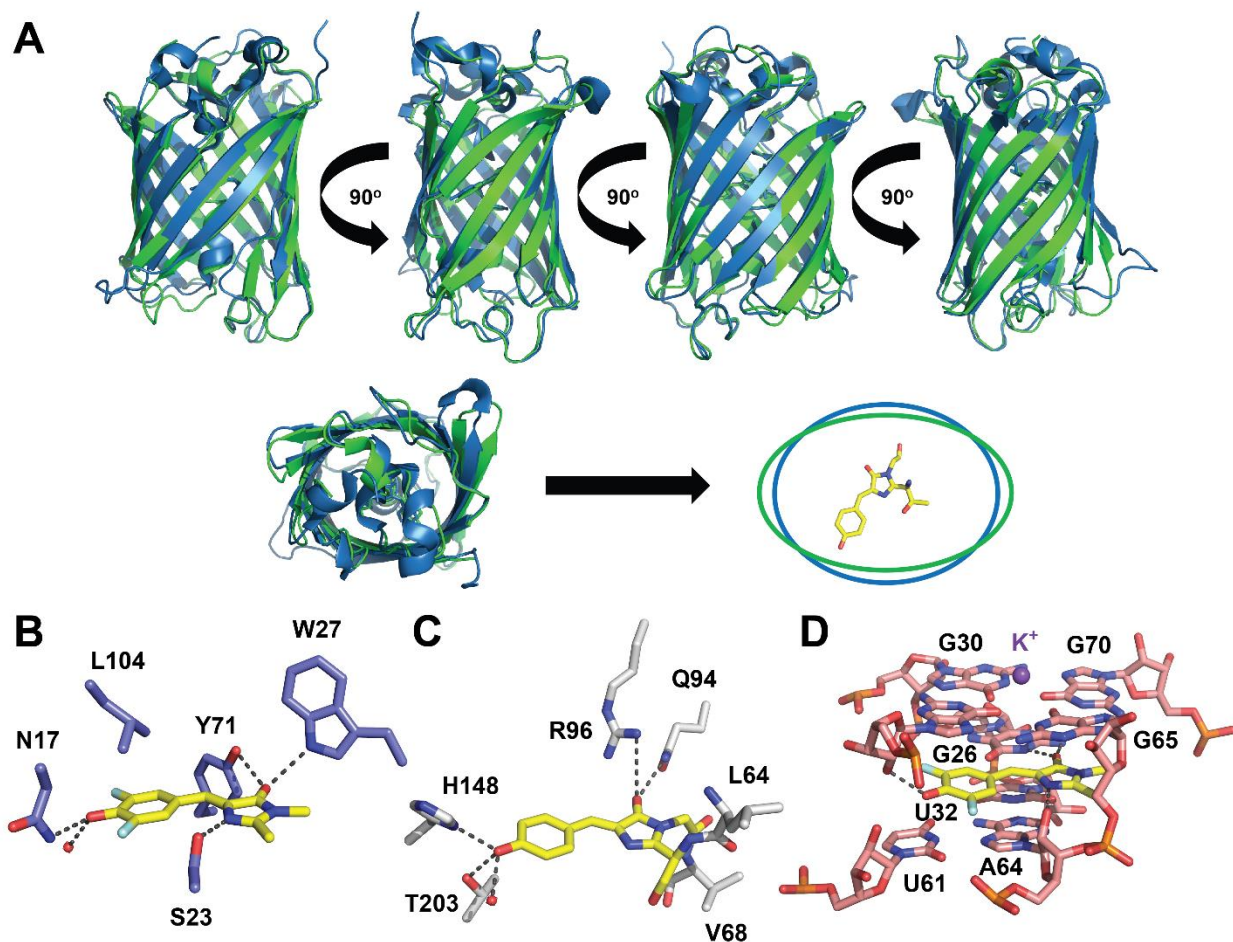
for the excited-state barriers  $E_P$  and  $E_I$ , preventing the FQY from being completely tied to color via electrostatics, as is the case for other photophysical properties<sup>23</sup>. If  $C_P$  equals  $C_I$ , there should be no preference for either isomerization pathway when  $\Delta\bar{\nu} = 0$ , corresponding to a maximum FQY (eq 5; Figure 3B, case 1). Since  $\Delta\bar{\nu} = 0$  also corresponds to the reddest possible absorption (eq 1), a combination of these two equations would suggest that the redder the chromophore, the higher the FQY by varying  $\Delta\bar{\nu}$ . However, we observe an apex in the trend that is not centered at  $\Delta\bar{\nu} = 0$  (Figure 3A), suggesting that  $C_P$  is not identical to  $C_I$ . Intuitively, the volume-demanding I twist experiences more steric hinderance than the P twist within the protein environment since the I ring is covalently anchored.

With eq 4, we can explain the apex position in the FQY (or excited-state barrier) vs driving force plot (Figure 3B). The sign of the driving force is defined positive when the P form is more stable than the I form, which is the case for all proteins studied so far<sup>23</sup> (Figure 1B). With zero differential sterics from the protein environment ( $C_P = C_I$ ; dashed lines) and zero driving force, the negative charge of the anionic chromophore is maximally delocalized and both exocyclic bonds are equally probable to twist upon excitation. This corresponds to the largest possible barrier when  $C_P = C_I$ , and the apex is located at  $\Delta\bar{\nu} = 0$  (Figure 3B, case 1). When the driving force becomes positive (right side of Figure 3B), electron density is reduced at the I bond (i.e., more single-bond character) upon excitation, and the I twist becomes more favorable<sup>44</sup> (Figure 3B, case 2). If the I ring is anchored inside the protein,  $C_I$  becomes larger than  $C_P$  (yellow arrows and solid lines in Figure 3B). Consequently, the apex shifts along the x-axis and lies at a positive driving force, as observed in Figure 3A, and it also increases along the y-axis due to the resulting constriction on bond rotation (Figure 3B, case 3). At that apex, the driving force from electrostatic influences matches the apex shift caused by differential steric interactions. However, when the steric effects are large enough to render  $k_{iso}$  uncompetitive with  $k_{fl}$  (Figure S3), the maximally allowed FQY is reached, and the apex for FQY cannot be detected. Note that the driving force at the apex is determined from the *differential* sterics ( $C_I - C_P$ ), while the barrier heights are affected by the *absolute* sterics ( $C_I$  or  $C_P$ ), so it is possible to have an apex location at zero driving force when steric hinderance to the P twist is comparable with I ring anchoring (Figure 3C).

**2.3. Applications, Generalizations, and Implications for Design.** This model allows us to quantitatively evaluate the contributions of sterics and electrostatics to excited-state catalysis. From Figure 3A, wild-type Dronpa2 sits at the apex among all Dronpa2 variants. As its FQY (~ 50%) is far from the maximally allowed 80%, this implies that the corresponding driving force (23.6 kcal/mol) offsets the differential sterics, so we can estimate the differential sterics as 31 kcal/mol ( $23.6 \times 2 \times 0.65$ , Figures 3C and 3D). For superfolder GFP, the apex (the monochlorinated variant, Figure 2A) lies at a driving force of 19.9 kcal/mol and approaches the FQY limit of 80%<sup>22</sup>. The corresponding differential sterics is 26 kcal/mol ( $= 19.9 \times 2 \times 0.65$ ). Combined with the fact that GFP has a higher apex FQY than Dronpa2, we can infer that the overall steric contribution should be higher for GFP than Dronpa2, but the differential sterics is also 5 kcal/mol smaller ( $= 31 - 26$ ) for GFP, leading to an apex located at a smaller driving force than Dronpa2 (Figure 3D). This is explained by a tighter  $\beta$ -barrel for GFP compared to Dronpa2, resulting in a more sterically hindered P twist (Figure 4A). Moreover, since the unmodified chromophore in the GFP environment possesses a driving force of 23.0 kcal/mol (Figure 2A, as opposed to 23.6 kcal/mol for that in Dronpa2), the Dronpa2 I-twist barrier is also lowered by 0.4 kcal/mol ( $= (23.6 - 23.0) \times 0.65$ ) electrostatically compared to the GFP counterpart. Therefore, both steric and electrostatic (to a lesser extent) effects work together in the GFP barrel to promote chromophore fluorescence, while Dronpa2 exhibits a higher photoisomerization efficiency (Figure 5A). For the Dronpa2 T159 isosteric series, the lengthened side chain creates more steric bulk to P twist and shifts the apex to a smaller driving force and higher FQY (Figure 2B), explaining why T159M appears as an outlier to the peaked trend.

This analysis can also explain why the *de novo* designed mFAPs (Figure 4B) failed to recapitulate avGFP's high FQYs (Figure 4C)<sup>12</sup> and more generally how an understanding of the energy landscape can provide guidance for the design of functional proteins. Original mFAPs utilize the same difluorinated chromophore as the RNA mimic Spinach (Figure 4D)<sup>45</sup> to encourage chromophore deprotonation, but fluorines lower the I-twist barrier as electron-withdrawing substituents<sup>22</sup>. In Spinach,  $\pi$ - $\pi$  stacking with G-quadruplexes effectively inhibits isomerization (Figure 4D)<sup>45,46</sup>, leading to a FQY of 72%. In mFAPs, however, the chromophore is neither anchored to the protein as in avGFP

(Figure 4C) nor motionally restricted. M27W is present in mFAP1 and mFAP2 to interact with the I ring via a hydrogen bond (Figure 4B), but this interaction is not sufficient to restore the maximal FQY. To further increase the FQYs, this analysis suggests the addition or removal of fluorines from the chromophore's I or P rings, respectively, and the introduction of aromatic amino acids near the chromophore's P ring to encourage  $\pi$ - $\pi$  stacking interactions. In fact, the newly installed  $-\text{CF}_3$  group on the I ring and L104H likely explains the much-improved FQY (23%) of chromophore-bound mFAP10<sup>13</sup>.

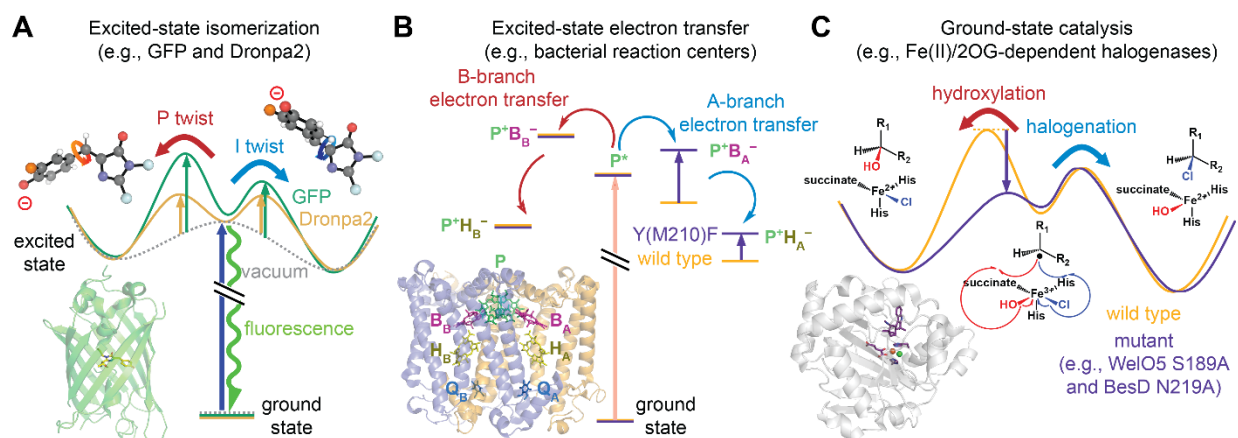


**Figure 4.** GFP chromophore (yellow) in various biomolecular environments. (A) Overlaid  $\beta$ -barrels of Dronpa2 (green, PDB: 6NQJ) and GFP (blue, PDB: 6OFK). The barrels are shown in different perspectives to illustrate the differences in dimensions. The overlaid ovals at the right bottom corner, color coded according to the proteins they represent, are exaggerated simplification for the cross sections of the barrels. P-twist motion clashes with residues along the wider dimension, for which GFP is tighter than Dronpa2. (B) mFAP1 (PDB: 6CZI). (C) avGFP (PDB: 2WUR). (D) Spinach (PDB: 4TS2). In panels B–D, hydrogen bonds associated with the chromophore are shown as dashed lines.



### 3. CONCLUSIONS

GFP is both green and fluorescent, while the free GFP chromophore in water is neither, so it is tempting to ascribe this drastic change in properties to the protein environment. However, the chromophore's ability to be green and fluorescent is already encoded in its PESs (i.e., energy landscape), and these properties can also be elicited using non-protein environments<sup>3,27</sup>. An analogous example is the relationship between an enzyme and its substrate. The availability of different reaction pathways and the potential for pathway selection, existing for numerous ground-state and excited-state enzymes<sup>47–50</sup>, are already inscribed in the PES(s) of the chromophore/substrate, illustrated by diverse examples in Figure 5. The protein environment can only stabilize the transition state of one particular pathway over another that is otherwise suppressed; it cannot then create new reactions. Therefore, to rationally design enzymes that are superior at catalyzing a reaction, it is important to sample a wide range of perturbations to substrates (or chromophores capable of structural change) and the environment's steric or electrostatic influences on the energetics of non-productive yet competitive pathways rather than only those that exhibit more desirable phenotypes<sup>51</sup>. Only when those less desirable cases are understood can we mechanistically deduce why the more productive pathway is not taken, guiding future design efforts to optimize the desired function.



**Figure 5.** Energetic control of competing pathways for excited- and ground-state catalysis by diverse protein environments. (A) GFP (green) and Dronpa2 (gold) protein environments suppress excited-state isomerization of the chromophore to different degrees compared to that in vacuum (gray), rendering GFP less photoisomerizable than



Dronpa2 (Figure 3D). (B) Y(M210)F mutant (purple) of *Rhodobacter sphaeroides* photosynthetic reaction center reveals that tyrosine at M210, which stabilizes the first intermediate, is in part responsible for the unidirectional excited-state electron transfer of wild type (orange)<sup>52,53</sup>. (C) Wild-type Fe(II)/2-oxoglutarate (2OG)-dependent halogenases (orange) chlorinate their substrates, but their intrinsic hydroxylating power can be unleashed upon mutation (purple)<sup>54,55</sup>. The default (blue) and the side pathways (red for all and green for panel A) are shown on the right and left for each panel, respectively. Energies are not drawn to scale.

## ACKNOWLEDGMENTS

We thank Jacob M. Kirsh and Steven D. E. Fried for useful comments. C.-Y.L. was supported by a Kenneth and Nina Tai Stanford Graduate Fellowship and the Taiwanese Ministry of Education. M.G.R. was supported by a Center for Molecular Analysis and Design graduate fellowship. This work was supported, in part, by NIH Grant GM118044 (to S.G.B.) and NSF CCI Phase I: Center for First Principles Design of Quantum Processes (CHE-1740645). Use of the Stanford Synchrotron Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-1542152.

## REFERENCES

- [1] Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* **2010**, *90*, 1103–1163.

- [2] Nienhaus, K.; Nienhaus, G. U. Fluorescent proteins for live-cell imaging with super-resolution. *Chem. Soc. Rev.* **2014**, *43*, 1088–110.
- [3] Walker, C. L.; Lukyanov, K. A.; Yampolsky, I. V.; Mishin, A. S.; Bommarius, A. S.; Duraj-Thatte, A. M.; Azizi, B.; Tolbert, L. M.; Solntsev, K. M. Fluorescence imaging using synthetic GFP chromophores. *Curr. Opin. Chem. Biol.* **2015**, *27*, 64–74.
- [4] Ferré-D'Amaré, A. R.; Truong, L. From fluorescent proteins to fluorogenic RNAs: Tools for imaging cellular macromolecules. *Protein Sci.* **2019**, *28*, 1374–1386.
- [5] Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp.* red fluorescent protein. *Nat. Biotechnol.* **2004**, *22*, 1567–1572.
- [6] Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.
- [7] Yoo, T. H.; Link, A. J.; Tirrell, D. A. Evolution of a fluorinated green fluorescent protein. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887–13890.
- [8] Sarkisyan, K. S.; Bolotin, D. A.; Meer, M. V.; Usmanova, D. R.; Mishin, A. S.; Sharonov, G. V.; Ivankov, D. N.; Bozhanova, N. G.; Baranov, M. S.; Soylemez, O.; Bogatyreva, N. S.; Vlasov, P. K.; Egorov, E. S.; Logacheva, M. D.; Kondrashov, A. S.; Chudakov, D. M.; Putintseva, E. V.; Mamedov, I. Z.; Tawfik, D. S.; Lukyanov, K. A.; Kondrashov, F. A. Local fitness landscape of the green fluorescent protein. *Nature* **2016**, *533*, 397–401.
- [9] Manna, P.; Hung, S.-T.; Mukherjee, S.; Friis, P.; Simpson, D. M.; Lo, M. N.; Palmer, A. E.; Jimenez, R. Directed evolution of excited state lifetime and brightness in FusionRed using a microfluidic sorter. *Integr. Biol.* **2018**, *10*, 516–526.
- [10] Saito, Y.; Oikawa, M.; Nakazawa, H.; Niide, T.; Kameda, T.; Tsuda, K.; Umetsu, M. Machine-learning-guided mutagenesis for directed evolution of fluorescent proteins. *ACS Synth. Biol.* **2018**, *7*, 2014–2022.
- [11] Alley, E. C.; Khimulya, G.; Biswas, S.; AlQuraishi, M.; Church, G. M. Unified rational protein engineering with sequence-based deep representation learning. *Nat. Methods* **2019**, *16*, 1315–1322.

- [12] Dou, J.; Vorobieva, A. A.; Sheffler, W.; Doyle, L. A.; Park, H.; Bick, M. J.; Mao, B.; Foight, G. W.; Lee, M. Y.; Gagnon, L. A.; Carter, L.; Sankaran, B.; Ovchinnikov, S.; Marcos, E.; Huang, P.-S.; Vaughan, J. C.; Stoddard, B. L.; Baker, D. De novo design of a fluorescence-activating  $\beta$ -barrel. *Nature* **2018**, *561*, 485–491.
- [13] Klima, J. C.; Doyle, L. A.; Lee, J. D.; Rappleye, M.; Gagnon, L. A.; Lee, M. Y.; Barros, E. P.; Vorobieva, A. A.; Dou, J.; Bremner, S.; Quon, J. S.; Chow, C. M.; Carter, L.; Mack, D. L.; Amaro, R. E.; Vaughan, J. C.; Berndt, A.; Stoddard, B. L.; Baker, D. Incorporation of sensing modalities into de novo designed fluorescence-activating proteins. *Nat. Commun.* **2021**, *12*, 856.
- [14] Ando, R.; Mizuno, H.; Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **2004**, *306*, 1370–1373.
- [15] Stiel, A. C.; Andresen, M.; Bock, H.; Hilbert, M.; Schilde, J.; Schönle, A.; Eggeling, C.; Egner, A.; Hell, S. W.; Jakobs, S. Generation of monomeric reversibly switchable red fluorescent proteins for far-field fluorescence nanoscopy. *Biophys. J.* **2008**, *95*, 2989–2997.
- [16] Grotjohann, T.; Testa, I.; Leutenegger, M.; Bock, H.; Urban, N. T.; Lavoie-Cardinal, F.; Willig, K. I.; Eggeling, C.; Jakobs, S.; Hell, S. W. Diffraction-unlimited all-optical imaging and writing with a photochromic GFP. *Nature* **2011**, *478*, 204–208.
- [17] El Khatib, M., Martins, A., Bourgeois, D., Colletier, J.-P. & Adam, V. Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Sci. Rep.* **6**, 18459 (2016).
- [18] Acharya, A.; Bogdanov, A. M.; Grigorenko, B. L.; Bravaya, K. B.; Nemukhin, A. V.; Lukyanov, K. A.; Krylov, A. I. Photoinduced chemistry in fluorescent proteins: Curse of blessing? *Chem. Rev.* **2017**, *117*, 758–795.
- [19] Bourgeois, D.; Adam, V. Reversible photoswitching in fluorescent proteins: A mechanistic view. *IUBMB Life* **2012**, *64*, 482–491.
- [20] Shcherbakova, D. M.; Sengupta, P.; Lippincott-Schwartz, J.; Verkhusha, V. V. Photocontrollable fluorescent proteins for superresolution imaging. *Annu Rev. Biophys.* **2014**, *43*, 303–329.
- [21] Zhou, X. X.; Lin, M. Z. Photoswitchable fluorescent proteins: Ten years of colorful chemistry and exciting applications. *Curr. Opin. Chem. Biol.* **2013**, *17*, 682–690.

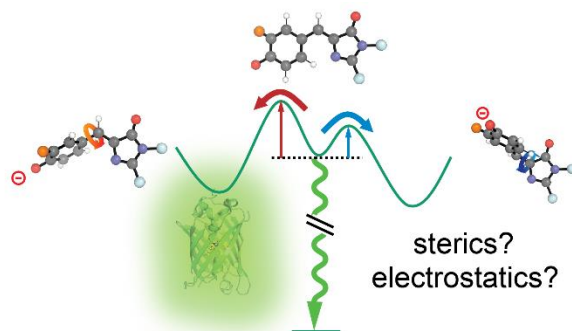
- [22] Romei, M. G.; Lin, C.-Y.; Mathews, I. I.; Boxer, S. G. Electrostatic control of photoisomerization pathways in proteins. *Science* **2020**, *367*, 76–79.
- [23] Lin, C.-Y.; Romei, M. G.; Oltrogge, L. M.; Mathews, I. I.; Boxer, S. G. Unified model for photophysical and electro-optical properties of green fluorescent proteins. *J. Am. Chem. Soc.* **2019**, *141*, 15250–15265.
- [24] Olsen, S.; Lamothe, K.; Martínez, T. J. Protonic gating of excited-state twisting and charge localization in GFP chromophores: A mechanistic hypothesis for reversible photoswitching. *J. Am. Chem. Soc.* **2010**, *132*, 1192–1193.
- [25] Martin, M. E.; Negri, F.; Olivucci, M. Origin, nature, and fate of the fluorescent state of the green fluorescent protein chromophore at the CASPT2//CASSCF resolution. *J. Am. Chem. Soc.* **2004**, *126*, 5452–5464.
- [26] Stiel, A. C.; Trowitzsch, S.; Weber, G.; Andresen, M.; Eggeling, C.; Hell, S. W.; Jakobs, S.; Wahl, M. C. 1.8 Å bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. *Biochem. J.* **2007**, *402*, 35–42.
- [27] Tolbert, L. M.; Baldridge, A.; Kowalik, J.; Solntsev, K. M. Collapse and recovery of green fluorescent protein chromophore emission through topological effects. *Acc. Chem. Res.* **2012**, *45*, 171–181.
- [28] Olsen, S. A modified resonance-theoretic framework for structure–property relationships in a halochromic oxonol dye. *J. Chem. Theory Comput.* **2010**, *6*, 1089–1103.
- [29] 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.
- [30] Poelwijk, F. J.; Kiviet, D. J.; Weinreich, D. M.; Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **2007**, *445*, 383–386.
- [31] 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.
- [32] However, when introducing substituents to the chromophore, the transition dipole moment and the extinction coefficient change due to covalent modification.

Changes in these parameters likely result in the observed asymmetric FQY trend (Figure 2A) when plotted against transition energy for Dronpa2 chromophore variants, while the corresponding excited-state barrier trend is symmetric (Figure 3A). Specifically, the methoxy modified chromophore (far red-shifted) has a higher  $k_{\text{fl}}$  than the nitro and trifluoro modified chromophores (far blue-shifted).

- [33] Striker, G.; Subramaniam, V.; Seidel, C. A. M.; Volkmer, A. Photochromicity and fluorescence lifetimes of green fluorescent protein. *J. Phys. Chem. B* **1999**, *103*, 8612–8617.
- [34] Byrdin, M.; Duan, C.; Bourgeois, D.; Brettel, K. A long-lived triplet state is the entrance gateway to oxidative photochemistry in green fluorescent proteins. *J. Am. Chem. Soc.* **2018**, *140*, 2897–2905.
- [35] One caveat of this statement is that the normal modes and their corresponding vibronic couplings are different for the chromophore in vacuum vs protein<sup>36</sup>, so  $k_{\text{IC}}$  in each case could be drastically different. Here we assume that since low-frequency vibrational modes are more important than higher frequency ones for internal conversion due to larger numbers of participating phonons<sup>37,38</sup>,  $k_{\text{IC}}$  should be roughly constant across protein environments with  $\beta$ -barrel scaffolds.
- [36] Bochenkova, A. V.; Andersen, L. H. Ultrafast dual photoresponse of isolated biological chromophores: link to the photoinduced mode-specific non-adiabatic dynamics in proteins. *Faraday Discuss.* **2013**, *163*, 297-319.
- [37] Englman, R.; Jortner, J. The energy gap law for radiationless transitions in large molecules. *Mol. Phys.* **1970**, *18*, 145-164.
- [38] Nitzan, A. *Chemical Dynamics in Condensed Phases*, 1<sup>st</sup> ed.; Oxford University Press: New York, 2006; pp 439-449.
- [39] Oscar, B. G.; Zhu, L.; Wolfenden, H.; Rozanov, N. D.; Chang, A.; Stout, K. T.; Sandwich, J. W.; Porter, J. J.; Mehl, R. A.; Fang, C. Dissecting optical response and molecular structure of fluorescent proteins with non-canonical chromophores. *Front. Mol. Biosci.* **2020**, *7*, 131.
- [40] 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.

- [41] Fried, S. D.; Boxer, S. G. Electric field and enzyme catalysis. *Annu. Rev. Biochem.* **2017**, *86*, 387–415.
- [42] Drobizhev, M.; Molina, R. S.; Callis, P. R.; Scott, J. N.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E. Local electric field controls fluorescence quantum yield of red and far-red fluorescent proteins. *Front. Mol. Biosci.* **2021**, *8*, 633217.
- [43] Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866–876.
- [44] Carrascoasa, E.; Bull, J. N.; Scholz, M. S.; Coughlan, N. J. A.; Olsen, S.; Wille, U.; Bieske, E. J. Reversible photoisomerization of the isolated green fluorescent protein chromophore. *J. Phys. Chem. Lett.* **2018**, *9*, 2647–2651.
- [45] You, M.; Jaffrey, S. R. Structure and mechanism of RNA mimics of green fluorescent protein. *Annu. Rev. Biophys.* **2015**, *44*, 187–206.
- [46] Huang, H.; Suslov, N. B.; Li, N.-S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **2014**, *10*, 686–691.
- [47] Schulten, K.; Hayashi, S. Quantum biology of retinal. In *Quantum Effects in Biology*; Mohseni, M.; Omar, Y.; Engel, G. S., Plenio, M. B. Eds.; Cambridge University Press: Cambridge, United Kingdom, 2014; pp 237–263.
- [48] Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res. Toxicol.* **2001**, *14*, 611–650.
- [49] Jasniewski, A. J.; Que, L., Jr. Dioxygen activation by nonheme diiron enzymes: Diverse dioxygen adducts, high-valent intermediates, and related model complexes. *Chem. Rev.* **2018**, *118*, 2554–2592.
- [50] Nicolet, Y. Structure–function relationships of radical SAM enzymes. *Nat. Catal.* **2020**, *3*, 337–350.
- [51] Mokhtari, D. A.; Appel, M. J.; Fordyce, P. M.; Herschlag, D. High throughput and quantitative enzymology in the genomic era. *Curr. Opin. Struct. Biol.* **2021**, *71*, 259–273.
- [52] Tamura, H.; Saito, K.; Ishikita, H. The origin of unidirectional charge separation in photosynthetic reaction centers: nonadiabatic quantum dynamics of exciton and charge in pigment–protein complexes. *Chem. Sci.* **2021**, *12*, 8131–8140.

- [53] Weaver, J. B.; Lin, C.-Y.; Faries, K. M.; Mathews, I. I.; Russi, S.; Holten, D.; Kirmaier, C.; Boxer, S. G. Photosynthetic reaction center variants made via genetic code expansion show Tyr at M210 tunes the initial electron transfer mechanism. 2021. ChemRxiv. <https://doi.org/10.33774/chemrxiv-2021-jfpj4-v2> (accessed Nov 20, 2021).
- [54] Mitchell, A. J.; Zhu, Q.; Maggiolo, A. O.; Ananth, N. R.; Hillwig, M. L.; Liu, X.; Boal, A. K. Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **2016**, *12*, 636–640.
- [55] Neugebauer, M. E.; Sumida, K. H.; Pelton, J. G.; McMurry, J. L.; Marchand, J. A.; Chang, M. C. Y. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **2019**, *15*, 1009–1016.



For Table of Contents Only